BMJ Open

The methodological quality of meta-analyses published in psychology and related fields: leads for enhancements

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-036349
Article Type:	Original research
Date Submitted by the Author:	12-Dec-2019
Complete List of Authors:	Leclercq, Victoria; University of Liege, Division of Public Health, Epidemiology and Health Economics Beaudart, Charlotte; University of Liege, Division of Public Health, Epidemiology and Health Economics Ajamieh, Sara; University of Liege, Division of Public Health, Epidemiology and Health Economics Tirelli, Ezio; University of Liege, Department of Psychology Bruyère, Olivier ; University of Liege, Division of Public Health, Epidemiology and Health Economics
Keywords:	EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

The methodological quality of meta-analyses published in psychology and related fields: leads for enhancements

Leclercq Victoria[1], Beaudart Charlotte[1], Ajamieh Sara[1,2], Tirelli Ezio[2], Bruyère Olivier[1]

[1]Division of Public Health, Epidemiology and Health Economics, University of Liège, Belgium

[2]Department of Psychology, University of Liège, Belgium

victoria.leclercq@uliege.be

c.beaudart@uliege.be

sajamieh@uliege.be

ezio.tirelli@uliege.be

olivier.bruyère@uliege.be

ABSTRACT

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

Design We evaluated a random sample of 206 MAs indexed in the PsycINFO database in 2016 through a cross-sectional study. Two authors independently extracted the methodologic characteristics of all MAs and checked their quality according to the 16 items of the AMSTAR2 (A MeaSurement Tool to Assess systematic Reviews) tool for MA critical appraisal. Moreover, we investigated the effect of mentioning 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.

Conclusions 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.

ARTICLE SUMMARY

- Some studies have highlighted methodological weaknesses in the conduct of systematic reviews (SRs) and meta-analysis (MAs) and we search to have an overview of methodological practice of MAs indexed in PsycINFO according to the tool AMSTAR2 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 PRISMA statement on the methodological quality of MAs**.**
- A sample of 206 Mas indexed in PsycINFO in 2016 and published in English was analyzed.

1 INTRODUCTION

Since the definition of meta-analyses (MAs) being introduced by Glass in 1976, MAs conducted in psychology and related fields have increased rapidly in number. There were more than 30 000 MAs indexed in 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[1].

However, the validity of MA conclusions relies upon the rigor 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 judgment calls that need to be made by the meta-analysist. Moreover, too many systematic reviews (SRs) and MAs are of low quality[1–5], 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[6–8], an inappropriate literature search[9–12], flaws in the statistical analyses[10,13–16] and an insufficient assessment of the risk of bias of individual studies[10,17,18].

To support researchers in the realization and reporting of MAs, two tools are commonly used. The first is PRISMA ("Preferred Reporting Items for Systematic Reviews and Meta-Analyses"), which was developed in 2009 by Liberati et al[19]. 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[20], 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 psychology 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 characterize the methodological characteristics of MAs published in psychology and related fields 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.

2 METHODS

Registration and protocol

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

Patients and public involvement

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

Samples, eligibility criteria and study selection

Our global methodology has previously been described[21]. 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 specialized in the field of behavioral 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 systematic review with a MA, indexed in the PsycINFO database, published between January 01, 2016, and December 31, 2016, and published in English. Two authors (V.L & C.B) screened the title and abstracts of the retrieved studies in order to exclude irrelevant articles and to ensure that only the studies that met the eligibility criteria were selected. 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 article selected on the basis of title and abstract one by one until we had a minimum of 100 articles per group. Two investigators, with the intervention of a third investigator in cases of disagreement, confirmed whether each article met the eligibility criteria. In the end, a random sample of 206 eligible studies was drawn for this meta-research study. 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 & SA) independently extracted all relevant data from the full texts of all selected articles in a standardized Microsoft Excel spreadsheet. The extraction form had been pretested on ten MAs. Data extraction disagreements between the two investigators were resolved by discussion (Median Kappa: 0.56±0.22), with the intervention of a third investigator if necessary. 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 "A MeaSurement Tool to Assess systematic Reviews 2"[20]. AMSTAR2 was a revision of the original AMSTAR instrument[22] developed by Shea et al in 2007, 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 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 2 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 gray 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 4 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 7 critical weaknesses that should reduce confidence in the findings of a review and 9 other items that are considered noncritical weaknesses, as presented in Table 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 hypothesized 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, 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 (JCR) from Thomson Reuters, the journal recommendation to use PRISMA obtained from the author instructions for each journal (Y/N) and whether there was an article word count limitation (Y/N, obtained from the author instructions for each journal).

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 summarized 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 not normal if the data met fewer than 3 of the 4 following conditions: the mean was close to the median, the Shapiro-Wilk normality test yielded a p-value ≤ 0.05, the curve of the variables followed the normal (or 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. Associations were quantified using odds ratios with 95% confidence intervals. All analyses were performed using SAS 9.4 software.

3 RESULTS

Search results

A total of 2159 potentially relevant MAs related to psychology and its related fields 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 2. 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 [23] (17, 74%), Mars (2, 9%) and Quorom (1, 4%). Finally, most articles were not available for open access (90.3%), and only one was a Cochrane MA.

Written by one 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 categorized as critically low quality, 8 as low quality, 2 as moderate quality and 1 as high quality. Only one MA [24] provided all the information on all 7 critical domains assessed and was considered high quality according to AMSTAR2. Two additional MAs [25,26] also provided all information on all 7 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 the classification advised by the AMSTAR2 tool (Table 1).

In Figure 1, we summarize the AMSTAR2 results for our 206 MAs. The most important items that were the least respected by our sample were:

- 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". Three-quarters 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: 1.89-12.37) 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: 2.46-29.90);
- study selection in duplicate (item 5; OR: 4.55; 95%CI: 2.52-8.21);
- a list of excluded studies (item 7; OR: 2.69; 95%CI: 1.06-6.86);
- a detailed description of the included studies (item 8; OR: 2.62; 95%CI: 1.44-4.76);
- a satisfactory technique for assessing the risk of bias in individual studies (item 9; OR: 4.48; 95%CI: 2.43-8.27);
- an appropriate method for the statistical combination of results (item 11; OR: 3.87; 95%CI: 1.50-10.04);
- an assessment of the potential impact of risk of bias in individual studies (item 12; OR: 5.17; 95%CI: 2.39-11.16);
- appropriate consideration of the risk of bias in primary studies when interpreting the results (item 13; OR: 6.34; 95%CI: 3.15-12.78);
- a satisfactory explanation for any heterogeneity found in the results (item 14; OR: 2.7; 95%CI: 1.38-5.27);
- an adequate investigation of publication bias (item 15; OR: 1.95; 95%CI: 1.06-3.59)
- a report of any potential conflict of interest sources (item 16; OR: 2.15; 95%CI: 1.14-4.04).

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.

4 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[27]. 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 characterize the methodological quality of MAs indexed in PsycINFO 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 psychology and related fields.

- First, no more than 11% of MAs had a research protocol available. However, several scientists[8,16,28] 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, minimizing the risk of bias, avoiding replicate studies and enhancing transparency[7]. 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 a systematic review 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 2 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 gray 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[10]. The search strategy is an essential step of the MA process since the comprehensiveness and completeness of the search[3,29]

is dependent on this strategy. Furthermore, other scientists have highlighted the need to improve research strategies for more comprehensive MAs[9,29].

- 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[17,30], 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 and her team 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[17]. However, there is still a lack of guidance to incorporate these risk of bias assessments into meta-analyses[17,18,31].

Regarding our second research question, the explicit mention of PRISMA suggested an improvement of the methodological quality of MAs. Three-quarters of items in the AMSTAR2 tool, including 6/7 of the critical items, were significantly more frequent in the MAs that explicitly mentioned PRISMA than in those that did not. However, it is recognized 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[21,27,30,32,33].

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[24] 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[25,26] 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[20], yet only a few studies from our sample of MAs mention it explicitly. Dechartres and her team stressed the association between publication characteristics and effect estimates[11] 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[11,12,34]. 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[10] but contradict previous studies that have highlighted several flaws in the application and interpretation of statistical analyses in MA[13,14,27,35]. Page et al identified some mistakes in the use of adequate statistical models, the sufficient exploration of subgroup analyses and sensitivity analyses[14]. 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 published in psychology and related fields with the newly developed AMSTAR2 tool[20]. 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 generalize our finding to MAs published in other years, in other

languages or in other databases. 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. Moreover, using AMSTAR2, we can investigate the methodological characteristics used to conduct the study (e.g. 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 (e.g. 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.

5 CONCLUSION

This research contributes to raising awareness among researchers about flaws in MAs published in psychology and related fields, which hopefully increases the adoption of more rigorous research practices. It is clear that meta-analytic 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.

KEYWORDS

Meta-analyses – AMSTAR – meta-research – methodological quality – PRISMA

FOOTNOTES

Conflict of interest and funding: The authors declare no conflict of interest. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author contributions: VL, CB, ET and OB conceived the study; VL, CB and SA participated in data collection; VL, CB and OB analyzed and interpreted data; VL, CB, ET and OB corrected the manuscripts. All co-authors read and approved the final version of the manuscript.

Data sharing statement: Data sets are available on request from the corresponding author.

BIBLIOGRAPHY

1. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nat Publ Gr*. 2018;555. doi:10.1038/nature25753.
2. 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.
3. Dieckmann NF, Malle BF, Bodner TE. An Empirical Assessment of Meta-Analytic Practice. 2009. doi:10.1037/a0015107.
4. Zhang H, Han J, Zhu Y-B, et al. Reporting and methodological qualities of published surgical meta-analyses. 2016. doi:10.1016/j.jclinepi.2015.06.009.
5. Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q*. 2016;94(3):485-514. doi:10.1111/1468-0009.12210.
6. Nuijten MB. Practical tools and strategies for researchers to increase replicability. *Dev Med Child Neurol*. October 2018. doi:10.1111/dmcn.14054.
7. Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Syst Rev*. 2018;7(1). doi:10.1186/s13643-018-0699-4.
8. 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;100:103-110. doi:10.1016/J.JCLINEPI.2018.01.003.
9. Franco JVA, Garrote VL, Escobar Liquitay CM, Vietto V. Identification of problems in search strategies in Cochrane Reviews. *Res Synth Methods*. 2018;9(3):408-416. doi:10.1002/jrsm.1302.
10. Ahn S, Ames AJ, Myers ND. A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Rev Educ Res*. 2012;82(4):436-476. doi:10.3102/0034654312458162.
11. Dechartres A, Atal I, Riveros C, Meerpohl J, Ravaud P. Association Between Publication Characteristics and Treatment Effect Estimates. *Ann Intern Med*. 2018;169(6):385. doi:10.7326/M18-1517.
12. Walpole SC. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. *J Clin Epidemiol*. 2019;111:127-134. doi:10.1016/j.jclinepi.2019.03.004.
13. Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods*. 2017. doi:10.1002/jrsm.1240.
14. 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.
15. Stanley TD, Carter EC, Doucouliagos H. What meta-analyses reveal about the replicability of psychological research. *Psychol Bull*. 2018;144(12):1325-1346. doi:10.1037/bul0000169.
16. 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(2):239-270. doi:10.1080/00273171003680187.
17. 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. doi:10.1016/j.jclinepi.2017.05.002.
18. Johnson BT, Low RE, MacDonald H V. Panning for the gold in health research: Incorporating studies' methodological quality in meta-analysis. *Psychol Health*. 2015;30(1):135-152. doi:10.1080/08870446.2014.953533.
19. 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(7):e1000100. doi:10.1371/journal.pmed.1000100.
20. 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:j4008. doi:10.1136/bmj.j4008.
21. 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. doi:10.1016/J.JCLINEPI.2019.06.014.
22. 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. doi:10.1186/1471-2288-7-10.
23. Stroup DF. Meta-analysis of Observational Studies in Epidemiology<SUBTITLE>A Proposal for Reporting</SUBTITLE>. *JAMA*. 2003;283(15):2008. doi:10.1001/jama.283.15.2008.
24. Coxeter P, Del Mar CB, McGregor L, Beller EM. Interventions to facilitate shared decision making to address antibiotic use for acute respiratory infections in primary care. *Cochrane database Syst Rev*. 2015;11(11):CD010907. doi:10.1002/14651858.CD010907.pub2.www.cochranelibrary.com.
25. 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(6):444-455. doi:10.1016/j.jaac.2016.03.007.
26. Cortese S, Moreira-Maia CR, St Fleur D, Morcillo-Peñalver C, Rohde LA, Faraone S V. Association between ADHD and obesity: A systematic review and meta-analysis. *Am J Psychiatry*. 2016;173(1):34-43. doi:10.1176/appi.ajp.2015.15020266.
27. Aytug ZG, Rothstein HR, Zhou W, Kern MC. Revealed or concealed? transparency of procedures, decisions, and judgment calls in meta-analyses. *Organ Res Methods*. 2012;15(1):103-133. doi:10.1177/1094428111403495.
28. Allers K, Hoffmann F, Mathes T, Pieper D. Systematic reviews with published protocols compared to those without: more effort, older search. *J Clin Epidemiol*. 2018;95:102-110. doi:10.1016/J.JCLINEPI.2017.12.005.
29. Yaylali IE, Alaçam T. Critical Assessment of Search Strategies in Systematic Reviews in Endodontics. *J Endod*. 2016;42:854-860. doi:10.1016/j.joen.2016.02.018.
30. Hohn RE, Slaney KL, Tafreshi D. Primary study quality in psychological meta-analyses: An empirical assessment of recent practice. *Front Psychol*. 2019;9(JAN). doi:10.3389/fpsyg.2018.02667.
31. Ahn S, Becker BJ. Incorporating Quality Scores in Meta-Analysis. *J Educ Behav Stat*. 2011;36(5):555-585. doi:10.3102/1076998610393968.
32. 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(AUG):1395. doi:10.3389/fpsyg.2017.01395.
33. Hutton B, Wolfe D, Moher D, Shamseer L. 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(2):46-52. doi:10.1136/eb-2017-102666.
34. Robson RC, Pham B, 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;106:121-135. doi:10.1016/j.jclinepi.2018.10.003.
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(1):116-134. doi:10.1037/met0000197.

Table 1. AMSTAR2 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 assessed in individual studies being included in the review (item 9)
- Appropriateness of meta-analytical methods (item 11)
- Consideration of risk of bias when interpreting the results of the review (item 13)
- Assessment of the presence and likely impact of publication bias (item 15)

Noncritical domains

- Research question and inclusion criteria based on the components of PICO (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 MA (item 12)
- Explanation for any heterogeneity observed in the results (item 14)
- Report any potential sources of conflict of interest (item 16)

Table 2. General characteristics of the MAs

	Characteristics	Category	n	Number (Percent)	Median (P25-P75)
Article					
	Mention of the use of a guideline other than PRISMA	Moose	23	17 (73.9)	
		Mars		2 (8.7)	
		Cochrane		1 (4.35)	
		Quorom		1 (4.35)	
		Strobe		1 (4.35)	
		Center 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)	
	Number of studies included in the first MA	1-3	206	11 (5.3)	
		4-9		57 (27.7)	
		≥10		138 (67)	
	Assessment of the risk of bias in individual studies	Yes	206	111 (53.9)	
	Tool used to assess the risk of bias	RoB tool	95	36 (37.9)	
		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)	
Authors					
	Number of authors	1	206	12 (5.8)	
		2-3		60 (29,1)	
		4-6		98 (47.6)	
		≥7		36 (17.5)	
	Continent of first author	Africa	206	1 (0.5)	
		America		64 (31.1)	
		Asia		41 (19.9)	
		Europe		70 (34)	
		Oceania		30 (14.5)	
	Country of first author	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)	
Journal					
	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)	
Methodological quality					
	AMSTAR 2 tool	High quality	206	1 (0.5)	
		Moderate quality		2 (1)	
		Low quality		8 (4)	
		Critically low quality		195 (94.5)	

Figure 1. Proportion of adherence to AMSTAR 2 items. ► : 7 critical domains identified by Amstar 2.

Figure 2. Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group

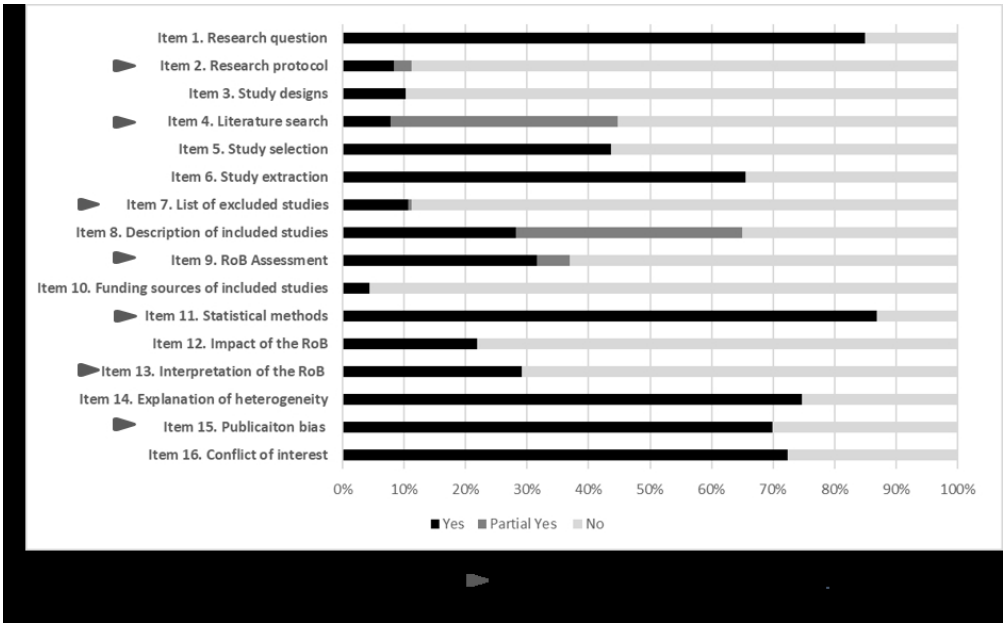


Figure 1. Proportion of adherence to AMSTAR 2 items. »: 7 critical domains identified by Amstar 2.

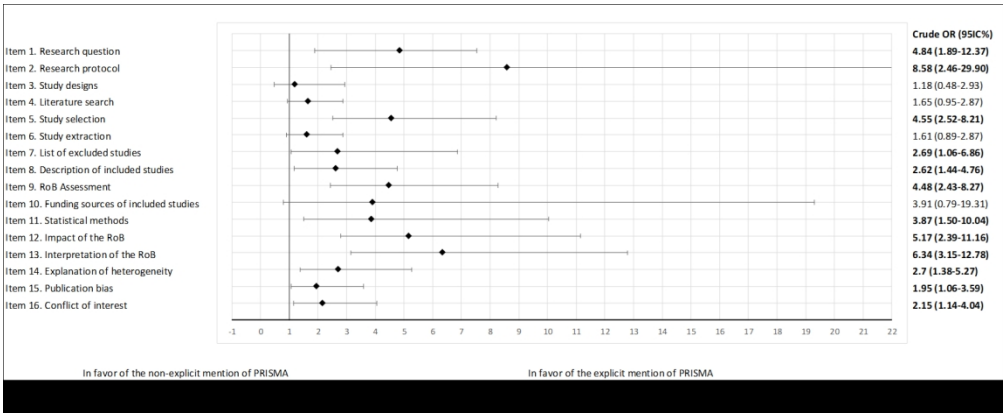


Figure 2. Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group

Register OSF

Study information

Title

Assessment of the reporting and methodological qualities and associated factors of a sample of meta-analyses recently indexed in PsycINFO (2016).

Authors

Victoria Leclercq – Charlotte Beaudart – Véronique Rabenda – Sara Ajamieh - Ezio Tirelli – Olivier Bruyère

Background

For scientists, searching for current best evidence has become a real challenge for scientists given the quasi limitless number of published articles (more than 1 270 000 in 2014, according to Thomson Reuteur's Web of Science). When facing a problematic implying a decision, scientists need documents and results oh high and reliable scientific value, as promoted by the evidence-based medicine movement (EBM). EBM is defined as the practice of medicine-based on knowledge and understanding of the literature in order to support clinical decisions (Guyatt et al. 2015). Following evidence hierarchy of EBM, systematic reviews (SRs) and meta-analyses (MAs) are considered the best level of evidence. Nowadays, in diverse disciplines, many researchers base their own research on the results of these SRs and MAs. The Cochrane collaboration adopted the definition of Antman (1992) and Oxman (1993) for the SR: "A systematic review attempts to collate all empirical evidence that pre-specified eligibility criteria in order to answer a specific research question" (Higgins & Green 2011) and the definition of Glass (1976) for MA : "Meta-analysis is the use of statistical methods to summarize the results of independent studies" (Higgins & Green 2011).

Some researchers have highlighted an increase in the publication rate of 2728% for SR (1024 articles in 1991 and 28 959 in 2014) and 2635% for MA (334 articles in 1991 and 9135 in 2014) (Ioannidis 2016). Several reasons could explain this phenomenon. In particular, fewer resources are necessary to perform SRs and MAs, which generally (yield) are worth high citation rates (contributing to increase the impact factor of the journal where they are published (Ioannidis 2016). However, an increasing number of studies have highlighted weaknesses in the design, conduct, analysis, and reporting of MAs published in many scientific fields (Zhu et al. 2016; Cullis et al. 2017; Peters et al. 2015; Zhang et al. 2016; Gagnier & Kellam 2013).

Two tools have been developed to evaluate the quality of the methodology ((AMSTAR, "A Measurement Tool to Assess systematic Reviews" (Shea et al. 2007)) and recently its update (AMSTAR 2 (Shea et al. 2017))) and another one for the quality of the reporting (PRISMA, "Preferred Reporting Items for Systematic Reviews and Meta-analyses" ((Moher et al. 2009)) of SRs and MAs. AMSTAR, a 11-item measuring tool aiming to assess the methodologic quality of MAs (Zhang et al. 2016), has been shown to be reliable and valid (Shea et al. 2009). AMSTAR 2, a 16-item measuring tool aiming to assess the methodologic quality of MAs of randomized and no randomized studies (Shea et al. 2017). PRISMA comprises a list of 27 items that are recommended to be used in the reporting of a MA in order to ensure that the article contains all relevant information (Moher et al. 2009). Several studies have already evaluated the quality of MAs published in specific medical fields such as surgery (Cullis et al. 2017; Zhang et al. 2016), depression (Zhu et al. 2016), orthopaedic (Gagnier & Kellam 2013) or even otorhinolaryngologic disorders (Peters et al. 2015). To the best of our knowledge, there are no such studies available in the field of psychological science.

In line with the EBM movement, the American Psychological Association (APA) has defined in 2006 the movement of Evidence-Based Practice in Psychology with the purpose "to promote effective psychological practice and enhances public health by applying empirically supported principles of psychological assessment, case formation, therapeutic relationship, and intervention"(American Psychological Association 2006). The American Psychological Association has brought out some benefits to the use of Reporting Standards whose the salutary effect on the way research has been conducted (Cooper 2008). The PRISMA statement could also have a positive effect on the methodological quality of the studies.

A growing meta-research literature has assessed the quality of empirical and experimental psychological studies in often large samples of articles (Ioannidis 2012; Bakker & Wicherts 2011; Oliveras et al. 2017; Stanley et al. 2017). It has revealed and quantified numerous methodological deficiencies, such as an inappropriate use of statistics, high rates of statistical mistakes, a frequent lack of statistical power (along with the neglect of effect size considerations) or the unambiguous presence of methodological biases, to mention but a few of them. Interestingly, a recent study conduct by Fanelli and co-workers (Fanelli, Costas, & Ioannidis, 2017) have highlighted differences in the risk of bias (poor estimate of the magnitude of effect size due to, for example, lower inclusion of grey literature, US effect or industry bias…) between the classical disciplines, the risk being highest in the social sciences (to which

psychology belongs). These differences could reflect dissimilar research practices documented in primary studies (e.g. higher publication bias in some disciplines) or distinct procedural choices in meta-analyses (e.g. lower inclusion of grey literature in some disciplines) (Fanelli et al., 2017).

To our knowledge, no studies have however been conducted in order to evaluate the quality of MAs published in the field of psychology. With this research project, our aim is to evaluate the quality of MAs and identify its associated factors published in the psychological or psychology-related field on the PsycINFO database during the year 2016.

Objectives

The objective of this research is to assess the factors associated with the quality of recent MAs indexed in PsycINFO for the year 2016 using two samples of MAs; one composed of MAs claiming to follow the PRISMA statement and the other one including MAs ignoring it.

Our research will be organized in three sub-studies:

1. The assessment of the reporting quality of MAs that claim to use the PRISMA statement compared to those that do not use PRISMA through the PRISMA statement;
2. The assessment of the methodological quality of MAs that claim to use the PRISMA statement compared to those that do not use PRISMA through the AMSTAR 2 tool;
3. The identification of potential factors associated with the quality of MAs.

Research questions

Based on our objectives, our research questions are the following:

1. What is the relationship between of the use of the PRISMA checklist on the reporting quality of MAs?
2. What is the relationship between of the use of the PRISMA checklist on the methodological quality of MAs?
3. What are the potential factors (e.g. publishing journal's impact factor, pre-registration of the study, experience of the first author…) associated with the quality of MAs?

Hypotheses

1. Comprehensive and transparent reporting is necessary to assessing the methodological quality of MAs (Page et al. 2016). They are some articles that highlighted that MAs have a poor reporting quality in the medical literature and the score of PRISMA Statement that were found is between 16,8 and 23/27 points ((Tunis et al. 2013; Fleming et al.

2013; Peters et al. 2015; Zhang et al. 2016; Gagnier & Kellam 2013; Adie et al. 2015; Zhu et al. 2016). We make the assumptions that the use of PRISMA statement improves the reporting quality of MAs using it. This is supported by another study showing that the PRISMA scores is higher by 1 point for the MAs for which authors claimed having used PRISMA statement (Zhu et al. 2016).

2. Meta-research has revealed that the psychological literature present an unsatisfactory level of methodological quality (Bakker & Wicherts 2011; Ioannidis 2012; Oliveras et al. 2017; Stanley et al. 2017). It is therefore possible that this is also true for the MA published in psychology and related fields. In recent studies analyzing the quality of Mas in a number of health-related fields AMSTAR scores have been found to fall between 3,7 and 7,8/11 points (Zhang et al. 2016; Adie et al. 2015; Gagnier & Kellam 2013; Klimo et al. 2014). Since the AMSTAR 2 tool was recently published, it is probable that no study has yet evaluated the quality of the MAs with this tool. Note that it is likely that the use of PRISMA statement exerts a positive influence on the quality of Mas using it. This is supported by a recent study on depression showing that the AMSTAR scores for the MAs for which authors claimed having used the PRISMA statement, reach an average of 0.4 point higher than those which did not use PRISMA (Zhu et al. 2016).
3. They are some potential factors that could correlate with (and possibly influence) the quality of MAs. More specifically, on the basis of previous meta-research studies we make the assumptions that the following factors are positively associated with the measures of quality of MAs: h-index of the first author (Cullis et al. 2017), experience of the principal author in MAs (Zhang et al. 2016), affiliation of the authors to a university (Cullis et al. 2017), publishing journal's impact factor (Cullis et al. 2017), PRISMA endorsement by the journal publishing (Cullis et al. 2017), funding sources described (Gagnier & Kellam 2013), Cochrane collaboration (Adie et al. 2015; Cullis et al. 2017; Zhu et al. 2016), number of pages of the manuscript (Adie et al. 2015; Cullis et al. 2017), pre-registration of the study (Cullis et al. 2017; Zhang et al. 2016; Zhu et al. 2016), non-Asian origine (Zhang et al. 2016) and meta-analyses of randomized controlled trials (Zhang et al. 2016; Zhu et al. 2016). Furthermore, we hypothesize that the following variables will be also associated with the quality of MAs: open access of the publication, open data (or data sharing), the field of psychology, number of individual studies in each MA, number of databases used, assessment of the quality of

individual studies and related tools used, pooling methods used to combine the data, assessment of the publication bias and related method used, assessment of the heterogeneity and related method used, the statistical software used and tendency of the conclusion.

Sampling plan

Existing Data

The tests can be considered as confirmatory given similar studies that have recently been conducted in the medical field.

Explanation of existing data

This is not applicable for our research protocol because no data have been collected so far.

