Methodological Quality of Systematic Reviews for Questions of Therapy and Prevention Published in the Urological Literature (2016–2021) Fails to Improve

Maylynn Ding,[1*] Jared Johnson,[2,3*] Onuralp Ergun,[2,3] Gustavo Ariel Alvez,[4] Philipp Dahm✉[2,3]

[1] McMaster University, Department of Surgery, Division of Urology, Hamilton, Canada [2] University of Minnesota, Department of Urology, Minneapolis, United States [3] Minneapolis VA Medical Center, Urology Section, Minneapolis, United States [4] Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

*Maylynn Ding and Jared Johnson contributed equally.

Abstract

Objectives Prior studies have suggested that few systematic reviews (SRs) published in the urological literature provide reliable evidence. We performed this study to provide a longitudinal analysis of the methodological quality of SRs published in 5 major urology journals over a 6-year period (2016–2021).

Methods As an extension of a prior study with a written a priori protocol, we systematically searched and analyzed all SRs related to questions of therapy or prevention published in the 5 major urology journals. Three independent reviewers working in pairs selected eligible studies and abstracted the data in duplicate. We used the updated Assessment of Multiple Systematic Reviews (AMSTAR-2) instrument to assess SR quality. We performed pre-planned statistical hypothesis testing by time period and journal of publication in SPSS Version 27.0.

Results Our updated search (2019–2021) identified 563 references of which 114 ultimately met inclusion criteria, which we added to the database of the prior 144 studies (2016–2018). Overall, among 258 SRs, only 6 (2.3%) and 9 SRs (3.5%), achieved a "high" (no critical weakness; up to one non-critical weakness) or "moderate" (no critical weakness; more than one non-critical weakness) confidence rating, respectively. Most SRs published had very low confidence rating (195; 75.6%). The proportion of studies with a high or moderate rating (6.1% versus 4.9%; $P = 0.481$) did not increase over time.

Conclusions Most SRs published in the urological literature continue to have serious methodological limitations and should not be relied upon. There is a critical need for greater awareness for established methodological standards.

Introduction

Systematic reviews (SRs) are foundational to evidence-based clinical practice, informing individual decision-making at the point of care and serving as the foundation of clinical practice guidelines and health policy decisions[1]. They are being published in increasing numbers, in part because of their ability to generate a large number of citations. As a result, there are areas of medicine where the number of SRs exceeds that of the contributing individual studies being synthesized[2]. In addition, methodological quality of many reviews is modest, thereby undermining the confidence that readers should be able to place in the results.

Key Words

Systematic reviews, meta-analysis, evidence-based medicine, AMSTAR

Competing Interests

None declared.

Article Information

Received on March 2, 2023
Accepted on May 13, 2023
This article has been peer reviewed.

Soc Int Urol J. 2023;4(5):415–422
DOI: 10.48083/WURA1857

This is an open access article under the terms of a license that permits non-commercial use, provided the original work is properly cited.
© 2023 The Authors. Société Internationale d'Urologie Journal, published by the Société Internationale d'Urologie, Canada.

Abbreviations

AMSTAR-2 Assessment of Multiple Systematic Reviews Version 2
AUA American Urological Association
EAU European Association of Urology
GRADE Grading of Recommendations Assessment, Development, and Evaluation
SRs systematic reviews

In a previous publication, we applied the updated Assessment of Multiple Systematic Reviews (AMSTAR-2) instrument[3] to assess the methodological quality of SRs in the urological literature from 2016–2018[4]. We found that only a small proportion of SRs achieved a "high" or "moderate" confidence rating according to AMSTAR-2 due to failure to meet basic quality criteria such as the provision of an a priori, registered protocol. We conducted this study to provide a longitudinal analysis of SRs quality over time by using identical methods to assess the subsequent 3 years (2019–2021).

