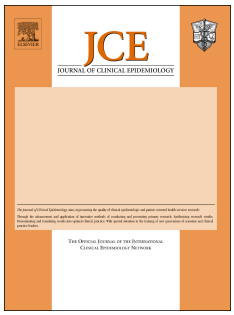


Updating systematic reviews can improve the precision of outcomes: a comparative study

Ya Gao, Kelu Yang, Yitong Cai, Shuzhen Shi, Ming Liu, Junhua Zhang, Jiarui Wu, Jinhui Tian, Fujian Song

PII: S0895-4356(19)31132-1

DOI: https://doi.org/10.1016/j.jclinepi.2020.05.019

Reference: JCE 10163

To appear in: *Journal of Clinical Epidemiology*

Received Date: 12 December 2019

Revised Date: 18 April 2020

Accepted Date: 13 May 2020

Please cite this article as: Gao Y, Yang K, Cai Y, Shi S, Liu M, Zhang J, Wu J, Tian J, Song F, Updating systematic reviews can improve the precision of outcomes: a comparative study, *Journal of Clinical Epidemiology* (2020), doi: https://doi.org/10.1016/j.jclinepi.2020.05.019.

Author Statement

Updating systematic reviews can improve the precision of outcomes: a comparative study

I have made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
I have drafted the work or revised it critically for important intellectual content; AND
I have approved the final version to be published; AND
I agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

The detailed author contributions are as follows:
YG and JT planned and designed the study. YG and JT developed search strategies. YG, KY, YC, and SS screened potential studies and extracted data from the included studies. YG, KY, ML, and JT managed the data and performed the statistical analysis. JT and FS conducted arbitration under disagreement and ensured that there were no errors. JZ, JW, and FS provided methodological support and helped to interpret findings. YG, KY, and JT wrote the first draft. YG, JT, and FS revised the draft. All authors approved the final version of the manuscript.

Jinhui Tian

Updating systematic reviews can improve the precision of outcomes: a comparative study

Ya Gao[1†], Kelu Yang[2†], Yitong Cai[2], Shuzhen Shi[1], Ming Liu[1], Junhua Zhang[3], Jiarui Wu[4], Jinhui Tian[1,2]*, Fujian Song[5]

[1]Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China

[2]Evidence-Based Nursing Center, School of Nursing, Lanzhou University, Lanzhou, China

[3]Evidence-Based Medicine Center, Tianjin University of Traditional Chinese Medicine, Tianjin, China

[4]Department of Clinical Chinese Pharmacy, School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing, China

[5]Public Health and Health Services Research, Norwich Medical School, University of East Anglia, Norwich, UK

[†]Ya Gao and Kelu Yang contributed equally to this work.

*** Correspondence to:** Jinhui Tian. Evidence-Based Medicine Centre, School of Basic Medical Sciences, Lanzhou University; Evidence-Based Nursing Center, School of Nursing, Lanzhou University, No.199, Donggang West Road, Lanzhou City,

730000, Gansu Province, China. (Email: tianjh@lzu.edu.cn)

Running title: Updating systematic reviews can improve the precision of outcomes

Abstract

Objectives: To investigate the main characteristics and the precision of outcomes between updated and original systematic reviews (SRs).

Study Design and Setting: We searched PubMed and Embase.com on 31 March 2019, and included 30 pairs of updated and original SRs. We calculated changes in outcomes and the precision of effect size estimates in updated SRs, compared with original SRs. Review Manager 5.3 software was adopted to create forest plots showing comparable outcomes.

Results: The average update time was 56.0 months, and incorporating new trials (23 SRs, 76.7%) was the main reason for the update. Compared with original SRs, 24 (80.0%) updated SRs included more randomized controlled trials (RCTs) and 22 (73.3%) updated SRs involved a larger number of patients. Of the 130 comparable outcomes, only three (2.3%) outcomes were observed with a significant change in three SR updates. No new data from RCTs were added to 36 (27.7%) outcomes during the update process. Of the 94 outcomes including new evidence, 83 (88.3%) showed an improvement in precision, 5 (5.3%) showed a decrease, and 6 (6.4%) did not exhibit changes in precision.

Conclusion: Updating SRs could increase the precision of most comparable outcomes, although the conclusions of almost all updated SRs were similar to original SRs.

Keywords: Meta-analysis; Update; Treatment effect, Evidence certainty, Outcome change; Meta-epidemiology

What is new?

Key findings

- Systematic reviews did not exhibit outcome changes during the update process.
- Updating systematic reviews could increase the precision of estimates of treatment effects.
- Most updated systematic reviews only briefly describe the reasons for updating and some cannot be identified as updated SRs.

What this adds to what was known?

- This study compared the general characteristics, statistical methods, and the precision of outcomes between updated SRs and original SRs.

What is the implication and what should change now?

- Researchers, journal editors, and peer reviewers should fully assess the need for updates and encourage authors to clarify the detailed reasons for the updates.
- Future research should fully consider multiple factors when updating systematic reviews to ensure that the results of systematic reviews better guide the clinical practice and avoid wasting resources and time.

1. Introduction

Systematic reviews (SRs) are fundamental scientific activities [1, 2], and help

clinicians keep up with the pace of medical literature by summarizing a large amount of evidence and explaining the differences in results between studies [3]. It was estimated that more than 8,000 SRs were indexed in Medline annually, which is equivalent to a threefold increase over the past decade [4, 5]. SRs are increasingly used in medical decision-making, pointing out the direction of future research, formulating clinical policies, and combining the best evidence with clinical practice [6, 7]. However, if the SR report is incomplete or the implementation method is flawed, the role of the SR in decision-making will be limited [8]. When health care workers apply the results of SRs in practice, they should first assess the validity of the SRs [9]. However, there is abundant evidence showed that the results of SRs may be outdated with the progress of research, and an SR is most useful when it is kept up to date [10-12].