Data collection procedures

Data collection

Data will be collected from a random sample of 200 MAs that will be divided into 2 groups (use of PRISMA vs no use of PRISMA).

Protocol selection of meta-analyses articles

All MAs performed on human subjects and published in English in 2016 in the electronic database PsycINFO will be searched. The electronic search strategy was developed with co-authors and the assistance of a librarian is available in *table 1*.

Table 1: Search strategy

1	meta analysis.md. (15886)
2	meta analysis/ (3940)
3	meta analys*.mp. (24573)
4	data pooling*.mp. (50)
5	2 or 3 or 4 (24599)
6	5 not 1 (10725)
7	1 or 6 (26611)
8	limit 7 to (English and human and yr="2016") (2159)

A total of 2159 potentially relevant MAs were identified in the PsycINFO database. Two investigators will independently review each title and abstract in order to exclude irrelevant articles and to only select the studies that meet inclusion criteria (full inclusion and exclusion criteria are available in Table 2). All discrepancies in opinion regarding the selection of articles

will be resolved through discussion and consensus between the two investigators; any persistent disagreement will be solved with the intervention of a third person (an expert).

Table 2 : eligibility criteria

Inclusion criteria
- Meta-analysis
- Articles published in the PsycINFO-database
- Published between 01.01.2016 to 31.12.2016
- English
Exclusion criteria
- Overview, review
- Meta-synthesis
- Qualitative meta-analysis
- Umbrella review
- Meta-analysis of meta-analyses
- Systematic review without meta-analysis
- Protocol of meta-analysis
- Network meta-analysis
- Activation likelihood Estimation Meta-analysis (ALE MA)
- Signed differential mapping meta-analysis (SMD MA)
- Voxel wise meta-analysis
- Individual patient data meta-analysis (IPD MA)
- Genetic association study (GWAS), genetic study
- Multi-level meta-analysis
- Update
- Letter, comment, abstract, chapter, erratum, dissertation or editorial journal

Choice of language for inclusion was based on expertise within our research team, due to budget constraints, limited time and resources.

A flowchart with the number of included studies will be elaborated. The reason of exclusion of articles will be presented at the step of full-text selection.

Once all MAs will be identified, we will randomly select a minimum of 200 articles as follows. All references of articles will be indexed in an Excel file and randomly assigned to a number. Then, we will rank the articles in ascending order. Two blinded researchers, with the intervention of a third researcher in case of disagreement, will classify MAs that meet inclusion criteria in either the group "with PRISMA" or "without PRISMA" until each group will contain a minimum of 100 MAs. Kappa statistics will be used to test inter-rater agreement.

Data extraction

Relevant data will be extracted from the full texts of all selected articles in a standardized Microsoft Excel spreadsheet by two independent investigators trained for this data extraction. We will record the following factors that might influence the quality of the MAs: characteristics of the manuscripts, characteristics of the study, objective(s) of the study, statistical analyses, characteristics of the protocol and items of PRISMA statement, AMSTAR tool and AMSTAR2 tool. All data extracted will be detailed in appendices. If any disagreements were to be observed between the two reviewers, they will be resolved by discussion, if necessary with the intervention of a third reviewer. Kappa statistics and absolute agreement (%) will be used to assess reproducibility.

Sample size

A first exploratory search in PsycINFO has yielded approximately 2000 articles, which are impossible to analyze for us in a reasonable period of time. We elected to randomly (see below for the method of randomization) select 200 MA articles (until each group, PRISMA and NO PRISMA, will contain a minimum of 100 MAs) from all eligible MAs published in 2016 and indexed in PsycINFO. There is no global MAs offering a synthetic effect size (of the published differences between the two samples) that could have been used to determine a priori a sample size allowing the detecting of a significant difference (power analyses). The chosen sample size can minimally detect a medium effect size (Cohen's d = 0.46065; as computed via G*Power) using a two-tailed Student t-test for independent groups taken at an alpha error probability of 0.05 and a power (1-beta error probability) of 0.90 (critical t = 1.9720). Note that smaller effects sizes cannot be detected (if existing) with such sample (n=100). The meaning and practical significance of the empirically obtained effect size will be discussed.

Sample size rationale

Considering the power analysis described above and constraints in terms of time, financial resources and staff, we will conduct this research on about 200 articles. We think that this will necessitate more than 500 hours of coding for each assessor.

Stopping rule

This is not applicable for our research protocol.

Variables

Manipulated variables

This cannot be applied to the present research protocol.

Measured variables

In order to verify our hypotheses, we will assess not only the quality of reporting and of conduct of MAs but also a set of variables to identify the potential factors that are associated with the quality of MAs.

Assessment of reporting quality

Eligible papers will first be assessed with the PRISMA statement. Each individual item of the PRISMA statement will be answered by "yes" or "no" response, respectively depending on the item being or not fulfilled and will be coded "1" or "0". The total score of the PRISMA statement is the addition of all items coded 1 with a maximum of 27 points.

Assessment of methodology quality

Eligible papers will then be assessed with AMSTAR 2 tool. Each individual item of the AMSTAR tool will be answered by a "yes", "partial yes" or "no" response, respectively depending on the item being or not fulfilled and will be coded "1", "0.5", "0". The total score given by the tool is the addition of all items coded 1 and 0.5 with a maximum of 16 points.

Eligible papers will also be assessed with the AMSTAR tool. Each individual item of the AMSTAR tool will be answered by a "yes" or "no" response, depending on the item being or not fulfilled and will be coded "1" or "0". The total score given by the tool is the addition of all items coded 1 with a maximum of 11 points.

Identification of potential factors associated with the quality of MA

All factors will be assessed by three types of variables: dichotomous variables, quantitative variables and text variables (open questions).

Dichotomous variables

Dichotomous variables will be coded as follows:

1: Yes, it features the characteristic that we seek.

0: No, it does not feature the characteristic that we seek.

99: Not reported, the characteristic that we seek is not available.

88: Not applicable.

The variables that are concerned are the following: author's experience in meta-analysis, affiliation of the authors to a university, PRISMA's recommendation by the journal, restriction of the word count by the journal, declaration of conflict of interest, declaration of funding sources, Cochrane collaboration, open access, open data, registration of the study in a database, evaluation of the quality of study, use of reporting or methodology guideline, the type of study (randomized controlled trials (RCT) or not), evaluation of publication bias, evaluation of heterogeneity, presence of a protocol and the conclusion supports the assumptions.

Quantitative variables

They will be encoded with numerical values and their units of measurements.

The variables that are concerned are the following: h-index (an author-level metric), number of authors, impact factor of the journal (which reflects the frequency with which the average article in a journal has been cited in a particular year) and number of database consulted.

Qualitative variables

The relevant variables are the following: the continent where the study was conducted (Europe, Asia, Africa, America, Oceania), number of study included in each MA (0-3; 4-9; ≥10) and the pooling method used (random effect model, fixed effect model or mixed effect model).

Text variables

The remaining variables are recorded as a text variable. The variables that are concerned are the following: the name of the tool used in order to assess the quality of individual study, the name of the database(s) searched, the PsycINFO classification, the name of the guideline used, the method used to evaluate the publication bias, the method used to assess the heterogeneity and the statistical software. These text variables will then be categorized.

If the data is not available, it will be coded 88.

Indices

This is not applicable for the present research protocol.

Design plan

Study type

This is an observational study. Data are collected from meta-analysis articles.

Blinding

The selection of MAs reviews by title and abstract and by full-text will be done independently by two investigators.

Study design

This is a cross-sectional study.

Randomization

We will randomly MAs articles to get a minimum of 100 MAs in each group. All references of articles (n= probably more than 2000) will be indexed in an Excel file and randomly assigned to a number. Then, we will rank the articles in ascending order. Two blinded researchers, with the intervention of a third researcher in case of disagreement will classify MAs that meet inclusion criteria in either group “with Prisma” and “without Prisma” until each group will contain a minimum of 100 articles MAs.

Analysis plan

Statistical model

The characteristics of all individual studies will first be presented. All quantitative variables that follow a normal distribution will be reported as mean and standard deviation and those that do not follow a normal distribution will be represented as median and percentile (P25 and P75). Distribution will be considered as normal if data meet 3 of the 4 following conditions: the mean is close to the median, the Shapiro-Wilk normality test yields a p-value </=0.05, the curve of the variable follows the normal (or Gaussian) distribution and the linearity of the QQ-Plots is respected. Qualitative and dichotomous variables will be reported as numbers and frequencies. The results of the quality assessment of MAs with the PRISMA statement, AMSTAR tool and AMSTAR 2 will be reported, for quantitative variable, as number and frequency for each item and as mean or as median for the total score. The data will be presented and analyzed using a

star chart. A star chart is a graphical tool that will allow us to represent and compare the percentages of item of PRISMA statement, AMSTAR tool and AMSTAR 2 met by the MAs.

To verify our first and second hypothesis, the reporting and methodological qualities of the individual studies will be compared between the studies that report using the PRISMA checklist and the studies that do not. Comparisons of means between the two groups will be calculated using the Student t-test if for independent groups if the score of PRISMA, AMSTAR and AMSTAR 2 are normally distributed and the Mann-Whitney test if the score of PRISMA, AMSTAR and AMSTAR 2 are not normally distributed. To test the association between the use of the PRISMA statement and the different items of PRISMA, AMSTAR and AMSTAR 2, we will be used a logistic regression.

To test our third hypothesis, factors (all the data detailed in the measured variables part) with potential influence on the quality of studies (mean score of AMSTAR, mean score of AMSTAR 2, independent variable) will be identified with a univariate linear regression. The variables with a p-value lower than 0.1 will be combined in stepwise backward multiple regression analysis. A p-value </=0.05 will be considered as significant.

Transformations

Each item of the PRISMA & AMSTAR checklists will be coded with the following meaning:

1 = Yes or not applicable

0 = No

We will also sum up all items coded 1, with a maximum score of 27 or 11, respectively.

Each item of AMSTAR 2 will be coded with the following meaning:

1 = Yes or not applicable

0.5 = Partial yes

0 = No

We will also sum up all items coded 1 and 0.5, with a maximum score of 16.

See appendix for the details of all transformations for each variable.

Follow-up analyses

All follow-up analyses are described above.

Inference criteria

Not applicable for our research protocol.

Data exclusion

No data will be excluded from our database.

Missing data

Missing data may have an impact on the analysis and on the interpretation of the results. Some of the extracted data may not be available (h-index, impact factor…). After data encoding, a quality control will be done, at database-level, in order to check for outliers, coding error and missing values. In case of incomplete information, we will contact the authors.

Exploratory analysis

The exploratory analyses will be considered, based on the results obtained.

If there is a statistically significant difference in quality between MAs which report using PRISMA and those which do not, we will consider carrying out the following analyses. A logistic regression will be carried out in order to describe the relation between the dichotomous dependent variables (PRISMA vs No PRISMA) and all potential explanatory variables (all the data are detailed in the measured variables part). The variables with a p-value lower than 0.1 will be combined in stepwise backward multiple logistic regression analysis. A p-value <0.05 will be considered significant on statistical analyses.

Scripts

Upload an analysis script with clear comments

Not available at the moment.

Other

We would like to acknowledge Pr. Anne-Françoise Donneau for interesting discussions about some aspects on the planned statistical analyses and Ms Nancy Durieux for her assistance in the building of our strategy, in terms of electronic literature search.

Bibliography

Adie, S. et al., 2015. Quality of Conduct and Reporting of Meta-analyses of Surgical Interventions. *Annals of Surgery*, 261(4), pp.685–694. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000658-201504000-00012 [Accessed June 14, 2017].

American Psychological Association, 2006. Evidence-Based Practice in Psychology APA Presidential Task Force on Evidence-Based Practice. *American Psychologist*, 61(4), pp.271–285. Available at: http://www.apa.org/pubs/journals/features/evidence-based-

statement.pdf [Accessed January 10, 2018].

Bakker, M. & Wicherts, J.M., 2011. The (mis)reporting of statistical results in psychology journals. Available at: https://link.springer.com/content/pdf/10.3758%2Fs13428-011-0089-5.pdf [Accessed July 31, 2017].

Cooper, H., 2008. Reporting Standards for Research in Psychology: Why Do We Need Them? What Might They Be? *American Psychologist*, 63(9), pp.839–851. Available at: https://www.apa.org/pubs/authors/jars.pdf [Accessed January 15, 2018].

Cullis, P.S., Gudlaugsdottir, K. & Andrews, J., 2017. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. *PLoS ONE*, 12(4). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5383307/pdf/pone.0175213.pdf [Accessed June 14, 2017].

Fanelli, D., Costas, R. & Ioannidis, J.P.A., Meta-assessment of bias in science. Available at: http://www.pnas.org/content/114/14/3714.full.pdf [Accessed December 18, 2017].

Fleming, P.S. et al., 2013. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. *The Angle Orthodontist*, 83(1), pp.158–163. Available at: http://www.angle.org/doi/10.2319/032612-251.1 [Accessed May 24, 2018].

Gagnier, J.J. & Kellam, P.J., 2013. Reporting and Methodological Quality of Systematic Reviews in the Orthopaedic Literature. *The Journal of Bone and Joint Surgery-American Volume*, 95(11), pp.e77-1–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23780547 [Accessed June 14, 2017].

Guyatt, G. et al., 2015. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Medicine*, Available at: https://jamaevidence.mhmedical.com/book.aspx?bookID=847 [Accessed September 1, 2017].

Higgins, J. & Green, S., 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011].* The Cochra., Available at: http://handbook.cochrane.org.

Ioannidis, J.P.A., 2016. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, 94(3), pp.485–514. Available at: http://doi.wiley.com/10.1111/1468-0009.12210.

Ioannidis, J.P.A., 2012. Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), pp.645–654. Available at: http://pps.sagepub.com/ [Accessed

July 31, 2017].

Klimo, P. et al., 2014. Methodology and reporting of meta-analyses in the neurosurgical literature. *Journal of neurosurgery*, 120(4), pp.796–810. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24460488.

Moher, D. et al., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). *Physical Therapy*, 89(9), pp.873–880. Available at: http://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1000097&type=printable [Accessed August 25, 2017].

Oliveras, I., Losilla, J.-M. & Vives, J., 2017. Methodological quality is underrated in systematic reviews and meta-analyses in health psychology. *Journal of Clinical Epidemiology*, 86, pp.59–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28499846 [Accessed November 28, 2017].

Page, M.J. et al., 2016. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study N. Low, ed. *PLOS Medicine*, 13(5), p.e1002028. Available at: http://dx.plos.org/10.1371/journal.pmed.1002028 [Accessed May 24, 2018].

Peters, J.P.M. et al., 2015. Reporting quality of systematic reviews and meta-analyses of otorhinolaryngologic articles based on the PRISMA statement. *PLoS ONE*, 10(8). Available at: http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0136540&type=printable [Accessed June 14, 2017].

Shea, B.J. et al., 2017. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, p.j4008. Available at: http://www.bmj.com/content/bmj/358/bmj.j4008.full.pdf [Accessed November 28, 2017].

Shea, B.J. et al., 2009. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), pp.1013–1020. Available at: http://ac.els-cdn.com/S0895435608003259/1-s2.0-S0895435608003259-main.pdf?_tid=7ad1b7aa-558b-11e7-a3dc-00000aacb361&acdnat=1497944501_0f82a189082ff8754f64a91d9670402e [Accessed June 20, 2017].

Shea, B.J. et al., 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*, 7, p.10. Available at: https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-7-10?site=bmcmedresmethodol.biomedcentral.com [Accessed August 25, 2017].

Stanley, T.D., Carter, E.C. & Doucouliagos, H., 2017. What Meta-Analyses Reveal about the Replicability of Psychological Research. Available at: http://www.deakin.edu.au/__data/assets/pdf_file/0007/1198456/WhatMeta-AnalysesReveal_WP.pdf [Accessed January 10, 2018].

Tunis, A.S. et al., 2013. Association of Study Quality with Completeness of Reporting: Have Completeness of Reporting and Quality of Systematic Reviews and Meta-Analyses in Major Radiology Journals Changed Since Publication of the PRISMA Statement? *Radiology*, 269(2), pp.413–426. Available at: http://pubs.rsna.org/doi/10.1148/radiol.13130273 [Accessed May 24, 2018].

Zhang, H. et al., 2016. Reporting and methodological qualities of published surgical meta-analyses. Available at: http://ac.els-cdn.com/S0895435615003121/1-s2.0-S0895435615003121-main.pdf?_tid=4e369bf4-50e8-11e7-9b21-00000aab0f26&acdnat=1497434613_6d09bcd5abe078df3344e91ace4b09d1 [Accessed June 14, 2017].

Zhu, Y. et al., 2016. Is the best evidence good enough: Quality assessment and factor analysis of meta-analyses on depression. *PLoS ONE*, 11(6). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4919061/pdf/pone.0157808.pdf [Accessed June 14, 2017].

Appendix
Explication of the data extraction form

Name	Explication	Description
Name of the reviewer	Name of the reviewer	Text
Study ID	Reference number of the article	Text
Inclusion of the article	Inclusion of the article based on the selection criteria	1 = Yes 0 = No
If excluded, indicate the reason of exclusion	Reason of exclusion	The reason of exclusion 88 = if not applicable
Use of PRISMA	The authors declared the use of PRISMA	1= Yes 0 = No

1. Characteristics of the manuscript

Name	Explication	Description
DOI of the article	Unique identifier of the article	Text
Year of publication	Publishing year of the manuscript	Text
Author's name	Name of the first author	Text
Author's h-index	H-Index of the first author (Scopus)	Quantitative variable
Author's experience	Number of meta-analyses from the same author(s) (Scopus)	Quantitative variable
Affiliation of the authors to a university	Affiliation of the authors to a university	1 = Yes 0 = No 99 = Not reported
Number of authors	Total number of authors	Quantitative variable
Contribution of authors	Details of the authors' contribution	1 = Yes 0 = No 99 = Not reported
Journal's name	Name of the journal	Text
Journal's Impact factor	The IF of the journal using the ISI Journal Citation Reports 2016 (http://isiknowledge.com)	Quantitative variable
Instruction for authors: PRISMA required?	The journal recommended to use PRISMA statement	1 = Yes 0 = No 99 = Not reported

Instruction for authors: page or word limitation?	Limitation of the number of pages or words	1 = Yes 0 = No 99 = Not reported
PsycINFO classification	Classification of the field of psychology based on the PsycINFO Content Classification Code System	2100 = General Psychology 2200 = Psychometrics & Statistics & Methodology 2300 = Human experimental Psychology 2400 = Animal Experimental & comparative Psychology 2500 = Physiological Psychology & Neuroscience 2600 = Psychology & The Humanities 2700 = Communication Systems 2800 = Developmental Psychology 2900 = Social Processes & Social Issues 3000 = Social Psychology 3100 = Personality Psychology 3200 = Psychological & Physical disorders 3300 = Health & Mental Health Treatment & Prevention 3400 = Professional Psychological & Health Personnel Issues 3500 = Educational Psychology 3600 = Industrial & Organizational Psychology 3700 = Sport Psychology & Leisure 3800 = Military Psychology 3900 = Consumer Psychology 4000 = Engineering & Experimental Psychology 4100 = Intelligent Systems 4200 = Forensic Psychology & Legal Issues
Corresponding author (email address)	Email of the author	Text

Conflict of interest described	Conflict of interest is described	1 = Yes 0 = No 99 = Not reported
Details of conflict of interest	If yes, brief description of conflict of interest	Text
Funding sources described	Funding sources are described	1 = Yes 0 = No 99 = Not reported
Funding sources	If yes, brief description of funding sources	Text
Cochrane collaboration	The study is a Cochrane collaboration	1 = Yes 0 = No
Number of page of manuscript	Total number of pages of the manuscript	Quantitative variable
Open access	The publication is open access?	1 = Yes 0 = No 99 = Not reported
Open data	The data is open access?	1 = Yes 0 = No 99 = Not reported

2. Characteristics of the study

Name	Explication	Description
Registration of the study?	The study was recorded in a specific database.	1 = Yes 0 = No 99 = Not reported
Number of the registration	Registration number of the study	Text 88 = if not applicable
Name of the registry	Name of the registry in which the meta-analysis has been registered	Text 88 = if not applicable
Date of submission of the manuscript	Date of submission of the manuscript	Text (month-year)
Date of publication of the manuscript	Publication date of the manuscript	Text (month-year)
Continent of origin of first author	Continent in which the study has been conducted	Europe – Asia – Africa – America - Oceania
Type of the individual study	Study design of the studies included in the MA	Observational study RCT All types Not specified
Number of databases searched	Number of databases consulted	Quantitative variable

Name of the database	Name of the database searched	Text
Quality of individual study is assessed?	Quality of individual study is assessed	1 = Yes 0 = No
Name of the tool used to assess the quality	Name of the tool used to assess the quality of individual studies	Text variable 88 = If not applicable
Reference to use of the guideline	Reference to use of a guideline	1 = Yes 0 = No
Name of the guideline used	Name of the guideline used (PRISMA, MOOSE, AMSTAR…)	Text 88 = If not applicable
Search strategy	Presence of the complete search strategy	1 = Yes 0 = No 99 = Not reported
Focus of review	Type of the field of psychology	Text

3. Objective of the study

Name	Explication	Description
Main objective	Aim of the study	Text
Primary outcome	Primary outcome of the study disclosed	Text
Secondary outcomes	Secondary outcome of the study disclosed	Text

4. Statistical analyses

Name	Explication	Description
Number of meta-analyses performed	Number of meta-analyses performed in the presented study	Quantitative variable
Number of studies included in each meta-analysis	Number of studies included in each meta-analysis performed in the study	0-3; 4-9; ≥10
Pooling methods	The pooling methods used to combine data	Fixed – Random - Mix
Assessment of the publication bias	The publication bias is evaluated	1 = Yes 0 = No
Method used to assess the publication bias	Method used to assess the publication bias	Text 88 = If not applicable

Assessment of the heterogeneity	The heterogeneity is evaluated	1 = Yes 0 = No
Heterogeneity	Method used to assess the heterogeneity	Text 88 = If not applicable

5. Protocol

Name	Explication	Description
Protocol	The protocol of the study is existent and available	1 = Yes 0 = No
Primary outcome	Primary outcome of the study	Text 88 = If not applicable
Secondary outcome	Secondary outcome of the study	Text 88 = If not applicable

6. Conclusion

Name	Explication	Description
Conclusion	Main conclusion of the study	Text
Trends of the conclusion	The conclusion supports the assumptions	1 = Yes 0 = No

7. PRISMA statement

Name	Explication	Description
P1	TITLE Title	1 = Yes 0 = No
P2	ABSTRACT Structured summary	1 = Yes 0 = No
P3	INTRODUCTION Rationale	1 = Yes 0 = No
P4	Objective	1 = Yes 0 = No
P5	METHODS Protocol and registration	1 = Yes 0 = No
P6	Eligibility criteria	1 = Yes 0 = No
P7	Information sources	1 = Yes 0 = No

P8	Search	1 = Yes 0 = No
P9	Study selection	1 = Yes 0 = No
P10	Data collection process	1 = Yes 0 = No
P11	Data items	1 = Yes 0 = No
P12	Risk of bias in individual studies	1 = Yes 0 = No
P13	Summary measures	1 = Yes 0 = No
P14	Synthesis of results / Planned methods of analysis	1 = Yes 0 = No
P15	Risk of bias across studies	1 = Yes 0 = No
P16	Additional analysis	1 = Yes 1 = Not applicable 0 = No
P17	RESULTS Study selection	1 = Yes 0 = No
P18	Study characteristics	1 = Yes 0 = No
P19	Risk of bias within studies	1 = Yes 0 = No
P20	Results of individual studies	1 = Yes 0 = No
P21	Synthesis of results	1 = Yes 0 = No
P22	Risk of bias across studies	1 = Yes 0 = No
P23	Additional analysis	1 = Yes 1 = Not applicable 0 = No
P24	DISCUSSION Summary of evidence	1 = Yes 0 = No
P25	Limitations	1 = Yes 0 = No
P26	Conclusions	1 = Yes 0 = No
P27	Funding	1 = Yes 0 = No
	Total score	Sum of all items coded "1"

8. Amstar tool

Name	Explication	Description
A1	Was an 'a priori' design provided?	1 = Yes 0 = No
A2	Was there duplicate study selection and data extraction?	1 = Yes 0 = No
A3	Was a comprehensive literature search performed?	1 = Yes 0 = No
A4	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	1 = Yes 0 = No
A5	Was a list of studies (included and excluded) provided?	1 = Yes 0 = No
A6	Were the characteristics of the included studies provided?	1 = Yes 0 = No
A7	Was the scientific quality of the included studies assessed and documented?	1 = Yes 0 = No
A8	Was the scientific quality of the included studies used appropriately in formulating conclusions?	1 = Yes 0 = No
A9	Were the methods used to combine the findings of studies appropriate?	1 = Yes 0 = No
A10	Was the likelihood of publication bias assessed?	1 = Yes 0 = No
A11	Was the conflict of interest included?	1 = Yes 0 = No
	Total score	Sum of all items coded "1"

9. Amstar 2 tool

Name	Explication	Description
AM1	Did the research questions and inclusion criteria for the review include the components of PICO?	1 = Yes 0 = No

AM2	Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	1 = Yes 0.5 = Partial Yes 0 = No
AM3	Did the review authors explain their selection of the study designs for inclusion in the review?	1 = Yes 0 = No
AM4	Did the review authors use a comprehensive literature search strategy?	1 = Yes 0.5 = Partial Yes 0 = No
AM5	Did the review authors perform study selection in duplicate?	1 = Yes 0 = No
AM6	Did the review authors perform data extraction in duplicate?	1 = Yes 0 = No
AM7	Did the review authors provide a list of excluded studies and justify the exclusions?	1 = Yes 0.5 = Partial Yes 0 = No
AM8	Did the review authors describe the included studies in adequate detail?	1 = Yes 0.5 = Partial Yes 0 = No
AM9	Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	1 = Yes 0.5 = Partial Yes 0 = No
AM10	Did the review authors report on the sources of funding for the studies included in the review?	1 = Yes 0 = No
AM11	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	1 = Yes 0 = No
AM12	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	1 = Yes 0 = No
AM13	Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	1 = Yes 0 = No
AM14	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	1 = Yes 0 = No

AM15	If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	1 = Yes 0 = No
AM16	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	1 = Yes 0 = No
	Total score	Sum of all items coded "1" or "0.5"

BMJ Open

The methodological quality of meta-analyses indexed in PsycINFO: leads for enhancements

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-036349.R1
Article Type:	Original research
Date Submitted by the Author:	10-Mar-2020
Complete List of Authors:	Leclercq, Victoria; University of Liege, Division of Public Health, Epidemiology and Health Economics Beaudart, Charlotte; University of Liege, Division of Public Health, Epidemiology and Health Economics Ajamieh, Sara; University of Liege, Division of Public Health, Epidemiology and Health Economics Tirelli, Ezio; University of Liege, Department of Psychology Bruyère, Olivier ; University of Liege, Division of Public Health, Epidemiology and Health Economics
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Research methods
Keywords:	EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

THE METHODOLOGICAL QUALITY OF META-ANALYSES INDEXED IN PSYCINFO: LEADS FOR ENHANCEMENTS

Short title (Max 70 characters)

The methodological quality of meta-analyses indexed in PsycINFO

Leclercq V[1], Beaudart C[1], Ajamieh S[1,2], Tirelli E[2], Bruyère O[1]

[1]Department of Public Health, Epidemiology and Health Economics, University of Liège, Belgium

[2]Department of Psychology, University of Liège, Belgium

Corresponding author: Victoria Leclercq; victoria.leclercq@uliege.be

ABSTRACT

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

Design We evaluated a random sample of 206 MAs indexed in the PsycINFO database in 2016 through a cross-sectional study. Two authors independently extracted the methodologic characteristics of all MAs and checked their quality according to the 16 items of the AMSTAR2 (A MeaSurement Tool to Assess systematic Reviews) tool for MA critical appraisal. Moreover, we investigated the effect of mentioning 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.

Discussion 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.

Strengths and limitations of this study

- Some studies have highlighted methodological weaknesses in the conduct of systematic reviews (SRs) and meta-analysis (MAs) and we search to have an overview of methodological practice of MAs indexed in PsycINFO according to the tool AMSTAR2 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 PRISMA statement on the methodological quality of MAs**.**
- A sample of 206 Mas indexed in PsycINFO in 2016 and published in English was analyzed.

