



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Living systematic reviews

Citation for published version:

Living Systematic Review Network, Thomas, J, Noel-Storr, A, Marshall, I, Wallace, B, McDonald, S, Mavergames, C, Glasziou, P, Shemilt, I, Synnot, A, Turner, T & Elliott, J 2017, 'Living systematic reviews: 2. Combining human and machine effort', *Journal of Clinical Epidemiology*, vol. 91, pp. 31-37.
<https://doi.org/10.1016/j.jclinepi.2017.08.011>

Digital Object Identifier (DOI):

[10.1016/j.jclinepi.2017.08.011](https://doi.org/10.1016/j.jclinepi.2017.08.011)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Clinical Epidemiology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Living systematic reviews: 2. Combining human and machine effort

James Thomas^{a,*}, Anna Noel-Storr^b, Iain Marshall^c, Byron Wallace^d, Steven McDonald^e,
Chris Mavergames^f, Paul Glasziou^g, Ian Shemilt^a, Anneliese Synnot^{e,h}, Tari Turner^e,
Julian Elliott^{e,i}, on behalf of the Living Systematic Review Network

^aEPPI-Centre, Department of Social Science, University College London, 18 Woburn Square, London, WC1H 0NR, UK

^bRadcliffe Department of Medicine, University of Oxford, Level 4, Academic Block, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

^cPrimary Care & Public Health Sciences, Kings College, Capital House, 42 Weston Street, London, UK

^dCollege of Computer and Information Science, Northeastern University, 360 Huntington Ave, Boston, MA 02115, USA

^eCochrane Australia, School of Public Health and Preventive Medicine, Monash University, 99 Commercial Road, Melbourne VIC 3004, Australia

^fCochrane, St Albans House, 57-59 Haymarket, London SW1Y 4QX, UK

^gCentre for Research on Evidence Based Practice, Bond University, 14 University Drive (Off Cottesloe Drive), Robina, QLD 4226, Australia

^hCentre for Health Communication and Participation, School of Psychology and Public Health, La Trobe University, Melbourne, Australia

ⁱDepartment of Infectious Diseases, Monash University and Alfred Hospital, 55 Commercial Rd, Melbourne VIC 3004, Australia

Accepted 17 August 2017; Published online 11 September 2017

Abstract

New approaches to evidence synthesis, which use human effort and machine automation in mutually reinforcing ways, can enhance the feasibility and sustainability of living systematic reviews. Human effort is a scarce and valuable resource, required when automation is impossible or undesirable, and includes contributions from online communities (“crowds”) as well as more conventional contributions from review authors and information specialists. Automation can assist with some systematic review tasks, including searching, eligibility assessment, identification and retrieval of full-text reports, extraction of data, and risk of bias assessment. Workflows can be developed in which human effort and machine automation can each enable the other to operate in more effective and efficient ways, offering substantial enhancement to the productivity of systematic reviews. This paper describes and discusses the potential—and limitations—of new ways of undertaking specific tasks in living systematic reviews, identifying areas where these human/machine “technologies” are already in use, and where further research and development is needed. While the context is living systematic reviews, many of these enabling technologies apply equally to standard approaches to systematic reviewing. © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Systematic review; Automation; Crowdsourcing; Citizen science; Machine learning; Text mining

1. Introduction

This is the second paper in a series of papers discussing the emerging field of living systematic reviews (Box 1). In this paper, we specifically focus on the ways

in which the use of new human and machine “technologies” can make the standard systematic review process more efficient.

Systematic reviews are a type of literature review, which adopt principles of scientific method to the task of finding and summarizing research. They aim to answer prespecified research questions using all relevant empirical evidence, using explicit and replicable methods and minimizing bias. They thus aim to provide trustworthy findings on which policy and practice decisions can be made [1]. A living systematic review is a systematic review which is continually updated in the light of new evidence as it becomes available (Box 2) [2]. Living systematic reviews represent an opportunity to rethink conventional review processes and take advantage of emerging approaches to reviewing which promise to increase efficiency [3]. In this paper, we describe how new “technologies” (which encompass both

Conflicts of interest: none.

Funding: The Living Systematic Review Network is supported by funding from Cochrane and the Australian National Health and Medical Research Council (Partnership Project grant APP1114605). J.T., A.N.-S., I.S., T.T., and J.E. receive funding from Cochrane (“Transform Project”) and Australian NHMRC (“Evidence Innovation Transforming the efficiency of systematic review”). J.T. is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) North Thames at Bart’s Health NHS Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

* Corresponding author. Tel.: +44 207 612 6844; fax: +44 207 612 6400.