The panel for updating guidance for the SRs group defines an update of SRs as a new version of the published SRs [10]. Updating SRs can include newly published studies and some gray literature, which can reduce publication bias and improve the credibility of the results of SRs [11-13]. Because systematic reviews are relatively more influential in evidence-based clinical decision-making compared to other study designs, it has been suggested that regularly updating of systematic reviews is needed to ensure that the evidentiary basis is current, complete, and as precise as possible [14-16]. Updated SRs may generate new and very important information, but the process of updating the SR is time-consuming and laborious, and may also inflate type I errors [10, 17, 18]. Previous studies have summarized methods and strategies

regarding when and how to update SRs [19], identified the signals for the need to update SRs [20], tested the effectiveness of simplified search strategies for updating SRs [21], and determined characteristics that needed to estimate the risk of conclusion changes in SR updates [22]. However, there was a lack of empirical evidence on differences in characteristics and outcome changes between updated and original SRs.

The primary objective of this study was to investigate the general characteristics of included SRs and compare these factors between updated SRs and original SRs of health care interventions. The secondary objective was to examine whether the updated SRs exhibited outcomes change and whether updated SRs improved the precision of outcomes. We did not attempt to explore the differences of methodological and reporting quality between updated SRs and original SRs, as these were already reported in our previous study [23].

2. Methods

2.1. Search strategy

We conducted comprehensive electronic searches in PubMed and Embase.com to identify updated SRs from inception to March 3, 2019. Search terms included the following words: “systematic review”, “meta-analysis”, “indirect comparison”, “indirect treatment”, “mixed treatment comparison”, “multiple treatment comparison”, and “update”. A combination of subject terms and keywords was used and we made

appropriate adjustments of vocabulary and grammar between different databases. We applied no restriction on publication date. The search strategies are presented in Appendix Word 1. Reference lists of relevant SRs were manually searched for potentially eligible studies.

2.2. Inclusion and exclusion criteria

Eligibility criteria are the same as those reported in our previous research [23]. Systematic reviews of interventions with or without meta-analysis that met the following criteria were included: (1) was a review article and explicitly described methods of study selection, and explicitly reported the methods of evidence synthesis [5, 24]; (2) all the original SRs and updated SRs included only randomized controlled trials (RCTs) or quasi-randomized controlled trials to evaluate clinical effects of health care interventions (including pharmaceutical agents, surgeries, and health technologies); and (3) all the original SRs and updated SRs were published in the English language.

We excluded following studies: (1) SRs that included both RCTs and nonrandomized studies or only included nonrandomized studies; (2) SRs did not focus on health care interventions such as etiology, diagnosis, and prognosis; (3) SRs did not clearly state "update" in the titles or articles; (4) the original SR or updated SR is a Cochrane review; (5) the second analysis of previous SR; (6) overviews of SRs, methodological reviews, umbrella overviews, scoping or rapid reviews, review

protocols, abstracts, conference proceedings, and letters to editors.

2.3. Study selection

The retrieved records were imported into EndNote X8 (Thomson Reuters (Scientific) LLC Philadelphia, PA, US) for management. The titles and abstracts of the identified records from the electronic database search were screened by two independent reviewers (Y.G. and Y.T.C.) to determine if they met the inclusion criteria. Then, the same two reviewers retrieved the full text of all possibly relevant studies and assessed the eligibility of each study according to the inclusion criteria. If we identified SRs with more than one updated version, the most recent one was included. Conflicts were resolved through discussions with a third reviewer (J.H.T.).

2.4. Data extraction

We developed a data extraction form using Microsoft Excel 2016 (Microsoft Corp, Redmond, WA, www.microsoft.com) through discussions with the review team and revised it after piloting on a random of five SRs. Then, one reviewer (Y.G., K.L.Y., Y.T.C., or S.Z.S) extracted data from the included SRs and a second reviewer (J.H.T.) checked the extracted data. The detailed data included: (1) general information: first author, year of publication, country of the corresponding author, number of authors, journal name, subject, details of inclusion criteria, number of

included studies, and number of participants; (2) quality of RCTs included in updated SRs and original SRs: random sequence generation, allocation concealment, blinding of participants and personnel, and blinding of outcome assessment; (3) main results: summary measures, effect size, 95% confidence intervals (95% CIs), and *P* value. Disagreements were resolved by consensus or by the discussion with a third reviewer (J.H.T.).

2.5. Data management and analysis

We compared the general characteristics between updated SRs and original SRs. Frequency and percentage were used for categorical variables, and median and interquartile range were used for continuous variables. We used the Chi-squared test or Fisher exact test (if a contingency table contained a cell with five or fewer events) to assess the differences in categorical data and nonparametric statistical approach (two-sample Wilcoxon rank-sum test) or Student t-test for continuous data [5, 24]. The analyses were conducted using IBM SPSS Statistics v. 24.0 (Armonk, NY: IBM Corp).

To determine whether updates exhibited an outcome change, we extracted comparative outcomes from the updated SRs and original SRs. Outcomes were considered comparable when the updated SR and the original SR evaluated the same interventions and outcome indicators. Then, we calculated the ratio of effect size, the difference of standardized mean differences, and the difference of mean differences to

determine whether there was a change in outcome using the extracted data of comparative outcomes. Analyses were conducted by using STATA (13.0; Stata Corporation, College Station, Texas, USA Stata). We also used Review Manager 5.3 software (Cochrane Collaboration, Oxford, UK) to create forest plots showing comparable outcomes to compare the precision of the results.

3. Results

3.1. Screening results

The systematic search in PubMed, Embase.com, and reference lists yielded 4997 records, with 2214 duplicates. After screening titles and abstracts of the remaining 2783 records, 2536 were considered irrelevant and excluded. Further assessing the full-text and retrieving the relevant original SRs of the remaining 247 records, a total of 60 SRs was included for the analyses, including 30 updated SRs and 30 original SRs. The flowchart of the screening process is presented in Appendix Figure 1. The full lists of included SRs can be found in Appendix Word 2.

3.2. General characteristics of included SRs

The included SRs were published between 1994 and 2018 with the majority of them (70.0%) published after 2007 (Figure 1). 60 SRs were conducted in 19 countries,

and the United Kindom ranked first, with 12 publications (Appendix Figure 2). The included SRs covered a wide range of disease categories, 10 (16.7%) focused on neoplasms, 8 (13.3%) related to diseases of the circulatory system, and 6 (10.0%) focused on diseases of the musculoskeletal system and connective tissue and diseases of the digestive system, respectively (Figure 2).