Methods

This study was an extension of a prior published study[4] with a written a priori protocol. Given its focus on methodology, it could not be registered in PROSPERO. We performed a comprehensive search in PubMed that indexes all 5 included journals (*BJU International, European Urology, The Journal of Urology, Urology,* and *World Journal of Urology*) to identify all SRs published either electronically or in print between January 1, 2016, and December 31, 2018, by using the SR search filter under Clinical Queries. We extended this search through June 30, 2021, with an overlapping search start date of 3 months (October 1 to December 31, 2018) to make sure we included all SRs from that year that may have been indexed late. The references were then imported into Rayyan, a dedicated online software program for study screening. We included studies that self-identified themselves as SRs in the title, abstract, or methods and were related to clinical questions of therapy and/or prevention. Individual participant data meta-analyses and co-published, much abbreviated, Cochrane reviews were included. Health technology assessment reviews by a funding agency (eg, National Institute for Health and Care Research) were also included. We excluded SRs that related to diagnostic test accuracy studies, prognosis, or cost-effectiveness. We further excluded narrative reviews (see **Table 1** for differences from systematic reviews) and clinical practice guidelines.

Three reviewers (M.D., J.J., and G.A.A.) working in pairs independently screened references in duplicate in 2 stages (title/abstract and full-text stage). Discrepancies were resolved by discussion and consensus; select discrepancies were resolved in consultation with the senior author (P.D.). Data abstraction was similarly performed by 2 of 3 reviewers (M.D., J.J., and G.A.A.) independently and in duplicate using a dedicated Google form based on the AMSTAR-2 instrument that was pilot-tested on 2 sets of 10 SRs up front. Discrepancies were once again resolved by discussion and consensus and secondary arbitration by the senior author (P.D.) in select cases. Each individual item was scored as "met," "partially met," or "not met," as per AMSTAR-2 guidance. For the confidence ratings, we collapsed the "met" and "partially met" categories. Individual studies were then rated as "high," "moderate," "low," or "critically low" quality based on the extent to which studies were determined to have met the 7 critical domains and 9 non-critical domains according to the AMSTAR-2 scoring guidance[3].

The primary outcome of this study was the proportion of studies classified into these 4 categories. Given the paucity of studies classified as "high" or "moderate,"

TABLE 1.

Differences between narrative and systematic reviews

Features	Narrative Review	Systematic Review
Question	Broad	Specific
Methods	Usually not specified	Prespecified in a written (ideally registered) protocol
Sources	Often not specified, focused on published literature alone, potentially biased	Comprehensive/exhaustive of multiple databases (published and unpublished literature), explicit search approach
Selection of included studies	Subjective, potentially biased	Criterion-based approach, uniformly applied
Evaluation	Variable	Criterion-based approach, rigorous critical evaluation
Synthesis	Usually only qualitative	Often quantitative (when appropriate)
Inferences	Sometimes evidence-based	Usually evidence-based

we collapsed these categories for some of the reporting. We used descriptive statistics to calculate proportions and corresponding confidence intervals. Pre-planned statistical analysis was performed by time period and journal of publication using SPSS, version 27.0 with a predefined alpha of 0.05.

Results

Our search identified 563 references for the 2019–2021 time period, of which 114 ultimately met inclusion criteria (see **Figure 1** for PRISMA flow diagram; see **Online Supplement 1** for lists of included and excluded studies). The largest contributor by journal of publication for this time period was the *World Journal of Urology* (33; 28.9%), whereas it had been *European Urology* (53; 36.8%) from 2016 to 2018 (**Table 2**). Oncology (53; 46.5%) and voiding

FIGURE 1.