E-mail address: james.thomas@ucl.ac.uk (J. Thomas).

What is new?

- The need to maintain an up-to-date, dynamic system for evidence synthesis can be facilitated using new technologies which comprise both human and machine effort.
- As well as standard review teams, systematic review activities can be broken down into “micro-tasks” and distributed across a wider group of people—including the involvement of citizen scientists through crowdsourcing.
- Machine automation can assist with some systematic review tasks, including routine searching, eligibility assessment, identification and retrieval of full-text reports, extraction of data, and risk of bias assessment.
- While the context is living systematic reviews, many of these enabling technologies apply equally to standard approaches to systematic reviewing.

computer technology and more efficient models of human contribution) can increase the efficiency and sustainability of the systematic review enterprise. We argue that human effort is a scarce and valuable resource which should be expended only where automation is impossible, impractical, or undesirable. Furthermore, for many of the repetitive and labor-intensive tasks of evidence synthesis, automation is increasingly preferable and viable [4–10]. Human effort can contribute in two ways: either by undertaking tasks in a specific review or by providing examples that can be used to “train” machines which can then automate (or semi-automate) the activity in question—sometimes across many reviews. We consider how human effort can be considered not simply in terms of traditional author teams, but in terms of communities—and “crowds”—of people who come together to curate knowledge in a given area. Rather than organize the paper in terms of the two families of

Box 1 Series of papers on living systematic reviews

- Living systematic reviews: 1. Introduction—the why, what, when, and how
- Living systematic reviews: 2. Combining human and machine effort
- Living systematic reviews: 3. Statistical methods for updating meta-analyses
- Living systematic reviews: 4. Living guideline recommendations

Box 2 Living systematic reviews

- A systematic review which is continually updated, incorporating relevant new evidence as it becomes available
- An approach to review updating, not a formal review methodology
- Can be applied to any type of review
- Use standard systematic review methods
- Explicit and a priori commitment to a predetermined frequency of search and review updating

technologies—human and machine—we consider each stage of the systematic review process and discuss ways in which these two technologies interact and operate in mutually supportive ways.

2. Opportunities for a different workflow

Systematic reviews are conventionally undertaken by a small team of trained researchers working in a highly labor-intensive—but time-limited—way. New ways of working aim to replace this with a less labor-intensive model in which the ongoing workflow is conducted by a wider community of individuals. The changes to review production we describe here are not required for living systematic reviews to be conducted, but are situated within a wider set of innovations in evidence synthesis from which living systematic reviews can draw to improve feasibility. For example, it is possible that systematic review production will evolve away from an individual, isolated endeavor, toward a dynamic and continuous research curation system in which communities of people work together to maintain an up-to-date evidence base in their areas of interest. By breaking work into micro-tasks, living systematic reviews may be conducted more efficiently among a wider range of people. Microtasks are discrete, small units of work, which can be done independently from one another. We describe key living systematic review microtasks alongside examples of new technologies and innovative ways of working to help accomplish them in [Table 1](#). Breaking up the living systematic review workload in this way allows the authorship team to take advantage of emerging automation systems to reduce the workload. It also makes more efficient use of the skill sets and time availability of contributors necessary for undertaking review tasks. The formation of review teams can be assisted using task-sharing platforms, such as Cochrane’s “Task Exchange” (taskexchange.cochrane.org/).

As Elliott et al. described [2], the work of a living systematic review begins with a traditional systematic review (which can also benefit, of course, from many of the efficiencies described here). We take the existence of the initial systematic review as our starting point and outline below

Table 1. The microtasks of a living systematic review, and how they may be automated or made more efficient

