RESEARCH ARTICLE

Research Synthesis Methods **WILEY**

The application of AMSTAR2 in 32 overviews of systematic reviews of interventions for mental and behavioural disorders: A cross-sectional study

Karina Karolina De Santis[1] ⓘ | **Robert C. Lorenz[2]** | **Meret Lakeberg[1]** |
Katja Matthias[3] ⓘ

[1]Department Prevention and Evaluation, Leibniz Institute for Prevention Research and Epidemiology- BIPS, Bremen, Germany

[2]Lise-Meitner Group for Environmental Neuroscience, Max Planck Institute for Human Development, Berlin, Germany

[3]Faculty of Electrical Engineering and Computer Science, University of Applied Sciences Stralsund, Stralsund, Germany

Correspondence
Karina Karolina De Santis, Leibniz Institute for Prevention Research and Epidemiology- BIPS, Achterstr. 30, 28359 Bremen, Germany.
Email: desantis@leibniz-bips.de

Abstract

'A measurement tool to assess systematic reviews, version 2' (AMSTAR2) is a 16-item tool to critically appraise systematic reviews (SRs) of healthcare interventions. This study aimed to assess the methods and outcomes of AMSTAR2 appraisals in overviews of SRs of interventions for mental and behavioural disorders. The cross-sectional study was conducted using 32 overviews of SRs selected from three electronic databases in January 2021. Data items included overview and SR characteristics and AMSTAR2 appraisal methods and outcomes. Data were extracted by two authors independently and narratively synthesised using descriptive statistics (means ± SD and relative frequencies). SR characteristics were compared based on AMSTAR2 appraisal outcomes using chi-square tests. The 32 overviews appraised SRs of predominantly non-pharmacological interventions for mental disorders. AMSTAR2 appraisals were reported as confidence ratings in 25/32 overviews or individual item scores in 24/32 overviews. Most SRs/overview were non-Cochrane (mean = 94%), included RCTs only (mean = 77%) and were published before AMSTAR2 release (mean = 79%). The confidence ratings derived in 25 overviews for 349 SRs were predominantly critically low (68%). Confidence ratings were similar for SRs with RCTs only versus RCTs+non-RCTs or SRs published before versus after AMSTAR2 release, while Cochrane SRs received more high+moderate than low+critically low confidence ratings ($p < 0.01$). Confidence ratings derived based on AMSTAR2 do not differentiate among SRs of healthcare interventions except for Cochrane SRs that fulfil the criteria for high confidence ratings. AMSTAR2 items should be consulted to avoid common weaknesses in future SRs.

KEYWORDS
AMSTAR2, evidence appraisal, overview, systematic review

What is already known?
- AMSTAR2 is a frequently used tool for appraisal of systematic reviews (SRs)

What is new?
- AMSTAR2 application was assessed in 32 overviews of SR of healthcare interventions
- Most SRs received critically low overall confidence ratings
- Confidence ratings depended on Cochrane status but not on RCT status or SR age

Potential impact for *Research Synthesis Methods* readers outside the authors' field
- AMSTAR2 does not differentiate among SRs based on RCT status or SR age
- AMSTAR2 items can provide helpful guidance to avoid common weaknesses in future SRs

1 | BACKGROUND

The number of systematic reviews (SRs) in healthcare is exponentially increasing such that there are already more SRs than primary trials in some clinical fields.[1] Multiple SRs on related research questions can be systematically identified and aggregated using a new method of research synthesis, an overview of SRs.[2] While compared to SRs the number of overviews is still relatively low, their popularity is also growing exponentially. For example, there were already 1826 overviews on any topic in PubMed on 21 July 2021, including 318 published in 2021 alone (Figure 1).

An important function of overviews is to systematically assess the quality of SRs.[2] This is particularly important because while SRs can guide future research and inform policy and clinical decisions, their quality varies considerably for recent examples see.[3–5] A critical appraisal of SRs of healthcare interventions can be conducted using 'A Measurement Tool to Assess Systematic Reviews, version 2' (AMSTAR2)[6] that is an open-access scale with 16-items. AMSTAR2 was developed to address some limitations of its predecessor, AMSTAR.[7] Specifically, the tool can be applied to SRs of randomised controlled trials (RCT) and non-RCTs and its items adhere to the current guidelines for SRs.[8] The new aspect of AMSTAR2 is the rating of the overall confidence in the results of the SR (high, moderate, low or critically low) derived based on a combination of scores on seven critical and nine non-critical items.[6]

Although AMSTAR2 has acceptable psychometric properties,[6,9,10] the rating of the overall confidence in SRs is difficult for two reasons. First, AMSTAR2 tends to assign the same (mostly low to critically low) ratings to SRs of various healthcare interventions in different clinical fields.[4,5,8,11–15] Second, AMSTAR2 users derive the overall ratings using different methods that are often not explicitly reported.[16] It is possible that both findings are related in that AMSTAR2 users attempt to improve the discriminatory power of the tool by selecting specific critical items according to their research questions, as advised by AMSTAR2 developers.[6] However, any modifications to the original seven-critical-item algorithm need to be explicitly reported by AMSTAR2 users to allow replication of their method to derive the overall confidence ratings.[16]

Since AMSTAR2 is still relatively new (published in late 2017), it remains unclear how AMSTAR2 users apply the tool. Thus, the aim of the current study was to descriptively assess AMSTAR2 appraisal methods and outcomes in studies with homogenous designs (overviews of SRs), objectives (appraisal of SRs) and field (interventions for mental and behavioural disorders).