PRISMA flow diagram of literature search

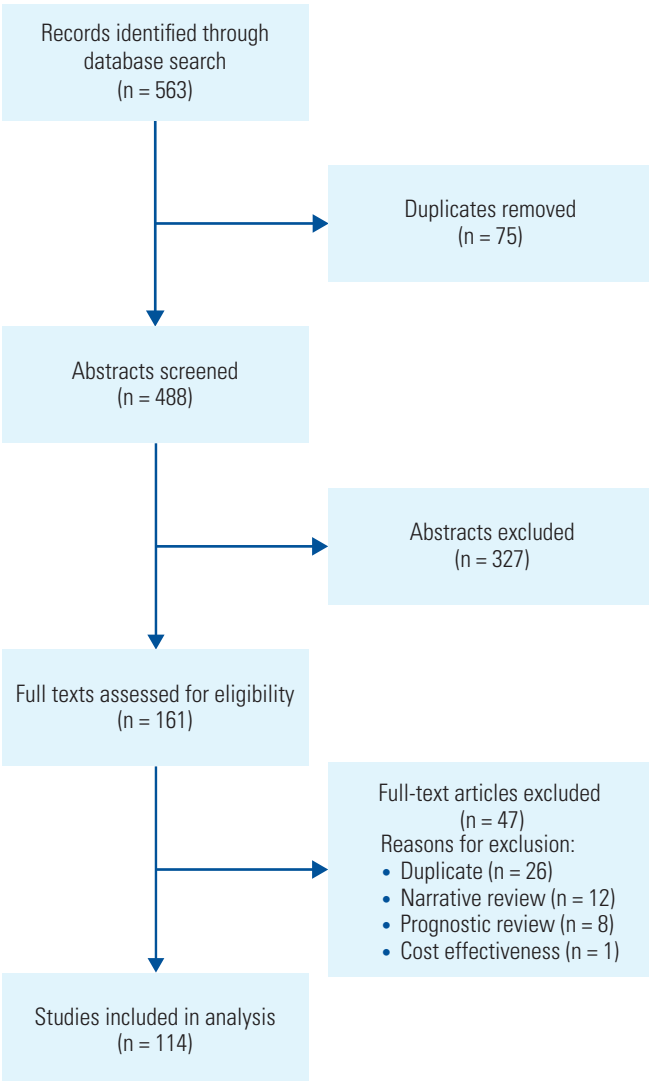


TABLE 2.

Table of baseline characteristics of 258 systematic reviews published between 2016 and 2021

Year of publication	Number (%)
2016	31 (12.0)
2017	42 (16.3)
2018	71 (27.5)
2019	32 (12.4)
2020	29 (11.2)
2021	53 (20.5)
Topic	
Oncology	117 (45.3)
Stones/Endourology	22 (8.5)
Laparoscopy	9 (3.5)
Trauma/Reconstruction	16 (6.2)
Voiding dysfunction	53 (20.5)
Infection/Inflammation	23 (8.9)
Infertility/Erectile dysfunction	18 (7.0)
Journal of publication	
BJU International	39 (15.1)
European Urology	78 (30.2)
Journal of Urology	31 (12.0)
Urology (Gold Journal)	53 (20.5)
World Journal of Urology	57 (22.1)
Included study designs	
Randomized controlled trials	91 (35.3)
Non-randomized controlled trials	44 (17.1)
All study designs	123 (47.7)
Number of included studies	
1–10	77 (29.8)
11–20	70 (27.1)
> 20	111 (43.0)

dysfunction (21; 18.4%) remained the 2 leading topic areas. Compared with the earlier time period, a larger proportion of studies (58.8% versus 38.9%; $P < 0.001$) included both randomized and non-randomized study designs. The number of included studies per SR was similar with medians of 16 (interquartile range: 9 to 31) and 20.0 (interquartile range: 8.75 to 42.75; $P = 0.264$) for the 2 time periods.

Figure 2 summarizes the reporting of the 7 critical AMSTAR-2 criteria comparing the 2 time periods. Overall, across both time periods, criterion #4, which refers to the comprehensiveness of the search, was met by the largest number of studies (241; 93.4%). Further details are summarized in Table 3: Of the 8 sub-criteria related to the comprehensiveness of the literature search, 2 of them, related to the search of trial registries (67.5% versus 31.9%; $P < 0.001$) and the search of the grey literature (37.7% versus 23.6%; $P = 0.014$) significantly improved. Criterion #7, which relates to the explanation and explicit referencing of all studies excluded at the full-text stage, was met by the lowest number of studies (28; 10.2%). Methodological quality changed significantly over time for 2 criteria: the proportion of studies with an a priori registered protocol (criterion #2) rose from 36.1% to 59.6% ($P < 0.001$) whereas those that considered the risk of bias when interpreting the study results (criterion #9) declined from 76.4% to 50.7% ($P < 0.001$). Other relatively large changes, for example, the decline in the proportion of studies assessing for the presence and likely impact of publication bias (criterion #15) were not statistically significant (here: $P = 0.106$).

With respect to non-critical criteria, criterion #1 that refers to an explicit PICO question was met by nearly all SRs (256; 99.2%) whereas criterion #10 which relates to the

FIGURE 2.

