

# BMJ Open Methodological quality of meta-analyses indexed in PsycINFO: leads for enhancements: a meta-epidemiological study

Victoria Leclercq <sup>1</sup>, Charlotte Beaudart,<sup>1</sup> Sara Ajamieh,<sup>1</sup> Ezio Tirelli,<sup>2</sup> Olivier Bruyère<sup>1</sup>

**To cite:** Leclercq V, Beaudart C, Ajamieh S, *et al*. Methodological quality of meta-analyses indexed in PsycINFO: leads for enhancements: a meta-epidemiological study. *BMJ Open* 2020;**10**:e036349. doi:10.1136/bmjopen-2019-036349

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-036349>).

Received 12 December 2019  
Revised 06 May 2020  
Accepted 09 June 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Division of Public Health, Epidemiology and Health Economics, University of Liege, Liege, Belgium

<sup>2</sup>Department of Psychology, University of Liege, Liege, Belgium

## Correspondence to

Victoria Leclercq;  
[victoria.leclercq@uliege.be](mailto:victoria.leclercq@uliege.be)

## ABSTRACT

**Objectives** Meta-analyses (MAs) are often used because they are lauded to provide robust evidence that synthesises information from multiple studies. However, the validity of MA conclusions relies on the procedural rigour applied by the authors. Therefore, this meta-research study aims to characterise the methodological quality and meta-analytic practices of MAs indexed in PsycINFO.

**Design** A meta-epidemiological study.

**Participants** We evaluated a random sample of 206 MAs indexed in the PsycINFO database in 2016.

**Primary and secondary outcomes** Two authors independently extracted the methodological characteristics of all MAs and checked their quality according to the 16 items of the A MeaSurement Tool to Assess systematic Reviews (AMSTAR2) tool for MA critical appraisal. Moreover, we investigated the effect of mentioning Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) on the methodological quality of MAs.

**Results** According to AMSTAR2 criteria, 95% of the 206 MAs were rated as critically low quality. Statistical methods were appropriate and publication bias was well evaluated in 87% and 70% of the MAs, respectively. However, much improvement is needed in data collection and analysis: only 11% of MAs published a research protocol, 44% had a comprehensive literature search strategy, 37% assessed and 29% interpreted the risk of bias in the individual included studies, and 11% presented a list of excluded studies. Interestingly, the explicit mentioning of PRISMA suggested a positive influence on the methodological quality of MAs.

**Conclusion** The methodological quality of MAs in our sample was critically low according to the AMSTAR2 criteria. Some efforts to tremendously improve the methodological quality of MAs could increase their robustness and reliability.

## INTRODUCTION

Since the definition of meta-analyses (MAs) being introduced by Glass in 1976, MAs conducted in behavioural and social sciences have increased rapidly in number. There were more than 30 000 MAs indexed in PsycINFO

## Strengths and limitations of this study

- Some studies have highlighted methodological weaknesses in the conduct of systematic reviews (SRs) and meta-analyses (MAs) and we search to have an overview of methodological practice of MAs indexed in PsycINFO according to the tool A MeaSurement Tool to Assess systematic Reviews which aimed to critically appraise SRs and MAs.
- Rather than solely focusing on methodological characteristics of MAs, this study investigates also the effect of the mentioning Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement on the methodological quality of MAs.
- A sample of 206 MAs indexed in PsycINFO in 2016 and published in English was analysed.
- Our findings cannot be generalised to MAs published in other years than 2016, in other languages than English or in other databases than PsycINFO.

in 2018. MAs are used extensively for clinical and policy decisions. They help to establish evidence-based practices and to resolve conflicting research findings.<sup>1</sup>

However, the validity of MA conclusions relies on the rigour of the procedures that authors applied and are subject to a range of biases. A particularly salient feature that impacts the conclusion of the MA is the number of decisions and judgement calls that need to be made by the meta-analysist. Moreover, too many systematic reviews (SRs) and MAs are of low quality,<sup>1-5</sup> as evidenced by the fact that numerous studies have highlighted methodological weaknesses in the conduct of MAs. Specifically, they found the absence of a well-developed research protocol,<sup>6-8</sup> an inappropriate literature search,<sup>9-12</sup> flaws in the statistical analyses<sup>10 13-16</sup> and an insufficient assessment of the risk of bias of individual studies.<sup>10 17 18</sup>

To support researchers in the realisation and reporting of MAs, two tools are

commonly used. The first is 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses' (PRISMA), which was developed by Liberati *et al.*<sup>19</sup> It is a statement proposed to enhance the reporting and transparency of the SR and MA. The second is AMSTAR2 ('A Measurement Tool to Assess systematic Reviews'), developed by Shea *et al* in 2017,<sup>20</sup> which is a critical appraisal tool to help with the methodological development and evaluation of SRs and MAs.

It is important to determine whether MAs published in behavioural and social sciences are conducted well and are trustworthy and to determine their methodological weaknesses. The review of the methodology of MAs and the identification of current practices could help to improve the methodological quality of MAs.

