Journal of Clinical Epidemiology 109 (2019) 136−141

Journal of Clinical Epidemiology

COMMENTARY

Software engineering principles address current problems in the systematic review ecosystem

Rabia Bashir*, Adam G. Dunn

Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Accepted 17 December 2018; Published online 21 December 2018

Abstract

Systematic reviewers are simultaneously unable to produce systematic reviews fast enough to keep up with the availability of new trial evidence while overproducing systematic reviews that are unlikely to change practice because they are redundant or biased. Although the transparency and completeness of trial reporting has improved with changes in policy and new technologies, systematic reviews have not yet benefited from the same level of effort. We found that new methods and tools used to automate aspects of systematic review processes have focused on improving the efficiency of individual systematic reviews rather than the efficiency of the entire ecosystem of systematic review production. We use software engineering principles to review challenges and opportunities for improving the interoperability, integrity, efficiency, and maintainability. We conclude by recommending ways to improve access to structured systematic review results. Major opportunities for improving systematic reviews will come from new tools and changes in policy focused on doing the right systematic reviews rather than just doing more of them faster.

Keywords: Software engineering; Systematic reviews as topic; Machine learning; Evidence synthesis; Trial registration; Updating systematic reviews

1. Current challenges in the systematic review ecosystem

Systematic reviews of clinical interventions play an important role in the policy and practice of health care and should provide an up-to-date synthesis of available trials and other clinical studies. Ensuring that evidence synthesis is current is particularly important for recently approved interventions, where the accumulation of new evidence might reveal safety issues and delays in their identification can cause harm [1]. Systematic reviews can also help to identify and mitigate the effect of publication and reporting biases [2], which result in delays in identifying safety issues. Despite rapid growth in the number of published systematic reviews, a substantial proportion is either redundant, or conflicted, or has little clinical value [3]. Ensuring that systematic reviews do what they are meant to do is an ongoing challenge in the area.

Systematic reviews are resource-intensive, and this hinders our ability to update them quickly enough to keep up with available evidence [4]. In response to this challenge, medical informatics specialists developed tools to support the automation of searching, screening, and synthesis of clinical evidence [5−9]. However, these tools have typically aimed to reduce the effort required to undertake individual systematic review processes. Less effort has been used to develop informatics tools and methods for identifying which systematic reviews and clinical questions should be prioritized for review [10]. New methods and guidelines in this space may benefit the broader systematic review ecosystem by helping to improve the allocation of resources to systematic reviews that are at highest risk of a change in results of conclusions. Current approaches used to decide when to update a systematic review interpret the existing evidence and the amount of time that has elapsed since the review was undertaken [11−13]. Although there are several examples of the tools that consider the use of new evidence to predict the risk of change in conclusions, examples that evaluate the effectiveness of their tools are rare [14−16]. From a 2016 assessment of the systematic review practices [3], it is evident that systematic reviewers still struggle with knowing in advance whether a systematic review is worthwhile.

Conflicts of interest: None.

* Corresponding author. Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia.
E-mail address: rabia.bashir@students.mq.edu.au (R. Bashir).

https://doi.org/10.1016/j.jclinepi.2018.12.014
0895-4356/

What is new?

Key findings

- Efforts aimed at improving systematic reviews have been focused on the quality or efficiency of performing individual reviews rather than on infrastructure to help avoid redundancy and monitor biases.

What this adds to what was known

- There are a range of innovations aimed at improving the completeness and timeliness of trial reporting, but connections across registries and bibliographic databases hinder systematic review production.
- Recent advances in the way trial study designs and results are represented in structured and machine-readable formats and stored in registries are not yet being fully used by systematic reviewers.

What is the implication and what should change now?

- Changes in policy and the culture of trial reporting could be expanded to cover systematic reviews, which could improve interoperability and efficiency.
- We propose establishing a centralized public repository for structured and machine-readable summaries of systematic reviews to match changes in the way clinical trials are registered and reported.

Our aim was to examine current issues in systematic reviews, looking specifically at how well recent proposals and developments are addressing current challenges.

2. How can software engineering principles help address these problems?

In software engineering practice, systems are built to meet requirements related to a set of quality of service attributes, which describe how well a system behaves when it is implemented. Here, we use quality of service attributes from software engineering as a lens through which to examine the entire systematic review ecosystem at once, identifying gaps and opportunities for improvement.

