Journal of Clinical Epidemiology 165 (2024) 111197

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

The 2-week systematic review (2weekSR) method was successfully blind-replicated by another team: a case study

Catalin Tufanaru[a,1], Didi Surian[a,1,*], Anna Mae Scott[b], Paul Glasziou[b], Enrico Coiera[a]

[a]*Australian Institute of Health Innovation, Level 6, Macquarie University, 75 Talavera Road, North Ryde, New South Wales 2109, Australia*
[b]*Institute for Evidence-Based Healthcare, Bond University, 14 University Drive, Robina, Queensland 4226, Australia*

Accepted 19 October 2023; Published online 23 October 2023

Abstract

Objective: To assess the replicability of a 2-week systematic review (index 2weekSR) created with the assistance of automation tools using the fidelity method.

Methods: A Preferred Reporting Items for Systematic reviews and Meta-Analyses compliant SR protocol was developed based on the published information of the index 2weekSR study. The replication team consisted of three reviewers. Two reviewers blocked off time during the replication. The total time to complete tasks and the meta-analysis results were compared with the index 2weekSR study. Review process fidelity scores (FSs) were calculated for review methods and outcomes. Barriers to completing the replication were identified.

Results: The review was completed over 63 person-hours (11 workdays/15 calendar days). A FS of 0.95 was achieved for the methods, with 3 (of 8) tasks only partially replicated, and an FS of 0.63 for the outcomes, with 6 (of 7) only partially replicated and one task was not replicated. Nonreplication was mainly caused by missing information in the index 2weekSR study that was not required in standard reporting guidelines. The replication arrived at the same conclusions as the original study.

Conclusion: A 2weekSR study was replicated by a small team of three reviewers supported by automation tools. Including additional information when reporting SRs should improve their replicability.

Keywords: 2-Week systematic review; Replication study; Automation tool; Replication fidelity; Barriers; Facilitators

1. Introduction

A recent study showed that a moderately sized systematic review (index SR) could be completed in 2 weeks (2weekSR) by protecting time to solely focus on SR tasks and supported by systematic review automation (SRA) tools [1]. Given that SRs are time and resource intensive that may require a median of five researchers and 41 weeks to submit to a journal [2], the 2weekSR study shows that it is feasible to complete an SR more efficiently assisted by automation tools. However, this important innovation currently has not been used widely to support the productivity of SRs and it has only been used by a few teams of researchers [3−5]. Given the novelty of the use of automation tools to produce 2weekSR efficiently, further assessment is important to ensure that the process is replicable and will not change SR conclusions.

More generally, testing the reproducibility of published research studies plays a crucial role in verifying the results of research [6−10]. There are many reasons for research results to be in error, including flaws in research methods, data collection, analysis, or publication [11−14]. Although exact replication may not be possible and replication may differ in innumerable ways from the original study, replication studies aim to achieve as much fidelity as possible to test whether the findings of previous studies were valid or were subject to experimental or reporting problems [15−17].

Therefore, this replication study aimed to:

1. Independently attempt to replicate the same methods used in a completed 2weekSR study to address the same research questions using only available information in published index SR and their supplementary materials [18].

Funding: None.

[1] Indicates joint first authorship.

* Corresponding author. Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Rd, New South Wales 2109, Australia. Tel.: +61-2-9850-2400; fax: +61-2-9850-2499.

E-mail address: didi.surian@mq.edu.au (D. Surian).

https://doi.org/10.1016/j.jclinepi.2023.10.013

What is new?

Key findings

- A 2-week systematic review (2weekSR) with the assistance of automation tools was successfully replicated with a high degree of fidelity score for the methods and a modest fidelity for the outcomes.

What this adds to what was known?

- The use of automation tools can improve the efficiency of producing SR.
- We describe methods to assess the replication fidelity of 2weekSR.

What is the implication and what should change now?

- Current reporting guidelines could be extended to support SRs assisted by automation tools.
- Sufficient information reported in published SRs will improve their replicability.

