In a pilot study, automated real-time systematic review updates were feasible, accurate, and work-saving.

Iain J. Marshall, Thomas A. Trikalinos, Frank Soboczenski, Hye Sun Yun, Gregory Kell, Rachel Marshall, Byron C. Wallace

PII: S0895-4356(22)00213-X
DOI: https://doi.org/10.1016/j.jclinepi.2022.08.013
Reference: JCE 10900

To appear in: Journal of Clinical Epidemiology

Received Date: 22 February 2022
Revised Date: 22 August 2022
Accepted Date: 29 August 2022

Please cite this article as: Marshall IJ, Trikalinos TA, Soboczenski F, Yun HS, Kell G, Marshall R, Wallace BC, In a pilot study, automated real-time systematic review updates were feasible, accurate, and work-saving, Journal of Clinical Epidemiology (2022), doi: https://doi.org/10.1016/j.jclinepi.2022.08.013.

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

© 2022 The Author(s). Published by Elsevier Inc.
In a pilot study, automated real-time systematic review updates were feasible, accurate, and work-saving

Iain J Marshall¹
Thomas A Trikalinos²
Frank Soboczenski¹
Hye Sun Yun³
Gregory Kell¹
Rachel Marshall⁴
Byron C Wallace³

1. School of Life Course and Population Sciences, King’s College London, London, UK
2. Center for Evidence Synthesis in Health, Brown University,
3. Khoury College of Computer Sciences, Northeastern University, Boston, US
4. Independent researcher, London, UK

Authors’ contributions
Conception: IJM, TAT, BCW; System design: IJM, TAT, RM, BCW; Software development: IJM, FS, HSY, GK, BCW; Data analysis: IJM; Initial draft of paper: IJM, TAT, FS, HSY, GK, RM, BCW; Review and agreement of final draft: all

Conflict of interest
The authors declare no conflicts of interest.

Funding
This work has been supported by National Institutes of Health (NIH) under the National Library of Medicine (NLM), grant R01-LM012086 and by the National Science Foundation (NSF) under Grant 1750978: "CAREER: Structured Scientific Evidence Extraction: Models and Corpora". The work has also been partially supported by the UK Medical Research Council (MRC), through its Skills Development Fellowship program, fellowship MR/N015185/1.
ABSTRACT

Objective
To describe and pilot a novel method for continuously identifying newly published trials relevant to a systematic review, enabled by combining artificial intelligence (AI) with human expertise.

Study Design and Setting
We used RobotReviewer LIVE to keep a review of Covid-19 vaccination trials updated from February–August 2021. We compared the papers identified by the system with those found by the conventional manual process by the review team.

Results
The manual update searches (last search date July 2021) retrieved 135 abstracts, of which 31 were included after screening (23% precision; 100% recall). By the same date, the automated system retrieved 56 abstracts, of which 31 were included after manual screening (55% precision; 100% recall). Key limitations of the system include that it is limited to searches of PubMed/MEDLINE, and considers only randomised controlled trial reports. We aim to address these limitations in future. The system is available as open-source software for further piloting and evaluation.

Conclusion
Our system identified all relevant studies, reduced manual screening work, and enabled rolling updates on publication of new primary research.
What is new?

Key findings

- We developed a system which continuously identifies newly published trial evidence relevant to a systematic review, enabled by combining artificial intelligence (AI) with human expertise.

- The semi-automated system found 100% of the relevant studies found by a conventional manual update, during a pilot, when updating a systematic review of covid vaccination trials.

What this adds to what is known

- Living systematic reviews have been proposed as a new model for keeping evidence syntheses updated. Most current living reviews rely on repeated manual update searches, which are time consuming and laborious.

- We show that using a hybrid AI/expert model could lead to lower latency updates, potentially reducing workload, and improving the currency of systematic reviews.

What is the implication, what should change now

- Systems which use AI to automatically notify systematic review authors of new evidence (‘push’ updates) are feasible, and should be piloted on a wider range of systematic reviews.

- Future research should examine how best to adapt these technologies for use in more complex reviews (particularly reviews of non-trial evidence, and those with complex inclusion criteria).

