

Journal Pre-proof

Methodological issues of systematic reviews and meta-analyses in the field of sleep medicine: A meta-epidemiological study

Chang Xu, Luis Furuya-Kanamori, Joey S.W. Kwong, Sheng Li, Yu Liu, Suhail A. Doi



PII: S1087-0792(21)00019-8

DOI: <https://doi.org/10.1016/j.smrv.2021.101434>

Reference: YSMRV 101434

To appear in: *Sleep Medicine Reviews*

Received Date: 30 March 2020

Revised Date: 18 September 2020

Accepted Date: 18 September 2020

Please cite this article as: Xu C, Furuya-Kanamori L, Kwong JSW, Li S, Liu Y, Doi SA, Methodological issues of systematic reviews and meta-analyses in the field of sleep medicine: A meta-epidemiological study, *Sleep Medicine Reviews*, <https://doi.org/10.1016/j.smrv.2021.101434>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 The Author(s). Published by Elsevier Ltd.

Methodological issues of systematic reviews and meta-analyses in the field of sleep medicine: A meta-epidemiological study

Chang Xu^a, Luis Furuya-Kanamori^b, Joey S.W. Kwong^c, Sheng Li^d, Yu Liu^e, Suhail A. Doi^a

^a Department of Population Medicine, College of Medicine, Qatar University, Doha, Qatar;

^b Research School of Population Health, Australian National University, Canberra, Australia;

^c JC School of Public Health and Primary Care, Faculty of Medicine, Chinese University of Hong Kong, Hong Kong, China;

^d Department of Biological Repositories, Human Genetics Resource Preservation Center of Hubei Province, Zhongnan Hospital of Wuhan University, China;

^e Gansu Provincial Maternity and Child-care Hospital, Gansu, China;

Corresponding author:

Dr. Chang Xu, Department of Population Medicine, College of Medicine, Qatar University, Al Jamiaa Street, P. O. Box 2713, Doha, Qatar; Email: xuchang2016@runbox.com

Acknowledgments: None.

Conflicts of interests: None.

Funding: LFK was supported by Australian National Health and Medical Research Council Early Career Fellowships (APP1158469).

Summary

An increasing number of systematic reviews and meta-analyses (SRMAs) have been published in the field of sleep medicine. We evaluated the methodological issues of these SRMAs. A protocol was developed in advance. Three databases were searched from inception to October 2019 for SRMAs published in major academic journals of sleep medicine that assessed healthcare interventions. The AMSTAR 2.0 instrument was used to evaluate the methodological issues and a multivariable regression analysis was conducted to investigate potential measures associated with methodological validity. We identified 163 SRMAs. The median number of missing safeguards of these SRMAs was 7 out of 16 (Interquartile range, IQR: 6-9), and on average, 2 of these missing safeguards were critical weaknesses. Our regression analysis suggested that SRMAs published in recent years ($\beta=0.16$; 95%CI: 0.08, 0.24; $p=0.002$), with the first author from Europe ($\beta=0.08$; 95%CI: 0.02, 0.14; $p=0.013$) tend to have higher relative methodological ranks. In conclusion, the methodological validity for current SRMAs in sleep medicine was poor. Further efforts to improve the methodological validity are needed.

Keywords: Evidence-based decision, systematic reviews, meta-analysis, methodology validity, meta-epidemiological study

Abbreviations: Assessment of Multiple Systematic Reviews: AMSTAR; Systematic review and meta-analysis: SRMA; Interquartile range: IQR; Population, Intervention, Comparison, and Outcome: PICO

1. Introduction

In the era of evidence-based medicine, credible evidence is the foundation upon which trustworthy decisions are built [1]. Systematic review and meta-analysis (SRMA) serves as an important source of evidence to support decision-making [2]. A systematic review summarizes findings from all available studies of the same topic, with or without a meta-analysis, and is expected to provide comprehensive evidence [3-5]. However, whether the evidence from a SRMA is credible largely rely on its design and conduct. These include but are not limited to 1) how the literature was searched and screened, 2) how the data were collected and analysed, and 3) how the results were interpreted and discussed. SRMAs involving methodological issues may generate non-credible results and mislead clinical practice [6].

In order to make a valid evaluation regarding methodological quality, several instruments have been developed in the past. These include the Sacks' checklist [7], the Overview Quality Assessment Questionnaire [8], the Assessment of Multiple Systematic Reviews (AMSTAR) [9], and the updated version of AMSTAR (AMSTAR 2.0) [10]. These instruments are widely used to assess methodological issues related to SRMAs. Jadad et al evaluated 50 SRMAs on the treatment of asthma and found that even after peer review there remain serious methodological flaws [11]. Xu et al investigated 529 dose-response meta-analyses and found that 87.9% of them were poorly designed and conducted [12]. These studies reveal that a large proportion of SRMAs may have poorly implemented safeguards that validate their conclusions.

Since SRMA was introduced to the field of sleep medicine, there has been an increasing number of SRMAs published during the past decades, some of which have been used as evidence in clinical guidelines (e.g. [13-15]) that governs physician's decisions, patients' behaviours, and administrators' policies. What makes things

worrisome is the question around how well these SRMAs were designed and conducted and therefore whether the evidence they produced is credible. In this review, we conducted a comprehensive assessment to examine the methodological issues as well as investigate potential mechanisms to improve SRMAs conducted in the field of sleep medicine.

2. Methods

2.1 Protocol

A protocol for the meta-epidemiological study was developed in advance to formulate the design and conduct of this study (*appendix 1*). A meta-epidemiological study is defined as a methodological survey that “*aims to evaluate trends and patterns in the literature with the overarching goal of improving the design, methods and conduct of future research*” [16-18]. The protocol contained details regarding the review question, eligibility criteria, literature search, screen, quality assessment, data collection, and data analysis. Some changes were made: First, we limited the inclusion criteria to focus on systematic reviews with meta-analysis on healthcare interventions as suggested by the reviewers; second, we replaced the pre-defined subgroup analysis with a regression analysis considering that the interaction test of the potential difference of the effects among groups is underpowered when there are 3 or more categories [19]; further, we added a *post-hoc* sensitivity analysis for the regression to test the robustness of the results.

2.2 Eligibility criteria

We included systematic reviews with meta-analyses or meta-analyses alone on

healthcare intervention published in the major academic journals of sleep medicine. We focused on healthcare intervention because AMSTAR 2.0 was designed to assess such types of systematic reviews [10]. Systematic reviews had to contain a quantitative synthesis, as the appropriate use of meta-analysis is part of the outcome of interest. The definition of systematic review has been clearly documented in the Cochrane handbook [20]. A meta-analysis refers to a statistical and quantitative synthesis of available findings of similar studies on the topic in question, which is generally regarded as a type of systematic review [21-24]. Overviews, scoping reviews, and narrative reviews were not considered since they differ from SRMAs [25]. Pooled analysis that do not use a regular literature search for at least one database were also not considered. Studies that consisted of original data plus a systematic review/overview/scoping review, again, were not considered. The primary outcome of the current review was the methodological flaws within the eligible studies. The secondary outcome was to examine the association between baseline characteristics (see data analysis part) and methodological weaknesses.

2.3 Literature search and screen

Literature search was conducted by one experienced researcher (XC). We searched for SRMAs published in academic journals in sleep medicine indexed in PubMed, Medline and Embase databases from inception to 22-Oct, 2019. We identified 23 related journals from SCImago Journal & Country Rank (<https://www.scimagojr.com/>) such as “Sleep”, “Sleep medicine reviews”, “Sleep medicine”. Of these we excluded four predatory journals (e.g. Journal of sleep disorders & therapy) based on the Beall’s list and 19 journals were included. Of these non-predatory journals, we used indexing status as an additional criterion (MedLine versus non-MedLine). A full list of the

journals and search strategy are presented in the appendix (*appendix 1*). Grey literature was not considered as we only aimed at peer reviewed SRMAs. We did not review the reference lists of eligible SRMAs since the sample would be sufficient and representative.

The Endnote X7 software was used to find duplicates. The Rayyan online app (<https://rayyan.qcri.org/>) was used for literature screening, which allows blinding of the raters to ensure the process was independent. Titles and abstracts were first screened by the lead author (XC) and those that were clearly not SRMAs were removed, a *post hoc* double-check of these excluded studies were performed by another author (LFK); the full-text of the remaining records were screened by two researchers (XC and LY) separately to make a further decision. Any disagreements were recorded and discussed until consensus was reached. The Cohen's kappa statistics was used for assessing inter-rater agreement [26, 27].

2.4 Data collection

Baseline characteristics such as first author's name, number of authors, year of publication, region of affiliation of the first author, number of studies included, use of reporting guideline, funding information, type of main meta-analysis used in systematic reviews, and journal of publication of each SRMAs were extracted. This was done by one researcher (LY) and double checked by another researcher (XC). This information could be directly extracted and therefore no missing data was expected.

Meta-analyses were categorized as either a standard meta-analysis or a special type of meta-analysis. A standard meta-analysis was defined by use of classical synthesis methods based on head-to-head comparisons; special type of meta-analysis

were those that involved more sophisticated assumptions and comparisons including diagnostic meta-analysis, dose-response meta-analysis, network meta-analysis, activation likelihood estimation meta-analysis, meta-analysis of prevalence, meta-analysis of means, meta-analysis of correlations, and meta-analysis of nucleotide polymorphism [28-40]. A detailed description of different types of meta-analyses is presented in Table S1.

