PERSPECTIVE
For reprint orders, please contact: reprints@futuremedicine.com

Modernizing the systematic review process to inform comparative effectiveness: tools and methods

Journal of **Comparative Effectiveness Research**

Byron C Wallace[1], Issa J Dahabreh[1], Christopher H Schmid[1], Joseph Lau[1] & Thomas A Trikalinos*[1]

[1]Center for Evidence-Based Medicine, Program in Public Health, Brown University, Providence, RI 02906, USA
*Author for correspondence: thomas_trikalinos@brown.edu

Systematic reviews are being increasingly used to inform all levels of healthcare, from bedside decisions to policy-making. Since they are designed to minimize bias and subjectivity, they are a preferred option to assess the comparative effectiveness and safety of healthcare interventions. However, producing systematic reviews and keeping them up-to-date is becoming increasingly onerous for three reasons. First, the body of biomedical literature is expanding exponentially with no indication of slowing down. Second, as systematic reviews gain wide acceptance, they are also being used to address more complex questions (e.g., evaluating the comparative effectiveness of many interventions together rather than focusing only on pairs of interventions). Third, the standards for performing systematic reviews have become substantially more rigorous over time. To address these challenges, we must carefully prioritize the questions that should be addressed by systematic reviews and optimize the processes of research synthesis. In addition to reducing the workload involved in planning and conducting systematic reviews, we also need to make efforts to increase the transparency, reliability and validity of the review process; these aims can be grouped under the umbrella of 'modernization' of the systematic review process.

KEYWORDS: comparative effectiveness research ▪ data mining ▪ health technology assessment ▪ informatics ▪ machine learning ▪ meta-analysis ▪ network meta-analysis ▪ statistical method

Comparative effectiveness research aims to inform decision-making by providing evidence on the benefits, harms and costs of all alternative interventions that are relevant to a particular clinical question [1]. Comparative effectiveness research comprises a diverse range of activities [101]: prioritizing which questions to study; generating primary data on the relative effectiveness and safety of interventions; synthesizing the totality of existing research with systematic reviews (appropriately called comparative effectiveness reviews) and meta-analyses (in particular multiple treatment or network meta-analysis); contextualizing research findings via decision and economic analysis; disseminating findings to consumers and stakeholders, including patients, physicians and other decision-makers; and putting findings into practice. Each of these components may inform the other; thus, comparative effectiveness reviews (systematic reviews of all clinically relevant alternative interventions) have an important role in the comparative effectiveness research agenda.

However, there is a pressing need to modernize systematic review methodology. One challenge is that the number of published biomedical articles, which is already vast, is growing at an increasing rate with no signs of plateauing in the foreseeable future. PubMed already contains more than 600,000 publications on clinical trials in humans (and more than 22 million articles in total); in 2010, an average of 75 new clinical trial reports were published every day [2]. Accordingly, a large number of systematic reviews are published every year [3].

Future Medicine part of fsg

Exacerbating this challenge of information overload, the standards for conducting systematic reviews and meta-analyses are more demanding now than they were only a few years ago. For example, the Institute of Medicine has recommended 21 standards and 82 elements of performance [102], the Methodological Expectations of Cochrane Intervention Reviews (MECIR) initiative includes almost 100 recommendations [4] and the internal guidance of organizations that routinely conduct systematic reviews is not far behind [5]. Reporting standards have been proposed by several groups reflecting methodological progress [6,7]. Strict adherence to all recommendations inevitably increases the effort involved in conducting systematic reviews or other research synthesis activities (e.g., evidence mapping and rapid reviews).

At the same time, as a result of the emphasis on the comparative effectiveness of alternative active interventions and increasing engagement of a broad range of stakeholders in developing and refining the focus of systematic reviews, the key questions that systematic reviews address are becoming more complex. Complex questions necessitate elaborate refinement and, sometimes, advanced methodologies for quantitative synthesis (meta-analysis).

Over a decade ago, Allen and Olkin calculated that a well-conducted systematic review and meta-analysis can take between 1000 and 2000 person-hours [8]. In our experience, the time to conduct a systematic review has not decreased in the last three decades. For example, evidence reports, prepared by Evidenced-Based Practice Centers (EPCs) for the US Agency for Healthcare Research and Quality, take more than 1 year to complete, and this has not changed in the last 15 years, despite any accumulated procedural experience and know-how in this continuously operating and well-respected program. The comparative effectiveness reviews introduced in 2005 are often even more demanding.

Challenges, owing to the growing volume of primary data, the higher methodological bar for evidence synthesis and the increasingly complex questions that are being addressed, may further prolong the completion of systematic reviews and efforts to keep them up-to-date. Rigorous processes for question refinement, including stakeholder engagement, further extend timelines. Long timelines generate high costs and can make systematic reviews unavailable to consumers who often need the information as soon as possible to inform their decisions [9–11]. For systematic reviews to be of maximum utility, the summarized evidence must be both useful and timely; the longer the process, the less timely the evidence (when published). Pragmatic resource constraints necessitate modernizing all stages of the systematic review pipeline to shorten the time necessary to produce comparative effectiveness reviews, minimize unnecessary replication of effort, solicit and capitalize on community input, and make the whole process more reliable and transparent.