2 | METHODS

We prospectively registered the study protocol[17] and adhere to 'The Strengthening the Reporting of Observational Studies in Epidemiology' (STROBE) guidelines.[18]

2.1 | Design and setting

We utilised a cross-sectional, observational design to descriptively assess how AMSTAR2 was applied in overviews of SRs of interventions for mental and behavioural disorders published in peer-reviewed journals.

2.2 | Data sources

Two authors conducted a title/abstract search for 'AMSTAR2' in Medline, Epistemonikos and CINHAL in January 2021. The inclusion and exclusion criteria, search strategy and study selection are reported in Appendix, Tables A1-A3 and Figure A1. The studies were independently selected by two authors according to the following inclusion criteria:
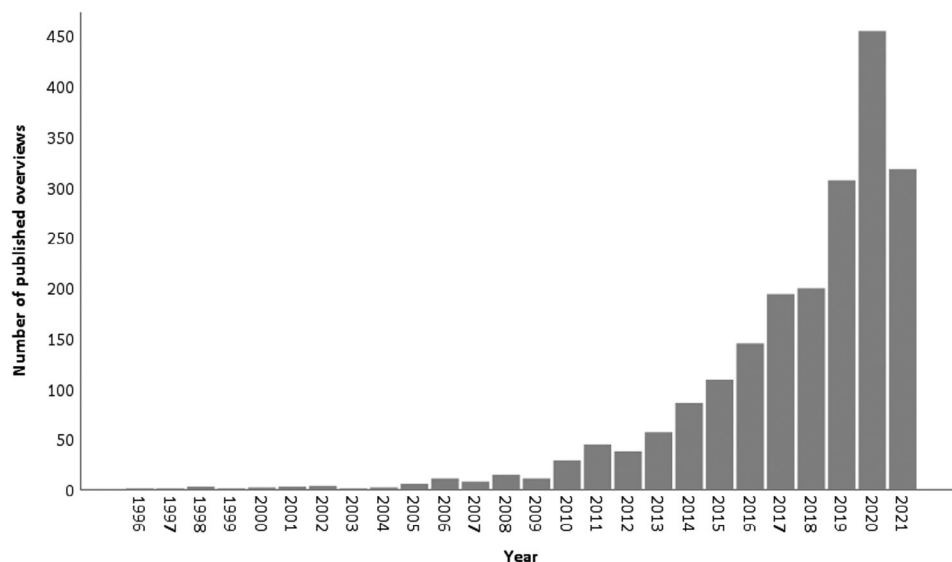
1. Study design: Overview of SRs,
2. PICO: (a) Population with mental and behavioural disorders (including neurological disorders due to their behavioural symptoms), (b) Intervention: Any pharmacological and/or non-pharmacological (complementary or alternative), (c) Control: Any or none, (d) Outcome: Any clinical outcome
3. AMSTAR2: At least one appraisal conducted.

Any discrepancies were resolved by consensus during discussion. Out of 352 sources from the electronic literature search, 32 overviews met the inclusion criteria for the current study. The list of included and excluded studies is reported in Appendix, Table A3.

2.3 | Variables (data items)/ measurement

Unlike specified in our protocol,[17] two (rather than one) authors independently coded all data from 32 overviews using a self-developed coding sheet in Microsoft-Excel. There were only few minor inconsistencies in coding that were resolved by consensus during discussion. The coded data items included:

1. Overview characteristics: Citation, author number and region, intervention and disorder type,
2. Characteristics of SRs included in the overviews: 'Recency' measured as a number of SRs published within 5 years of overview publication year, 'Cochrane status' measured as the number of Cochrane SRs, 'randomised controlled trial (RCT) status' measured as the number of SRs with RCTs only

and 'SR age' measured as the year of SR publication relative to AMSTAR2 release year (up to 2018 vs. 2018 onwards),
3. AMSTAR2 appraisal details (methods and outcomes) in the overviews.

2.4 | Data analysis

As specified in the protocol, all data from 32 overviews were narratively synthesised using descriptive statistics. Scale variables were expressed as mean $\pm$ SD and range (minimum - maximum). Nominal variables were expressed as relative frequencies computed out of 32 overviews. The overall confidence ratings derived based on seven critical and nine non-critical items were narratively synthesised using descriptive statistics as explained above. In addition to the protocol, chi-square tests computed in IBM-SPSS24 were used to compare SR characteristics based on AMSTAR2 appraisal outcomes (high+moderate vs. low+critically low overall confidence ratings).