1 INTRODUCTION

Since the definition of meta-analyses (MAs) being introduced by Glass in 1976, MAs conducted in behavioral and social sciences have increased rapidly in number. There were more than 30 000 MAs indexed in 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[1].

However, the validity of MA conclusions relies upon the rigor 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 judgment calls that need to be made by the meta-analysist. Moreover, too many systematic reviews (SRs) and MAs are of low quality[1–5], 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[6–8], an inappropriate literature search[9–12], flaws in the statistical analyses[10,13–16] and an insufficient assessment of the risk of bias of individual studies[10,17,18].

To support researchers in the realization and reporting of MAs, two tools are commonly used. The first is PRISMA ("Preferred Reporting Items for Systematic Reviews and Meta-Analyses"), which was developed in 2009 by Liberati et al[19]. 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[20], 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 behavioral 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 characterize 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.

2 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 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[21]. 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 specialized in the field of behavioral 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 systematic review with a MA, indexed in the PsycINFO database, published between January 01, 2016, and December 31, 2016, and published in English. In total, 2159 records were identified. Two authors (V.L & C.B) 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 flowchart in 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 & SA) independently extracted all relevant data from the full texts of all selected articles in a standardized Microsoft Excel spreadsheet. The extraction form had been pretested on ten MAs. Data extraction disagreements between the two investigators were resolved by discussion (Median Kappa with P25 and P75: 0.56(0.29-0.76), with the intervention of a third investigator if necessary. 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 "A MeaSurement Tool to Assess systematic Reviews 2"[20]. AMSTAR2 was a revision of the original AMSTAR instrument[22] developed by Shea et al in 2007, 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 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 2 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 gray 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 4 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 7 critical weaknesses that should reduce confidence in the findings of a review and 9 other items that are considered noncritical weaknesses, as presented in Table 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 hypothesized 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 (JCR) 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 summarized 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 3 of the 4 following conditions: the mean was close to the median, the Shapiro-Wilk normality test yielded a P-value ≥ 0.05, the curve of the variables followed the Gaussian distribution and the linearity of the QQ-Plots was respected. In this study, we made

the assumptions 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. 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. Associations were quantified using odds ratios with 95% confidence intervals. A Bonferroni correction was used to adjust the results for multiple testing (16 tests, p-value<0.003). All analyses were performed using SAS 9.4 software.

Patient and public involvement

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

3 RESULTS

Search results

A total of 2159 potentially relevant MAs related to behavioral 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 2. 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 [23] (17, 74%), Mars (2, 9%) and Quorom (1, 4%). Finally, most articles were not available for open access (90.3%), and only one was a Cochrane MA.

Written by one 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 categorized as critically low quality, 8 as low quality, 2 as moderate quality and 1 as high quality. Only one MA [24] provided all the information on all 7 critical domains assessed and was considered high quality according to AMSTAR2. Two additional MAs [25,26] also provided all information on all 7 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 the classification advised by the AMSTAR2 tool (Table 1).

In Figure 1, we summarize 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". Three-quarters 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: 1.89-12.37) 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: 2.46-29.90);
- study selection in duplicate (item 5; OR: 4.55; 95%CI: 2.52-8.21);
- a list of excluded studies (item 7; OR: 2.69; 95%CI: 1.06-6.86);
- a detailed description of the included studies (item 8; OR: 2.62; 95%CI: 1.44-4.76);

- a satisfactory technique for assessing the risk of bias in individual studies (item 9; OR: 4.48; 95%CI: 2.43-8.27);
- an appropriate method for the statistical combination of results (item 11; OR: 3.87; 95%CI: 1.50-10.04);
- an assessment of the potential impact of risk of bias in individual studies (item 12; OR: 5.17; 95%CI: 2.39-11.16);
- appropriate consideration of the risk of bias in primary studies when interpreting the results (item 13; OR: 6.34; 95%CI: 3.15-12.78);
- a satisfactory explanation for any heterogeneity found in the results (item 14; OR: 2.7; 95%CI: 1.38-5.27);
- an adequate investigation of publication bias (item 15; OR: 1.95; 95%CI: 1.06-3.59)
- a report of any potential conflict of interest sources (item 16; OR: 2.15; 95%CI: 1.14-4.04).

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.

4 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[27]. 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 characterize the methodological quality of MAs indexed in PsycINFO 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 behavioral and social sciences.

- First, no more than 11% of MAs had a research protocol available. However, several scientists[8,16,28] 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, minimizing the risk of bias, avoiding replicate studies and enhancing transparency[7]. 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 a systematic review 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 2 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 gray 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[10]. The search strategy is an essential step of the MA process since the comprehensiveness and completeness of the search[3,29]

is dependent on this strategy. Furthermore, other scientists have highlighted the need to improve research strategies for more comprehensive MAs[9,29].

- 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[17,30], 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 and her team 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[17]. However, there is still a lack of guidance to incorporate these risk of bias assessments into meta-analyses[17,18,31].

Regarding our second research question, the explicit mention of PRISMA suggested an improved methodological quality of MAs. Three-quarters of the items in the AMSTAR2 tool, including 6/7 of the critical items, were significantly more frequent in the MAs that explicitly mentioned PRISMA than in those that did not. However, it is recognized 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. Moreover, these results should be interpreted with caution because after adjustment for multiple testing with the Bonferroni correction, fewer items were found to be statistically significant. 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[21,27,30,32,33].

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[24] 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[25,26] 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[20], yet only a few studies from our sample of MAs mention it explicitly. Dechartres and her team stressed the association between publication characteristics and effect estimates[11] 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[11,12,34]. 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[10] but contradict previous studies that have highlighted several flaws in the application and interpretation of statistical analyses in MA[13,14,27,35]. Page et al identified some mistakes in the use of adequate statistical models, the sufficient exploration of subgroup analyses and sensitivity analyses[14]. 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[20]. 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 generalize 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 Kappa value indicated a moderate agreement but subjectivity related to data extraction is limited since all data has been extracted in duplicate. Moreover, using AMSTAR2, we can investigate the methodological characteristics used to conduct the study (e.g. 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 (e.g. 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.

5 CONCLUSION

This research contributes to raising awareness among researchers about flaws in MAs published in behavioral and social sciences fields, which hopefully increases the adoption of more rigorous research practices. It is clear that meta-analytic 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.

KEYWORDS

Meta-analyses – AMSTAR – meta-research – methodological quality – PRISMA

FOOTNOTES

Conflict of interest and funding: The authors declare no conflict of interest. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors' contribution: VL, CB, ET and OB conceived the study; VL, CB and SA participated in data collection; VL, CB and OB analyzed and interpreted data; VL, CB, ET and OB corrected the manuscripts. All co-authors read and approved the final version of the manuscript.

Data sharing statement: Data are available in a public, open access repository: https://osf.io/hjybx/.

Legends

Figure 1. Proportion of adherence to AMSTAR2 items.► : 7 critical domains identified by AMSTAR2.

Figure 2. Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group. *Items statistically significant with the Bonferroni correction for multiple testing ($p \leq 0.003$).

BIBLIOGRAPHY

1. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nat Publ Gr*. 2018;555. doi:10.1038/nature25753.
2. 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.
3. Dieckmann NF, Malle BF, Bodner TE. An Empirical Assessment of Meta-Analytic Practice. 2009. doi:10.1037/a0015107.
4. Zhang H, Han J, Zhu Y-B, et al. Reporting and methodological qualities of published surgical meta-analyses. 2016. doi:10.1016/j.jclinepi.2015.06.009.
5. Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q*. 2016;94(3):485-514. doi:10.1111/1468-0009.12210.
6. Nuijten MB. Practical tools and strategies for researchers to increase replicability. *Dev Med Child Neurol*. October 2018. doi:10.1111/dmcn.14054.
7. Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Syst Rev*. 2018;7(1). doi:10.1186/s13643-018-0699-4.
8. 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;100:103-110. doi:10.1016/J.JCLINEPI.2018.01.003.
9. Franco JVA, Garrote VL, Escobar Liquitay CM, Vietto V. Identification of problems in search strategies in Cochrane Reviews. *Res Synth Methods*. 2018;9(3):408-416. doi:10.1002/jrsm.1302.
10. Ahn S, Ames AJ, Myers ND. A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Rev Educ Res*. 2012;82(4):436-476. doi:10.3102/0034654312458162.
11. Dechartres A, Atal I, Riveros C, Meerpohl J, Ravaud P. Association Between Publication Characteristics and Treatment Effect Estimates. *Ann Intern Med*. 2018;169(6):385. doi:10.7326/M18-1517.
12. Walpole SC. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. *J Clin Epidemiol*. 2019;111:127-134. doi:10.1016/j.jclinepi.2019.03.004.
13. Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods*. 2017. doi:10.1002/jrsm.1240.
14. 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.
15. Stanley TD, Carter EC, Doucouliagos H. What meta-analyses reveal about the replicability of psychological research. *Psychol Bull*. 2018;144(12):1325-1346. doi:10.1037/bul0000169.
16. 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(2):239-270. doi:10.1080/00273171003680187.
17. 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. doi:10.1016/j.jclinepi.2017.05.002.
18. Johnson BT, Low RE, MacDonald H V. Panning for the gold in health research: Incorporating studies' methodological quality in meta-analysis. *Psychol Health*. 2015;30(1):135-152. doi:10.1080/08870446.2014.953533.
19. 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(7):e1000100. doi:10.1371/journal.pmed.1000100.
20. 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:j4008. doi:10.1136/bmj.j4008.
21. 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. doi:10.1016/J.JCLINEPI.2019.06.014.
22. 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. doi:10.1186/1471-2288-7-10.
23. Stroup DF. Meta-analysis of Observational Studies in Epidemiology<SUBTITLE>A Proposal for Reporting</SUBTITLE>. *JAMA*. 2003;283(15):2008. doi:10.1001/jama.283.15.2008.
24. Coxeter P, Del Mar CB, McGregor L, Beller EM. Interventions to facilitate shared decision making to address antibiotic use for acute respiratory infections in primary care. *Cochrane database Syst Rev*. 2015;11(11):CD010907. doi:10.1002/14651858.CD010907.pub2.www.cochranelibrary.com.
25. 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(6):444-455. doi:10.1016/j.jaac.2016.03.007.
26. Cortese S, Moreira-Maia CR, St Fleur D, Morcillo-Peñalver C, Rohde LA, Faraone S V. Association between ADHD and obesity: A systematic review and meta-analysis. *Am J Psychiatry*. 2016;173(1):34-43. doi:10.1176/appi.ajp.2015.15020266.
27. Aytug ZG, Rothstein HR, Zhou W, Kern MC. Revealed or concealed? transparency of procedures, decisions, and judgment calls in meta-analyses. *Organ Res Methods*. 2012;15(1):103-133. doi:10.1177/1094428111403495.
28. Allers K, Hoffmann F, Mathes T, Pieper D. Systematic reviews with published protocols compared to those without: more effort, older search. *J Clin Epidemiol*. 2018;95:102-110. doi:10.1016/J.JCLINEPI.2017.12.005.
29. Yaylali IE, Alaçam T. Critical Assessment of Search Strategies in Systematic Reviews in Endodontics. *J Endod*. 2016;42:854-860. doi:10.1016/j.joen.2016.02.018.
30. Hohn RE, Slaney KL, Tafreshi D. Primary study quality in psychological meta-analyses: An empirical assessment of recent practice. *Front Psychol*. 2019;9(JAN). doi:10.3389/fpsyg.2018.02667.
31. Ahn S, Becker BJ. Incorporating Quality Scores in Meta-Analysis. *J Educ Behav Stat*. 2011;36(5):555-585. doi:10.3102/1076998610393968.
32. 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(AUG):1395. doi:10.3389/fpsyg.2017.01395.
33. Hutton B, Wolfe D, Moher D, Shamseer L. 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(2):46-52. doi:10.1136/eb-2017-102666.
34. Robson RC, Pham B, 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;106:121-135. doi:10.1016/j.jclinepi.2018.10.003.
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(1):116-134. doi:10.1037/met0000197.

Table 1. AMSTAR2 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 assessed in individual studies being included in the review (item 9)
- Appropriateness of meta-analytical methods (item 11)
- Consideration of risk of bias when interpreting the results of the review (item 13)
- Assessment of the presence and likely impact of publication bias (item 15)

Noncritical domains

- Research question and inclusion criteria based on the components of PICO (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 MA (item 12)
- Explanation for any heterogeneity observed in the results (item 14)
- Report any potential sources of conflict of interest (item 16)

Table 2. General characteristics of the MAs

	Characteristics	Category	n	Number (Percent)	Median (P25-P75)
Article					
	Mention of the use of a guideline other than PRISMA	Moose	23	17 (73.9)	
		Mars		2 (8.7)	
		Cochrane		1 (4.35)	
		Quorom		1 (4.35)	
		Strobe		1 (4.35)	
		Center 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)	
	Number of studies included in the first MA	1-3	206	11 (5.3)	
		4-9		57 (27.7)	
		≥10		138 (67)	
	Assessment of the risk of bias in individual studies	Yes	206	111 (53.9)	
	Tool used to assess the risk of bias	RoB tool	95	36 (37.9)	
		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)	
Authors					
	Number of authors	1	206	12 (5.8)	
		2-3		60 (29,1)	
		4-6		98 (47.6)	
		≥7		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)	
	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)	
Journal					
	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)	
Methodological quality					
	AMSTAR 2 tool	High quality	206	1 (0.5)	
		Moderate quality		2 (1)	
		Low quality		8 (4)	
		Critically low quality		195 (94.5)	

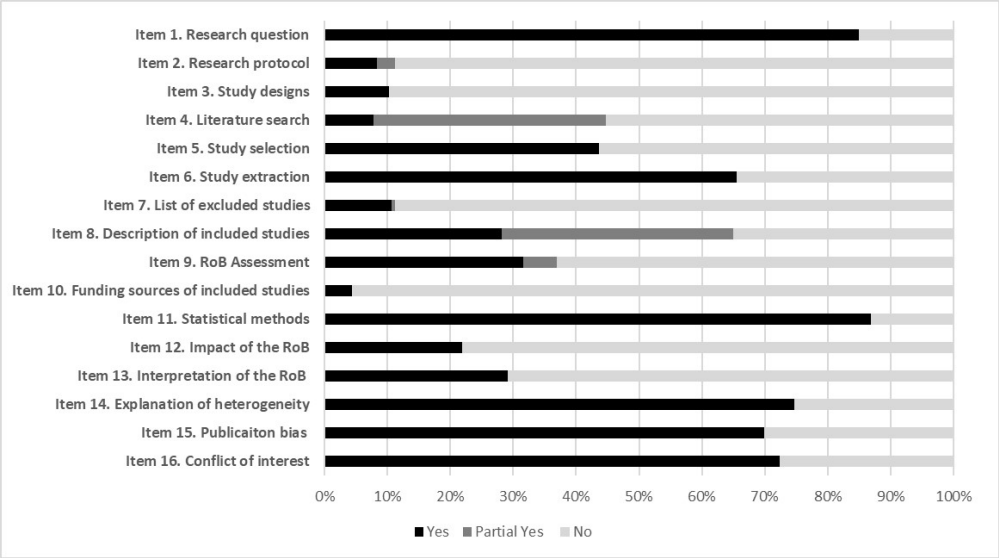


Figure 1. Proportion of adherence to AMSTAR2 items. >> : 7 critical domains identified by AMSTAR2.

98x54mm (300 x 300 DPI)

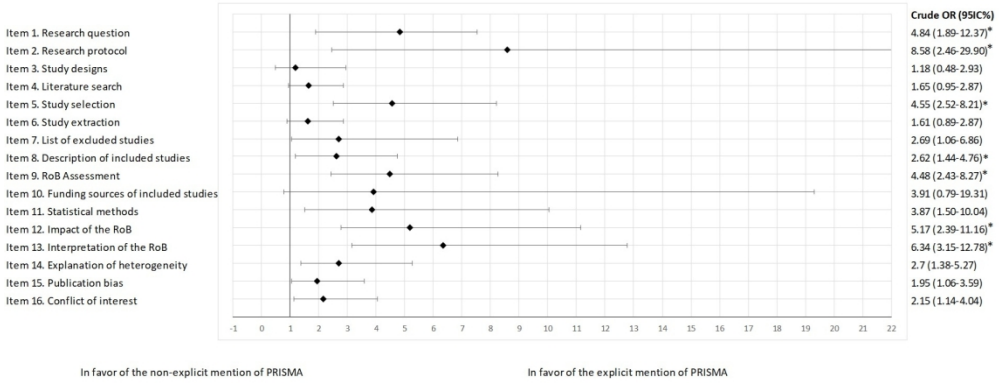


Figure 2. Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group. *Items statistically significant with the Bonferroni correction for multiple testing (p≤0.003).

161x66mm (300 x 300 DPI)

Register OSF

Study information

Title

Assessment of the reporting and methodological qualities and associated factors of a sample of meta-analyses recently indexed in PsycINFO (2016).

Authors

Victoria Leclercq – Charlotte Beaudart – Véronique Rabenda – Sara Ajamieh - Ezio Tirelli – Olivier Bruyère

Background

For scientists, searching for current best evidence has become a real challenge for scientists given the quasi limitless number of published articles (more than 1 270 000 in 2014, according to Thomson Reuteur's Web of Science). When facing a problematic implying a decision, scientists need documents and results oh high and reliable scientific value, as promoted by the evidence-based medicine movement (EBM). EBM is defined as the practice of medicine-based on knowledge and understanding of the literature in order to support clinical decisions (Guyatt et al. 2015). Following evidence hierarchy of EBM, systematic reviews (SRs) and meta-analyses (MAs) are considered the best level of evidence. Nowadays, in diverse disciplines, many researchers base their own research on the results of these SRs and MAs. The Cochrane collaboration adopted the definition of Antman (1992) and Oxman (1993) for the SR: "A systematic review attempts to collate all empirical evidence that pre-specified eligibility criteria in order to answer a specific research question" (Higgins & Green 2011) and the definition of Glass (1976) for MA : "Meta-analysis is the use of statistical methods to summarize the results of independent studies" (Higgins & Green 2011).

Some researchers have highlighted an increase in the publication rate of 2728% for SR (1024 articles in 1991 and 28 959 in 2014) and 2635% for MA (334 articles in 1991 and 9135 in 2014) (Ioannidis 2016). Several reasons could explain this phenomenon. In particular, fewer resources are necessary to perform SRs and MAs, which generally (yield) are worth high citation rates (contributing to increase the impact factor of the journal where they are published (Ioannidis 2016). However, an increasing number of studies have highlighted weaknesses in the design, conduct, analysis, and reporting of MAs published in many scientific fields (Zhu et al. 2016; Cullis et al. 2017; Peters et al. 2015; Zhang et al. 2016; Gagnier & Kellam 2013).

Two tools have been developed to evaluate the quality of the methodology ((AMSTAR, "A Measurement Tool to Assess systematic Reviews" (Shea et al. 2007)) and recently its update (AMSTAR 2 (Shea et al. 2017))) and another one for the quality of the reporting (PRISMA, "Preferred Reporting Items for Systematic Reviews and Meta-analyses" ((Moher et al. 2009)) of SRs and MAs. AMSTAR, a 11-item measuring tool aiming to assess the methodologic quality of MAs (Zhang et al. 2016), has been shown to be reliable and valid (Shea et al. 2009). AMSTAR 2, a 16-item measuring tool aiming to assess the methodologic quality of MAs of randomized and no randomized studies (Shea et al. 2017). PRISMA comprises a list of 27 items that are recommended to be used in the reporting of a MA in order to ensure that the article contains all relevant information (Moher et al. 2009). Several studies have already evaluated the quality of MAs published in specific medical fields such as surgery (Cullis et al. 2017; Zhang et al. 2016), depression (Zhu et al. 2016), orthopaedic (Gagnier & Kellam 2013) or even otorhinolaryngologic disorders (Peters et al. 2015). To the best of our knowledge, there are no such studies available in the field of psychological science.

In line with the EBM movement, the American Psychological Association (APA) has defined in 2006 the movement of Evidence-Based Practice in Psychology with the purpose "to promote effective psychological practice and enhances public health by applying empirically supported principles of psychological assessment, case formation, therapeutic relationship, and intervention"(American Psychological Association 2006). The American Psychological Association has brought out some benefits to the use of Reporting Standards whose the salutary effect on the way research has been conducted (Cooper 2008). The PRISMA statement could also have a positive effect on the methodological quality of the studies.

A growing meta-research literature has assessed the quality of empirical and experimental psychological studies in often large samples of articles (Ioannidis 2012; Bakker & Wicherts 2011; Oliveras et al. 2017; Stanley et al. 2017). It has revealed and quantified numerous methodological deficiencies, such as an inappropriate use of statistics, high rates of statistical mistakes, a frequent lack of statistical power (along with the neglect of effect size considerations) or the unambiguous presence of methodological biases, to mention but a few of them. Interestingly, a recent study conduct by Fanelli and co-workers (Fanelli, Costas, & Ioannidis, 2017) have highlighted differences in the risk of bias (poor estimate of the magnitude of effect size due to, for example, lower inclusion of grey literature, US effect or industry bias…) between the classical disciplines, the risk being highest in the social sciences (to which

psychology belongs). These differences could reflect dissimilar research practices documented in primary studies (e.g. higher publication bias in some disciplines) or distinct procedural choices in meta-analyses (e.g. lower inclusion of grey literature in some disciplines) (Fanelli et al., 2017).

To our knowledge, no studies have however been conducted in order to evaluate the quality of MAs published in the field of psychology. With this research project, our aim is to evaluate the quality of MAs and identify its associated factors published in the psychological or psychology-related field on the PsycINFO database during the year 2016.

Objectives

The objective of this research is to assess the factors associated with the quality of recent MAs indexed in PsycINFO for the year 2016 using two samples of MAs; one composed of MAs claiming to follow the PRISMA statement and the other one including MAs ignoring it.

Our research will be organized in three sub-studies:

1. The assessment of the reporting quality of MAs that claim to use the PRISMA statement compared to those that do not use PRISMA through the PRISMA statement;
2. The assessment of the methodological quality of MAs that claim to use the PRISMA statement compared to those that do not use PRISMA through the AMSTAR 2 tool;
3. The identification of potential factors associated with the quality of MAs.

Research questions

Based on our objectives, our research questions are the following:

1. What is the relationship between of the use of the PRISMA checklist on the reporting quality of MAs?
2. What is the relationship between of the use of the PRISMA checklist on the methodological quality of MAs?
3. What are the potential factors (e.g. publishing journal's impact factor, pre-registration of the study, experience of the first author…) associated with the quality of MAs?

Hypotheses

1. Comprehensive and transparent reporting is necessary to assessing the methodological quality of MAs (Page et al. 2016). They are some articles that highlighted that MAs have a poor reporting quality in the medical literature and the score of PRISMA Statement that were found is between 16,8 and 23/27 points ((Tunis et al. 2013; Fleming et al.

2013; Peters et al. 2015; Zhang et al. 2016; Gagnier & Kellam 2013; Adie et al. 2015; Zhu et al. 2016). We make the assumptions that the use of PRISMA statement improves the reporting quality of MAs using it. This is supported by another study showing that the PRISMA scores is higher by 1 point for the MAs for which authors claimed having used PRISMA statement (Zhu et al. 2016).

2. Meta-research has revealed that the psychological literature present an unsatisfactory level of methodological quality (Bakker & Wicherts 2011; Ioannidis 2012; Oliveras et al. 2017; Stanley et al. 2017). It is therefore possible that this is also true for the MA published in psychology and related fields. In recent studies analyzing the quality of Mas in a number of health-related fields AMSTAR scores have been found to fall between 3,7 and 7,8/11 points (Zhang et al. 2016; Adie et al. 2015; Gagnier & Kellam 2013; Klimo et al. 2014). Since the AMSTAR 2 tool was recently published, it is probable that no study has yet evaluated the quality of the MAs with this tool. Note that it is likely that the use of PRISMA statement exerts a positive influence on the quality of Mas using it. This is supported by a recent study on depression showing that the AMSTAR scores for the MAs for which authors claimed having used the PRISMA statement, reach an average of 0.4 point higher than those which did not use PRISMA (Zhu et al. 2016).
3. They are some potential factors that could correlate with (and possibly influence) the quality of MAs. More specifically, on the basis of previous meta-research studies we make the assumptions that the following factors are positively associated with the measures of quality of MAs: h-index of the first author (Cullis et al. 2017), experience of the principal author in MAs (Zhang et al. 2016), affiliation of the authors to a university (Cullis et al. 2017), publishing journal's impact factor (Cullis et al. 2017), PRISMA endorsement by the journal publishing (Cullis et al. 2017), funding sources described (Gagnier & Kellam 2013), Cochrane collaboration (Adie et al. 2015; Cullis et al. 2017; Zhu et al. 2016), number of pages of the manuscript (Adie et al. 2015; Cullis et al. 2017), pre-registration of the study (Cullis et al. 2017; Zhang et al. 2016; Zhu et al. 2016), non-Asian origine (Zhang et al. 2016) and meta-analyses of randomized controlled trials (Zhang et al. 2016; Zhu et al. 2016). Furthermore, we hypothesize that the following variables will be also associated with the quality of MAs: open access of the publication, open data (or data sharing), the field of psychology, number of individual studies in each MA, number of databases used, assessment of the quality of

individual studies and related tools used, pooling methods used to combine the data, assessment of the publication bias and related method used, assessment of the heterogeneity and related method used, the statistical software used and tendency of the conclusion.

Sampling plan

Existing Data

The tests can be considered as confirmatory given similar studies that have recently been conducted in the medical field.

Explanation of existing data

This is not applicable for our research protocol because no data have been collected so far.

Data collection procedures

Data collection

Data will be collected from a random sample of 200 MAs that will be divided into 2 groups (use of PRISMA vs no use of PRISMA).

Protocol selection of meta-analyses articles

All MAs performed on human subjects and published in English in 2016 in the electronic database PsycINFO will be searched. The electronic search strategy was developed with co-authors and the assistance of a librarian is available in *table 1*.

Table 1: Search strategy

1	meta analysis.md. (15886)
2	meta analysis/ (3940)
3	meta analys*.mp. (24573)
4	data pooling*.mp. (50)
5	2 or 3 or 4 (24599)
6	5 not 1 (10725)
7	1 or 6 (26611)
8	limit 7 to (English and human and yr="2016") (2159)

A total of 2159 potentially relevant MAs were identified in the PsycINFO database. Two investigators will independently review each title and abstract in order to exclude irrelevant articles and to only select the studies that meet inclusion criteria (full inclusion and exclusion criteria are available in Table 2). All discrepancies in opinion regarding the selection of articles

will be resolved through discussion and consensus between the two investigators; any persistent disagreement will be solved with the intervention of a third person (an expert).

Table 2 : eligibility criteria

Inclusion criteria
- Meta-analysis
- Articles published in the PsycINFO-database
- Published between 01.01.2016 to 31.12.2016
- English
Exclusion criteria
- Overview, review
- Meta-synthesis
- Qualitative meta-analysis
- Umbrella review
- Meta-analysis of meta-analyses
- Systematic review without meta-analysis
- Protocol of meta-analysis
- Network meta-analysis
- Activation likelihood Estimation Meta-analysis (ALE MA)
- Signed differential mapping meta-analysis (SMD MA)
- Voxel wise meta-analysis
- Individual patient data meta-analysis (IPD MA)
- Genetic association study (GWAS), genetic study
- Multi-level meta-analysis
- Update
- Letter, comment, abstract, chapter, erratum, dissertation or editorial journal

Choice of language for inclusion was based on expertise within our research team, due to budget constraints, limited time and resources.

A flowchart with the number of included studies will be elaborated. The reason of exclusion of articles will be presented at the step of full-text selection.

Once all MAs will be identified, we will randomly select a minimum of 200 articles as follows. All references of articles will be indexed in an Excel file and randomly assigned to a number. Then, we will rank the articles in ascending order. Two blinded researchers, with the intervention of a third researcher in case of disagreement, will classify MAs that meet inclusion criteria in either the group "with PRISMA" or "without PRISMA" until each group will contain a minimum of 100 MAs. Kappa statistics will be used to test inter-rater agreement.