Table 1 summarized the characteristics of the included SRs. 20 original SRs and 26 updated SRs published in journals with impact factors between 0.0 and 9.0, and there was no significant difference between the updated SRs and original SRs. 70.0% of the SRs completed by one to six authors. The median number of RCTs included in original SRs was 13.5, and the median number of samples was 1625. The updated SRs included a median of 18.5 RCTs involving a median number of 2419 patients, and significant differences were observed in the number of included RCTs and samples between the updated SRs and original SRs (Table 1). Only 36.7% of the original SRs conducted subgroup analyses, 40.0% performed sensitivity analyses, and no original SRs performed meta-regression analyses. As for updated SRs, 53.3% performed subgroup analyses, 33.3% conducted sensitivity analyses, and 6.7% conducted meta-regression analyses. The publication bias was evaluated in seven original SRs and six updated SRs, and significant publication bias was reported in two original SRs and three updated SRs. Of the 60 SRs, only two pairs of updated and original SRs used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) [25] tool to assess the quality of evidence (Table 2).

Table 1 The characteristics of the included SRs

Characteristics	Original SRs (n=30)	Updated SRs (n=30)	*P* value
Journal impact factor			
0.0 to 3.0	8(26.7)	11(36.7)	0.405
3.1 to 6.0	8(26.7)	11(36.7)	0.405
6.1 to 9.0	4(13.3)	4(13.3)	1.000
> 9.0	9(30.0)	4(13.3)	0.117
Non-SCI	1(3.3)	0(0.0)	1.000
Number of authors			
1 to 3 authors	10(33.3)	8(26.7)	0.573
4 to 6 authors	11(36.7)	13(43.3)	0.598
7 or more authors	9(30.0)	9(30.0)	1.000
Number of RCTs included: median (IQR)	13.5(7.75, 29.25)	18.5(9.75, 30)	0.010
Number of patients included: median (IQR)	1625(662, 6800)	2419(955, 8898)	0.004

SRs, systematic reviews; SCI, science citation index; RCTs, randomized controlled trials; IQR, interquartile range.

Table 2 The statistical analyses of the included SRs

Category	Original SRs	Updated SRs	Frequency	Proportion (%)
Subgroup analysis conducted?	Yes	Yes	10	33.3
	No	Yes	6	20.0
	Yes	No	1	3.3
	No	No	5	16.7
	Did not conduct meta-analyses	No	2	6.7
	Did not conduct meta-analyses	Did not conduct meta-analyses	6	20.0
Sensitivity analysis conducted?	Yes	Yes	7	23.3
	No	Yes	3	10.0
	Yes	No	5	16.7
	No	No	7	23.3
	Did not conduct meta-analyses	Yes	1	3.3
	Did not conduct meta-analyses	No	1	3.3
	Did not conduct meta-analyses	Did not conduct meta-analyses	6	20.0
Meta-regression analysis conducted?	Yes	Yes	0	0.0
	No	Yes	2	6.7
	Yes	No	0	0.0
	No	No	20	66.7
	Did not conduct meta-analyses	No	2	6.7
	Did not conduct meta-analyses	Did not conduct meta-analyses	6	20.0
Publication bias assessed?	Yes	Yes	5	16.7
	No	Yes	1	3.3
	Yes	No	2	6.7
	No	No	22	73.3
GRADE used	Yes	Yes	2	6.7
	No	Yes	1	3.3
	Yes	No	0	0.0
	No	No	27	90.0

SRs, systematic reviews; GRADE, Grading of Recommendations Assessment, Development and Evaluation.

3.3. Reasons for updating and time interval of update

The 30 original SRs were updated mainly for the following four reasons: to incorporate new trials (23 SRs, 76.7%), to assess whether new evidence will change the conclusion of the original SR (2 SRs, 6.7%), to assess detailed interventions (1 SR, 3.3%), and original SR being criticized for its methodology and interpretation of results (1 SR, 3.3%). Twenty-five original SRs were updated once and five were updated two times. For SRs with one update, the median of the update interval was 43.5 months. For SRs with two updates, the median update time was 51.0 months from the original SR to the first version of updated SR, and 79.0 months from the first version of updated SR to the latest updated SR. Overall, The average update time was 56.0 months, Table 3.

Table 3 Reasons for updating and time interval of update

Category	Frequency	Proportion (%)
Reasons for updating (n=30)		
Incorporate new trials	23	76.7
Assess whether new evidence will change the conclusion of original SR	2	6.7
Assess detailed interventions	1	3.3
Original SR was criticized for its methodology and interpretation of results	1	3.3
Not reported	3	10.0
Time interval of update (month) (n=30)		
SRs with 1 update (n=25)		
median (IQR)	43.5(27.25,68.75)	
Unclear	5	20.0
SRs with 2 updates (n=5)		
Original SR to first updated SR: median (IQR)	51.0(39.5,71.0)	
First updated SR to latest updated SR: median (IQR)	79.0(36.5,110.5)	
Original SR to latest updated SR: median (IQR)	131.0(76.0,181.0)	
Overall: mean (IQR) (n=25)	56.0(34.0,79.5)	

SRs, systematic reviews; IQR, interquartile range.

3.4. Differences between updated SRs and original SRs

Five (16.7%) updated SRs were published in a higher impact factor journal compared to original SRs, but half of the updated SRs were published in lower impact factor journals. Compared with original SRs, 14 (46.7%) updated SRs included more authors, 24 (80.0%) updated SRs included more RCTs, and 22 (73.3%) updated SRs involved a larger number of patients. As for changes in inclusion criteria, the most common aspect was the outcome (9, 75.0%), other factors included population (4, 33.3%), intervention (6, 50.0%), and comparison (7, 58.3%). Appendix Table 1 presents the details of the inclusion criteria. Considering literature search, 14 (46.7%) pairs searched the same number of databases, and 16 (53.3%) updated SRs searched the years that the original SRs covered. More than 80.0% updated SRs incorporated RCTs that previously included in the original SRs, and more than 60.0% updated SRs included no less than 90.0% of the previously included RCTs. The changed information between updated SRs and original SRs is summarized in Table 4.