3 | RESULTS

3.1 | Sample size

The current study includes the data from 32 overviews.

3.2 | Characteristics of 32 overviews

The characteristics of each overview are reported in Appendix, Table A4. The characteristics of all 32 overviews

TABLE 1 Characteristics of 32 overviews

Variable	Frequency	% of 32
Publication year		
2019	8	25%
2020	22	69%
2021	2	6%
Corresponding author region		
Asia	20	64%
Europe	8	21%
Australia	2	7%
North America	2	7%
Diagnosis type (ICD-10)		
Mental	21	66%
Behavioural (neurological)	11	34%
Intervention type		
Non-pharmacological, complementary or alternative	27	84%
Pharmacological	3	9%
Both	2	6%
Included SRs by Cochrane status[a]		
Non-Cochrane	19	59%
Cochrane and non-Cochrane	13	41%
Included SRs by RCT status		
RCTs	19	59%
RCTs and non-RCTs	12	38%
Non-RCTs	1	3%
Included SRs by age (publication year relative to AMSTAR2 release)		
All published before AMSTAR2 up to 2018	5	16%
At least one published after AMSTAR2 in 2018 onwards	27	84%

Note: [a]Cochrane status also includes Campbell SRs.
Abbreviations: AMSTAR2, A Measurement Tool to Assess Systematic Reviews; ICD-10, international classification of diseases; RCT, randomised controlled trial; SR, systematic review.

are shown in Table 1. The overviews were written by 2–14 authors (mean = 6, SD = 3), published in 2019–2021 and most (64%) originated from Asia. The focus of the overviews was predominantly on non-pharmacological interventions (84%) for mental disorders (66%). Most overviews included only non-Cochrane SRs (59%) and SRs of RCTs only (59%). Some overviews (16%) included SRs published up to 2018 before AMSTAR2 release while others (84%) included at least one SR that was published in 2018 onwards after AMSTAR2 release.

TABLE 2 Characteristics of SRs appraised with AMSTAR2 in 32 overviews

Variable	Mean ± SD	Min - Max
% SRs/overview by recency (published within 5 years since overview publication year)	54 ± 21%	7 – 100%
% SRs/overview with non-Cochrane status	94 ± 11%	56 – 100%
% SRs/overview with RCTs only status	77 ± 33%	0 – 100%
% SRs/overview by age (publication year up to 2018 before AMSTAR2 release)	79 ± 16%	33 – 100%

Abbreviations: AMSTAR2, A Measurement Tool to Assess Systematic Reviews; RCT, randomised controlled trial; SR, systematic review.

3.3 | Characteristics of SRs in 32 overviews

The characteristics of SRs appraised in all overviews are shown in Table 2. The overviews appraised 4–64 SRs (mean = 16, SD = 12). The SRs were published in 1997–2020 and about half were recent (published within 5 years of the overview publication year; mean = 54% of SRs/overview). Most SRs were non-Cochrane (mean = 94% of SRs/overview), included RCTs only (mean = 77% of SRs/overview) and were published before AMSTAR2 release (mean = 79% of SRs/overview).

3.4 | AMSTAR2 appraisal methods and outcomes in 32 overviews

The AMSTAR2 appraisal methods are shown in Table 3. While most overviews had no information about the study protocol (56%), most of those with a protocol adhered to their planned appraisal methods (62%). The appraisals were typically conducted independently by at least two authors (84%). AMSTAR2 appraisals were reported as confidence ratings in 25/32 (78%) overviews. Such confidence ratings were derived based on seven critical and nine non-critical items, as suggested by AMSTAR2 developers.[6] Other overviews also reported the overall confidence ratings without specifying the method of derivation (6%) or using the AMSTAR2 website (3%), while 12% of overviews did not derive the overall confidence ratings. Most overviews (75%) also discussed SR weaknesses and/or strengths using the individual item scores. Appraisals were also conducted using other tools, such as the Grading of Recommendations, Assessment, Development and

TABLE 3 AMSTAR2 appraisal methods in 32 overviews

Variable	Frequency	% of 32
Overview protocol		
Yes published/accessible online	13	41%
If yes, adherence to methods listed in the protocol	8/13	(62%)
If yes, no adherence to methods listed in the protocol	3/13	(23%)
If yes, adherence unclear	2/13	(15%)
Yes unpublished/not accessible online	1	3%
No information	18	56%
Appraisal conduct		
Independently by at least two authors	27	84%
One author	2	6%
Unclear	3	9%
Overall confidence rating		
Four categories based on seven critical items as suggested by AMSTAR2 developers	25	78%
Four categories, method unclear (not reported)	2	6%
Four categories based on AMSTAR2 website	1	3%
No overall rating	4	12%
Scores on individual AMSTAR2 items discussed		
Yes	24	75%
No	8	25%
Other appraisal methods in addition to AMSTAR2		
GRADE	17	53%
GRADE and ROBIS	2	6%
None	13	41%

Note: Values in brackets are computed out of 13.
Abbreviations: AMSTAR2, A Measurement Tool to Assess Systematic Reviews; GRADE, Grading of Recommendations, Assessment, Development and Evaluations; RCT, randomised controlled trial; ROBIS, Risk of Bias in Systematic Reviews; SR, systematic review.