Data extraction

Relevant data will be extracted from the full texts of all selected articles in a standardized Microsoft Excel spreadsheet by two independent investigators trained for this data extraction. We will record the following factors that might influence the quality of the MAs: characteristics of the manuscripts, characteristics of the study, objective(s) of the study, statistical analyses, characteristics of the protocol and items of PRISMA statement, AMSTAR tool and AMSTAR2 tool. All data extracted will be detailed in appendices. If any disagreements were to be observed between the two reviewers, they will be resolved by discussion, if necessary with the intervention of a third reviewer. Kappa statistics and absolute agreement (%) will be used to assess reproducibility.

Sample size

A first exploratory search in PsycINFO has yielded approximately 2000 articles, which are impossible to analyze for us in a reasonable period of time. We elected to randomly (see below for the method of randomization) select 200 MA articles (until each group, PRISMA and NO PRISMA, will contain a minimum of 100 MAs) from all eligible MAs published in 2016 and indexed in PsycINFO. There is no global MAs offering a synthetic effect size (of the published differences between the two samples) that could have been used to determine a priori a sample size allowing the detecting of a significant difference (power analyses). The chosen sample size can minimally detect a medium effect size (Cohen's d = 0.46065; as computed via G*Power) using a two-tailed Student t-test for independent groups taken at an alpha error probability of 0.05 and a power (1-beta error probability) of 0.90 (critical t = 1.9720). Note that smaller effects sizes cannot be detected (if existing) with such sample (n=100). The meaning and practical significance of the empirically obtained effect size will be discussed.

Sample size rationale

Considering the power analysis described above and constraints in terms of time, financial resources and staff, we will conduct this research on about 200 articles. We think that this will necessitate more than 500 hours of coding for each assessor.

Stopping rule

This is not applicable for our research protocol.

Variables

Manipulated variables

This cannot be applied to the present research protocol.

Measured variables

In order to verify our hypotheses, we will assess not only the quality of reporting and of conduct of MAs but also a set of variables to identify the potential factors that are associated with the quality of MAs.

Assessment of reporting quality

Eligible papers will first be assessed with the PRISMA statement. Each individual item of the PRISMA statement will be answered by "yes" or "no" response, respectively depending on the item being or not fulfilled and will be coded "1" or "0". The total score of the PRISMA statement is the addition of all items coded 1 with a maximum of 27 points.

Assessment of methodology quality

Eligible papers will then be assessed with AMSTAR 2 tool. Each individual item of the AMSTAR tool will be answered by a "yes", "partial yes" or "no" response, respectively depending on the item being or not fulfilled and will be coded "1", "0.5", "0". The total score given by the tool is the addition of all items coded 1 and 0.5 with a maximum of 16 points.

Eligible papers will also be assessed with the AMSTAR tool. Each individual item of the AMSTAR tool will be answered by a "yes" or "no" response, depending on the item being or not fulfilled and will be coded "1" or "0". The total score given by the tool is the addition of all items coded 1 with a maximum of 11 points.

Identification of potential factors associated with the quality of MA

All factors will be assessed by three types of variables: dichotomous variables, quantitative variables and text variables (open questions).

Dichotomous variables

Dichotomous variables will be coded as follows:

1: Yes, it features the characteristic that we seek.

0: No, it does not feature the characteristic that we seek.

99: Not reported, the characteristic that we seek is not available.

88: Not applicable.

The variables that are concerned are the following: author's experience in meta-analysis, affiliation of the authors to a university, PRISMA's recommendation by the journal, restriction of the word count by the journal, declaration of conflict of interest, declaration of funding sources, Cochrane collaboration, open access, open data, registration of the study in a database, evaluation of the quality of study, use of reporting or methodology guideline, the type of study (randomized controlled trials (RCT) or not), evaluation of publication bias, evaluation of heterogeneity, presence of a protocol and the conclusion supports the assumptions.

Quantitative variables

They will be encoded with numerical values and their units of measurements.

The variables that are concerned are the following: h-index (an author-level metric), number of authors, impact factor of the journal (which reflects the frequency with which the average article in a journal has been cited in a particular year) and number of database consulted.

Qualitative variables

The relevant variables are the following: the continent where the study was conducted (Europe, Asia, Africa, America, Oceania), number of study included in each MA (0-3; 4-9; ≥10) and the pooling method used (random effect model, fixed effect model or mixed effect model).

Text variables

The remaining variables are recorded as a text variable. The variables that are concerned are the following: the name of the tool used in order to assess the quality of individual study, the name of the database(s) searched, the PsycINFO classification, the name of the guideline used, the method used to evaluate the publication bias, the method used to assess the heterogeneity and the statistical software. These text variables will then be categorized.

If the data is not available, it will be coded 88.

Indices

This is not applicable for the present research protocol.

Design plan

Study type

This is an observational study. Data are collected from meta-analysis articles.

Blinding

The selection of MAs reviews by title and abstract and by full-text will be done independently by two investigators.

Study design

This is a cross-sectional study.

Randomization

We will randomly MAs articles to get a minimum of 100 MAs in each group. All references of articles (n= probably more than 2000) will be indexed in an Excel file and randomly assigned to a number. Then, we will rank the articles in ascending order. Two blinded researchers, with the intervention of a third researcher in case of disagreement will classify MAs that meet inclusion criteria in either group “with Prisma” and “without Prisma” until each group will contain a minimum of 100 articles MAs.

Analysis plan

Statistical model

The characteristics of all individual studies will first be presented. All quantitative variables that follow a normal distribution will be reported as mean and standard deviation and those that do not follow a normal distribution will be represented as median and percentile (P25 and P75). Distribution will be considered as normal if data meet 3 of the 4 following conditions: the mean is close to the median, the Shapiro-Wilk normality test yields a p-value </=0.05, the curve of the variable follows the normal (or Gaussian) distribution and the linearity of the QQ-Plots is respected. Qualitative and dichotomous variables will be reported as numbers and frequencies. The results of the quality assessment of MAs with the PRISMA statement, AMSTAR tool and AMSTAR 2 will be reported, for quantitative variable, as number and frequency for each item and as mean or as median for the total score. The data will be presented and analyzed using a

star chart. A star chart is a graphical tool that will allow us to represent and compare the percentages of item of PRISMA statement, AMSTAR tool and AMSTAR 2 met by the MAs.

To verify our first and second hypothesis, the reporting and methodological qualities of the individual studies will be compared between the studies that report using the PRISMA checklist and the studies that do not. Comparisons of means between the two groups will be calculated using the Student t-test if for independent groups if the score of PRISMA, AMSTAR and AMSTAR 2 are normally distributed and the Mann-Whitney test if the score of PRISMA, AMSTAR and AMSTAR 2 are not normally distributed. To test the association between the use of the PRISMA statement and the different items of PRISMA, AMSTAR and AMSTAR 2, we will be used a logistic regression.

To test our third hypothesis, factors (all the data detailed in the measured variables part) with potential influence on the quality of studies (mean score of AMSTAR, mean score of AMSTAR 2, independent variable) will be identified with a univariate linear regression. The variables with a p-value lower than 0.1 will be combined in stepwise backward multiple regression analysis. A p-value </=0.05 will be considered as significant.

Transformations

Each item of the PRISMA & AMSTAR checklists will be coded with the following meaning:

1 = Yes or not applicable

0 = No

We will also sum up all items coded 1, with a maximum score of 27 or 11, respectively.

Each item of AMSTAR 2 will be coded with the following meaning:

1 = Yes or not applicable

0.5 = Partial yes

0 = No

We will also sum up all items coded 1 and 0.5, with a maximum score of 16.

See appendix for the details of all transformations for each variable.

Follow-up analyses

All follow-up analyses are described above.

Inference criteria

Not applicable for our research protocol.

Data exclusion

No data will be excluded from our database.

Missing data

Missing data may have an impact on the analysis and on the interpretation of the results. Some of the extracted data may not be available (h-index, impact factor…). After data encoding, a quality control will be done, at database-level, in order to check for outliers, coding error and missing values. In case of incomplete information, we will contact the authors.

Exploratory analysis

The exploratory analyses will be considered, based on the results obtained.

If there is a statistically significant difference in quality between MAs which report using PRISMA and those which do not, we will consider carrying out the following analyses. A logistic regression will be carried out in order to describe the relation between the dichotomous dependent variables (PRISMA vs No PRISMA) and all potential explanatory variables (all the data are detailed in the measured variables part). The variables with a p-value lower than 0.1 will be combined in stepwise backward multiple logistic regression analysis. A p-value <0.05 will be considered significant on statistical analyses.

Scripts

Upload an analysis script with clear comments

Not available at the moment.

Other

We would like to acknowledge Pr. Anne-Françoise Donneau for interesting discussions about some aspects on the planned statistical analyses and Ms Nancy Durieux for her assistance in the building of our strategy, in terms of electronic literature search.

Bibliography

Adie, S. et al., 2015. Quality of Conduct and Reporting of Meta-analyses of Surgical Interventions. *Annals of Surgery*, 261(4), pp.685–694. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000658-201504000-00012 [Accessed June 14, 2017].

American Psychological Association, 2006. Evidence-Based Practice in Psychology APA Presidential Task Force on Evidence-Based Practice. *American Psychologist*, 61(4), pp.271–285. Available at: http://www.apa.org/pubs/journals/features/evidence-based-

statement.pdf [Accessed January 10, 2018].

Bakker, M. & Wicherts, J.M., 2011. The (mis)reporting of statistical results in psychology journals. Available at: https://link.springer.com/content/pdf/10.3758%2Fs13428-011-0089-5.pdf [Accessed July 31, 2017].

Cooper, H., 2008. Reporting Standards for Research in Psychology: Why Do We Need Them? What Might They Be? *American Psychologist*, 63(9), pp.839–851. Available at: https://www.apa.org/pubs/authors/jars.pdf [Accessed January 15, 2018].

Cullis, P.S., Gudlaugsdottir, K. & Andrews, J., 2017. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. *PLoS ONE*, 12(4). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5383307/pdf/pone.0175213.pdf [Accessed June 14, 2017].

Fanelli, D., Costas, R. & Ioannidis, J.P.A., Meta-assessment of bias in science. Available at: http://www.pnas.org/content/114/14/3714.full.pdf [Accessed December 18, 2017].

Fleming, P.S. et al., 2013. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. *The Angle Orthodontist*, 83(1), pp.158–163. Available at: http://www.angle.org/doi/10.2319/032612-251.1 [Accessed May 24, 2018].

Gagnier, J.J. & Kellam, P.J., 2013. Reporting and Methodological Quality of Systematic Reviews in the Orthopaedic Literature. *The Journal of Bone and Joint Surgery-American Volume*, 95(11), pp.e77-1–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23780547 [Accessed June 14, 2017].

Guyatt, G. et al., 2015. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Medicine*, Available at: https://jamaevidence.mhmedical.com/book.aspx?bookID=847 [Accessed September 1, 2017].

Higgins, J. & Green, S., 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011].* The Cochra., Available at: http://handbook.cochrane.org.

Ioannidis, J.P.A., 2016. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, 94(3), pp.485–514. Available at: http://doi.wiley.com/10.1111/1468-0009.12210.

Ioannidis, J.P.A., 2012. Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), pp.645–654. Available at: http://pps.sagepub.com/ [Accessed

July 31, 2017].

Klimo, P. et al., 2014. Methodology and reporting of meta-analyses in the neurosurgical literature. *Journal of neurosurgery*, 120(4), pp.796–810. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24460488.

Moher, D. et al., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). *Physical Therapy*, 89(9), pp.873–880. Available at: http://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1000097&type=printable [Accessed August 25, 2017].

Oliveras, I., Losilla, J.-M. & Vives, J., 2017. Methodological quality is underrated in systematic reviews and meta-analyses in health psychology. *Journal of Clinical Epidemiology*, 86, pp.59–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28499846 [Accessed November 28, 2017].

Page, M.J. et al., 2016. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study N. Low, ed. *PLOS Medicine*, 13(5), p.e1002028. Available at: http://dx.plos.org/10.1371/journal.pmed.1002028 [Accessed May 24, 2018].

Peters, J.P.M. et al., 2015. Reporting quality of systematic reviews and meta-analyses of otorhinolaryngologic articles based on the PRISMA statement. *PLoS ONE*, 10(8). Available at: http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0136540&type=printable [Accessed June 14, 2017].

Shea, B.J. et al., 2017. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, p.j4008. Available at: http://www.bmj.com/content/bmj/358/bmj.j4008.full.pdf [Accessed November 28, 2017].

Shea, B.J. et al., 2009. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), pp.1013–1020. Available at: http://ac.els-cdn.com/S0895435608003259/1-s2.0-S0895435608003259-main.pdf?_tid=7ad1b7aa-558b-11e7-a3dc-00000aacb361&acdnat=1497944501_0f82a189082ff8754f64a91d9670402e [Accessed June 20, 2017].

Shea, B.J. et al., 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*, 7, p.10. Available at: https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-7-10?site=bmcmedresmethodol.biomedcentral.com [Accessed August 25, 2017].

Stanley, T.D., Carter, E.C. & Doucouliagos, H., 2017. What Meta-Analyses Reveal about the Replicability of Psychological Research. Available at: http://www.deakin.edu.au/__data/assets/pdf_file/0007/1198456/WhatMeta-AnalysesReveal_WP.pdf [Accessed January 10, 2018].

Tunis, A.S. et al., 2013. Association of Study Quality with Completeness of Reporting: Have Completeness of Reporting and Quality of Systematic Reviews and Meta-Analyses in Major Radiology Journals Changed Since Publication of the PRISMA Statement? *Radiology*, 269(2), pp.413–426. Available at: http://pubs.rsna.org/doi/10.1148/radiol.13130273 [Accessed May 24, 2018].

Zhang, H. et al., 2016. Reporting and methodological qualities of published surgical meta-analyses. Available at: http://ac.els-cdn.com/S0895435615003121/1-s2.0-S0895435615003121-main.pdf?_tid=4e369bf4-50e8-11e7-9b21-00000aab0f26&acdnat=1497434613_6d09bcd5abe078df3344e91ace4b09d1 [Accessed June 14, 2017].

Zhu, Y. et al., 2016. Is the best evidence good enough: Quality assessment and factor analysis of meta-analyses on depression. *PLoS ONE*, 11(6). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4919061/pdf/pone.0157808.pdf [Accessed June 14, 2017].

Appendix
Explication of the data extraction form

Name	Explication	Description
Name of the reviewer	Name of the reviewer	Text
Study ID	Reference number of the article	Text
Inclusion of the article	Inclusion of the article based on the selection criteria	1 = Yes 0 = No
If excluded, indicate the reason of exclusion	Reason of exclusion	The reason of exclusion 88 = if not applicable
Use of PRISMA	The authors declared the use of PRISMA	1= Yes 0 = No

1. Characteristics of the manuscript

Name	Explication	Description
DOI of the article	Unique identifier of the article	Text
Year of publication	Publishing year of the manuscript	Text
Author's name	Name of the first author	Text
Author's h-index	H-Index of the first author (Scopus)	Quantitative variable
Author's experience	Number of meta-analyses from the same author(s) (Scopus)	Quantitative variable
Affiliation of the authors to a university	Affiliation of the authors to a university	1 = Yes 0 = No 99 = Not reported
Number of authors	Total number of authors	Quantitative variable
Contribution of authors	Details of the authors' contribution	1 = Yes 0 = No 99 = Not reported
Journal's name	Name of the journal	Text
Journal's Impact factor	The IF of the journal using the ISI Journal Citation Reports 2016 (http://isiknowledge.com)	Quantitative variable
Instruction for authors: PRISMA required?	The journal recommended to use PRISMA statement	1 = Yes 0 = No 99 = Not reported

Instruction for authors: page or word limitation?	Limitation of the number of pages or words	1 = Yes 0 = No 99 = Not reported
PsycINFO classification	Classification of the field of psychology based on the PsycINFO Content Classification Code System	2100 = General Psychology 2200 = Psychometrics & Statistics & Methodology 2300 = Human experimental Psychology 2400 = Animal Experimental & comparative Psychology 2500 = Physiological Psychology & Neuroscience 2600 = Psychology & The Humanities 2700 = Communication Systems 2800 = Developmental Psychology 2900 = Social Processes & Social Issues 3000 = Social Psychology 3100 = Personality Psychology 3200 = Psychological & Physical disorders 3300 = Health & Mental Health Treatment & Prevention 3400 = Professional Psychological & Health Personnel Issues 3500 = Educational Psychology 3600 = Industrial & Organizational Psychology 3700 = Sport Psychology & Leisure 3800 = Military Psychology 3900 = Consumer Psychology 4000 = Engineering & Experimental Psychology 4100 = Intelligent Systems 4200 = Forensic Psychology & Legal Issues
Corresponding author (email address)	Email of the author	Text

Conflict of interest described	Conflict of interest is described	1 = Yes 0 = No 99 = Not reported
Details of conflict of interest	If yes, brief description of conflict of interest	Text
Funding sources described	Funding sources are described	1 = Yes 0 = No 99 = Not reported
Funding sources	If yes, brief description of funding sources	Text
Cochrane collaboration	The study is a Cochrane collaboration	1 = Yes 0 = No
Number of page of manuscript	Total number of pages of the manuscript	Quantitative variable
Open access	The publication is open access?	1 = Yes 0 = No 99 = Not reported
Open data	The data is open access?	1 = Yes 0 = No 99 = Not reported

2. Characteristics of the study

Name	Explication	Description
Registration of the study?	The study was recorded in a specific database.	1 = Yes 0 = No 99 = Not reported
Number of the registration	Registration number of the study	Text 88 = if not applicable
Name of the registry	Name of the registry in which the meta-analysis has been registered	Text 88 = if not applicable
Date of submission of the manuscript	Date of submission of the manuscript	Text (month-year)
Date of publication of the manuscript	Publication date of the manuscript	Text (month-year)
Continent of origin of first author	Continent in which the study has been conducted	Europe – Asia – Africa – America - Oceania
Type of the individual study	Study design of the studies included in the MA	Observational study RCT All types Not specified
Number of databases searched	Number of databases consulted	Quantitative variable

Name of the database	Name of the database searched	Text
Quality of individual study is assessed?	Quality of individual study is assessed	1 = Yes 0 = No
Name of the tool used to assess the quality	Name of the tool used to assess the quality of individual studies	Text variable 88 = If not applicable
Reference to use of the guideline	Reference to use of a guideline	1 = Yes 0 = No
Name of the guideline used	Name of the guideline used (PRISMA, MOOSE, AMSTAR…)	Text 88 = If not applicable
Search strategy	Presence of the complete search strategy	1 = Yes 0 = No 99 = Not reported
Focus of review	Type of the field of psychology	Text

3. Objective of the study

Name	Explication	Description
Main objective	Aim of the study	Text
Primary outcome	Primary outcome of the study disclosed	Text
Secondary outcomes	Secondary outcome of the study disclosed	Text

4. Statistical analyses

Name	Explication	Description
Number of meta-analyses performed	Number of meta-analyses performed in the presented study	Quantitative variable
Number of studies included in each meta-analysis	Number of studies included in each meta-analysis performed in the study	0-3; 4-9; ≥10
Pooling methods	The pooling methods used to combine data	Fixed – Random - Mix
Assessment of the publication bias	The publication bias is evaluated	1 = Yes 0 = No
Method used to assess the publication bias	Method used to assess the publication bias	Text 88 = If not applicable

Assessment of the heterogeneity	The heterogeneity is evaluated	1 = Yes 0 = No
Heterogeneity	Method used to assess the heterogeneity	Text 88 = If not applicable

5. Protocol

Name	Explication	Description
Protocol	The protocol of the study is existent and available	1 = Yes 0 = No
Primary outcome	Primary outcome of the study	Text 88 = If not applicable
Secondary outcome	Secondary outcome of the study	Text 88 = If not applicable

6. Conclusion

Name	Explication	Description
Conclusion	Main conclusion of the study	Text
Trends of the conclusion	The conclusion supports the assumptions	1 = Yes 0 = No

7. PRISMA statement

Name	Explication	Description
P1	TITLE Title	1 = Yes 0 = No
P2	ABSTRACT Structured summary	1 = Yes 0 = No
P3	INTRODUCTION Rationale	1 = Yes 0 = No
P4	Objective	1 = Yes 0 = No
P5	METHODS Protocol and registration	1 = Yes 0 = No
P6	Eligibility criteria	1 = Yes 0 = No
P7	Information sources	1 = Yes 0 = No

P8	Search	1 = Yes 0 = No
P9	Study selection	1 = Yes 0 = No
P10	Data collection process	1 = Yes 0 = No
P11	Data items	1 = Yes 0 = No
P12	Risk of bias in individual studies	1 = Yes 0 = No
P13	Summary measures	1 = Yes 0 = No
P14	Synthesis of results / Planned methods of analysis	1 = Yes 0 = No
P15	Risk of bias across studies	1 = Yes 0 = No
P16	Additional analysis	1 = Yes 1 = Not applicable 0 = No
P17	RESULTS Study selection	1 = Yes 0 = No
P18	Study characteristics	1 = Yes 0 = No
P19	Risk of bias within studies	1 = Yes 0 = No
P20	Results of individual studies	1 = Yes 0 = No
P21	Synthesis of results	1 = Yes 0 = No
P22	Risk of bias across studies	1 = Yes 0 = No
P23	Additional analysis	1 = Yes 1 = Not applicable 0 = No
P24	DISCUSSION Summary of evidence	1 = Yes 0 = No
P25	Limitations	1 = Yes 0 = No
P26	Conclusions	1 = Yes 0 = No
P27	Funding	1 = Yes 0 = No
	Total score	Sum of all items coded "1"

8. Amstar tool

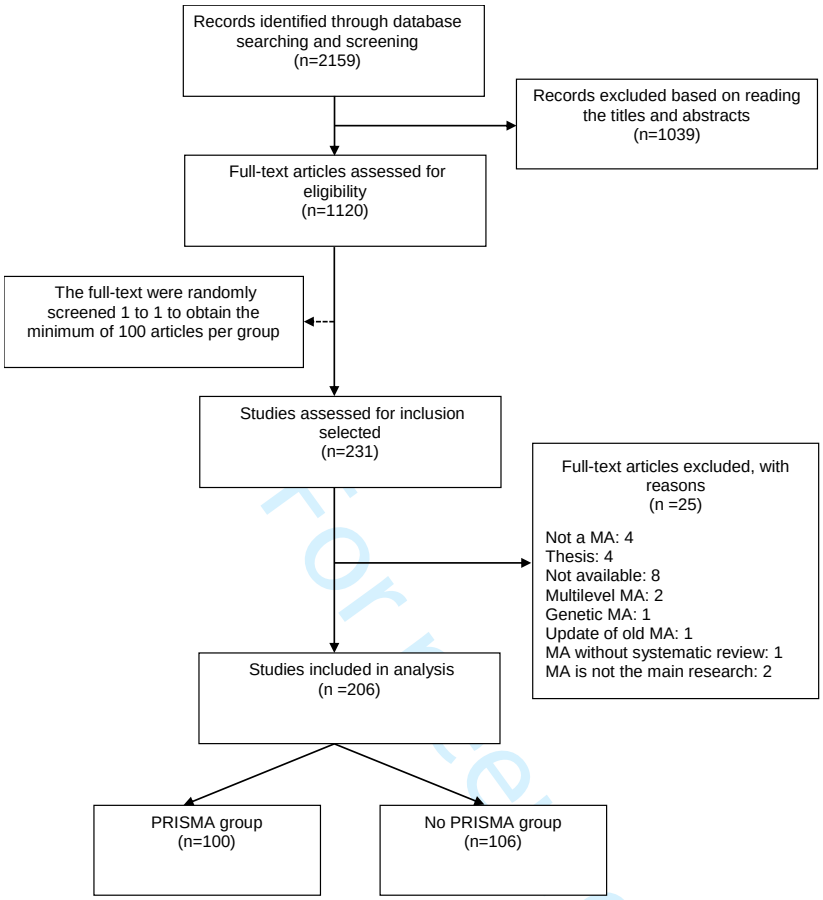
Name	Explication	Description
A1	Was an 'a priori' design provided?	1 = Yes 0 = No
A2	Was there duplicate study selection and data extraction?	1 = Yes 0 = No
A3	Was a comprehensive literature search performed?	1 = Yes 0 = No
A4	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	1 = Yes 0 = No
A5	Was a list of studies (included and excluded) provided?	1 = Yes 0 = No
A6	Were the characteristics of the included studies provided?	1 = Yes 0 = No
A7	Was the scientific quality of the included studies assessed and documented?	1 = Yes 0 = No
A8	Was the scientific quality of the included studies used appropriately in formulating conclusions?	1 = Yes 0 = No
A9	Were the methods used to combine the findings of studies appropriate?	1 = Yes 0 = No
A10	Was the likelihood of publication bias assessed?	1 = Yes 0 = No
A11	Was the conflict of interest included?	1 = Yes 0 = No
	Total score	Sum of all items coded "1"

9. Amstar 2 tool

Name	Explication	Description
AM1	Did the research questions and inclusion criteria for the review include the components of PICO?	1 = Yes 0 = No

AM2	Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	1 = Yes 0.5 = Partial Yes 0 = No
AM3	Did the review authors explain their selection of the study designs for inclusion in the review?	1 = Yes 0 = No
AM4	Did the review authors use a comprehensive literature search strategy?	1 = Yes 0.5 = Partial Yes 0 = No
AM5	Did the review authors perform study selection in duplicate?	1 = Yes 0 = No
AM6	Did the review authors perform data extraction in duplicate?	1 = Yes 0 = No
AM7	Did the review authors provide a list of excluded studies and justify the exclusions?	1 = Yes 0.5 = Partial Yes 0 = No
AM8	Did the review authors describe the included studies in adequate detail?	1 = Yes 0.5 = Partial Yes 0 = No
AM9	Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	1 = Yes 0.5 = Partial Yes 0 = No
AM10	Did the review authors report on the sources of funding for the studies included in the review?	1 = Yes 0 = No
AM11	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	1 = Yes 0 = No
AM12	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	1 = Yes 0 = No
AM13	Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	1 = Yes 0 = No
AM14	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	1 = Yes 0 = No

AM15	If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	1 = Yes 0 = No
AM16	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	1 = Yes 0 = No
	Total score	Sum of all items coded "1" or "0.5"



Supplementary file 2: Flowchart illustrating the MAs Selection

BMJ Open

The methodological quality of meta-analyses indexed in PsycINFO: leads for enhancements

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-036349.R2
Article Type:	Original research
Date Submitted by the Author:	09-Apr-2020
Complete List of Authors:	Leclercq, Victoria; University of Liege, Division of Public Health, Epidemiology and Health Economics Beaudart, Charlotte; University of Liege, Division of Public Health, Epidemiology and Health Economics Ajamieh, Sara; University of Liege, Division of Public Health, Epidemiology and Health Economics Tirelli, Ezio; University of Liege, Department of Psychology Bruyère, Olivier ; University of Liege, Division of Public Health, Epidemiology and Health Economics
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Research methods, Evidence based practice
Keywords:	EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

THE METHODOLOGICAL QUALITY OF META-ANALYSES INDEXED IN PSYCINFO: LEADS FOR ENHANCEMENTS

Short title (Max 70 characters)

The methodological quality of meta-analyses indexed in PsycINFO

Leclercq V[1], Beaudart C[1], Ajamieh S[1,2], Tirelli E[2], Bruyère O[1]

[1]Division of Public Health, Epidemiology and Health Economics, University of Liège, Belgium

[2]Department of Psychology, University of Liège, Belgium

Corresponding author: Victoria Leclercq; victoria.leclercq@uliege.be

ABSTRACT

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

Design We evaluated a random sample of 206 MAs indexed in the PsycINFO database in 2016 through a cross-sectional study. Two authors independently extracted the methodologic characteristics of all MAs and checked their quality according to the 16 items of the AMSTAR2 (A MeaSurement Tool to Assess systematic Reviews) tool for MA critical appraisal. Moreover, we investigated the effect of mentioning 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.

Discussion 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.

Strengths and limitations of this study

- Some studies have highlighted methodological weaknesses in the conduct of systematic reviews (SRs) and meta-analysis (MAs) and we search to have an overview of methodological practice of MAs indexed in PsycINFO according to the tool AMSTAR2 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 PRISMA statement on the methodological quality of MAs**;**
- A sample of 206 Mas indexed in PsycINFO in 2016 and published in English was analyzed;
- Only a random sample of studies indexed in PsycINFO, published in 2016 and in English was included. Therefore, we cannot generalize our findings to MAs published in other years, in other languages or in other databases.

