

## **Systematic reviews: importance of assessing quality of studies in the review process and tools to be used**

**Sriyani Perera<sup>1</sup>**

### **Abstract**

Systematic Reviews (SRs) are considered as a useful tool to be adopted in the approaches of Evidence-Based Practice (EBP) in Social Sciences & Humanities, and more widely in Evidence-Based Medicine (EBM) in Healthcare. In these processes of EBM and EBP, where ‘informed decision making’ based on the ‘best evidence’ is the key factor, Systematic Reviews are found to be an important source of information. The methodology of conducting a systematic review, involves several steps where Assessing Quality of the studies that are selected to include in the review is one of the most important steps, as it is the vital element which determines the validity and the creditability of the end product or the conclusion of a systematic review. Therefore, in SRs where assessing quality is not included in the procedure, the validity of the findings of the SR will be questionable. The quality of a research study can be evaluated through various facts and measures. This paper attempts to emphasize on the importance of quality appraisal of the studies in a systematic review and identifies the tools /standards that could be used for the purpose in the SR process.

**Keywords:** Systematic Reviews, Quality appraisal of research, Assessing Quality of studies, Methodological quality, Quality appraisal tools, Risks of bias

---

<sup>1</sup>*Former Senior Assistant Librarian, Medical library, University of Peradeniya  
sriyanip88@gmail.com*

## **Introduction**

It is evident globally, that Evidence-Based Practice (EBP) in Social Sciences & Humanities and, similarly Evidence-Based Medicine (EBM) in Healthcare are gaining momentum as the best practice in the processes especially of policy making and formulation of guidelines. In both processes of EBP and EBM, the practitioners use and rely mainly on the available evidence pertaining to the context of the situation, as the baseline. It is in this context that Systematic Reviews become important in EBP and EBM, merely due to the fact that conducting a systematic review is the best option to follow, in order to finding the “best evidence” on a given situation or a topic in question. By definition, Systematic Review is a review of literature pertaining to a clearly formulated question, using **systematic and explicit methods to: identify, assess and select** studies for the review, and also to **collect, analyze and synthesize data** from those studies, aiming to **present a valid conclusion**, which eventually would be the ‘best evidence’ that can be drawn from the review, in order to help answer the question (Glasziou, Irwig , Bain, and Colditz, 2001).

The procedure of a SR involves several steps: formulation of research question based on the PICOS format (Population, Intervention, Comparison, Outcomes, Study design or setting), identifying search terms, search strategy, screening studies, assessing quality and selection of studies, extraction of data, analyzing data, synthesizing and interpretation of findings of individual studies, and finally reporting the review (Glasziou et al., 2001; Perera, 2017). For the methodological quality of a SR to be in the expected level, all these steps of the procedure should be followed while the reviewers should aim to be very systematic and transparent while maintaining the scientific rigor at each step of the SR procedure (Perera, 2017).

While all these steps are equally important in the SR process, the reviewers should pay special attention on the importance of evaluating the quality of individual studies that are included in the systematic review. However, it has been observed in some systematic review articles from various parts of the world, that authors have paid little attention to this important aspect: assessment of quality of studies. These systematic reviews lack the step of assessing quality

of studies in the procedure. Researchers have commented that only a minority of reviews assessed the methodological quality of included studies (Audige, Bhandari, Griffin and Middleton, 2004; Golder and Loke, 2006). According to a study on Reviews by Juni, Witschi, Bloch, and Egger (1999), only 40% of SRs appeared to have used some form of quality appraisal of studies. This situation implies the methodological flaws in the procedure of such Systematic Reviews.

While this article does not present a research study, one of its objectives is to make the potential reviewers / researchers in a SR, understand the importance of “assessing quality of research studies” in the SR process and to make them aware of what is all about it. In the Sri Lankan context, it is a concept which seems to be unknown than known, among the researchers, not only in the LIS field, as well as in other fields of study where SRs are considered as a very important tool to be used in many scenarios; for example, Healthcare. Further, especially for a novice researcher and even otherwise, the process of assessing quality would appear as a difficult task. This may be the reason why some reviewers have overlooked this step during the SR procedure. Therefore, another objective of this article is to provide the potential reviewers with some clues on the quality appraisal of research studies, so that it could be a directive for guidance.

‘Assessing quality of research studies’ is a topic on which lengthy discussions can be presented with technical details of a wide range of important elements pertaining to the ‘quality’ of a study. However, in this article, only a brief account on the topic will be presented with the intention of providing the researchers, an insight into this important aspect of a research study.