Evaluations (GRADE) approach and/or the Risk of Bias in Systematic Reviews (ROBIS) tool (59%).

3.5 | Overall confidence ratings in 25 overviews

Out of 32 overviews, 25 overviews reported the overall confidence ratings based on seven critical and nine non-critical items, while seven other overviews either did not

TABLE 4 Overall confidence ratings derived based on seven critical and nine non-critical items on AMSTAR2 in 25 overviews

Variable	Mean ± SD	Min – Max
% SRs/overview rated as high	4 ± 7%	0 – 27%
% SRs/overview rated as moderate	4 ± 12%	0 – 56%
% SRs/overview rated as low	17 ± 21%	0 – 71%
% SRs/overview rated as critically low	63 ± 28%	14 – 100%

Abbreviations: AMSTAR2, A Measurement Tool to Assess Systematic Reviews; SR, systematic review.

explain the methods of derivation or did not derive the overall confidence ratings. Out of 25 overviews, nine addressed a unique diagnosis and/or intervention and thus included only unique SRs (Appendix, Table A5). Further 16 overviews were grouped according to diagnosis and intervention and ordered by their publication year and month. The included SRs were visually compared among overviews in each group. Out of 16 overviews, six overviews used partially overlapping SRs that were identified and included only once in the final SR count. Out of 380 SRs appraised in 25 overviews, 349 SRs were unique (included in only one overview; Appendix, Table A5).

The overall confidence ratings for 349 unique SRs in 25 overviews are shown in Table 4 and Figure 2. The majority of 25 overviews included SRs with critically low confidence ratings (mean = 63% of SRs/overview), while SRs with other confidence ratings were less frequent (low ratings: mean = 17% of SRs/overview; moderate ratings: mean = 4% of SRs/overview; high ratings: mean = 4% of SRs/overview).

3.6 | Overall confidence ratings depending on SR characteristics in 25 overviews

The overall confidence ratings were further inspected based on the characteristics of 349 unique SRs appraised in 25 overviews. The overall confidence ratings based on seven critical and nine non-critical items for all 349 unique SRs are shown in Figure 3 and in Appendix, Table A6. The 349 SRs in 25 overviews received critically low (68%), low (21%), moderate (4%) or high (6%) overall confidence ratings. Regardless of the confidence rating, most SRs included RCTs only and were published before AMSTAR2 release (up to 2018). However, SRs rated 'high' were mostly Cochrane (77%), while SRs rated 'critically low' were mostly non-Cochrane (99%). We conducted Pearson chi-square (X^2)

FIGURE 2 Overall confidence ratings derived based on seven critical and nine non-critical items on AMSTAR2 for 349 SRs in 25 overviews. Abbreviations: AMSTAR2, A Measurement Tool to Assess Systematic Reviews; SR, systematic review
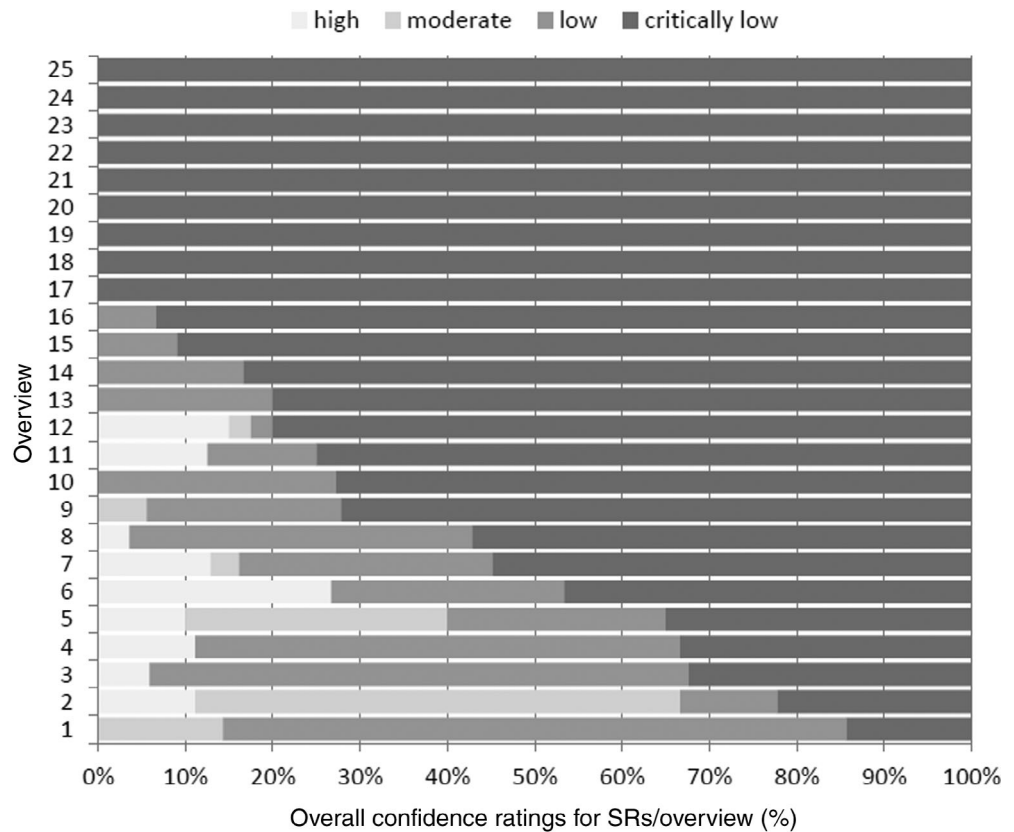
analyses to compare the characteristics of SRs with high or moderate versus low or critically low confidence ratings. These analyses showed that the confidence ratings were similar for SRs with RCTs only (86% in 'high+moderate' vs. 80% in 'low+critically low' groups; $X^2 = 0.86$, df = 1, p = 0.353) and for SRs published before AMSTAR2 release (76% in 'high+moderate' vs. 78% in 'low+critically low' groups; $X^2 = 0.07$, df = 1, $p = 0.795$). However, a higher proportion of Cochrane SRs received high or moderate (60%) than low or critically low (2%) confidence ratings ($X^2 = 142.12$, df = 1, $p < 0.01$).