In this article we review the steps of the systematic review process and discuss opportunities to optimize them. Some improvements will streamline processes to remove unnecessary redundancies; some will necessitate the development of new, publicly available resources [12,13]; and others will require the development and application of novel methodologies and tools [14,15]. Thus, modernization will require both technical innovations and changes to the processes of conducting comparative effectiveness reviews. We will discuss some of the efforts that are already underway and that show, at least, preliminary promise.

Systematic review pipeline

As shown in Figure 1, a systematic review follows a well-defined process that aims to minimize bias and subjectivity by completing a series of steps:

- For a given topic, one first generates clear and answerable key questions that describe the populations, interventions, comparators, outcomes and study designs of interest. Often, multiple distinct, but complementary, key questions are of interest, and these can be organized in an analytic framework – that is, a graphical depiction of the logic behind the questions that the systematic review aims to address. A protocol is then developed that outlines the implementation details for the steps that follow;
- Subsequently, one tries to identify studies that fulfill the eligibility criteria dictated by the key questions and prespecified in the review protocol. At a minimum, this involves searching electronic databases of published research. However, it can also include manually searching through selected journals, conference proceedings and unpublished research. Typically, search strategies are constructed to be as sensitive as practically feasible (they cast a broad net to identify as many eligible studies as possible);

- The next step involves screening the citations identified by the searches to select those that appear to be relevant to the systematic review on the basis of their titles and/or abstracts. The full texts of these citations are then retrieved and, if they fulfill the protocol's eligibility criteria, they are included in the systematic review;
- Each eligible paper is reviewed in depth. This includes extracting predefined pieces of data from each paper, including information for quantitative analysis (meta-analysis), and evaluating the methodological validity and reporting clarity of each paper;
- The penultimate step includes synthesizing the information extracted from eligible studies. Qualitative synthesis is always possible, but quantitative analyses of the extracted data can often provide additional insights. Depending on the type of available data and the questions being addressed, meta-analysis methods range from simple to fairly advanced statistical techniques;
- Finally, the findings of the systematic review are organized and reported to the consumers of systematic reviews, including physicians, researchers, payers and other decision-makers, and patients and caregivers.

One way of modernizing the systematic review pipeline is to optimize and improve each step in the process in turn, keeping in mind how the steps interact. Table 1 lists potential examples of step-specific optimizations; some of these are still in prototype, while others have already been deployed. In the rest of this article, we primarily focus on steps 3–5, as these are the most amenable to technical innovation and have been the focus of our own work on modernizing systematic reviews in recent years. That said, there is ongoing research interest in topic preparation [16,17], search strategy development [18] and in dissemination of evidence synthesis products [19].

As can be seen in Table 1, computational tools will play a key role in optimizing the systematic review process, particularly in the identification and extraction steps. We review some work in this direction in the section 'Optimizing literature screening'. Specifically, we review work on optimizing citation screening, including practical software tools and data-mining methods that look to semi-automate the task. We discuss the potential of such tools to optimize the otherwise daunting task of keeping published (or 'living') reviews and synopses of whole domains ('field synopses') up-to-date. We also consider work on reducing the effort involved in data