2.5 Evaluation of methodological issues

The AMSTAR 2.0 instrument (<https://amstar.ca/Amstar-2.php>) was used to evaluate the potential methodological issues of eligible SRMAs [10]. The validity and reliability of this instrument has been critically assessed [41]. The AMSTAR 2.0 instrument (*appendix 2*) consists of 16 methodological items for SRMA, seven of which (items 2, 4, 7, 9, 11, 13, and 15) have been flagged as critical items [10]. In the current study, we did not consider item 2 as a critical item because the importance of protocol registration for methodological validity is not well verified [42]. Methodological weaknesses in any of the remaining six items would have greater impact on the validity of the SRMA and therefore the conclusions of the study.

The global methodology rating of a SRMA has routinely been judged by how many critical and non-critical weakness were identified, for example, high quality has been denoted as presence of none or only one non-critical weakness and critical low quality as two or more critical weakness [10]. However, such a judgement is somewhat arbitrary and anchors the assessment to the tool we have used. In order to make this universally valid, we used the *relative quality rank* as an alternative to measure the global methodological rating. This was done by enumerating items implemented out of 16 and creating a relative quality rank by dividing each

enumerated count of safeguards by the maximum count across the SRMAs. The best SRMA thus has a rank of 1 (which serves as the anchor) and all lesser values are below this (range zero to 1).

There were two (“Yes” or “No”) possible responses for each item, except for items 2, 4, 7, 8, and 9 where three possible responses (“Yes”, “Partial Yes”, “No”) were available to rate the extent of a SRMA’ adherence to the criterion. If an item was rated as “No”, it was regarded as a weakness for the SRMAs. If no information was provided, a “No” response was rated [10, 43]. For item 8 (describes the characteristics of included studies adequately), there was no clear indicators to distinguish “Partial Yes” (described all components but not in details) and “Yes” (described all in details), thus we contacted the principle investigator of AMSTAR for clarification but did not receive a response. Therefore, we rated all eligible SRMAs that described required components of characteristics as “Partial Yes” for item 8 to make a conservative evaluation. The enumerated counts considered “Yes” as 1 and “Partial Yes” as 0.5 while “No” was 0.

The lead author (XC), took charge of the assessment of methodological issues using the AMSTAR 2.0 tool. To ensure the quality of the process, at most 15 SRMAs were scheduled for assessment each day. A careful cross-checking process was utilised after the evaluation of all eligible SRMAs was completed. Then these records were double-checked by another researcher (LY). Any disagreements were discussed with two other methodologists (LFC and SD).

2.6 Data analysis

The baseline information of the SRMAs (e.g. author number, region) was qualitatively summarized. A bar chart was used to describe the adherence for each item with the

proportion of each response (“Yes”, “Partial Yes”, and “No”). For methodological issues, we focused on two separate aspects: a) the numbers of total weakness and critical weakness for each SRMA and b) the relative rank of all items and critical items of each SRMA.

In order to investigate potential measures to improve the methodological validity, we established a weighted least square regression for the *relative quality rank* against four predefined variables. The best SRMA thus has a rank of 1 (which serves as the anchor) and all lesser values are below this (range zero to 1). The predefined variables were: 1) region of affiliation of the first author (America, European, and Asia-pacific), 2) year of publication (≤ 2009 , 2010-2017, 2018-present), 3) number of authors (≤ 4 , 5-7, ≥ 8), and 4) use of reporting guideline (Yes, No). We categorized year of publication based on the year of release of AMSTAR (2009) and AMSTAR 2.0 (2017) [9, 10]. The number of authors were categorized by the quartiles. We did not use funding information as a dependent variable because it was already contained in the AMSTAR 2.0, which would break the *i.i.d* assumption of regression analysis [44]. Considering that SRMAs published in the same journal may have clustering on the methodological issues, a cluster robust-error variance was used in regression analysis [45].

A *post hoc* sensitivity analysis was employed under the consideration that the detection of publication bias (item 15) could be difficult or not defined for SRMAs with special types of meta-analyses (e.g. network meta-analysis) in the current period due to methodological constraints. Item 15 of AMSTAR 2.0 may not be well suited for these SRMAs. Therefore, we recomputed the *relative quality ranks* by removing SRMAs with special type meta-analysis and repeated the regression analysis to see if the results remained stable. The analyses were conducted using Stata 14.0/SE

(StataCorp, College Station, TX) with confidence level set at 0.95.

3. Results

3.1 Baseline characteristics

The literature search identified 1,630 records, of which 936 were identified as duplicates. We further excluded 104 records by screening the titles and abstracts (*appendix 1*). Of the remaining 590 records screened by full-text, 353 were SRMAs. Of which, we identified 163 that focused on healthcare interventions and were included in the analysis (Figure 1). The kappa statistic was 0.66 between the two raters. A detailed description of the screening process, list of included studies, and list of excluded studies including the reasons for exclusion are available in *appendix 1*.

Baseline characteristics of the included studies are presented in Table 1 and *appendix 3*. For the 163 eligible SRMAs, most of which were published in 2010 and after (90.80%), only 9.20% were published before 2010. In terms of region of the first author, 38.65% (n=63), 33.13% (n=54), and 28.22% (n=46) were from Asia-Pacific, America, and Europe, respectively. The median number of authors was 5 (interquartile range, IQR: 4 to 7); there were 37.42% SRMAs with 1 to 4 authors, 48.47% with 5 to 7 authors, and 14.11% with 8 or more authors.

The majority of the meta-analyses within these SRMAs were standard meta-analyses (n=157, 96.32%), and only 6 (3.68%) were special type meta-analyses. For SRMAs with special type meta-analyses, 5 were network meta-analysis and 1 was activation likelihood estimation. About half (48.47%) of the SRMAs referred to the use of a reporting guideline (e.g. PRISMA [46]).

The median number of included studies in each SRMA was 13 (IQR: 8 to 23),

most of them included more than 10 studies (n=107, 65.64%). In terms of funding information, 87 (53.37%) were supported by non-profit (government or institute) funding, 4 (2.45%) were supported by profit (industry) funding, 25 (15.34%) did not receive funding, and 47 (28.83%) did not report funding information.

3.2 Detailed methodological issues

The details of evaluation of the methodological issues for each SRMA are presented in *appendix 3*. Figure 2 presents the adherence to each methodological item.

Issue 1. Research questions and inclusion criteria

Most of the SRMAs (n=152, 93.25%; 95%CI: 88.32%, 96.19%) presented a clear research question and inclusion criteria in light of the population, intervention, comparison, and outcome (PICO). However, there were still some SRMAs that failed to clarify this (n=11, 6.75%; 95%CI: 3.81%, 11.68%), of which seven failed to provide a clear comparison, four did not clearly specify the population and one did not specify both intervention and comparison.

Issue 2. Protocol registration

Protocol registrations were identified in 28 (17.18%; 95%CI: 12.16%, 23.71%) SRMAs. There were 10 SRMAs that reported a protocol was developed in advance, but failed to provide it, and we decided to rate these as “No”. For the 28 with accessible protocol, eight failed to develop a meta-analysis plan, a plan for investigating source of heterogeneity, and justify any changes from the protocol.

Issue 3. Study designs for inclusion

There were only 12 (7.36%; 95%CI: 4.26%, 12.42%) SRMAs that reported the

reasons why certain study designs were included. Of which, one explained it in the abstract, and 11 explained it in the introduction or methods section. The majority (n=151, 92.64%; 95%CI: 87.58%, 95.74%) of the SRMAs failed to report the reason. For the 12 SRMAs, three stated why only randomized controlled trials were included, seven reported why both randomized controlled trials and non-randomized studies of interventions were included, while the rest explained why only non-randomized studies were included.

Issue 4*. Literature search (Critical Item)

In total, 19 (11.66%; 95%CI: 7.59%, 17.49%) of the SRMAs used a comprehensive literature search strategy that satisfies all the components required (rated as "Yes"). In addition, 125 (76.69%; 95%CI: 69.63%, 82.52%) of the SRMAs searched two or more databases, provided keywords or strategy, and justified any limitations, which met the minimal requirement (rated as "Partial Yes"). However, there were 11.66% (95%CI: 7.59%, 17.49%; n=19) of the SRMAs failed to use a comprehensive literature search (rated as "No"). The reasons were: 13 of them only searched one database and six did not provide keywords or search strategy.

Issue 5. Duplicate study selection, literature screen

There were 73.01% (95%CI: 65.72%, 79.24%; n=119) of the SRMAs that stated that the study selection process was conducted by two reviewers independently (rated as "Yes"). It is notable that, of the 119 meta-analyses, only eight provided objective evidence (e.g. kappa statistic) that the process involved two reviewers.

Issue 6. Duplicate data extraction

Similarly, 63.19% (95%CI: 55.56%, 70.21%; n=103) of the SRMAs stated that the data extraction process was conducted by two reviewers independently (rated as “Yes”). Again, only two of the 103 SRMAs provided objective evidence (kappa statistic) that the process involves two reviewers. More than one-third (n=60, 36.81%; 95%CI: 29.79%, 44.44%) of the SRMAs failed to perform study selection in duplicate.

Issue 7*. Study exclusion and justification (Critical Item)

Only 23 (14.11%; 95%CI: 9.59%, 20.28%;) SRMAs provided a list of excluded studies and justify the reasons of exclusions, and 3 (1.84%; 95%CI: 0.63%, 5.27%) provided a list of excluded studies but without the reasons for exclusion. The remaining 137 (84.05%; 95%CI: 77.66%, 88.88%) failed to provide a full list of excluded studies, although two of them list a part of the excluded studies.