Review task	Method/microtask	Potential methods for efficiency gain
Team formation	Use a wider range of personnel than may traditionally be the case	Crowdsourcing (e.g., Cochrane Crowd) Task-sharing platforms (e.g., TaskExchange)
Search	Running search on bibliographic databases	Automatic, continuous database search with push notification Database aggregators (e.g., HDAS, Epistemonikos) Notification from clinical trial registries Automatic retrieval of full-text papers (e.g., CrossRef)
Eligibility assessment	Selecting studies for inclusion	Machine-learning classifier Crowdsourced inclusion decisions
Data extraction or collection	Extracting information on characteristics of the participants, interventions, outcomes	Machine-learning information-extraction systems (e.g., RobotReviewer, ExaCT) Linkage of existing structured data sources (e.g., clinical trials registries) Automated structured data extraction tools for PDFs (e.g., ContentMine, Graph2Data)
	Assessing risks of bias	Machine learning–assisted risk of bias tools (e.g., RobotReviewer)
Synthesis	Entering data into meta-analysis software	Structured data extraction tools, which automatically provide data in suitable format for statistical analysis
	Conducting meta-analyses	Continuous analysis updating based on availability of structured extracted data
	Report writing and updating conclusions	Templated reporting of some report items Statistical surveillance of key analysis results, with threshold set for potential conclusion change
Supportive systems that reduce duplication of effort	Data sharing and reuse	Using standard descriptors for studies across all systematic reviews (e.g., linkeddata.cochrane.org)

Abbreviation: HDAS, Health Database Advanced Search.

the ways in which human and machine technologies can work together to maintain the review as a living entity. Throughout, we illustrate our argument with examples of tools and processes. This is not intended to be an exhaustive overview, and readers are referred to the Systematic Review ToolBox (<http://systematicreviewtools.com/>) which indexes a wide range of relevant tools.

2.1. Database searching and eligibility assessment

Identifying studies for inclusion is one of the most labor-intensive and time-consuming tasks of the systematic review process [11,12]. The component tasks include the following: the searching of electronic bibliographic databases, downloading the results, uploading them into software for citation management, deduplicating records, and screening them independently for eligibility. Guidance recommends that every citation is checked by two people to reduce the possibility that studies will be missed by accident [13]. This standard process has remained largely unchanged since systematic reviews first appeared.

2.1.1. Database searching

In a living systematic review, the search for bibliographic material needs to be streamlined, changing the traditional “pull” for bibliographic records (as carried out within most systematic reviews) to a “push” model. In the “push” model, automated searches are run regularly and reviewers are alerted to the presence of new potentially

relevant research evidence (rather than needing to run searches for themselves). A schematic showing how the push model could work in practice is shown in Fig. 1.

The databases covered by these automated searches need to cover the locations where relevant literature may be found comprehensively and there are two potential limitations in our current information infrastructure: (1) not all databases support the regular running of specific searches (i.e., auto-alerts); and (2) many databases do not offer an open application programming interface (API) for third-party software to connect to. Such software can, for example, enable reviewers to conduct searches on these databases within bespoke systematic review software.

Systems which offer the possibility of regular comprehensive searches are emerging and include the Epistemonikos database ([epistemonikos.org](#), developed by the Epistemonikos Foundation) and the Health Database Advanced Search (HDAS—[hdas.nice.org.uk](#)—developed by the National Institute for Health and Care Excellence). These systems enable the user to search multiple databases simultaneously: Epistemonikos by conducting its own broad searches across a wide range of health databases and adding results to a large corpus of health research; and HDAS by using APIs to search multiple databases when the user conducts a search—or periodically, when “alerts” are set up.

In situations where a given database is necessary for a living systematic review, but is not supported by the kinds of services outlined previously, manual searching is

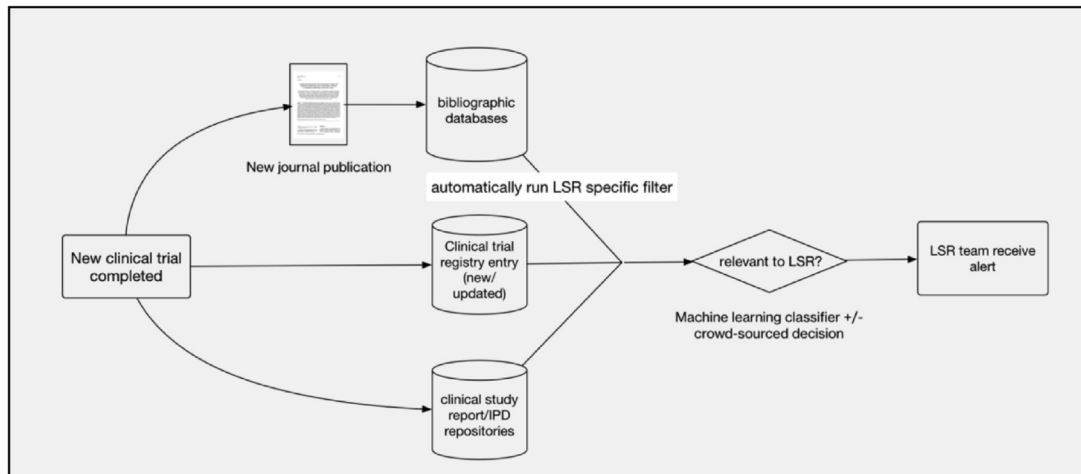


Fig. 1. The push model for adding new evidence to a living systematic review. Note that once the system has been set up, the surveillance process can run continuously and automatically without the need for manual intervention.

