La Torre G, Backhaus I, Mannocci A./Senses Sci 2015; 2 (2):31-35
doi: 10.14616/sands-2015-2-3135

Rating for narrative reviews: concept and development of the International Narrative Systematic Assessment tool

Giuseppe La Torre[1], Insa Backhaus[1,2], Mannocci Alice[1]

[1] Department of Public Health and Infectious Diseases, Sapienza University of Rome, P.le A. Moro, 5 - 00185 Rome, Italy
[2] Maastricht University

*Corresponding author Prof. Giuseppe La Torre, Department of Public Health and Infectious Diseases, Sapienza University of Rome, Piazzale Aldo Moro 5 – 00185 Roma – Italy. E-mail: giuseppe.latorre@uniroma1.it

Article history
Received: May 31, 2015
Accepted: June 21, 2015
Published: June 30, 2015

Abstract

Background: Narrative reviews differ from systematic reviews in numerous ways. Generally, they tend to be mainly descriptive, do not involve a systematic search of the literature, and thereby often focus on a subset of studies of a certain topic chosen based on availability or author selection. These are typical review articles that can be found in most journals. The quality rating of the selected studies is another crucial point. Surprisingly, at this point of time there are no measures that judge the quality of narrative reviews. The goal of this study was to design an instrument providing an easy and convenient tool for the quality assessment of narrative reviews for systematic reviews (International Narrative Systematic assessment, INSA)
Methods: Web searches on PubMed database and Google (considering the firsts 5 pages) have been conducted in February 2015 by two independent reviewers. No restriction language restriction was applied. The tool designed is an adapted version based on the available criteria for the quality assessment of systematic reviews. The tool reviewed are: the AMSTAR tool (Shea et al, 2007) and Overview Quality Assessment Questionnaire (OQAQ) by Oxman and Guyatt (1991).
Results: Eleven epidemiologist and physicians in public health participated in the study. The study results suggest that the instrument content is sufficiently comprehensive. The reviewers expressed strong support for the instrument's content for assessing the quality of narrative reviews and are consent about the usefulness of the tool.
Conclusion: The INSA tool is a valid measure of the quality of research overviews.

Keywords: *Narrative reviews, international assessment tool, quality assessment.*

Introduction

Research studies typically fall into one of two main categories: primary research and secondary research. Secondary research, also known as desk research, involves the summary and synthesis of existing research. Primary research on the contrary collects data, for instance, from, research subjects or experiments. In particular the study designs considered in this category are: narrative reviews, systematic review and meta-analysis.

It was between the 1970s and 1980s, when systematic reviews and meta-analyses began to evolve and emerge in a variety of health fields (Fox, 2011). Since then they established to a valuable and useful source for clinicians and public health professionals. Systematic reviews, also known as systematic overviews or evidence summaries, use systematic and precise methods to recognize, select, and critically evaluate relevant studies (McGown & Sampson, 2005). They should be based on a protocol so that they can be replicated if necessary. Often, systematic reviews include a meta-analysis component, which involves using statistical techniques to synthesize the data from several studies into a single quantitative estimate or summary effect size (Petticrew & Roberts, 2006).

Narrative reviews differ from systematic reviews in numerous ways. Generally, they tend to be mainly descriptive, do not involve a systematic search of the literature, and thereby often focus on a subset of studies of a certain topic chosen based on availability or author selection (Lindsay S. 2011). Although they are more prone to selection bias and often do not even meet central criteria to help alleviate bias, these are typical review articles that can be found in most journals.

In the last decade the amount of published studies raised significantly and consequently it became more and more important to summarize the evidence of primary research. Moreover, in the course of time researchers did not only realize that performing systematic reviews are necessary but also that the quality of those matters essentially. In fact, during the last years many tools evaluating primary research studies, such as the JADAD scale for randomized controls trials (Jadad, 1996) or the Newcastle Ottawa Scale for observational studies (Ottawa Hospital Research Institute, 2014), evolved. In addition, tools for judging the quality of systematic reviews established during the last years. Sacks (1989), Oxman and Guyatt (1991) were among the first researchers to propose criteria for assessing the scientific quality of research overviews. Nowadays more than 24 tools are available, but the most updated tool, adapted in 2007, is the "AMSTAR" tool. AMSTAR stand for "assessment of multiple systematic reviews and it consists of 11 items and has good face and content validity for measuring the methodological quality of systematic reviews (Shea, 2007).

In order to perform a systematic review of reviews, sometimes it may not only be essential to include systematic reviews, but also narrative ones. The quality rating of the selected studies is a crucial point. Surprisingly, at this point of time there are no measures that judge the quality of narrative reviews. The goal of this study was to design an instrument providing an easy and convenient tool for the quality assessment of narrative reviews for systematic reviews (International Narrative Systematic Assessment tool, INSA).