Issue 8. Description of included studies

In total, 139 (85.28%; 95%CI: 79.03%, 89.90%) SRMAs provided a clear description of population (P), intervention (I), comparator (C), outcome (O), and study design (S) of the included studies. However, there were 24 (14.72%; 95%CI: 10.1%, 20.97%) SRMAs that failed to describe all of these components. In details, 13 out of 24 SRMAs failed to specify the study design of the included studies, four did not provide any description on PICOS, four did not describe the comparators, one did not describe the outcomes, and two did not describe at least two of the component (IC: 1, ICO: 1).

Issue 9*. Risk of bias assessment

There were 89 (54.60%; 95%CI: 46.94%, 62.05%) SRMAs that adequately used a risk of bias assessment for all of the important biases. In addition, 35 (21.47%; 95%CI: 15.86%, 28.39%) of the SRMAs assessed part of the important biases (Randomized controlled trial: concealed allocation and blinding; Non-randomized studies of interventions: confounding and selection bias), which met the minimal requirement. The remaining 39 (23.93%; 95%CI: 18.03%, 31.03%) failed to achieve minimal requirements, of which, five did not report the assessment results, one did not report which tool was used, and 33 did not assess the risk of bias.

Issue 10. Report on the sources of funding for included studies

The majority of the SRMAs (n=154, 94.48%; 95%CI: 89.84%, 97.97%) failed to report the funding information of included studies, and only 9 (5.52%; 95%CI: 2.93%, 10.16%) reported this item.

Issue 11*. Methods for statistical combination

For the statistical methods of combination, we identified 19 (11.66%; 7.57%, 17.49%) SRMAs with methodological issues to pool the data. The main problem was that most of them (n=15) incorrectly combined different types of studies together (e.g. cohort and cross-sectional study), of which, three also had other problems, for example, did not consider confounding and heterogeneity. In addition, two did not report the method of how the data were synthesized; one used fixed-effect model and did not consider heterogeneity; and one did not report how adjustments for confounding were handled.

Issue 12. Assess potential impact of risk of bias on the results

Three methods out of those available [47] were used by the authors to incorporate risk of bias into the results in 20.86% (95%CI: 15.33%, 27.73%) of SRMAs. These included, stratification (n=22), meta-regression (n=3), and included only low risk bias studies (n=9). However, the remaining majority of SRMAs (n=129, 79.14%; 95%CI: 72.27%, 84.67%) failed to assess the potential impact of risk bias on the results.

Issue 13*. Results interpretation with risk of bias

Again, most of the SRMAs (n=130, 79.75%; 95%CI: 72.60%, 85.47%) did not discuss risk of bias with the results interpretation notwithstanding if bias was incorporated into results or not. We documented 33 (20.25%; 95%CI: 14.80%, 27.07%) SRMAs did consider risk of bias in results interpretation of which 15 were those that had adjusted results for bias as reported above. Thus, most of these SRMAs failed to assess the potential impact of risk bias on the results.

Issue 14. Exploring and explanation of heterogeneity

There were 107 (65.64%; 95%CI: 58.06%, 72.50%) of the SRMAs that had a low between study heterogeneity or had some heterogeneity and attempted to explore the source of heterogeneity and discussed the potential impact on the conclusions. There were 56 SRMAs (34.36%; 95%CI: 27.50%, 41.94%) had some heterogeneity but did not explore the source.

Issue 15*. Investigation and discussion of publication bias

There were 76 (46.63%; 95%CI: 39.14%, 54.28%) SRMAs that investigated publication bias and discussed the potential influence on the results (rated as “Yes”).

In addition, 23 (14.11%; 95%CI: 9.59%, 20.28%) investigated publication bias but failed to discuss the potential influence (rated as “Partial Yes”). As much as 64 (39.26%; 95%CI: 32.09%, 46.92%) did not include investigation of publication bias (rated as “No”). Amongst the 64, 58 did not detect publication bias, and six did not provide results of publication bias. The reasons were recorded by nine SRMAs that the number of included studies were too small to assess publication bias.

Issue 16. Report sources of conflict of interest

Source of conflict of interest was reported by the majority of the SRMAs (n=158, 96.93%; 95%CI: 93.02%, 98.68%).

3.3 Rating of each issue and global confidence

Figure 3 presents the ranking for each item in light of the proportion of number of “No”. The majority of the SRMAs have a poor validity on protocol registration (item 2), study designs for inclusion (item 3), study exclusion and justification (item 7), report on the sources of funding for included studies (item 10), assess potential impact of risk of bias on the results (item 12), results interpretation with risk of bias (item 13); Of these, two of them (item 7 and 13) were critical important domains. In addition, about one-fourth to two-fifth of the SRMAs have a poor validity on study selection (item 5), data extraction (item 6), risk of bias assessment (item 9), heterogeneity exploring and explanation (item 14), and investigation and discussion of publication bias (item 15); again, two of them (item 9 and 15) were critical important domains. For the remaining five methodological items (item 1, 4, 8, 11, and 16), the validity was well-quantified by most of the SRMAs.

Figure 4 presents the distribution of number of total weakness and critical

weakness for these SRMAs. The median number of total and critical weakness were 7 (Inter Quartile Range, IQR: 6-9) and 2 (IQR: 2-3).

For total items, the best SRMA had a safeguard count of 12.5 and thus was regarded as the anchor for relative ranks. The median relative rank was 0.64, with the first quartile as 0.52 and the third quartile as 0.72 (Figure 4). This indicated that the top quartile SRMAs had up to a third (0-28%) of methodological safeguards missing. For six critical items, the best SRMA had a count of 6 meaning that all the six critical items were well adhered to. The median relative rank was 0.5, with the first quartile as 0.33 and the third quartile as 0.58 (Figure 4). This indicated that the top quartile SRMAs had up to almost half (0-42%) methodological safeguards missing.

3.4 Regression analysis

Our regression analysis suggested that, for total relative ranks, studies with first author from the Europe (estimated $\beta=0.08$; 95%CI: 0.02, 0.14; $p=0.013$), published more recently (e.g. 2010 to 2017 vs. 2009 and before, estimated $\beta=0.16$; 95%CI: 0.08, 0.24; $p=0.002$), and involves more authors (≥ 8 vs. ≤ 4 ; estimated $\beta=0.06$; 95%CI: 0.01, 0.11; $p=0.026$) were associated with higher relative quality ranks. These associations were similar for critical items: estimated β were 0.09 (95%CI: 0.03, 0.14; $p=0.007$) for studies with first author from the Europe, 0.14 (95%CI: 0.005, 0.27; $p=0.043$) for studies published between 2010 to 2017. While the association were no longer observed for studies with 8 or more authors (estimated $\beta=0.01$; 95%CI: -0.08, 0.09; $p=0.865$), Table 2. Sensitivity analysis, after the exclusion six SRMAs with special type meta-analysis, suggested the associations were mostly stable (Table 3), except for the publication data (estimated $\beta=0.14$; 95%CI: -0.004, 0.28; $p=0.056$).

4. Discussion

In this review, we comprehensively evaluated the methodological shortcomings of published SRMAs of healthcare intervention in the field of sleep medicine. Our results suggested that most of these SRMAs have 7 or more methodological issues and 2 of which, on average, were critical issues. These issues mainly were with study inclusion and exclusion and risk of bias assessment and interpretation. By summarizing the relative ranks, we found that the majority of the SRMAs were of much lower quality than the best SRMA, and this is more serious for the six critical items. And even for the best one, there were still 3 methodological items that were not well adhered to.

Results from the regression analysis suggest that SRMAs published in recent years tend to have higher quality ranks. This finding indicates that the methodological quality of SRMAs improved over the years. We further observed that SRMAs with the first author from Europe and the Asia-Pacific region tend to have higher quality ranks, especially for critical items. However, we did not observe a stable improvement on the methodological quality over the years for critical items.

We found that use of a reporting guideline did not help to increase the methodological validity of SRMAs for both global quality and critical items. This could be expected because reporting guidelines were primarily designed to help authors remember all items that need to be reported, rather than to conduct a SRMA [48]. However, some of the methodological issues were highly correlated with reporting problems, for example description of the inclusion criteria, study section, baseline characteristic, and conflict of interests. The role of reporting guideline on the methodological validity should be further investigated through well-designed experimental studies.

The findings of the current study concurred with reviews from other fields (e.g. urology, bariatrics, general surgery) [49-54]. For example, Corbyons et al conducted a survey on SRMAs published in urology and their findings suggested that the methodological quality of these studies was suboptimal [49]; Storman et al found that 99% of the published systematic reviews/meta-analyses in bariatric surgery were critically low on methodological quality [50]. These findings revealed that, many SRMAs may have serious methodological issues.

In this review, we did not use the rating scheme recommended by AMSTAR 2.0 to rate the methodology confidence of eligible, instead, the relative rank method was utilized. In addition to the reason we mentioned earlier (i.e. subjective judgment), the rating scheme of this instrument is not sensitive to distinguish the confidence of SRMAs with critical low quality – all SRMAs with two or more critical issues were rated as critical low. Indeed, a SRMAs with 2 critical issues might be more credible than one with 3 critical issues. The relative rank method provides a better solution to rate the confidence and can avoid such problems.