Table 4 Changed information between updated SRs and original SRs

Category	Frequency	Proportion (%)
Whether the impact factors change?		
Increase	5	16.7
Reduce	15	50.0
No change	10	33.3
Whether the number of authors changes?		
Increase	14	46.7
Reduce	7	23.3
No change	9	30.0
Whether the number of included RCTs changes?		

Increase	24	80.0
Reduce	4	13.3
No change	2	6.7
Whether the number of included patients change?		
Increase	22	73.3
Reduce	4	13.3
No change	1	3.3
Unclear	3	10.0
Change of inclusion criteria (Yes) (n=30)	12	40.0
Changed factors of inclusion criteria (n=12)		
Population	4	33.3
Intervention	6	50.0
Comparison	7	58.3
Outcome	9	75.0
Whether to change the number of databases searched (n=30)		
Change		
Add 10 databases	1	3.3
Add 3 to 4 databases	3	10.0
Add 1 to 2 databases	3	10.0
Reduce 1 database	4	13.3
Reduce 2 databases	3	10.0
No change	14	46.7
Unclear	2	6.7
Whether to search the years that original SRs retrieved (n=30)		
Yes	16	53.3
No	12	40.0
Unclear	2	6.7
Whether to include RCTs included in the original SRs (Yes) (n=30)	25	83.3
Proportion of previously RCTs of original SRs is included in the updated SRs (%) (n=25)		
0 to 30	1	4.0
31 to 50	0	0.0
51 to 70	2	8.0
71 to 90	3	12.0
91 to 99	2	8.0
100	17	68.0

SRs, systematic reviews; RCTs, randomized controlled trials.

3.5. Methodological quality of RCTs included in original SRs and RCTs newly included in updated SRs

Twenty-one (70.0%) original SRs and 23 (76.7%) updated SRs assessed the methodological quality of included RCTs, although most SRs did not provide results of individual items. Eight updated SRs and nine original SRs reported the result of random sequence generation for 179 RCTs. There was a significant difference between previously included RCTs and newly included RCTs on high-risk results ($P < 0.001$), but no significant differences were found between them on unclear risk ($P = 0.529$) and low-risk results ($P = 0.163$). The previously included RCTs and newly included RCTs had the similar methodological quality in items of allocation concealment ($P > 0.05$). Considering the blinding of participants and personnel, relatively more newly included RCTs were rated as low risk (45.6%), compared with those previously included (25.5%). Compared to updated SRs, original SRs included less high-risk quality RCTs in terms of blinding of outcome assessment ($P = 0.002$) (Table 5).

Table 5. Comparisons of methodological quality of previously included RCTs in original SRs and newly included RCTs in updated SRs

Category	Frequency	Proportion (%)	*P* value
Random sequence generation (Reported in 8 updated SRs and 9 original SRs)			
High risk			<0.001
Previously included RCTs (n=126)	1	0.8	
Newly included RCTs (n=53)	9	17.0	
Unclear risk			0.529
Previously included RCTs (n=126)	73	57.9	
Newly included RCTs (n=53)	28	52.8	
Low risk			0.163
Previously included RCTs (n=126)	52	41.3	
Newly included RCTs (n=53)	16	30.2	
Allocation concealment (Reported in 8 updated SRs and 6 original SRs)			
High risk			0.252
Previously included RCTs (n=113)	10	8.8	

Newly included RCTs (n=70)	10	14.3	
Unclear risk			>0.078
Previously included RCTs (n=113)	79	69.9	
Newly included RCTs (n=70)	40	57.1	
Low risk			>0.259
Previously included RCTs (n=113)	24	21.2	
Newly included RCTs (n=70)	20	28.6	
Blinding of participants and personnel (Reported in 10 updated SRs and 10 original SRs)			
High risk			>0.238
Previously included RCTs (n=141)	48	34.0	
Newly included RCTs (n=90)	24	26.7	
Unclear risk			<0.001
Previously included RCTs (n=141)	57	40.4	
Newly included RCTs (n=90)	15	16.7	
Low risk			<0.002
Previously included RCTs (n=141)	36	25.5	
Newly included RCTs (n=90)	41	45.6	
Blinding of outcome assessment (Reported in 4 updated SRs and 3 original SRs)			
High risk			<0.002
Previously included RCTs (n=31)	0	0	
Newly included RCTs (n=25)	7	28.0	
Unclear risk			>0.113
Previously included RCTs (n=31)	19	61.3	
Newly included RCTs (n=25)	10	40.0	
Low risk			>0.602
Previously included RCTs (n=31)	12	38.7	
Newly included RCTs (n=25)	8	32.0	

SRs, systematic reviews; RCTs, randomized controlled trials.

3.6. Results of comparable outcomes and change of precision

We identified 130 comparable outcomes from 16 among the 30 pairs of included SRs, of which 97 (74.6%) were outcomes of the binary data and 33 (25.4%) were outcomes of the continuous data. As for the 130 comparable outcomes, four (3.1%) changed measures of effect size from the odds ratio to risk ratio, and the remaining 126 (96.9%) used the same measures of effect size. Estimates of the updated SRs

were consistent with the original SRs for 127 (97.7%) comparable outcomes, and only three (2.3%) outcomes were observed with a significant change ($P < 0.05$) in three SR updates (Table 6, Appendix Figures 3-6, Appendix Table 2). However, there were no new data from RCTs for 36 (27.7%) outcomes during the update process. Of the 94 comparative outcomes that included new evidence, the width of the 95% confidence interval for 83 (88.3%) outcomes was narrowed. The standard error of 19 (20.2%) comparable outcomes were reduced between 0.1000 and 1.2000, 32 (34.0%) were reduced between 0.0499 and 0.0100, and 5 (5.3%) were increased between 0.0001 and 0.0600 (Table 6, Appendix Table 3). We further compared results of newly included and previously included RCTs in the updated SRs, for 47 comparative outcomes. Six outcomes were found with a statistical difference between the evidence of new RCTs and previously included RCTs of original SRs(Table 6, Appendix Figures 7-10).