Materials and Methods

Search strategy

Items were generated from a review of the literature. Web searches on Google (considering the firsts 5 pages) and PubMed database have been conducted in February 2015 by three independent reviewers and no restriction language restriction was applied. The search terms used were the following: quality appraisal tool, narrative review, validity assessment reviews, and measure of quality, scientific quality, quality literature. A second review was performed using the following key words: measure of quality, scientific quality, systematic review. The items were defined through an interactive process of discussions, pretesting and revision.

Development of the instrument

Due to an unsuccessful first search, which did not lead to any indication on whether a rating tool for narrative reviews has already been developed a new rating tool was established using the results of the second search. The tool designed is an adapted version based on the available criteria for the quality assessment of systematic reviews. The tool reviewed is:

- The AMSTAR tool (Shea et al, 2007);
- Overview Quality Assessment Questionnaire (OQAQ) by Oxman and Guyatt (1991).

It is to be noted that the criteria developed by Sacks et al. (1987) were also taken into consideration. Nonetheless, Sacks and colleagues focused mainly on the quality of meta-analyses; therefore none of the criteria seemed to be applicable for the assessment of narrative reviews. The purpose of the criteria (Tab. 1) is to assess certain aspects of the scientific quality of research overviews. Respectively, two items from OQAQ and two from AMSTAR were considered as useful.

Table 1. Criteria for assessing scientific quality of narrative reviews (INSA) International Narrative Systematic Assessment tool.

Background of the study clearly explained / state of the art
Objective is clear[a]
Description/Motivation of selection of studies[a]
Description of the characteristics of the included studies is clear in the paper[b]
Presentation of results (paragraphs, tables, synthezing of data)
Conclusion is clear
The author(s) declare(s) that there is or not conflict of interest regarding the publication of the article

[a]Oxman AD et al.[b] Shea BJ et al.

In addition, another three items to determine whether the description of the background, the results and conclusion was clear were established. Finally, this resulted in a 'point system', in which a study is judged on seven criteria. For each criterion one point can be assigned. A review awarded a total of 5 points will be considered good. Approach for assessing the validity of the criteria. We invited 14 clinical epidemiologists and public health doctors, experts in literature search and synthesis, that had no prior knowledge about the study design. A set of three narrative reviews, concerning the association between tobacco smoking and multiple sclerosis (Wingerchuk 2012; Pugliatti 2008; Jafari 2011)

Table 2. Scores derived from the own judgment (1st round) and the INSA tool (2nd round)

	Wingerchunk1	Wingerchunk2	Jafari1	Jafari2	Pugliatti1	Pugliatti2
Mean	5.36	5.64	6	5,64	4.09	3.27
Median	5	6	6	6	4	3
Std. Deviation	1.027	0.809	1.414	0.924	1.044	1.348
Minimum	4	4	2	4	3	2
Maximum	7	7	7	7	6	5

were assigned to each researcher and they were encouraged to rate the quality of the paper. A score from 1 to 7 was referred (1=low quality, 7= high quality). After that the researchers received a copy of the INSA instrument and applied this to the same articles.

Following this, the researchers were asked finally to judge the instrument according to the adapted Feinstein's criteria (Feinstein 1987, Yeung et al, 2014, Rowe & Oxman, 1993). Feinstein has suggested 13 criteria for the "sensibility" of an instrument. Each item was rated on a scale of 1 to 7, where 1 signified that the instrument was not meeting its goals and 7 signified that goals were fully met.

Statistical analysis

The Wilcoxon test was applied to evaluate the concordance/consensus between the researchers' score and INSA score. Bland Altman Plot has been chosen to analyze the difference between the first rating of the reviewers and the second rating of the reviewers. The Bland Altman plot has been proven to be useful method of data plotting used in analyzing the agreement between two different measures. In this graphical method the differences between the two techniques are plotted against the averages of the two techniques. Horizontal lines are drawn at the mean difference, and at the limits of agreement, which are defined as the mean difference plus and minus 1.96 times the standard deviation of the differences.

Moreover, the median value and the range of each Feinstein's criteria were calculated.

Results

Eleven epidemiologist and public health doctors consented and participated in the study, the majority of whom had several years of experience examining and writing systematic reviews. Our study results suggest that the instrument content is sufficiently comprehensive.

There were no significant differences between the first and the second way of judgment for the three papers reviewed) ($p = 0.426$; $p = 0.524$ and $p = 0.116$, respectively see order in Tab. 2).

The Bland Altman plots (Fig. 1) reveal that for all the three papers reviewed there are situations that go out of the limits of agreement, both over and under the limits.