Therefore, our current meta-research study attempts to address the following aims:

- ▶ To characterise the methodological characteristics of MAs indexed in PsycINFO according to AMSTAR2.
- ▶ To investigate the effect of the mention of PRISMA on the methodological quality of MAs according to AMSTAR2.
- ▶ To identify potential factors associated with the quality of MAs.

In this study, we made the hypothesis that the methodological quality of MAs indexed in PsycINFO was unsatisfactory using the AMSTAR2 tool and that the use of PRISMA could influence the presence of the different AMSTAR2 items. Specifically, we made the hypothesis that the MAs will present more often a satisfactory research question and inclusion criteria based on the components of a Population, Intervention, Comparison, Outcome (PICO) (item 1) if the MAs authors mention the PRISMA statement. This hypothesis was tested for each of the 16 AMSTAR2 items.

## METHODS

### Registration and protocol

We carried out this study in accordance with a research protocol, which is available on the Open Science Framework: <https://osf.io/hjybx/> or in online supplementary file 1. This study is the second part of a larger project assessing reporting and methodological quality of MAs.

### Samples, eligibility criteria and study selection

Our global methodology has previously been described.<sup>21</sup> Briefly, we wished to identify all MAs published in 2016 and indexed in PsycINFO. For that, we developed a systematic search to identify all MAs indexed in the electronic database PsycINFO (via Ovid) and published in 2016. This database was developed by the American Psychological Association and is specialised in the field of behavioural and social sciences. The electronic search strategy was developed with coauthors and the assistance of a skilled librarian. Then, we defined the eligibility criteria to conduct the study selection process. To be included in our sample, studies needed to be SR with a MA, indexed

in the PsycINFO database, published between 1 January 2016 and 31 December 2016, and published in English. In total, 2159 records were identified. Two authors (VL and CB) screened the title and abstracts of the retrieved studies in order to exclude irrelevant articles (n=1039) and to ensure that only the studies that met the eligibility criteria were selected (n=1120). Discrepancies in study selection were resolved by a third investigator. After the first selection process, to be able to investigate the effect of the mention of PRISMA on the methodological quality of MAs, we decided to have two samples with a minimum of 100 MAs in each group: one was composed of MAs claiming that they followed the PRISMA statement and the other included MAs that did not. To reach our sample goal, we randomly selected the full texts of the articles selected on the basis of their title and abstract, one by one, until we had a minimum of 100 articles per group. To do this, all articles references (n=1120) were indexed in an Excel file and randomly assigned to a number. Then, articles were ranked in ascending order. Afterward, two investigators, with the intervention of a third investigator in cases of disagreement, confirmed whether each article met the eligibility criteria, until a minimum of 100 studies per group were selected. A random sample of 206 eligible studies was drawn for this meta-research study. The selection procedure is illustrated in a flow chart in online supplementary file 2. The list of included and excluded studies can be found at <https://osf.io/hjybx/>

### Data extraction

To retrieve the data for our analyses, two investigators (VL and SA) independently extracted all relevant data from the full texts of all selected articles in a standardised Microsoft Excel spreadsheet. The extraction form had been pretested on ten MAs. Data extraction disagreements between the two investigators were resolved by discussion with the intervention of a third investigator if necessary. The inter-rater reliability between the two investigators was calculated with Cohen's kappa (median value with IQR of 0.66 (0.40–0.75)) and the Gwet's AC1 (median value with IQR of 0.77 (0.69–0.88)) both suggesting a substantial agreement.<sup>22</sup> Our primary concern was the methodological characteristics of the MAs. Furthermore, we extracted the data about the general characteristics of the MAs and the factors potentially associated with MA quality.

### Methodological characteristics appraisal

The methodological characteristics of the MAs were assessed using the tool AMSTAR2.<sup>20</sup> AMSTAR2 was a revision of the original AMSTAR instrument<sup>23</sup> developed by Shea *et al*, which was designed to appraise SRs and MAs. The relevance of all 11 original items was confirmed and some were refined. The AMSTAR2 tool is now composed of 16 items and is structured around the key sequential steps in the conduct of an MA. Each individual item is defined by a set of subitems to ensure that the item is completed. Each item was answered with a 'yes', 'partial

## Box 1 A Measurement Tool to Assess systematic Reviews 2 tool

### Critical domains

- ▶ Protocol registered before commencement of the review (item 2).
- ▶ Adequacy of the literature search (item 4).
- ▶ Justification for excluded studies (item 7).
- ▶ Risk of bias (RoB) assessed in individual studies being included in the review (item 9).
- ▶ Appropriateness of meta-analytical methods (item 11).
- ▶ Consideration of RoB when interpreting the results of the review (item 13).
- ▶ Assessment of the presence and likely impact of publication bias (item 15).