1 INTRODUCTION

Since the definition of meta-analyses (MAs) being introduced by Glass in 1976, MAs conducted in behavioral and social sciences have increased rapidly in number. There were more than 30 000 MAs indexed in 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[1].

However, the validity of MA conclusions relies upon the rigor 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 judgment calls that need to be made by the meta-analysist. Moreover, too many systematic reviews (SRs) and MAs are of low quality[1–5], 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[6–8], an inappropriate literature search[9–12], flaws in the statistical analyses[10,13–16] and an insufficient assessment of the risk of bias of individual studies[10,17,18].

To support researchers in the realization and reporting of MAs, two tools are commonly used. The first is PRISMA ("Preferred Reporting Items for Systematic Reviews and Meta-Analyses"), which was developed in 2009 by Liberati et al[19]. 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[20], 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 behavioral 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 characterize 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 PICO (item 1) if the MAs authors mention the PRISMA statement. This hypothesis, was tested for each of the 16 AMSTAR2 items.

2 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 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[21]. 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 specialized in the field of behavioral 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 systematic review with a MA, indexed in the PsycINFO database, published between January 01, 2016, and December 31, 2016, and published in English. In total, 2159 records were identified. Two authors (V.L & C.B) 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 flowchart in 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 & SA) independently extracted all relevant data from the full texts of all selected articles in a standardized 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 interquartile range of 0.66 [0.40-0.75]) and the Gwet's AC1 (median value with interquartile range of 0.77 [0.69-0.88]) both suggesting a substantial agreement[22], 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 "A MeaSurement Tool to Assess systematic Reviews 2"[20]. AMSTAR2 was a revision of the original AMSTAR instrument[23] developed by Shea et al in 2007, 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 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 2 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 gray 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 4 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 7 critical weaknesses that should reduce confidence in the findings of a review and 9 other items that are considered noncritical weaknesses, as presented in Table 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 hypothesized 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 (JCR) 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 summarized 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 3 of the 4 following conditions: the mean was close to the median, the Shapiro-Wilk normality test yielded a P-value ≥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 dichotomizing the AMSTAR2 items was performed as a sensitivity analysis. Associations were quantified using odds ratios with 95% confidence intervals. A Bonferroni correction was used to adjust the results for multiple testing (16 tests, p-value<0.003). All analyses were performed using SAS 9.4 software.

Patient and public involvement

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

3 RESULTS

Search results

A total of 2159 potentially relevant MAs related to behavioral 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 2. 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 [24] (17, 74%), Mars (2, 9%) and Quorom (1, 4%). Finally, most articles were not available for open access (90.3%), and only one was a Cochrane MA.

Written by one 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 categorized as critically low quality, 8 as low quality, 2 as moderate quality and 1 as high quality. Only one MA [25] provided all the information on all 7 critical domains assessed and was considered high quality according to AMSTAR2. Two additional MAs [26,27] also provided all information on all 7 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 the classification advised by the AMSTAR2 tool (Table 1).

In Figure 1, we summarize 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: 1.90-12.37) 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: 2.46-29.90);
- study selection in duplicate (item 5; OR: 4.55; 95%CI: 2.52-8.21);
- a detailed description of the included studies (item 8; OR: 2.62; 95%CI: 1.44-4.76);

- a satisfactory technique for assessing the risk of bias in individual studies (item 9; OR: 4.48; 95%CI: 2.43-8.27);
- an assessment of the potential impact of risk of bias in individual studies (item 12; OR: 5.17; 95%CI: 2.39-11.16);
- appropriate consideration of the risk of bias in primary studies when interpreting the results (item 13; OR: 6.34; 95%CI: 3.15-12.78).

The results of the sensitivity analysis, performed without dichotomizing the responses modality of AMSTAR2 (yes, partial yes and no) using a logistic regression, showed similar results (Table 1 in 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.

4 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[28]. 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 characterize the methodological quality of MAs indexed in PsycINFO 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 behavioral and social sciences.

- First, no more than 11% of MAs had a research protocol available. However, several scientists[8,16,29] 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, minimizing the risk of bias, avoiding replicate studies and enhancing transparency[7]. 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 a systematic review 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 2 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 gray 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[10]. The search strategy is an essential step of the MA process since the comprehensiveness and completeness of the search[3,30]

is dependent on this strategy. Furthermore, other scientists have highlighted the need to improve research strategies for more comprehensive MAs[9,30].

- 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[17,31], 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 and her team 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[17]. However, there is still a lack of guidance to incorporate these risk of bias assessments into meta-analyses[17,18,32].

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 recognized 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[21,28,31,33,34].

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[25] 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[26,27] 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[20], yet only a few studies from our sample of MAs mention it explicitly. Dechartres and her team stressed the association between publication characteristics and effect estimates[11] 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[11,12,35]. 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[10] but contradict previous studies that have highlighted several flaws in the application and interpretation of statistical analyses in MA[13,14,28,36]. Page et al identified some mistakes in the use of adequate statistical models, the sufficient exploration of subgroup analyses and sensitivity analyses[14]. 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[20]. 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 generalize 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 AC1 was presented along with the Cohen's Kappa. Although Cohen's Kappa is more widely used, Gwet's AC1 is a more robust alternative (less sensitive to data distribution and number of observation)[22]. Moreover, using AMSTAR2, we can investigate the methodological characteristics used to conduct the study (e.g. 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 (e.g. 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.

5 CONCLUSION

This research contributes to raising awareness among researchers about flaws in MAs published in behavioral and social sciences fields, which hopefully increases the adoption of more rigorous research practices. It is clear that meta-analytic 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.

KEYWORDS

Meta-analyses – AMSTAR – meta-research – methodological quality – PRISMA

FOOTNOTES

Conflict of interest and funding: The authors declare no conflict of interest. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors' contribution: VL, CB, ET and OB conceived the study; VL, CB and SA participated in data collection; VL, CB and OB analyzed and interpreted data; VL, CB, ET and OB corrected the manuscripts. All co-authors read and approved the final version of the manuscript.

Data sharing statement: Data are available in a public, open access repository: https://osf.io/hjybx/.

Legends

Figure 1. Proportion of adherence to AMSTAR2 items. ► : 7 critical domains identified by AMSTAR2.

Figure 2. Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group. *Items statistically significant with the Bonferroni correction for multiple testing ($p \leq 0.003$).

BIBLIOGRAPHY

1. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nat Publ Gr*. 2018;555. doi:10.1038/nature25753
2. 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
3. Dieckmann NF, Malle BF, Bodner TE. An Empirical Assessment of Meta-Analytic Practice. 2009. doi:10.1037/a0015107
4. Zhang H, Han J, Zhu Y-B, et al. Reporting and methodological qualities of published surgical meta-analyses. 2016. doi:10.1016/j.jclinepi.2015.06.009
5. Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q*. 2016;94(3):485-514. doi:10.1111/1468-0009.12210
6. Nuijten MB. Practical tools and strategies for researchers to increase replicability. *Dev Med Child Neurol*. October 2018. doi:10.1111/dmcn.14054
7. Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Syst Rev*. 2018;7(1). doi:10.1186/s13643-018-0699-4
8. 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;100:103-110. doi:10.1016/J.JCLINEPI.2018.01.003
9. Franco JVA, Garrote VL, Escobar Liquitay CM, Vietto V. Identification of problems in search strategies in Cochrane Reviews. *Res Synth Methods*. 2018;9(3):408-416. doi:10.1002/jrsm.1302
10. Ahn S, Ames AJ, Myers ND. A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Rev Educ Res*. 2012;82(4):436-476. doi:10.3102/0034654312458162
11. Dechartres A, Atal I, Riveros C, Meerpohl J, Ravaud P. Association Between Publication Characteristics and Treatment Effect Estimates. *Ann Intern Med*. 2018;169(6):385. doi:10.7326/M18-1517
12. Walpole SC. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. *J Clin Epidemiol*. 2019;111:127-134. doi:10.1016/j.jclinepi.2019.03.004
13. Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods*. 2017. doi:10.1002/jrsm.1240
14. 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
15. Stanley TD, Carter EC, Doucouliagos H. What meta-analyses reveal about the replicability of psychological research. *Psychol Bull*. 2018;144(12):1325-1346. doi:10.1037/bul0000169
16. 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(2):239-270. doi:10.1080/00273171003680187
17. 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. doi:10.1016/j.jclinepi.2017.05.002
18. Johnson BT, Low RE, MacDonald H V. Panning for the gold in health research: Incorporating studies' methodological quality in meta-analysis. *Psychol Health*. 2015;30(1):135-152. doi:10.1080/08870446.2014.953533
19. 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(7):e1000100. doi:10.1371/journal.pmed.1000100
20. 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:j4008. doi:10.1136/bmj.j4008
21. 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. doi:10.1016/J.JCLINEPI.2019.06.014
22. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. 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):1-7. doi:10.1186/1471-2288-13-61
23. 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. doi:10.1186/1471-2288-7-10
24. Stroup DF. Meta-analysis of Observational Studies in Epidemiology<SUBTITLE>A Proposal for Reporting</SUBTITLE>. *JAMA*. 2003;283(15):2008. doi:10.1001/jama.283.15.2008
25. Coxeter P, Del Mar CB, McGregor L, Beller EM. Interventions to facilitate shared decision making to address antibiotic use for acute respiratory infections in primary care. *Cochrane database Syst Rev*. 2015;11(11):CD010907. doi:10.1002/14651858.CD010907.pub2.www.cochranelibrary.com
26. 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(6):444-455. doi:10.1016/j.jaac.2016.03.007
27. Cortese S, Moreira-Maia CR, St Fleur D, Morcillo-Peñalver C, Rohde LA, Faraone S V. Association between ADHD and obesity: A systematic review and meta-analysis. *Am J Psychiatry*. 2016;173(1):34-43. doi:10.1176/appi.ajp.2015.15020266
28. Aytug ZG, Rothstein HR, Zhou W, Kern MC. Revealed or concealed? transparency of procedures, decisions, and judgment calls in meta-analyses. *Organ Res Methods*. 2012;15(1):103-133. doi:10.1177/1094428111403495
29. Allers K, Hoffmann F, Mathes T, Pieper D. Systematic reviews with published protocols compared to those without: more effort, older search. *J Clin Epidemiol*. 2018;95:102-110. doi:10.1016/J.JCLINEPI.2017.12.005
30. Yaylali IE, Alaçam T. Critical Assessment of Search Strategies in Systematic Reviews in Endodontics. *J Endod*. 2016;42:854-860. doi:10.1016/j.joen.2016.02.018
31. Hohn RE, Slaney KL, Tafreshi D. Primary study quality in psychological meta-analyses: An empirical assessment of recent practice. *Front Psychol*. 2019;9(JAN). doi:10.3389/fpsyg.2018.02667
32. Ahn S, Becker BJ. Incorporating Quality Scores in Meta-Analysis. *J Educ Behav Stat*. 2011;36(5):555-585. doi:10.3102/1076998610393968
33. 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(AUG):1395. doi:10.3389/fpsyg.2017.01395
34. Hutton B, Wolfe D, Moher D, Shamseer L. 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(2):46-52. doi:10.1136/eb-2017-102666
35. Robson RC, Pham B, 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;106:121-135. doi:10.1016/j.jclinepi.2018.10.003
36. 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(1):116-134. doi:10.1037/met0000197

Table 1. AMSTAR2 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 assessed in individual studies being included in the review (item 9)
- Appropriateness of meta-analytical methods (item 11)
- Consideration of risk of bias 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 (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 MA (item 12)
- Explanation for any heterogeneity observed in the results (item 14)
- Report any potential sources of conflict of interest (item 16)

Table 2. General characteristics of the MAs

Characteristics	Category	n	Number (Percent)	Median (P25-P75)
Article				
Mention of the use of a guideline other than PRISMA	Moose	23	17 (73.9)	
	Mars		2 (8.7)	
	Cochrane		1 (4.35)	
	Quorom		1 (4.35)	
	Strobe		1 (4.35)	
	Center 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)	
Number of studies included in the first MA	1-3	206	11 (5.3)	
	4-9		57 (27.7)	
	≥10		138 (67)	
Assessment of the risk of bias in individual studies	Yes	206	111 (53.9)	
Tool used to assess the risk of bias	RoB tool	95	36 (37.9)	
	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)	
Authors				
Number of authors	1	206	12 (5.8)	
	2-3		60 (29,1)	
	4-6		98 (47.6)	
	≥7		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)	
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)	
Journal				
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)	
Methodological quality				
AMSTAR 2 tool	High quality	206	1 (0.5)	
	Moderate quality		2 (1)	
	Low quality		8 (4)	
	Critically low quality		195 (94.5)	

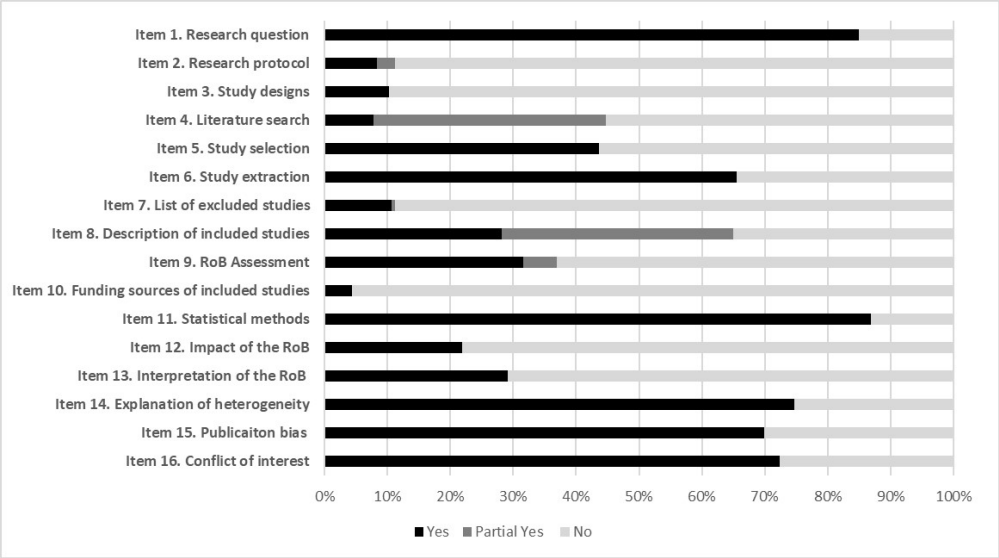


Figure 1. Proportion of adherence to AMSTAR2 items. >> : 7 critical domains identified by AMSTAR2.

98x54mm (300 x 300 DPI)

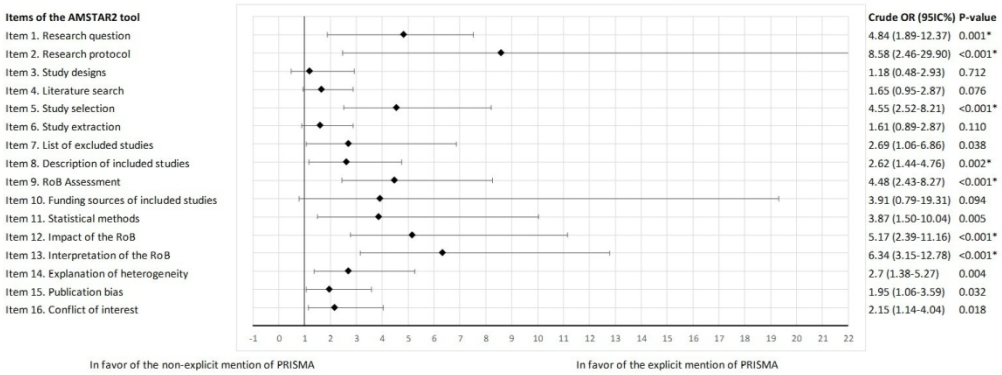


Figure 2. Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group. *Items statistically significant with the Bonferroni correction for multiple testing (p≤0.003).

128x48mm (300 x 300 DPI)

Register OSF

Study information

Title

Assessment of the reporting and methodological qualities and associated factors of a sample of meta-analyses recently indexed in PsycINFO (2016).

Authors

Victoria Leclercq – Charlotte Beaudart – Véronique Rabenda – Sara Ajamieh - Ezio Tirelli – Olivier Bruyère

Background

For scientists, searching for current best evidence has become a real challenge for scientists given the quasi limitless number of published articles (more than 1 270 000 in 2014, according to Thomson Reuteur's Web of Science). When facing a problematic implying a decision, scientists need documents and results oh high and reliable scientific value, as promoted by the evidence-based medicine movement (EBM). EBM is defined as the practice of medicine-based on knowledge and understanding of the literature in order to support clinical decisions (Guyatt et al. 2015). Following evidence hierarchy of EBM, systematic reviews (SRs) and meta-analyses (MAs) are considered the best level of evidence. Nowadays, in diverse disciplines, many researchers base their own research on the results of these SRs and MAs. The Cochrane collaboration adopted the definition of Antman (1992) and Oxman (1993) for the SR: "A systematic review attempts to collate all empirical evidence that pre-specified eligibility criteria in order to answer a specific research question" (Higgins & Green 2011) and the definition of Glass (1976) for MA : "Meta-analysis is the use of statistical methods to summarize the results of independent studies" (Higgins & Green 2011).

Some researchers have highlighted an increase in the publication rate of 2728% for SR (1024 articles in 1991 and 28 959 in 2014) and 2635% for MA (334 articles in 1991 and 9135 in 2014) (Ioannidis 2016). Several reasons could explain this phenomenon. In particular, fewer resources are necessary to perform SRs and MAs, which generally (yield) are worth high citation rates (contributing to increase the impact factor of the journal where they are published (Ioannidis 2016). However, an increasing number of studies have highlighted weaknesses in the design, conduct, analysis, and reporting of MAs published in many scientific fields (Zhu et al. 2016; Cullis et al. 2017; Peters et al. 2015; Zhang et al. 2016; Gagnier & Kellam 2013).

Two tools have been developed to evaluate the quality of the methodology ((AMSTAR, "A Measurement Tool to Assess systematic Reviews" (Shea et al. 2007)) and recently its update (AMSTAR 2 (Shea et al. 2017))) and another one for the quality of the reporting (PRISMA, "Preferred Reporting Items for Systematic Reviews and Meta-analyses" ((Moher et al. 2009)) of SRs and MAs. AMSTAR, a 11-item measuring tool aiming to assess the methodologic quality of MAs (Zhang et al. 2016), has been shown to be reliable and valid (Shea et al. 2009). AMSTAR 2, a 16-item measuring tool aiming to assess the methodologic quality of MAs of randomized and no randomized studies (Shea et al. 2017). PRISMA comprises a list of 27 items that are recommended to be used in the reporting of a MA in order to ensure that the article contains all relevant information (Moher et al. 2009). Several studies have already evaluated the quality of MAs published in specific medical fields such as surgery (Cullis et al. 2017; Zhang et al. 2016), depression (Zhu et al. 2016), orthopaedic (Gagnier & Kellam 2013) or even otorhinolaryngologic disorders (Peters et al. 2015). To the best of our knowledge, there are no such studies available in the field of psychological science.

In line with the EBM movement, the American Psychological Association (APA) has defined in 2006 the movement of Evidence-Based Practice in Psychology with the purpose "to promote effective psychological practice and enhances public health by applying empirically supported principles of psychological assessment, case formation, therapeutic relationship, and intervention"(American Psychological Association 2006). The American Psychological Association has brought out some benefits to the use of Reporting Standards whose the salutary effect on the way research has been conducted (Cooper 2008). The PRISMA statement could also have a positive effect on the methodological quality of the studies.

A growing meta-research literature has assessed the quality of empirical and experimental psychological studies in often large samples of articles (Ioannidis 2012; Bakker & Wicherts 2011; Oliveras et al. 2017; Stanley et al. 2017). It has revealed and quantified numerous methodological deficiencies, such as an inappropriate use of statistics, high rates of statistical mistakes, a frequent lack of statistical power (along with the neglect of effect size considerations) or the unambiguous presence of methodological biases, to mention but a few of them. Interestingly, a recent study conduct by Fanelli and co-workers (Fanelli, Costas, & Ioannidis, 2017) have highlighted differences in the risk of bias (poor estimate of the magnitude of effect size due to, for example, lower inclusion of grey literature, US effect or industry bias…) between the classical disciplines, the risk being highest in the social sciences (to which

psychology belongs). These differences could reflect dissimilar research practices documented in primary studies (e.g. higher publication bias in some disciplines) or distinct procedural choices in meta-analyses (e.g. lower inclusion of grey literature in some disciplines) (Fanelli et al., 2017).

To our knowledge, no studies have however been conducted in order to evaluate the quality of MAs published in the field of psychology. With this research project, our aim is to evaluate the quality of MAs and identify its associated factors published in the psychological or psychology-related field on the PsycINFO database during the year 2016.

Objectives

The objective of this research is to assess the factors associated with the quality of recent MAs indexed in PsycINFO for the year 2016 using two samples of MAs; one composed of MAs claiming to follow the PRISMA statement and the other one including MAs ignoring it.

Our research will be organized in three sub-studies:

1. The assessment of the reporting quality of MAs that claim to use the PRISMA statement compared to those that do not use PRISMA through the PRISMA statement;
2. The assessment of the methodological quality of MAs that claim to use the PRISMA statement compared to those that do not use PRISMA through the AMSTAR 2 tool;
3. The identification of potential factors associated with the quality of MAs.

Research questions

Based on our objectives, our research questions are the following:

1. What is the relationship between of the use of the PRISMA checklist on the reporting quality of MAs?
2. What is the relationship between of the use of the PRISMA checklist on the methodological quality of MAs?
3. What are the potential factors (e.g. publishing journal's impact factor, pre-registration of the study, experience of the first author…) associated with the quality of MAs?

Hypotheses

1. Comprehensive and transparent reporting is necessary to assessing the methodological quality of MAs (Page et al. 2016). They are some articles that highlighted that MAs have a poor reporting quality in the medical literature and the score of PRISMA Statement that were found is between 16,8 and 23/27 points ((Tunis et al. 2013; Fleming et al.

2013; Peters et al. 2015; Zhang et al. 2016; Gagnier & Kellam 2013; Adie et al. 2015; Zhu et al. 2016). We make the assumptions that the use of PRISMA statement improves the reporting quality of MAs using it. This is supported by another study showing that the PRISMA scores is higher by 1 point for the MAs for which authors claimed having used PRISMA statement (Zhu et al. 2016).

2. Meta-research has revealed that the psychological literature present an unsatisfactory level of methodological quality (Bakker & Wicherts 2011; Ioannidis 2012; Oliveras et al. 2017; Stanley et al. 2017). It is therefore possible that this is also true for the MA published in psychology and related fields. In recent studies analyzing the quality of Mas in a number of health-related fields AMSTAR scores have been found to fall between 3,7 and 7,8/11 points (Zhang et al. 2016; Adie et al. 2015; Gagnier & Kellam 2013; Klimo et al. 2014). Since the AMSTAR 2 tool was recently published, it is probable that no study has yet evaluated the quality of the MAs with this tool. Note that it is likely that the use of PRISMA statement exerts a positive influence on the quality of Mas using it. This is supported by a recent study on depression showing that the AMSTAR scores for the MAs for which authors claimed having used the PRISMA statement, reach an average of 0.4 point higher than those which did not use PRISMA (Zhu et al. 2016).
3. They are some potential factors that could correlate with (and possibly influence) the quality of MAs. More specifically, on the basis of previous meta-research studies we make the assumptions that the following factors are positively associated with the measures of quality of MAs: h-index of the first author (Cullis et al. 2017), experience of the principal author in MAs (Zhang et al. 2016), affiliation of the authors to a university (Cullis et al. 2017), publishing journal's impact factor (Cullis et al. 2017), PRISMA endorsement by the journal publishing (Cullis et al. 2017), funding sources described (Gagnier & Kellam 2013), Cochrane collaboration (Adie et al. 2015; Cullis et al. 2017; Zhu et al. 2016), number of pages of the manuscript (Adie et al. 2015; Cullis et al. 2017), pre-registration of the study (Cullis et al. 2017; Zhang et al. 2016; Zhu et al. 2016), non-Asian origine (Zhang et al. 2016) and meta-analyses of randomized controlled trials (Zhang et al. 2016; Zhu et al. 2016). Furthermore, we hypothesize that the following variables will be also associated with the quality of MAs: open access of the publication, open data (or data sharing), the field of psychology, number of individual studies in each MA, number of databases used, assessment of the quality of

individual studies and related tools used, pooling methods used to combine the data, assessment of the publication bias and related method used, assessment of the heterogeneity and related method used, the statistical software used and tendency of the conclusion.

Sampling plan

Existing Data

The tests can be considered as confirmatory given similar studies that have recently been conducted in the medical field.

Explanation of existing data

This is not applicable for our research protocol because no data have been collected so far.

Data collection procedures

Data collection

Data will be collected from a random sample of 200 MAs that will be divided into 2 groups (use of PRISMA vs no use of PRISMA).

Protocol selection of meta-analyses articles

All MAs performed on human subjects and published in English in 2016 in the electronic database PsycINFO will be searched. The electronic search strategy was developed with co-authors and the assistance of a librarian is available in *table 1*.

Table 1: Search strategy

1	meta analysis.md. (15886)
2	meta analysis/ (3940)
3	meta analys*.mp. (24573)
4	data pooling*.mp. (50)
5	2 or 3 or 4 (24599)
6	5 not 1 (10725)
7	1 or 6 (26611)
8	limit 7 to (English and human and yr="2016") (2159)

A total of 2159 potentially relevant MAs were identified in the PsycINFO database. Two investigators will independently review each title and abstract in order to exclude irrelevant articles and to only select the studies that meet inclusion criteria (full inclusion and exclusion criteria are available in Table 2). All discrepancies in opinion regarding the selection of articles

will be resolved through discussion and consensus between the two investigators; any persistent disagreement will be solved with the intervention of a third person (an expert).

Table 2 : eligibility criteria

Inclusion criteria
- Meta-analysis
- Articles published in the PsycINFO-database
- Published between 01.01.2016 to 31.12.2016
- English
Exclusion criteria
- Overview, review
- Meta-synthesis
- Qualitative meta-analysis
- Umbrella review
- Meta-analysis of meta-analyses
- Systematic review without meta-analysis
- Protocol of meta-analysis
- Network meta-analysis
- Activation likelihood Estimation Meta-analysis (ALE MA)
- Signed differential mapping meta-analysis (SMD MA)
- Voxel wise meta-analysis
- Individual patient data meta-analysis (IPD MA)
- Genetic association study (GWAS), genetic study
- Multi-level meta-analysis
- Update
- Letter, comment, abstract, chapter, erratum, dissertation or editorial journal

Choice of language for inclusion was based on expertise within our research team, due to budget constraints, limited time and resources.

A flowchart with the number of included studies will be elaborated. The reason of exclusion of articles will be presented at the step of full-text selection.

Once all MAs will be identified, we will randomly select a minimum of 200 articles as follows. All references of articles will be indexed in an Excel file and randomly assigned to a number. Then, we will rank the articles in ascending order. Two blinded researchers, with the intervention of a third researcher in case of disagreement, will classify MAs that meet inclusion criteria in either the group "with PRISMA" or "without PRISMA" until each group will contain a minimum of 100 MAs. Kappa statistics will be used to test inter-rater agreement.

Data extraction

Relevant data will be extracted from the full texts of all selected articles in a standardized Microsoft Excel spreadsheet by two independent investigators trained for this data extraction. We will record the following factors that might influence the quality of the MAs: characteristics of the manuscripts, characteristics of the study, objective(s) of the study, statistical analyses, characteristics of the protocol and items of PRISMA statement, AMSTAR tool and AMSTAR2 tool. All data extracted will be detailed in appendices. If any disagreements were to be observed between the two reviewers, they will be resolved by discussion, if necessary with the intervention of a third reviewer. Kappa statistics and absolute agreement (%) will be used to assess reproducibility.

Sample size

A first exploratory search in PsycINFO has yielded approximately 2000 articles, which are impossible to analyze for us in a reasonable period of time. We elected to randomly (see below for the method of randomization) select 200 MA articles (until each group, PRISMA and NO PRISMA, will contain a minimum of 100 MAs) from all eligible MAs published in 2016 and indexed in PsycINFO. There is no global MAs offering a synthetic effect size (of the published differences between the two samples) that could have been used to determine a priori a sample size allowing the detecting of a significant difference (power analyses). The chosen sample size can minimally detect a medium effect size (Cohen's d = 0.46065; as computed via G*Power) using a two-tailed Student t-test for independent groups taken at an alpha error probability of 0.05 and a power (1-beta error probability) of 0.90 (critical t = 1.9720). Note that smaller effects sizes cannot be detected (if existing) with such sample (n=100). The meaning and practical significance of the empirically obtained effect size will be discussed.