We did not consider protocol registration as a critical item. Our previous study suggested that developing a protocol in advance although of benefit to improve reporting, may not represent the methodological quality of SRMAs well [55]. Waugh [42] has pointed out six issues with such registration 1) confidentiality of research ideas, 2) deterrence of others from similar research that may be of higher quality, 3) no clear relevance of information requested in PROSPERO, 4) no clear benefit in terms of precedence in registration, 5) no clear mandate from the academic community, and 5) finally the cost of time spent versus effectiveness of the process. This whole concept needs revisiting to assess its fitness for purpose.

Based on current findings and our experiences, we proposed some

recommendations about dos and don'ts of SRMAs beyond the AMSTAR 2.0 instrument: 1) when starting a SRMA, it is helpful to design and conduct it according to a well-designed instrument (e.g. AMSTAR 2.0 [10]); 2) when including both observational and experimental studies in a SRMA, it is not recommended to incorporate data of these two types of studies together as the former would introduce risk of reverse causality; 3) it is highly recommended to explain the selection of effect estimator (e.g. odds ratio, risk ratio) to measure the effects in the meta-analysis and how the effect estimators were dealt with when difference estimators were used by these studies; 4) it is recommended to use two or more weighting methods as sensitivity analysis when the effect was small but statistically significant; 5) if applicable, a dose-response gradient should be investigated; 6) when measuring publication bias, P-value driven methods (e.g. Egger's test, rank correlation test [56]) are discouraged as these are dependent on the number of studies included in a meta-analysis, instead non-P-value driven methods (e.g. LFK index [57]) should be used.

In this review, we employed a comprehensive evaluation of SRMAs in sleep medicine, to the best of our knowledge, this is the first study that focuses on the methodological issues of SRMAs in this field. We collected nearly all published SRMAs of healthcare intervention in the field of sleep medicine, therefore, our findings have a high level of representativeness. We acknowledge that our review had some limitations. In this study, the literature search was based on 19 academic journals of sleep medicine that there was no doubt that some related studies published in other journals (e.g. general journals) were not included, which may bring some selection bias on the results. Previous study had document that the methodology quality of meta-analysis may differs from general journals and specialist journals [58].

However, it is difficult to identify meta-analyses on this topic (sleep) from other journals. Further, our study put focus on the methodological validity of SRMAs, while neglected the importance of the quality of individual studies included in these SRMAs. The quality of these original studies is also very important. In addition, as we mentioned earlier, some methodology tips may not well reflected and covered in the AMSTAR 2.0, which may affect the validity of current survey. Moreover, the screen and assessment processes, although were strict, may still at risk of systematic errors since both were of somewhat subjective. These limitations should be highlighted and merit attention in the results interpretation.

5. Conclusions

In conclusion, the methodological validity of SRMAs of healthcare intervention was suboptimal when measured by AMSTAR 2.0 in the field of sleep medicine. Although the it has improved over time, methodological confidence was lacking for most of these SRMAs. Based on current findings, we advocate a critical evaluation on the methodological validity of a SRMA before it can be used as clinical evidence.

Practice Points

1. Most of the systematic reviews have 7 or more methodological issues and 2 of which, on average, were critical issues. These issues can have serious impact on the credibility of the evidence.
2. Relative ranks are likely a better quality-assessment scheme than the absolute judgments commonly used.

Research Agenda

Future studies should:

1. Undertake a critical evaluation of the methodological validity of a systematic review before using it as clinical evidence.
2. More focus should be put on the methodological validity of the systematic reviews and meta-analyses rather than simple checklists followed by inexperienced researchers.
3. Rigorous guidelines for the methodology for different types of meta-analysis are needed to help systematic reviews' authors to improve the methodological quality of what they create.

Authors' contributions: XC conceived and designed the study, conducted the literature search, analyzed the data, plotted the figures and tables, and drafted the manuscript; XC and LY contributed to the quality assessment and data collection; LFK and SD provided statistical guidance and edited the draft; LFK, SD, JK, LS provided further comments for the manuscript. All authors approved the final version.

Journal Pre-proof

Reference

*[1] Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is "quality of evidence" and why is it important to clinicians? *BMJ* 2008; 336(7651):995-8.

*[2] Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nature* 2018;555(7695):175-182.

*[3] Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126(5):376-80.

[4] Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux PJ, Prasad K, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA* 2014;312(2):171-9.

[5] Noordzij M, Hooft L, Dekker FW, Zoccali C, Jager KJ. Systematic reviews and meta-analyses: when they are useful and when to be careful. *Kidney Int* 2009;76(11):1130-6.

[6] Stone J, Gurunathan U, Glass K, Munn Z, Tugwell P, Doi SAR. Stratification by quality induced selection bias in a meta-analysis of clinical trials. *J Clin Epidemiol* 2019; 107:51-59.

[7] Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987;316(8):450-5.

[8] Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44:1271-8.

[9] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; 7:10.

*[10] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2:

a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017; 358: j4008.

[11]Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, et al. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMJ* 2000; 320(7234):537-40.

[12]Xu C, Liu Y, Jia PL, Li L, Liu TZ, Cheng LL, et al. The methodological quality of dose-response meta-analyses needed substantial improvement: a cross-sectional survey and proposed recommendations. *J Clin Epidemiol* 2019; 107:1-11.

[13]Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur Heart J*;32(12):1484-92.

[14]Liu TZ, Xu C, Rota M, Cai H, Zhang C, Shi MJ, et al. Sleep duration and risk of all-cause mortality: A flexible, non-linear, meta-regression of 40 prospective cohort studies. *Sleep Med Rev* 2017; 32:28-36.

[15]Fatima Y, Doi SA, Mamun AA. Longitudinal impact of sleep on overweight and obesity in children and adolescents: a systematic review and bias-adjusted meta-analysis. *Obes Rev* 2015; 16:137–49.

[16] Gu AP, Gu CN, Ahmed AT, et al. Sham surgical procedures for pain intervention result in significant improvements in pain: systematic-review and meta-analysis: Metaepidemiologic research requires reporting requirements. *J Clin Epidemiol*. 2017; 87:108.

[17] Puljak L, Makaric ZL, Buljan I, Pieper D. What is a meta-epidemiological study? Analysis of published literature indicated heterogeneous study designs and definitions. *J Comp Eff Res*. 2020;9(7):497-508.

[18] Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical

methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med.* 2002;21(11):1513-1524.

[19]Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA* 2014;311(4):405-11.

*[20] Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org. Accessed on 7th Oct 2018.

[21]Shadish WR, Lecy JD. The meta-analytic big bang. *Res Synth Methods* 2015; 6(3):246-64.

[22]Villar J, Carroli G, Belizán JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995; 345(8952):772-6.

[23]Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. *Contemp Clin Trials* 2015; 45(Pt A):130-8.

[24]Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model. *Contemp Clin Trials* 2015; 45(Pt A):123-9.

*[25] Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018;18(1):143.

[26]Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20:37-46.

[27]McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22(3):276-82.

[28]Doi SAR, Furuya-Kanamori L, Toft E, Musa OAH, Islam N, Clark J, et al.

Metformin in pregnancy to avert gestational diabetes in women at high risk: Meta-analysis of randomized controlled trials. *Obes Rev* 2020;21(1): e12964.

[29]McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, and the PRISMA-DTA Group, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018; 319(4):388-396.

[30]Bauckneht M, Chincarini A, De Carli F, et al. Presynaptic dopaminergic neuroimaging in REM sleep behavior disorder: A systematic review and meta-analysis. *Sleep Med Rev* 2018; 41:266-274.

[31]Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013;159(2):130-7.

[32]Ballesio A, Aquino MRJV, Feige B, et al. The effectiveness of behavioural and cognitive behavioural therapies for insomnia on depressive and fatigue symptoms: A systematic review and network meta-analysis. *Sleep Med Rev* 2018; 37:114-129.

[33]Eickhoff SB, Bzdok D, Laird AR, Kurth F, Fox PT. Activation likelihood estimation meta-analysis revisited. *Neuroimage* 2012;59(3):2349-61.

[34]Javaheipour N, Shahdipour N, Noori K, Zarei M, Camilleri JA, Laird AR, et al. Functional brain alterations in acute sleep deprivation: An activation likelihood estimation meta-analysis. *Sleep Med Rev* 2019; 46:64-73.

[35]Barendregt JJ, Doi SA, Lee YY, Norman RE, Vos T. Meta-analysis of prevalence. *J Epidemiol Community Health* 2013; 67(11):974-8.

[36]Furuya-Kanamori L, Bell KJL, Clark J, Glasziou P, Doi SAR. Prevalence of Differentiated Thyroid Cancer in Autopsy Studies Over Six Decades: A Meta-Analysis. *J Clin Oncol* 2016; 34(30):3672-3679.

[37]Galland BC, Taylor BJ, Elder DE, Herbison P. Normal sleep patterns in infants

and children: a systematic review of observational studies. *Sleep Med Rev*;16(3):213-22.

[38] Dewald JF, Meijer AM, Oort FJ, Kerkhof GA, Bögels SM. The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. *Sleep Med Rev* 2010;14(3):179-89.

[39] Manning AK, LaValley M, Liu CT, Rice K, An P, Liu Y, et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP \times environment regression coefficients. *Genet Epidemiol* 2011; 35(1):11-8.

[40] Capittini C, De Silvestri A, Terzaghi M, Scotti V, Rebuffi C, Pasi A, et al. Correlation between HLA-DQB1*06:02 and narcolepsy with and without cataplexy: approving a safe and sensitive genetic test in four major ethnic groups. A systematic meta-analysis. *Sleep Med* 2018; 52:150-157.

