Adding Up the Evidence: Systematic Reviews and Meta-Analyses

Marlies Noordzij^a Carmine Zoccali^c Friedo W. Dekker^b Kitty J. Jager^a

^aERA-EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, and ^bDepartment of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands; ^cCNR-IBIM, Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Renal and Transplantation Unit, Ospedali Riuniti, Reggio Calabria, Italy

Key Words

Systematic review • Meta-analysis • Epidemiology • Statistics • Nephrology

Abstract

A systematic review aims to synthesize all available individual studies on a certain topic and uses explicit and reproducible methods for searching the literature, while a meta-analysis is a mathematical synthesis of the results of these individual studies. Because of the explosion of information in the scientific literature, these study designs can be useful tools to summarize the knowledge on a particular subject. In addition, combining individual studies in a meta-analysis increases statistical power, resulting in more precise effect estimates. Even though all parts of the specific methodology of systematic reviews include steps to minimize bias, both investigators and readers should be aware of potential biases like poor study quality and heterogeneity between studies. This paper explains how systematic reviews and meta-analyses should be performed and how to critically appraise them, based on an example from the nephrology literature.

Copyright © 2011 S. Karger AG, Basel

Introduction

In recent times, systematic reviews and meta-analyses have gained tremendous popularity in the medical literature. Also in the field of nephrology these study designs are published more frequently as is shown in figure 1. Why is this? To put it simply, both systematic reviews and meta-analyses are papers summarizing other papers. Systematic reviews aim to provide a systematic overview of study results on a certain subject by including all available information from similar individual studies. They are a tool to communicate otherwise unmanageable quantities of research and may be helpful for clinical decisions, and as the foundation for evidence-based practice guidelines, economic evaluations, and future research agendas. So, instead of reading a huge amount of research papers, often reporting conflicting results, clinicians could read one systematic review to get a quick and clear overview of the current state of the art on a certain topic.

Many people use the terms systematic review and meta-analysis interchangeably, but this is incorrect. Strictly speaking, a meta-analysis is a possible component of a systematic review, a specific statistical strategy for summarizing the results of several individual studies in one

single effect estimate. In the current paper, we explain the reasons for undertaking systematic reviews and meta-analyses and the basic principles of both approaches based on an example from the nephrology literature. We also discuss how to critically appraise papers reporting these studies.

An Example

Preeclampsia is a systemic disorder that occurs only during pregnancy, and is defined as the development of proteinuria and hypertension after 20 weeks of gestation. It is a disease unique to humans and affects 5–8% of all pregnancies, which makes preeclampsia the most common medical complication during pregnancy [1]. Recently, McDonald et al. [2] noticed that although preeclampsia occurs frequently, there was uncertainty about the occurrence and natural history of subsequent kidney disease. They therefore decided to undertake a systematic review and meta-analysis to determine whether women with a history of preeclampsia are at risk of subsequent kidney disease. Many systematic reviews include only randomized controlled trials (RCTs). However, because it was impossible to perform an RCT to answer their research question, these investigators chose to systematically search the literature for cohort studies and case-control studies examining kidney outcomes including microalbuminuria, proteinuria, serum creatinine level and glomerular filtration rate, in women who had had preeclampsia during their pregnancy, compared to women with unaffected pregnancies. How should they get started? In the next paragraphs we will explain which strategy to follow when performing a systematic review of the literature.

How to Perform a Systematic Review

There are different ways of reviewing the literature on a certain topic. The first possibility is to perform a review that is not systematic; these reviews are called narrative reviews or traditional reviews. Narrative reviews usually give a comprehensive overview of a topic, rather than addressing a specific question such as how effective a treatment is for a particular condition. Conclusions from different primary studies are often drawn into one overall conclusion supplemented by the reviewer's own experience and existing theories. This makes narrative reviews subjective and therefore prone to bias and error. Narrative reviews do not include a meta-analysis and do not

Fig. 1. The number of published meta-analyses between January 1, 1990 and December 31, 2009, found on MEDLINE via PubMed using the search term 'kidney'.

often report on how the search for literature was carried out or how it was decided which studies were relevant to include.