4 | DISCUSSION

4.1 | Main findings

Our study assessed how AMSTAR2 was applied to appraise SRs of interventions for mental and behavioural disorders in 32 overviews published in 2019–2021. Most overviews derived the overall confidence ratings and/or discussed SR weaknesses using the scores on individual AMSTAR2 items. The confidence ratings assigned to 349 unique SRs in 25 overviews were predominantly critically low. Confidence ratings were similar for SRs with RCTs only versus RCTs+non-RCTs or in SRs published before versus after AMSTAR2 release, while Cochrane

SRs received more high or moderate than low or critically low confidence ratings.

4.2 | Application of AMSTAR2

We show that most overviews followed the recommendations of AMSTAR2 developers in that two authors independently appraised the SRs and derived the overall confidence ratings according to the seven-critical-item algorithm.[6] However, in some overviews the appraisal methods were unclear or the authors did not derive the overall confidence ratings. Thus, the current results support a call for better reporting first proposed for AMSTAR[19] and also applicable to AMSTAR2.[16]

It is disappointing that the overall confidence in most SRs of interventions for mental and behavioural disorders is low to critically low as already shown in various clinical fields.[4,5,8,11–15] While authors of SRs are primarily responsible for the quality of their work, the 'gatekeepers' of scientific publishing (peer-reviewers and journal editors) are equally responsible for guiding SR authors through the publication process, especially if the authors are new in the field of research synthesis or are not English speakers.