Proportion of studies meeting critical AMSTAR-2 criteria comparing 2016–18 with 2019–2021

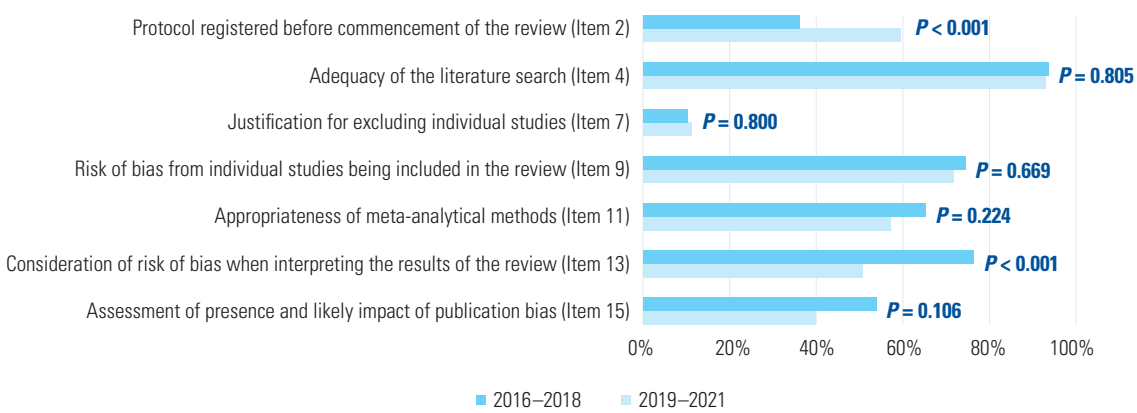


TABLE 3.

Details of reported literature search informing adequacy of the literature search (criterion #4)

Sub-criteria related to adequacy of search	Number met/Applicable SR (%)		
	2016–2018	2019–2021	*P*-value
Minimum of 2 databases searched	96.5	94.7	0.480
Provided key words and/or search strategy	97.9	95.6	0.289
Justified search strategy restrictions (eg, language)	40.3	36.0	0.479
Searched references/bibliographies	61.8	69.3	0.210
Searched trial/study registries	31.9	67.5	< 0.001
Included/consulted content experts	20.8	21.1	0.966
Searched "grey literature"	23.6	37.7	0.014
Search conducted within 24 months of review completion	97.2	96.5	0.737

reporting of the funding sources of the individual studies being synthesized was met by the lowest number of SRs (19; 7.4%; **Figure 3**). Statistically significant changes over time in terms of improvement were seen for criterion #3 (explanation provided for included study designs) which improved from 54.9% to 73.7% ($P = 0.001$). Methodological quality declined based on criterion #8 (description of included studies in adequate detail) and criterion #12 (potential impact of risk of bias assessed if a meta-analysis was performed), which declined from 88.2% to 65.8% ($P < 0.001$) and 74.5% to 56.4% ($P = 0.008$), respectively.

Overall, only 6 (2.3%) and 9 SRs (3.5%), achieved a "high" (no critical weakness; up to one non-critical weakness) or "moderate" (no critical weakness; more than one non-critical weakness) confidence rating, respectively. Most SRs published had "very low" confidence ratings (195; 75.6%). The proportion of studies with a "high/moderate" confidence rating increased only slightly from 4.9% to 6.1% from 2016–2018 to 2019–2021($P = 0.481$). *BJU International* (6/39; 15.4%) and the *Journal of Urology* (3/31; 9.7%) had a higher proportion of SRs with a "high/moderate" confidence rating than the other 3 journals, namely the *World Journal of Urology* (3/57; 5.3%), *Urology* (1/53; 1.9%), and *European Urology* (1/78; 1.3%; **Figure 4**).

Discussion

Statement of principal findings

This study provides the first longitudinal assessment of SRs published in the urological literature based on the updated AMSTAR-2 instrument. The main finding of this study is that approximately 3 of 4 SRs published in 5 major general urology journals continue to have a "critically low" confidence rating. Only about 1 in 20 reviews had a "high" or "moderate" confidence rating, with some journals faring better than others. Among the 7 critical criteria of the AMSTAR-2 instrument, criterion #7, which refers to the need to fully reference and explain the exclusion of studies at the full-text literature review stage was met by approximately 1 in 10 studies and did not improve over time. Although less than half of SRs had an a priori protocol (criterion #2), compliance for this criterion improved considerably by 2019–2021 to nearly 6 in 10 studies.