*[41] Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;62(10):1013-20.

*[42] Waugh N. Misgivings about PROSPERO in solving the problem of duplicate systematic reviews. *BMJ* 2013; 347: f5496.

[43] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-12.

[44] Sun J, Tong X, He X. Regression analysis of panel count data with dependent observation times. *Biometrics* 2007;63(4):1053-9.

[45] Xu C, Doi SA. The robust-error meta-regression method for dose-response meta-analysis. *International Journal of Evidence-Based Healthcare* 2018; 16(3):138-144.

*[46] Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine* 2009; 6(7): e1000097.

*[47] Stone JC, Glass K, Munn Z, Tugwell P, Doi SAR. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. *J Clin Epidemiol* 2020; 117:36-45.

[48] Moher D, Weeks L, Ocampo M, Seely D, Sampson M, Altman DG, et al. Describing reporting guidelines for health research: a systematic review. *J Clin Epidemiol* 2011; 64(7):718-42.

[49] Corbyons K, Han J, Neuberger MM, Dahm P. Methodological Quality of Systematic Reviews Published in the Urological Literature from 1998 to 2012. *J Urol* 2015;194(5):1374-9.

[50] Storman M, Storman D, Jasinska KW, Swierz MJ, Bala MM. The quality of systematic reviews/meta-analyses published in the field of bariatrics: A cross-sectional systematic survey using AMSTAR 2 and ROBIS. *Obes Rev* 2020. doi: 10.1111/obr.12994.

[51] Dixon E, Hameed M, Sutherland F, Cook DJ, Doig C. Evaluating meta-analyses in the general surgical literature: a critical appraisal. *Ann Surg* 2005;241(3):450-9.

[52] Gómez-García F, Ruano J, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, et al. Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool. *J Clin Epidemiol* 2017; 92:79-88.

[53] Greenberg DR, Richardson MT, Tijerina JD, Bass MB, Eisenberg ML. The Quality of Systematic Reviews and Meta-Analyses in Erectile Dysfunction Treatment and Management Published in the Sexual Medicine Literature. *J Sex Med*

2019;16(3):394-401.

[54] Kattan S, Lee SM, Kohli MR, Setzer FC, Karabucak B. Methodological Quality Assessment of Meta-analyses in Endodontics. *J Endod* 2018;44(1):22-31.

[55] Xu C, Cheng LL, Liu Y, Jia PL, Gao MY, Zhang C. Protocol registration or development may benefit the design, conduct and reporting of dose-response meta-analysis: empirical evidence from a literature survey. *BMC Med Res Methodol* 2019;19(1):78.

[56] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-634.

[57] Furuya-Kanamori L, Xu C, Lin L, Doan T, Chu H, Thalib L, et al. P value-driven methods were underpowered to detect publication bias: analysis of Cochrane review meta-analyses. *J Clin Epidemiol* 2020; 118:86-92.

[58] Xu C, Liu Y, Zhang C, Kwong JSW, Zhou JG, Ge L, et al. An overview on the methodological and reporting quality of dose-response meta-analysis on cancer prevention. *J Cancer Res Clin Oncol*. 2019;145(5):1201-1211.

Figure Legends

Figure 1. The literature screen process.

Figure 2. Methodology adherence of eligible systematic reviews.

Figure 3. Rating for the proportion of weakness of each item.

Figure 4. Histogram for the distribution of the number of total and critical weakness as well as the relative ranks of the methodological quality (Y-axis is the frequency).

Table 1. Baseline characteristics of included systematic reviews.

Baseline characteristics	All publication (N=163)
Year	
~2009	15 (9.20%)
2010~2017	88 (53.99%)
2018~present	60 (36.81%)
Region of first author	
America	54 (33.13%)
Asia-Pacific	63 (38.65%)
European	46 (28.22%)
Author number [median (Q1, Q3)]	
< 5	61 (37.42%)
5~7	79 (48.47%)
>=8	23 (14.11%)
Type of meta-analysis	
Generic meta-analysis	157 (96.32%)
Special type meta-analysis	6 (3.68%)
Type of study included	
RCTs	90 (55.21%)
RCTs and NRSI	47 (28.83%)
NRSI	24 (14.72%)
Not reported	2 (1.23%)
Use of reporting guidance	
Yes	79 (48.47%)
No	84 (51.53%)
Protocol	
Yes, and accessible	28 (17.18%)
Yes, but not provided	10 (6.13%)
No	125 (76.69%)
Studies eligible for meta-analysis [median (Q1, Q3)]	
< 10	56 (34.36%)
10-29	76 (46.63%)
>=30	31 (19.02%)
Funding	
Non-profit funding	87 (53.37%)
Profit funding	4 (2.45%)
No funding	25 (15.34%)
Not reported	47 (28.83%)

NRSI: non-randomized study of intervention (e.g. non-randomized controlled, pre-post study).

Table 2. Regression analysis for relative quality ranks of the methodology to four pre-defined variables.

Variables	Systematic reviews with standard meta-analysis			
	<i>Total items</i>	<i>P</i>	<i>Critical items</i>	<i>P</i>
Region				
America	Reference		Reference	
Asia-pacific	0.06 (-0.01, 0.13)	0.075	0.09 (0.03, 0.14)	0.007
European	0.08 (0.02, 0.14)	0.013	0.12 (0.07, 0.18)	0.001
Year of publication				
2009 and before	Reference		Reference	
2010 to 2017	0.16 (0.08, 0.24)	0.002	0.14 (0.005, 0.27)	0.043
2018 to present	0.12 (0.05, 0.19)	0.004	0.08 (-0.07, 0.23)	0.245
Number of authors				
<=4	Reference		Reference	
5 to 7	0.03 (-0.02, 0.07)	0.217	0.00 (-0.05, 0.05)	0.980
>=8	0.06 (0.01, 0.11)	0.026	0.01 (-0.08, 0.09)	0.865
Use of reporting guidance				
No	Reference		Reference	
Yes	0.05 (-0.005, 0.10)	0.071	0.04 (-0.03, 0.11)	0.246

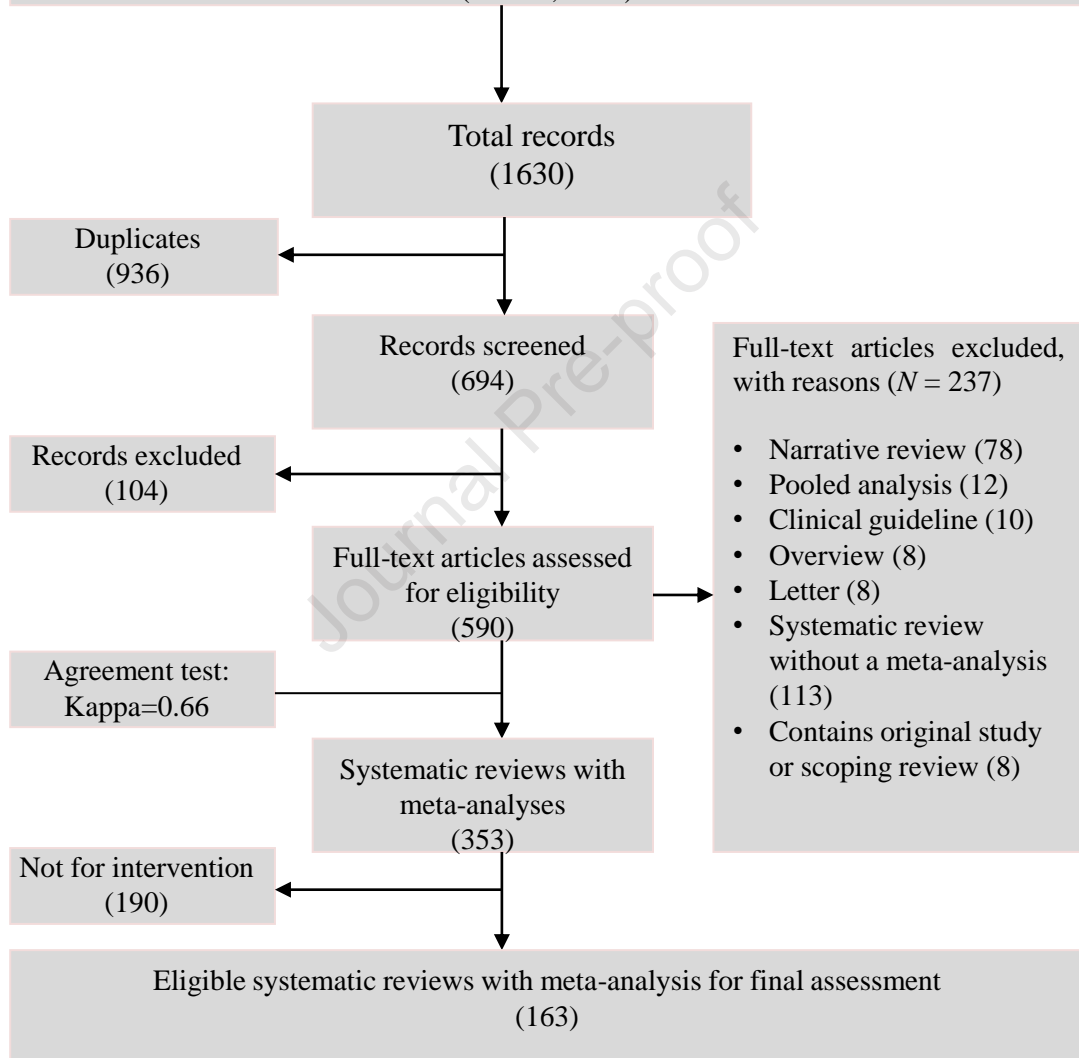
Note: The relative quality ranks were normally distributed. Joint P for Skewness and Kurtosis was 0.3245 for ranks of total items, and 0.9146 for ranks of critical items.