### Non-critical domains

- ▶ Research question and inclusion criteria based on the components of PICO (Population, Intervention, Comparison, Outcome) (item 1).
- ▶ Explanation for the selection of the study designs included in the review (item 3).
- ▶ Study selection performed in duplicate (item 5).
- ▶ Study extraction performed in duplicate (item 6).
- ▶ Description of the included studies in adequate detail (item 8).
- ▶ Report of the sources of funding for the included studies (item 10).
- ▶ Assessment of the impact of RoB in individual studies on the results of the meta-analyses (item 12).
- ▶ Explanation for any heterogeneity observed in the results (item 14).
- ▶ Report any potential sources of conflict of interest (item 16).

yes' or 'no' response, depending on whether the item was fulfilled. For example, when evaluating item 4, 'Did the review authors use a comprehensive literature search strategy?', to obtain a 'partial yes', it was required that the MA consulted at least two databases, provided the keywords and justified the publication restriction. To obtain a 'yes', it was required that the MA authors searched the reference lists of the included studies, searched study registries, consulted an expert, searched for grey literature and conducted the research within 24 months of completion of the review. To critically assess the methodological quality of MAs, the use of a global score is not recommended, and the authors of the tool advised classification of the MAs into four categories of quality: critically low, low, moderate and high. The suggested classification is based on the presence or absence of critical domains. The tool identifies seven critical weaknesses that should reduce confidence in the findings of a review and nine other items that are considered noncritical weaknesses, as presented in [box 1](#).

When the MA presented 'more than one critical flaw with or without noncritical weaknesses', the quality was considered critically low. When the review had 'one critical flaw with or without noncritical weaknesses', the quality was considered low. When the review had 'no critical flaws and more than one noncritical weaknesses', the quality was considered moderate. When the review had 'no critical flaws and ≤one noncritical weakness', the quality was considered high.

### General characteristics of the MAs and potential factors

From each study, some general characteristics of the MAs related to the journal, authors and included articles were extracted; these characteristics were the ones that we hypothesised could impact the methodological quality.

The article information included the mention of the use of PRISMA (Y/N), the mention of the use of a guideline other than PRISMA (Y/N), the availability of open access (Y/N), a protocol registration (Y/N), if the MA was a Cochrane study (Y/N), the presence of a search strategy (Y/N), restriction to the English language (Y/N), the use of statistical software (Y/N and which one), the number of studies included in the first MA, the assessment of the risk of bias in the individual studies (Y/N) and the tool used to assess the risk of bias and the design of the studies included in the MA.

The extracted author information included the number of authors, the continent and the country of the first author workplace, the H index of the first author and of the last author, the first author's experience with MAs (obtained from a search of Scopus to investigate the number of MA publications the author had previously coauthored), the affiliation of the first author to a university (Y/N), the contribution of the authors (Y/N), the declaration of the conflict of interest (Y/N) and the management of the conflict of interest.

The extracted journal information included the impact factor according to the 2016 Journal Citation Report from Thomson Reuters, the journal recommendation to use PRISMA obtained from the author instructions available in 2017 for each journal (Y/N) and whether there was an article word count limitation (Y/N, obtained from the author instructions for each journal available in 2017).

### Data analysis

We used descriptive statistics to assess the general characteristics of the MAs and to present the methodological quality of the MAs by showing compliance with AMSTAR2 and the potential factors associated with the quality of MAs. We summarised data as frequency and percentage values for categorical items and as median and P25–P75 values for continuous items. None of the quantitative variables followed a normal distribution. The distribution was considered normal if the data met three of the four following conditions: the mean was close to the median, the Shapiro-Wilk normality test yielded a  $p \geq 0.05$ , the curve of the variables followed the Gaussian distribution and the linearity of the QQ-Plots was respected. A univariate logistic regression was used to test the association between the explicit mention of PRISMA (Y/N, dependent variable) and the adherence of different AMSTAR2 items. Specifically, to evaluate the association between the mention of PRISMA and the quality of studies according to AMSTAR2, all AMSTAR2 items rated 'partial yes' (items 2, 4, 7, 8 and 9) were considered 'yes' for the analysis. Then, a univariate logistic regression without dichotomising the AMSTAR2 items was performed as a sensitivity analysis. Associations were quantified using ORs with 95%

CIIs. A Bonferroni correction was used to adjust the results for multiple testing (16 tests,  $p < 0.003$ ). All analyses were performed using SAS V.9.4 software.

### Patient and public involvement

There was no patient or public involvement in the whole process of conducting this research.

## RESULTS

### Search results

A total of 2159 potentially relevant MAs related to behavioural and social sciences were identified from PsycINFO during 2016. Of these, a random sample of 206 MAs was included in our analyses.

### General characteristics of the MAs

The main characteristics of the 206 MAs that qualified for this analysis are illustrated in [table 1](#). The majority of the MAs (67%) included more than 10 studies in their main analyses. Of the 206 studies, 97 (47%) included observational studies, and 60 (29%) included interventional studies. Reporting guidelines other than PRISMA were used by 23 (11%) MAs and included MOOSE<sup>24</sup> (17, 74%), Mars (2, 9%) and Quorum (1, 4%). Finally, most articles were not available for open access (90.3%), and only one was a Cochrane MA.

