Palmeri V., Colamesta V., La Torre G. /Senses Sci 2016; 3 (3):235-241
doi: 10.14616/sands-2016-3-235241

Evaluation of methodological quality of studies

*Vanessa Palmeri, Vittoria Colamesta, Giuseppe La Torre

Affiliation: Department of Public Health and Infection Diseases, "Sapienza" - University of Rome

*Corresponding author: Dr. Vanessa Palmeri, Department of Public Health and Infection Diseases, Piazzale Aldo Moro, 5, Rome, Italy.

Article history
Received: September 3, 2016
Accepted: September 28, 2016
Published: September 30, 2016

Abstract

Evidence-Based Medicine (EBM) is a systematic approach to medical question, integrating individual clinical expertise with the best available clinical evidence from systematic research.
Systematic reviews are key in practice of EBM and are to the peak of the evidence pyramid. In performing a systematic review/meta-analysis and overview of systematic reviews, the evaluation of methodological quality of the included studies represents an important step to identify poor quality studies that can give a falsification of effect's esteems. The aim of the present article is to analyze the rating scales about methodological quality of observational studies, experimental studies, economic analysis, studies about diagnostic tests and systematic review.

Keywords: *Evidence-Based Medicine , methodological quality, systematic review, observational studies, clinical trial.*

Introduction

The scientific articles are classified in the "*evidence pyramid*" from the base to the peak of pyramid in: expert opinion, case report, series of cases, cross-sectional studies, case-control studies, cohort studies, consensus conference, clinical studies with control group, clinical studies with randomized control group (RCT), systematic revisions and meta-analysis [1].
Two kinds of validated checklists exist, depending on study design taken in consideration, to evaluate:

- methodological quality of study;
- reporting of study.

Evaluation of methodological quality represents an important step in performing a systematic review/meta-analysis and overview of systematic reviews. This evaluation must be conduct by at least two researchers independently by application of standardized scales. It is important to identify poor quality studies because they can give a falsification of effect's esteems.

The aim of this article is to analyze in detail the rating scales about methodological quality of observational studies, experimental studies, economic analysis, studies about diagnostic tests and systematic review while tools for evaluation of studies reporting will only mention.

Evaluation of methodological quality of observational studies

Observational studies can be realized in several settings and include different study designs, in particular: cross-sectional, case-control and cohort. The document named the Strengthening The Reporting of Observational Studies in Epidemiology (STROBE) represents a guide to standardization of observational study reporting and it consists of a checklist of 22 items considered essential for a correct description of cross-sectional, case-control and cohort study. Items included in the checklist examine several section of the article: title and abstract, introduction, methods, results, discussion and the additional data (about financial sources).

Eighteen items are shared by the three study designs, whereas four (two into methods section and two into results section) are specific for each different design of studies [2].

This checklist be born as tool for a correct reporting of observational studies and doesn't represent a tool for the evaluation of methodological quality of the same study included in a systematic review.

The evaluation of the methodological quality is essential for a correct comprehension of the observational study. For this aim the **Newcastle-Ottawa Scale (NOS)** has been developed for the cohort and case-control studies.

This scale is born by collaboration between the University of Newcastle, Australia, and University of Ottawa, Canada, with the aim of obtain an easy and useful tool for the evaluation of quality of non-randomized studies (cohort and case-control) in a systematic review for a best interpretation of the meta-analysis' results.

This scale does not assign a score but uses a star system (for a maximum of nine stars) to identify high quality studies.

The NOS includes 4 questions about quality of selection of groups in the study, 1 question about comparability of groups and 3 questions about exposure or outcome for case-control and cohort studies, respectively.

It is possible to give a maximum of one star to each question into the sections of selection and exposure/outcome and a maximum of two stars into the section of comparability [3].

In **Table 1** the questions of the NOS for the case-control studies and cohort studies are reported.

For the cross-sectional studies, instead, in a systematic review Herzog et al. have adapted the NOS for cohort study in order to realize an evaluation of methodological quality for prevalence studies.