Strengths and weaknesses of the study

This study represents an extension of a previously published study with a written protocol. The use of the same methods and data abstraction form, and the involvement of 2 of the authors of the prior study facilitated consistent data abstraction and criteria interpretation. The assessment of SR methodological quality was performed using the AMSTAR-2, a widely used, validated instrument[3]. We once again pilot-tested the data abstraction form and completed all assessments independently and in duplicate. One potential limitation was that investigators were not blinded to the journal of publication and authors for each study, which could have potentially biased the ratings. Only the SRs published in 5 major urology journals were assessed in this study, therefore omitting reviews published in other urology journals or non-specialty journals. Our goal was not to conduct an exhaustive evaluation of all SRs in urology but rather examine the quality of the reviews published in 5 of the most prominent journals in our field. The AMSTAR-2 instrument was designed and validated only to appraise the methodological rigor of SRs related to questions of therapy and prevention[3]. Therefore, the conclusions of this study are only applicable to these types of SRs and not, for example, those for questions of diagnosis, prognosis, or cost-effectiveness. Fortunately, SRs for

FIGURE 3.

Percentage Proportion of studies meeting non-critical AMSTAR-2 criteria comparing 2016–18 with 2019–2021

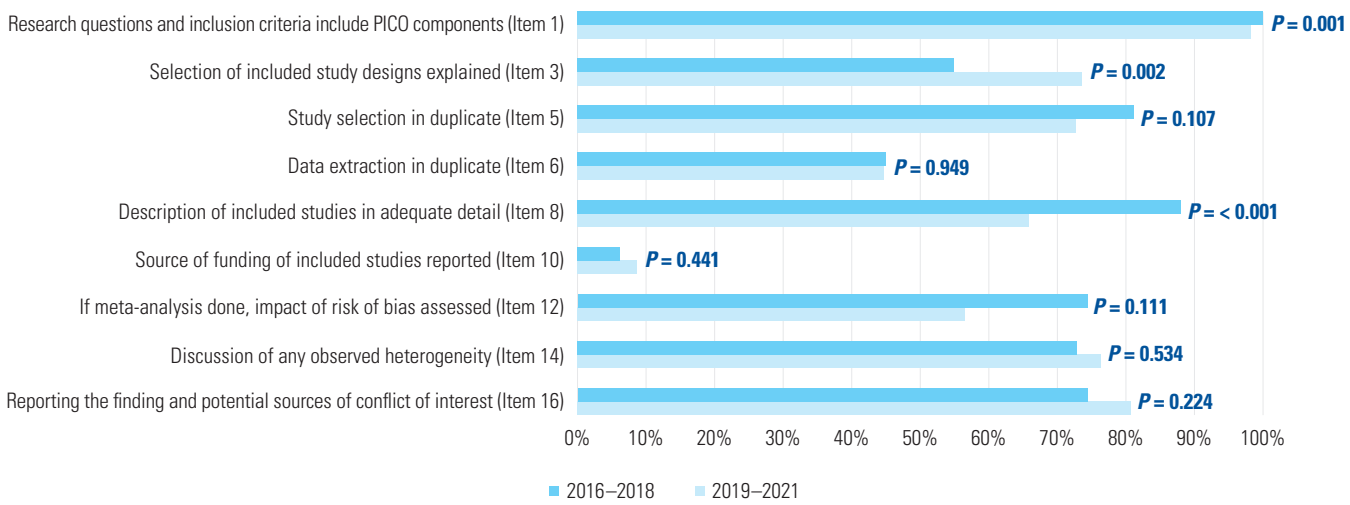
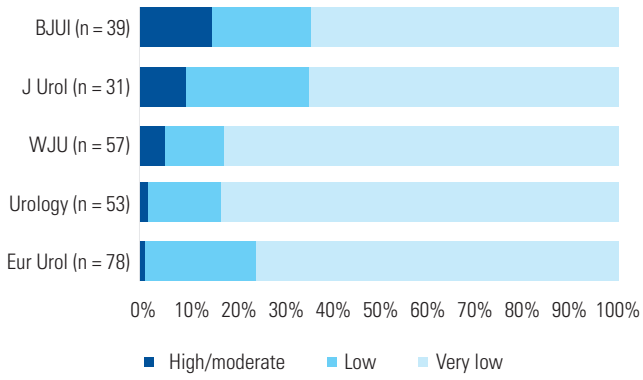


FIGURE 4.