Besides one reviewer, all stated that the INSA tool was useful for assessing the quality of narrative.

For most of the 13 questions used to assess sensibility the mean rating was 5 or greater, indicating general satisfaction with the instrument and the reviewers expressed strong support for the instrument's content for assessing the quality of narrative reviews. The issues dealt with in each of the 13 items are summarized in Table 3.

Discussion

The number of published studies has should be increase steadily over the past years. Narrative review have established as an important source of summarized evidence as they give a broad overview of primary literature published, relieving readers and also researcher of some of the burden of appraising the primary studies on their own. However, these can vary their methodological quality. This was the first study to generate a quality assessment tool for narrative reviews. With increasing number of studies it became more and more important to summarize the evidence of primary research and during the last years many tools evaluating primary research studies and tools for judging the quality of systematic reviews have been developed established. Based on the criteria outlined in the AMSTAR tool (Shea et al, 2007) and by Overview Quality Assessment Questionnaire (OQAQ), a quality assessment tool including 7 items has been established. Although the number of eleven participants appears to be sufficient for pilot studies examining doubts and clarity of an instrument's content, as indicated by a similar study, in which fifteen participants were recruited for, the results should be viewed with caution (Oxman & Guyatt, 1991). There was general consensus among all participants that

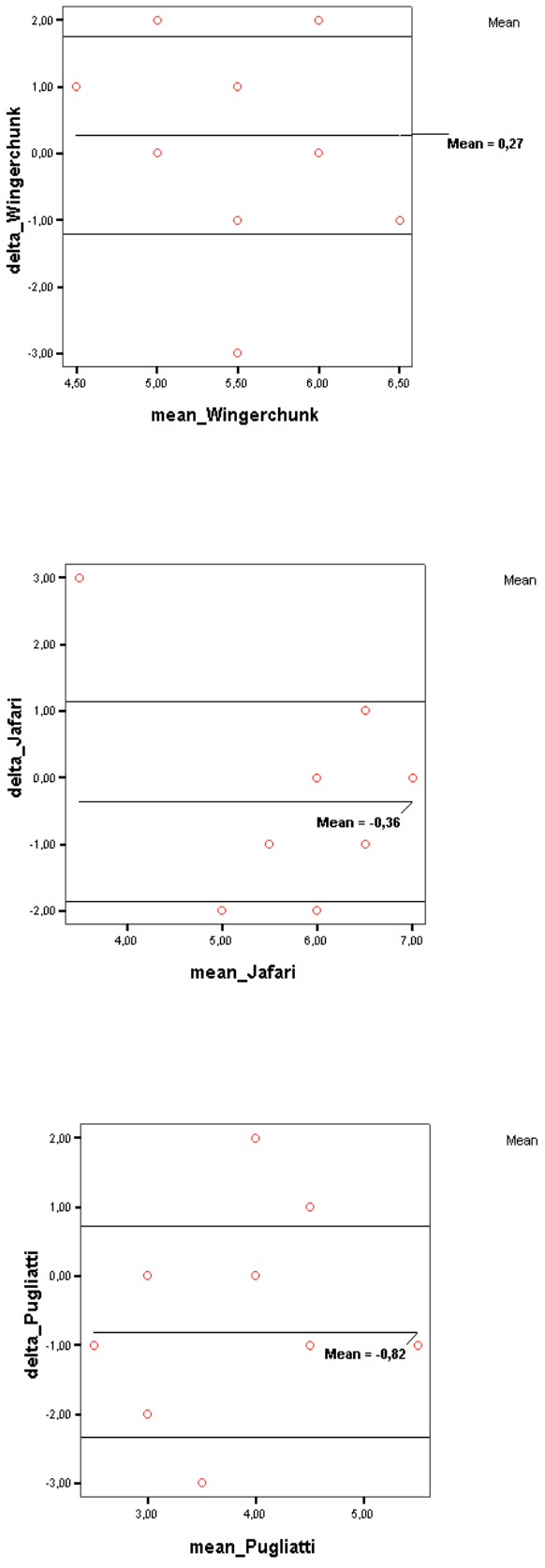

Figure 1. The Bland Altman plots concerning the judgment of the three papers.

INSA tool is useful in assessing the quality of narrative reviews. Item 3 appeared to be a rather difficult item, since it has to do with how likely it is that information needed to make the ratings will be included in a published overview, for instance, how likely is it that authors will report the selection process of studies included search, for instance, how the authors chose the articles to be included from among those, which may have not been included. Most participants stated could not assign a point to any of the reviews. This reflects the problem that most published narrative reviews currently do not report their methods.