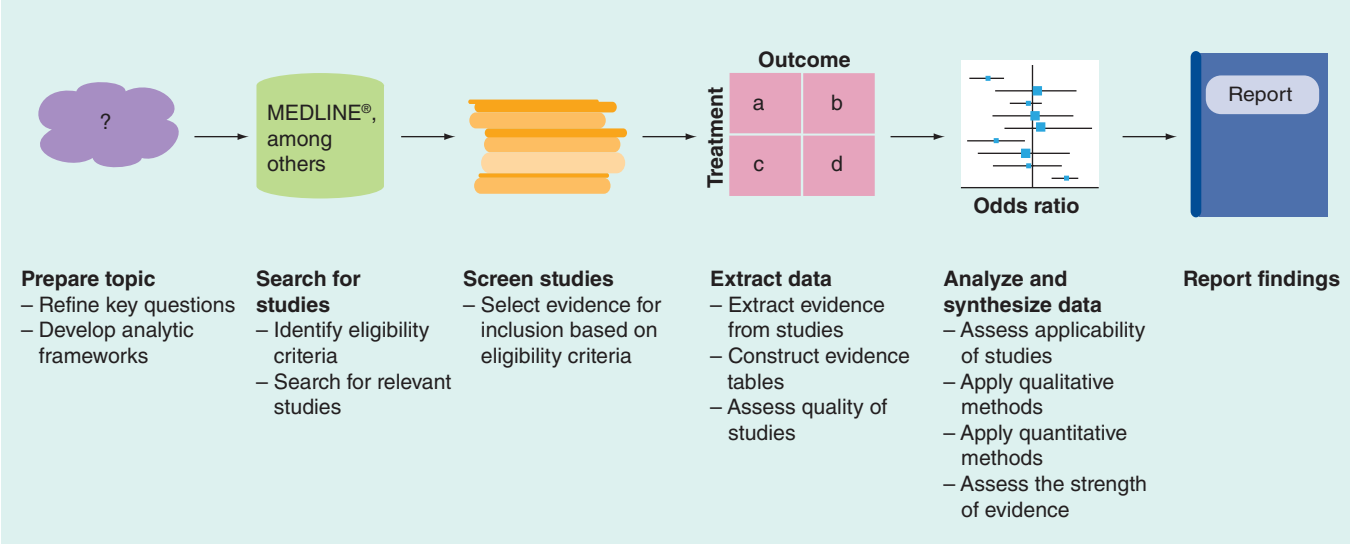

Figure 1. Systematic review pipeline. Researchers first formulate the precise clinical question to be answered and from this they derive a search query. This query is used to retrieve potentially eligible studies from databases of literature, such as MEDLINE®. Only a fraction of the studies retrieved via this search will be eligible for inclusion in the review; researchers must read through all of the retrieved citations to determine which studies are available. Once the set of eligible studies is identified, researchers extract the data elements of interest from the text. The final step is to statistically synthesize these extracted elements to gain an overall estimate. 'Downstream' work (keeping the systematic review current) is not shown here.

Table 1. Nonexhaustive list of examples of potential targets for modernizing systematic reviews.

Where in the pipeline?	Examples of targets for optimization or advancement of methods (nonexhaustive)
Across all steps	Comprehensive software platform for conducting systematic reviews in 'real time' Development of methods for crowd sourcing conduct of systematic reviews
Individual steps	
Prepare topic	Improvement of formal methods for prioritizing which topics to review Coordination of review efforts on a national or global scale
Search for studies	Software tools to facilitate formulating searches Machine translation for accessing databases in various languages
Screen studies	Software to facilitate abstract screening (facilitating collaborative, remote screening over the web) Software tools for semi-automating citation screening Develop a framework for guiding the selection of studies for review, particularly in topics where randomized comparative trials are scarce
Extract data	Online tools to facilitate form generation and data extraction Repository of extracted data for reuse Software tools for semi-automating data extraction
Analyze and synthesize data	Advancement of statistical methods for evidence synthesis Easy-to-use software tools for qualitative synthesis and meta-analysis More widespread use of modeling to extrapolate beyond the follow-up horizon of clinical trials, the incorporation of evidence from 'nonstandard' research designs and the synthesis of data on surrogate and patient-relevant clinical outcomes Incorporation of primary (observational) data analyses conducted in parallel with systematic reviews to 'fill the gaps' in clinical-trial evidence
Report findings	Software tools for rapid reporting of completed reviews Formal methods (e.g., value of information analyses) for identifying knowledge gaps and future research needs based on completed reviews

extraction, including emerging work on semi-automating data extraction and 'crowd sourcing' data extraction via an online repository. The described tools independently address different steps in the pipeline. Eventually, however, we envision the development of an integrated systematic review 'platform' that guides reviewers through the entire process, from query formulation and literature searching to analysis and report generation.

Modernization is not only a question of saving time. Published studies are increasingly diverse in terms of their design; for example, which interventions are compared often varies between studies. Network meta-analysis comprises a family of methods that estimate the relative effectiveness of two treatments, even when they have not been directly compared [20,21]. We discuss network meta-analysis as an example of statistical advances that have emerged to address questions on the comparative effectiveness and safety of patient management strategies, and argue that software tools to facilitate the conduct of network meta-analysis are needed. We conclude with an outline of potential future steps toward modernizing the conduct of systematic reviews.

Optimizing literature screening

Owing to the vast volume of published biomedical literature, it is difficult to keep up with new information, both in terms of conducting new reviews and in terms of updating existing reviews [2]. Much of this workload is due to the requirement that systematic reviews be as comprehensive as possible, encouraging broad literature searches. Reviewers must then screen all of the articles retrieved from such searches to assess their relevance. This translates into a huge amount of effort. Printing abstracts out and screening hard copies is not ideal.

Software tools, both commercial (and available for a fee) and open-source (and hence available free of charge), are now available to facilitate this process. DistillerSR™ [22] and EPPI-Reviewer [23] are commercial, web-based offerings that allow multiple users to screen abstracts for a given review (in addition to abstract screening, they also facilitate full-text screening and data extraction). We have also developed open-source web-based software for abstract screening called abstrackr [24,103,104]. All of these programs allow data to be imported from a variety of sources (e.g., PubMed ID lists or citations exported from popular

reference-manager software) and support multiple participants simultaneously screening abstracts. In addition, they provide a digital 'paper trail' of who screened what, and which abstracts have been designated as relevant (i.e., which have been 'screened in'). Each of these software tools comes with unique advantages and drawbacks, and other similar tools exist; a systematic survey of such technologies is outside of the scope of this paper.

