

Elsevier required licence: © <2017>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

# The art of systematic reviews

Author: Arianne Verhagen

## ***Aim***

The general aim of a systematic review is to get an overview of the existing evidence/literature from all relevant studies on a specific topic (i.e. review question). Given the increasing amount of research information it becomes almost impossible for clinicians to keep up with the published evidence. Also synthesising and critically appraise all the evidence becomes impossible for clinicians. Therefore systematic reviews aim to help e.g. clinicians (or patients, decision makers etc) to provide them with a clear overview of the existing knowledge.

## ***History***

Systematic synthesis of research findings and critical appraisal of the literature in a systematic way was first mentioned in 1975 as a 'meta analysis' [Glass and Smith 1979]. Next, the term 'systematic review' appeared around 1990. The origin of the term can be attributed to a meeting organised by the British Medical Journal and the Cochrane Centre in London in 1993 [Chalmers and Altman, 1995]. Here the members of this meeting expressed their concern that reviews on clinical treatments lack the use of 'scientific principles' and could therefore be potentially biased or even harmful [Chalmers and Altman, 1995, Boel 2011].

Already in 1972 Archie Cochrane wrote a book in which he urged health care practitioners to practice evidence-based medicine [Cochrane 1972]. The term evidence based health care was later defined by David Sackett as 'the conscientious, explicit, judicious use of current best evidence in making decisions about the care of individual patients.' [Sackett 1996]. Since then the term 'systematic review' has gained momentum in being adopted for a state of the art literature review. Systematic reviews of randomized controlled trials became key in the practice of evidence-based medicine [Chalmers 1995]. Today systematic reviews and meta-analysis (a systematic review where the review authors use statistical analysis to combine the results of multiple studies into one effect estimate) are an essential element in professional guidelines and guides health care decisions.

## ***Cochrane Collaboration***

In 1992 the Cochrane Collaboration opened its first center in Oxford. Since the launch of the Cochrane Collaboration the Cochrane systematic review set the standards of review methodology and still many innovative methodologists update the rigorous methodology and statistical analysis to ensure high quality systematic reviews and meta-analysis. Today the Cochrane Collaboration is an international network of over 37,000 researchers, academics, practitioners and consumers (patients) in over 130 countries, 14 centers, 53 review groups and 14 methods groups, all committed to the principles of managing healthcare knowledge in such a way that it is quality assured, accessible, and cumulative [Cochrane Collaboration].

### ***Systematic review as study design.***

As the label indicates a systematic review follows a systematic and structured approach in a way that it ensures an objective and transparent research synthesis, with the aim of minimizing bias. The formulation of the research question is essential and can comprise treatment effectiveness, prognosis, etiology, diagnostic strategies etc. A typical formulation of the research question of effectiveness/efficacy studies is called the PICO: P for patients/participants, I for the intervention under study, C for control intervention as a comparison, O for outcome. Within the Cochrane these outcomes are always patient-related or patient-reported outcome measures. An example of a PICO is: “Effectiveness (systematic review of (randomized) clinical trials) of spinal manipulative therapy (I) compared to all kinds of control interventions (C) for acute low back pain (P) on pain and functioning (O) [Rubinstein 2012]. For other research designs the PICO is slightly different, for instance for diagnostic accuracy reviews the P is for patients, the I for index test, the C for comparison test or reference test and O for outcome (or diagnosis). An example of such a PICO is: “Diagnostic accuracy of physical examination (I) compared to imaging or surgery (C) in low back pain patients (P) suspected of lumbar disc herniation (O)” [van der Windt 2010]. The research question guides a thorough search for all relevant literature. Restriction for languages or dates of publications is unusual as it conflicts with the general aim “to get an overview of the existing evidence”. Study selection, assessment of the risk of bias (or methodological quality) of the included studies, the data extraction and the analysis all follow a strict protocol ensuring optimal transparency and a minimal risk of bias. Frequently systematic reviews use statistical techniques (statistical pooling) to combine the results of multiple studies, which provide one, more precise, estimate, and than it is called a meta-analysis. When a statistical pooling cannot be done due to e.g. clinical heterogeneity of the studies (not combining apples and oranges) or the high risk of bias in the individual studies, some systematic reviews use a scoring system (levels of evidence) to come to a conclusion instead of a statistical pooling. Recently this statistical pooling is followed by a system that determines what the quality of the evidence is, meaning how much do we trust that the effect estimate is true. The system to do so is called GRADE Grading of Recommendations Assessment, Development, and Evaluation (GRADE) [Gyatt 2008].