For rating tools it is of great importance to reduce unnecessary burden placed on the user resulting from inappropriate instrument format, length and clarity. Inappropriate administrative burden demanding too much time and training might make the instrument impractical and unusable (Yueng et al., 2015). With respect to feasibility most participants stated that the instruction were clear and that the tool did not require too much effort. Acceptability is the extent to which prospective users approve the content of the instrument and the interpretation the assessment results. Without adequate evaluation of the acceptability of an instrument content and subsequent interpretation of assessment results, the construct being assessed may be over or underrepresented (Auger et al., 2006). The scoring scale, for instance, should be adequately adapted with the instrument's purpose and reflect the nature of the construct of interest and the context in which the instrument will eventually be used. Besides one participant all examiners stated that the instrument is included important items necessary for the assessment of narrative reviews and that the tool captured all elements of needed for a good narrative review. When asked about item redundancies, two participants noted that items were redundant.

However, the results from the present study must be interpreted with caution largely due to the relatively small sample size. The quantitative results of this study might not be generalizable to all reviewers who may use this tool. Hence, future work should be conducted on larger samples of raters to endorse the results of this study.

Conclusion

Based on the results of the statistical analysis the study demonstrated a general acceptance of the INSA tool for assessing the quality of narrative reviews. Finally, this study that is unique has developed the first rating tool for the quality assessment of narrative reviews. It can be used by readers and editors of journals to identify scientifically rigorous overviews and thus to judge their qualities which is important for drawing conclusions.

Table 3. Sensibility Assessment Completed by 11 reviewers.

Question Topic and Corresponding Number From Sensibility Questionnaire	Median	Range
1. The assessment tool is useful for evaluating the quality of narrative reviews	6	3 – 7
2. The instructions were clear and easy to understand	6	3 – 7
3. The assessment tool took too long to complete	1	1 – 6
4. The assessment tool required too much effort to complete	1	1 – 6
5. The overall assessment tool makes sense to me	6	5 – 7
6. I was able to assign a rating for all assessment criteria	6	5 – 7
7. The descriptions for each assessment criteria were clear and easy to understand	6	4 – 7
8. The scale (response categories) for each assessment criteria was adequate	6	5 – 7
9. I was consistently able to find appropriate ratings	6	5 – 7
10. The narrative review assessment tool was able to capture all elements of needed for a good narrative review.	6	5 – 7
11. The assessment tool included important items that are necessary for assessing the quality of narrative reviews	7	1 – 7
12. The assessment tool included items that were repetitive or redundant	1	1 – 7
13. There were assessment criteria missing from this assessment tool that should be included	1.5	1 – 6

References

1. Auger C, Demers L, Swaine B. Making sense of pragmatic criteria for the selection of geriatric rehabilitation measurement tools. Arch Gerontol Geriatr 2006; 43: 65-83
2. Fox DM. Systematic review and health policy: the influence of a project on perinatal case since 1988. Milbank Q 2011; 89(3): 425-49
3. McGowan J, Sampson M. Systematic reviews need systematic searchers. Journal of the Medical Library Association 2005; 93(1): 74-80.
4. Petticrew M, Roberts H. Systematic reviews in the social sciences: A practical guide. John Wiley & Sons, 2008.
5. Uman LS. Systematic Reviews and Meta-Analyses. J Can Acad Child Adolesc Psychiatry 2011; 20(1): 57–59.
6. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Controlled Clinical Trials. 1996; 17:1–12
7. Ohri.ca, (2015). Ottawa Hospital Research Institute. [online] Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [Accessed 22 Feb. 2015].
8. Sacks HS, Chalmers TC, Smith H Jr. Randomized versus historical assignment in controlled clinical trials. N E&l J Med 1983; 309: 1353-61.
9. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44(11): 1271-8.
10. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM.Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007; 7:10.
11. Feinstein AR. Clinimetrics. New Haven, Conn. Yale University Press; 1987: 141-66.
12. Rowe BH, Oxman AD. An assessment of the sensibility of a quality-of-life instrument. Am J Emerg Med 1993; 11:374e80.
13. Yeung E, Woods N, Dubrowski A, Hodges B, Carnahan H. Sensibility of a new instrument to assess clinical reasoning in post-graduate orthopaedic manual physical therapy education. Manual therapy 2015; 20(2): 303-12.
14. Wingerchuk DM. Smoking: effects on multiple sclerosis susceptibility and disease progression. Ther Adv Neurol Disord 2012; 5(1): 13–22.
15. Pugliatti M, Harbo HF, Holmøy T, Kampman MT, Myhr K-M, Riise T, Wolfson C. Environmental risk factors in multiple sclerosis. Acta Neurol Scand 2008: 117 (Suppl. 188): 34–40.
16. Jafari N, Hintzen RQ. The association between cigarette smoking and multiple sclerosis. Journal of the Neurological Sciences 2011; 311 : 78–85.