For these reasons, performing a systematic review is usually a better choice. In order to reduce bias a formal, rigorous methodology for systematic reviews has been developed. This predefined and explicit methodology includes steps to minimize bias in all parts of the process: identifying relevant studies, selecting them for inclusion, and collecting and combining their data. The method consists of 6 or 7 steps, which are described below and summarized in table 1.

Step 1: Defining a Research Question

The first step in performing a systematic review is defining a clear and specific research question. To be sure that the question is sufficiently concise, there are some models available that can be applied. For example, the PICO (Patients/Population, Intervention, Control/Comparison, and Outcomes) model is used in evidence-based medicine and can be applied to formulate a clinical question specifying exactly which treatments will be compared in what type of subjects and which outcome(s) will be studied. The PICO model is especially eligible for RCTs and for some research questions it is therefore not possible to adapt the PICO headings. In these cases, an alternative model such as ECLIPSE may be more appropriate [3].

Step 2: Literature Search

The second step is to perform a comprehensive search of the literature based on the research question defined

Table 1. Summary of the steps in performing a systematic review

Step	Activity	Tools
1	Define an appropriate research question	the PICO (Patient, Intervention, Control, Outcome) model or alternative models, such as ECLIPSE, are models for formulating clinical research questions
2	Literature search	literature can be found via databases, such as MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL, via the Cochrane Library)
3	Study selection	relevant articles should be selected based on predefined inclusion and exclusion criteria; a flow diagram can be used to summarize this process
4	Assessment of (methodological) quality of included studies	<ul style="list-style-type: none">– by 2 persons independently– several checklists are available
5	Data extraction	<ul style="list-style-type: none">– by 2 persons independently– a data abstraction form can be created for this purpose– an overview of the extracted data should be presented in a table in the Results section
6	Meta-analysis (optional)	<ul style="list-style-type: none">– statistical pooling can be performed using software packages such as Review Manager (RevMan), SAS and S-Plus– results can be presented in a forest plot
7	Formulate conclusions to answer the research question	in the case of data which prove to be insufficient or of poor quality, recommendations for future research can be made

in step 1. It is important to find all potentially relevant studies and, where possible, also unpublished data via trial registers. This is especially relevant because incomplete searching of the literature may particularly omit studies with smaller, negative, or null effect sizes. It is important that the published studies that are found constitute an unbiased sample of all studies performed on the subject. However, in general studies reporting statistically significant and positive effects of a treatment are more likely to be submitted and published than work with nonsignificant or negative results. This phenomenon is called publication bias [4].

Literature can be found in several databases, such as MEDLINE, EMBASE, CINAHL and the Cochrane Central Register of Controlled Trials (CENTRAL, via the Cochrane Library). MEDLINE is freely available on the Internet and searchable via PubMed. A useful feature of MEDLINE is that the records are indexed with Medical Subject Headings (MeSH), which provides a consistent way to retrieve information that may use different terminology for the same concepts [5]. Unfortunately, searching only PubMed may be insufficient, especially in search strategies related to drug treatment. It can be helpful to consult an experienced (medical) librarian when defining a comprehensive search strategy.

Step 3: Study Selection

In most cases, the comprehensive literature search performed in step 2 yields a large amount of articles that can potentially answer the research question. However, although the search strategy is specified in detail, usually a considerable part of the studies that were found initially will not fulfill the inclusion criteria. The next step is therefore to select studies that are eligible for inclusion based on predefined inclusion and exclusion criteria. It is important that the decision on whether to include a study or not, is made regardless of its results or quality. The applied search strategy and its yield should be reported in detail in the systematic review to guarantee reproducibility. A convenient and clear method to do so is including a flow diagram providing information about the number of studies identified, included, and excluded and the reasons for exclusion. For instance, in the aforementioned study on preeclampsia and subsequent kidney disease, the authors identified 1,243 nonduplicate studies after searching MEDLINE and EMBASE. From these 1,243 articles, 116 were potentially eligible and were selected to undergo full-text article review. After this review process, only 7 cohort studies remained for inclusion in the systematic review [2].