necessary—which may be an opportunity for community contribution. Because running a database search is a specific time-limited task that can be accomplished without the expenditure of significant time if individuals are each only responsible for one or two databases, the task of searching can be distributed efficiently across multiple individuals. Thus, the manual searching of individual databases and the downloading of search results are tasks that can be assigned to people who may not be involved in any other part of the living systematic review process. The automated and manual search results can then be deduplicated together, the process thus combining the efficiencies of automation (where possible) with distributed manual effort. This is not to say that specialist information science expertise is not required because the creation—and maintenance—of a good search strategy will always require specialist human skills; but that, once a search has been created, components of regularly running and downloading results may be efficiently distributed across a community.

2.1.2. Eligibility assessment

Once potentially relevant studies have been identified through database searches and other sources, they need to be checked for their eligibility to the review. Here a number of human and machine enablers can contribute to the efficiency of this process: generic classification, review-specific classification, and crowdsourcing.

“Text classification” is a standard machine-learning problem in which the aim is to categorize texts (in this case, biomedical texts) into groups of interest (here, citations that do and do not meet the review’s inclusion criteria). When used for eligibility assessment, text classification entails the use of machine learning to exclude irrelevant citations automatically. It usually operates on the titles and abstracts of citations identified through database searches and provides a probability score that a given citation is, or is not, of interest. Machine-learning classifiers need to “learn”

from gold standard classifications, which have usually been generated by human experts.

For example, the Cochrane Crowd, a citizen science platform which enables anyone to contribute to reviews via microtasks (see the following), has collectively classified over 415,000 records as to whether they describe a randomized controlled trial (RCT) or not. These decisions have been used to build a machine-learning model which can predict how likely a new citation is to be describing an RCT.

When trained with large amounts of high-quality data, as in the case of the aforementioned RCT classifier, classifiers can be very accurate. The RCT model, for example, is able to exclude 60–80% of irrelevant records retrieved from a database search while maintaining a sensitivity of over 99% [5]. The impact of this in a typical review is depicted in Fig. 2. The area covered by the gray square depicts the articles retrieved in the search and the yellow/orange rectangle the number of RCTs retrieved in that search. The classifier can reduce the burden of screening for reviewers to only those found in the pink rectangle, for the loss of less

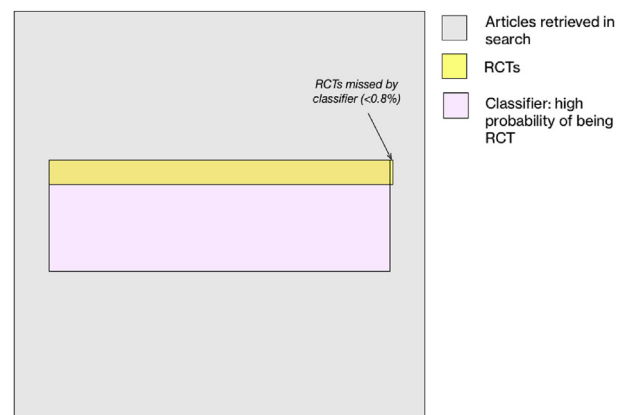


Fig. 2. Scaled rectangle diagram illustrating screening workload saved when using the RCT classifier developed from the data annotated by Cochrane Crowd. RCT, randomized controlled trial.

than 1% of the RCTs retrieved. This is a nice example of a human/machine workflow operating in mutually reinforcing ways: as the Cochrane Crowd screens more citations, the RCT Classifier can become more accurate; and as the Classifier becomes more accurate, it is able to make the Cochrane Crowd more efficient, by removing from the process those citations that are clearly not RCTs.

Thus, any given systematic review can take advantage of the considerable workload savings that these “generic” classifiers offer—which aim to apply one or more of the eligibility criteria of the review in question. For example, a generic RCT classifier is suitable for use in all reviews which only include RCTs but is unsuitable for use in reviews which also include observational studies, as it would eliminate these. The creation of generic classifiers requires the existence of suitable training material.