Proportion of studies with "high/moderate", "low" or "critically low" confidence rating by journal of publication



questions of therapy and prevention are the most common form of evidence synthesis. Lastly, AMSTAR-2 assesses only the quality of the methodological "handiwork" used to conduct a given SR, not how confident we can be in its result, which is the focus of frameworks such as Grading of Recommendations Assessment, Development, and Evaluation (GRADE) that assess the confidence in the estimates of effect[5]. Therefore, an SR with a "high" or "moderate" confidence rating (reflecting the methodological rigor that has gone into its development) may nevertheless provide evidence of very low certainty (according to GRADE) by virtue of the methodological limitations of the included studies, as well as issues related to indirection, indirectness, imprecision, and possible publication bias[6].

Strengths and weaknesses in relation to other studies

Previous studies have longitudinally assessed the methodological quality of SRs published in the urological literature from 1998 to 2015, documenting both a large increase in the number of reviews published each year and their modest quality[6–8]. Han et al. found mean AMSTAR scores ± standard deviations of 4.8 ± 2.0, 5.4 ± 2.3 and 4.8 ± 2.4 for the 1998–2008 (n = 57), 2009–2012 (n = 113), and 2013–2015 (n = 125) time periods, respectively suggesting no improvement over time[8]. Ding et al.[4] were the first to apply the updated AMSTAR-2 instrument that introduced confidence ratings ("high," "moderate," "low," and "critically low") to replace the AMSTAR score on a scale of 0 to 11 scale (with higher scores reflecting higher methodological quality)and found that most SRs published in the urological literature had a "low" or "critically low" confidence rating; the current study indicates that this has not changed.

Two published studies have assessed systematic review quality of urology-relevant studies with a specific clinical focus area. O'Kelly et al.[9] assessed 227 SRs in clinical pediatric urology published from 1992 to 2018, thereby applying it retrospectively to a time period before the AMSTAR-2 instrument had been developed. They found no study with a "high" confidence rating. "Moderate," "low," and "critically low" ratings were determined in 15%, 65%, and 20% of SRs, respectively. These results would indicate that 80% of SRs had no more than one critical flaw, which would be considerably better than what we found in a general sample of more recent SRs from some of the same journals. This would suggest that pediatric urology authors are more compliant than those in other fields of urology. However, we also worry that this discrepancy may be due to different interpretation and application of the AMSTAR-2 criteria in scoring SRs. Second, Bole et al.[10] investigated 17 studies on Peyronie's disease and rated 65% (11/17) of studies to have a "critically low" confidence rating. These findings correspond largely with ours, as did the authors' conclusions that many SR's "fail to meet accepted methodological criteria"[10].

Studies from outside the field of urology using AMSTAR-2 have also found serious methodological limitations[11]. For example, Dettori et al. assessed 28 SRs from 2018 related to spinal surgery and rated most (26/28; 93%) as "critically low" and the remainder (2/28; 7%) as "low"[12]. Martinez-Monedoro et al. found that nearly all (95%) of SRs published in the 10 highest-impact otolaryngology journals from 2012 to 2017 were reviews of "critically low" confidence[13]. Yu et al. evaluated 141 SRs in surgery and found 2.8%, 2.1%, 5.7% and 89.5% to have a "high," "moderate," "low," and "critically low" confidence rating, respectively[14]. In a methodological review of SRs from 2010 to 2019 related to advanced cancer, most (230/261; 85.1%) had a "critically low" confidence rating[15]. This study also provided detailed information on what criteria were not met, which closely mirrored our findings: The majority (209/261; 80.1%) of studies were classified as "critically low" because of lack of an a priori protocol (criterion #2; 222, 85.1%) and failure to reference excluded full-text studies and providing justifications for exclusion (criterion #4; 218, 83.5%). Lastly, in a comparative study of recent updates of previously published SRs, 96.7% had a rating of "critically low"[16]. In aggregate, these studies provide compelling evidence that our findings are not unique to the specialty of urology, but the issue of poor SR quality is widespread in the health sciences.