2. Compare the replication study with the index 2weekSR study for time to complete review tasks and review outcomes, that is, meta-analysis results.
3. Ascertain the achieved fidelity for review methods and identified outcomes.
4. Identify barriers to replicating automation-assisted reviews, including the completeness and accuracy of method descriptions in study protocols.

2. Materials and methods

2.1. Systematic review replication process

2.1.1. Accessing the original protocol

The index SR study selected for replication examined the benefits of increased fluid intake to prevent urinary tract infection (UTI). The review was reported in a clinical journal, and additional technical information about its methods appeared in a second article [1,19]. Supplementary files associated with the second methodological paper were unavailable online and were requested from and provided by the corresponding author (A.M.S.) of that report. This supplementary material is made available in Supplementary Appendix 3. The authors of the index 2weekSR study (A.M.S., P.G.) were only informed of the replication study after the replication was completed to ensure the replication was independent and proceeded only based on published information.

2.1.2. SR replication protocol

To conduct the replication study, a Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2009 compliant SR protocol was developed that described the different steps and associated methods reported in the index review, covering study selection (screening), data extraction, risk of bias assessment, and data synthesis (Supplementary Appendix 2) [20–25]. The SR replication protocol was used as a guide in replicating the systematic review. The SR replication protocol had the following intentional deviations from the index 2weekSR study:

1. Tasks related to the writing of a systematic review manuscript, submission to a journal, and revision after peer review were not part of the replication assessment [1].
2. The replication team consisted of three individuals, while the index SR study had four reviewers. Two reviewers (C.T., D.S.) identified the review questions and Population, Intervention, Comparison, and Outcome (PICO) elements, and undertook screening, data extraction, and risk of bias assessment. The third reviewer (E.C.) resolved conflicts in the interpretation of methodology or outcomes of review steps in the protocol.
3. The review questions and PICO were formulated by two reviewers (C.T., D.S.) based on information extracted from the abstract and method sections of the index study, while for the index SR study these were designed by clinician researchers.

2.1.3. Replication project management

As with the index 2weekSR study, two reviewers (C.T., D.S.) blocked off their time from other projects during the replication and focused on the completion of SR tasks planned for each day. Daily administrative meetings were held to resolve any issues in each SR task. The time needed by each reviewer to complete an SR task was recorded. A list of SR tasks and automation tools used to assist is provided in Supplementary Appendix 1.

2.1.4. Inclusion criteria

We conducted a systematic review of randomized controlled trials (RCTs) to assess the impact of increased fluid intake on UTI recurrence, antimicrobial use, and UTI symptoms in individuals at risk for UTIs. We included RCTs targeting individuals at risk for UTIs of any age and sex, and who were ambulatory (i.e., noncatheterized). To be included, RCTs needed to assess the impact of increased fluids, for example, water, D-mannose dissolved in fluid, or juice. We excluded RCTs if controls used antimicrobials or cranberry in nonliquid form (tablet, powder, supplement, or fruit).

2.1.5. Search strategy

A systematic search of the literature was performed, limited to studies published up to January 21, 2019, the date when the last database searches were performed in the index SR study, with no language restriction. The index SR study developed a search strategy by conducting a word-frequency analysis of seven manually selected target reviews, using these keywords to draft a search in PubMed/Medline, and then refining the search using the SearchRefiner tool [26]. The index SR study did not provide details of these seven studies, so a search strategy was designed from scratch by the replication team (C.T.) (Supplementary Appendix 1).

Next, a preliminary search was conducted looking for existing SRs using the PubMed, Epistemonikos, and PROSPERO databases. Finally, full searches were undertaken on three databases (PubMed, Cochrane CENTRAL, and EMBASE) and two clinical trial registries (Clinicaltrials.gov and WHO ICTRP). The EMBASE search strategy was modified to accommodate the Ovid search platform instead of using Embase.com, which was used in the index SR team.

2.1.6. Screening and data extraction

Two reviewers (C.T., D.S.) independently screened titles and abstracts against the inclusion criteria. The reviewers retrieved full texts and screened these for inclusion. Any disagreements were resolved by discussion or reference to the third reviewer. Forward and backward citation analyses were performed using SCOPUS.