This scale takes into consideration 7 items and, as for cohort studies, evaluates quality of selection of groups in study (4 questions), comparability of groups (1 question) and outcome of interest (2 questions). It's possible to attribute a maximum of ten stars: a maximum of five stars in the section of selection, a maximum of two stars into sectionof comparability and a maximum of three stars to the part relative to the outcome of interest [4].

In the **Table 2** the questions of NOS adapted for cross-sectional studies are reported.

Table 1: The Newcastle-Ottawa Scale (NOS) for case-control studies and for cohort studies

NEWCASTLE-OTTAWA SCALE (NOS)	
CASE-CONTROL STUDIES	**COHORT STUDIES**
Selection	Selection
1) <u>Is the case definition adequate</u>? a) yes, with independent validation * b) yes, eg record linkage or based on self-reports c) no description	1) <u>Representativeness of the exposed cohort</u> a) truly representative of the average _____ (describe) in the community * b) somewhat representative of the average _____ in the community * c) selected group of users eg nurses, volunteers d) no description of the derivation of the cohort

2) Representativeness of the cases a) consecutive or obviously representative series of b) cases * c) b) potential for selection biases or not stated	2) Selection of the non-exposed cohort a) drawn from the same community as the exposed cohort * b) drawn from a different source c) no description of the derivation of the non-exposed cohort
3) Selection of Controls a) community controls * b) hospital controls c) no description	3) Ascertainment of exposure a) secure record (eg surgical records) * b) structured interview * c) written self report d) no description
4) Definition of Controls a) no history of disease (endpoint) * b) no description of source	4) Demonstration that outcome of interest was not present at start of study a) yes * b) no
Comparability	Comparability
1) Comparability of cases and controls on the basis of the design or analysis a) study controls for _____ (Select the most important factor.) * b) study controls for any additional factor * (This criteria could be modified to indicate specific control for a second important factor.)	1) Comparability of cohorts on the basis of the design or analysis a) study controls for _____ (select the most important factor) * b) study controls for any additional factor * (This criteria could be modified to indicate specific control for a second important factor.)
Exposure	**Outcome**
1) Ascertainment of exposure a) secure record (eg surgical records) * b) structured interview where blind to case/control status * c) interview not blinded to case/control status d) written self-report or medical record only e) no description	1) Assessment of outcome a) independent blind assessment * b) record linkage * c) self-report d) no description
2) Same method of ascertainment for cases and controls a) yes * b) no	2) Was follow-up long enough for outcomes to occur a) yes (select an adequate follow up period for outcome of interest) * b) no
3) Non-Response rate a) same rate for both groups * b) non respondents described c) rate different and no designation	3) Adequacy of follow up of cohorts a) complete follow up - all subjects accounted for * b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ % (select an adequate %) follow up, or description provided of those lost) * c) follow up rate < ____% (select an adequate %) and no description of those lost d) no statement

Table 2: The Newcastle-Ottawa Scale (NOS) adapted to the cross-sectional studies

NEWCASTLE-OTTAWA SCALE (NOS) CROSS-SECTIONAL STUDIES
Selection
1) Representativeness of the sample: a) Truly representative of the average in the target population. * (all subjects or random sampling) b) Somewhat representative of the average in the target population. * (non-random sampling) c) Selected group of users. d) No description of the sampling strategy.
2) Sample size: a) Justified and satisfactory. * b) Not justified.
3) Non-respondents: a) Comparability between respondents and non-respondents characteristics is established, and the response rate is satisfactory. * b) The response rate is unsatisfactory, or the comparability between respondents and non-respondents is unsatisfactory. c) No description of the response rate or the characteristics of the responders and the non-responders.
4) Ascertainment of the exposure (risk factor): a) Validated measurement tool. ** b) Non-validated measurement tool, but the tool is available or described.* c) No description of the measurement tool.
Comparability
1) The subjects in different outcome groups are comparable, based on the study design or analysis. Confounding factors are controlled. a) The study controls for the most important factor (select one). * b) The study control for any additional factor. *
Outcome
1) Assessment of the outcome: a) Independent blind assessment. ** b) Record linkage. ** c) Self-report. * d) No description.
2) Statistical test: a) The statistical test used to analyze the data is clearly described and appropriate, and the measurement of the association is presented, including confidence intervals and the probability level (p value). * b) The statistical test is not appropriate, not described or incomplete.