Systematic reviewing is now a respectable research design, comparable to other research designs such as randomized clinical trials or (prospective) cohort studies. First, there is a registry where authors can register their protocol of a systematic review (PROSPERO). The aim of this registry is comparable with the aim of e.g. registries of randomized clinical trials, namely to avoid duplication, to ensure that reviews are executed as planned, prevent outcome reporting bias and that protocol deviations are discussed. Next, the quality of the systematic review itself can be assessed with quality appraisal instruments (Amstar, Robis), comparable with all quality appraisal instruments for all kinds of other research designs [Shea 2009; Whiting 2016]. Finally, reporting guidelines (prisma) ensure that systematic reviews are reported as best as possible comparable with reporting guidelines for all kinds of other study designs [www.equator.com].

### ***Strengths***

There are several benefits of systematic reviews. First of all it provides an overview of the existing evidence. This implies that researchers ideally should only design new studies (and find money to execute these new studies) after evaluating the body of knowledge via systematic reviews. Two of the founding fathers of evidence based medicine (Ian Chalmers and Paul Glasziou) therefore posted the term 'research waste' where the authors argue that research (and money) waste could be avoided if all research was preceded by a systematic assessment of the existing evidence [Chalmers 2009, 2014]. In case the evidence is clear, new studies should not be executed. This is still not common practice today.

Secondly, decision makers or health care insurance companies frequently use systematic reviews to influence current practice. Some interventions get easily implemented, for instance a Cochrane review that showed that act as usual was as effective as bed rest for patients with a suspected lumbar disc herniation [Hagen 2000]. The implementation of this change in interventions from bed rest to act as usual or exercise happened rather rapidly as apparently patients, care givers and decision makers liked it. On the other hand systematic reviews showing that diagnostic imaging in musculoskeletal disorders, more specifically low back pain, does not lead to better patient outcomes [Ghou 2009; Karel 2015]. Due to various factors, such as patient preferences, difficulty in changing doctors behavior, costs (earnings) this finding does not get easily implemented by itself. This is because not performing diagnostic imaging actually involves de-implementation of an intervention in which patients and doctors have high trusts, although there are effective de-implementation strategies for this intervention [Jenkins 2015].

Lastly, publishing a systematic review can boost careers. It can help journals to improve their impact factor as high quality systematic reviews are frequently cited. The same holds for the career of a scientist, as their careers depend on publications in high impact journals and frequent citations of their publication that increase their H-index. In conclusion, performing and publishing systematic reviews is sexy.

### ***Weaknesses***

To be of value systematic reviews need to be as up to date as possible. This increases the risk of duplication, which are potentially wasted research efforts. Today for the majority of topics in musculoskeletal disorders there is more than one systematic review available. For instance, in the field of low back pain various systematic reviews exist evaluating the effectiveness of spinal manipulations. Frequently the number of studies included in these reviews varies and so does the conclusion. This conflicts with the primary aim, as it does not help clinicians and patients in deciding about clinical management.

Scientists sometimes think a systematic review is a rather easy to do research design, therefore frequently systematic reviews lack methodological rigor and quality. For instance, during the development of a neck pain guideline for physiotherapists we found four systematic reviews on dry needling [Bier 2017]. All four systematic reviews suffered from severe methodological problems, as well as their conclusions in the abstract were much more positive for dry needling as the actual conclusions in the main text of the publication. Therefore we searched for relevant trials within these systematic reviews that answered our study question in the guideline.

Already in 2007 a study found that of 100 systematic reviews monitored, 7% needed updating at the time of publication, another 4% within a year, and another 11% within 2 years; this figure was higher in rapidly changing fields of medicine, especially cardiovascular medicine [Shojania 2007].