A data extraction form for study characteristics and outcome data was designed (C.T.) utilizing the structure of tables in the index SR study, as described in the index SR materials. Two reviewers (C.T., D.S.) extracted the following data from included studies: authors, year, country, number of randomized participants, mean age in years, percentage of female participants, UTI history for trial participants, intervention details, comparators, duration of follow-up in months, primary outcomes (UTIs), and secondary outcomes (antimicrobial use, UTI symptoms).

2.1.7. Risk of bias

Two reviewers (C.T., D.S.) independently assessed the risk of bias for each included study using the criteria outlined in the Cochrane Handbook, and as described in the index SR study. Results from an automated tool for assessing the risk of bias, RobotReviewer [27], were compared with the results from a manual risk of bias assessment. All disagreements were resolved by consensus or by referring to the third reviewer (E.C.).

2.1.8. Data synthesis and meta-analysis

A meta-analysis was performed independently by two reviewers (C.T., D.S.). Review Manager 5 was used to calculate the treatment effect. Odds ratios or rate ratios for dichotomous outcomes were used. A random effects model was used and I^2 statistic was calculated to measure the heterogeneity of the included trials. Different time points ($\leq$6 months and 12 months) were compared for the primary outcome and the number of patients with UTIs.

2.2. Replication fidelity

The fidelity of replication was assessed in two SR domains, that is, methods and outcomes and was characterized using a fidelity score (FS) [28]. Eight activities in the methods domain were scored: i) search strategy, ii) database searches, iii) deduplicating of search results, iv) screening, v) citation analysis, vi) data extraction, vii) risk of bias assessment, and viii) synthesis and meta-analysis. Seven activities in the outcomes were analyzed: i) publications identified in searches, ii) publications identified through forward and backward citation analyses, iii) included studies, iv) automated risk of bias assessment, v) manual risk of bias assessment, vi) data extraction from included studies, and vii) meta-analysis.

The FS was calculated using one of five fidelity ratings: 4 (identical), 3 (substitutable), 2 (in-class), 1 (augmented), and 0 (not replicated) [28]. The FSs were calculated for each of the eight activities in the methods domain and the seven activities in the outcomes domain. For every activity in each domain, raw scores were averaged if there were more than one value (e.g., identical (4) plus augmented (1) would receive a score of 2.5). The total FS for each domain was calculated by dividing the total raw scores by the total maximum possible raw score for the respective domain.

2.3. Facilitators and barriers

Facilitators and barriers to completing the 2weekSR replication study were mapped onto four domains in the Theoretical Domains Framework, that is, i) knowledge; ii) memory, attention, and decision process; iii) environmental context and resources; and iv) technology and infrastructure.

3. Results

The systematic review replication was completed over 15 calendar days (11 workdays between November 2 and November 16, 2020) (Table 1), compared to 12 calendar days (9 workdays between January 21 and February 1, 2019) in the index 2weekSR study. Reviewers (C.T., D.S.) took 63 person-hours (3,809 person-minutes) to complete the replication of 2weekSR study. Table 1 reports the time required for each SR task replication.

The replication team identified six included studies [29−34]. All these six studies were also among the eight included studies in the index SR study [29−36]. Based on the content in the manuscripts, the replication team found that the other two studies [35 36] did not satisfy