Two hypotheses have been proposed to explain the abundance of the critically low ratings on AMSTAR2: (1) AMSTAR2 classification is too conservative leading to
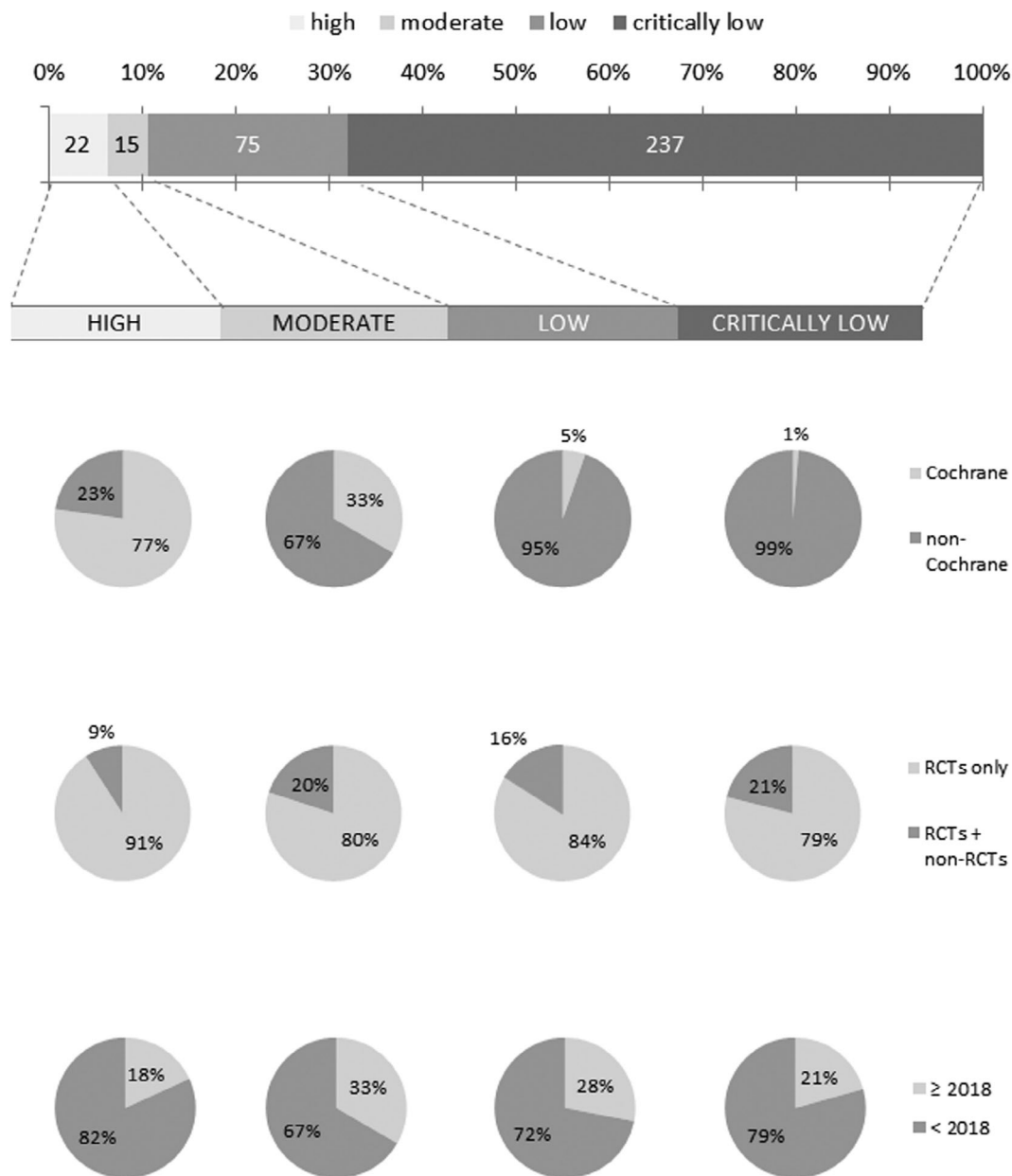
FIGURE 3 Overall confidence ratings and characteristics of 349 SRs in 25 overviews. In the upper part of the figure, the stacked bar graph shows the overall confidence ratings derived based on seven critical and nine non-critical items on AMSTAR2 for 349 SRs included in all 25 overviews. In the lower part of the figure, the pie charts show the characteristics of SRs (Cochrane status, RCT status and SR age) stratified by the overall confidence ratings. Abbreviations: AMSTAR2, A Measurement Tool to Assess Systematic Reviews; RCT, randomised controlled trial; SR, systematic review

a floor effect and little discriminative power of the tool,[11] (2) AMSTAR2 classification detects a very high methodological quality (for example, associated with Cochrane SRs).[12] The current study provides some evidence in favour of both hypotheses. On the one hand, the floor effect on AMSTAR2 is likely because most SRs in one field received critically low ratings independent of the quality of data (SRs with RCTs only vs. RCTs+non-RCTs) or SR age (published before vs. after AMSTAR2 release). On the other hand, AMSTAR2 appears to detect SRs with

high methodological quality because it assigned more high or moderate than low or critically low ratings to Cochrane SRs.

4.3 | Suggestions for using AMSTAR2

We propose three suggestions for the users of AMSTAR2 that could reduce the potential floor effect on the tool and/or improve the interpretation of confidence ratings.

1. Decide if AMSTAR2 is a suitable appraisal tool. Overview authors should choose the appropriate critical appraisal tool while planning their overview. SRs are typically appraised to identify common problems, including methodological limitations and/or poor reporting.[3,4,13–15,20–25] While the focus of AMSTAR2 is on the methodological quality, this tool is highly sensitive to poor reporting. Specifically, inadequate reporting of only 2/16 items can be sufficient to assign a critically low confidence rating to a SR. Thus, when interpreting AMSTAR2 scores, the overview authors need to address the possibility that SRs could receive critically low confidence ratings for poor reporting rather than for poor methodological quality. Furthermore, the aim of some overviews may be to assess the risk of bias in SRs rather than to focus on the methodological quality. In this case, other tools, such as ROBIS,[26] may be more appropriate than AMSTAR2 to appraise SRs.

2. Adapt the selection of the critical items. As suggested by AMSTAR2 developers, the users could decide a priori about deletion, addition or replacement of the critical items and adapt the overall confidence ratings to the specific research question.[6] While this approach could reduce the potential floor effect on AMSTAR2, derivation of the overall confidence ratings using different methods could lead to confusion in comparing such ratings among studies. Sufficient reporting of appraisal methods using AMSTAR2 are essential to prevent such confusion.[16] Furthermore, our unpublished study suggests that five critical and two non-critical AMSTAR2 items are difficult to score in SRs of interventions for psychological or neurological disorders. Thus, revisions to the scoring guidelines may be required for such items. Interestingly, our choice of items for potential revisions is closely aligned with importance ratings of AMSTAR2 items according to 242 experts in SR/meta-analysis.[27] The experts rated the items on the appropriateness of statistical analyses, adequacy of the literature search and assessment of the risk of bias and heterogeneity as the most important items on AMSTAR2,[27] while we selected these items as candidates for possible revisions.