### **The quality of research studies**

The term ‘study quality’ seems to have different interpretations for different study designs as well as in different fields of sciences. In a Systematic Review it takes two folds: internal validity and the external validity of the research studies that are used in the review (Petticrew and Roberts, 2008). Internal validity is the extent to which a study aims to avoid the methodological /system errors (biases) in the design, conduct, analysis and reporting etc. External

validity of a study is the extent to which the findings of the study could be generalized to other settings (Petticrew and Roberts, 2008). However, there is no clear definition available for ‘quality’ of research studies, albeit, experimental evidence shows that studies with problems in the study design and execution have been subjected to criticism about the validity/quality of the findings of those studies (Higgins, Altman and Sterne, 2011). Therefore, it implies that quality of a study relates to **the extent to which the study design, conduct, analysis of data and reporting**, have been appropriate or good enough, in answering the issue in the research question (Higgins et al., 2011). Further it also relates to the degree of possible risk of bias in the study design (Centre for Reviews and Dissemination, 2009). In this context, quality appraisal of any study should consider factors such as appropriateness of the study design to the research question, avoiding possible risks of bias (example: selection bias, publication bias, measurement bias, reporting bias etc.), choice of analysis, choice of outcome measures, how to deal with confounding and generalizability of the findings.

### **Why quality appraisal of studies is important in a SR?**

Systematic reviews are carried out for a purpose, with an intention of using its conclusion as the ‘best evidence’, pertaining to a given situation/question. Therefore, one might be concerned about the reliability of the reviewed research from which the ‘best evidence’ was drawn, and also about the reasons why some articles of the studies addressing the same topic / question, are not included in the review. Reviewers need to address both these issues by explaining their judgments based on the assessments of study quality and applicability of the findings of individual studies (Glasziou et al., 2001).

Further, in an effort to find the ‘best evidence’ through a systematic review, the validity and the reliability of the conclusion of a systematic review becomes very important. The extent of the strength of the final evidence drawn from a SR is of much value when it is used for making recommendations which is the ultimate purpose of conducting a systematic review in a given situation. Therefore, the validity of data and results of individual studies included, is an essential requirement of a systematic review, which would ultimately influence the analysis and interpretation of data and making conclusions of the review.

For this to be realized, the reviewers should make sure that the **quality of the studies** included in the review have met the required / expected level. It helps make sure that the final outcome of the review is reliable, and that it is drawn from findings of good quality research studies done on the topic in question. Therefore, quality appraisal of individual studies becomes an essential component in the SR process.

### **Types / designs of research studies**

A comprehensive search on a given topic to find the relevant literature in a SR, would yield all types of research studies which are of varying designs, types, and conducted in diverse settings by different authors. Similarly, 'Literature' reveals that research publications from all fields of sciences such as social sciences, education, economics, psychology, pure and applied sciences and healthcare include findings from various types of studies with varying designs. For example:

- Systematic Reviews and Meta-analysis,
- Randomized Controlled Trials, Randomized Cross-over Trials, Cluster Randomized Trials
- Non-Randomized Controlled Trials
- Quasi-experimental studies
- Cohort studies, Prospective and Observational studies, Case-control studies,
- Cross-sectional studies , surveys
- Case reports and case series etc.

Which type of research design is appropriate for a particular study is determined by the nature of the research question. These study designs could be quantitative or qualitative in nature, depending on the outcomes and the parameters to be measured in the research study. For example, experimental studies could contain qualitative or quantitative data (Petticrew and Roberts, 2008).

## **Assessing quality of research studies**

As mentioned earlier, in the context of a Systematic Review, primary studies included in the review would be of varying designs as well as of different levels in quality with respect to methodology. Evaluating the quality of such diverse studies in SRs may or may not be a uniform process, as different methods of quality appraisal would be needed for different types / designs of studies.

Systematic Reviews are conducted adopting a protocol driven methodology. Therefore the review protocol can be considered as a guiding tool specially to avoid methodological bias and minimize errors as it defines in advance, the details of every aspect of study designs (Glasziou et al.,2001) including characteristics of study data, methods for data collection, analysis and, features to be assessed etc., regarding all the studies included in the review. Further, the method to be used for assessing quality of studies is specified in the review protocol and the reviewers are compelled to follow the protocol.