Table 6 Summaries of comparable outcomes

Category	Frequency	Proportion (%)
Number of SR pairs with comparable outcomes (n=30 Pairs)	16	100.0
Number of comparable outcomes	130	100.0
Category of comparable outcomes (n=130)		
Binary data	97	74.6
Continuous data	33	25.4
Measures of the comparable outcomes (n=130)		
HR	18	13.8
OR	3	2.3
RR	72	55.4
OR and RR	4	3.1
SMD	12	9.2
MD	21	16.2
Whether the measures of effect size change (n=130)		
Change OR to RR	4	3.1
No change	126	96.9

Whether the comparable outcomes change (n=130)		
Change	3	2.3
Binary data	2	1.5
Continuous data	1	0.8
No change	127	97.7
Binary data	95	73.1
Continuous data	32	24.6
Whether the comparable outcomes added new RCTs in the updated SRs (n=130)		
Yes	94	72.3
No	36	27.7
Whether the precision of the comparable outcomes change between updated SRs and original SRs (n=94)		
Improve	83	88.3
Binary data	58	61.7
Continuous data	25	26.6
Reduce	5	5.3
Binary data	3	3.2
Continuous data	2	2.1
No change	6	6.4
Binary data	3	3.2
Continuous data	3	3.2
Change of standard error of comparable outcomes (n=94)		
-1.2000 to -0.1000	19	20.2
-0.0999 to -0.0500	19	20.2
-0.0499 to -0.0100	32	34.0
-0.0099 to -0.0001	13	13.8
0.0000 to 0.0600	11	11.7
Is there a statistical difference between the evidence of newly RCTs and previously RCTs in the comparable outcomes (n=47)		
Yes	6	12.8
No	41	87.2

SRs, systematic reviews; RCTs, randomized controlled trials; HR, hazard ratio; OR, odds ratio; RR, risk ratio; SMD, standard mean difference; MD, mean difference.

4. Discussion

4.1. Summary of findings

We conducted a comprehensive literature search and identified 30 pairs of updated and original SRs, which were published in 48 journals between 1994 and

2018. There were mainly four reasons for the authors to update their original SRs, and the most common reason was to incorporate new trials. Only ten pairs of updated and original SRs conducted subgroup analyses, no more than 25.0% pairs of SRs performed sensitivity analyses, and only two SRs conducted meta-regression analyses. It is well known that heterogeneity has a crucial impact on the interpretation of the results of SRs, while sensitivity analysis, subgroup analysis, and meta-regression analysis are the main methods for exploring heterogeneity. Therefore, future research should conduct these analyses to explore heterogeneity to make the results of the meta-analysis more credible. GRADE system is a tool that can be used to rate the quality of evidence of SRs and other bodies of evidence, and the Cochrane Collaboration recommended authors to use GRADE for all important outcomes in their SRs [25, 26]. The previous study has indicated that reviews with a high level of certainty in the results assessed by the GRADE tool were less likely to change by adding new evidence [10]. GRADE can help interpret changes in results between original and updated systematic reviews, as well as help decide whether to update SRs [10]. However, the current study found only 8.3% of the included SRs used the GRADE to assess the quality of evidence. Thus, the application of GRADE needs to be further promoted.

All the included SRs reported inclusion criteria in terms of specified participants, interventions, comparisons, and outcomes. The inclusion criteria for most updated SRs remain unchanged, although more detailed inclusion criteria were applied in some. More than 70.0% SRs assessed the risk of bias of included RCTs, but no more

than 35.0% SRs provided the detailed result of individual items. The updated SRs incorporated more RCTs with high-risk of bias regarding random sequence generation and blinding of outcome assessment, compared to original SRs. However, compared with RCTs previously included in original SRs, the newly included RCTs more often clarified the blinding of participants and personnel. These, to some extent, indicated that the quality of previously included RCTs and newly included RCTs was generally similar.

We identified 130 comparable outcomes from sixteen pairs of original and updated SRs. Among the 130 outcomes, only three were found to have statistical differences between the updated and original SRs. The statistical difference was changed for one binary outcome after changing the measure of effect size from the odds ratio to risk ratio without new evidence added during the update process. The other two outcomes incorporated new RCTs addition to RCTs previously included in the original SRs. The new evidence may contribute to the change in the overall effect size. There were no changes for 96.9% of the outcomes in the SR updates, including those (27.7%) that did not integrate new evidence during the update process. This result is consistent with findings from a previous study [22], in which Bashir et al. applied classification trees to model the risk of conclusion changes in SR updates and found that 68.0% of the SRs did not exhibit a change in the conclusion in their updates. Therefore, we can conclude that the results of most updated SRs are generally consistent with results of the original SRs.

Based on a comparison of standard errors of effect size estimates for 94

comparable outcomes, we observed that 88.3% of the outcomes showed an improvement in precision. Outcomes with fewer randomized controlled trials are expected to have wider confidence intervals [27]. Therefore, the incorporation of new trials in updated SRs will increase the sample size and improve the precision of the results. However, the precision was reduced in five outcomes after incorporating new evidence in updated SRs. Six outcomes did not exhibit a change in precision, but their effect size may have changed. Overall, the precision of effect size estimates could be improved for most comparable outcomes in updated SRs.

4.2. Implications for future research

A previous study, published in 2007 by Moher et al, summarized the methods and strategies describing when and how to update systematic reviews [19]. The current study found that the updated SRs could increase the precision of most outcomes, although almost all the results of updated SRs were consistent with the results of the original SRs.

The average update time of SRs was 56.0 months, which may not be optimal. The Cochrane collaboration recommended that SRs should be updated based on need and priority [28]. It is difficult to know when a given review might become out of date, but tools have been available to help determine when a review needs to be updated [10, 28, 29]. Previous studies indicated that updating meta-analyses may result in biased treatment estimates and inflated type I error rates [30]. Therefore, we

should take the appropriate approach to update SRs at the appropriate time interval. Otherwise, it will not only waste resources and time but also produce inaccurate results and even mislead clinical practice. Future research should fully consider the need for updating, including treatment characteristics, statistical methods, clinical endpoints, availability of resources, and the impact of treatment on public health when updating SRs to ensure that the results of SRs better guide the clinical practice [19]. Recently, a novel approach to SR updating, living systematic reviews have been developed [31], by continually incorporating new evidence as it becomes available. Although there are many challenges in statistical methods, production processes, peer and editorial review, and publication [31-33], living SRs may be particularly important in fields where research evidence is rapidly emerging, current evidence is uncertain, and new research may change policy or practice decisions [28, 32]. Future reviewers can reasonably adopt living SRs to reduce the time of translating results of new research into health practice and reduce the waste of research investment by society [32].