Written by 1 to 32 authors, most MAs came from either Europe (34%, with authors mainly coming from England and the Netherlands) or America (31.1%, with a large proportion of authors from the USA), followed by Asia (19.9%, where most MAs were conducted in China). The first MA authors had a median H index of 5 (2–11) with a median experience in MAs of 2 (1–5) and the last authors had a median H index of 22 (10–35). Almost all of the first authors were academics (91.3%). Of the 129 studies that declared the presence or absence of the conflicts of interest in our sample, 114 stated that the authors had no conflicts of interest to declare, and 15 described how they handled these conflicts.

The median impact factor of the journals in which the MAs were published was 3.3 (2.3–5.2). Additionally, nearly 30% of the MAs were published in a journal that recommended the use of PRISMA guidelines. In more than 63% of the MAs, the number of words in the article was limited.

### Methodological characteristics of the MAs

Across our sample of 206 MAs, according to the classification advised by AMSTAR2, 195 MAs were categorised as critically low quality, 8 as low quality, 2 as moderate quality and 1 as high quality. Only one MA<sup>25</sup> provided all the information on all seven critical domains assessed and was considered high-quality according to AMSTAR2. Two additional MAs<sup>26 27</sup> also provided all information on all seven critical domains assessed but had more than one noncritical weakness; they were considered moderate quality. The other MAs in our sample (98.5%) lacked

information in one or more critical domains and were considered low (4%) and critically low quality (94.5%) according to the classification advised by the AMSTAR2 tool ([box 1](#)).

In [figure 1](#), we summarise the AMSTAR2 results for our 206 MAs. The most important items that were the least respected by our sample were:

- ▶ An adequate information about the research protocol (item 2; yes: 8.3% and partial yes: 2.9%).
- ▶ A justification for the selection of the study design for the included studies (item 3; 10.2%).
- ▶ An adequate literature search (item 4; yes: 7.77% and partial yes: 36.9%).
- ▶ An adequate assessment of the risk of bias (item 9; yes: 31.5% and partial yes: 5.3%).
- ▶ Adequate reporting of the sources of funding for the studies included in the MA (item 10; only 4.4% reported this item).
- ▶ An adequate interpretation of the risk of bias (item 13; 23%).

However, some items were met by more than three quarters of the MAs:

- ▶ An appropriate research question with, ideally, the components of PICO (item 1; 85%).
- ▶ The use of appropriate methods for statistical analyses (item 11; 86.7%).
- ▶ A satisfactory explanation for any heterogeneity found in the results (item 14; 74.8%).

### Association of the explicit mention of PRISMA and methodological characteristics

The results of the univariate logistic regression that assessed the effect of the explicit mention of the PRISMA statement on the methodological characteristics of all AMSTAR2 items are presented in [figure 2](#). For the purpose of this analysis, all ‘partial yes’ items were considered ‘yes’. After applying the Bonferroni correction for multiple testing, almost half of the AMSTAR2 items were encountered with a significantly greater frequency in the MAs that explicitly mentioned PRISMA than in those that did not. The probability of having a good research question (item 1, OR 4.84; 95% CI 1.90 to 12.37) was significantly higher in the MAs with an explicit mention of PRISMA than in those not mentioning PRISMA. This observation was the same for some other items.

- ▶ Information about the research protocol (item 2, OR 8.58; 95% CI 2.46 to 29.90).
- ▶ Study selection in duplicate (item 5; OR 4.55; 95% CI 2.52 to 8.21).
- ▶ A detailed description of the included studies (item 8; OR 2.62; 95% CI 1.44 to 4.76).
- ▶ A satisfactory technique for assessing the risk of bias in individual studies (item 9; OR 4.48; 95% CI 2.43 to 8.27).
- ▶ An assessment of the potential impact of risk of bias in individual studies (item 12; OR 5.17; 95% CI 2.39 to 11.16).

**Table 1** General characteristics of the MAs

Characteristics	Category	N	No. (Per cent)	Median (P25–P75)
<b>Article</b>				
Mention of the use of a guideline other than PRISMA	Moose	23	17 (73.9)	
	Mars		2 (8.7)	
	Cochrane		1 (4.35)	
	Quorum		1 (4.35)	
	Strobe		1 (4.35)	
	Centre for reviews and dissemination		1 (4.35)	
Open access	Yes	206	20 (9.7)	
Protocol registration	Yes	206	15 (7.3)	
Cochrane MA	Yes	206	1 (0.5)	
Presence of a search strategy	Yes	206	81 (39.3)	
Presence of a linguistic bias	Yes	206	96 (46.6)	
Use of statistical software	Yes		170 (82.5)	
Statistical software used	CMA	193	87 (45.1)	
	STATA		30 (15.6)	
	Revman		29 (15)	
	SPSS		17 (8.8)	
	R		10 (5.2)	
	SAS		4 (2.1)	
	Other		17 (8.8)	
No. of studies included in the first MA	1–3 studies	206	11 (5.3)	
	4–9 studies		57 (27.7)	
	≥10 studies		138 (67)	
Assessment of the risk of bias (RoB) in individual studies	Yes	206	111 (53.9)	
Tool used to assess the risk of bias	RoB tool	95	36 (37.9)	
	Newcastle-Ottawa Scale (NOS)		14 (14.7)	
	Downs and Black		6 (6.3)	
	Jadad		5 (5.3)	
	Pedro		5 (5.3)	
	Quadas		5 (5.3)	
	Other		24 (25.3)	
Design of the included studies	Experimental		60 (29.1)	
	Observational		97 (47.1)	
	All types		18 (8.7)	
	Not specified		31 (15.1)	
<b>Authors</b>				
No. of authors	1 author	206	12 (5.8)	
	2–3 authors		60 (29.1)	
	4–6 authors		98 (47.6)	
	≥7 authors		36 (17.5)	
Continent of first author (workplace)	Africa	206	1 (0.5)	
	America		64 (31.1)	
	Asia		41 (19.9)	
	Europe		70 (34)	
	Oceania		30 (14.5)	