Table 1. Reviewers and allocated time for each SR task

		The index 2weekSR [1]		Our replication 2weekSR	
Task no.	Tasks	Number reviewers involved	Total person time in mins (percent)	Number reviewers involved	Total person time in mins (percent)
0	Daily administrative meetings	4	380 (21.3%)	2	130 (3.4%)
1	Formulate review question	4	120 (3.7%)	2	38 (1.0%)
2	Find previous/upcoming SRs	1	63 (2.0%)	1	60 (1.6%)
3	Write the protocol	4	175 (5.4%)	2	63 (1.6%)
4	Design systematic search	1	109 (3.4%)	1	208 (5.5%)
5	Design data extraction form and pilot	1	72 (2.2%)	1	70 (1.8%)
6	Run systematic search	1	72 (2.2%)	1	133 (3.5%)
7	Deduplicate results	1	16 (0.5%)	1	66 (1.7%)
8	Screen abstracts	2	404 (12.6%)	2	324 (8.5%)
9	Obtain full text	2	41 (1.3%)	2	162 (4.3%)
10	Screen full text	2	187 (5.8%)	2	264 (6.9%)
11	Screen trial registries	2	123 (3.8%)	2	324 (8.5%)
12	Citation analysis	1	30 (0.9%)	1	68 (1.8%)
13	Screen citation analysis	2	171 (5.3%)	2	326 (8.6%)
14	Extract data	2	461 (14.3%)	2	832 (21.8%)
15	Assess risk of bias	2	323 (10.0%)	2	641 (16.8%)
16	Synthesis and meta-analysis	2	167 (5.2%)	2	100 (2.6%)
	Total	4	3,219 person-minutes (54 person-hours)	2	3,809 person-minutes (63 person-hours)

the inclusion criteria, that is, interventions that involved an increased fluid intake. In one study [35], the additional information about interventions that involved an increased fluid intake was not explicitly mentioned in the manuscript and it was only available in the brochures that the authors distributed to the participants, which also, were not publicly available. The brochure was made available to the reviewers of the index SR study after they contacted the authors of the respective paper. The replication team was unaware of this additional information as the replication team did not request the brochure from the respective manuscript.

Of the six studies, one study was not meta-analyzed as it did not report the impact of increased fluid intake [34]. The replication team initially identified nine additional studies [37–45]. However, these nine additional studies were later excluded after the full-text screening step as they only involved interventions with the same amount of fluid. These nine additional studies were also not reported in the index SR. Table S2.4 in Supplementary Appendix 2 compares the included studies in the index 2weekSR and the replication study.

3.1. Replication fidelity results

A high degree of fidelity (FS 0.95) in the systematic review methods indicates that the replication study design was very similar to the index 2weekSR study's methods (Table 2). Three (of 8) activities could only be partially replicated because of insufficient information in the original SR, such as information about the initial set of seven potentially relevant articles used in the original SR to generate the keywords, terms identified through word frequency analysis, and subject terms (MeSH terms) could not be identified, inaccessible Embase.com from our university, and insufficient information about the exact "PICO in title" screening technique used in the index 2weekSR. The other five tasks in methods were replicated. Fidelity was modest (FS 0.63) for the systematic review outcomes, with the majority of results (6 out of 7) partially replicated and one (results of the automated risk of bias assessment using RobotReviewer) not replicated (Table 2). The results of the automated risk of bias assessment using RobotReviewer and data extracted from included studies could not be replicated because of insufficient information from the index 2weekSR study. Supplementary Appendix 2 provides details of what information that was insufficient for the replication team to perform exact replication and the raw scores for the fidelity results.

3.2. Facilitators and barriers

The identified facilitators and barriers (Supplementary Appendix 1) in the Theoretical Domains Framework were similar to those reported in the index 2weekSR study [1].