Semi-automating screening

In addition to practical tools that assist with the logistics of abstract screening, there has been a great deal of research into using machine-learning methods to automate the task, thereby reducing reviewer workload. The task of identifying relevant studies from a pool of query results may be viewed as a classification task, in which the aim is to induce a model to automatically discriminate between 'relevant' and 'irrelevant' studies [14,15,25,26]. This is similar to the way a spam filter decides whether incoming e-mail is legitimate or not.

Preliminary results suggest that such methods can substantially reduce the number of citations that must be manually screened without greatly affecting sensitivity. However, for such methods to be of practical use, further empirical data are needed to demonstrate that computerized screening will not result in missed relevant studies for reviews conducted in a wide variety of topics. Furthermore, the tools must be made readily available to end-users, who will generally not have the wherewithal to implement machine-learning methods themselves. We are working on integrating methods for semi-automation into abstrackr; the tool already prioritizes screening abstracts that are likely to be relevant (according to the model). In the near future, we hope to allow the computer to automatically exclude irrelevant studies – that is, decide which citations require human judgment.

Keeping 'live synopses' up-to-date

There has been a movement toward online, continuously curated resources that summarize the state of the evidence using systematic review methods. These are sometimes called 'live synopses'. Live synopses are particularly popular in genetics, a field in which evidence accumulates rapidly. Four such resources in genetics are PDGene [27], AlzGene [28], SzGene [29] and COPDdb [30] for genetic determinants of Parkinson's disease, Alzheimer's disease, schizophrenia and chronic obstructive pulmonary disease, respectively. Live synopses are not limited to genetics; the Tufts CEA Registry, for example, summarizes information from published cost–effectiveness analyses. Methods for determining when to update such synopses and how to do so in a statistically appropriate manner are active research questions [31,32].

Live synopses, such as the above, require a considerable informatics infrastructure: a suitable database must be created and then populated; a front-end web for end-users that includes tools for exploring the data must be developed; and mechanisms for adding new studies and analyses must be put in place. Furthermore, as far as the authors are aware, the code for the informatics infrastructure of the aforementioned live synopses is not publicly available. Thus, an open-source effort to develop a flexible framework for serving live synopses may save future redundant effort.

The technologies previously mentioned for semi-automating screening may be particularly useful here, since the set of abstracts manually screened for the original review forms a large training corpus from which a classifier may 'learn' (estimate its parameters). Our recent work suggests that the workload involved in the literature-search phase of updating existing systematic reviews can be halved without missing relevant literature [33].

We therefore envision a web application for live synopses that provides services for both consumers and contributors of data. For example, the application would allow consumers to retrieve the data already curated, perhaps by performing meta-analyses of specific subsets of the data (e.g., as COPDdb allows). In addition, the application would incorporate functionality to make maintaining the database more efficient. The tool may, for example, search PubMed and automatically identify newly published studies that might meet eligibility criteria. The system may then e-mail the group maintaining the live synopsis to notify them of newly published and potentially eligible studies.

Making data extraction more efficient

Data extraction, the task of manually extracting the variables that are eventually synthesized and other clinically relevant attributes, requires a huge amount of time and resources. Compounding the problem, data elements from the same studies are often extracted multiple times by different groups, duplicating efforts. It would thus be efficient to maintain a repository of all of the data extracted from published articles to avoid

such wasteful redundancies. This is the aim of the Systematic Review Data Repository (SRDR) [34], an online extraction tool and data repository that facilitates the extraction process and makes the extracted data accessible to researchers worldwide, free of charge. In this way, extracted data will be searchable and reusable.

This collaborative tool (and others like it) may reduce redundant labor carried out by systematic reviewers working in isolation. Consider, for example, that multiple groups around the world may (re-)extract data from the same study independently, clearly representing wasted effort. Open, collaborative online tools thus represent an attractive means of reducing such wasted effort, as long as the community adopts them.

By distributing the extraction task across independent groups of researchers, the SRDR platform provides one means of 'crowd sourcing' the labor in systematic reviews. This distributed approach might improve data quality by allowing the community to curate and verify data over time, amending incorrect records when necessary. Applying crowd-sourcing to systematic review work is a controversial idea and would be riddled with complexity. A detailed discussion regarding the potential merits and downsides of this approach is beyond the scope of this article. While we have touted what we view as the advantages of SRDR, a potential drawback to this tool is the complexity inherent to such a flexible extraction system. We also note that other tools for performing data extraction exist, for example, Distiller SR [22], which integrates this functionality with screening. However, the biggest advantage of SRDR, in our view, is its potential as a repository; mitigating duplicated effort in abstraction has huge potential, should its promise be realized.