Sample size rationale

Considering the power analysis described above and constraints in terms of time, financial resources and staff, we will conduct this research on about 200 articles. We think that this will necessitate more than 500 hours of coding for each assessor.

Stopping rule

This is not applicable for our research protocol.

Variables

Manipulated variables

This cannot be applied to the present research protocol.

Measured variables

In order to verify our hypotheses, we will assess not only the quality of reporting and of conduct of MAs but also a set of variables to identify the potential factors that are associated with the quality of MAs.

Assessment of reporting quality

Eligible papers will first be assessed with the PRISMA statement. Each individual item of the PRISMA statement will be answered by "yes" or "no" response, respectively depending on the item being or not fulfilled and will be coded "1" or "0". The total score of the PRISMA statement is the addition of all items coded 1 with a maximum of 27 points.

Assessment of methodology quality

Eligible papers will then be assessed with AMSTAR 2 tool. Each individual item of the AMSTAR tool will be answered by a "yes", "partial yes" or "no" response, respectively depending on the item being or not fulfilled and will be coded "1", "0.5", "0". The total score given by the tool is the addition of all items coded 1 and 0.5 with a maximum of 16 points.

Eligible papers will also be assessed with the AMSTAR tool. Each individual item of the AMSTAR tool will be answered by a "yes" or "no" response, depending on the item being or not fulfilled and will be coded "1" or "0". The total score given by the tool is the addition of all items coded 1 with a maximum of 11 points.

Identification of potential factors associated with the quality of MA

All factors will be assessed by three types of variables: dichotomous variables, quantitative variables and text variables (open questions).

Dichotomous variables

Dichotomous variables will be coded as follows:

1: Yes, it features the characteristic that we seek.

0: No, it does not feature the characteristic that we seek.

99: Not reported, the characteristic that we seek is not available.

88: Not applicable.

The variables that are concerned are the following: author's experience in meta-analysis, affiliation of the authors to a university, PRISMA's recommendation by the journal, restriction of the word count by the journal, declaration of conflict of interest, declaration of funding sources, Cochrane collaboration, open access, open data, registration of the study in a database, evaluation of the quality of study, use of reporting or methodology guideline, the type of study (randomized controlled trials (RCT) or not), evaluation of publication bias, evaluation of heterogeneity, presence of a protocol and the conclusion supports the assumptions.

Quantitative variables

They will be encoded with numerical values and their units of measurements.

The variables that are concerned are the following: h-index (an author-level metric), number of authors, impact factor of the journal (which reflects the frequency with which the average article in a journal has been cited in a particular year) and number of database consulted.

Qualitative variables

The relevant variables are the following: the continent where the study was conducted (Europe, Asia, Africa, America, Oceania), number of study included in each MA (0-3; 4-9; ≥10) and the pooling method used (random effect model, fixed effect model or mixed effect model).

Text variables

The remaining variables are recorded as a text variable. The variables that are concerned are the following: the name of the tool used in order to assess the quality of individual study, the name of the database(s) searched, the PsycINFO classification, the name of the guideline used, the method used to evaluate the publication bias, the method used to assess the heterogeneity and the statistical software. These text variables will then be categorized.

If the data is not available, it will be coded 88.

Indices

This is not applicable for the present research protocol.

Design plan

Study type

This is an observational study. Data are collected from meta-analysis articles.

Blinding

The selection of MAs reviews by title and abstract and by full-text will be done independently by two investigators.

Study design

This is a cross-sectional study.

Randomization

We will randomly MAs articles to get a minimum of 100 MAs in each group. All references of articles (n= probably more than 2000) will be indexed in an Excel file and randomly assigned to a number. Then, we will rank the articles in ascending order. Two blinded researchers, with the intervention of a third researcher in case of disagreement will classify MAs that meet inclusion criteria in either group “with Prisma” and “without Prisma” until each group will contain a minimum of 100 articles MAs.

Analysis plan

Statistical model

The characteristics of all individual studies will first be presented. All quantitative variables that follow a normal distribution will be reported as mean and standard deviation and those that do not follow a normal distribution will be represented as median and percentile (P25 and P75). Distribution will be considered as normal if data meet 3 of the 4 following conditions: the mean is close to the median, the Shapiro-Wilk normality test yields a p-value </=0.05, the curve of the variable follows the normal (or Gaussian) distribution and the linearity of the QQ-Plots is respected. Qualitative and dichotomous variables will be reported as numbers and frequencies. The results of the quality assessment of MAs with the PRISMA statement, AMSTAR tool and AMSTAR 2 will be reported, for quantitative variable, as number and frequency for each item and as mean or as median for the total score. The data will be presented and analyzed using a

star chart. A star chart is a graphical tool that will allow us to represent and compare the percentages of item of PRISMA statement, AMSTAR tool and AMSTAR 2 met by the MAs.

To verify our first and second hypothesis, the reporting and methodological qualities of the individual studies will be compared between the studies that report using the PRISMA checklist and the studies that do not. Comparisons of means between the two groups will be calculated using the Student t-test if for independent groups if the score of PRISMA, AMSTAR and AMSTAR 2 are normally distributed and the Mann-Whitney test if the score of PRISMA, AMSTAR and AMSTAR 2 are not normally distributed. To test the association between the use of the PRISMA statement and the different items of PRISMA, AMSTAR and AMSTAR 2, we will be used a logistic regression.

To test our third hypothesis, factors (all the data detailed in the measured variables part) with potential influence on the quality of studies (mean score of AMSTAR, mean score of AMSTAR 2, independent variable) will be identified with a univariate linear regression. The variables with a p-value lower than 0.1 will be combined in stepwise backward multiple regression analysis. A p-value </=0.05 will be considered as significant.

Transformations

Each item of the PRISMA & AMSTAR checklists will be coded with the following meaning:

1 = Yes or not applicable

0 = No

We will also sum up all items coded 1, with a maximum score of 27 or 11, respectively.

Each item of AMSTAR 2 will be coded with the following meaning:

1 = Yes or not applicable

0.5 = Partial yes

0 = No

We will also sum up all items coded 1 and 0.5, with a maximum score of 16.

See appendix for the details of all transformations for each variable.

Follow-up analyses

All follow-up analyses are described above.

Inference criteria

Not applicable for our research protocol.

Data exclusion

No data will be excluded from our database.

Missing data

Missing data may have an impact on the analysis and on the interpretation of the results. Some of the extracted data may not be available (h-index, impact factor…). After data encoding, a quality control will be done, at database-level, in order to check for outliers, coding error and missing values. In case of incomplete information, we will contact the authors.

Exploratory analysis

The exploratory analyses will be considered, based on the results obtained.

If there is a statistically significant difference in quality between MAs which report using PRISMA and those which do not, we will consider carrying out the following analyses. A logistic regression will be carried out in order to describe the relation between the dichotomous dependent variables (PRISMA vs No PRISMA) and all potential explanatory variables (all the data are detailed in the measured variables part). The variables with a p-value lower than 0.1 will be combined in stepwise backward multiple logistic regression analysis. A p-value <0.05 will be considered significant on statistical analyses.

Scripts

Upload an analysis script with clear comments

Not available at the moment.

Other

We would like to acknowledge Pr. Anne-Françoise Donneau for interesting discussions about some aspects on the planned statistical analyses and Ms Nancy Durieux for her assistance in the building of our strategy, in terms of electronic literature search.

Bibliography

Adie, S. et al., 2015. Quality of Conduct and Reporting of Meta-analyses of Surgical Interventions. *Annals of Surgery*, 261(4), pp.685–694. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000658-201504000-00012 [Accessed June 14, 2017].

American Psychological Association, 2006. Evidence-Based Practice in Psychology APA Presidential Task Force on Evidence-Based Practice. *American Psychologist*, 61(4), pp.271–285. Available at: http://www.apa.org/pubs/journals/features/evidence-based-

statement.pdf [Accessed January 10, 2018].

Bakker, M. & Wicherts, J.M., 2011. The (mis)reporting of statistical results in psychology journals. Available at: https://link.springer.com/content/pdf/10.3758%2Fs13428-011-0089-5.pdf [Accessed July 31, 2017].

Cooper, H., 2008. Reporting Standards for Research in Psychology: Why Do We Need Them? What Might They Be? *American Psychologist*, 63(9), pp.839–851. Available at: https://www.apa.org/pubs/authors/jars.pdf [Accessed January 15, 2018].

Cullis, P.S., Gudlaugsdottir, K. & Andrews, J., 2017. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. *PLoS ONE*, 12(4). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5383307/pdf/pone.0175213.pdf [Accessed June 14, 2017].

Fanelli, D., Costas, R. & Ioannidis, J.P.A., Meta-assessment of bias in science. Available at: http://www.pnas.org/content/114/14/3714.full.pdf [Accessed December 18, 2017].

Fleming, P.S. et al., 2013. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. *The Angle Orthodontist*, 83(1), pp.158–163. Available at: http://www.angle.org/doi/10.2319/032612-251.1 [Accessed May 24, 2018].

Gagnier, J.J. & Kellam, P.J., 2013. Reporting and Methodological Quality of Systematic Reviews in the Orthopaedic Literature. *The Journal of Bone and Joint Surgery-American Volume*, 95(11), pp.e77-1–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23780547 [Accessed June 14, 2017].

Guyatt, G. et al., 2015. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Medicine*, Available at: https://jamaevidence.mhmedical.com/book.aspx?bookID=847 [Accessed September 1, 2017].

Higgins, J. & Green, S., 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011].* The Cochra., Available at: http://handbook.cochrane.org.

Ioannidis, J.P.A., 2016. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, 94(3), pp.485–514. Available at: http://doi.wiley.com/10.1111/1468-0009.12210.

Ioannidis, J.P.A., 2012. Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), pp.645–654. Available at: http://pps.sagepub.com/ [Accessed

July 31, 2017].

Klimo, P. et al., 2014. Methodology and reporting of meta-analyses in the neurosurgical literature. *Journal of neurosurgery*, 120(4), pp.796–810. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24460488.

Moher, D. et al., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). *Physical Therapy*, 89(9), pp.873–880. Available at: http://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1000097&type=printable [Accessed August 25, 2017].

Oliveras, I., Losilla, J.-M. & Vives, J., 2017. Methodological quality is underrated in systematic reviews and meta-analyses in health psychology. *Journal of Clinical Epidemiology*, 86, pp.59–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28499846 [Accessed November 28, 2017].

Page, M.J. et al., 2016. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study N. Low, ed. *PLOS Medicine*, 13(5), p.e1002028. Available at: http://dx.plos.org/10.1371/journal.pmed.1002028 [Accessed May 24, 2018].

Peters, J.P.M. et al., 2015. Reporting quality of systematic reviews and meta-analyses of otorhinolaryngologic articles based on the PRISMA statement. *PLoS ONE*, 10(8). Available at: http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0136540&type=printable [Accessed June 14, 2017].

Shea, B.J. et al., 2017. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, p.j4008. Available at: http://www.bmj.com/content/bmj/358/bmj.j4008.full.pdf [Accessed November 28, 2017].

Shea, B.J. et al., 2009. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), pp.1013–1020. Available at: http://ac.els-cdn.com/S0895435608003259/1-s2.0-S0895435608003259-main.pdf?_tid=7ad1b7aa-558b-11e7-a3dc-00000aacb361&acdnat=1497944501_0f82a189082ff8754f64a91d9670402e [Accessed June 20, 2017].

Shea, B.J. et al., 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*, 7, p.10. Available at: https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-7-10?site=bmcmedresmethodol.biomedcentral.com [Accessed August 25, 2017].

Stanley, T.D., Carter, E.C. & Doucouliagos, H., 2017. What Meta-Analyses Reveal about the Replicability of Psychological Research. Available at: http://www.deakin.edu.au/__data/assets/pdf_file/0007/1198456/WhatMeta-AnalysesReveal_WP.pdf [Accessed January 10, 2018].

Tunis, A.S. et al., 2013. Association of Study Quality with Completeness of Reporting: Have Completeness of Reporting and Quality of Systematic Reviews and Meta-Analyses in Major Radiology Journals Changed Since Publication of the PRISMA Statement? *Radiology*, 269(2), pp.413–426. Available at: http://pubs.rsna.org/doi/10.1148/radiol.13130273 [Accessed May 24, 2018].

Zhang, H. et al., 2016. Reporting and methodological qualities of published surgical meta-analyses. Available at: http://ac.els-cdn.com/S0895435615003121/1-s2.0-S0895435615003121-main.pdf?_tid=4e369bf4-50e8-11e7-9b21-00000aab0f26&acdnat=1497434613_6d09bcd5abe078df3344e91ace4b09d1 [Accessed June 14, 2017].

Zhu, Y. et al., 2016. Is the best evidence good enough: Quality assessment and factor analysis of meta-analyses on depression. *PLoS ONE*, 11(6). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4919061/pdf/pone.0157808.pdf [Accessed June 14, 2017].

Appendix
Explication of the data extraction form

Name	Explication	Description
Name of the reviewer	Name of the reviewer	Text
Study ID	Reference number of the article	Text
Inclusion of the article	Inclusion of the article based on the selection criteria	1 = Yes 0 = No
If excluded, indicate the reason of exclusion	Reason of exclusion	The reason of exclusion 88 = if not applicable
Use of PRISMA	The authors declared the use of PRISMA	1= Yes 0 = No

1. Characteristics of the manuscript

Name	Explication	Description
DOI of the article	Unique identifier of the article	Text
Year of publication	Publishing year of the manuscript	Text
Author's name	Name of the first author	Text
Author's h-index	H-Index of the first author (Scopus)	Quantitative variable
Author's experience	Number of meta-analyses from the same author(s) (Scopus)	Quantitative variable
Affiliation of the authors to a university	Affiliation of the authors to a university	1 = Yes 0 = No 99 = Not reported
Number of authors	Total number of authors	Quantitative variable
Contribution of authors	Details of the authors' contribution	1 = Yes 0 = No 99 = Not reported
Journal's name	Name of the journal	Text
Journal's Impact factor	The IF of the journal using the ISI Journal Citation Reports 2016 (http://isiknowledge.com)	Quantitative variable
Instruction for authors: PRISMA required?	The journal recommended to use PRISMA statement	1 = Yes 0 = No 99 = Not reported

Instruction for authors: page or word limitation?	Limitation of the number of pages or words	1 = Yes 0 = No 99 = Not reported
PsycINFO classification	Classification of the field of psychology based on the PsycINFO Content Classification Code System	2100 = General Psychology 2200 = Psychometrics & Statistics & Methodology 2300 = Human experimental Psychology 2400 = Animal Experimental & comparative Psychology 2500 = Physiological Psychology & Neuroscience 2600 = Psychology & The Humanities 2700 = Communication Systems 2800 = Developmental Psychology 2900 = Social Processes & Social Issues 3000 = Social Psychology 3100 = Personality Psychology 3200 = Psychological & Physical disorders 3300 = Health & Mental Health Treatment & Prevention 3400 = Professional Psychological & Health Personnel Issues 3500 = Educational Psychology 3600 = Industrial & Organizational Psychology 3700 = Sport Psychology & Leisure 3800 = Military Psychology 3900 = Consumer Psychology 4000 = Engineering & Experimental Psychology 4100 = Intelligent Systems 4200 = Forensic Psychology & Legal Issues
Corresponding author (email address)	Email of the author	Text

Conflict of interest described	Conflict of interest is described	1 = Yes 0 = No 99 = Not reported
Details of conflict of interest	If yes, brief description of conflict of interest	Text
Funding sources described	Funding sources are described	1 = Yes 0 = No 99 = Not reported
Funding sources	If yes, brief description of funding sources	Text
Cochrane collaboration	The study is a Cochrane collaboration	1 = Yes 0 = No
Number of page of manuscript	Total number of pages of the manuscript	Quantitative variable
Open access	The publication is open access?	1 = Yes 0 = No 99 = Not reported
Open data	The data is open access?	1 = Yes 0 = No 99 = Not reported

2. Characteristics of the study

Name	Explication	Description
Registration of the study?	The study was recorded in a specific database.	1 = Yes 0 = No 99 = Not reported
Number of the registration	Registration number of the study	Text 88 = if not applicable
Name of the registry	Name of the registry in which the meta-analysis has been registered	Text 88 = if not applicable
Date of submission of the manuscript	Date of submission of the manuscript	Text (month-year)
Date of publication of the manuscript	Publication date of the manuscript	Text (month-year)
Continent of origin of first author	Continent in which the study has been conducted	Europe – Asia – Africa – America - Oceania
Type of the individual study	Study design of the studies included in the MA	Observational study RCT All types Not specified
Number of databases searched	Number of databases consulted	Quantitative variable

Name of the database	Name of the database searched	Text
Quality of individual study is assessed?	Quality of individual study is assessed	1 = Yes 0 = No
Name of the tool used to assess the quality	Name of the tool used to assess the quality of individual studies	Text variable 88 = If not applicable
Reference to use of the guideline	Reference to use of a guideline	1 = Yes 0 = No
Name of the guideline used	Name of the guideline used (PRISMA, MOOSE, AMSTAR…)	Text 88 = If not applicable
Search strategy	Presence of the complete search strategy	1 = Yes 0 = No 99 = Not reported
Focus of review	Type of the field of psychology	Text

3. Objective of the study

Name	Explication	Description
Main objective	Aim of the study	Text
Primary outcome	Primary outcome of the study disclosed	Text
Secondary outcomes	Secondary outcome of the study disclosed	Text

4. Statistical analyses

Name	Explication	Description
Number of meta-analyses performed	Number of meta-analyses performed in the presented study	Quantitative variable
Number of studies included in each meta-analysis	Number of studies included in each meta-analysis performed in the study	0-3; 4-9; ≥10
Pooling methods	The pooling methods used to combine data	Fixed – Random - Mix
Assessment of the publication bias	The publication bias is evaluated	1 = Yes 0 = No
Method used to assess the publication bias	Method used to assess the publication bias	Text 88 = If not applicable

Assessment of the heterogeneity	The heterogeneity is evaluated	1 = Yes 0 = No
Heterogeneity	Method used to assess the heterogeneity	Text 88 = If not applicable

5. Protocol

Name	Explication	Description
Protocol	The protocol of the study is existent and available	1 = Yes 0 = No
Primary outcome	Primary outcome of the study	Text 88 = If not applicable
Secondary outcome	Secondary outcome of the study	Text 88 = If not applicable

6. Conclusion

Name	Explication	Description
Conclusion	Main conclusion of the study	Text
Trends of the conclusion	The conclusion supports the assumptions	1 = Yes 0 = No

7. PRISMA statement

Name	Explication	Description
P1	TITLE Title	1 = Yes 0 = No
P2	ABSTRACT Structured summary	1 = Yes 0 = No
P3	INTRODUCTION Rationale	1 = Yes 0 = No
P4	Objective	1 = Yes 0 = No
P5	METHODS Protocol and registration	1 = Yes 0 = No
P6	Eligibility criteria	1 = Yes 0 = No
P7	Information sources	1 = Yes 0 = No

P8	Search	1 = Yes 0 = No
P9	Study selection	1 = Yes 0 = No
P10	Data collection process	1 = Yes 0 = No
P11	Data items	1 = Yes 0 = No
P12	Risk of bias in individual studies	1 = Yes 0 = No
P13	Summary measures	1 = Yes 0 = No
P14	Synthesis of results / Planned methods of analysis	1 = Yes 0 = No
P15	Risk of bias across studies	1 = Yes 0 = No
P16	Additional analysis	1 = Yes 1 = Not applicable 0 = No
P17	RESULTS Study selection	1 = Yes 0 = No
P18	Study characteristics	1 = Yes 0 = No
P19	Risk of bias within studies	1 = Yes 0 = No
P20	Results of individual studies	1 = Yes 0 = No
P21	Synthesis of results	1 = Yes 0 = No
P22	Risk of bias across studies	1 = Yes 0 = No
P23	Additional analysis	1 = Yes 1 = Not applicable 0 = No
P24	DISCUSSION Summary of evidence	1 = Yes 0 = No
P25	Limitations	1 = Yes 0 = No
P26	Conclusions	1 = Yes 0 = No
P27	Funding	1 = Yes 0 = No
	Total score	Sum of all items coded "1"

8. Amstar tool

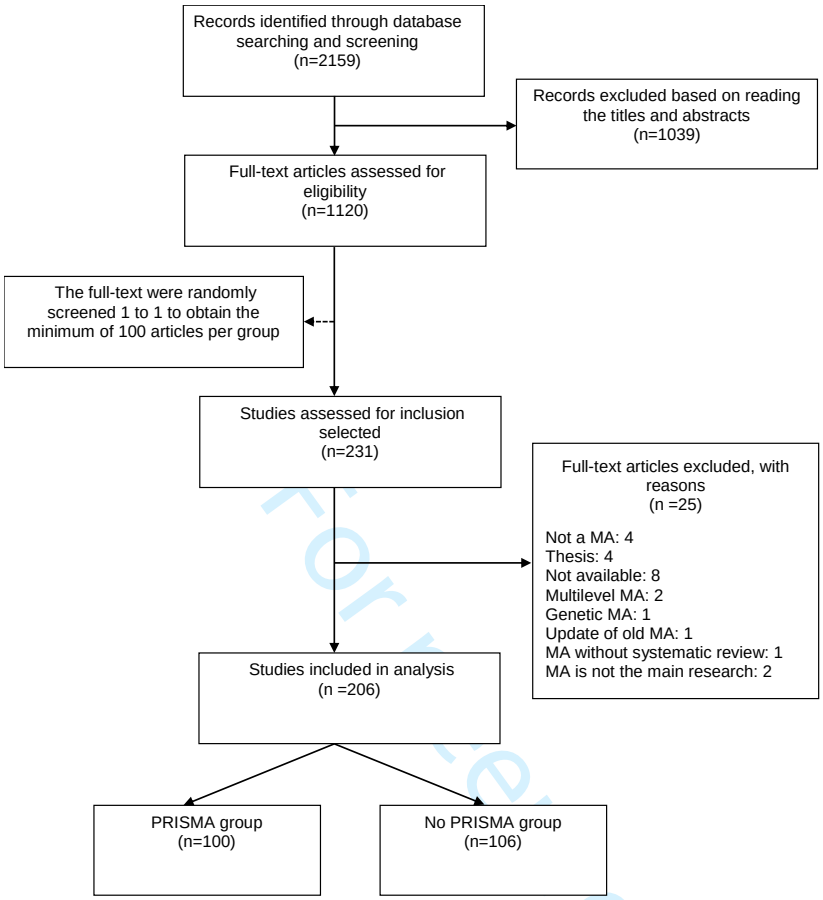
Name	Explication	Description
A1	Was an 'a priori' design provided?	1 = Yes 0 = No
A2	Was there duplicate study selection and data extraction?	1 = Yes 0 = No
A3	Was a comprehensive literature search performed?	1 = Yes 0 = No
A4	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	1 = Yes 0 = No
A5	Was a list of studies (included and excluded) provided?	1 = Yes 0 = No
A6	Were the characteristics of the included studies provided?	1 = Yes 0 = No
A7	Was the scientific quality of the included studies assessed and documented?	1 = Yes 0 = No
A8	Was the scientific quality of the included studies used appropriately in formulating conclusions?	1 = Yes 0 = No
A9	Were the methods used to combine the findings of studies appropriate?	1 = Yes 0 = No
A10	Was the likelihood of publication bias assessed?	1 = Yes 0 = No
A11	Was the conflict of interest included?	1 = Yes 0 = No
	Total score	Sum of all items coded "1"

9. Amstar 2 tool

Name	Explication	Description
AM1	Did the research questions and inclusion criteria for the review include the components of PICO?	1 = Yes 0 = No

AM2	Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	1 = Yes 0.5 = Partial Yes 0 = No
AM3	Did the review authors explain their selection of the study designs for inclusion in the review?	1 = Yes 0 = No
AM4	Did the review authors use a comprehensive literature search strategy?	1 = Yes 0.5 = Partial Yes 0 = No
AM5	Did the review authors perform study selection in duplicate?	1 = Yes 0 = No
AM6	Did the review authors perform data extraction in duplicate?	1 = Yes 0 = No
AM7	Did the review authors provide a list of excluded studies and justify the exclusions?	1 = Yes 0.5 = Partial Yes 0 = No
AM8	Did the review authors describe the included studies in adequate detail?	1 = Yes 0.5 = Partial Yes 0 = No
AM9	Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	1 = Yes 0.5 = Partial Yes 0 = No
AM10	Did the review authors report on the sources of funding for the studies included in the review?	1 = Yes 0 = No
AM11	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	1 = Yes 0 = No
AM12	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	1 = Yes 0 = No
AM13	Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	1 = Yes 0 = No
AM14	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	1 = Yes 0 = No

AM15	If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	1 = Yes 0 = No
AM16	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	1 = Yes 0 = No
	Total score	Sum of all items coded "1" or "0.5"



Supplementary file 2: Flowchart illustrating the MAs Selection

Supplementary file 3

Table 1. Sensitivity analyses : Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group using univariate logistic regression with the three responses modality of AMSTAR2 (Yes, partial yes and no).

AMSTAR2's items		**OR**	**95%CI**	**p-value**
Item 1. Research question	Yes	4.84	1.90-12.37	0.001*
Item 2. Research protocol	Yes	6.01	1.67-21.62	0.006
	Partial yes	NE	NE	0.933
Item 3. Study designs	Yes	1.19	0.48-2.93	0.711
Item 4. Literature search	Yes	1.71	0.59-4.90	0.321
	Partial yes	1.64	0.91-2.94	0.098
Item 5. Study selection	Yes	4.55	2.52-8.21	<0.0001*
Item 6. Study extraction	Yes	1.61	0.90-2.87	0.110
Item 7. List of excluded studies	Yes	2.53	0.98-6.49	0.054
	Partial yes	NE	NE	0.958
Item 8. Description of included studies	Yes	4.44	2.12-9.33	<0.0001*
	Partial yes	1.80	0.93-3.50	0.083
Item 9. RoB Assessment	Yes	4.42	2.32-8.42	<0.0001*
	Partial yes	4.87	1.23-19.25	0.024
Item 10. Funding sources of included studies	Yes	3.91	0.79-19.31	0.094
Item 11. Statistical methods	Yes	3.87	1.49-10.04	0.005
Item 12. Impact of the RoB	Yes	5.17	2.39-11.16	<0.0001*
Item 13. Interpretation of the RoB	Yes	6.34	3.15-12.78	<0.0001*
Item 14. Explanation of heterogeneity	Yes	2.70	1.38-5.27	0.004
Item 15. Publication bias	Yes	1.95	1.06-3.59	0.032
Item 16. Conflict of interest	Yes	2.15	1.14-4.04	0.018

Reference group for all items is "No"; NE : Not estimable, Calculation of the OR is impossible because of the data distribution. *Items statistically significant with the Bonferroni correction for multiple testing ($p \leq 0.003$).

BMJ Open

The methodological quality of meta-analyses indexed in PsycINFO: leads for enhancements – a meta-epidemiological study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-036349.R3
Article Type:	Original research
Date Submitted by the Author:	06-May-2020
Complete List of Authors:	Leclercq, Victoria; University of Liege, Division of Public Health, Epidemiology and Health Economics Beaudart, Charlotte; University of Liege, Division of Public Health, Epidemiology and Health Economics Ajamieh, Sara; University of Liege, Division of Public Health, Epidemiology and Health Economics Tirelli, Ezio; University of Liege, Department of Psychology Bruyère, Olivier ; University of Liege, Division of Public Health, Epidemiology and Health Economics
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Research methods, Evidence based practice
Keywords:	EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

THE METHODOLOGICAL QUALITY OF META-ANALYSES INDEXED IN PSYCINFO: LEADS FOR ENHANCEMENTS – a meta-epidemiological study

Short title (Max 70 characters)

The methodological quality of meta-analyses indexed in PsycINFO

Leclercq V[1], Beaudart C[1], Ajamieh S[1,2], Tirelli E[2], Bruyère O[1]

[1]Division of Public Health, Epidemiology and Health Economics, University of Liège, Belgium

[2]Department of Psychology, University of Liège, Belgium

Corresponding author: Victoria Leclercq; victoria.leclercq@uliege.be

ABSTRACT

Objectives Meta-analyses (MAs) are often used because they are lauded to provide robust evidence that synthesizes information from multiple studies. However, the validity of MA conclusions relies upon the procedural rigor applied by the authors. Therefore, this meta-research study aims to characterize 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 methodologic characteristics of all MAs and checked their quality according to the 16 items of the AMSTAR2 (A MeaSurement Tool to Assess systematic Reviews) tool for MA critical appraisal. Moreover, we investigated the effect of mentioning 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.