Various methods have been used for quality appraisal of research studies, depending on the field of study and the study design. The quality of research studies is considered to be prime important in Health/Medical fields specially in healthcare, and evaluating research studies has become a widespread process, so is conducting systematic reviews in healthcare. Due to this reason, in the field of health sciences, the standards and guidelines have been formulated and well-defined tools have been designed for assessing quality of studies by various systematic review centres / institutes (Jesson, Matheson, and Lacey, 2011). Reviewers in other fields of sciences, have attempted to adapt to these guidelines in research studies of their topics. These quality measurement tools have also been used to derive standards / tools for research studies in fields such as management studies, social science studies and multidisciplinary studies etc. with certain limitations (Jesson et al., 2011)). It is, therefore, worthwhile to note that, tools or checklists to make a judgment on the quality of studies are available in all fields of studies, mostly through websites and free of charge.

The important element in ‘assessing quality’ is to evaluate the methodology of the primary studies. Therefore, in a Systematic Review, it is the process of assessing the methods used, and findings of individual studies those included in

the review. This is normally done by examining pre-defined key features of the study at various stages of the procedure that were described in the review protocol. Standards for SRs emphasize that quality appraisal of studies be carried out independently by more than one reviewer so that the judgments can be compared, checked and conflicts could be resolved through consensus.

Guidelines on quality appraisal in Biomedical Sciences, use the hierarchy of study designs as a model to set standards while it is also seen applied in other fields of applied research (Jesson et al., 2011). In Medical Sciences, Randomized Control Trials (RCTs) are considered to be the study design of highest quality, therefore is known as the gold standard, whereas non-randomized, qualitative and narrative studies are considered to be the lowest in quality (Glasziou et al., 2001). However, applying this judgment in some other fields of research such as policy research and social science studies, may not be possible.

Standards for assessment of risk of bias in Randomized Controlled Trials (RCTs) have been developed taking into consideration, all the dimensions to prevent risk of bias or minimize errors, which are to be fulfilled by each study as a requirement. Similarly, for other types of Non-Randomized Studies (NRS), assessment of risk of bias is considered with much caution because of the diversity of different study design features. This is because in some situations, the quality criteria might deprive the review authors of making use of important evidence and useful recommendations from certain studies. Potential biases are appeared to be more in NRS than in RCTs therefore researchers should pay more attention specifically to possible bias due to selection of subjects for the NRS. The best and convenient method for quality appraisal is to use a validated checklist or a tool designed for the purpose, which are available through various sources.

### **Tools and checklists for quality appraisal of studies**

Various instruments have been developed by different Organizations and Institutes for use in appraisal of quality of research studies. Most of these are available in the form of checklists or scales. Identification of a ‘best tool’ that can be applied for all types of studies is not feasible or possible. Therefore,

different tools are needed to assess studies of different designs from different disciplines. Quality assessment tools have been derived from studies in both health/medical sciences (Higgins et al., 2011) and social / multidisciplinary sciences (Campbell, and Stanley, 1966). Deeks et al. (2003) identifies a range of checklists for use in systematic reviews.

While this article does not intend to present a listing of all the instruments available for assessing quality of research, following commonly used tools are presented as examples:

- The Cochrane Collaboration's tool is a widely used checklist for assessing risk of bias in randomized trials in health sciences (Higgins et al. , 2011)
- NHS CRD (National Health Services, Centre for Research and Dissemination) Report 4 for case control studies (Petticrew and Roberts, 2008)
- The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomized studies in meta-analyses (Wells et al., 2012)
- Downs and Black Scale is an extensively validated tool for randomized and non-randomized studies (Downs and Black, 1998)
- Cowley checklist to assess comparative studies (Cowley, 1995).
- Checklists from EPOC (The Cochrane Effective Practice and Organization of Care) Group for SRs of financial, educational, organizational, behavioral and policy making studies (Petticrew and Roberts, 2008).
- The Jadad Scale for Randomized Control Trials (Jadad, 1998)
- Chalmer's Checklist for RCTs (Chalmers et al.,1981)
- Cook and Campbell's checklist for non-randomized (quasi-experimental) studies (Cook, and Campbell, 1979)
- The Maryland Scientific Methods Scale (SMS) for systematic reviews from certain disciplines (Farrington, Gottfredson, Sherman, and Welsh, 2002)
- Thomas Quality Assessment Tool for Quantitative studies (Thomas, 2003)

These checklists have been designed identifying the major issues causing bias for each type of study. Main purpose is to examine the extent to which these bias elements influence the findings of the study, which is the final evidence that will be drawn from the research. Review authors could use the quality appraisal results depict through these checklist to guide the assessment on the study to determine how good or bad the study is in overall quality.