Another potentially fruitful research direction aimed at reducing human effort is the automatic extraction of relevant clinical elements from free-text. The idea is to automatically map free-text to clinically relevant variables – that is, those elements in which reviewers are interested. This would effectively semi-automate data abstraction. This is a very difficult problem from a data-mining perspective, but some progress has already been made. Many information-extraction systems have been developed for biomedical texts in general [35]. More specific to the task of automated data extraction for systematic reviews, Kiritchenko *et al.* describe a system that finds sentences containing relevant data elements with relatively high recall and precision [36]. Similar systems have also recently been proposed by other researchers [37,38]. These represent only initial attempts and an automated extraction system is probably far off. However, tools that assist reviewers (e.g., by highlighting areas of the text likely to contain inferred relevant clinical elements of interest) may be developed in the near future to help with extraction.

Finally, we note briefly that an additional issue in performing data extraction is the existence of articles published in languages other than English. Translating these to English prior to extraction would be costly and time-consuming; therefore, in practice, such studies are often discarded. Recent work by Balk *et al.* suggests that machine translation – that is, using computerized methods to automatically translate other languages into English, is now sufficient to facilitate accurate data extraction, at least in some languages [39]. Specifically, translation usually required minimal resources and, for Portuguese and German language articles, extractors had approximately 60% agreement, compared with 80% for articles written in English (performance was worse for Spanish, Chinese and Hebrew language articles). This may allow groups to include non-English studies in reviews at a relatively minor cost.

■ Modernizing statistical models for meta-analysis

Early meta-analytic methods have addressed the simple scenario in which two treatments are to be compared with respect to a single outcome over a specific time period. In reality, however, the data are more complex. Consider that, for a given condition, multiple treatments may be available from different drug companies, and all might claim to be the best. How can we know which is the best?

One could attempt a series of systematic reviews, each examining a pair of drugs, but this then requires the investigators to carefully piece the results together to form a coherent picture. This is made complicated by demographic differences between the patients involved in each comparison, variation in the eras in which the comparisons were made, the handling of multi-arm trials that encompass different formulations and/or different dosages, as well as statistical issues, such as multiple testing. Perhaps most importantly, some pairs of treatments may never have been directly compared with each other.

Network meta-analysis provides a framework for elegantly addressing these challenges,

providing valid estimation in a single model under explicitly formulated assumptions. Such a model allows incorporation of indirect evidence, increases the precision of estimation and permits principled ranking of treatments [20].

Network meta-analysis is predicated on the idea of indirect effect estimation. If treatment A is compared with treatment B, and B is compared with some other treatment C, we can estimate the relative efficacy of A to C indirectly based on their comparisons with the common reference B. This is illustrated in Figure 2, which shows an example of a network meta-analysis of ten alternative interventions for patients with chronic lymphocytic leukemia.

With a network meta-analysis, one can obtain estimates of the treatment effect between interventions that have never been directly compared, by taking into account relative comparisons with other treatments in the network. In the example illustrated in Figure 2, chlorambucil (a cheap, well-studied drug) has not been compared in head-to-head trials with pentostatin–rituximab-based chemoimmunotherapy (a recently developed, more expensive therapy). Nevertheless, by analyzing all of the data together, one can obtain an indirect treatment effect between these two treatments. For a technical tutorial on network meta-analysis, we refer the interested reader to existing resources [20,40]. We note that network meta-analysis is an emerging method and, hence, its limitations may not yet be fully appreciated. It also necessitates a substantial amount of effort for data collection.

An important practical consideration in network meta-analysis and related advanced methods is their availability to researchers. Easy-to-use, graphically driven software exists for basic meta-analysis, for example Comprehensive Meta-Analysis® [41] and Meta-Analyst® [42]; however, advanced (multivariate) analyses generally require using advanced statistical software, none of which is customized for the application. Bayesian models can be fit with