Continued



Table 1 Continued

Characteristics	Category	N	No. (Per cent)	Median (P25–P75)
Country of first author (workplace)	USA	206	49 (23.8)	
	Australia		26 (12.6)	
	China		22 (10.7)	
	England		22 (10.7)	
	Netherlands		15 (7.3)	
	Canada		13 (6.3)	
	Germany		11 (5.3)	
	Other (<11 reviews/country, 25 countries)		48 (23.3)	
H index of first author		205		5 (2–12)
H index of last author		195		22 (10–35)
Experience with MAs of the first author	Years	206		2 (1–5)
Affiliation of the first author	University	206	189 (91.8)	
Declaration of conflicts of interest	Yes	206	129 (62.6)	
Management of conflicts of interest	None	206	114 (55.3)	
	Described how they managed		15 (7.3)	
	Not indicated		77 (37.4)	
<b>Journal</b>				
Journal impact factor (2016)	0.0–5.0	200	148 (71.9)	
	5.1–10.0		45 (21.8)	
	10.1–15.0		1 (0.5)	
	>15.0		5 (2.4)	
	No impact factor		7 (3.4)	
Impact factor		200		3.3 (2.3–5.2)
PRISMA-endorsing journal	Yes	206	61 (29.6)	
Limitation of words	Yes	206	130 (63.1)	
<b>Methodological quality</b>				
AMSTAR 2 tool	High quality	206	1 (0.5)	
	Moderate quality		2 (1)	
	Low quality		8 (4)	
	Critically low quality		195 (94.5)	

AMSTAR2, A Measurement Tool to Assess systematic Reviews; MAs, meta-analyses; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

- ▶ Appropriate consideration of the risk of bias in primary studies when interpreting the results (item 13; OR 6.34; 95% CI 3.15 to 12.78).

The results of the sensitivity analysis, performed without dichotomising the responses modality of AMSTAR2 (yes, partial yes and no) using a logistic regression, showed similar results in online supplementary file 3.

#### Potential factors associated with the quality of MAs

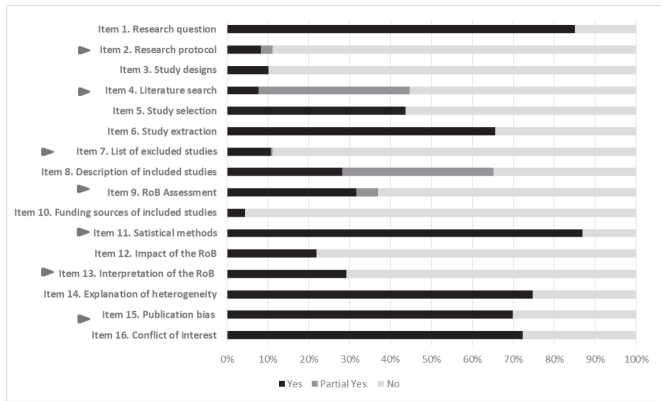
In our research protocol, we planned to identify the potential factors (impact factor, country, statistic software...) associated with the methodological quality of MAs according to the criteria advised by AMSTAR2. However, the data obtained did not allow us to identify

factors associated with good MAs, since almost all of the MAs (95%) were considered to be poor quality.

#### DISCUSSION

The credibility of MAs in research is based on the use of rigorous methodology. As is the case for individual studies, methodological choices may influence the results and conclusions of MAs.<sup>28</sup> With this study, we aim to provide a global overview of the methodological characteristics of MAs indexed in the PsycINFO database and to draw attention to specific deficiencies in conducting MAs.