Evaluation of methodological quality of experimental studies

The experimental studies are commonly known as randomized controlled trials (RCT) and are considered the gold standard for the evaluation of efficacy of medical interventions thanks to capacity of avoid or minimize biases. However, in absence of a clear and adequate reporting, it's difficult to evaluate the reliability and soundness of trial results and to find necessary data to realize a systematic review. Moreover incongruous reporting and design can bring to a distortion of efficacy's estimate of analyzed treatment. With the aim of improve the quality of RCT's reporting the CONSORT statement (*Consolidated Standards of Reporting Trials*) has been published, in 1996, and recently reviewed. The CONSORT Statement is a guideline for the reporting of every trials and includes a checklist of 25 items and a flow-chart. The items of the checklist represent fundamental points that have to be reported in an RCT while the flow-chart provides data about participants in

different phases of trial (recruitment, allocation, follow-up, analysis) [5]. The CONSORT Statement directs just the RCT's reporting, does not include recommendations about planning, conduction and analysis of a trial and is not a tool for evaluation of quality.

A commonly used and easy-to-use tool for the evaluation of methodological quality of RCT is the **Jadad scale**. The Jadad scale is composed by five questions and analyzes the randomization, double-blind and lost to follow-up. It provides a total score that can range from 0 to 5, where 0 is a study of low quality and 5 corresponds to the maximum possible quality. It can be considered of good quality a trial that obtains a score of at least 3 [6].

Evaluation of methodological quality of economic analysis

The economic evaluation in healthcare provides for a comparative analysis, under profile of costs and outcomes, between at least two different alternatives. They are always more used in Public Health, at the aim of correctly allocate sources, always more limited, so it's fundamental to effectuate a critic evaluation of these studies. The British Medical Journal (BMJ), in 1996, has instituted a group of work and developed a checklist of 35 items for the evaluation of methodological quality [7]. The checklist is subdivided into the following categories: design of study (7 questions), collection of data (14 questions), analysis and interpretation (14 questions).

Table 3: Jadad Scale for the evaluation of methodological quality of randomized clinical trials

JADAD SCALE
1. Was the study described as randomized? a. Yes (1 point) b. No (0 point) c.
2. Was the study described as double blind? a. Yes (1 point) b. No (0 point)
3. Was there a description of withdrawals and dropouts? a. Yes (1 point) b. No (0 point)
From the total points scored with the three questions, assign a more point or subtract it if *4. Randomization is correct* a. Yes (1 point) b. No (-1 point)
5. Blindness is appropriated a. Yes (1 point) b. No (-1 point)

To every question the answers are expected:"yes", "no", "uncertain" and for some items "not appropriate".
This scale gives equal weight to every question, not taking into account the importance of each question, as in the case in the scale Chiou et al. The Chiou scale is a validated tool, articulated in 16 items of quantity type examining the methodological aspects that maysignificantly influence the evaluation results.

The Chiou scale provide a score that may vary from 0 to 100 and is the weighted sum of points given to each item [8].

Considered the different importance that questions can have into valuation of methodological quality a new system of scores that attribute a weight to the questions of the original Drummond's checklist has been developed [9].

Evaluation of methodological quality of studies about diagnostic tests

Studies about diagnostic tests are very complex and aim to evaluate diagnostic accuracy and/or appropriateness of a test.

In 2003 a first tool to evaluate the study's quality about diagnostic tests has been developed - QUADAS (*Quality Assessment of Diagnostic Accuracy Studies*) – which was subsequently updated, in 2011 (QUADAS-2). QUADAS-2 is a grid subdivided in 4 categories: selection of the patient, index test (the new evaluated test), gold standard of reference and patient's flow in the study and timing of the index test and of gold standard. All categories are valuated in terms of risk of bias (that can be low, high or uncertain) and, moreover, the first three of them are also judged in terms of applicability (low, high or uncertain preoccupation of applicability).