There is no consensus or standard for list of quality of service attributes in software engineering. Therefore, the definitions of what constitutes a quality of service attribute vary [17], but four important attributes are common across most lists: interoperability, integrity, efficiency, and maintainability. We use these to frame an evaluation of current systematic review practices, encompassing the technologies, data, and resources used to produce them. For each attribute, we examine how well systematic review practices currently meet expectations and discuss the emerging initiatives and technologies that are aimed at addressing deficiencies.

2.1. Interoperability

In software engineering, interoperability relates to the capability of a system to interact with other systems. If we consider the set of all systematic reviews and the processes for producing them as our system, then interoperability is how well systematic reviews connect with trial registries, bibliographic databases, other sources of trial and study information, and the guidelines and summaries that make use of systematic reviews in practice. In software engineering, common approaches for ensuring interoperability might include using standardized data formats to ensure frictionless communication. Despite efforts to improve interoperability in the registration and reporting of clinical studies, much less effort has been spent on connecting systematic reviews to the sources of information they use or the guidelines and policies that depend on them.

Some work has been done to improve the connectivity between systems in ways that may support interoperability with systematic reviews. There are different processes that have been used to establish interoperability between trial registries and bibliographic databases [18]. New methods have also been proposed to improve the proportion of trial registrations that include machine-readable links to bibliographic databases [19,20]. However, systematic reviews are typically not reported in ways that make it easy to establish links to trial registries and bibliographic databases. The evidence transfer between published trials, registries, and systematic reviews is largely ad hoc and unstructured. This is because there is a lack of standardization and interoperability to enable cross-study analyses [21].

In 2005, Sim et al. [22] proposed the use of structured and computable reporting of trial results specifically to enhance interoperability and transparency, but progress in the space has taken many years [23]. Recently, Zarin et al. [24] suggested a further step toward interoperability by proposing the use of ClinicalTrials.gov as a central location for linking trials via their unique registry identifiers to their protocols, published results, and to any systematic reviews in which they are included. In a similar way, structured representations of systematic review registrations and results could improve interoperability through better interfacing with structured representation of trials.

2.2. Integrity

In software engineering, integrity is defined by the completeness and consistency of the data that are maintained by the system. For the systematic review ecosystem

to demonstrate completeness and consistency, it would need to ensure that systematic reviews answering the same clinical question and specifying the same inclusion and exclusion criteria would include the same studies, and those studies would represent all relevant studies at the time of searching.

Incomplete representations of available evidence in systematic reviews are especially problematic when the studies that are included capture a biased subset of what is available. For example, where negative efficacy results are unpublished [25], or where safety outcomes are missing from reporting [26], systematic reviews may overestimate the efficacy and underestimate the harms of new interventions. Statistical methods used to detect or account for publication bias are of limited value [27].

Interoperability also affects data integrity at external level where the links between registration and publication of included studies are missing. This situation makes hard to spot the outcome reporting biases [28−31] and selective reporting or missing outcome data [32,33].

To address this challenge, systematic reviewers need to be able to efficiently access the complete results of trials. However, not every clinical trial gets published, which means that other sources of trial reporting become important. ClinicalTrials.gov in particular represents a very large source of structured summary results and may provide information for trials earlier and more completely. Studies examining the impact of searching for trial results in places other than in bibliographic databases conclude that trial registries are of some value [34,35]. To be comprehensive, systematic reviewers need to consider all sources of clinical trial results information including bibliographic databases, ClinicalTrials.gov, and clinical study reports available directly from investigators, this is often challenging because of a lack of transparency in trial reporting and the effort required to search and screen multiple databases.

2.3. *Efficiency*

In software engineering, efficiency is defined as the degree to which a system performs without wasting resources. An efficient systematic review system is one that quickly incorporates new clinical evidence in systematic reviews without undertaking unnecessary effort or producing redundant systematic reviews. Our recent work showed that a substantial proportion of systematic review updates are not targeting the clinical questions where the evidence accumulates faster [36]. These results are aligned with the broader perspective that many systematic reviews are redundant and poorly focused where they are most needed [3]. We think that problems with efficiency in the systematic review ecosystem may contribute to the slow detection of safety issues in new interventions [1]. Wherever resources are being wasted on redundant systematic reviews, they could instead be targeted at updating systematic reviews with signals from recently reported trials. A study performed by Takwoingi et al. [15] is an example of a decision tool that estimates the risk of conclusion change in systematic review update in advance of allocating resources to a systematic review. If we can improve tools of this type to work across a broader range of systematic reviews, we could improve the efficiency of the system by better targeting resources at clinical questions where conclusions are more likely to change.