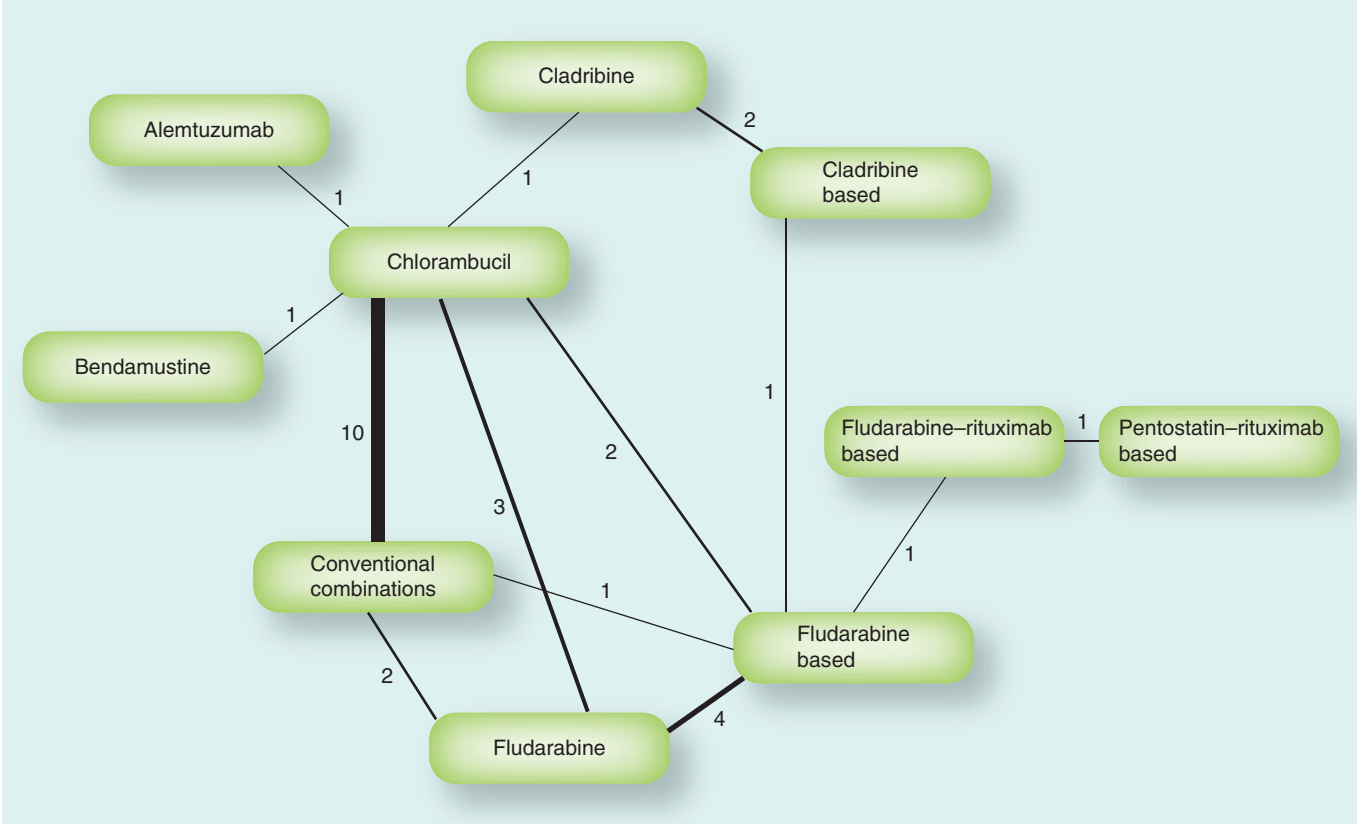

Figure 2. Example of a network of alternative interventions for treating patients with chronic lymphocytic leukemia, a hematological malignancy. The outcome is overall survival. The boxes represent the ten alternative interventions. Lines connect treatments that have been compared head-to-head in randomized controlled trials. Next to each line is the number of trials that compared the respective interventions, which is also encoded by the thickness of the line. The position of the boxes is arbitrary and chosen to enhance visibility.

one of the Bayesian inference Using Gibbs Sampling (BUGS) implementations, such as WinBugs/OpenBugs [43] or Just Another Gibbs Sampler (JAGS) [44].

That users must craft their own solutions with general-purpose software is a problem, as it hinders widespread adoption of multivariate meta-analytic methods. To address this problem, we have started work on Open Meta-Analyst [45], an open-source, cross-platform program that provides a graphical user interface for performing meta-analysis, but uses R software and JAGS to perform statistical calculations. The hope is that, in this way, cutting-edge statistical methods can be made available to end-users who may not have the skills necessary to implement them themselves. Of course, care must be taken to not simply treat such methods as a 'black-box'; users will therefore need to be trained in the interpretation of the output of the software.

Conclusion & future perspective

Systematic reviews, meta-analyses and other tools of evidence-based medicine have become established in the medical community as an invaluable approach to inform healthcare practices and policies. However, while systematic reviews have become more widely accepted, and hence are in greater demand than ever before, the evidence base has massively expanded and standards for conducting reviews have been raised. This means that producing reviews requires more labor than ever before. Furthermore, the complexity of the evidence has increased as stakeholders ask increasingly complex questions, necessitating more sophisticated statistical methods to provide meaningful summaries. Finally, pragmatic constraints on available funding necessitate that researchers surmount these obstacles with fewer resources. In light of these changes, we must modernize the practice of systematic reviews.

In this article, we have highlighted specific areas in which progress is being made toward modernization. These include methods and tools for optimizing the abstract-screening and data-extraction phase of reviews, as well as the development of new statistical models that are more appropriate for addressing increasingly complex questions. However, the existing work we have reviewed is only a start. Indeed, as outlined in Table 1, opportunities for further modernization abound. To start with, we need to conduct research into formalizing criteria regarding what clinical questions ought to be prioritized for review. Even given a specific research question, in our view, the process of search-strategy formulation is currently suboptimal, (although some guidelines exist [18]). Methods and tools that formalize and optimize the query formulation and refinement process have the potential to make this step more rigorous and efficient. We note that further empirical assessments are needed to judge the utility of all of the innovations reviewed in this article. For example, while steps have been made toward optimizing the subsequent step of abstract screening (reviewed above), methods for semi-automating the screening step require further empirical validation before wide adoption.