The main objective of this study was to characterise the methodological quality of MAs indexed in PsycINFO



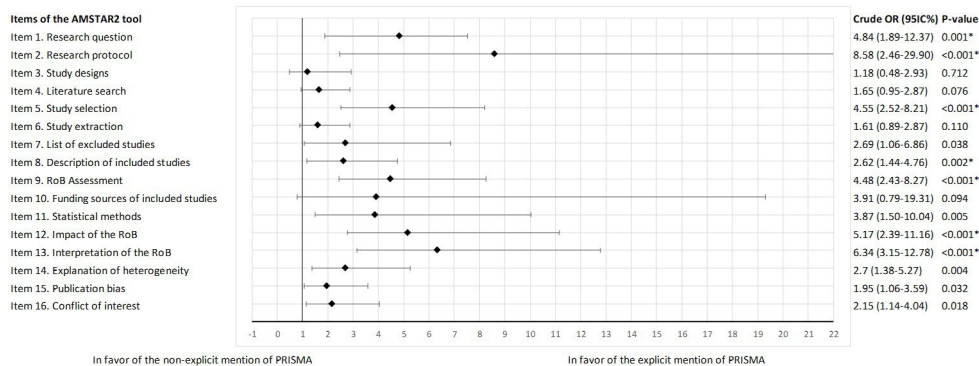
**Figure 1** Proportion of adherence to A MeaSurement Tool to Assess systematic Reviews (AMSTAR2) items. > : 7 critical domains identified by AMSTAR2. RoB, risk of bias.

according to the AMSTAR2 criteria. It appeared that the methodological quality of most of the sampled MAs was critically low, with many serious flaws. We found that the weaknesses were due to a lack of consistency in the methods used to perform the MAs in behavioural and social sciences.

- ▶ First, no more than 11% of MAs had a research protocol available. However, several scientists<sup>8 16 29</sup> highlighted the fact that an SR with an a priori research protocol was associated with increased quality and better elaborated and reported reviews. The many benefits of publishing a research protocol a priori include anticipating all the methodological steps, minimising the risk of bias, avoiding replicate studies and enhancing transparency.<sup>7</sup> These results should to be interpreted with caution because the registration of the research protocol is a relatively recent practice. However, the recommendation to use a research protocol to conduct an SR was already presented in the PRISMA statement in 2009 and in the first version of AMSTAR in 2007.
- ▶ Second, less than 37% of MAs provided a satisfactory literature search (according to AMSTAR2, satisfaction of the first part of item 4 included a search in

a minimum of two databases, a list of keywords and a justification of the publication restriction) and less than 8% provided a complete search (according to AMSTAR2, satisfaction of the last part of item 4 included a search of the reference lists of included studies, a search of study registries, a search for grey literature, the consultation of an expert and conducting the research within 24 months of completion of the review). Our results also showed that very few studies implemented all available methods to find all the individual studies, as also reported by Ahn *et al.*<sup>10</sup> The search strategy is an essential step of the MA process since the comprehensiveness and completeness of the search<sup>3 30</sup> is dependent on this strategy. Furthermore, other scientists have highlighted the need to improve research strategies for more comprehensive MAs.<sup>9 30</sup>

- ▶ Third, the presence of a list of studies excluded at the step of full-text selection was an AMSTAR requirement that was very rarely found in non-Cochrane MAs, as evidenced by the fact that only 11% of our sample provided the excluded studies list and related reasons of exclusion.
- ▶ Finally, only one-third of MAs used a satisfactory technique for assessing the risk of bias in the individual studies included in the MA. Furthermore, consistent with previous studies,<sup>17 31</sup> only one-fifth of our sample assessed the potential impact of the risk of bias in individual studies on the results of the MA, and less than one-third of MAs accounted for the risk of bias when interpreting the results. More specifically, Oliveras *et al* identified several possible methods to take into account the risk of bias of the studies included in the research synthesis when exploring the association between the effects size and the risk of bias, such as sensitivity analyses, cumulative MAs in order of quality, quality-based subgroup analyses, meta-regression and bias adjustment models.<sup>17</sup> However, there is still a lack of guidance to incorporate these risk of bias assessments into MAs.<sup>17 18 32</sup>



**Figure 2** Impact of the explicit mention of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) on the methodological characteristics of MAs: the non-explicit mention of PRISMA group versus the explicit mention of PRISMA group. \*Items statistically significant with the Bonferroni correction for multiple testing ( $p \leq 0.003$ ). AMSTAR2, MeaSurement Tool to Assess systematic Reviews; MAs, meta-analyses; RoB, risk of bias.

Regarding our second research question, the explicit mention of PRISMA suggested an improved methodological quality of MAs. Almost half of the items in the AMSTAR2 tool were significantly more frequent in the MAs that explicitly mentioned PRISMA than in those that did not. However, it is recognised that the accuracy of ORs may be variable due to variations in CIs widths between items. This difference can be explained by the variation in occurrence of the events of the different items. Even so, the explicit mention of PRISMA suggested a positive influence on the methodological quality of MAs indexed in PsycINFO. Moreover, the completeness of reporting helped with the evaluation of the robustness of MA results, but MA reporting still needs to be improved.<sup>21 28 31 33 34</sup>

Concerning the methodological quality of MAs and the potentially associated factors, no conclusion could be drawn. As identified in our sample, with the classification suggested by the AMSTAR2 tool, the majority of MAs were considered low quality. Furthermore, even though potential factors could be identified in relation to the quality of MAs, some characteristics of the MAs were still suggested to be interesting. The only MA considered high quality according to AMSTAR2<sup>25</sup> was a Cochrane collaboration review. This collaboration is considered the reference for conducting a meta-analysis due to its methodological requirements. The two other studies considered moderate quality<sup>26 27</sup> had the same first author and were published in journals with high impact factors of 6.442 and 14.176.