## **Conclusion**

Conclusion of a Systematic Review is based on the best/highest quality evidence that would be distilled from a large pool of research evidence. This will be possible only when the procedure of the SR specifies a standardized and valid method to select relevant articles of acceptable quality for inclusion in the review (Glasziou et al., 2001). In complying with this, it is expected in a systematic review that, after assessing a study using above mentioned checklists or scales, a summary report of the appraisal results is prepared. Regarding qualitative studies, reviewers should be able to weigh the studies against such reports of critical appraisal to decide whether the study is **high, medium or low** in quality (Petticrew and Roberts, 2008), so that it can be taken into consideration when conducting a SR. Similarly, in quantitative research, weighing studies can be done using a scoring system for each study type so that the judgment on the quality can be made, based on the total score gained by each study by applying a pre-defined threshold score. Using these checklists reviewers are allowed to assess to what extent the bias elements such as selection bias and measurement bias etc. have been avoided in research studies. Based on these quality measurements, review authors would be in a position to identify and select from a pool of studies, only the studies that meet the quality threshold, therefore of acceptable quality and relevance. The findings of such quality studies can then be included in the synthesis of conclusion in the systematic review which would be of much reliability.

## References:

- Audige L, Bhandari M, Griffin D, Middleton P, R. B. (2004). Systematic reviews of nonrandomized clinical studies in the orthopaedic literature. *Clinical Orthopaedics and Related Research*, 249–257. Retrieved from [http://handbook.cochrane.org/chapter\\_13/13\\_including\\_non\\_randomized\\_studies.htm](http://handbook.cochrane.org/chapter_13/13_including_non_randomized_studies.htm)
- Campbell, D. and Stanley, J. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Centre for Reviews and Dissemination, U. of Y. (2009). *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. Retrieved from [https://www.york.ac.uk/media/crd/Systematic\\_Reviews.pdf](https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf)
- Chalmers, T., Smith, H., Blackburn, C., Silverman, B., Schroeder, B., & Reitman, D., et al. (1981). *A method for assessing the quality of a randomized controlled trial*. *Controlled Clinical Trials*. 2, 31–49.
- Cook, T. and Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago,,: Rand McNally.
- Cowley DE. (1995). Prostheses for primary total hip replacement. A critical appraisal of the literature. *Int J Technol Assess Health Care*, 11, 770–778.
- Deeks, J. et. a. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessment*, 7(27), 101–107.
- Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, 52(6), 377–384. <https://doi.org/10.1136/jech.52.6.377>
- Farrington, D., Gottfredson, D., Sherman, L., and Welsh, B. (2002). *The Maryland Scientific Methods Scale*. In *Evidence-based crime prevention*. London: Routledge, ( and D. M. (eds. . L. Sherman, D. Farrington, B. Welsh, ed.).

Glasziou P., Irwig L., Bain C., C. G. (2001). *Systematic Reviews in Health Care : A Practical Guide*. Cambridge University Press.

Golder S, Loke Y, M. H. (2006). Room for improvement? A survey of the methods used in systematic reviews of adverse effects. *BMC Medical Research Methodology*, 3–6.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Online)*, 343(7829), 1–9. <https://doi.org/10.1136/bmj.d5928>

Higgins Julian PT Higgins, Douglas G Altman and Jonathan AC Sterne, E. (2011). Assessing risk of bias in included studies. In Julian PT Higgins and Sally Green editors (Ed.), *Cochrane Handbook for Systematic Reviews of Interventions 5.1.0*. Retrieved from <https://handbook-5-1.cochrane.org/>

Jadad, A. (1998). *Randomised controlled trials: A users guide*. London: BMJ Books.

Jesson, J.K., Matheson, L. &Lacey, F. . (2011). *Doing your literature review: traditional and systematic techniques*. London: sage publishers.

Juni, P., Witschi, A., Bloch, R., and Egger, M. (1999). . The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association*, 282, 1054–1060.

Perera, P. A. S. H. (2017). Systematic Reviews: Significance of Search Strategy. *Journal of the University Librarians Association of Sri Lanka*, 20(1), 58–72. <https://doi.org/http://doi.org/10.4038/jula.v20i1.7896>

Petticrew, M., & Roberts, H. (2008). Systematic Reviews in the Social Sciences: A Practical Guide. In *Systematic Reviews in the Social Sciences: A Practical Guide*. <https://doi.org/10.1002/9780470754887>

Thomas, H. (2003). *Quality assessment tool for quantitative studies*. *Effective*

*Public Health Practice Project*. Retrieved from  
<http://www.hamilton.ca/phcs/ephpp/Research/Tools/QualityTool2003.pdf>

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2012). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. (*Available from: URL: [Http://Www.Ohri.ca/Programs/Clinical\\_epidemiology/Oxford.Asp](Http://Www.Ohri.ca/Programs/Clinical_epidemiology/Oxford.Asp)*), 2012. <https://doi.org/10.2307/632432>