Efforts to optimize data extraction are only just emerging. Further research (which will require more empirical data) is needed to assess the benefits and drawbacks of 'crowd sourcing' data extraction, as in the SRDR effort. If this and similar efforts are to be successful, the systematic review community will need to address thorny issues of data integrity, maintenance, ownership and the problem of academic credit. While some recent work has considered the task of semi-automating data extraction, innovations in data mining will be required to improve the performance of such approaches. Practical tools that facilitate data extraction are sorely in need of development.

Statistical pooling of complex data brings its own challenges. Novel methods for advanced meta-analysis are necessary to deal with the increasingly common multivariate datasets. Furthermore, if they are to be useful, such emerging meta-analysis methods will need to be made widely available to end-users, ideally via graphically based, easy-to-use software.

Owing to the high level of abstraction, part of the reason systematic reviews remain labor-intensive is because the current subtasks comprising the systematic-review pipeline are still mostly performed manually (Figure 1). Moreover, computerized tools for making these steps more efficient are disjointed, requiring necessary but time-consuming data migration. For example, a researcher may first search PubMed, then somehow import the retrieved studies into a suitable Microsoft® Excel file to track screening. Once studies are identified, one may extract relevant variables from the articles in an ACCESS database. Performing a quantitative analysis would require another tool, and formatting

results for presentation yet another tool. Currently, no single tool facilitates each step in the pipeline, at least to our knowledge. Clearly, this process could be streamlined by using a common platform that provides tools for each step in the pipeline.

Evaluating and summarizing clinical evidence and using the analyses for patient care or health policy decisions are complex activities. The explosion of published biomedical literature, and the increasingly complex nature of the questions being asked, has made performing syntheses more complicated still. Much work remains to be carried out to modernize the practice of evidence-based medicine.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Background

- Systematic reviews are an indispensable tool for informing medical decision-making and they have gained wide acceptance in this regard.
- The exponential growth of biomedical literature, coupled with more stringent guidelines, has made producing reviews increasingly laborious.
- New informatics tools and statistical methods are needed to modernize the practice of systematic reviews.

The systematic review pipeline

- A systematic review follows a series of well-defined steps, including topic development, search formulation and query refinement, abstract screening, data extraction, statistical synthesis and reporting of results.
- There is room in each of these steps for modernization. Methodological innovations have modernized the steps of abstract screening through statistical synthesis; these aim to increase efficiency and to provide appropriate statistical models for increasingly complex data.

Optimizing literature screening

- Literature screening is a timely endeavor; this problem is exacerbated by the rapidly growing biomedical literature base.
- Software tools exist to ease the process and screening may soon be semi-automated using methods from machine learning.

Data extraction

- Data extraction refers to the process of abstracting the data elements to be synthesized for a review from the free-text of articles. It is a very time-consuming process.
- Software tools exist to ease this task, and methods from natural language processing may one day automate or semi-automate this task.
- A repository of abstracted data would allow for reuse of abstracted elements, mitigating duplicated effort and increasing efficiency.

Modernizing statistical models for meta-analysis

- More complex (multivariate) statistical methods have been developed to handle increasingly complex data (and questions). More work in this direction is needed.

References

1 Institute of Medicine. *Initial National Priorities for Comparative Effectiveness Research*. National Academies Press, Washington, DC, USA (2009).

2 Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med.* 7(9), e1000326 (2010).

3 Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med.* 4(3), e78 (2007).

4 Higgins J, Churchill R, Tovey D, Lasserson T, Chandler J. Update on the MECIR project: Methodological Expectations for Cochrane Intervention. *Cochrane Methods* 2, 2 (2011).

5 Agency for Healthcare Research and Quality. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Agency for Healthcare Research and Quality, Rockville, MD, USA (2008).

6 Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med.* 151(4), 264–269, W64 (2009).

7 Stroup DF, Berlin JA, Morton SC *et al.* Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 283(15), 2008–2012 (2000).

8 Allen IE, Olkin I. Estimating time to conduct a meta-analysis from number of citations retrieved. *JAMA* 282(7), 634–635 (1999).

9 Shojania KG, Sampson M, Ansari MT *et al.* *Updating Systematic Reviews*. Agency for Healthcare Research and Quality, Rockville, MD, USA (2007).

10 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.* 147(4), 224–233 (2007).

11 Sampson M, Shojania KG, Garritty C, Horsley T, Ocampo M, Moher D. Systematic reviews can be produced and published faster. *J. Clin. Epidemiol.* 61(6), 531–536 (2008).

12 Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat. Genet.* 40(2), 124–125 (2008).

13 Yu W, Clyne M, Dolan SM *et al.* GAPscreener: an automatic tool for screening human genetic association literature in PubMed using the support vector machine technique. *BMC Bioinformatics* 9, 205 (2008).

14 Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics* 11, 55 (2010).

15 Cohen AM, Hersh WR, Peterson K, Yen PY. Reducing workload in systematic review preparation using automated citation classification. *JAMA* 13(2), 206–219 (2006).