Table 2. Replication fidelity for the systematic review methods and outcomes

Domain	Raw score [28]	Description
Methods		
1. Developing a search strategy	3.0	The search strategy from the index SR study for PubMed and Cochrane CENTRAL was used, while a modified search strategy was used for EMBASE in the replication study.
2. Systematic databases searching	3.7	The same platforms were used for PubMed, Cochrane CENTRAL, and Clinical Trial Registries, while the Ovid platform (rather than Embase.com as in the index SR study) was used for EMBASE search.
3. Deduplicating search results using SRA de-duplicator	4.0	Same method.
4. Screening	3.7	RobotSearch was used to identify RCTs as in the index 2weekSR study. The standard manual screening was performed by two authors (C.T., D.S.) to screen the articles (instead of using the specific 'PICO in title' screening technique as mentioned in the index 2weekSR study).
5. Citation analysis	4.0	Same method.
6. Data extraction	4.0	Same method.
7. Risk of bias assessment	4.0	Same method.
8. Synthesis and meta-analysis	4.0	Same method.
Total raw score	30.4	
Fidelity score	30.4/32 = 0.95	
Outcomes		
1. Publications identified through databases searches	1.5	The index SR study reported 1,081 publications retrieved from databases searches, while there were 799 publications in this replication study. Information on search results from Clinical Trial Registries was not available in the index 2weekSR study.
2. Publications identified through forward and backward citations analyses	3.0	The index SR study reported 613 publications identified (before deduplication), while there were 875 publications in this replication study.
3. Included studies	2.7	There were eight included studies in the index SR study. The replication study also identified the eight included studies and nine new studies. Only six studies were included, whereas five studies were meta-analyzed. One same study was not meta-analyzed as in the index SR study.
4. Results of the automated risk of bias assessment using RobotReviewer	0.0	Information on the results generated by RobotReviewer could not be identified in the index SR study.
5. Results of the manual risk of bias assessment	3.5	The overall percentages across all included studies were reported as in the index SR study and can be compared to the results from the replication study.
6. Data extracted from included studies	3.5	Some extracted data were the same as the index 2weekSR study.
7. Meta-analysis results	3.5	The results of the meta-analysis were almost identical to the index SR study, and the conclusions of SR were not changed.
Total raw score	17.7	
Fidelity score	17.7/28 = 0.63	

Abbreviations: SRA, systematic review automation; PICO, population, intervention, comparison, and outcome.

In the knowledge and skills domains, the replication team (C.T., D.S., E.C.) had extensive experience in performing systematic reviews, and one of the reviewers (C.T.) was an accredited JBI Comprehensive Systematic Review (CSR) trainer, which was an important contribution to the methodological expertise in the replication. Two replication team members (C.T., E.C.) had a clinical background, which contributed to the clinical understanding of the review topic. The ability to understand the Italian language of one reviewer (E.C.) helped verify data extracted from an article in Italian.

In the memory, attention, and decision process domain, the replication team found difficulties in fully blocking off time to focus solely on this project compared to the index 2weekSR study. However, the replication team focused predominantly on the replication study. The barriers in the environmental context included noisy surroundings were also experienced by reviewers in the index 2weekSR study.

Table 3. Factors that hampered the replicability and their requirement in reporting guidelines

Domain	Factors	Reporting guidelines
Methods	The provenance of search terms used in the search strategy	Not required
	List of studies used for the forward and backward citation analysis	Not required
	Details of list of studies/trials before deduplication	Not required
	Details of list of studies used for automated screening (e.g., RobotSearch) for identifying randomized controlled trials (RCTs)	Not required
Outcomes	Search results retrieved from each database	Not required
	The number of references identified from citation searching	Required by PRISMA 2020
	Results of the automated risk of bias assessment (e.g., using RobotReviewer)	Not required

In the resources domain, the unavailability of access to Embase.com was a barrier. The technology and infrastructure barriers involved were similar to those reported in the index 2weekSR study, that is, nonfunctional website (i.e., RobotReviewer was not functional for some time during the replication process; A RobotReviewer run by one of the replication reviewers [C.T.] reported a non-RCT study for one of the included studies but reported an RCT study when it was run by the other reviewer [D.S.]); software incompatibilities (i.e., SRA Helper tool did not work in EndNote); and common operating system/software issues (i.e., Windows and Microsoft Office updates added time to the process and introduced problems after installing the updates).

3.3. Factors that hampered the replicability

Nonreplicable elements were mainly due to missing information in the index SR study (Table 3). This information is currently not required by standard reporting guidelines such as PRISMA 2009, PRISMA 2020, and PRIMA-S. For example, information about the initial set of seven potentially relevant articles to generate the keywords and MeSH terms was not reported in the index SR study. Consequently, the reported search strategy for PubMed and Cochrane CENTRAL databases was used directly in the replication process. The search strategy for the EMBASE database was modified to accommodate the OVID platform because the replication team did not have access to the EMBASE platform used in the original SR. Screening of articles and trials was not based on the ''PICO in title'' process reported in the original SR due to insufficient information about the exact keywords used. None of the modifications made in the replication study changed the results or conclusions of the original systematic review.