- Journal publishers should investigate models for rapid updating, to enable automated live updates of review status to be published.
Background

For many health conditions and treatments, evidence accumulates rapidly.[1,2] Systematic reviews identify, appraise and synthesise all empirical evidence on healthcare topics, and are therefore invaluable for making clinical decisions and informing policy. However, most reviews are static publications, which can become quickly out of date as new primary research is published.[3] For the reader, it is currently impossible to determine whether any particular systematic review is up-to-date, or whether new important new research was published after the searches were conducted. For authors, it is unclear whether it is worth the effort of updating their review, given uncertainty about whether new evidence exists which might change their conclusions.[4] For commissioners and policy makers, it is unclear when and whether to fund updates of systematic reviews.

As an example, consider the topic of COVID-19 treatments, or vaccines. New studies are being rapidly conducted and published on these topics. A ‘static’ systematic review on either topic, with a search date of six months ago (from the time of this writing) is likely to have missed critical new findings, and failed to provide an account of the current science. Given the pace of new published trial evidence in COVID-19, a conventional systematic review would likely become outdated before it was ever published.

*Living systematic reviews* have been proposed as one model for keeping rigorous syntheses current with evolving evidence.[5,6] The idea is to update syntheses as new evidence emerges, ideally with low latency. For COVID-19 specifically, a number of living reviews are currently being maintained on both treatment and vaccines.[7,8] To date, living systematic reviews have been achieved by repeating a conventional systematic review update on a frequent basis (updating searches, say, monthly or weekly), screening the results and extracting data.[9] This process still depends upon review teams having to actively run searches and find new studies (a “pull” model); and will result in some lag between manual
search and identification of relevant studies. In addition, conventional database searching can yield large numbers of abstracts which require screening.

The process of conducting the search, and screening the results to identify potentially relevant abstracts is a large proportion of the work to conduct an SR. The findings of this work (whether there are new studies identified or not) is important to readers and policy makers. The main mechanism to provide this information to users is to publish a ‘full’ update. This process, particularly for ‘empty updates’ is time consuming. There is a need to identify new evidence relevant to existing systematic reviews in a more efficient and less manual way. In addition, there is a need to have a formal way to represent the currency of existing systematic reviews, based on whether all relevant evidence has been incorporated.

There has been much recent research attention on how to use artificial intelligence (AI) systems to automate (or semi-automate: where AI systems are combined with human experts) living updates.[10,11] The most advanced technology in this respect is the use of machine learning (ML) to prioritise studies for screening, which has found to be accurate and efficient in a number of methodological studies,[12–14] and is available at the time of writing in several systematic review authoring tools.[15,16]

Here, we describe a hybrid system that integrates machine learning and natural language processing (NLP) methods with human expertise to translate static systematic reviews into living reviews. The system automatically monitors research databases for new, relevant research to a systematic review, and notifies the review authors. This “push” model differs fundamentally from the standard approach to updating reviews, which depends on review authors taking the initiative to periodically search for newly published evidence. We present a formative evaluation of the system, comparing the reliability of (semi-)automatic systematic review updates prospectively with traditional manual update searches for an ongoing systematic review of COVID-19 vaccine efficacy. The system — a collection of trained models and a prototype web interface with which to interact with them — is free and open-
source. This work constitutes a step toward translating the idea of living reviews into practice.

Methods

Before using RobotReviewer Live, users develop a review question, and manually (via baseline full systematic manual search, and manual screen of abstracts) identify baseline included studies. RobotReviewer Live can be used prospectively with new reviews (to replace the need for update searches prior to publication), or to bring existing reviews up to date – the only requirement to use RobotReviewer Live is the availability of the baseline search results and baseline included abstracts.
To start the process, the user registers their review on RobotReviewer Live using the user interface shown in Figure 1. The system then surveils the medical literature for newly published RCTs likely to meet the inclusion criteria of the systematic review. Topic experts screen the matching abstracts as and when they are identified, and their inclusion decisions produce live status updates of the systematic review. We illustrate the steps of the updating process in Figure 2, and describe each in detail below. These steps are run continuously after publication of the initial review.
FIGURE 2: Model for real-time updates of systematic review in response to new evidence.

