


LETTER TO THE EDITOR
Psychometric measurements of AMSTAR 2 in a sample of meta-analyses indexed in PsycINFO

AMSTAR (A MeaSurement Tool to Assess systematic Reviews) has been developed by Shea et al. in 2007 [1] to assess the methodological quality of systematic reviews and meta-analyses (MA). In 2017, Shea et al. proposed a revised version of this tool [2]. Recently, Lorenz et al. [3] published, in this journal, an important article in which they confirmed the validity of this revised tool and highlighted the moderate inter-rater reliability (IRR) of AMSTAR 2 in a sample of 60 MAs published between 2012 and 2017 in the field of depression.

In the present letter, we propose to assess reliability between the initial AMSTAR and its revised form, that is, AMSTAR 2, and to measure IRR of the latter in a larger sample of 206 MAs published in psychology and related field. The data were collected previously as part of a larger project assessing the reporting completeness in MAs published in this journal [4].

Following the methodology used in Lorenz's study, we calculated the Fleiss' kappa (K) statistic for dichotomous items and linear weighted kappa for categorical items to determine the IRR of AMSTAR 2. Then, we used two analytical approaches to measure the convergent validity between AMSTAR and AMSTAR 2. First, we evaluated the agreement between the overall rating for AMSTAR 2 and AMSTAR. Second, the relation between the number of items fulfilled by each MA with both tools was evaluated with a Spearman's correlation.

Across our sample of 206 MAs, 195 MAs were categorized as critically low, eight as low, two as moderate, and one as high using the classification advised by AMSTAR 2. With the AMSTAR tool and the classification advised by Lorenz, 51 MAs out of 206 were categorized as low quality, 131 as medium, and 24 as high (Fig. 1A). With regards to the IRR, our results were consistent with the results of Lorenz's study; the IRR (K) for each of the items of AMSTAR2 is in the range of 0.12–0.81. AMSTAR2 has a moderate IRR with a median K of 0.56

(P25–75: 0.30–0.76). Also in agreement with Lorenz, we found some degree of concordance between AMSTAR and AMSTAR 2. All MAs that have not been rated as critically low with AMSTAR 2 were considered as moderate or high quality when rated with AMSTAR (Fig. 1B). Finally, the correlation between the two tools indicated a strong positive association ($r = 0.84$; $P < 0.001$) (Fig. 1C).

In conclusion, consistent with the results from Lorenz's study, the IRR for AMSTAR 2 between the two raters found in our study indicated a moderate agreement in a sample of 206 MAs published in psychology and related fields. Our results also highlight that AMSTAR 2 is subject to a floor effect, as evidenced by the fact that 95% of our sample was rated as critically low, which is the lowest category proposed by the tool. Therefore, the discriminative capacity of this tool is not optimal, and additional investigations of the AMSTAR 2, aiming to take this issue into account, should be encouraged.

Victoria Leclercq*
Charlotte Beaudart
Division of Public Health
Epidemiology and Health Economics
University of Liège
Liege, Belgium

Ezio Tirelli
Department of Psychology
University of Liège
Liege, Belgium

Olivier Bruyère
Division of Public Health
Epidemiology and Health Economics
University of Liège
Liege, Belgium

*Corresponding author. Division of Public Health, Epidemiology and Health Economics, University of Liège, Liege, Belgium. Tel.: +32 (0)4 366 25 20; fax: +32 (0)4 366 28 12.
E-mail address: victoria.leclercq@uliege.be (V. Leclercq)

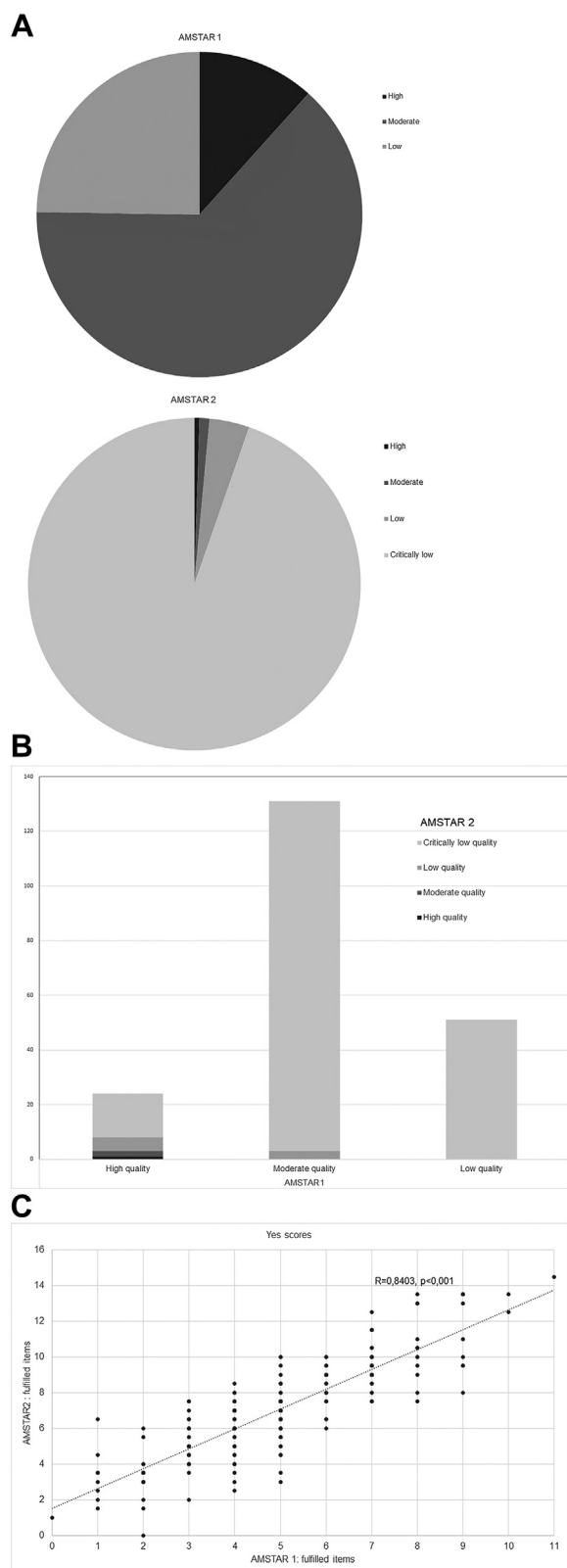


Fig. 1. Validity results. Panel A present the pie chart of the overall judgments for AMSTAR et AMSTAR 2. Panel B present the bar chart illustrating the concordance of the overall judgments for both tools. Panel C present the correlation of the number of items fulfilled between the two tools.

References

- [1] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
- [2] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. Amstar 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017j4008.
- [3] Lorenz RC, Matthias K, Pieper D, Wegewitz U, Morche J, Nocon M, et al. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. *J Clin Epidemiol* 2019;114:133–40.
- [4] Leclercq V, Beaudart C, Ajamieh S, Rabenda V, Tirelli E, Bruyère O. Meta-analyses indexed in PsycINFO had a better completeness of reporting when they mention PRISMA. *J Clin Epidemiol* 2019;115:46–54.

<https://doi.org/10.1016/j.jclinepi.2019.10.005>