Our study also revealed that the overall quality of RCTs was low, and most SRs did not provide the result of individual items of quality assessment. Therefore, further research should clearly provide the results of quality assessment and incorporate high-quality RCTs to provide more reliable evidence. Majority of the included SRs did not conduct subgroup analyses, sensitivity analysis, and meta-regression analysis, so the exploration of heterogeneity still needs to be improved. Most updated SRs only briefly described reasons for updating and some SRs cannot be identified as updated

SRs which should be considered in the peer-reviewing and editorial process. Researchers, journal editors, and peer reviewers should fully assess the need for updates and encourage authors to clarify the detailed reasons for the updates.

4.3. Strengths and limitations

To be the best of our knowledge, this is the first study to investigate the general characteristics, statistical analyses, and methodological quality of included RCTs of updated SRs and original SRs, and to compare the precision of outcomes between updated and original SRs. We also assessed changes in outcomes after including new evidence in updated SRs. There are also some limitations in our study. First, only SRs of RCTs published in English were enrolled, so that findings may not apply to SRs published in other languages and SRs of other types such as cohort studies and observational studies [5]. Second, although this study included 60 SRs in total, the number of comparable SRs and comparable outcomes was rather small, and the available evidence was not sufficient for conducting stratified analyses. Third, outcome changes and differences in precision may be affected by different systematic review methods, although we have compared the population, intervention, comparison, outcomes and statistical methods between updated SRs and original SRs to make the outcomes comparable, and performed analyses using the comparable outcomes. Fourth, analyses associated with effect estimates of comparable outcomes were restricted as some outcomes did not integrate original and new evidence during the

update process. Finally, we did not include Cochrane systematic reviews because of time and resource restrictions in the current study, and we plan to conduct a study to compare the characteristics and outcomes of Cochrane reviews between the updated SRs and original SRs.

5. Conclusions

Updating SRs could increase the precision of estimates of treatment effects for most comparable outcomes, although the results of almost all updated SRs were similar to the original SRs. Future research is required to fully consider multiple factors when updating SR to ensure that the results of SRs better guide the clinical practice and avoid wasting resources and time.

Abbreviations

SR: Systematic review; RCT: randomized controlled trial; GRADE, Grading of Recommendations Assessment, Development and Evaluation.

Acknowledgments

None.

Funding

None.

Authors' contributions

YG and JT planned and designed the study. YG and JT developed search strategies. YG, KY, YC, and SS screened potential studies and extracted data from the included studies. YG, KY, ML, and JT managed the data and performed the statistical analysis. JT and FS conducted arbitration under disagreement and ensured that there were no errors. JZ, JW, and FS provided methodological support and helped to interpret findings. YG, KY, and JT wrote the first draft. YG, JT, and FS revised the draft. All authors approved the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

[1] Mulrow CD. Rationale for systematic reviews. BMJ (Clinical research ed). 1994;309:597-9.

[2] Gao Y, Ge L, Ma X, Shen X, Liu M, Tian J. Improvement needed in the network geometry and inconsistency of Cochrane network meta-analyses: a cross-sectional survey. Journal of clinical epidemiology. 2019;113:214-27.

[3] Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Annals of internal medicine. 1997;126:376-80.

[4] Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS medicine. 2016;13:e1002028.

[5] Ge L, Tian JH, Li YN, Pan JX, Li G, Wei D, et al. Association between prospective registration and overall reporting and methodological quality of systematic reviews: a meta-epidemiological study. Journal of clinical epidemiology. 2018;93:45-55.

[6] Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet (London, England). 2014;383:166-75.

[7] Cohen AM, Ambert K, McDonagh M. Studying the potential impact of automated document classification on scheduling a systematic review update. BMC medical informatics and decision making. 2012;12:33.

[8] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet (London,

England). 2014;383:267-76.

[9] 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;312:171-9.

[10] Garner P, Hopewell S, Chandler J, MacLehose H, Schunemann HJ, Akl EA, et al. When and how to update systematic reviews: consensus and checklist. BMJ (Clinical research ed). 2016;354:i3507.

[11] Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. The Cochrane database of systematic reviews. 2007:Mr000011.

[12] Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. Jama. 2005;294:2203-9.

[13] Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:831-6.

[14] Chalmers I, Enkin M, Keirse MJ. Preparing and updating systematic reviews of randomized controlled trials of health care. The Milbank quarterly. 1993;71:411-37.

[15] Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Annals of internal medicine. 2007;147:224-33.

[16] Créquit P, Trinquart L, Yavchitz A, Ravaud P. Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of

lung cancer. BMC medicine. 2016;14:8.

[17] Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. Statistics in medicine. 2011;30:903-21.

[18] Nikolakopoulou A, Mavridis D, Egger M, Salanti G. Continuously updated network meta-analysis and statistical monitoring for timely decision-making. Statistical methods in medical research. 2018;27:1312-30.

[19] Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, et al. A systematic review identified few methods and strategies describing when and how to update systematic reviews. Journal of clinical epidemiology. 2007;60:1095-104.

[20] Chung M, Newberry SJ, Ansari MT, Yu WW, Wu H, Lee J, et al. Two methods provide similar signals for the need to update systematic reviews. Journal of clinical epidemiology. 2012;65:660-8.

[21] Rice M, Ali MU, Fitzpatrick-Lewis D, Kenny M, Raina P, Sherifali D. Testing the effectiveness of simplified search strategies for updating systematic reviews. Journal of clinical epidemiology. 2017;88:148-53.

[22] Bashir R, Surian D, Dunn AG. The risk of conclusion change in systematic review updates can be estimated by learning from a database of published examples. Journal of clinical epidemiology. 2019;110:42-9.