Implications for clinicians and policymakers

SRs have a preeminent role in evidence-based clinical practice[17]: They not only are used by healthcare providers to inform decision-making for individual patients, but also provide the foundation for evidence-based clinical practice guidelines such as those by the American Urological Association (AUA) and the European Association of Urology (EAU) as well as policy

decision-making[18,19]. For that reason, they need to provide a trustworthy synthesis of the individual studies informing a given clinical question. Our study indicates that this is rarely the case. SRs do not provide reliable evidence summaries due to the authors' failure to adhere to established methodological standards. This contributes to avoidable research waste, with the potential to distract and misguide patient care decisions[2,20]. It therefore appears imperative to raise awareness for the criteria that determine the quality of a SR. These include a written protocol that is registered in advance and describes all relevant aspects of the SR development including the search, study selection, risk of bias assessment, meta-analysis (if appropriate) and interpretation. Ideally, the authors should report the certainty of evidence on a per-outcome basis using GRADE, which has become the de facto standard for determining the confidence in the estimates of effect and has been used in evidence synthesis underpinning the AUA guidelines[5]. It is encouraging to see that the proportion of SRs in the urological literature with a protocol has improved substantially, likely due to the explicit requirements by individual journals[21,22]. Introduction of an AMSTAR-2 checklist that editors could apply to every SR submission to any of these major urological journals might help address this issue.

Unanswered questions and future research

Multiple studies have highlighted the issue of low SR quality across many arenas of medicine including urology but appear to have been unable to impact a meaningful improvement. Future studies that identify the barriers to higher quality SR production and more rigorous evaluation on the part of authors and editors may be helpful for identifying barriers for change. Authorship analyses that seek to identify whether the involvement of individuals with specific methodological expertise or information specialist engagement[23] improves SR quality may yield opportunities for targeted interventions. Lastly, similar studies should continue to monitor the SR quality until there is assurance of consistent, higher quality content generation.

Conclusions

Most SRs published in the urological literature continue to have serious methodological limitations and should not be relied upon. There is a critical need for raising awareness among SR authors, journal editors, and consumers of the urological literature on established standards for high-quality SR to promote improvement.

Acknowledgements

Funding: Authors report no funding for this study.

Ethical Compliance: As a methodological review of the published literature this study was not subject to IRB approval.

Author Contributions in CRediT Format:

Maylynn Ding: Data abstraction, writing – review & editing; **Jared Johnson:** Data abstraction, writing, review, & editing; **Onuralp Ergun:** writing – review & editing**; Gustavo Ariel Alvez:** Data abstraction; **Philipp Dahm:** Conceptualization, analyses, writing – drafting, review & editing. **Data Availability:** The data collected and analyzed during this study are available from the corresponding author upon request.

References

1. Dickersin K. Health-care policy. To reform U.S. health care, start with systematic reviews. *Science*.2010 Jul 30;329(5991):516–517. doi: 10.1126/science.1189330

2. Ioannidis JP The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q.2016 Sep;94(3):485–514.doi: 10.1111/1468-0009.12210

3. 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 Sep 21;358:j4008. doi: 10.1136/bmj.j4008.

4. Ding M, Soderberg L, Jung JH, Dahm P. Low methodological quality of systematic reviews published in the urological literature (2016–2018). Urology.2020 Apr;138:5–10.doi: 10.1016/j.urology.2020.01.004.

5. Gonzalez-Padilla DA, Dahm P. Evidence-based urology: understanding GRADE methodology. *Eur Urol Focus*.2021 Nov;7(6):1230–1233. doi: 10.1016/j.euf.2021.09.014

6. Canfield SE, Dahm P. Rating the quality of evidence and the strength of recommendations using GRADE. *World J Urol*.2011 Jun;29(3):311–7. doi: 10.1007/s00345-011-0667-2.

