



STUDY PROTOCOL

REVISED Data extraction methods for systematic review (semi)automation: A living review protocol [version 2; peer review: 2 approved]

Lena Schmidt ¹, Babatunde K. Olorisade ¹, Luke A. McGuinness ¹, James Thomas ², Julian P. T. Higgins ¹

¹Bristol Medical School, University of Bristol, Bristol, BS8 2PS, UK

²UCL Social Research Institute, University College London, London, WC1H 0AL, UK

v2 First published: 25 Mar 2020, 9:210
<https://doi.org/10.12688/f1000research.22781.1>
 Latest published: 08 Jun 2020, 9:210
<https://doi.org/10.12688/f1000research.22781.2>

Abstract

Background: Researchers in evidence-based medicine cannot keep up with the amounts of both old and newly published primary research articles. Support for the early stages of the systematic review process – searching and screening studies for eligibility – is necessary because it is currently impossible to search for relevant research with precision. Better automated data extraction may not only facilitate the stage of review traditionally labelled ‘data extraction’, but also change earlier phases of the review process by making it possible to identify relevant research. Exponential improvements in computational processing speed and data storage are fostering the development of data mining models and algorithms. This, in combination with quicker pathways to publication, led to a large landscape of tools and methods for data mining and extraction.

Objective: To review published methods and tools for data extraction to (semi)automate the systematic reviewing process.

Methods: We propose to conduct a living review. With this methodology we aim to do constant evidence surveillance, bi-monthly search updates, as well as review updates every 6 months if new evidence permits it. In a cross-sectional analysis we will extract methodological characteristics and assess the quality of reporting in our included papers.

Conclusions: We aim to increase transparency in the reporting and assessment of automation technologies to the benefit of data scientists, systematic reviewers and funders of health research. This living review will help to reduce duplicate efforts by data scientists who develop data mining methods. It will also serve to inform systematic reviewers about possibilities to support their data extraction.

Keywords

Data Extraction, Natural Language Processing, Reproducibility, Systematic reviews, Text mining

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	2
version 2 (revision) 08 Jun 2020		✓ report
version 1 25 Mar 2020	✓ report	↑ ? report

1 **Emma McFarlane** ¹, National Institute for Health and Care Excellence, Manchester, UK

2 **Matt Carter** ¹, Bond University, Robina, Australia

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Lena Schmidt (lena.schmidt@bristol.ac.uk)

Author roles: **Schmidt L:** Conceptualization, Methodology, Project Administration, Software, Visualization, Writing – Original Draft Preparation;

Olorisade BK: Conceptualization, Methodology, Software, Writing – Review & Editing; **McGuinness LA:** Conceptualization, Methodology, Software, Writing – Review & Editing; **Thomas J:** Methodology, Supervision, Writing – Review & Editing; **Higgins JPT:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by the National Institute for Health Research [RM-SR-2017-09-028; NIHR Systematic Review Fellowship to LS and DRF-2018-11-ST2-048; NIHR Doctoral Research Fellowship to LAM]. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. LS funding ends in September 2020, but ideally further updates to this review will continue after this date.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Schmidt L *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Schmidt L, Olorisade BK, McGuinness LA *et al.* **Data extraction methods for systematic review (semi)automation: A living review protocol [version 2; peer review: 2 approved]** F1000Research 2020, 9:210 <https://doi.org/10.12688/f1000research.22781.2>

First published: 25 Mar 2020, 9:210 <https://doi.org/10.12688/f1000research.22781.1>

REVISED Amendments from Version 1

In the second version we incorporated changes based on the first two peer reviews. These changes included a clarification that the scope of data extraction that we are interested in goes beyond PICO's, we elaborated on the rationale of looking at data extraction specifically, we clarified the objectives and added some more recent related literature. We also clarified the proposed schedule for updating the review in living mode. We have also added a new author James Thomas.