3. Inspect individual AMSTAR2 items. Some overviews in our study did not derive the overall confidence ratings or, rather than focusing entirely on the poor confidence ratings, addressed the SR weaknesses based on the individual AMSTAR2 items. In fact, the aggregated (sum or percentage) score based on fulfilled items could provide a better differentiation among SRs than the overall confidence ratings.[9] The focus on individual AMSTAR2 items could guide authors to

avoid common weaknesses throughout the preparation of future SRs.[8]

4.4 | Limitations

There were several limitations in the current study. First, our search strategy could have missed some relevant overviews if the authors did not report that they have used AMSTAR2 in title or abstract. Although it cannot be ruled out, it is unlikely that such missing overviews would majorly influence the outcomes of our meta-research study. Second, our results are based on overviews in one clinical field and may not reflect how researchers conduct AMSTAR2 appraisals in other fields and/or if such appraisals lead to other (better) overall confidence ratings. Third, we have not conducted sensitivity analyses to assess the AMSTAR2 appraisal outcomes based on different critical items or compared to other tools used in some overviews. For example, excluding two typical weaknesses (lack of a protocol or lack of a list of excluded studies) from critical items could improve the critically low confidence ratings for some SRs. Confidence ratings could also depend on other factors, such as SR publication source (academic journal vs. dissertation; peer-reviewed vs. not peer-reviewed; published internationally in English vs. published in a local language). Fourth, although we have otherwise adhered to our protocol, we conducted unplanned, univariate chi-square tests to compare the characteristics of SRs with high or moderate versus low or critically low confidence ratings. These analyses confirmed our descriptive conclusion that confidence ratings depended on Cochrane status but not on RCT status or SR age. Once sufficient volume of data is available, confidence ratings could also be compared multivariately based on different methods of AMSTAR2 appraisals (for example, based on seven critical items or other critical items, conducted by hand or using the website, conducted by one or two authors) and multiple SR characteristics. Fifth, while 75% of overviews discussed the scores on individual AMSTAR2 items, we have not coded any further details, such as whether sum or percentage scores were computed for fulfilled items.

5 | CONCLUSION

AMSTAR2 was applied by authors of 32 overviews to appraise SRs of interventions for mental and behavioural disorders. Appraisal outcomes were reported either as confidence ratings or individual item scores were discussed. Most SRs received critically low confidence ratings. Confidence ratings were similar for SRs with RCTs

only versus RCTs+non-RCTs or in SRs published before versus after AMSTAR2 release, while most Cochrane SRs fulfilled the criteria for high or moderate confidence ratings. Except for Cochrane SRs, AMSTAR2 does not differentiate among SRs of healthcare interventions. The potential floor effect could be reduced if AMSTAR2 users select specific critical items according to their research questions, although this approach will not allow to compare AMSTAR2 ratings among studies. AMSTAR2 items should be consulted to avoid common weaknesses in future SRs.

ACKNOWLEDGMENT

CONFLICT OF INTEREST

AUTHOR CONTRIBUTIONS

Karina De Santis: Conceptualisation; Project administration; Supervision; Data curation; Data analysis; Visualisation; Writing - original draft; Writing - review & editing; Robert Lorenz: Data curation; Visualisation; Writing - review & editing. Meret Lakeberg: Data curation; Visualisation; Writing - review & editing. Katja Matthias: Conceptualisation; Data curation; Data analysis; Writing - review & editing.

ETHICS STATEMENT

The protocol for this study was prospectively registered in OSF [17].

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Karina Karolina De Santis ⬤ https://orcid.org/0000-0001-7647-6767
Katja Matthias ⬤ https://orcid.org/0000-0002-7934-0256