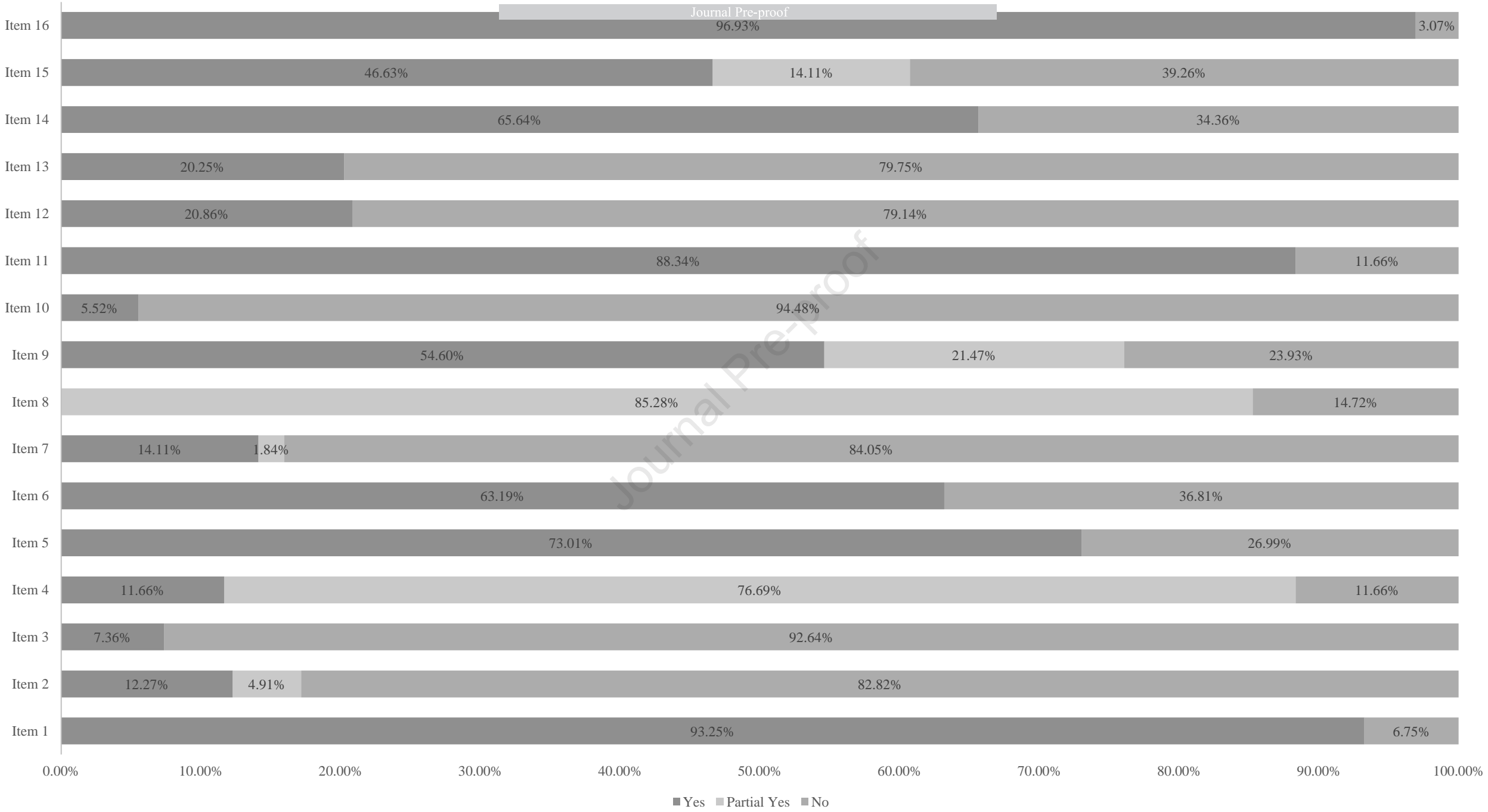
Table 3. Sensitivity analysis of systematic reviews with meta-analysis after excluding 6 special type meta-analyses.

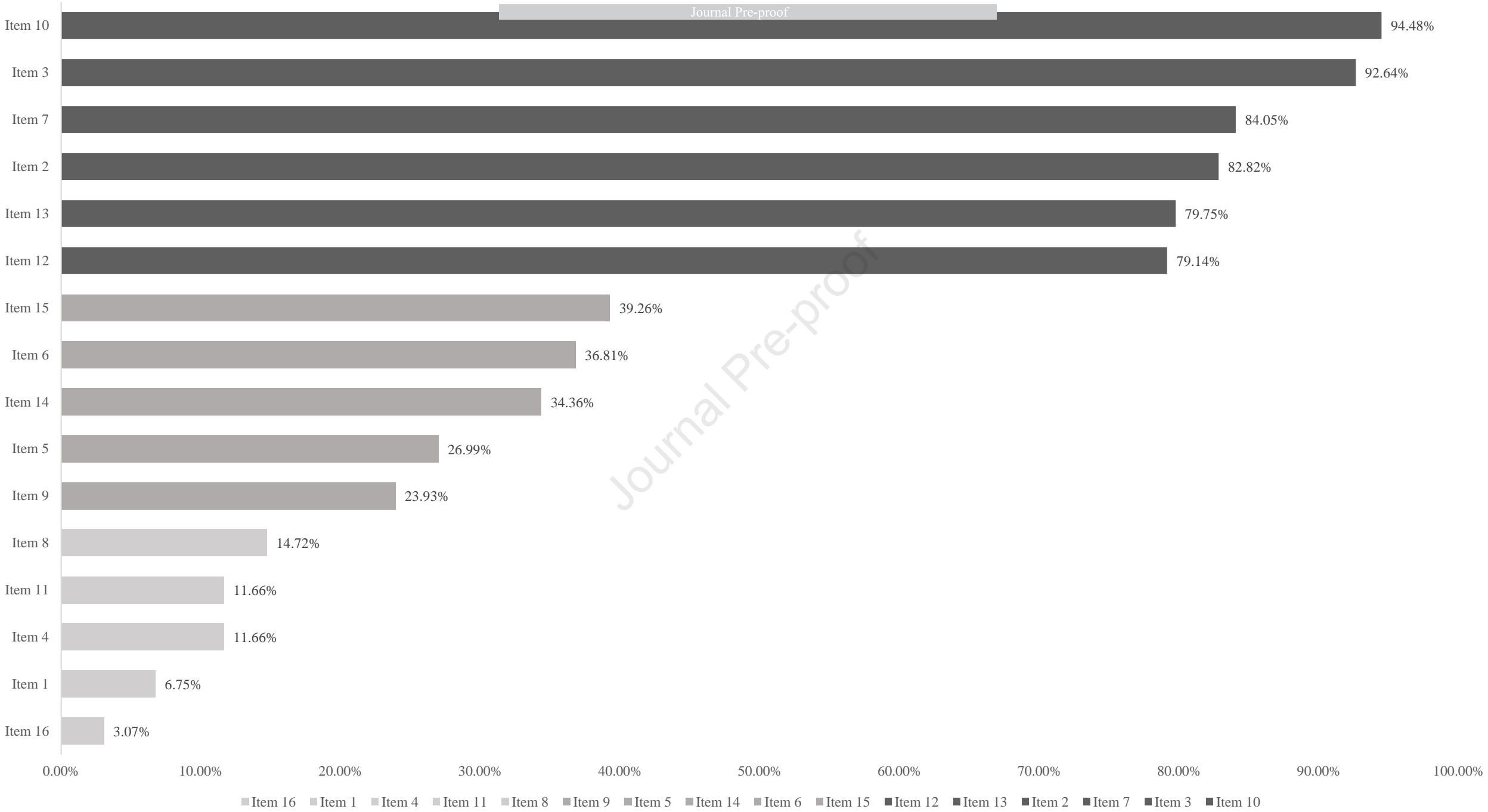
Variables	Primary analysis				Sensitivity analysis			
	Total items	P	Critical items	P	Total items	P	Critical items	P
Region								
America	Reference		Reference		Reference		Reference	
Asia-pacific	0.06 (-0.01, 0.13)	0.075	0.09 (0.03, 0.14)	0.007	0.07 (0.004, 0.13)	0.04	0.08 (0.04, 0.12)	0.003
European	0.08 (0.02, 0.14)	0.013	0.12 (0.07, 0.18)	0.001	0.08 (0.04, 0.13)	0.002	0.12 (0.07, 0.17)	0.001
Year of publication								
2009 and before	Reference		Reference		Reference		Reference	
2010 to 2017	0.16 (0.08, 0.24)	0.002	0.14 (0.005, 0.27)	0.043	0.14 (0.05, 0.23)	0.008	0.14 (-0.004, 0.28)	0.056
2018 to present	0.12 (0.05, 0.19)	0.004	0.08 (-0.07, 0.23)	0.245	0.10 (0.02, 0.18)	0.025	0.08 (-0.08, 0.24)	0.276
Number of authors								
<=4	Reference		Reference		Reference		Reference	
5 to 7	0.03 (-0.02, 0.07)	0.217	0.00 (-0.05, 0.05)	0.980	0.02 (-0.02, 0.06)	0.259	0.00 (-0.05, 0.05)	0.930
>=8	0.06 (0.01, 0.11)	0.026	0.01 (0.08, 0.09)	0.865	0.05 (0.005, 0.10)	0.034	0.00 (-0.08, 0.08)	0.981
Use of reporting guidance								
No	Reference		Reference		Reference		Reference	
Yes	0.05 (-0.005, 0.10)	0.071	0.04 (-0.03, 0.11)	0.246	0.06 (-0.01, 0.12)	0.08	0.04 (-0.03, 0.11)	0.225

The relative quality ranks were normally distributed (joint P for Skewness and Kurtosis were 0.3909 and 0.9566 for ranks total items and critical items in sensitivity analysis).