Our results also highlight that AMSTAR2 is subject to floor effects because 95% of our sample was rated as critically low, which is the lowest category proposed by the tool. The discriminative capacity of this tool is not optimal, and the relevance of the choice of critical or noncritical items and the composition of these items can raise some questions. For example, one of AMSTAR's requirements for item 4, 'comprehensive literature search strategy', is the presence of a publication restrictions' justification,<sup>20</sup> yet only a few studies from our sample of MAs mention it explicitly. Dechartres *et al* stressed the association between publication characteristics and effect estimates<sup>11</sup> and confirmed that restricting a search to published studies may lead to an overestimation of treatment effects with possible repercussions on the conclusion of the MA. In contrast, the effect of the language bias (narrowing the selection to articles written in English only) on the results of an MA is controversial.<sup>11 12 35</sup> This is consistent with the literature, as the importance of this criterion (publication restriction justification) on the methodological quality of MA is still being questioned. However, this criterion played an important role in the assessment of MA quality with AMSTAR2. In contrast, items concerning the use of appropriate methods for the statistical combination of results (item 11) and the assessment of heterogeneity (item 14) may not be precise enough. For example, there is no item concerned with the use of one-way sensitivity analyses to test the robustness of the results. This failure could lead to overestimation of the use of relevant

statistical methods in our sample, as evidenced by the fact that 87% of our sample used appropriate methods for the statistical combination of results (item 11). Our results are consistent with the study conducted by Ahn<sup>10</sup> but contradict previous studies that have highlighted several flaws in the application and interpretation of statistical analyses in MA.<sup>13 14 28 36</sup> Page *et al* identified some mistakes in the use of adequate statistical models, the sufficient exploration of subgroup analyses and sensitivity analyses.<sup>14</sup> Consequently, additional investigations of the AMSTAR2 tool should be encouraged to improve it.

To the best of our knowledge, this study is the first to evaluate the methodological characteristics of MAs indexed in PsycINFO with the newly developed AMSTAR2 tool.<sup>20</sup> Our study has some limitations that should be taken into account. First, only a random sample of studies indexed in PsycINFO, published in 2016 and in English, was included. Therefore, we cannot generalise our finding to MAs published in other years, in other languages or in other databases. Further researches evaluating other databases and considering different years of publication could be relevant as new perspective. Second, the methodological quality of MAs depends on the descriptions made by the authors in the publication and may not be an accurate reflection of what actually occurred during the review process. Finally, there are some limitations regarding the use of AMSTAR2 as a tool to evaluate the methodological quality of MAs, which is rigorous and comprehensive tool. First, considering that the MAs in this study were published before 2017, the quality of MAs did not meet the new quality standards. Second, our agreement coefficient indicated a substantial agreement, indeed subjectivity related to data extraction is limited since all data has been extracted in duplicate. The Gwet's ACI was presented along with the Cohen's Kappa. Although Cohen's kappa is more widely used, Gwet's ACI is a more robust alternative (less sensitive to data distribution and number of observation).<sup>22</sup> Moreover, using AMSTAR2, we can investigate the methodological characteristics used to conduct the study (eg, The authors consulted two databases to be the most exhaustive) but we cannot investigate the adequacy of the methodological choice to the specific context of the review (eg, did the authors consult the appropriate databases to answer their research questions). Finally, without a priori excellent expertise in the research question of the study, the use of AMSTAR2 ensures a partial assessment of the research quality. No tool is perfect but AMSTAR2 allows us to have an overview of the methodological characteristic of MAs.

## CONCLUSION

This research contributes to raising awareness among researchers about flaws in MAs published in behavioural and social sciences fields, which hopefully increases the adoption of more rigorous research practices. It is clear that meta-analytical practices can be improved. If some critical items identified with AMSTAR2 were given more



consideration, the published MAs could make a leap in methodological quality and thus gain robustness and reliability. Furthermore, validation of the AMSTAR2 tool and the relevance of the choice of critical or noncritical items established to rate the overall confidence in the results of MAs with AMSTAR 2 opens new leads for further investigation.