1. Finding new clinical trial reports
As a first step, we monitor PubMed daily for new clinical trial reports via the Trialstreamer database.[17] Trialstreamer identifies all reports of randomised controlled trials (RCTs) via a validated machine learning model (recall 0.97, precision 0.52). Abstracts describing RCTs then go on to detailed automatic extraction of trial characteristics (descriptors of participants, interventions, and outcomes), sample sizes, author conclusions, and indicators of methodological bias. We have described the NLP data extraction methods used in Trialstreamer in detail previously.[17]

2. Broad topic filter
In this step we move from the set of all RCTs (>750,000 at the time of writing) to a topic-focused (but highly sensitive) set for subsequent machine classification. This topic-focused set of RCTs is created by selecting broad topic terms relevant to the question of the systematic review. The available terms are derived from the MeSH vocabulary alongside an indicator of whether the term describes the Population, Interventions, or Outcomes of the trial. The RobotReviewer LIVE interface allows reviewers to select relevant terms using an autocompleting text box.

The topic-focused set might only include articles that mention a particular condition of interest. The topic-focused set is assumed to be much broader than the set of studies that would be retrieved with a conventional search for a specific systematic review question. In the case of the COVID-19 vaccines review, we included articles in the topic-focused set only if the abstracts contained a mention of COVID-19 or a synonym from a synonym list we generated automatically by minimally processing terms from the Unified Medical Language System (UMLS) Metathesaurus. We have described the method for generating these terms previously.\[17\]

3. Machine learning inclusion decisions

Our goal in this step is to automatically filter out the vast majority of irrelevant articles. We have found previously that machine learning models with sufficient recall for systematic reviews (which aim to retrieve all research fulfilling their inclusion criteria) will, even in the best case, retrieve a high fraction of false positives. We therefore aim to develop a model with near 100% recall, but add a later screening step by a human expert to remove false positives. A lower precision is therefore acceptable so long as the volume of articles for manual screening is manageable. To achieve this, we train a classification layer on top of ‘BERT’-based[18] representations of input articles. BERT (Bidirectional Encoder Representations from Transformers) is a multi-layer neural network language model, which is ‘pre-trained’ using a large volume of unlabeled plain text documents (for example, the full contents of Wikipedia, and large collections of books freely available on the internet). Here specifically, we use the BERT variant BioMed RoBERTa, which is optimised for scientific
research articles by conducting the pre-training on a large collection of scientific articles obtained from Semantic Scholar.[19][20] We make use of the human inclusion and exclusion decisions from the original systematic review to train this model.

In the task of abstract screening for systematic reviews, there are typically far fewer relevant than irrelevant citations (that is, most candidate articles retrieved via search will not meet review eligibility criteria). This creates class imbalance[21] in the training set, which can in turn result in poor model sensitivity, because overall predictive loss can be largely minimised by predicting that all instances belong to the majority class (i.e., that all abstracts are irrelevant). Following prior work[22] on methods for achieving a better balance between sensitivity and specificity in imbalanced scenarios, we re-sample the data to induce a balanced distribution during model training. We construct balanced batches during Stochastic Gradient Descent (SGD) using weighted sampling, such that minority examples (relevant citations) are assigned weights inversely proportional to the prevalence of the minority class.

We trained our BioMed RoBERTa model for five epochs using SGD with a learning rate of $10^{-3}$ and momentum of 0.9, yielding a final model that recalled 100% of relevant articles with 40% precision (AUROC 0.97) when evaluated on 10% of the dataset which was held out from training. Our model code is available on our project GitHub page.¹

4. Validation of results by systematic review authors

If the steps above retrieve new potentially relevant articles, systematic review authors are notified by email and invited to screen new abstracts for relevance. This step aims to remove any false positives (i.e., ineligible articles deemed relevant by the model). While conventional systematic review searches might include hundreds or thousands of articles for manual

¹ [https://github.com/bwallace/RobotScreen/]
review, the automatic system (in Step 3) aims to remove the vast majority of these articles. In the case of our example topic (which was subject to particularly high rates of research and publication during the study period), the system identified on average three potentially eligible abstracts per week which were then pushed to the review’s lead author. Review authors can screen the new studies by signing on to the website (Figures 3 and 4).