7. 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 Nov;194(5):1374–9. doi: 10.1016/j.juro.2015.05.085

8. Han JL, Gandhi S, Bockoven CG, Nararyan VM, Dahm P. The landscape of systematic reviews in urology (1998 to 2015): an assessment of methodological quality. *BJU Int*.2017 Apr;119(4):638–649. doi: 10.1111/bju.13653

9. O'Kelly F, DeCotiis K, Aditya I, Braga LH, Koyle MA. Assessing the methodological and reporting quality of clinical systematic reviews and meta-analyses in paediatric urology: can practices on contemporary highest levels of evidence be built? J Pediatr Urol.2020 Apr;16(2):207–217. doi: 10.1016/j.jpurol.2019.12.002

10. Bole R, Gottlich HC, Ziegelmann MJ, Corrigan D, Levine LA, Mulhall JP, et al. a critical analysis of reporting in systematic reviews and meta-analyses in the peyronie's disease literature. J Sex Med.2022 Apr;19(4):629–640. doi: 10.1016/j.jsxm.2022.01.008.

11. Bojcic R, Todoric M, Puljak L. Adopting AMSTAR 2 critical appraisal tool for systematic reviews: speed of the tool uptake and barriers for its adoption. BMC Med Res Methodol.2022 Apr 10;22(1):104. doi: 10.1186/s12874-022-01592-y

12. Dettori JR, Skelly AC, Brodt ED. Critically low confidence in the results produced by spine surgery systematic reviews: an AMSTAR-2 evaluation from 4 spine journals. Global Spine J.2020;10:667–673. doi: 10.1177/2192568220917926

13. Martinez-Monedero R, Danielian A, Angajala V, Dinalo JE, Kezerian EJ. Methodological quality of systematic reviews and meta-analyses published in high-impact otolaryngology journals. Otolaryngol Head Neck Surg.2020 Nov;163(5):892–905.doi: 10.1177/0194599820924621

14. Yu J, Yang Z, Zhang Y, Cui Y, Tang J, Hirst A, et al. The methodological quality on systematic reviews of surgical randomised controlled trials: a cross-sectional survey. Asian J Surg.2022 Oct;45(10):1817–1822. doi: 10.1016/j.asjsur.2021.10.021

15. Siemens W, Schwarzer G, Rohe MS, Buroh S, Meerpohl JJ, Becker G. Methodological quality was critically low in 9/10 systematic reviews in advanced cancer patients-a methodological study. J Clin Epidemiol.2021;136: 84, 2021 2021 Aug;136:84–95 doi: 10.1016/j.jclinepi.2021.03.010

16. 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. J Clin Epidemiol.2020 Mar;119:36–46. doi: 10.1016/j.jclinepi.2019.11.012

17. Tsen g TY, Dahm P, Poolman RW, Preminger GM, Canales BJ, Montori VM, et al. How to use a systematic literature review and meta-analysis. *J Urol*.2008 Oct;180(4):1249–56 doi: 10.1016/j.juro.2008.06.046

18. Faraday M, Hubbard H, Kosiak B, Dmochowski R, et al. Staying at the cutting edge: a review and analysis of evidence reporting and grading; the recommendations of the American Urological Association. *BJU Int*.2009 Aug;104(3):294–297. doi: 10.1111/j.1464-410X.2009.08729.x

19. Knoll T, Omar MI, Maclennan S, Hernández V, Canfield S, Yuan Y, et al. Key steps in conducting systematic reviews for underpinning clinical practice guidelines: methodology of the European Association of Urology. *Eur Urol*.2018; Feb;73(2):290–300. doi: 10.1016/j.eururo.2017.08.016.

20. 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*.2014 Jan 18;383(9913):267–276. doi: 10.1016/S0140-6736(13)62228-X

21. Dahm P. Raising the bar for systematic reviews with Assessment of Multiple Systematic Reviews (AMSTAR). *BJU Int*.2017 Feb;119(2):193. doi: 10.1111/bju.13754

22. Jung JH, Dahm P. Reaching for the stars - rating the quality of systematic reviews with the Assessment of Multiple Systematic Reviews (AMSTAR) 2. *BJU Int*.2018 Nov;122(5):717–718. doi: 10.1111/bju.14571

23. Koffel JB. Use of recommended search strategies in systematic reviews and the impact of librarian involvement: a cross-sectional survey of recent authors. PLoS One.2015 May 4;10(5):e0125931. doi: 10.1371/journal.pone.0125931