[23] Gao Y, Cai Y, Yang K, Liu M, Shi S, Chen J, et al. Methodological and reporting quality in non-Cochrane systematic review updates could be improved: a comparative study. Journal of clinical epidemiology. 2020;119:36-46.

[24] Tian J, Zhang J, Ge L, Yang K, Song F. The methodological and reporting quality

of systematic reviews from China and the USA are similar. Journal of clinical epidemiology. 2017;85:50-8.

[25] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ (Clinical research ed). 2008;336:924-6.

[26] Yao L, Sun R, Chen YL, Wang Q, Wei D, Wang X, et al. The quality of evidence in Chinese meta-analyses needs to be improved. Journal of clinical epidemiology. 2016;74:73-9.

[27] Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. BMJ (Clinical research ed). 2011;343:d4909.

[28] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 [updated July 2019]. Cochrane, 2019. Available from www.training.cochrane.org/handbook.

[29] Martínez García L, Pardo-Hernandez H, Superchi C, Niño de Guzman E, Ballesteros M, Ibargoyen Roteta N, et al. Methodological systematic review identifies major limitations in prioritization processes for updating. Journal of clinical epidemiology. 2017;86:11-24.

[30] Borm GF, Donders AR. Updating meta-analyses leads to larger type I errors than publication bias. Journal of clinical epidemiology. 2009;62:825-30.e10.

[31] Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JP, Mavergames C, et al.

Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS medicine. 2014;11:e1001603.

[32] Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction-the why, what, when, and how. Journal of clinical epidemiology. 2017;91:23-30.

[33] Simmonds M, Salanti G, McKenzie J, Elliott J. Living systematic reviews: 3. Statistical methods for updating meta-analyses. Journal of clinical epidemiology. 2017;91:38-46.

Figure legends

Figure 1. Published years of the included SRs.

Figure 2. Categories of disease of included SRs (according to ICD-11). Not applicable indicates that the topic of included SRs does not focus on diseases, such as the use of frozen plasma and the promotion of exclusive breastfeeding.

Supplementary files

Appendix Word 1. Search strategies of PubMed and Embase.

Appendix Word 2. List of included SRs.

Appendix Table 1. Summary of inclusion criteria.

Appendix Table 2. Summary of comparable outcomes and comparison of updated SRs and original SRs in comparable outcomes.

Appendix Table 3. Comparison of the precision of comparable outcomes between updated SRs and original SRs.

Appendix Figure 1. The flowchart of the screening process.

Appendix Figure 2. Countries of the included SRs.

Appendix Figure 3. Comparable outcomes between updated SRs and original SRs in dichotomous variables of the (A) hazard ratio, (B) odds ratio, (C) risk ratio.

Appendix Figure 4. (Appendix Figure 4A, Appendix Figure 4B). Comparable outcomes between updated SRs and original SRs in dichotomous variables of the risk ratio.

Appendix Figure 5. Comparable outcomes between updated SRs and original SRs in continuous variables of the standardized mean difference.

Appendix Figure 6. Comparable outcomes between updated SRs and original SRs in continuous variables of the mean difference.

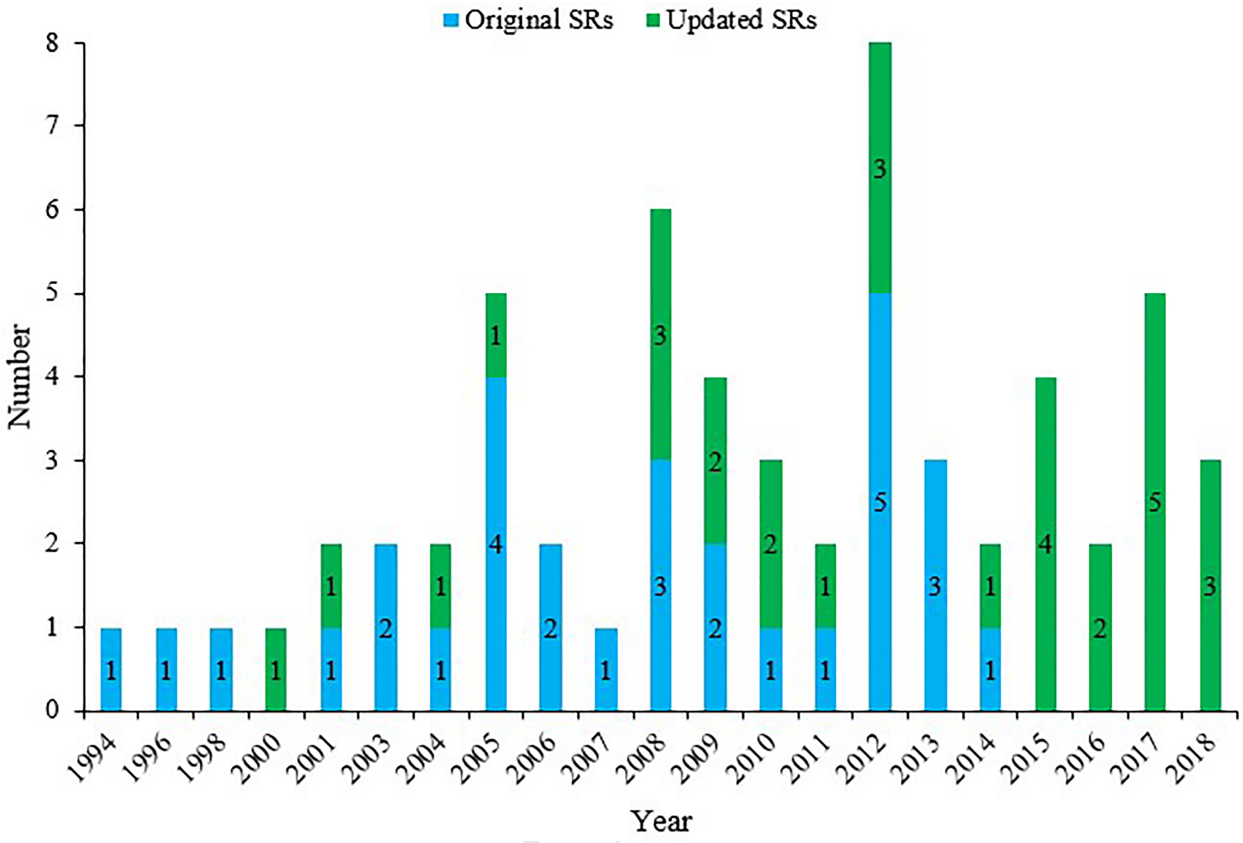
Appendix Figure 7. Comparable outcomes between previously included RCTs and newly included RCTs in dichotomous variables of the (A) hazard ratio, (B) odds ratio.