Researchers have proposed a number of different approaches for deciding if and when a systematic review should be updated [11−13]. Novel approaches in this space use information from previous examples of systematic review updates to signal when a systematic review may be at risk of a change in results or conclusions [15,37]. Factors used to predict which reviews are at high risk include the amount of time since the review was last updated, the number of trials and participants in the previous update, the attributes of the primary meta-analysis, or simple information about new and potentially relevant trials.

2.4. *Maintainability*

In software engineering, maintainability describes the capacity for a system to cope with or adapt to changes in its environment. For systematic reviews, this corresponds to the ability to adapt changes in the ways evidence is produced, synthesized, and disseminated, and how the systematic reviews are used by health providers, patients, and policy-makers. Maintainability is a challenge because it involves dealing with any changes in the resources used to produce systematic reviews, including trial registries, bibliographic databases, as well as the culture and funding of systematic reviewers. It also includes any changes to the way the results of systematic reviews are disseminated to stakeholders, such as summary reports, policy briefs, guidelines, and through news and social media.

One such change in environment has come from the perceived value of systematic reviews. Evidence about associations between conclusions with conflicts of interest and funding [38,39] suggests that industry groups may be using systematic reviews as marketing tools. In addition, the growth in the number of available journals has made it easier to publish systematic reviews that are redundant or capture a biased subset of the available evidence. Each of these changes in practice may have introduced challenges to the credibility of systematic reviews. A number of guidelines have been made available for managing the expected norms for systematic reviews [40−43], but the use of these guidelines remains low [44].

3. Where are the best places to focus efforts now?

The substantial growth in the number of systematic reviews being produced suggests a recognition of their value

in improving policy and practice in medicine, but this growth has also created challenges in interoperability, integrity, efficiency, and maintainability that have not yet been fully addressed. To meet these challenges, we suggest the combined efforts of the systematic review and medical informatics communities.

We recommend the expanded use of standardized data formats for representing trials through their registrations and all forms of reporting including published articles, structured summary results, clinical study reports, and individual participant data [24]. The use of standardized representations of trial results data would benefit the interoperability of the system by providing a more complete representation of trial results available, which would in turn help to monitor and mitigate publication and reporting biases among prospectively registered trials. This will improve interoperability across trials that will have flow-on effects on systematic reviews—improving efficiency and integrity in the system by making it easier for systematic reviewers to account for all available trial evidence rather than just the subset of studies and outcomes reported in published articles. However, this would not be useful for systematic reviews based on old trials published before the inception of trial registration. Therefore, we can only recommend that the practice should be promoted for all newly registered trials, matching the requirements of some new policies in certain countries.

Part of the challenge of knowing when a systematic review is needed comes from keeping track of other similar systematic reviews that have been registered or published. Despite the substantial improvements that have been made to the transparency and completeness of trial reporting, systematic reviews have not seen the same level of scrutiny and development [45]. There is currently no centralized repository for identifying published systematic reviews along with information about their inclusion criteria and the outcomes they examine. Such a system would help systematic reviewers quickly ascertain if a given clinical question has been addressed by other systematic reviews and would help manage the ecosystem by monitoring the conclusions of systematic reviews with equivalent inclusion and exclusion criteria. The availability of all key attributes in a structured format could be used to quickly evaluate overlapping systematic reviews [46], helping to avoid redundancy and improving both efficiency and maintainability. PROSPERO is a registry for systematic reviews [47]. In the same way that ClinicalTrials.gov started as a way to prospectively register trials and was expanded to include summary results data for registered trials, PROSPERO would be a logical choice for expanding to include structured summary results data for systematic reviews. Echoing the governance used to define required information in ClinicalTrials.gov, required information in a systematic review registry could be based on PRISMA and its extensions and could include links to identifiers of included trials and studies.

Biases in systematic reviews are complex. There is growing evidence of biases caused by delayed or missing publications, selective outcome reporting, differences in funding over time after new interventions are marketed, the financial conflicts of interest, and others. A registry of systematic reviews and their results would not immediately solve each of these problems, but if it were coupled with changes in policy and practice, it could support improved surveillance and reduce duplication of effort.