Step 4: Quality Assessment

Now the (methodological) quality of all included studies should be assessed, preferably by two persons independently of each other to avoid errors and to warrant objectivity. If the two assessors disagree, they need to resolve discrepancies by discussion or by consulting a third, independent reviewer. Depending on the study design of the included studies, different items can be chosen to assess their quality. Domains that are considered to be indicative of the quality of RCTs are allocation concealment, i.e. the treatment to be allocated is not known before the patient is entered into the study, the blinding of the patients and investigators, the proportion of patients that completed the study (study attrition), and whether the analyses were performed according to the intention-to-treat principle. Several checklists for the appraisal of study quality are available [6].

Step 5: Data Extraction

Subsequently, data need to be extracted from the included articles. The best way to do this is by using a data extraction form that has been developed before the data extraction starts and has been piloted by extracting data from a few included articles. Like the quality assessment, the data extraction should be performed by two persons independently because results from the individual studies could be interpreted differently. In the results section of the systematic review, a table summarizing the extracted data should be presented.

Step 6: Meta-Analysis (Optional)

This sixth step in the systematic review process is optional, because a meta-analysis can only be included if the data from the individual studies are similar (homogeneous) enough. In the next section we will explain how to perform a meta-analysis.

Step 7: Answering the Research Question

The final step is to answer the research question based on the results found and to draw conclusions. In addition, recommendations for future research can be made. For example, in the final conclusion of their systematic review and meta-analysis, McDonald et al. [2] answer their research question: women who have had preeclampsia are indeed at substantial risk of subsequent microalbuminuria and proteinuria. They also conclude that the high number of pregnancies affected by preeclampsia yields a large number of women with a propensity for the development of kidney disease [2].

How to Perform a Meta-Analysis

Principles of Meta-Analysis

Many systematic reviews summarize both qualitative and quantitative results. If the included individual studies are sufficiently similar, a meta-analysis could be performed as part of the systematic review (step 6). A meta-analysis represents a summary of the quantitative results in the individual studies, which is also called statistical pooling.

Individual studies may find inconclusive results due to relatively small sample size. By combining a number of individual studies in a meta-analysis, the study power can be increased substantially, resulting in effect estimates that are more precise. In general, weight factors are applied in the statistical pooling [7]. These weight factors are based on sample size and the number of events in the included studies and as a result, large studies and those with a lot of events get more weight than smaller studies or studies with fewer events. This means that if a very large study with a strong and significant effect has been published, combining its results with those of small studies may be of little use, since the pooled effect estimate will be almost entirely driven by the effect of the largest study. In these cases, a meta-analysis adds neither information nor strength to the results of a properly performed, powerful single study. Some investigators even 'filter' studies to be included in their meta-analysis based on the number of studied patients. On the one hand, focusing on larger studies can be an approach to reduce publication bias, because small studies are more susceptible to it. On the other hand, removing small studies from a meta-analysis diminishes the strength of a systematic literature search.

Each well-performed meta-analysis should be preceded by a systematic review of the literature as described above. It is also possible to perform statistical pooling without a systematic approach, simply by combining the results from more than one individual study. However, although such a combined analysis will provide a more precise effect estimate than an analysis of any one of the included individual studies, it is prone to biases that arise from the nonsystematic study selection process and therefore may produce a misleading result.