**FIGURE 3** — The RobotReviewer LIVE ‘Dashboard’ to monitor progress of all of a user’s reviews.
FIGURE 4 — Interface for review authors to validate new studies; their include/exclude decisions are automatically and instantly incorporated into the published live status update

5. Publication of live status update

We automatically publish a live update, which makes use of the latest information from both the automated and manual evidence screening (see example in Figure 2). This text is designed to be displayed as an additional section in the structured abstract, with the header “Automatic updates”. We display the full abstract including the live update section on our website, and also make this available via a REST API so that external journal publishers...
could opt to display a live, updated version of the abstract as part of the primary research article in future.

We provide meta-data about new studies (including numbers of studies screened, and how many were deemed relevant by the topic expert, and numbers of trial participants). This numerical meta-data is collected from our screening records, and from the structured data in the Trialstreamer database (which has been automatically extracted using NLP models), and displayed following a template.

As part of this step, we also have explored the use of automatic narrative summaries of newly included studies. We aimed to produce a brief summary of the new studies' findings to be presented alongside the templated metadata described in the main paper. We provide further details about this method and results in the Appendix.

**Evaluation: Prospective case study with a COVID-19 vaccination review**

We evaluated the system prospectively in comparison to a conventional manually updated living systematic review on COVID-19 vaccination evidence. The baseline full systematic manual searches for this review were completed and screened on February 9th, 2021.

We ran our comparative evaluation from February 9th 2021 to August 1st 2021. During this period, the review authors performed conventional manual update searches, and we ran the semi-automated system in parallel. We calculated recall with respect to the combined set of included articles from the manual and automatic update systems. Screening of the abstracts found by RobotReviewer LIVE was done by an independent member of the review team, who was not involved in the screening of the manual update searches.

Due to time taken to screen abstracts on the manual update, the last manual update search done during the evaluation period was on July 1st 2021. The “push” model used by our
automated system, where smaller numbers of abstracts were sent to be screened on the day of publication, meant that there was close to no lag between abstract publication and screening, and the live status updates included abstracts published up to and including August 1st 2021. We present results separately until July 1st (which represent a direct comparison of automatic update search performance versus conventional manual update searches at intervals), and from July 1st until August 1st 2021 (which evaluate any advantage in screening efficiency with automation) to allow a fair comparison.

Results

We present a PRISMA flow diagram comparing the screening approaches in Figure 5. The baseline (manual) version of the review search was conducted in February 2021. This yielded 4493 abstracts of which 38 both met eligibility criteria and were reports of RCTs.

![PRISMA flow diagram of the live review.](image)

Manual update searches retrieved 135 abstracts; by contrast the automated system retrieved 56. Both strategies resulted in the same 31 included abstracts after screening.
Discussion

We have presented a system for identifying new evidence to include in systematic reviews, and for producing live abstract updates on the currency of systematic reviews. RobotReviewer Live combines AI (ML/NLP) with human expertise, and allows new studies to be incorporated in published review reports quickly after publication. We have made the software, ML models, and data needed to implement the system freely available as open source software. We also provide a prototype of RobotReviewer Live that features a simple user interface, which should allow systematic review authors to produce live updates for their existing “static” systematic reviews. This prototype is also available as open-source code.

We provide an easy-to-use interface to allow experts to validate the automatic search results — potentially providing substantial efficiencies in the updating process, while still providing the assurances afforded by expert verification. In practice, converting a new conventional systematic review to a “living” equivalent using the system could be done in a matter of minutes. We make the technology available as open source, together with a REST API to enable live updates to be used inline in published journal articles, embedded in the websites of third party publishers. Even where a review is not actively kept up to date, this may allow interested individuals to see estimates of the amount of relevant evidence published since the time said review was completed. In the future, this platform may also permit “crowdsourced” maintenance of systematic reviews.