Finally, we recommend continued pressure on systematic reviewers, funders, and journal editors to maintain expected standards for prospective registration and reporting for trials and systematic reviews. Although the rates are improving, prospective trial registration is still not fully enforced across all medical journals [48], and even where systematic reviews are registered, they may not account for substantial changes in outcomes when published [49]. Similarly, investigators of trials should also understand their responsibility to update trial registries with current information after completion, and more countries and funding organizations should consider requiring the timely reporting of structured and machine-readable summary results for trials they fund.

4. Conclusion

Using software engineering principles, we examined the challenges and opportunities that currently face the systematic review community. We found that the major improvements in the transparency and integrity of trial reporting have not yet been fully translated into improvements in the transparency and integrity of systematic reviews, although there are several nascent changes that should be encouraged. We also found that many of the new developments from the field of clinical research informatics tend to be more often focused on improving the efficiency of individual systematic reviews and less often focused on broader notions of efficiency that currently lead to the overproduction of redundant and unnecessary reviews. Although efforts in the area of machine learning and information retrieval are likely to be of value to systematic reviewers, there are further opportunities for clinical research informatics to address these broader challenges to efficiency associated with redundant and misleading reviews. The main reason for producing a systematic review is to guide the practice of health care. To meet the current challenges associated with systematic reviews, we recommend renewed focus on new tools and changes in policy that help systematic reviewers do the right systematic reviews at the right time.

Acknowledgments

This work was funded by Macquarie University Postgraduate Scholarship.

References

[1] Downing NS, Shah ND, Aminawung JA, Pease AM, Zeitoun J-D, Krumholz HM, et al. Postmarket safety events among novel therapeutics approved by the US food and drug administration between 2001 and 2010. JAMA 2017;317:1854−63.

[2] Page MJ, McKenzie JE, Higgins JP. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. BMJ Open 2018;8:e019703.

[3] Ioannidis J. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 2016; 94:485−514.

[4] Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How Quickly Do systematic reviews go out of date? A survival analysis. Ann Intern Med 2007;147:224−33.

[5] Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Syst Rev 2015;4:78.

[6] Miwa M, Thomas J, O'Mara-Eves A, Ananiadou S. Reducing systematic review workload through certainty-based screening. J Biomed Inform 2014;51:242−53.

[7] O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev 2015;4:5.

[8] Shemilt I, Simon A, Hollands GJ, Marteau TM, Ogilvie D, O'Mara-Eves A, et al. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. Res Synth Methods 2014;5:31−49.

[9] Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev 2014;3:74.

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

[11] Barrowman NJ, Fang M, Sampson M, Moher D. Identifying null meta-analyses that are ripe for updating. BMC Med Res Methodol 2003;3:13.

[12] Higgins J, Green S, Scholten RJ. Maintaining reviews: updates, amendments and feedback. Cochrane handbook for systematic reviews of interventions: Cochrane book series. England: John Wiley and Sons Ltd; 2008:31−49.

[13] Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR. Evidence-based sample size calculations based upon updated meta-analysis. Stat Med 2007;26:2479−500.

[14] Langan D, Higgins JP, Gregory W, Sutton AJ. Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis. J Clin Epidemiol 2012;65:511−9.

[15] Takwoingi Y, Hopewell S, Tovey D, Sutton AJ. A multicomponent decision tool for prioritising the updating of systematic reviews. BMJ 2013;347:f7191.

[16] Welsh E, Stovold E, Karner C, Cates C. Cochrane Airways Group reviews were prioritized for updating using a pragmatic approach. J Clin Epidemiol 2015;68:341−6.

[17] Chung L, do Prado Leite JCS. On non-functional requirements in software engineering. In: Borgida AT, Chaudhri VK, Giorgini P, Yu ES, editors. Conceptual modeling: foundations and applications. New York, NY: Springer; 2009:363−79.

[18] Bashir R, Bourgeois FT, Dunn AG. A systematic review of the processes used to link clinical trial registrations to their published results. Syst Rev 2017;6:123.

[19] Dunn AG, Coiera E, Bourgeois FT. Unreported links between trial registrations and published articles were identified using document similarity measures in a cross-sectional analysis of ClinicalTrials.gov. J Clin Epidemiol 2018;95:94−101.