**Contributors** Authors' contribution: VL, CB, ET and OB conceived the study; VL, CB and SA participated in data collection; VL, CB and OB analysed and interpreted data; VL, CB, ET and OB corrected the manuscripts. All coauthors read and approved the final version of the manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository: <https://osf.io/hjybx/>.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Victoria Leclercq <http://orcid.org/0000-0002-7595-0268>

#### REFERENCES

- Gurevitch J, Koricheva J, Nakagawa S, *et al*. Meta-Analysis and the science of research synthesis. *Nature* 2018;555:175–82.
- 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.
- Dieckmann NF, Malle BF, Bodner TE. An empirical assessment of meta-analytic practice. *Review of General Psychology* 2009;13:101–15.
- Zhang H, Han J, Zhu Y-B, *et al*. Reporting and methodological qualities of published surgical meta-analyses. *J Clin Epidemiol* 2016;70:4–16.
- Ioannidis JPA. The mass production of redundant, misleading, and Conflicted systematic reviews and meta-analyses. *Milbank Q* 2016;94:485–514.
- Nuijten MB. Practical tools and strategies for researchers to increase replicability. *Dev Med Child Neurol* 2019;61:535–9.
- Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in prospero: 30,000 records and counting. *Syst Rev* 2018;7:32.
- Sideri S, Papageorgiou SN, Eliades T. Registration in the International prospective register of systematic reviews (prospero) of systematic review protocols was associated with increased review quality. *J Clin Epidemiol* 2018;100:103–10.
- Franco JVA, Garrote VL, Escobar Liquitay CM, *et al*. Identification of problems in search strategies in Cochrane reviews. *Res Synth Methods* 2018;9:408–16.
- Ahn S, Ames AJ, Myers ND. A review of meta-analyses in education: methodological strengths and weaknesses. *Rev Educ Res* 2012;82:436–76.
- Dechartres A, Atal I, Riveros C, *et al*. Association between publication characteristics and treatment effect estimates: a Meta-epidemiologic study. *Ann Intern Med* 2018;169:385.
- Walpole SC. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. *J Clin Epidemiol* 2019;111:127–34.
- Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods* 2017.
- 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.
- Stanley TD, Carter EC, Doucouliagos H. What meta-analyses reveal about the replicability of psychological research. *Psychol Bull* 2018;144:1325–46.
- Cafri G, Kromrey JD, Brannick MT. A Meta-Meta-Analysis: empirical review of statistical power, type I error rates, effect sizes, and model selection of meta-analyses published in psychology. *Multivariate Behav Res* 2010;45:239–70.
- Oliveras I, Losilla J-M, Vives J. Methodological quality is underrated in systematic reviews and meta-analyses in health psychology. *J Clin Epidemiol* 2017;86:59–70.
- Johnson BT, Low RE, MacDonald HV. Panning for the gold in health research: incorporating studies' methodological quality in meta-analysis. *Psychol Health* 2015;30:135–52.
- Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6:e1000100.
- 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.
- Leclercq V, Beaudart C, Ajamieh S, *et al*. Meta-Analyses indexed in PsycINFO had a better completeness of reporting when they mention PRISMA. *J Clin Epidemiol* 2019;115:46–54.
- Wongpakaran N, Wongpakaran T, Wedding D, *et al*. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol* 2013;13:1–7.
- 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;7:10.
- Stroup DF. Meta-Analysis of observational studies in Epidemiology. A proposal for reporting. *JAMA* 2008;283.
- Coxeter P, Del Mar CB, McGregor L, *et al*. Interventions to facilitate shared decision making to address antibiotic use for acute respiratory infections in primary care. *Cochrane Database Syst Rev* 2015;11:CD010907.
- Cortese S, Ferrin M, Brandeis D, *et al*. Neurofeedback for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *J Am Acad Child Adolesc Psychiatry* 2016;55:444–55.
- Cortese S, Moreira-Maia CR, St Fleur D, *et al*. Association between ADHD and obesity: a systematic review and meta-analysis. *Am J Psychiatry* 2016;173:34–43.
- Aytug ZG, Rothstein HR, Zhou W, *et al*. Revealed or concealed? transparency of procedures, decisions, and judgment calls in meta-analyses. *Organ Res Methods* 2012;15:103–33.
- Allers K, Hoffmann F, Mathes T, *et al*. Systematic reviews with published protocols compared to those without: more effort, older search. *J Clin Epidemiol* 2018;95:102–10.
- Yaylali IE, Alaçam T, Alaşam T. Critical Assessment of Search Strategies in Systematic Reviews in Endodontics. *J Endod* 2016;42:854–60.
- Hohn RE, Slaney KL, Tafreshi D. Primary study quality in psychological meta-analyses: an empirical assessment of recent practice. *Front Psychol* 2018;9:2667.
- Ahn S, Becker BJ. Incorporating quality scores in meta-analysis. *Journal of Educational and Behavioral Statistics* 2011;36:555–85.
- Schalken N, Rietbergen C. The reporting quality of systematic reviews and meta-analyses in industrial and organizational psychology: a systematic review. *Front Psychol* 2017;8:1395.
- Hutton B, Wolfe D, Moher D, *et al*. Reporting guidance considerations from a statistical perspective: overview of tools to enhance the rigour of reporting of randomised trials and systematic reviews. *Evid Based Ment Health* 2017;20:46–52.
- Robson RC, Pham Ba', Hwee J, *et al*. Few studies exist examining methods for selecting studies, abstracting data, and appraising quality in a systematic review. *J Clin Epidemiol* 2019;106:121–35.
- Augusteijn HEM, van Aert RCM, van Assen MALM. The effect of publication bias on the Q test and assessment of heterogeneity. *Psychol Methods* 2019;24:116–34.