Research protocol available on the Open Science Framework: https://osf.io/hjybx/

Strengths and limitations of this study

- Some studies have highlighted methodological weaknesses in the conduct of systematic reviews (SRs) and meta-analysis (MAs) and we search to have an overview of methodological practice of MAs indexed in PsycINFO according to the tool AMSTAR2 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 PRISMA statement on the methodological quality of MAs;
- A sample of 206 Mas indexed in PsycINFO in 2016 and published in English was analyzed;
- Our findings cannot be generalized to MAs published in other years than 2016, in other languages than English or in other databases than PsycINFO.

1 INTRODUCTION

Since the definition of meta-analyses (MAs) being introduced by Glass in 1976, MAs conducted in behavioral and social sciences have increased rapidly in number. There were more than 30 000 MAs indexed in 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[1].

However, the validity of MA conclusions relies upon the rigor 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 judgment calls that need to be made by the meta-analysist. Moreover, too many systematic reviews (SRs) and MAs are of low quality[1–5], 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[6–8], an inappropriate literature search[9–12], flaws in the statistical analyses[10,13–16] and an insufficient assessment of the risk of bias of individual studies[10,17,18].

To support researchers in the realization and reporting of MAs, two tools are commonly used. The first is PRISMA ("Preferred Reporting Items for Systematic Reviews and Meta-Analyses"), which was developed in 2009 by Liberati et al[19]. 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[20], 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 behavioral 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 characterize 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 PICO (item 1) if the MAs authors mention the PRISMA statement. This hypothesis, was tested for each of the 16 AMSTAR2 items.

2 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 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[21]. 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 specialized in the field of behavioral 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 systematic review with a MA, indexed in the PsycINFO database, published between January 01, 2016, and December 31, 2016, and published in English. In total, 2159 records were identified. Two authors (V.L & C.B) 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 flowchart in 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 & SA) independently extracted all relevant data from the full texts of all selected articles in a standardized 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 interquartile range of 0.66 [0.40-0.75]) and the Gwet's AC1 (median value with interquartile range of 0.77 [0.69-0.88]) both suggesting a substantial agreement[22], 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 "A MeaSurement Tool to Assess systematic Reviews 2"[20]. AMSTAR2 was a revision of the original AMSTAR instrument[23] developed by Shea et al in 2007, 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 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 2 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 gray 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 4 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 7 critical weaknesses that should reduce confidence in the findings of a review and 9 other items that are considered noncritical weaknesses, as presented in Table 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 hypothesized 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 (JCR) 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 summarized 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 3 of the 4 following conditions: the mean was close to the median, the Shapiro-Wilk normality test yielded a P-value ≥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 dichotomizing the AMSTAR2 items was performed as a sensitivity analysis. Associations were quantified using odds ratios with 95% confidence intervals. A Bonferroni correction was used to adjust the results for multiple testing (16 tests, p-value<0.003). All analyses were performed using SAS 9.4 software.

Patient and public involvement

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

3 RESULTS

Search results

A total of 2159 potentially relevant MAs related to behavioral 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 2. 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 [24] (17, 74%), Mars (2, 9%) and Quorom (1, 4%). Finally, most articles were not available for open access (90.3%), and only one was a Cochrane MA.

Written by one 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 categorized as critically low quality, 8 as low quality, 2 as moderate quality and 1 as high quality. Only one MA [25] provided all the information on all 7 critical domains assessed and was considered high quality according to AMSTAR2. Two additional MAs [26,27] also provided all information on all 7 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 the classification advised by the AMSTAR2 tool (Table 1).

In Figure 1, we summarize 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: 1.90-12.37) 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: 2.46-29.90);
- study selection in duplicate (item 5; OR: 4.55; 95%CI: 2.52-8.21);
- a detailed description of the included studies (item 8; OR: 2.62; 95%CI: 1.44-4.76);

- a satisfactory technique for assessing the risk of bias in individual studies (item 9; OR: 4.48; 95%CI: 2.43-8.27);
- an assessment of the potential impact of risk of bias in individual studies (item 12; OR: 5.17; 95%CI: 2.39-11.16);
- appropriate consideration of the risk of bias in primary studies when interpreting the results (item 13; OR: 6.34; 95%CI: 3.15-12.78).

The results of the sensitivity analysis, performed without dichotomizing the responses modality of AMSTAR2 (yes, partial yes and no) using a logistic regression, showed similar results (Table 1 in 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.

4 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[28]. 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 characterize the methodological quality of MAs indexed in PsycINFO 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 behavioral and social sciences.

- First, no more than 11% of MAs had a research protocol available. However, several scientists[8,16,29] 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, minimizing the risk of bias, avoiding replicate studies and enhancing transparency[7]. 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 a systematic review 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 2 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 gray 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[10]. The search strategy is an essential step of the MA process since the comprehensiveness and completeness of the

search[3,30] is dependent on this strategy. Furthermore, other scientists have highlighted the need to improve research strategies for more comprehensive MAs[9,30].

- 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[17,31], 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 and her team 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[17]. However, there is still a lack of guidance to incorporate these risk of bias assessments into meta-analyses[17,18,32].

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 recognized 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[21,28,31,33,34].

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[25] 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[26,27] 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[20], yet only a few studies from our sample of MAs mention it explicitly. Dechartres and her team stressed the association between publication characteristics and effect estimates[11] 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[11,12,35]. 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[10] but contradict previous studies that have highlighted several flaws in the application and interpretation of statistical analyses in MA[13,14,28,36]. Page et al identified some mistakes in the use of adequate statistical models, the sufficient exploration of subgroup analyses and sensitivity analyses[14]. 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[20]. 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 generalize 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 AC1 was presented along with the Cohen's Kappa. Although Cohen's Kappa is more widely used, Gwet's AC1 is a more robust alternative (less sensitive to data distribution and number of observation)[22]. Moreover, using AMSTAR2, we can investigate the methodological characteristics used to conduct the study (e.g. 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 (e.g. 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.

5 CONCLUSION

This research contributes to raising awareness among researchers about flaws in MAs published in behavioral and social sciences fields, which hopefully increases the adoption of more rigorous research practices. It is clear that meta-analytic 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.

KEYWORDS

Meta-analyses – AMSTAR – meta-research – methodological quality – PRISMA

FOOTNOTES

Conflict of interest: The authors declare no conflict of interest.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors' contribution: VL, CB, ET and OB conceived the study; VL, CB and SA participated in data collection; VL, CB and OB analyzed and interpreted data; VL, CB, ET and OB corrected the manuscripts. All co-authors read and approved the final version of the manuscript.

Data sharing statement: Data are available in a public, open access repository: https://osf.io/hjybx/.

Legends

Figure 1. Proportion of adherence to AMSTAR2 items. ► : 7 critical domains identified by AMSTAR2.

Figure 2. Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group. *Items statistically significant with the Bonferroni correction for multiple testing ($p \leq 0.003$).

BIBLIOGRAPHY

1. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nat Publ Gr*. 2018;555. doi:10.1038/nature25753
2. 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
3. Dieckmann NF, Malle BF, Bodner TE. An Empirical Assessment of Meta-Analytic Practice. 2009. doi:10.1037/a0015107
4. Zhang H, Han J, Zhu Y-B, et al. Reporting and methodological qualities of published surgical meta-analyses. 2016. doi:10.1016/j.jclinepi.2015.06.009
5. Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q*. 2016;94(3):485-514. doi:10.1111/1468-0009.12210
6. Nuijten MB. Practical tools and strategies for researchers to increase replicability. *Dev Med Child Neurol*. October 2018. doi:10.1111/dmcn.14054
7. Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Syst Rev*. 2018;7(1). doi:10.1186/s13643-018-0699-4
8. 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;100:103-110. doi:10.1016/J.JCLINEPI.2018.01.003
9. Franco JVA, Garrote VL, Escobar Liquitay CM, Vietto V. Identification of problems in search strategies in Cochrane Reviews. *Res Synth Methods*. 2018;9(3):408-416. doi:10.1002/jrsm.1302
10. Ahn S, Ames AJ, Myers ND. A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Rev Educ Res*. 2012;82(4):436-476. doi:10.3102/0034654312458162
11. Dechartres A, Atal I, Riveros C, Meerpohl J, Ravaud P. Association Between Publication Characteristics and Treatment Effect Estimates. *Ann Intern Med*. 2018;169(6):385. doi:10.7326/M18-1517
12. Walpole SC. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. *J Clin Epidemiol*. 2019;111:127-134. doi:10.1016/j.jclinepi.2019.03.004
13. Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods*. 2017. doi:10.1002/jrsm.1240
14. 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
15. Stanley TD, Carter EC, Doucouliagos H. What meta-analyses reveal about the replicability of psychological research. *Psychol Bull*. 2018;144(12):1325-1346. doi:10.1037/bul0000169
16. 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(2):239-270. doi:10.1080/00273171003680187
17. 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. doi:10.1016/j.jclinepi.2017.05.002
18. Johnson BT, Low RE, MacDonald H V. Panning for the gold in health research: Incorporating studies' methodological quality in meta-analysis. *Psychol Health*. 2015;30(1):135-152. doi:10.1080/08870446.2014.953533
19. 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(7):e1000100. doi:10.1371/journal.pmed.1000100
20. 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:j4008. doi:10.1136/bmj.j4008
21. 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. doi:10.1016/J.JCLINEPI.2019.06.014
22. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. 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):1-7. doi:10.1186/1471-2288-13-61
23. 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. doi:10.1186/1471-2288-7-10
24. Stroup DF. Meta-analysis of Observational Studies in Epidemiology<SUBTITLE>A Proposal for Reporting</SUBTITLE>. *JAMA*. 2003;283(15):2008. doi:10.1001/jama.283.15.2008
25. Coxeter P, Del Mar CB, McGregor L, Beller EM. Interventions to facilitate shared decision making to address antibiotic use for acute respiratory infections in primary care. *Cochrane database Syst Rev*. 2015;11(11):CD010907. doi:10.1002/14651858.CD010907.pub2.www.cochranelibrary.com
26. 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(6):444-455. doi:10.1016/j.jaac.2016.03.007
27. Cortese S, Moreira-Maia CR, St Fleur D, Morcillo-Peñalver C, Rohde LA, Faraone S V. Association between ADHD and obesity: A systematic review and meta-analysis. *Am J Psychiatry*. 2016;173(1):34-43. doi:10.1176/appi.ajp.2015.15020266
28. Aytug ZG, Rothstein HR, Zhou W, Kern MC. Revealed or concealed? transparency of procedures, decisions, and judgment calls in meta-analyses. *Organ Res Methods*. 2012;15(1):103-133. doi:10.1177/1094428111403495
29. Allers K, Hoffmann F, Mathes T, Pieper D. Systematic reviews with published protocols compared to those without: more effort, older search. *J Clin Epidemiol*. 2018;95:102-110. doi:10.1016/J.JCLINEPI.2017.12.005
30. Yaylali IE, Alaçam T. Critical Assessment of Search Strategies in Systematic Reviews in Endodontics. *J Endod*. 2016;42:854-860. doi:10.1016/j.joen.2016.02.018
31. Hohn RE, Slaney KL, Tafreshi D. Primary study quality in psychological meta-analyses: An empirical assessment of recent practice. *Front Psychol*. 2019;9(JAN). doi:10.3389/fpsyg.2018.02667
32. Ahn S, Becker BJ. Incorporating Quality Scores in Meta-Analysis. *J Educ Behav Stat*. 2011;36(5):555-585. doi:10.3102/1076998610393968
33. 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(AUG):1395. doi:10.3389/fpsyg.2017.01395
34. Hutton B, Wolfe D, Moher D, Shamseer L. 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(2):46-52. doi:10.1136/eb-2017-102666
35. Robson RC, Pham B, 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;106:121-135. doi:10.1016/j.jclinepi.2018.10.003
36. 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(1):116-134. doi:10.1037/met0000197

Table 1. AMSTAR2 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 assessed in individual studies being included in the review (item 9)
- Appropriateness of meta-analytical methods (item 11)
- Consideration of risk of bias 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 (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 MA (item 12)
- Explanation for any heterogeneity observed in the results (item 14)
- Report any potential sources of conflict of interest (item 16)

Table 2. General characteristics of the MAs

	Characteristics	Category	n	Number (Percent)	Median (P25-P75)
Article					
	Mention of the use of a guideline other than PRISMA	Moose	23	17 (73.9)	
		Mars		2 (8.7)	
		Cochrane		1 (4.35)	
		Quorom		1 (4.35)	
		Strobe		1 (4.35)	
		Center 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)	
	Number of studies included in the first MA	1-3	206	11 (5.3)	
		4-9		57 (27.7)	
		≥10		138 (67)	
	Assessment of the risk of bias in individual studies	Yes	206	111 (53.9)	
	Tool used to assess the risk of bias	RoB tool	95	36 (37.9)	
		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)	
Authors					
	Number of authors	1	206	12 (5.8)	
		2-3		60 (29,1)	
		4-6		98 (47.6)	
		≥7		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)	
	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)	
Journal					
	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)	
Methodological quality					
	AMSTAR 2 tool	High quality	206	1 (0.5)	
		Moderate quality		2 (1)	
		Low quality		8 (4)	
		Critically low quality		195 (94.5)	

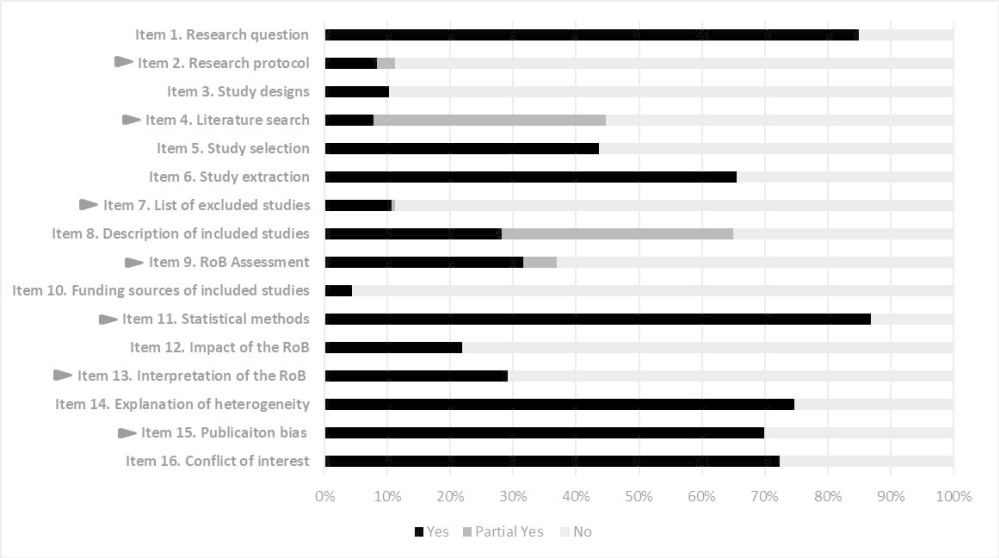


Figure 1. Proportion of adherence to AMSTAR2 items. >> : 7 critical domains identified by AMSTAR2.

98x54mm (300 x 300 DPI)

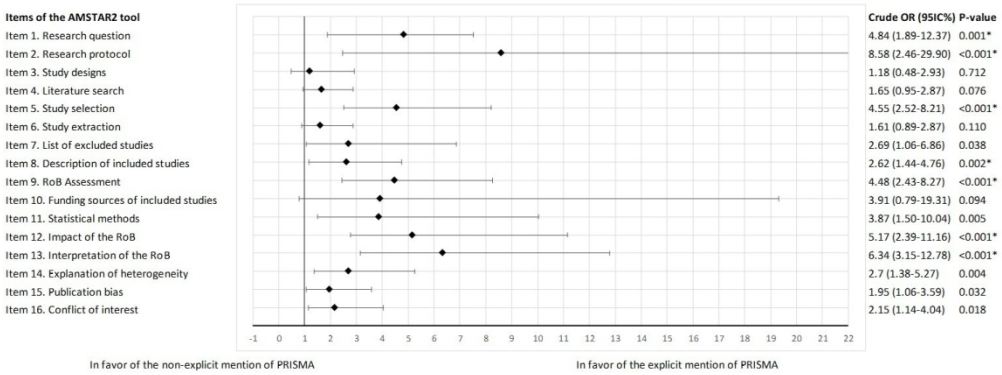


Figure 2. Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group. *Items statistically significant with the Bonferroni correction for multiple testing (p≤0.003).

128x48mm (300 x 300 DPI)

Register OSF

Study information

Title

Assessment of the reporting and methodological qualities and associated factors of a sample of meta-analyses recently indexed in PsycINFO (2016).

Authors

Victoria Leclercq – Charlotte Beaudart – Véronique Rabenda – Sara Ajamieh - Ezio Tirelli – Olivier Bruyère

Background

For scientists, searching for current best evidence has become a real challenge for scientists given the quasi limitless number of published articles (more than 1 270 000 in 2014, according to Thomson Reuteur's Web of Science). When facing a problematic implying a decision, scientists need documents and results oh high and reliable scientific value, as promoted by the evidence-based medicine movement (EBM). EBM is defined as the practice of medicine-based on knowledge and understanding of the literature in order to support clinical decisions (Guyatt et al. 2015). Following evidence hierarchy of EBM, systematic reviews (SRs) and meta-analyses (MAs) are considered the best level of evidence. Nowadays, in diverse disciplines, many researchers base their own research on the results of these SRs and MAs. The Cochrane collaboration adopted the definition of Antman (1992) and Oxman (1993) for the SR: "A systematic review attempts to collate all empirical evidence that pre-specified eligibility criteria in order to answer a specific research question" (Higgins & Green 2011) and the definition of Glass (1976) for MA : "Meta-analysis is the use of statistical methods to summarize the results of independent studies" (Higgins & Green 2011).

Some researchers have highlighted an increase in the publication rate of 2728% for SR (1024 articles in 1991 and 28 959 in 2014) and 2635% for MA (334 articles in 1991 and 9135 in 2014) (Ioannidis 2016). Several reasons could explain this phenomenon. In particular, fewer resources are necessary to perform SRs and MAs, which generally (yield) are worth high citation rates (contributing to increase the impact factor of the journal where they are published (Ioannidis 2016). However, an increasing number of studies have highlighted weaknesses in the design, conduct, analysis, and reporting of MAs published in many scientific fields (Zhu et al. 2016; Cullis et al. 2017; Peters et al. 2015; Zhang et al. 2016; Gagnier & Kellam 2013).

Two tools have been developed to evaluate the quality of the methodology ((AMSTAR, "A Measurement Tool to Assess systematic Reviews" (Shea et al. 2007)) and recently its update (AMSTAR 2 (Shea et al. 2017))) and another one for the quality of the reporting (PRISMA, "Preferred Reporting Items for Systematic Reviews and Meta-analyses" ((Moher et al. 2009)) of SRs and MAs. AMSTAR, a 11-item measuring tool aiming to assess the methodologic quality of MAs (Zhang et al. 2016), has been shown to be reliable and valid (Shea et al. 2009). AMSTAR 2, a 16-item measuring tool aiming to assess the methodologic quality of MAs of randomized and no randomized studies (Shea et al. 2017). PRISMA comprises a list of 27 items that are recommended to be used in the reporting of a MA in order to ensure that the article contains all relevant information (Moher et al. 2009). Several studies have already evaluated the quality of MAs published in specific medical fields such as surgery (Cullis et al. 2017; Zhang et al. 2016), depression (Zhu et al. 2016), orthopaedic (Gagnier & Kellam 2013) or even otorhinolaryngologic disorders (Peters et al. 2015). To the best of our knowledge, there are no such studies available in the field of psychological science.

In line with the EBM movement, the American Psychological Association (APA) has defined in 2006 the movement of Evidence-Based Practice in Psychology with the purpose "to promote effective psychological practice and enhances public health by applying empirically supported principles of psychological assessment, case formation, therapeutic relationship, and intervention"(American Psychological Association 2006). The American Psychological Association has brought out some benefits to the use of Reporting Standards whose the salutary effect on the way research has been conducted (Cooper 2008). The PRISMA statement could also have a positive effect on the methodological quality of the studies.

A growing meta-research literature has assessed the quality of empirical and experimental psychological studies in often large samples of articles (Ioannidis 2012; Bakker & Wicherts 2011; Oliveras et al. 2017; Stanley et al. 2017). It has revealed and quantified numerous methodological deficiencies, such as an inappropriate use of statistics, high rates of statistical mistakes, a frequent lack of statistical power (along with the neglect of effect size considerations) or the unambiguous presence of methodological biases, to mention but a few of them. Interestingly, a recent study conduct by Fanelli and co-workers (Fanelli, Costas, & Ioannidis, 2017) have highlighted differences in the risk of bias (poor estimate of the magnitude of effect size due to, for example, lower inclusion of grey literature, US effect or industry bias…) between the classical disciplines, the risk being highest in the social sciences (to which

psychology belongs). These differences could reflect dissimilar research practices documented in primary studies (e.g. higher publication bias in some disciplines) or distinct procedural choices in meta-analyses (e.g. lower inclusion of grey literature in some disciplines) (Fanelli et al., 2017).

To our knowledge, no studies have however been conducted in order to evaluate the quality of MAs published in the field of psychology. With this research project, our aim is to evaluate the quality of MAs and identify its associated factors published in the psychological or psychology-related field on the PsycINFO database during the year 2016.

Objectives

The objective of this research is to assess the factors associated with the quality of recent MAs indexed in PsycINFO for the year 2016 using two samples of MAs; one composed of MAs claiming to follow the PRISMA statement and the other one including MAs ignoring it.

Our research will be organized in three sub-studies:

1. The assessment of the reporting quality of MAs that claim to use the PRISMA statement compared to those that do not use PRISMA through the PRISMA statement;
2. The assessment of the methodological quality of MAs that claim to use the PRISMA statement compared to those that do not use PRISMA through the AMSTAR 2 tool;
3. The identification of potential factors associated with the quality of MAs.

Research questions

Based on our objectives, our research questions are the following:

1. What is the relationship between of the use of the PRISMA checklist on the reporting quality of MAs?
2. What is the relationship between of the use of the PRISMA checklist on the methodological quality of MAs?
3. What are the potential factors (e.g. publishing journal's impact factor, pre-registration of the study, experience of the first author…) associated with the quality of MAs?

Hypotheses

1. Comprehensive and transparent reporting is necessary to assessing the methodological quality of MAs (Page et al. 2016). They are some articles that highlighted that MAs have a poor reporting quality in the medical literature and the score of PRISMA Statement that were found is between 16,8 and 23/27 points ((Tunis et al. 2013; Fleming et al.

2013; Peters et al. 2015; Zhang et al. 2016; Gagnier & Kellam 2013; Adie et al. 2015; Zhu et al. 2016). We make the assumptions that the use of PRISMA statement improves the reporting quality of MAs using it. This is supported by another study showing that the PRISMA scores is higher by 1 point for the MAs for which authors claimed having used PRISMA statement (Zhu et al. 2016).

2. Meta-research has revealed that the psychological literature present an unsatisfactory level of methodological quality (Bakker & Wicherts 2011; Ioannidis 2012; Oliveras et al. 2017; Stanley et al. 2017). It is therefore possible that this is also true for the MA published in psychology and related fields. In recent studies analyzing the quality of Mas in a number of health-related fields AMSTAR scores have been found to fall between 3,7 and 7,8/11 points (Zhang et al. 2016; Adie et al. 2015; Gagnier & Kellam 2013; Klimo et al. 2014). Since the AMSTAR 2 tool was recently published, it is probable that no study has yet evaluated the quality of the MAs with this tool. Note that it is likely that the use of PRISMA statement exerts a positive influence on the quality of Mas using it. This is supported by a recent study on depression showing that the AMSTAR scores for the MAs for which authors claimed having used the PRISMA statement, reach an average of 0.4 point higher than those which did not use PRISMA (Zhu et al. 2016).
3. They are some potential factors that could correlate with (and possibly influence) the quality of MAs. More specifically, on the basis of previous meta-research studies we make the assumptions that the following factors are positively associated with the measures of quality of MAs: h-index of the first author (Cullis et al. 2017), experience of the principal author in MAs (Zhang et al. 2016), affiliation of the authors to a university (Cullis et al. 2017), publishing journal's impact factor (Cullis et al. 2017), PRISMA endorsement by the journal publishing (Cullis et al. 2017), funding sources described (Gagnier & Kellam 2013), Cochrane collaboration (Adie et al. 2015; Cullis et al. 2017; Zhu et al. 2016), number of pages of the manuscript (Adie et al. 2015; Cullis et al. 2017), pre-registration of the study (Cullis et al. 2017; Zhang et al. 2016; Zhu et al. 2016), non-Asian origine (Zhang et al. 2016) and meta-analyses of randomized controlled trials (Zhang et al. 2016; Zhu et al. 2016). Furthermore, we hypothesize that the following variables will be also associated with the quality of MAs: open access of the publication, open data (or data sharing), the field of psychology, number of individual studies in each MA, number of databases used, assessment of the quality of

individual studies and related tools used, pooling methods used to combine the data, assessment of the publication bias and related method used, assessment of the heterogeneity and related method used, the statistical software used and tendency of the conclusion.

Sampling plan

Existing Data

The tests can be considered as confirmatory given similar studies that have recently been conducted in the medical field.

Explanation of existing data

This is not applicable for our research protocol because no data have been collected so far.

Data collection procedures

Data collection

Data will be collected from a random sample of 200 MAs that will be divided into 2 groups (use of PRISMA vs no use of PRISMA).

Protocol selection of meta-analyses articles

All MAs performed on human subjects and published in English in 2016 in the electronic database PsycINFO will be searched. The electronic search strategy was developed with co-authors and the assistance of a librarian is available in *table 1*.

Table 1: Search strategy

1	meta analysis.md. (15886)
2	meta analysis/ (3940)
3	meta analys*.mp. (24573)
4	data pooling*.mp. (50)
5	2 or 3 or 4 (24599)
6	5 not 1 (10725)
7	1 or 6 (26611)
8	limit 7 to (English and human and yr="2016") (2159)

A total of 2159 potentially relevant MAs were identified in the PsycINFO database. Two investigators will independently review each title and abstract in order to exclude irrelevant articles and to only select the studies that meet inclusion criteria (full inclusion and exclusion criteria are available in Table 2). All discrepancies in opinion regarding the selection of articles

will be resolved through discussion and consensus between the two investigators; any persistent disagreement will be solved with the intervention of a third person (an expert).

Table 2 : eligibility criteria

Inclusion criteria
- Meta-analysis
- Articles published in the PsycINFO-database
- Published between 01.01.2016 to 31.12.2016
- English
Exclusion criteria
- Overview, review
- Meta-synthesis
- Qualitative meta-analysis
- Umbrella review
- Meta-analysis of meta-analyses
- Systematic review without meta-analysis
- Protocol of meta-analysis
- Network meta-analysis
- Activation likelihood Estimation Meta-analysis (ALE MA)
- Signed differential mapping meta-analysis (SMD MA)
- Voxel wise meta-analysis
- Individual patient data meta-analysis (IPD MA)
- Genetic association study (GWAS), genetic study
- Multi-level meta-analysis
- Update
- Letter, comment, abstract, chapter, erratum, dissertation or editorial journal

Choice of language for inclusion was based on expertise within our research team, due to budget constraints, limited time and resources.

A flowchart with the number of included studies will be elaborated. The reason of exclusion of articles will be presented at the step of full-text selection.

Once all MAs will be identified, we will randomly select a minimum of 200 articles as follows. All references of articles will be indexed in an Excel file and randomly assigned to a number. Then, we will rank the articles in ascending order. Two blinded researchers, with the intervention of a third researcher in case of disagreement, will classify MAs that meet inclusion criteria in either the group "with PRISMA" or "without PRISMA" until each group will contain a minimum of 100 MAs. Kappa statistics will be used to test inter-rater agreement.