Available Software

Meta-analyses can be performed using several statistical software packages such as SAS and S-Plus. In the literature, manuals can be found explaining how these software programs can be applied [8]. In addition, the soft-

Fig. 2. Example of a forest plot presenting the risk of albuminuria as dichotomous variable (**a**) and as continuous outcome (**b**) (in mg/24 h) at follow-up in women with a history of preeclampsia compared to women without preeclampsia, adapted with permis-

sion from the study by McDonald et al. [2]. Effects in individual trials are depicted as squares with 95% CIs; sizes of the squares indicate the weights of each study in the analysis. Effects of meta-analyses are depicted as diamonds. f/u = Follow-up.

ware package Review Manager (RevMan) that can be obtained at no cost via the Cochrane Collaboration can be used for statistical pooling [9].

Forest Plots

The results of a meta-analysis can be presented in a forest plot. A forest plot is a graphical display designed to illustrate the relative strength of treatment effects in multiple quantitative scientific studies addressing the same question. It was developed for use in medical research as a means of graphically representing a meta-analysis of the results of RCTs. These plots usually present the effect estimates of all individual studies as squares with 95% con-

fidence intervals (CIs) and the pooled effect estimate of all studies combined as a diamond with 95% CI. An example of a forest plot from the study by McDonald et al. [2] on kidney disease after preeclampsia is shown in figure 2.

Critical Appraisal of a Systematic Review

Especially for clinicians, reading a systematic review instead of a large pile of original articles saves a lot of precious time. When reading a systematic review it is, however, important to critically appraise the quality of the review: Does it include all important original articles?

Has the quality of included studies been assessed? Are the subjects and interventions in the included studies sufficiently similar to be combined? These and many more questions can be asked when reading a systematic review. Below we discuss the most important issues to keep in mind when critically appraising such a paper [10].

Reporting

For the sake of reproducibility it is essential to report systematic reviews and meta-analyses in a clear and transparent manner. To improve the quality of reporting meta-analyses, particularly of RCTs, the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA, formerly QUORUM) statement was published in 2009 [11]. This statement includes a checklist describing the preferred way to present the abstract, introduction, methods, results, and discussion sections of a report of a meta-analysis. Recommendations on the reporting of meta-analyses of observational studies are published in a similar proposal (MOOSE) [12]. There is also a tool available to judge the methodological quality of systematic reviews, which is called the AMSTAR measurement tool [13].

Study Quality

As mentioned before, the quality of individual studies included in a systematic review can be assessed using a checklist for study quality. However, sometimes essential details on study quality are not reported or difficult to ascertain based on the original publication and in these cases it is very difficult or even impossible to judge the study quality. For that reason, the Consolidated Standards of Reporting Trials (CONSORT) statement and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement were introduced in 2001 and 2007, respectively [7, 14]. The CONSORT statement helps authors to improve the reporting of an RCT, thereby enabling readers to understand a trial's conduct and to assess the validity of its results, while the STROBE statement provides assistance in the reporting of observational studies. Both statements can also be helpful to find items indicative of study quality and may be worthwhile to be included on a checklist for study quality.

All studies that meet the inclusion criteria, also those with a poor quality, should be included in a systematic review. However, in meta-analyses the inclusion of reports of low-quality studies may be seriously misleading because they are at risk of bias and one may consider ruling out those low-quality studies from the analysis. An alternative approach is to report results of the meta-analysis stratified for study quality.

Heterogeneity

Another potential problem in a systematic review is heterogeneity, which may also be a consequence of including poor-quality studies. To be able to combine the results of multiple individual studies, the studies need to be sufficiently similar (homogeneous). If studies are clinically or methodologically too diverse, for example studies with different types of participants, follow-up, or treatment modes, the results of a meta-analysis may be meaningless. Besides this problem of clinical heterogeneity, there may also be variability in the effects evaluated in the individual studies, which is called statistical heterogeneity.