PubMed, Medline, and Embase for 19 journals in sleep medicine: PubMed (833), Medline and Embase (997)
(22-Oct, 2019)



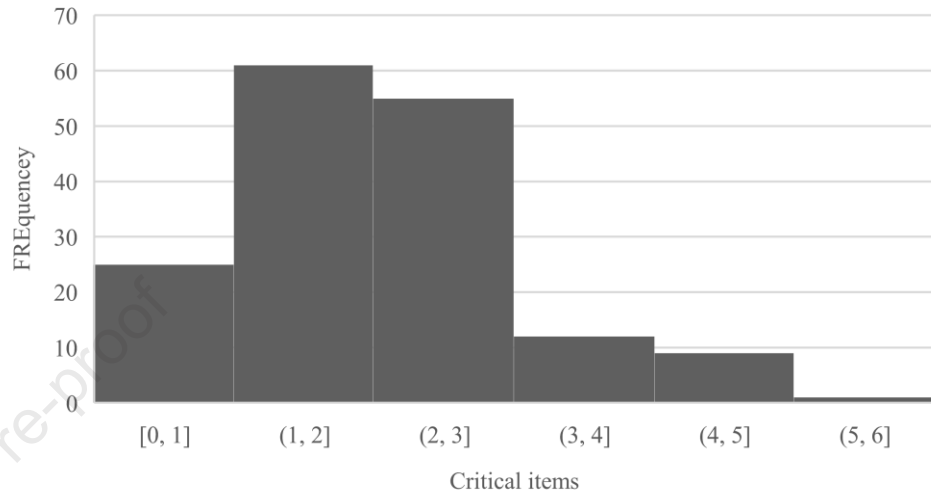
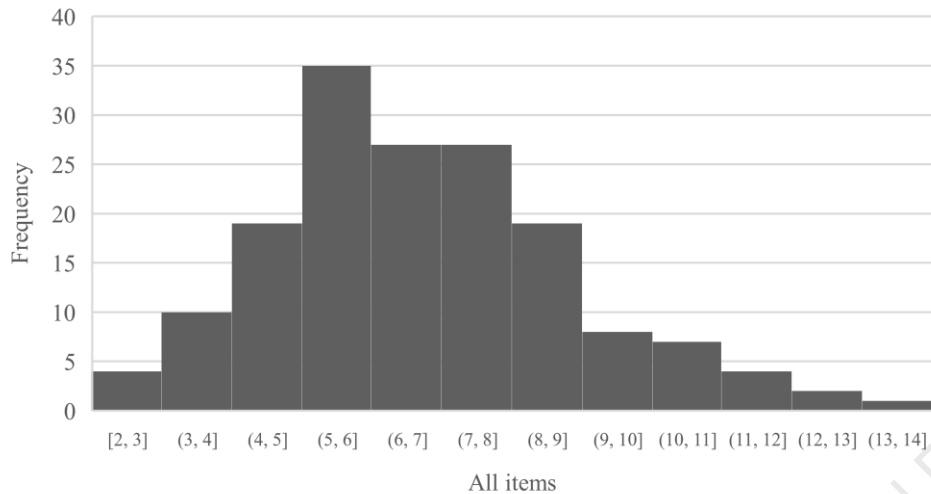




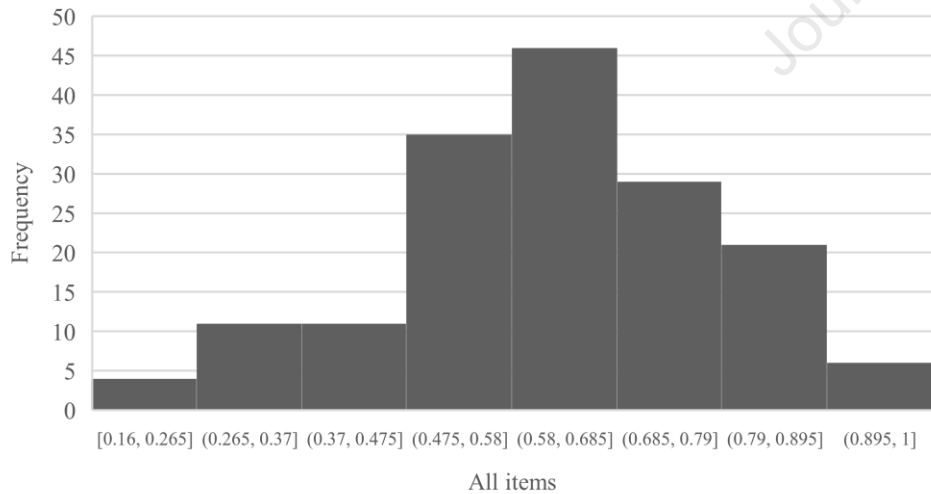
Number of “No”

Journal Pre-proof

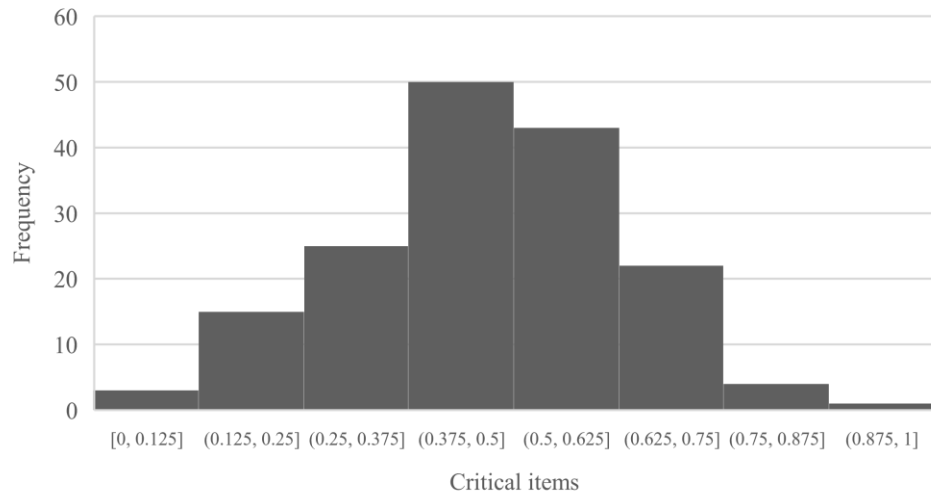
Number of “No”



Relative ranks



Relative ranks



18th Sep 2020

Michael V. Vitiello, PhD
Editor-in-Chief, Sleep Medicine Reviews

Re: Methodological issues of systematic reviews and meta-analyses in the field of sleep medicine: A meta-epidemiological study (SMRV-D-20-00042R1)

Many thanks for giving us the opportunity to respond to the reviewers' comments. Our specific responses are detailed below.

Reviewer #1:

I thank the authors for their responses and appreciate the effort that went into revisions of this work. There remain points to be clarified and issues resolved before the reader can make an informed decision as to the findings of this research.

I still have concerns about methodology part of this research. In response to my earlier concern, you clarified that this is an epidemiological survey on methodological quality of published meta-analytic reviews of healthcare intervention in the field of sleep medicine, where meta-analyses of RCT, NRCT, and "not reported" original studies were included. You further stated that "We conducted the study according to AMSTAR 2.0 where relevant items were applicable. There is no need for us, in principle, to follow AMSTAR when reporting this paper."

The clarifications are needed to the following items:

1. Re statement: "The current study does not aim to assess healthcare interventions, but rather a survey on methodological quality. Therefore, we only defined the "P" and "O". "P" was systematic reviews with meta-analysis of healthcare intervention, "O" was the methodological quality." AND "By going through the list, several studies were narrative reviews (References 97, 304, 347, 352, etc). What was your rationale to include them?"

a. Please clarify inconsistencies in inclusion/exclusion criteria- systematic reviews with meta-analyses, OR meta-analyses OR both. This decision has significant implication for your study searches, selection, appraisal, and results.

Response: We have revised the description of the inclusion criteria, now it says "We included systematic reviews with meta-analyses, or meta-analyses alone of healthcare intervention..."

b. Please provide definition of "survey on methodological quality" as it concerns methods used in epidemiological research applied in this study. This should also be reflected in the title and in manuscript's method section, as per AMSTAR 2.0.

Response: We have changed the term “survey on methodological quality” to “meta-epidemiological study”, and added the definition of “meta-epidemiological study”. This change is reflected in the title and methods section in the revised manuscript.

In a recent study (*J Comp Eff Res. 2020;9(7):497-508.*), the term “meta-epidemiological study” refers to a study type that aims to evaluate trends and patterns in the literature with the overarching goal of improving the design, methods and conduct of future research.