Related systems have been developed and evaluated; notably the Cochrane ‘Evidence Pipeline’, and Centralised Search Service.[23–25] These projects also monitor research databases (using a combination of machine learning identification of RCTs and crowdsourcing) and notify Cochrane review groups (who each typically manage tens of systematic reviews on a common clinical theme) when new research is published relevant to
their theme. By contrast, our system is designed to manage updates for individual systematic reviews.

In our prospective case study, the automated method identified all of the includable abstracts found manually. We continued to run the automated system for an additional month after the evaluation period (until August 1st), since the review team conducted the manual update search earlier than we had expected. In this month, the automated system found 12 additional abstracts which were deemed includable. This illustrates the advantage of the low latency “push” screening model, especially for topics such as Covid-19 vaccination, with rapid publication rates.

One criticism of systematic review automation tools previously is that they are often found as discrete, scattered pieces of academic code which require substantial technical expertise to use in practice.[11,26] To overcome this problem, we have produced an easy-to-use web interface which should allow users to create a “living” version of a systematic review with minimal effort (Figures 3 and 4).

This technology is still emerging, and users should be aware of important limitations. Although the performance on this case study is strong, the evaluation review is ideal for such technology. The review question is precise, and concerns a well-defined intervention and health condition, both of which are easy to capture in the structured vocabularies used in the Trialstreamer database. In the midst of a pandemic, there are also large numbers of eligible studies being published (whereas precision is likely to reduce in any search as the prevalence of eligible studies decreases — no matter whether manual or automated). We have presented a single case study, and it is likely that performance will vary particularly for more complex reviews.
Currently, we make use of the Trialstreamer database, which at present is limited to articles describing RCTs. We intend to make additional article types available in future; at present the system is limited for use to systematic reviews of intervention trials due to the data sources used. At present, we make use of articles from PubMed only — we are unable to access additional proprietary databases such as EMBASE which might (modestly) harm the recall of the system.[27] Overall, although the individual components of the system have been extensively validated, this report describes the only validation using a conventional systematic review as a comparator. The reliability of the system in general (particularly for reviews that deviate substantially from the format of the current evaluation) requires further study.

Conclusion

Manually updating systematic reviews is time consuming and laborious, meaning many conventionally produced reviews become quickly out-of-date. We hope that further evaluation and development of the ideas and methods presented here will bring the goal of dynamic publication of live evidence synthesis updates a step closer into practice.

References

1. Marshall IJ, L’Esperance V, Marshall R, Thomas J, Noel-Storr A, Soboczenski F, et al. State of the evidence: a survey of global disparities in clinical trials. BMJ Glob Health. 2021;6: e004145. doi:10.1136/bmjgh-2020-004145
2. Bastian H, Glasziou P, Chalmers I. Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up? PLOS Med. 2010;7: e1000326. doi:10.1371/journal.pmed.1000326
3. 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: 224–233. doi:10.7326/0003-4819-147-4-200708210-00179
4. Garner P, Hopewell S, Chandler J, MacLehose H, Schünemann HJ, Akl EA, et al. When and how to update systematic reviews: consensus and checklist. BMJ. 2016;354: i3507. doi:10.1136/bmj.i3507
5. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JPT, Mavergames C, et al. Living
Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap. PLOS Med. 2014;11: e1001603. doi:10.1371/journal.pmed.1001603

6. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. J Clin Epidemiol. 2017.

7. Siemieniuk RA, Bartoszko JJ, Ge L, Zeraatkar D, Izcoivich A, Kum E, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. BMJ. 2020;370: m2980. doi:10.1136/bmj.m2980

8. Boutron I, Chaimani A, Meerpohl JJ, Hróbjartsson A, Devane D, Rada G, et al. The COVID-NMA Project: Building an Evidence Ecosystem for the COVID-19 Pandemic. Ann Intern Med. [cited 29 Nov 2021]. doi:10.7326/M20-5261

9. Living Evidence Network. Guidance for the production and publication of Cochrane living systematic reviews: Cochrane Reviews in living mode. Cochrane; 2019 Dec. Available: https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf

10. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. J Clin Epidemiol. 2017;91: 31–37. doi:10.1016/j.jclinepi.2017.08.011

11. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev. 2019;8: 163. doi:10.1186/s13643-019-1074-9