4. Discussion

4.1. Limitations and risks of bias

This replication study has several limitations. First, it did not replicate tasks related to writing the manuscript, submission to a journal, and revision after peer review (Table 1). The replication introduced other deviations from the index 2weekSR study [1], such as the number of reviewers and the formulation of the review questions and PICO. Result deviations from the original study were also observed in testing the reproducibility of SRs [46]. Although these deviations did not affect the conclusions of the replicated SR, there may have been biases that impacted the reported results.

Second, although all the authors of the index 2weekSR study were not informed at the beginning of the replication process, the replication team member E.C. knew A.M.S. and P.G. before the replication process. This may introduce biases as the replication team might have been exposed to information that could shape this study.

Third, although the authors of the index 2weekSR study did not participate in the replication process at any point from conceptualization to meta-analysis, they were involved in writing this manuscript and provided clarification on issues. There is a risk that this could bias the analysis, for example, the exclusion of studies that were initially included which changed the meta-analysis but not the SR conclusions.

Fourth, the replication team members (C.T., D.S., E.C.) had access to the index 2weekSR and the original SR prior to the replication study. Although the review questions were developed by the replication team only based on the methods in the Abstract, Background, and Method sections of the original SR, there is a risk that the information from the Results sections might be exposed to the replication team, which could bias the replication process. A stricter replication process is needed to minimize the risk of bias.

4.2. Future directions

Standard reporting guidelines provide a minimum set of items for reporting systematic reviews, meta-analysis of randomized trials, and other types of reviews. As these guidelines are continuously updated to mitigate replicability issues, for example, PRISMA 2020 is more comprehensive than PRISMA 2009, the most recent reporting guidelines should be used. Work is also needed to extend the current reporting guidelines to support SRs assisted by automation tools.

Standard methods for more accurate comparisons between the index and replication studies are also important to assess the fidelity of replication studies. This study only assesses two domains, that is, eight tasks in methods and seven outcomes, using fidelity ratings reported by Coiera and Tong (2021) [28]. Further refinements in assessing domains of interest will better characterize the degree of replication fidelity.

5. Conclusions

The outcomes of a moderate size SR completed in 2 weeks by four reviewers supported by automation tools were able to be replicated in 2 weeks by a small team of three reviewers. Sufficient information reported in published SRs will improve replicability.

CRediT authorship contribution statement

Catalin Tufanaru: Conceptualization, Methodology, Writing − original draft, Validation, Formal analysis, Investigation, Resources, Data curation. **Didi Surian:** Writing − original draft, Validation, Formal analysis, Investigation, Resources, Data curation, Visualization, Writing − review & editing. **Anna Mae Scott:** Writing − review & editing. **Paul Glasziou:** Writing − review & editing. **Enrico Coiera:** Conceptualization, Methodology, Writing − review & editing, Supervision, Project administration.

Data availability

All data relevant to the study are included in the article or uploaded as online supplementary material. All data relevant to the analysis are reported in the article.

Declaration of competing interest

A.M.S. and P.G. are the coinventors of the 2weekSR method but were not aware of and not involved in the replication study at the time of its conduct.

Acknowledgments

This work is dedicated to the memory of Catalin Tufanaru, who passed away on July 30, 2021, while this manuscript was being prepared. We thanked Jeremy Cullis from Macquarie University Library for help in designing a comprehensive EMBASE search strategy.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2023.10.013.