2. You indicated that you used The AMSTAR 2.0 instrument to evaluate the methodological issues of each study included in this research. The AMSTAR was developed to enable appraisal of systematic reviews of randomised and non-randomised studies of healthcare interventions.

a. Please provide rationale for using this framework to appraise narrative reviews that applied quantitative synthesis of multiple studies, and studies that included mixed data (i.e., longitudinal studies, cross-sectional., etc.).

Response: Narrative reviews often use qualitative summary or a meta-ethnographic technique to qualitatively synthesize the findings (<https://guides.library.uab.edu/c.php?g=63689&p=409774>). They do not involve a quantitative synthesis. In another word, a review with quantitative synthesis is either a systematic review with meta-analysis, or a pooled analysis (differs from meta-analysis, as it does not have a regular literature search). As we have clarified a pooled analysis was not included.

There were few SR/MAs that included mixed data such as longitudinal studies or cross-sectional one, while in the healthcare intervention part, they only use experimental study for the analysis (e.g. *sleep Breath. 2016;20(2):719-731.*). For longitudinal studies mentioned in the included SR/MAs, these were based on post-pre design, so they can be considered quasi-experimental (i.e. NRSI).

The different study designs do not impact the appropriateness of using AMSTAR 2.0. There were several items related to this point (e.g. item 3, item 11, item 12), while there is a very clear criteria to assess RCTs and NRSIs separately. For example, for item 11, both cohort and cross-section studies should adjust the potential confounding, and studies with different designs should not combine together. Longitudinal study is analytic design, while cross-sectional is observational design, so could not be combine together. As we reported in the results “*The main problem was that most of them (n=18) incorrectly combined different types of studies together (e.g. cohort and cross-sectional study)...*”

3. Some studies included in your reviews were systematic reviews with meta-analyses and some were meta-analyses. You also stated in response to my earlier comment (revision 1) that "References 97, 304, 347, and 352 included quantitative synthesis of multiple studies (i.e. meta-analyses)" and therefore were included. Please elaborate on how rating for the proportion of weakness of each item of the AMSSTAR 2.0 varied

based on each different study methods.

Response: Please see our response to Q1a. A systematic review may not include a meta-analysis, while a meta-analysis is always a type of systematic review. In our study, we included systematic reviews with meta-analyses or meta-analyses alone, which means they were all systematic reviews, and all contain meta-analyses. In essential, they are the same.

4. The revised version of the manuscript STILL includes numerous reviews within the 169 selected that do not concern healthcare interventions. Some of them are listed below:

1. Brockmann, P.E., P. Bertrand, and J.A. Castro-Rodriguez, Influence of asthma on sleep disordered breathing in children: A systematic review. *Sleep Medicine Reviews*, 2014. 18(5): p. 393-397.
2. Kendzerska, T., et al., Incident cardiovascular events and death in individuals with restless legs syndrome or periodic limb movements in sleep: A systematic review. *Sleep*, 2017. 40(3): p. zsx013.
3. Leong, R.L.F., et al., The effects of sleep on prospective memory: A systematic review and metaanalysis. *Sleep Medicine Reviews*, 2019. 47: p. 18-27.
4. Ma, N., et al., How acute total sleep loss affects the attending brain: A meta-analysis of neuroimaging studies. *Sleep*, 2015. 38(2): p. 233-240.
5. Morgenthaler, T.I., et al., High school start times and the impact on high school students: What we know, and what we hope to learn. *Journal of Clinical Sleep Medicine*, 2016. 12(12): p. 1681-1689.
6. Pires, G.N., et al., Effects of acute sleep deprivation on state anxiety levels: a systematic review and meta-analysis. *Sleep Medicine*, 2016. 24: p. 109-118.
7. Schripf, M., et al., The effect of sleep deprivation on pain perception in healthy subjects: A metaanalysis. *Sleep Medicine*, 2015. 16(11): p. 1313-1320.

Please reflect on the probability of a random error. (i.e., mistakes).

Response: Thanks for this comment. We have carefully checked all the 169 SR/MAs, and removed those (n=6) which did not include healthcare intervention.

From the 7 references pointed out by the reviewer, 4 were removed, while the other 3 included healthcare intervention (references 4, 6, 7). We also removed other 2 studies (*Bowers JM et al Sleep Health 3 (2017) 423–431; Sassani A 2004;27(3):453-8.*) as they did not contain healthcare interventions.

We re-analyze the data and no substantial changes were observed.

5. Table 1 refers to descriptions of the different types of meta-analyses and related citations. It is not clear why this table was presented in the main body of the manuscript as most types of presented meta-analyses are not relevant to your research question. Also, citations presented in your examples are not found within your selected and analysed 169

articles. Please elaborate on the rationale to present this table in the main text.

Response: We presented table 1 in the main body because we believe it will be useful for readers to understand different types of meta-analyses. According to your suggestion, we now put it into the supplementary file (Table S1).

For the citations in the examples, we did not use the 169 articles, as you can see in the appendix data file that the majority of the 169 articles were generic meta-analysis and only 6 were special type meta-analysis (5 network meta-analyses, and 1 coordinate-based meta-analysis), which we were not suitable to show the different types of meta-analysis.

6. The process of study selection, analyses, visual presentation of the results, as well as citations of included studies can only be found in the supplementary material. This creates difficulty to the readership to follow research analyses and results reported by researchers. Transparent reporting, linkage of information that made a basis for your reporting, included but not limiting raw data that made aggregation in Table 2, Table S1 (baseline characteristics, standard meta-analyses OR special type meta-analyses, study origin, reporting guidelines, funding information, etc.) should be available. For example, when a reader faces statement like this one: "There were 90 (53.25%) SRMAs that adequately used a risk of bias assessment for all of the important biases.... The remaining 42 (24.85%) failed to achieve minimal requirements, of which, 5 did not report the assessment results", he/she would like to connect these statements to specific studies, which is currently not possible. Not only reporting numbers do not make any application, but also the extent of risk of bias is unknown. It is anticipated that included studies are cited in the main body of the manuscript and are linked to reporting of the results.

Response: We now present the raw data of study baseline characteristics in the supplementary file (appendix 3). In addition, the process of study selection (Figure 1) and sensitivity analyses (Table 3) are now in the main text.

The included studies were not cited in the main body due to the restriction (*A maximum of 100 references requested by the journal's policy*) on the number of references.

7. Your limitation section should be revisited. Qualitative studies, non-systematic reviews, publications within other but major journals, and systematic reviews of observational data were falling outside the score of your research, and therefore it is unclear why you reported these features within your limitation section. The focus of your limitation section should be on the credibility of your research, search criteria, validation of keywords for searches, omission of attention to the risk of bias assessment of the individual studies that were included in each systematic review with meta-analyses, type of intervention, etc., and overall "methodological survey" design and conduct, to let the reader be informed how well your results, including pooled estimates, were protected against misleading results.

Response: We have revised the limitation section accordingly, we now focus on the credibility and validity of our study.

Minor issues:

1. Clarification is needed on what "inadequate methodological validity" refers to in the conclusion section. How adequacy was defined?

Response: We have revised the conclusion, now it says "*In conclusion, the methodological validity of SRMAs of healthcare intervention was suboptimal when measured by AMSTAR 2.0 in the field of sleep medicine*".

2. Clarification is needed on the statement in the conclusion, which does not seem to be supported by results of your research: "Finally, perhaps registration of systematic review should be removed from AMSTAR 2.0 as this item does not impact on the methodological quality of systematic reviews". Did you perform sensitivity analyses? Please provide rationale to support this conclusion.

Response: We have removed this sentence.

3. Incomplete citation:

Hornyak, M., et al., Efficacy and safety of dopamine agonists in restless legs syndrome. *Sleep Medicine*.

Response: We revised this (in appendix 1) and the full citation is "*Hornyak, M., et al., Efficacy and safety of dopamine agonists in restless legs syndrome. Sleep Medicine. Sleep Med. 2012;13(3):228-236.*". We also checked other citations one by one in the appendix file.

4. Inconsistency in writing style (i.e., "Amongst the 67, sixty-two did not detect publication bias, and five did not provide results of publication bias". Generally, numbers below 10 are written in words, 10 and above in numbers.

Response: It has been revised throughout the manuscript.

In summary, while the revised paper improved significantly in term of reporting from the initial version, excluding many studies, I am left with concern about quality of databases' searches, methods of the manuscript, and the interpretability of the research findings, given that studies included in this work (n=169) were not reported in the main manuscript, and were not linked to results. These should be points to further refinement of the work by referencing included studies in the main body of the manuscript and linking the key results to the data and citations, so results can be directly inspected. By doing so, you increase the likelihood that the results you report are correct, and this also provides means to extend the research on the quality of systematic reviews and meta-analyses in the future. In the current state, it remains unclear whether this work is protected against misleading results. Moreover, the journal's requirements of transparent reporting and reproducibility are not met.

Response: Thanks for your valuable comments on our revised manuscript, we have revised manuscript carefully.

For the studies included in this work, we have added the citation in the appendix data file

(*appendix 3*, we added two additional column named “citation” and “citation refers to reference listed in *appendix 1*” for included studies) to make it easier to follow.

In *appendix 1* “list of included studies” the list of included studies is reported in the same order as in *appendix 3*.

As mentioned before, we did not include the references in the main text because of the journal’s policy on number of references

Reviewer #2: I am satisfied with the changes made to the manuscript.

Response: Thanks for your positive comment.

Journal Pre-proof