A statistical test for heterogeneity tells whether such variability is due to genuine differences underlying the results of the studies (heterogeneity), or whether the variation in findings is due to chance alone. The test for heterogeneity is called Cochran's Q, which is similar to a χ^2 test for which the p value can be interpreted ($p < 0.05$ indicates the presence of heterogeneity). Another test for heterogeneity that is commonly applied is the I^2 statistic. The I^2 represents the percentage of total variation across studies caused by heterogeneity rather than chance and an advantage of I^2 is that it can be calculated and compared across meta-analyses with different sizes and types [15]. An I^2 of 0% indicates no observed heterogeneity and larger values show increasing heterogeneity. A strict categorization of I^2 is not available and would not be appropriate for all circumstances, but I^2 values of 25, 50 and 75% are often termed low, moderate and high heterogeneity, respectively [15].

In the aforementioned study by McDonald et al. [2], heterogeneity ranged from 0% (for the outcome measure glomerular filtration rate) to 30% (for microalbuminuria as dichotomous outcome, fig. 2a) and 86% (for microalbuminuria as continuous outcome, fig. 2b). The authors attempted to explain the heterogeneity observed by exploring differences in methodology between the included studies, such as duration and completeness of follow-up and the source of kidney outcomes [2]. Moreover, they applied a random-effects model to perform their meta-analysis. In most cases, including the meta-analysis by McDonald et al. [2], random-effects modeling is the most appropriate approach to data analysis [8, 16]. Yet in certain (very selected) situations, fixed-effects modeling may be more appropriate.

Updating

Published systematic reviews may become obsolete. In the years after the publication of a systematic review new

and important primary studies may be published. Shojania et al. [17] showed that 23% of a sample of 100 meta-analyses were out of date 2 years after their publication, and 15% 1 year after their publication. This study emphasizes that it is important to realize this when reading a systematic review that was not very recently performed. In these cases we recommend to perform a simple literature search to retrieve important original papers published on the subject in the years after the publication of the systematic review.

Conclusions

In conclusion, systematic reviews and meta-analyses are useful tools to increase statistical power and to summarize the current 'state of the art' of a certain topic. Although the explicit and reproducible methodology of systematic reviews includes steps to minimize bias in all parts of the process, when reading a systematic review one should be aware of potential biases such as poor quality of included studies and heterogeneity between studies.

References

- 1 Ray JG, Vermeulen MJ, Schull MJ, Redelmeier DA: Cardiovascular health after maternal placental syndromes (CHAMPS): population-based retrospective cohort study. *Lancet* 2005;366:1797–1803.
- 2 McDonald SD, Han Z, Walsh MW, Gerstein HC, Devereaux PJ: Kidney disease after preeclampsia: a systematic review and meta-analysis. *Am J Kidney Dis* 2010;55:1026–1039.
- 3 Wildridge V, Bell L: How CLIP became ECLIPSE: a mnemonic to assist in searching for health policy/management information. *Health Info Libr J* 2002;19:113–115.
- 4 Tripepi G, Jager KJ, Dekker FW, Wanner C, Zoccali C: Bias in clinical research. *Kidney Int* 2008;73:148–153.
- 5 <http://medline.cos.com/>.
- 6 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S: Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16:62–73.
- 7 Moher D, Schulz KF, Altman D: The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials 2001. *Explore (NY)* 2005;1:40–45.
- 8 DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–188.
- 9 <http://www.cochrane.org/>.
- 10 Palmer SC, Navaneethan SD, Strippoli GF: How to read a nephrology systematic review. *Nephrology (Carlton)* 2010;15:158–164.
- 11 Moher D, Liberati A, Tetzlaff J, Altman DG: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
- 12 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;283:2008–2012.
- 13 Shea BJ, Grimshaw JM, Wells GA, et al: Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
- 14 Moher D, Schulz KF, Altman D: The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987–1991.
- 15 Higgins JP, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–560.
- 16 Lau J, Ioannidis JP, Schmid CH: Summing up evidence: one answer is not always enough. *Lancet* 1998;351:123–127.
- 17 Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D: How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007;147:224–233.