References

[1] Clark J, Glasziou P, Mar CD, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. J Clin Epidemiol 2020;121:81−90.
[2] Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7(2):e012545.
[3] Yadav K, Krzyzaniak N, Alexander C, Scott AM, Clark J, Glasziou P, et al. The impact of antibiotics on clinical response over time in uncomplicated cellulitis: a systematic review and meta-analysis. Infection 2022;50(4):859−71.
[4] Chaboyer W, Coyer F, Harbeck E, Thalib L, Latimer S, Wan CS, et al. Oedema as a predictor of the incidence of new pressure injuries in adults in any care setting: a systematic review and meta-analysis. Int J Nurs Stud 2022;128:104189.
[5] Meloncelli N, Young A, Christoffersen A, Rushton A, Zhelnov P, Wilkinson SA, et al. Co-designing nutrition interventions with consumers: a scoping review. J Hum Nutr Diet 2023;36:1045−67.
[6] Landy JF, Jia ML, Ding IL, Viganola D, Tierney W, Dreber A, et al. Crowdsourcing hypothesis tests: making transparent how design choices shape research results. Psychol Bull 2020;146(5):451−79.
[7] Platt JR. Strong Inference: certain systematic methods of scientific thinking may produce much more rapid progress than others. Science 1964;146:347−53.
[8] Zwaan RA, Etz A, Lucas RE, Donnellan MB. Making replication mainstream. Behav Brain Sci 2017;41:e120.
[9] Nosek BA, Lakens D. Registered reports: a method to increase the credibility of published results. Soc Psychol 2014;45:137−41.
[10] Nakagawa S, Parker TH. Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. BMC Biol 2015;13:88.
[11] Coiera E, Ammenwerth E, Georgiou A, Magrabi F. Does health informatics have a replication crisis? J Am Med Inform Assoc 2018;25:963−8.
[12] Coles NA, Tiokhin L, Scheel AM, Isager PM, Lakens D. The costs and benefits of replication studies. Behav Brain Sci 2018;41:e124.
[13] Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 2011;10(9):712.
[14] Jr CMF, Huivin R, Aranda L, Pandis N, Alarcon M. The search and selection for primary studies in systematic reviews published in dental journals indexed in MEDLINE was not fully reproducible. J Clin Epidemiol 2018;98:53−61.
[15] Schmidt S. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. Rev Gen Psychol 2009;13(2):90−100.
[16] Ioannidis JPA. Why most published research findings are false. PLoS Med 2005;2(8):e124.
[17] Page MJ, Moher D, Fidler FM, Higgins JP, Brennan SE, Haddaway NR, et al. The REPRISE project: protocol for an evaluation of REProducibility and Replicability in Syntheses of Evidence. Syst Rev 2021;10(1):112.
[18] Puljak L, Pieper D. Replicability in the context of systematic reviews: a call for a framework with considerations regarding duplication, overlap, and intentionality. J Clin Epidemiol 2022;142:313−4.
[19] Scott AM, Clark J, Mar CD, Glasziou P. Increased fluid intake to prevent urinary tract infections: systematic review and meta-analysis. Br J Gen Pract 2020;70(692):e200−7.