REFERENCES

1. Niforatos JD, Weaver M, Johansen ME. Assessment of publication trends of systematic reviews and randomized clinical trials, 1995 to 2017. *JAMA Intern Med*. 2019;179(11):1593-1594. doi:10.1001/jamainternmed.2019.3013
2. Pollock M, Fernandes R, Becker L, Pieper D, Hartling L. Chapter V: overviews of reviews. In: Higgins J, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.2* https://training.cochrane.org/handbook/current/chapter-v; Cochrane; 2021.
3. Kedzior K, Kaplan I. Tai chi and Parkinson's disease (PD): a systematic overview of the scientific quality of the past systematic reviews. *Complement Ther Med*. 2019;46:144-152. doi:10.1016/j.ctim.2019.08.008
4. Matthias K, Rissling O, Pieper D, et al. The methodological quality of systematic reviews on the treatment of adult major depression needs improvement according to AMSTAR 2: a cross-sectional study. *Heliyon*. 2020;6(9):e04776. doi:10.1016/j.heliyon.2020.e04776
5. O'Connor DB. The future of health behaviour change interventions: opportunities for open science and personality research. *Health Psychol Rev*. 2020;14(1):176-181. doi:10.1080/17437199.2019.1707107
6. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008. doi:10.1136/bmj.j4008
7. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7(1):1-7. doi:10.1186/1471-2288-7-10
8. De Santis KK, Kaplan I. Assessing the quality of systematic reviews in healthcare using AMSTAR and AMSTAR2: a comparison of scores on both scales. *Z Psychol*. 2020;228(1):36-42. doi:10.1027/2151-2604/a000397
9. Lorenz RC, Matthias K, Pieper D, et al. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. *J Clin Epidemiol*. 2019;114:133-140. doi:10.1016/j.jclinepi.2019.05.028
10. Pieper D, Puljak L, Gonzalez-Lorenzo M, Minozzi S. Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. *J Clin Epidemiol*. 2019;108:26-33. doi:10.1016/j.jclinepi.2018.12.004
11. Leclercq V, Beaudart C, Tirelli E, Bruyere O. Psychometric measurements of AMSTAR 2 in a sample of meta-analyses indexed in PsycINFO. *J Clin Epidemiol*. 2020;119:144-145. doi:10.1016/j.jclinepi.2019.10.005
12. Lorenz RC, Matthias K, Pieper D, et al. AMSTAR 2 overall confidence rating: lacking discriminating capacity or requirement of high methodological quality? *J Clin Epidemiol*. 2020;119:142-144. doi:10.1016/j.jclinepi.2019.10.006
13. Leclercq V, Beaudart C, Ajamieh S, Tirelli E, Bruyère O. Methodological quality of meta-analyses indexed in PsycINFO: leads for enhancements: a meta-epidemiological study. *BMJ Open*. 2020;10(8):e036349. doi:10.1136/bmjopen-2019-036349
14. Antony J, Brar R, Khan PA, et al. Interventions for the prevention and management of occupational stress injury in first responders: a rapid overview of reviews. *Syst Rev*. 2020;9(1):121. doi:10.1186/s13643-020-01367-w
15. Rios P, Cardoso R, Morra D, et al. Comparative effectiveness and safety of pharmacological and non-pharmacological interventions for insomnia: an overview of reviews. *Syst Rev*. 2019;8(1):281. doi:10.1186/s13643-019-1163-9

16. Pieper D, Lorenz RC, Rombey T, et al. Authors should clearly report how they derived the overall rating when applying AMSTAR 2-a cross-sectional study. *J Clin Epidemiol*. 2021;129: 97-103. doi:10.1016/j.jclinepi.2020.09.046

17. De Santis K, Lorenz R, Matthias K. Application of AMSTAR 2 in overviews of systematic reviews of interventions for mental and behavioural disorders: A cross-sectional study (protocol registration). https://osf.io/g5437/ 2020.

18. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10):e296. doi:10.1371/journal. pmed.0040296

19. Pieper D, Koensgen N, Breuing J, Ge L, Wegewitz U. How is AMSTAR applied by authors - a call for better reporting. *BMC Med Res Methodol*. 2018;18(1):56. doi:10.1186/s12874-018-0520-z

20. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med*. 2007;4(3):e78. doi:10.1371/journal. pmed.0040078

21. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med*. 2016;13(5): e1002028. doi:10.1371/journal.pmed.1002028

22. Page MJ, Altman DG, Shamseer L, et al. Reproducible research practices are underused in systematic reviews of biomedical interventions. *J Clin Epidemiol*. 2018;94:8-18. doi:10.1016/j. jclinepi.2017.10.017

23. Oliveras I, Losilla JM, Vives J. Methodological quality is underrated in systematic reviews and meta-analyses in health psychology. *J Clin Epidemiol*. 2017;86:59-70. doi:10.1016/j.jclinepi.2017.05.002

24. Kedzior K, Seehoff H. Common problems with meta-analysis in published reviews on major depressive disorders (MDD): A systematic review. Res Syn Conf. https://cutt.ly/FgtAcAi. 2018.

25. Page MJ, Altman DG, McKenzie JE, et al. Flaws in the application and interpretation of statistical analyses in systematic reviews of therapeutic interventions were common: a cross-sectional analysis. *J Clin Epidemiol*. 2018;95:7-18. doi:10.1016/j.jclinepi.2017.11.022

26. Whiting P, Savović J, Higgins JPT, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-234. doi:10.1016/j.jclinepi.2015.06.005

27. Leclercq V, Hiligsmann M, Parisi G, Beaudart C, Tirelli E, Bruyère O. Best-worst scaling identified adequate statistical methods and literature search as the most important items of AMSTAR2 (a measurement tool to assess systematic reviews). *J Clin Epidemiol*. 2020;128:74-82. doi:10.1016/j.jclinepi.2020.08.011

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.