[20] Goodwin TR, Skinner MA, Harabagiu SM. Automatically linking registered clinical trials to their published results with deep highway networks. AMIA Jt Summits Transl Sci Proc 2018;2017:54.

[21] van Valkenhoef G, Tervonen T, de Brock B, Hillege H. Deficiencies in the transfer and availability of clinical trials evidence: a review of existing systems and standards. BMC Med Inform Decis Mak 2012; 12:95.

[22] Sim I, Detmer DE. Beyond trial registration: a global trial bank for clinical trial reporting. PLoS Med 2005;2:e365.

[23] Altman DG. Making research articles fit for purpose: structured reporting of key methods and findings. Trials 2015;16:53.

[24] Zarin DA, Tse T. Sharing individual participant data (IPD) within the context of the trial reporting system (TRS). PLoS Med 2016;13: e1001946.

[25] Schmucker C, Schell LK, Portalupi S, Oeller P, Cabrera L, Bassler D, et al. Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. PLoS One 2014;9:e114023.

[26] Golder S, Loke YK, Wright K, Norman G. Reporting of adverse events in published and unpublished studies of health care interventions: a systematic review. PLoS Med 2016;13:e1002127.

[27] Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. Stat Med 2015;34:343−60.

[28] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA 2009;302:977−84.

[29] Pranić S, Marušić A. Changes to registration elements and results in a cohort of Clinicaltrials. gov trials were not reflected in published articles. J Clin Epidemiol 2016;70:26−37.

[30] Su C-X, Han M, Ren J, Li W-Y, Yue S-J, Hao Y-F, et al. Empirical evidence for outcome reporting bias in randomized clinical trials of acupuncture: comparison of registered records and subsequent publications. Trials 2015;16:28.

[31] Tang E, Ravaud P, Riveros C, Perrodeau E, Dechartres A. Comparison of serious adverse events posted at ClinicalTrials. gov and published in corresponding journal articles. BMC Med 2015;13: 189.

[32] Jones CW, Keil LG, Holland WC, Caughey MC, Platts-Mills TF. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. BMC Med 2015;13:282.

[33] Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 2010;340:c365.

[34] Baudard M, Yavchitz A, Ravaud P, Perrodeau E, Boutron I. Impact of searching clinical trial registries in systematic reviews of pharmaceutical treatments: methodological systematic review and reanalysis of meta-analyses. BMJ 2017;356:j448.

[35] Jones CW, Keil LG, Weaver MA, Platts-Mills TF. Clinical trials registries are under-utilized in the conduct of systematic reviews: a cross-sectional analysis. Syst Rev 2014;3:126.

[36] Bashir R, Dunn AG. Do systematic review updates target questions where evidence accumulates faster? Evidence Live. Oxford, UK; 2017 June 21-22.

[37] Bashir R, Surian D, Dunn AG. An empirically-defined decision tree to predict systematic reviews at risk of change in conclusion. 25th Cochrane Colloquium. Edingburgh, Scotland; 2018 September 16-18.

[38] Bes-Rastrollo M, Schulze MB, Ruiz-Canela M, Martinez-Gonzalez MA. Financial conflicts of interest and reporting bias regarding the association between sugar-sweetened beverages and weight gain: a systematic review of systematic reviews. PLoS Med 2013;10:e1001578.

[39] Dunn AG, Arachi D, Hudgins J, Tsafnat G, Coiera E, Bourgeois FT. Financial conflicts of interest and conclusions about neuraminidase inhibitors for influenza: an analysis of systematic reviews. Ann Intern Med 2014;161:513−8.

[40] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264−9.

[41] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;349:g7647.

[42] Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016;69:225−34.

[43] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008.

[44] 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 Med 2016;13:e1002028.

[45] Trinquart L, Dunn AG, Bourgeois FT. Registration of published randomized trials: a systematic review and meta-analysis. BMC Med 2018;16:173.

[46] Riva N, Puljak L, Moja L, Ageno W, Schünemann H, Magrini N, et al. Multiple overlapping systematic reviews facilitate the origin of disputes: the case of thrombolytic therapy for pulmonary embolism. J Clin Epidemiol 2018;97:1−13.

[47] Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev 2012;1:2.

[48] Dal-Ré R, Ross JS, Marušić A. Compliance with prospective trial registration guidance remained low in high-impact journals and has implications for primary end point reporting. J Clin Epidemiol 2016;75:100−7.

[49] Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. Syst Rev 2018;7:32.