[20] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.
[21] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264−9.
[22] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097.
[23] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Open Med 2009;3:e123−30.
[24] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006−12.
[25] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2009;8(5):336−41.
[26] searchrefiner: A query visualisation and understanding tool for systematic reviews. Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM), page 1839-1942. https://doi.org/10.1145/3269206.3269215
[27] Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc 2016;23:193−201.
[28] Coiera E, Tong HL. Replication studies in the clinical decision support literature−frequency, fidelity, and impact. J Am Med Inform Assoc 2021;13:1815−25.
[29] Kranjčec B, Papeš D, Altarac S. D-mannose powder for prophylaxis of recurrent urinary tract infections in women: a randomized clinical trial. World J Urol 2014;32(1):79−84.
[30] Hooton TM, Vecchio M, Iroz A, Higgins JP, Brennan SE, Haddaway NR, et al. Effect of increased daily water intake in premenopausal women with recurrent urinary tract infections: a randomized clinical trial. JAMA Intern Med 2018;178(11):1509−15.
[31] Kontiokari T, Sundqvist K, Nuutinen M, Pokka T, Koskela M, Uhari M. Randomised trial of cranberrylingonberry juice and Lactobacillus GG drink for the prevention of urinary tract infections in women. BMJ 2001;322:1571.
[32] Ferrara P, Romaniello L, Vitelli O, Gatto A, Serva M, Cataldi L. Cranberry juice for the prevention of recurrent urinary tract infections: a randomized controlled trial in children. Scand J Urol Nephrol 2009;43(5):369−72.
[33] Handeland M, Grude N, Torp T, Slimestad R. Black chokeberry juice (Aronia melanocarpa) reduces incidences of urinary tract infection among nursing home residents in the long term–a pilot study. Nutr Res 2014;34(6):518−25.
[34] Stapleton AE, Dziura J, Hooton TM, Cox ME, Yarova-Yarovaya Y, Chen S, et al. Recurrent urinary tract infection and urinary Escherichia coli in women ingesting cranberry juice daily: a randomized controlled trial. Mayo Clin Proc 2012;87(2):143−50.
[35] Temiz Z, Cavdar I. The effects of training and the use of cranberry capsule in preventing urinary tract infections after urostomy. Complement Ther Clin Pract 2018;31:111−7.
[36] DE Leo V, Cappelli V, Massaro MG, Tosti C, Morgante G. Evaluation of the effects of a natural dietary supplement with cranberry, Noxamicina® and D-mannose in recurrent urinary infections in perimenopausal women. Minerva Ginecol 2017;69(4):336−41.
[37] Barbosa-Cesnik C, Brown MB, Buxton M, Zhang L, DeBusscher J, Foxman B. Cranberry juice fails to prevent recurrent urinary tract infection: results from a randomized placebo-controlled trial. Clin Infect Dis 2011;52:23−30.
[38] Cowan CC, Hutchison C, Cole T, et al. A randomised double-blind placebo-controlled trial to determine the effect of cranberry juice on decreasing the incidence of urinary symptoms and urinary tract infections in patients undergoing radiotherapy for cancer of the bladder or cervix. Clin Oncol 2012;24(2):e31−8.
[39] Maki KC, Kaspar KL, Khoo C, Derrig LH, Schild AL, Gupta K. Consumption of a cranberry juice beverage lowered the number of clinical urinary tract infection episodes in women with a recent history of urinary tract infection. Am J Clin Nutr 2016;103:1434−42.
[40] McMurdo MET, Bissett LY, Price RJG, Phillips G, Crombie IK. Does ingestion of cranberry juice reduce symptomatic urinary tract infections in older people in hospital? A double-blind, placebo-controlled trial. Age Ageing 2005;34:256−61.
[41] Stothers L. A randomized trial to evaluate effectiveness and cost effectiveness of naturopathic cranberry products as prophylaxis against urinary tract infection in women. Can J Urol 2002;9(3):1558−62.
[42] Salo J, Kontiokari T, Helminen M, Korppi M, Nieminen T, Pokka T. Randomized trial of cranberry juice for the prevention of recurrences of urinary tract infections in children. Clin Microbiol Infect 2010;16:S385−6.
[43] Takahashi S, Hamasuna R, Yasuda M, Arakawa S, Tanaka K, Ishikawa K, et al. A randomized clinical trial to evaluate the preventive effect of cranberry juice (UR65) for patients with recurrent urinary tract infection. J Infect Chemother 2013;19(1):112−7.
[44] Wan K-S, Liu C-K, Lee W-K, Ko M-C, Huang C-S. Cranberries for preventing recurrent urinary tract infections in uncircumcised boys. Altern Ther Health Med 2016;22(6):20−3.
[45] Wing DA, Rumney PJ, Preslicka CW, Chung JH. Daily cranberry juice for the prevention of asymptomatic bacteriuria in pregnancy: a randomized, controlled pilot study. J Urol 2008;180:1367−72.
[46] Pieper D, Heß S, Jr CMF. A new method for testing reproducibility in systematic reviews was developed, but needs more testing. BMC Med Res Methodol 2021;21:157.