At the time of writing, there are good classifiers available for RCTs [5], and members of the author team are currently developing classifiers for systematic reviews and economic evaluations; but there are no high-performing classifiers to identify, for example, diagnostic test accuracy studies because training data are not available. Thus, not all systematic reviews will be able to take advantage of the workload savings afforded using generic classification.

All is not lost though, as machine learning can assist many systematic reviews through review-specific classification. These are custom-built models that do not rely on there being large amounts of data external to the review that delineates its area of interest (i.e., the RCT example, above). Instead, review-specific models learn directly from the human screening decisions that were made when the original review was conducted. Here, it is possible to build a model that directly predicts the probability that a given study is, or is not, relevant to the review. This is potentially more specific to the review in question than the generic classification example, as such a classifier would implicitly combine the study-type specific classification (e.g., the RCT classifier) with topic-specific classification (i.e., condition, population, etc.). There are some important considerations. First, the most significant drawback to using this type of classifier is the lack of training data, although there are techniques for using machine learning even in entirely new reviews where humans and computers operate in a workflow which prioritizes the relevant items for manual screening while simultaneously improving the accuracy of the machine model [14]. Even if reviewers have previously considered, for example, 10,000 citations, this is far less than might be used for building a generic classifier, and its accuracy is likely to be lower. Second, it is difficult to know how reliable the classifier will be when assessing “unseen” citations, as it will not have had the benefit of “seeing” the wide variety of relevant citations that might be available. Thus, the generic classifiers are highly accurate with respect to the task they have been trained to perform (performing general categorizations), but they will not be able to classify abstracts in a way that directly aligns

with the nuances of inclusion criteria defined by individual reviews. By contrast, review-specific classifiers will tend to be highly attuned to the context of the criteria for a review, but may be less accurate, due to the lack of training material. Both are developed by humans and machines operating in mutually beneficial ways, illustrating one of the key themes of this paper.

Automation is currently unable to perform the entirety of eligibility assessment, and it is here that humans in the form of “crowds” are particularly useful. In May 2016, Cochrane launched *Cochrane Crowd*, an online platform where contributors can sign up to help identify and describe health evidence (Fig. 3). To date, more than 6,000 people have signed up and together have identified over 33,000 reports of randomized controlled trials for Cochrane’s Central register of controlled trials (CENTRAL). The Crowd is supported by brief, interactive online training, and accuracy is ensured using an agreement algorithm that requires each record to be classified multiple times before it is either submitted to CENTRAL or rejected. Evaluations of Crowd performance have shown 99% crowd sensitivity (the Crowd’s ability to identify RCTs correctly) with respect to a reference standard set of annotations, and 99% crowd specificity (the Crowd’s collective ability to correctly identify the records that should be rejected).

Cochrane Crowd represents an important shift in the way Cochrane seeks and manages information about studies within the organization, as it facilitates a move away from the traditional siloed approach to identifying evidence on a peer-review basis, to an upstream model that makes the best use of human effort, providing new opportunities for contributors to play a very real practical role in managing the data deluge.

We should note that the performance metrics mentioned previously have suggested that, for example, the recall of 99% of relevant citations may be “good enough” for the purposes for a systematic review, and that this may be at variance to an important principle of systematic reviews that all relevant evidence should be included. It is true that, in some reviews, even the loss of one study may change conclusions, and so systems and processes which cannot guarantee 100% recall need to be looked at critically. However, we would argue that compromises between what is ideal and what is possible to achieve in terms of the breadth of searches conducted are already part of systematic review practice; that highly intensive human effort is unlikely to result in perfect decisions, as humans make mistakes, especially when fatigued by significant time spent on repetitive tasks; and that sensitive electronic searches are only one of the ways in which eligible studies are identified. For these reasons, we consider that the imperfect, but extremely high, recall achieved in the aforementioned examples is likely to be as good as, if not better, than that often achieved using more conventional approaches.