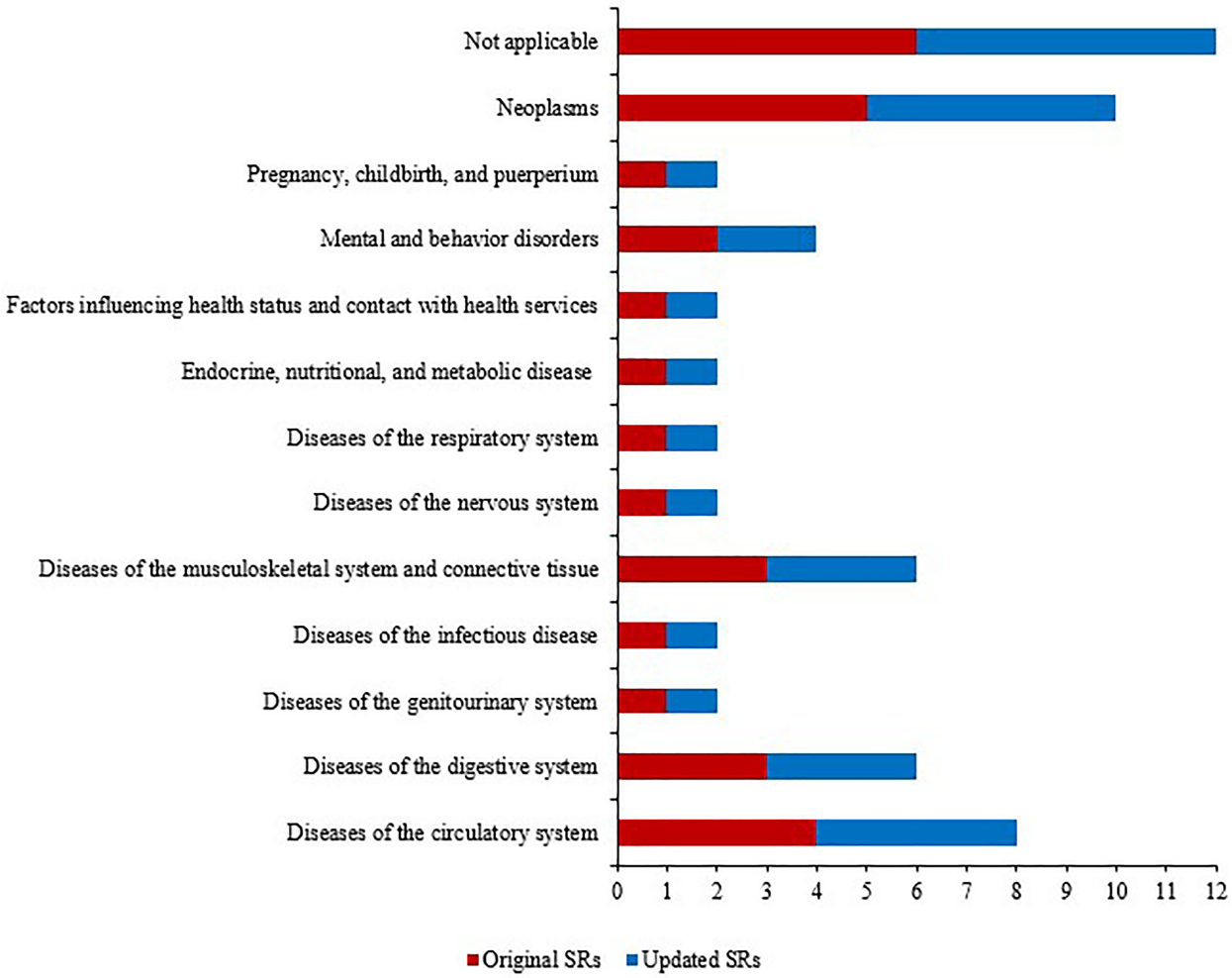
Appendix Figure 8. (Appendix Figure 8A, Appendix Figure 8B). Comparable outcomes between previously included RCTs and newly included RCTs in dichotomous variables of the risk ratio.

Appendix Figure 9. Comparable outcomes between previously included RCTs and newly included RCTs in continuous variables of the standardized mean difference.

Appendix Figure 10. (Appendix Figure 10A, Appendix Figure 10B). Comparable outcomes between previously included RCTs and newly included RCTs in continuous

variables of the mean difference.


Original SRs
Updated SRs
Number
Year
0
1
2
3
4
5
6
7
8
1994
1996
1998
2000
2001
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018


Not applicable
Neoplasms
Pregnancy, childbirth, and puerperium
Mental and behavior disorders
Factors influencing health status and contact with health services
Endocrine, nutritional, and metabolic disease
Diseases of the respiratory system
Diseases of the nervous system
Diseases of the musculoskeletal system and connective tissue
Diseases of the infectious disease
Diseases of the genitourinary system
Diseases of the digestive system
Diseases of the circulatory system
0 1 2 3 4 5 6 7 8 9 10 11 12
Original SRs
Updated SRs

What is new?

Key findings

- Systematic reviews did not exhibit outcome changes during the update process.
- Updating systematic reviews could increase the precision of estimates of treatment effects.
- Most updated systematic reviews only briefly describe the reasons for updating and some cannot be identified as updated SRs.

What this adds to what was known?

- This study compared the general characteristics, statistical methods, and the precision of outcomes between updated SRs and original SRs.

What is the implication and what should change now?

- Researchers, journal editors, and peer reviewers should fully assess the need for updates and encourage authors to clarify the detailed reasons for the updates.
- Future research should fully consider multiple factors when updating systematic reviews to ensure that the results of systematic reviews better guide the clinical practice and avoid wasting resources and time.