## Referenties

- Cochrane AL (1972) *Effectiveness and efficiency: random reflections on health services*. London: Royal Society of Medicine Press.
- Glass GV, Smith ML (1979) Meta-analysis of research on the relationship of class-size and achievement. *Educational Evaluation and Policy Analysis* 1: 2-16.
- Oakley A et al (2005) The politics of evidence and methodology: lessons from the EPPI-Centre. *Evidence & Policy* 1 (1): 5-31.
- Sackett DL et al (1996) Evidence based medicine: what it is and what it isn't. *British Medical Journal* 312: 71-2.
- Smith M, Glass G, Miller T (1980) *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.
- **Boell, Sebastian K; Cecez-Kecmanovic, Dubravka (2011). Are Systematic Reviews better, less biased and of higher quality?. European Conference on Information Systems 2011 (ECIS 2011). Paper 223. [online] <http://aisel.aisnet.org/ecis2011/223>**
- 26 Moher, David; Tetzlaff, Jennifer; Tricco, Andrea C.; Sampson, Margaret; Altman, Douglas G. (2007). "Epidemiology and Reporting Characteristics of Systematic Reviews". *PLoS Medicine*. 4 (3): e78
- Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. "How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis". *Annals of Internal Medicine*. 2007;147 (4): 224-33.
- Savoie I, Helmer D, Green CJ, Kazanjian A. "Beyond Medline: reducing bias through extended systematic review search". *International Journal of Technology Assessment in Health Care*. 2003 19 (1): 168-78.
- Roberts, I; Ker, K; Edwards, P; Beecher, D; Manno, D; Sydenham, E (3 June 2015). "The knowledge system underpinning healthcare is not fit for purpose and must change.". *BMJ (Clinical research ed.)*. 350: h2463. doi:10.1136/bmj.h2463. PMID 26041754.
- Altman, DG (29 January 1994). "The scandal of poor medical research.". *BMJ (Clinical research ed.)*. 308 (6924): 283-4. doi:10.1136/bmj.308.6924.283. PMID 8124111.
- Shapiro, S (1 November 1994). "Meta-analysis/Shmeta-analysis.". *American Journal of Epidemiology*. 140 (9): 771-8. PMID 7977286.
- Page, MJ; McKenzie, JE; Kirkham, J; Dwan, K; Kramer, S; Green, S; Forbes, A (Oct 1, 2014). "Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions.". *The Cochrane database of systematic reviews*. 10: MR000035. doi:10.1002/14651858.MR000035.pub2. PMID 25271098.
- **Jump up** ^ "Retraction Of Scientific Papers For Fraud Or Bias Is Just The Tip Of The Iceberg". *IFL Science!*. Retrieved 29 June 2015.
- **Jump up** ^ "Retraction and republication for Lancet Resp Med tracheostomy paper". *Retraction Watch*. Retrieved 29 June 2015.
- **Jump up** ^ "BioMed Central retracting 43 papers for fake peer review". *Retraction Watch*.
- ^ **Jump up to:** <sup>a</sup> <sup>b</sup> Hagen-Zanker, Jessica; Duvendack, Maren; Mallett, Richard; Slater, Rachel; Carpenter, Samuel; Tromme, Mathieu (January 2012). "Making systematic reviews work for international development research". Overseas Development Institute.
- 
- Greenhalgh, T. and Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *British Medical Journal*. 331(7524). 1064-1065
- Cochrane Collaboration: [www.cochrane.org](http://www.cochrane.org); last assessed Nov 2016.
- Chalmers I, Bracken M, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014; 383: 156-65.
- Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ* 2013; 346: f2304.
- Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010; 7: e1000326.
- Jones A, Conroy E, Williamson P, Clarke M, Gamble C. The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. *BMC Med Res Methodol* 2013; 13: 50.
- Egger M, Jüni P, Bartlett C, Hoenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol*

Assess 2003; 7: 1–76.

Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009; **374**: 86–89.

- Guyatt GH, Oxman AD, Vist GE et al (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336:924-926
- Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol.* 2009 Oct;62(10):1013-20.
- Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R; ROBIS group. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol.* 2016 Jan;69:225-34.
- Hagen KB, Hilde G, Jamtvedt G, Winnem M. Bed rest for acute low back pain and sciatica. *Cochrane Database Syst Rev.* 2000;(2):CD001254.
- Jenkins HJ, Hancock MJ, French SD, Maher CG, Engel RM, Magnussen JS. Effectiveness of interventions designed to reduce the use of imaging for low-back pain: a systematic review. *CMAJ.* 2015 Apr 7;187(6):401-8.
- Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet.* 2009 Feb 7;373(9662):463-72.
- Karel YH, Verkerk K, Endenburg S, Metselaar S, Verhagen AP. Effect of routine diagnostic imaging for patients with musculoskeletal disorders: A meta-analysis. *Eur J Intern Med.* 2015 Oct;26(8):585-95.