After the titles and abstracts of potentially relevant studies have been assessed, the full texts of reports are then

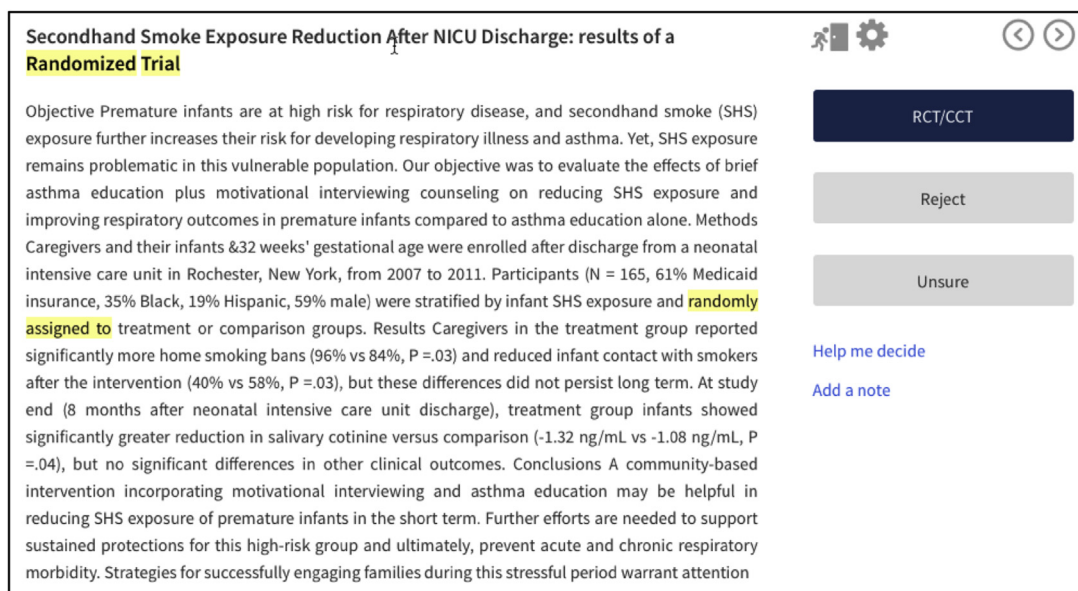


Fig. 3. The Cochrane Crowd user interface.

retrieved for final selection and inclusion in the review; again, humans and crowds can assist in this process. Services such as the CrossRef API (<http://search.crossref.org/>) can be used to automate the discovery of the locations of full-text reports and papers and, although automation can help, it often takes human effort and judgment to track down all papers, and navigate permissions on subscription content.

The assessment of full-text papers for inclusion in the review is not a stage in the workflow where automation has yet been developed, probably because the potential for workload reduction here is so much less than is the case for citation screening. The assessment of reports to check whether they are relevant for the review is a task that can be shared out across a distributed team though, thus keeping any one individual's workload to an acceptable minimum.

Finally, the prospective identification of relevant studies is increasingly possible thanks to the growth of data deposited in trials registers, where studies are registered before their commencement. In the future, systematic reviews could benefit from an alerting mechanism to let members know when a trial might be expected to report—and possibly communication with authors to ensure no relevant papers are missed. The utilization of trial registries, such as ClinicalTrials.gov, may be an important enabler for review efficiency, given that they can contain quite detailed outcome data, in a more structured and reanalyzable way than publications. Although we may be some way off being able to reuse outcome data in meta-analyses, having detailed and machine-readable data on outcomes will enable us, for example, to determine eligibility at a more granular level and check for outcome reporting bias more effectively.

2.2. Data extraction/collection and risk of bias assessment

Once studies have been checked for their eligibility and included in the review, information about the study, including study characteristics and results data, needs to be collected about them in a standardized way and their risks of bias, relating to how they have been conducted and assessed.

Automation technologies here are still being developed, but they include the extraction of structured data from tables and graphs in PDF files, thus saving time and also increasing reliability (e.g., ContentMine, <http://contentmine.org/> and Graph2Data, <https://github.com/EPPI-Centre/Graph2Data>).

Assessment of the studies' risk of bias can also be partially automated, given sufficient human-created training material, as exemplified by the RobotReviewer tool, which can assess the risk of bias associated with RCTs with an accuracy similar to that of humans [6]. RobotReviewer has "learned" to apply the Cochrane Risk of Bias tool (version 1) from more than 12,000 manually completed assessments included in Cochrane reviews. The forthcoming update of the Risk of Bias tool to version 2 presents a challenge for future machine-learning tools because there currently exist few (or no) human-conducted examples to learn from; this may be another opportunity for a human-machine workflow to be created, to develop training data for automation, which then makes the manual task more efficient in the future.

