

REVIEW ARTICLES

# Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties

Dawid Pieper<sup>a,\*</sup>, Roland Brian Buechter<sup>b</sup>, Lun Li<sup>c</sup>, Barbara Prediger<sup>a</sup>, Michaela Eikermann<sup>a</sup>

<sup>a</sup>Institute for Research in Operative Medicine, Department for Evidence-based health services research, Ostmerheimer Str. 200 (building 38), 51109 Cologne, Witten/Herdecke University, Germany

<sup>b</sup>Institute for Quality and Efficiency in Health Care (IQWiG), Department for Health Information, Mediapark 8 (KölnTurm), 50670 Cologne, Germany

<sup>c</sup>Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, 199 Donggang West Road, Lanzhou, Gansu 730000, China

Accepted 23 December 2014; Published online 30 December 2014

## Abstract

**Objectives:** To summarize all available evidence on measurement properties in terms of reliability, validity, and feasibility of the Assessment of Multiple Systematic Reviews (AMSTAR) tool, including R(evised)-AMSTAR.

**Study Design and Setting:** MEDLINE, EMBASE, Psycinfo, and CINAHL were searched for studies containing information on measurement properties of the tools in October 2013. We extracted data on study characteristics and measurement properties. These data were analyzed following measurement criteria.

**Results:** We included 13 studies, four of them were labeled as validation studies. Nine articles dealt with AMSTAR, two articles dealt with R-AMSTAR, and one article dealt with both instruments. In terms of interrater reliability, most items showed a substantial agreement (> 0.6). The median intraclass correlation coefficient (ICC) for the overall score of AMSTAR was 0.83 (range 0.60–0.98), indicating a high agreement. In terms of validity, ICCs were very high with all but one ICC lower than 0.8 when the AMSTAR score was compared with scores from other tools. Scoring AMSTAR takes between 10 and 20 minutes.

**Conclusion:** AMSTAR seems to be reliable and valid. Further investigations for systematic reviews of other study designs than randomized controlled trials are needed. R-AMSTAR should be further investigated as evidence for its use is limited and its measurement properties have not been studied sufficiently. In general, test–retest reliability should be investigated in future studies. © 2015 Elsevier Inc. All rights reserved.

**Keywords:** Systematic reviews; Psychometrics; Reliability and validity; Observer variation; Epidemiologic methods; Clinimetrics; Measurement properties

## 1. Introduction

Systematic reviews (SRs) are the cornerstone of evidence-based health care as they can provide the highest level of evidence. Thus, access to methodologically sound SRs is crucial to health care professionals and researchers. Over the years, many tools were developed to help users assess the methodological quality of SRs. In this context, it is necessary to distinguish between methodological quality and reporting quality. A reporting quality checklist includes items irrelevant to a review's methodological quality and does not involve judgments regarding risk of bias. Consequently,

an SR can have good quality of reporting but low methodological quality; when the quality of reporting is poor, on the other hand, it is difficult to judge the methodological quality. The Overview Quality Assessment Questionnaire (OQAQ) [1,2] and the Assessment of Multiple Systematic Reviews (AMSTAR) [3–5] are the most widely used tools for the assessment of methodological quality for SRs. Two surveys of overviews (SRs of reviews) found both instruments to be used most often in this context [6,7]. The OQAQ was used in 25% and 11%, and AMSTAR was used in 10% and 9% of all overviews (analysis also includes reviews that had not assessed the quality of SRs), respectively. AMSTAR, which was published in 2007, is growing in popularity, possibly because of its availability, but also because it reflects methodological developments that have occurred after the development of the OQAQ in 1991. It was developed based on the OQAQ and the checklist by Sacks et al. [8] and consists of 11 items, each of which is categorized into a standardized

Conflict of interest: None.

Funding: None.

\* Corresponding author. Tel.: +49-(221)98957-42; fax: +49-(2302) 926-44201.

E-mail address: dawid.pieper@uni-wh.de (D. Pieper).

**What is new?****Key findings**

- The Assessment of Multiple Systematic Reviews (AMSTAR) seem to be reliable and valid.
- Although often used, measurement properties of R-AMSTAR are rather unexplored.