16 Hoomans T, Seidenfeld J, Basu A, Meltzer D. *Systematizing the Use of Value of Information Analysis in Prioritizing Systematic Reviews.* Agency for Healthcare Research and Quality, Rockville, MD, USA (2012).

17 Chang SM, Carey TS, Kato EU, Guise JM, Sanders GD. Identifying research needs for improving healthcare. *Ann. Intern. Med.* 157(6), 439–445 (2012).

18 Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. *J. Clin. Epidemiol.* 62(9), 944–952 (2009).

19 Matchar DB, Westermann-Clark EV, McCrory DC *et al.* Dissemination of Evidence-based Practice Center reports. *Ann. Intern. Med.* 142(12 Pt 2), 1120–1125 (2005).

20 Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat. Med.* 21(16), 2313–2324 (2002).

21 Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat. Med.* 15(24), 2733–2749 (1996).

22 DistillerSR™. Evidence Partners Incorporated, Ontaria, Canada (2001).

23 EPPI-Reviewer 4. Evidence for Policy and Practice Information and Co-ordinating Centre, London, UK (2012).

24 Wallace B, Small K, Brodley C, Lau J, Trikalinos T. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In: *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. Association for Computing Machinery, NY, USA, 819–824 (2012).

25 Wallace BC, Small K, Brodley CE, Trikalinos TA. Active learning for biomedical citation screening. Presented at: *The ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. Washington, DC, USA, 25–28 July 2010.

26 Bekhuis T, Demner-Fushman D. Towards automating the initial screening phase of a systematic review. *Stud. Health Technol. Inform.* 160(Pt 1), 146–150 (2010).

27 Lill CM, Roehr JT, Mcqueen MB *et al.* Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database. *PLoS Genet.* 8(3), e1002548 (2012).

28 Bertram L, Mcqueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat. Genet.* 39(1), 17–23 (2007).

29 Allen NC, Bagade S, Mcqueen MB *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat. Genet.* 40(7), 827–834 (2008).

30 Castaldi PJ, Cho MH, Cohn M *et al.* The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. *Hum. Mol. Genet.* 19(3), 526–534 (2010).

31 Garritty C, Tsertsvadze A, Tricco AC, Sampson M, Moher D. Updating systematic reviews: an international survey. *PLoS One* 5(4), e9914 (2010).

32 Shekelle PG, Newberry SJ, Wu H *et al.* *Identifying Signals for Updating Systematic Reviews: A Comparison of Two Methods.* Agency for Healthcare Research and Quality, Rockville, MD, USA (2011).

33 Wallace BC, Small K, Brodley CE *et al.* Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining. *Genet. Med.* 14(7), 663–669 (2012).

34 Ip S, Hadar N, Keefe S *et al.* A web-based archive of systematic review data. *Syst. Rev.* 1(1), 15 (2012).

35 Simpson MS, Demner-Fushman D. Biomedical text mining: a survey of recent progress. *Mining Text Data* 465–517 (2012).

36 Kiritchenko S, De Bruijn B, Carini S, Martin J, Sim I. ExaCT: automatic extraction of clinical trial characteristics from journal publications. *BMC Med. Inform. Decis. Mak.* 10, 56 (2010).

37 Boudin F, Nie JY, Bartlett JC, Grad R, Pluye P, Dawes M. Combining classifiers for robust PICO element detection. *BMC Med. Inform. Decis. Mak.* 10, 29 (2010).

38 Kim SN, Martinez D, Cavedon L, Yencken L. Automatic classification of sentences to support Evidence Based Medicine. *BMC Bioinformatics* 12(Suppl. 2), S5 (2011).

39 Balk E, Chung M, Hadar N *et al.* *Accuracy of Data Extraction of Non-English Language Trials with Google Translate.* Agency for Healthcare Research and Quality, Rockville, MD, USA (2012).

40 Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res. Synthesis Methods* 3(2), 80–97 (2012).

41 Borenstein M, Hedges L, Higgins J, Rothstein H. *Comprehensive Meta-analysis Version 2 Manual.* Biostat, Englewood, NJ, USA (2005).

42 Wallace BC, Schmid CH, Lau J, Trikalinos TA. Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. *BMC Med. Res. Methodol.* 9, 80 (2009).

43 Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Stat. Med.* 28(25), 3049–3067 (2009).

44 Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Presented at: *3rd International Workshop on Distributed Statistical Computing.* Vienna, Austria, 20–22 March 2003.

45 Wallace B, Dahabreh I, Trikalinos T, Lau J, Trow P, Schmid C. Closing the gap between methodologists and end-users: R as a computational back-end. *J. Stat. Software* 49(5), 1–15 (2012).

Websites

101 Cochrane Handbook for Systematic Reviews of Interventions version 5.1.0. www.cochrane.org/training/cochrane-handbook

102 Finding what works in healthcare: standards for systematic reviews. http://books.nap.edu/openbook.php?record_id=13059

103 Abstrackr. http://abstrackr.tuftscaes.org

104 Abstrackr source code repository. https://github.com/bwallace/abstrackr-web