If a study is judged with a low risk of bias or with low preoccupation of applicability in every category, this study can be considered globally with a low risk of bias or with low preoccupation of applicability. If, instead, a study is considered with a high or uncertain risk of bias or with a high or uncertain preoccupation of applicability in just one of the categories, this study must be valuated globally as with a high risk of bias or with a high preoccupation of applicability.

Results can be presented in a table, indicating the risk of bias and the applicability in a graphic way, or they can be summarized in a bar graph [10].

Evaluation of methodological quality of systematic reviews

Systematic reviews are scientific articles that aim to summarize the available evidences in response to a specific question research, clearly formulated, with the use of a systematic and explicit methodology.

Systematic reviews can, moreover, include a quantitative analysis, called meta-analysis, that allows to combine results of the studies included in the systematic review.

Considering the importance of these studies, a correct reporting on the basis of PRISMA Statement (*Preferred Reporting Items for Systematic Reviews and Meta-Analysis*) is fundamental [11].

It comprehends a checklist of 27 items and a flow-chart of 4 phases. PRISMA represents a guide for the reporting of systematic review and is not a tool for the evaluation of quality [12]

To this aim, in 2007, the checklist AMSTAR (*Assessment of Multiple Systematic Reviews*) has been developed. It includes 11 questions in which there are provided the following answers: yes, no, it's impossible to answer, not applicable. The questions take into consideration several aspects of the methodology used for performing a systematic review [13].

Table 4: AMSTAR (Assessment of Multiple Systematic Reviews) checklist for the evaluation of systematic review's quality

AMSTAR (Assessment of Multiple Systematic Reviews)
1. Was an 'a priori' design provided?
2. Was there duplicate study selection and data extraction?
3. Was a comprehensive literature search performed?
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
5. Was a list of studies (included and excluded) provided?
6. Were the characteristics of the included studies provided?
7. Was the scientific quality of the included studies assessed and documented?
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
9. Were the methods used to combine the findings of studies appropriate?
10. Was the likelihood of publication bias assessed?
11. Was the conflict of interest included?

Conclusion

According to evidence-based medicine (EBM), a critical analysis of a study, including the methodological quality and reporting of the study, aims to evaluate the internal validity, clinical relevance and the applicability of a published study. The primary studies take on an important role performing a systematic review to synthetize the results and to avoid distortion of effect's esteems.

Considering the important role that systematic reviews, identifying, evaluating and summarizing the findings of all relevant studies on a well-defined health-related question, can have in decision making, it's clear the importance that they are valid and the adopted methodology avoid or reduce potential bias.

References

1. Evidence-Based medicine Working Group. Evidence Based Medicine. A new approach to teaching the practice of medicine. JAMA 1992 Nov 4; 268 (17): 2420-5.
2. Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP; for the STROBE Initiative. The Strengthening The Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. Lancet. 2007 Oct 20;370(9596):1453-7.
3. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Disponibile all'indirizzo web: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm.
4. Hergoz R, Àlvarez-Pasquin MJ, Dìaz C, del Barrio JL, Estrada JM,Gilà. Are healthcare workers'intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. BMC Public Health. 2013 Feb 19;13:154.
5. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: update guidelines for reporting parallel group randomized trials. BMJ2010; 340:c332.
6. Jadad AR, Moore RA, Carrol D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996; 17: 1-12.
7. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submission to the BMJ. The BMJ Economic Evaluation Working Party. BMJ. 1996 Aug 3;313(7052):275-83.
8. Chiou CF, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD, Yu HT, Keeler EB, Henning JM, Ofman JJ. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care. 2003 Jan; 41 (1):32-44.
9. La Torre G, Nicolotti N, De Waure C, Ricciardi W. Development of a weighted scale to assess the quality of cost-effectiveness studies and an application to the economic evaluations of tetravalent HPV vaccine. J Public Health. 2011; 19:103-11.
10. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang, MM Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.
11. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta analyses: the PRISMA statement. PLoS Med. 2009 Jul 21; 6(7):e1000097.
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche P, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 6 e1000100. Doi:10.1371/journal.pmed.1000100.
13. Shea BJ, Grimshaw JM, Wells GA, Boers M, Anderrson 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 Feb 15;7:10.