12. O’Marra-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. doi:10.1186/2046-4053-4-5

13. Shemilt I, Simon A, Hollands GJ, Marteau TM, Ogilvie D, O’Marra-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. doi:https://doi.org/10.1002/jrsm.1093

14. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an Interactive Machine Learning System in an Evidence-based IHI ‘12. New York, NY, USA: ACM; 2012. pp. 819–824. doi:10.1145/2110363.2110464

15. Hamel C, Kelly SE, Thavorn K, Rice DB, Wells GA, Hutton B. An evaluation of DistillerSR’s machine learning-based prioritization tool for title/abstract screening – impact on reviewer-relevant outcomes. BMC Med Res Methodol. 2020;20: 256. doi:10.1186/s12874-020-01129-1

16. Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. Syst Rev. 2020;9: 73. doi:10.1186/s13643-020-01324-7

17. Marshall IJ, Nye B, Kuiper J, Noel-Storr A, Marshall R, Maclean R, et al. Trialstreamer: A living, automatically updated database of clinical trial reports. JAMIA. 2020; ocaa163.

18. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv181004805 Cs. 2018 [cited 8 Nov 2018]. Available: http://arxiv.org/abs/1810.04805

19. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXiv190711692 Cs. 2019 [cited 15 Oct 2021]. Available: http://arxiv.org/abs/1907.11692

20. Gururangan S, Marasović A, Swayamdipta S, Lo K, Beltagy I, Downey D, et al. Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics; 2020. pp. 8342–8360. doi:10.18653/v1/2020.acl-main.740

21. The class imbalance problem: A systematic study | Intelligent Data Analysis. [cited 15 Oct 2021]. Available: https://dl.acm.org/doi/10.5555/1293951.1293954

22. Wallace BC, Small K, Brodley CE, Trikalinos TA. Class Imbalance, Redux. 2011 IEEE 11th International Conference on Data Mining. 2011. pp. 754–763.
23. Thomas J, McDonald S, Noel-Storr A, Shemilt I, Elliott J, Mavergames C, et al. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. J Clin Epidemiol. 2020;0. doi:10.1016/j.jclinepi.2020.11.003

24. Noel-Storr AH, Dooley G, Wisniewski S, Glanville J, Thomas J, Cox S, et al. Cochrane Centralised Search Service showed high sensitivity identifying randomized controlled trials: A retrospective analysis. J Clin Epidemiol. 2020;127: 142–150. doi:10.1016/j.jclinepi.2020.08.008

25. Noel-Storr A, Dooley G, Elliott J, Steele E, Shemilt I, Mavergames C, et al. An evaluation of Cochrane Crowd found that crowdsourcing produced accurate results in identifying randomized trials. J Clin Epidemiol. 2021;133: 130–139. doi:10.1016/j.jclinepi.2021.01.006

26. O’Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Wolfe MS. Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR). Syst Rev. 2018;7: 3. doi:10.1186/s13643-017-0667-4

27. Marshall I, Marshall R, Wallace B, Brassey J, Thomas J. Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. J Clin Epidemiol. 2018. doi:10.1016/j.jclinepi.2018.12.015
• We developed a hybrid human expert/artificial intelligence system to keep systematic reviews up to date
• The system continuously surveils PubMed/MEDLINE for new relevant articles, and notifies review authors
• A living abstract is made available, which shows the status of the review in real-time
• In a pilot, the system was effective and reduced workload in a systematic review of covid-19 vaccination studies
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
CRediT author statement:

Iain J Marshall: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing – Review and Editing, Supervision, Project Administration, Funding Acquisition
Thomas A Trikalinos: Conceptualization, Methodology, Validation, Writing – Review and Editing, Supervision, Funding Acquisition
Frank Soboczenski: Conceptualization, Methodology, Software, Validation, Writing – Review and Editing
Hye Sun Yun: Methodology, Software, Writing – Review and Editing
Gregory Kell: Methodology, Software, Writing – Review and Editing
Rachel Marshall: Conceptualization, Methodology, Writing – Review and Editing
Byron C Wallace: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing – Review and Editing, Supervision, Project Administration, Funding Acquisition