Data extraction

Relevant data will be extracted from the full texts of all selected articles in a standardized Microsoft Excel spreadsheet by two independent investigators trained for this data extraction. We will record the following factors that might influence the quality of the MAs: characteristics of the manuscripts, characteristics of the study, objective(s) of the study, statistical analyses, characteristics of the protocol and items of PRISMA statement, AMSTAR tool and AMSTAR2 tool. All data extracted will be detailed in appendices. If any disagreements were to be observed between the two reviewers, they will be resolved by discussion, if necessary with the intervention of a third reviewer. Kappa statistics and absolute agreement (%) will be used to assess reproducibility.

Sample size

A first exploratory search in PsycINFO has yielded approximately 2000 articles, which are impossible to analyze for us in a reasonable period of time. We elected to randomly (see below for the method of randomization) select 200 MA articles (until each group, PRISMA and NO PRISMA, will contain a minimum of 100 MAs) from all eligible MAs published in 2016 and indexed in PsycINFO. There is no global MAs offering a synthetic effect size (of the published differences between the two samples) that could have been used to determine a priori a sample size allowing the detecting of a significant difference (power analyses). The chosen sample size can minimally detect a medium effect size (Cohen's d = 0.46065; as computed via G*Power) using a two-tailed Student t-test for independent groups taken at an alpha error probability of 0.05 and a power (1-beta error probability) of 0.90 (critical t = 1.9720). Note that smaller effects sizes cannot be detected (if existing) with such sample (n=100). The meaning and practical significance of the empirically obtained effect size will be discussed.

Sample size rationale

Considering the power analysis described above and constraints in terms of time, financial resources and staff, we will conduct this research on about 200 articles. We think that this will necessitate more than 500 hours of coding for each assessor.

Stopping rule

This is not applicable for our research protocol.

Variables

Manipulated variables

This cannot be applied to the present research protocol.

Measured variables

In order to verify our hypotheses, we will assess not only the quality of reporting and of conduct of MAs but also a set of variables to identify the potential factors that are associated with the quality of MAs.

Assessment of reporting quality

Eligible papers will first be assessed with the PRISMA statement. Each individual item of the PRISMA statement will be answered by "yes" or "no" response, respectively depending on the item being or not fulfilled and will be coded "1" or "0". The total score of the PRISMA statement is the addition of all items coded 1 with a maximum of 27 points.

Assessment of methodology quality

Eligible papers will then be assessed with AMSTAR 2 tool. Each individual item of the AMSTAR tool will be answered by a "yes", "partial yes" or "no" response, respectively depending on the item being or not fulfilled and will be coded "1", "0.5", "0". The total score given by the tool is the addition of all items coded 1 and 0.5 with a maximum of 16 points.

Eligible papers will also be assessed with the AMSTAR tool. Each individual item of the AMSTAR tool will be answered by a "yes" or "no" response, depending on the item being or not fulfilled and will be coded "1" or "0". The total score given by the tool is the addition of all items coded 1 with a maximum of 11 points.

Identification of potential factors associated with the quality of MA

All factors will be assessed by three types of variables: dichotomous variables, quantitative variables and text variables (open questions).

Dichotomous variables

Dichotomous variables will be coded as follows:

1: Yes, it features the characteristic that we seek.

0: No, it does not feature the characteristic that we seek.

99: Not reported, the characteristic that we seek is not available.

88: Not applicable.

The variables that are concerned are the following: author's experience in meta-analysis, affiliation of the authors to a university, PRISMA's recommendation by the journal, restriction of the word count by the journal, declaration of conflict of interest, declaration of funding sources, Cochrane collaboration, open access, open data, registration of the study in a database, evaluation of the quality of study, use of reporting or methodology guideline, the type of study (randomized controlled trials (RCT) or not), evaluation of publication bias, evaluation of heterogeneity, presence of a protocol and the conclusion supports the assumptions.

Quantitative variables

They will be encoded with numerical values and their units of measurements.

The variables that are concerned are the following: h-index (an author-level metric), number of authors, impact factor of the journal (which reflects the frequency with which the average article in a journal has been cited in a particular year) and number of database consulted.

Qualitative variables

The relevant variables are the following: the continent where the study was conducted (Europe, Asia, Africa, America, Oceania), number of study included in each MA (0-3; 4-9; ≥10) and the pooling method used (random effect model, fixed effect model or mixed effect model).

Text variables

The remaining variables are recorded as a text variable. The variables that are concerned are the following: the name of the tool used in order to assess the quality of individual study, the name of the database(s) searched, the PsycINFO classification, the name of the guideline used, the method used to evaluate the publication bias, the method used to assess the heterogeneity and the statistical software. These text variables will then be categorized.

If the data is not available, it will be coded 88.

Indices

This is not applicable for the present research protocol.

Design plan

Study type

This is an observational study. Data are collected from meta-analysis articles.

Blinding

The selection of MAs reviews by title and abstract and by full-text will be done independently by two investigators.

Study design

This is a cross-sectional study.

Randomization

We will randomly MAs articles to get a minimum of 100 MAs in each group. All references of articles (n= probably more than 2000) will be indexed in an Excel file and randomly assigned to a number. Then, we will rank the articles in ascending order. Two blinded researchers, with the intervention of a third researcher in case of disagreement will classify MAs that meet inclusion criteria in either group “with Prisma” and “without Prisma” until each group will contain a minimum of 100 articles MAs.

Analysis plan

Statistical model

The characteristics of all individual studies will first be presented. All quantitative variables that follow a normal distribution will be reported as mean and standard deviation and those that do not follow a normal distribution will be represented as median and percentile (P25 and P75). Distribution will be considered as normal if data meet 3 of the 4 following conditions: the mean is close to the median, the Shapiro-Wilk normality test yields a p-value </=0.05, the curve of the variable follows the normal (or Gaussian) distribution and the linearity of the QQ-Plots is respected. Qualitative and dichotomous variables will be reported as numbers and frequencies. The results of the quality assessment of MAs with the PRISMA statement, AMSTAR tool and AMSTAR 2 will be reported, for quantitative variable, as number and frequency for each item and as mean or as median for the total score. The data will be presented and analyzed using a

star chart. A star chart is a graphical tool that will allow us to represent and compare the percentages of item of PRISMA statement, AMSTAR tool and AMSTAR 2 met by the MAs.

To verify our first and second hypothesis, the reporting and methodological qualities of the individual studies will be compared between the studies that report using the PRISMA checklist and the studies that do not. Comparisons of means between the two groups will be calculated using the Student t-test if for independent groups if the score of PRISMA, AMSTAR and AMSTAR 2 are normally distributed and the Mann-Whitney test if the score of PRISMA, AMSTAR and AMSTAR 2 are not normally distributed. To test the association between the use of the PRISMA statement and the different items of PRISMA, AMSTAR and AMSTAR 2, we will be used a logistic regression.

To test our third hypothesis, factors (all the data detailed in the measured variables part) with potential influence on the quality of studies (mean score of AMSTAR, mean score of AMSTAR 2, independent variable) will be identified with a univariate linear regression. The variables with a p-value lower than 0.1 will be combined in stepwise backward multiple regression analysis. A p-value </=0.05 will be considered as significant.

Transformations

Each item of the PRISMA & AMSTAR checklists will be coded with the following meaning:

1 = Yes or not applicable

0 = No

We will also sum up all items coded 1, with a maximum score of 27 or 11, respectively.

Each item of AMSTAR 2 will be coded with the following meaning:

1 = Yes or not applicable

0.5 = Partial yes

0 = No

We will also sum up all items coded 1 and 0.5, with a maximum score of 16.

See appendix for the details of all transformations for each variable.

Follow-up analyses

All follow-up analyses are described above.

Inference criteria

Not applicable for our research protocol.

Data exclusion

No data will be excluded from our database.

Missing data

Missing data may have an impact on the analysis and on the interpretation of the results. Some of the extracted data may not be available (h-index, impact factor…). After data encoding, a quality control will be done, at database-level, in order to check for outliers, coding error and missing values. In case of incomplete information, we will contact the authors.

Exploratory analysis

The exploratory analyses will be considered, based on the results obtained.

If there is a statistically significant difference in quality between MAs which report using PRISMA and those which do not, we will consider carrying out the following analyses. A logistic regression will be carried out in order to describe the relation between the dichotomous dependent variables (PRISMA vs No PRISMA) and all potential explanatory variables (all the data are detailed in the measured variables part). The variables with a p-value lower than 0.1 will be combined in stepwise backward multiple logistic regression analysis. A p-value <0.05 will be considered significant on statistical analyses.

Scripts

Upload an analysis script with clear comments

Not available at the moment.

Other

We would like to acknowledge Pr. Anne-Françoise Donneau for interesting discussions about some aspects on the planned statistical analyses and Ms Nancy Durieux for her assistance in the building of our strategy, in terms of electronic literature search.

Bibliography

Adie, S. et al., 2015. Quality of Conduct and Reporting of Meta-analyses of Surgical Interventions. *Annals of Surgery*, 261(4), pp.685–694. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000658-201504000-00012 [Accessed June 14, 2017].

American Psychological Association, 2006. Evidence-Based Practice in Psychology APA Presidential Task Force on Evidence-Based Practice. *American Psychologist*, 61(4), pp.271–285. Available at: http://www.apa.org/pubs/journals/features/evidence-based-

statement.pdf [Accessed January 10, 2018].

Bakker, M. & Wicherts, J.M., 2011. The (mis)reporting of statistical results in psychology journals. Available at: https://link.springer.com/content/pdf/10.3758%2Fs13428-011-0089-5.pdf [Accessed July 31, 2017].

Cooper, H., 2008. Reporting Standards for Research in Psychology: Why Do We Need Them? What Might They Be? *American Psychologist*, 63(9), pp.839–851. Available at: https://www.apa.org/pubs/authors/jars.pdf [Accessed January 15, 2018].

Cullis, P.S., Gudlaugsdottir, K. & Andrews, J., 2017. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. *PLoS ONE*, 12(4). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5383307/pdf/pone.0175213.pdf [Accessed June 14, 2017].

Fanelli, D., Costas, R. & Ioannidis, J.P.A., Meta-assessment of bias in science. Available at: http://www.pnas.org/content/114/14/3714.full.pdf [Accessed December 18, 2017].

Fleming, P.S. et al., 2013. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. *The Angle Orthodontist*, 83(1), pp.158–163. Available at: http://www.angle.org/doi/10.2319/032612-251.1 [Accessed May 24, 2018].

Gagnier, J.J. & Kellam, P.J., 2013. Reporting and Methodological Quality of Systematic Reviews in the Orthopaedic Literature. *The Journal of Bone and Joint Surgery-American Volume*, 95(11), pp.e77-1–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23780547 [Accessed June 14, 2017].

Guyatt, G. et al., 2015. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Medicine*, Available at: https://jamaevidence.mhmedical.com/book.aspx?bookID=847 [Accessed September 1, 2017].

Higgins, J. & Green, S., 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011].* The Cochra., Available at: http://handbook.cochrane.org.

Ioannidis, J.P.A., 2016. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, 94(3), pp.485–514. Available at: http://doi.wiley.com/10.1111/1468-0009.12210.

Ioannidis, J.P.A., 2012. Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), pp.645–654. Available at: http://pps.sagepub.com/ [Accessed

July 31, 2017].

Klimo, P. et al., 2014. Methodology and reporting of meta-analyses in the neurosurgical literature. *Journal of neurosurgery*, 120(4), pp.796–810. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24460488.

Moher, D. et al., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). *Physical Therapy*, 89(9), pp.873–880. Available at: http://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1000097&type=printable [Accessed August 25, 2017].

Oliveras, I., Losilla, J.-M. & Vives, J., 2017. Methodological quality is underrated in systematic reviews and meta-analyses in health psychology. *Journal of Clinical Epidemiology*, 86, pp.59–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28499846 [Accessed November 28, 2017].

Page, M.J. et al., 2016. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study N. Low, ed. *PLOS Medicine*, 13(5), p.e1002028. Available at: http://dx.plos.org/10.1371/journal.pmed.1002028 [Accessed May 24, 2018].

Peters, J.P.M. et al., 2015. Reporting quality of systematic reviews and meta-analyses of otorhinolaryngologic articles based on the PRISMA statement. *PLoS ONE*, 10(8). Available at: http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0136540&type=printable [Accessed June 14, 2017].

Shea, B.J. et al., 2017. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, p.j4008. Available at: http://www.bmj.com/content/bmj/358/bmj.j4008.full.pdf [Accessed November 28, 2017].

Shea, B.J. et al., 2009. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), pp.1013–1020. Available at: http://ac.els-cdn.com/S0895435608003259/1-s2.0-S0895435608003259-main.pdf?_tid=7ad1b7aa-558b-11e7-a3dc-00000aacb361&acdnat=1497944501_0f82a189082ff8754f64a91d9670402e [Accessed June 20, 2017].

Shea, B.J. et al., 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*, 7, p.10. Available at: https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-7-10?site=bmcmedresmethodol.biomedcentral.com [Accessed August 25, 2017].

Stanley, T.D., Carter, E.C. & Doucouliagos, H., 2017. What Meta-Analyses Reveal about the Replicability of Psychological Research. Available at: http://www.deakin.edu.au/__data/assets/pdf_file/0007/1198456/WhatMeta-AnalysesReveal_WP.pdf [Accessed January 10, 2018].

Tunis, A.S. et al., 2013. Association of Study Quality with Completeness of Reporting: Have Completeness of Reporting and Quality of Systematic Reviews and Meta-Analyses in Major Radiology Journals Changed Since Publication of the PRISMA Statement? *Radiology*, 269(2), pp.413–426. Available at: http://pubs.rsna.org/doi/10.1148/radiol.13130273 [Accessed May 24, 2018].

Zhang, H. et al., 2016. Reporting and methodological qualities of published surgical meta-analyses. Available at: http://ac.els-cdn.com/S0895435615003121/1-s2.0-S0895435615003121-main.pdf?_tid=4e369bf4-50e8-11e7-9b21-00000aab0f26&acdnat=1497434613_6d09bcd5abe078df3344e91ace4b09d1 [Accessed June 14, 2017].

Zhu, Y. et al., 2016. Is the best evidence good enough: Quality assessment and factor analysis of meta-analyses on depression. *PLoS ONE*, 11(6). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4919061/pdf/pone.0157808.pdf [Accessed June 14, 2017].

Appendix
Explication of the data extraction form

Name	Explication	Description
Name of the reviewer	Name of the reviewer	Text
Study ID	Reference number of the article	Text
Inclusion of the article	Inclusion of the article based on the selection criteria	1 = Yes 0 = No
If excluded, indicate the reason of exclusion	Reason of exclusion	The reason of exclusion 88 = if not applicable
Use of PRISMA	The authors declared the use of PRISMA	1= Yes 0 = No

1. Characteristics of the manuscript

Name	Explication	Description
DOI of the article	Unique identifier of the article	Text
Year of publication	Publishing year of the manuscript	Text
Author's name	Name of the first author	Text
Author's h-index	H-Index of the first author (Scopus)	Quantitative variable
Author's experience	Number of meta-analyses from the same author(s) (Scopus)	Quantitative variable
Affiliation of the authors to a university	Affiliation of the authors to a university	1 = Yes 0 = No 99 = Not reported
Number of authors	Total number of authors	Quantitative variable
Contribution of authors	Details of the authors' contribution	1 = Yes 0 = No 99 = Not reported
Journal's name	Name of the journal	Text
Journal's Impact factor	The IF of the journal using the ISI Journal Citation Reports 2016 (http://isiknowledge.com)	Quantitative variable
Instruction for authors: PRISMA required?	The journal recommended to use PRISMA statement	1 = Yes 0 = No 99 = Not reported

Instruction for authors: page or word limitation?	Limitation of the number of pages or words	1 = Yes 0 = No 99 = Not reported
PsycINFO classification	Classification of the field of psychology based on the PsycINFO Content Classification Code System	2100 = General Psychology 2200 = Psychometrics & Statistics & Methodology 2300 = Human experimental Psychology 2400 = Animal Experimental & comparative Psychology 2500 = Physiological Psychology & Neuroscience 2600 = Psychology & The Humanities 2700 = Communication Systems 2800 = Developmental Psychology 2900 = Social Processes & Social Issues 3000 = Social Psychology 3100 = Personality Psychology 3200 = Psychological & Physical disorders 3300 = Health & Mental Health Treatment & Prevention 3400 = Professional Psychological & Health Personnel Issues 3500 = Educational Psychology 3600 = Industrial & Organizational Psychology 3700 = Sport Psychology & Leisure 3800 = Military Psychology 3900 = Consumer Psychology 4000 = Engineering & Experimental Psychology 4100 = Intelligent Systems 4200 = Forensic Psychology & Legal Issues
Corresponding author (email address)	Email of the author	Text

Conflict of interest described	Conflict of interest is described	1 = Yes 0 = No 99 = Not reported
Details of conflict of interest	If yes, brief description of conflict of interest	Text
Funding sources described	Funding sources are described	1 = Yes 0 = No 99 = Not reported
Funding sources	If yes, brief description of funding sources	Text
Cochrane collaboration	The study is a Cochrane collaboration	1 = Yes 0 = No
Number of page of manuscript	Total number of pages of the manuscript	Quantitative variable
Open access	The publication is open access?	1 = Yes 0 = No 99 = Not reported
Open data	The data is open access?	1 = Yes 0 = No 99 = Not reported

2. Characteristics of the study

Name	Explication	Description
Registration of the study?	The study was recorded in a specific database.	1 = Yes 0 = No 99 = Not reported
Number of the registration	Registration number of the study	Text 88 = if not applicable
Name of the registry	Name of the registry in which the meta-analysis has been registered	Text 88 = if not applicable
Date of submission of the manuscript	Date of submission of the manuscript	Text (month-year)
Date of publication of the manuscript	Publication date of the manuscript	Text (month-year)
Continent of origin of first author	Continent in which the study has been conducted	Europe – Asia – Africa – America - Oceania
Type of the individual study	Study design of the studies included in the MA	Observational study RCT All types Not specified
Number of databases searched	Number of databases consulted	Quantitative variable

Name of the database	Name of the database searched	Text
Quality of individual study is assessed?	Quality of individual study is assessed	1 = Yes 0 = No
Name of the tool used to assess the quality	Name of the tool used to assess the quality of individual studies	Text variable 88 = If not applicable
Reference to use of the guideline	Reference to use of a guideline	1 = Yes 0 = No
Name of the guideline used	Name of the guideline used (PRISMA, MOOSE, AMSTAR…)	Text 88 = If not applicable
Search strategy	Presence of the complete search strategy	1 = Yes 0 = No 99 = Not reported
Focus of review	Type of the field of psychology	Text

3. Objective of the study

Name	Explication	Description
Main objective	Aim of the study	Text
Primary outcome	Primary outcome of the study disclosed	Text
Secondary outcomes	Secondary outcome of the study disclosed	Text

4. Statistical analyses

Name	Explication	Description
Number of meta-analyses performed	Number of meta-analyses performed in the presented study	Quantitative variable
Number of studies included in each meta-analysis	Number of studies included in each meta-analysis performed in the study	0-3; 4-9; ≥10
Pooling methods	The pooling methods used to combine data	Fixed – Random - Mix
Assessment of the publication bias	The publication bias is evaluated	1 = Yes 0 = No
Method used to assess the publication bias	Method used to assess the publication bias	Text 88 = If not applicable

Assessment of the heterogeneity	The heterogeneity is evaluated	1 = Yes 0 = No
Heterogeneity	Method used to assess the heterogeneity	Text 88 = If not applicable

5. Protocol

Name	Explication	Description
Protocol	The protocol of the study is existent and available	1 = Yes 0 = No
Primary outcome	Primary outcome of the study	Text 88 = If not applicable
Secondary outcome	Secondary outcome of the study	Text 88 = If not applicable

6. Conclusion

Name	Explication	Description
Conclusion	Main conclusion of the study	Text
Trends of the conclusion	The conclusion supports the assumptions	1 = Yes 0 = No

7. PRISMA statement

Name	Explication	Description
P1	TITLE Title	1 = Yes 0 = No
P2	ABSTRACT Structured summary	1 = Yes 0 = No
P3	INTRODUCTION Rationale	1 = Yes 0 = No
P4	Objective	1 = Yes 0 = No
P5	METHODS Protocol and registration	1 = Yes 0 = No
P6	Eligibility criteria	1 = Yes 0 = No
P7	Information sources	1 = Yes 0 = No

P8	Search	1 = Yes 0 = No
P9	Study selection	1 = Yes 0 = No
P10	Data collection process	1 = Yes 0 = No
P11	Data items	1 = Yes 0 = No
P12	Risk of bias in individual studies	1 = Yes 0 = No
P13	Summary measures	1 = Yes 0 = No
P14	Synthesis of results / Planned methods of analysis	1 = Yes 0 = No
P15	Risk of bias across studies	1 = Yes 0 = No
P16	Additional analysis	1 = Yes 1 = Not applicable 0 = No
P17	RESULTS Study selection	1 = Yes 0 = No
P18	Study characteristics	1 = Yes 0 = No
P19	Risk of bias within studies	1 = Yes 0 = No
P20	Results of individual studies	1 = Yes 0 = No
P21	Synthesis of results	1 = Yes 0 = No
P22	Risk of bias across studies	1 = Yes 0 = No
P23	Additional analysis	1 = Yes 1 = Not applicable 0 = No
P24	DISCUSSION Summary of evidence	1 = Yes 0 = No
P25	Limitations	1 = Yes 0 = No
P26	Conclusions	1 = Yes 0 = No
P27	Funding	1 = Yes 0 = No
	Total score	Sum of all items coded "1"

8. Amstar tool

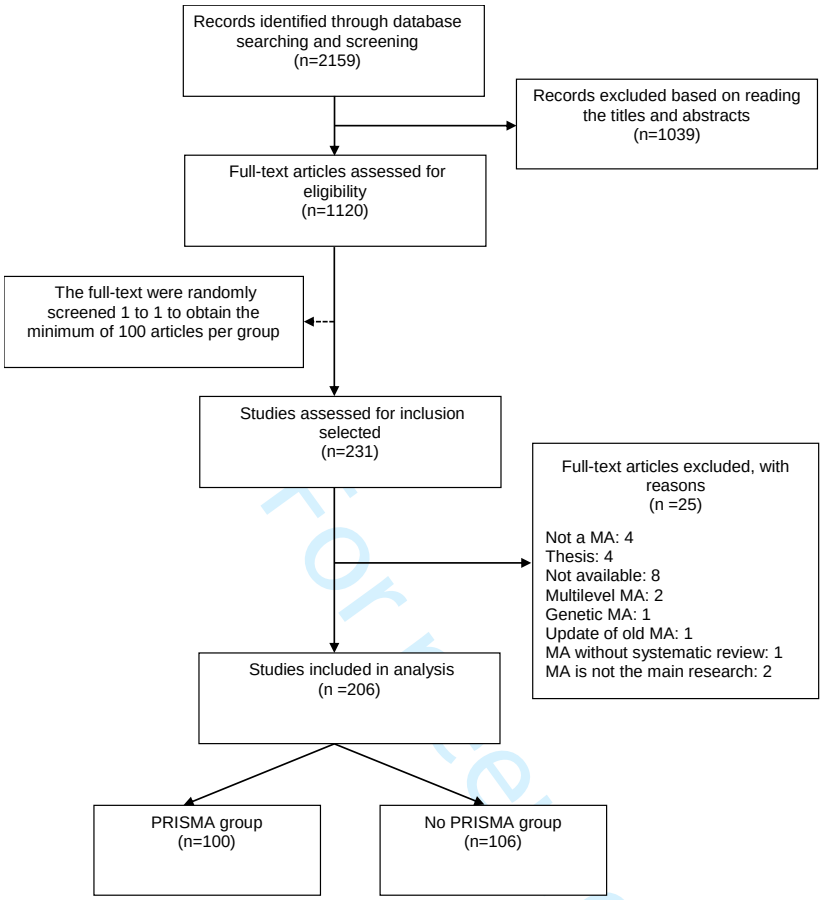
Name	Explication	Description
A1	Was an 'a priori' design provided?	1 = Yes 0 = No
A2	Was there duplicate study selection and data extraction?	1 = Yes 0 = No
A3	Was a comprehensive literature search performed?	1 = Yes 0 = No
A4	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	1 = Yes 0 = No
A5	Was a list of studies (included and excluded) provided?	1 = Yes 0 = No
A6	Were the characteristics of the included studies provided?	1 = Yes 0 = No
A7	Was the scientific quality of the included studies assessed and documented?	1 = Yes 0 = No
A8	Was the scientific quality of the included studies used appropriately in formulating conclusions?	1 = Yes 0 = No
A9	Were the methods used to combine the findings of studies appropriate?	1 = Yes 0 = No
A10	Was the likelihood of publication bias assessed?	1 = Yes 0 = No
A11	Was the conflict of interest included?	1 = Yes 0 = No
	Total score	Sum of all items coded "1"

9. Amstar 2 tool

Name	Explication	Description
AM1	Did the research questions and inclusion criteria for the review include the components of PICO?	1 = Yes 0 = No

AM2	Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	1 = Yes 0.5 = Partial Yes 0 = No
AM3	Did the review authors explain their selection of the study designs for inclusion in the review?	1 = Yes 0 = No
AM4	Did the review authors use a comprehensive literature search strategy?	1 = Yes 0.5 = Partial Yes 0 = No
AM5	Did the review authors perform study selection in duplicate?	1 = Yes 0 = No
AM6	Did the review authors perform data extraction in duplicate?	1 = Yes 0 = No
AM7	Did the review authors provide a list of excluded studies and justify the exclusions?	1 = Yes 0.5 = Partial Yes 0 = No
AM8	Did the review authors describe the included studies in adequate detail?	1 = Yes 0.5 = Partial Yes 0 = No
AM9	Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	1 = Yes 0.5 = Partial Yes 0 = No
AM10	Did the review authors report on the sources of funding for the studies included in the review?	1 = Yes 0 = No
AM11	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	1 = Yes 0 = No
AM12	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	1 = Yes 0 = No
AM13	Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	1 = Yes 0 = No
AM14	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	1 = Yes 0 = No

AM15	If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	1 = Yes 0 = No
AM16	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	1 = Yes 0 = No
	Total score	Sum of all items coded "1" or "0.5"



Supplementary file 2: Flowchart illustrating the MAs Selection

Supplementary file 3

Table 1. Sensitivity analyses : Impact of the explicit mention of PRISMA on the methodological characteristics of MAs: the non-explicit mention of PRISMA group vs the explicit mention of PRISMA group using univariate logistic regression with the three responses modality of AMSTAR2 (Yes, partial yes and no).

AMSTAR2's items		**OR**	**95%CI**	**p-value**
Item 1. Research question	Yes	4.84	1.90-12.37	0.001*
Item 2. Research protocol	Yes	6.01	1.67-21.62	0.006
	Partial yes	NE	NE	0.933
Item 3. Study designs	Yes	1.19	0.48-2.93	0.711
Item 4. Literature search	Yes	1.71	0.59-4.90	0.321
	Partial yes	1.64	0.91-2.94	0.098
Item 5. Study selection	Yes	4.55	2.52-8.21	<0.0001*
Item 6. Study extraction	Yes	1.61	0.90-2.87	0.110
Item 7. List of excluded studies	Yes	2.53	0.98-6.49	0.054
	Partial yes	NE	NE	0.958
Item 8. Description of included studies	Yes	4.44	2.12-9.33	<0.0001*
	Partial yes	1.80	0.93-3.50	0.083
Item 9. RoB Assessment	Yes	4.42	2.32-8.42	<0.0001*
	Partial yes	4.87	1.23-19.25	0.024
Item 10. Funding sources of included studies	Yes	3.91	0.79-19.31	0.094
Item 11. Statistical methods	Yes	3.87	1.49-10.04	0.005
Item 12. Impact of the RoB	Yes	5.17	2.39-11.16	<0.0001*
Item 13. Interpretation of the RoB	Yes	6.34	3.15-12.78	<0.0001*
Item 14. Explanation of heterogeneity	Yes	2.70	1.38-5.27	0.004
Item 15. Publication bias	Yes	1.95	1.06-3.59	0.032
Item 16. Conflict of interest	Yes	2.15	1.14-4.04	0.018

Reference group for all items is "No"; NE : Not estimable, calculation of the OR is not estimable because the occurrence of the outcome of interest is small and there was a zero value in the contingency table.
*Items statistically significant with the Bonferroni correction for multiple testing ($p \leq 0.003$).