**What this adds to what was known?**

- Provide insights from other samples of reviews and groups of authors, which help establish the generalizability of the findings from the first validation studies.

**What is the implication and what should change now?**

- AMSTAR should be further investigated in a wider range of systematic reviews with varying study designs.
- More evidence is needed with respect to R-AMSTAR.
- Reporting of studies on measurement properties of AMSTAR and R-AMSTAR should be improved.

set of four possible responses: “yes,” “no,” “can’t answer,” or “not applicable.”

After its introduction, the developers of AMSTAR published two studies to validate the instrument [3,5]. Most authors who use AMSTAR refer to one or both of these articles reasoning that it was found to be valid, reliable, and easy to use [9]. According to the developers of AMSTAR, it can be applied to a wide variety of SRs, although they recognize that it has only been tested on SRs of randomized controlled trials (RCTs) evaluating treatment interventions. A second research group revised AMSTAR (R-AMSTAR) to quantify the quality by assigning a quality score to each SR [10]. Rating the risk of bias for a specific study is crucial to grade the confidence for a body of evidence. However, this requires reliable and valid instruments to assess the risk of bias for different types of research designs. Additional validation studies of AMSTAR might have been conducted over the last couple of years. These data provide valuable new information on the measurement properties of this tool and provide insights from other samples of reviews and groups of authors, which help establish the generalizability of the findings from the first validation studies. Thus, it is timely to re-evaluate the measurement properties of AMSTAR in the light of these findings.

The objective of this article was to summarize all available evidence on measurement properties in terms of reliability, validity, and feasibility of the AMSTAR (including R-AMSTAR) tool.

**2. Methods**

We performed a systematic literature search to identify all relevant publications containing information on measurement properties of AMSTAR. MEDLINE (via PubMed), EMBASE (via EMBASE), Psycinfo (via EBSCO), and CINAHL (via EBSCO) were searched from inception through October 2013. We searched for the terms “Assessment of Multiple Systematic Reviews,” “AMSTAR,” or “R-AMSTAR” in the title or abstract. Reference lists of relevant articles were inspected to identify additional articles. No language restrictions were applied.

For consideration in this SR, the following inclusion criteria were applied: empirical study; AMSTAR/R-AMSTAR was applied to more than one review; and the study reported measurement properties in terms of reliability, validity, or feasibility. We followed the Consensus-based Standards for the selection of health Measurement Instruments initiative where reliability is defined as “the degree to which the measurement is free from measurement error” and validity is defined as “the degree to which an instrument truly measures the construct(s) it purports to measure” [11]. Feasibility refers to whether users are able to respond to the questions of the instrument [12]. However, feasibility is not a measurement property. Studies that did not specifically focus on measurement properties were also included (eg, SRs of effectiveness reporting the interrater reliability).

All titles and abstracts were screened independently by two reviewers (D.P., B.P., and M.E.). The full texts of potentially eligible articles were then assessed for inclusion independently by two reviewers (D.P., R.B.B., and B.P.), with the final inclusion decision determined by consensus or discussion. Disagreements at all steps of screening were resolved by discussion or by an independent third reviewer (M.E.).

Data extraction was conducted by one reviewer (D.P.) into structured summary tables. A second reviewer double checked for accuracy (R.B.B., L.L., and M.E.). (The data extraction form is available on request from the corresponding author.) For each publication, we extracted data on year of publication, study objective, applied tool (AMSTAR/R-AMSTAR), number of evaluated reviews, number of reviewers, and measurement properties. In case of any uncertainty regarding the data provided in the studies, we tried to contact the authors for clarification. We made additional calculations based on the results in the studies, where possible. Specifically, we calculated the median interrater reliability for each item and the corresponding range.

To the best of our knowledge, there is currently no guideline for the evaluation of measurement properties of critical appraisal tools for study types. Thus, in our interpretation of the data, we decided to follow measurement criteria [12–14], without applying strict decision rules for measurement properties (eg, interrater reliability has to be  $\geq 0.9$ ).

Download English Version:

<https://daneshyari.com/en/article/10513541>

Download Persian Version:

<https://daneshyari.com/article/10513541>

[Daneshyari.com](https://daneshyari.com)