One option under consideration is whether automatic data extraction (e.g., by RobotReviewer) might replace one of two human reviewers (because many systematic review tasks are performed in duplicate to reduce error and bias). This semi-automation strategy might still substantially reduce workload

but leaves all output being manually verified. The impact of such a strategy on efficiency and data accuracy is unclear. The results of real-world pilot studies will help determine how to best fit automation tools into the workflow.

2.3. Synthesis and reporting

Automation technologies exist for generating sections of a review based on templates, based on its quantitative findings (e.g., RevMan HAL: schizophrenia.cochrane.org/revman-hal-v4). However, this area is very much under development, and it is unlikely that most systematic reviews will use this kind of automation. This said, there may be a role for automation in determining whether a review may need updating or should be prioritized for immediate updating, by automatically identifying the number of participants in the study (and, potentially, the direction of its findings), and estimating how likely it is that a study of its size would be able to change the result of the systematic review significantly. For example, if the review already contained 80 trials, the appearance of one additional small trial would be unlikely to alter the review's conclusions; but a large new trial in an analysis which did not contain many studies might well require more urgent attention. (See also paper 3 in this series for a discussion on specific statistical issues relating to updating meta-analyses.)

In the future, systems that assist authors and peer reviewers will be of benefit to living systematic reviews, and there are no fully developed systems yet available. Important features include the automatic incorporation of new studies in the appropriate analyses, and the ability to highlight to authors and readers, those sections which have recently changed. Such systems will also aid peer reviewers in highlighting the newly incorporated data and enabling them to focus their attention on the sections of the review that have changed in the light of new evidence.

2.4. Looking ahead to new research evidence surveillance systems

We have examined here the possibilities for human and machines to contribute to the workflow in an individual living systematic review. However, it may be that real gains in efficiency will come when reviewers work together at the research curation stage, developing mutually supportive systems that reduce duplication of effort (e.g., a new study is assessed once for its relevance to multiple living systematic reviews). Critical here will be the development not only of standards which enable systems to interchange data efficiently (e.g., linkeddata.cochrane.org/) but also of material and cultural incentives to share research data openly for the public good. Fig. 1 reflects the growing openness of trial data, in its inclusion of an individual participant data (IPD) repository, although we should not allow a focus on IPD to obscure wider issues of improving trial transparency [15].

3. Conclusion

We have provided an overview of the ways in which human and machine technologies can combine to create workflows that enable living systematic reviews to be maintained with relatively limited time input from any one individual. Automation technologies are ready for use in the early stages of a living systematic review—in assessing the eligibility of studies for inclusion—but require further research and development in the latter stages. While the context for this series of papers is “living” systematic reviews, many of these enabling technologies apply equally to standard systematic reviews.

References

- [1] Higgins JPT, Green S. editors, Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available at www.cochrane-handbook.org.
- [2] Elliott JH, Synnot A, Turner T, Simmonds M, Akl E, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when and how. *J Clin Epidemiol* 2017;91:23–30.
- [3] 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(2):e1001603.
- [4] Thomas J, Noel-Storr A, Elliott J. Human and machine effort in Project Transform: how intersecting technologies will help us to identify studies reliably, efficiently and at scale. *Cochrane Methods* 2016;1:37–41.
- [5] Wallace BC, Noel-Storr A, Marshall IJ, Cohen AM, Smalheiser NR, Thomas J. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. *J Am Med Inform Assoc* 2017;24:1165–8.
- [6] Marshall I, 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.
- [7] Tsafnat G, Dunn A, Glasziou P, Coiera E. The automation of systematic reviews. *BMJ* 2013;346:f139.
- [8] Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. *Syst Rev* 2014;3(1):74.
- [9] 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(1):5.
- [10] Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. *Res Synth Methods* 2011;2(1):1–14.
- [11] 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(9):e1000326.
- [12] Shemilt I, Khan N, Park S, Thomas J. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews. *Syst Rev* 2016;5(1):140.
- [13] Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological Expectations of Cochrane Intervention Reviews (MECIR). Standards for the conduct and reporting of new Cochrane Intervention Reviews, reporting of protocols and the planning, conduct and reporting of updates. London: Cochrane; 2016.
- [14] Wallace BC, Trikalinos TA, Lau J, Brodley CE, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics* 2010;11(1):55.
- [15] Hoffmann T, Glasziou P, Beller E, Goldacre B, Chalmers I. Focus on sharing individual patient data distracts from other ways of improving trial transparency. *BMJ* 2017;357:j2782.