Any further responses from the reviewers can be found at the end of the article

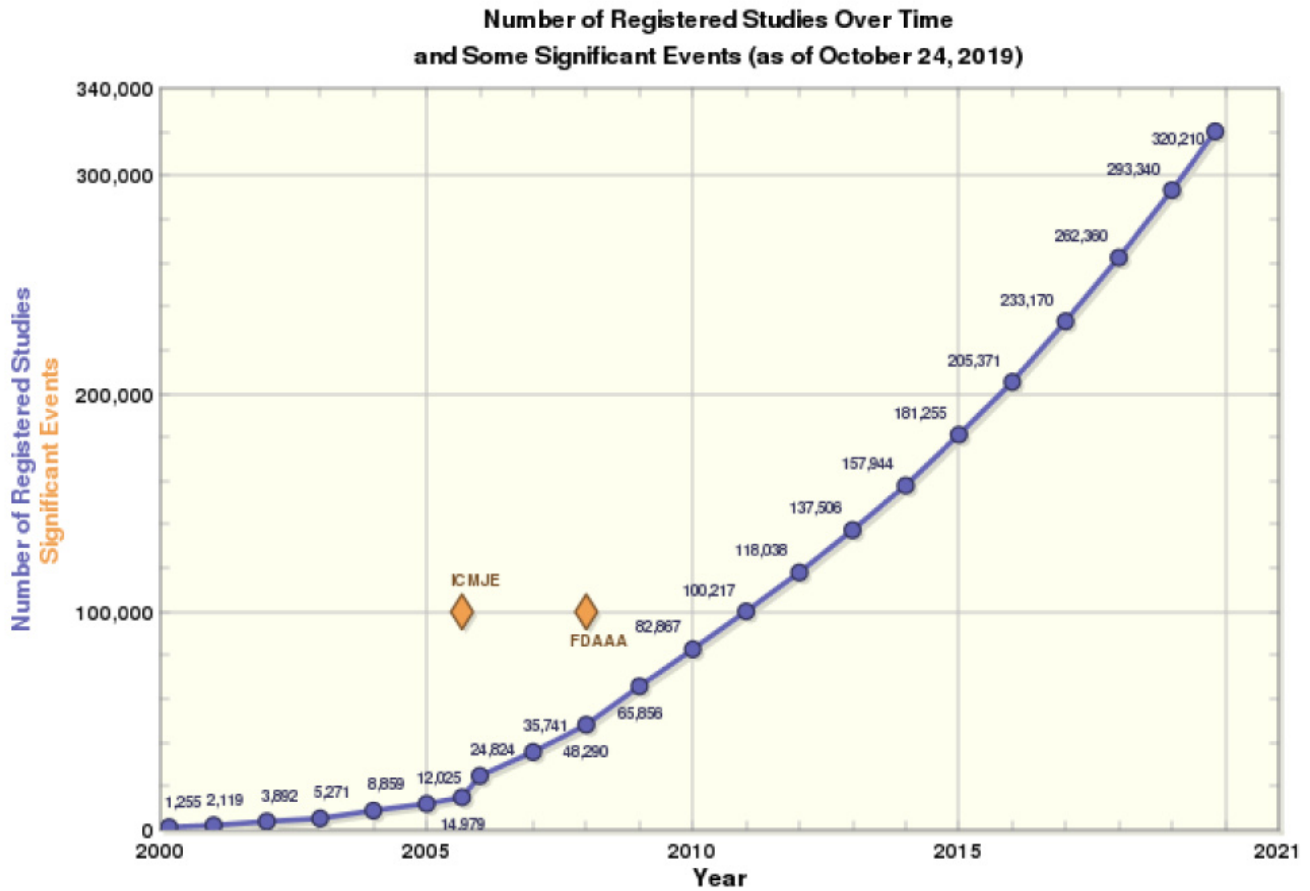
processing of information related to the PICO framework (Population, Intervention, Comparator, Outcome). A 2017 analysis of 195 systematic reviews investigated the workload associated with authoring a review. On average, the analysed reviews took 67 weeks to write and publish. Although review size and the number of authors varied between the analysed reviews, the authors concluded that supporting the reviewing process with technological means is important in order to save thousands of personal working hours of trained and specialised staff¹. The potential workload for systematic reviewers is increasing, because the evidence base of clinical studies that can be reviewed is growing rapidly (Figure 1). This entails not only a need to publish new reviews, but also to commit to them and to continually keep the evidence up to date.

Introduction
Background

Research on systematic review (semi)automation sits at the interface between evidence-based medicine and data science. The capacity of computers for supporting humans increases, along with the development of processing power and storage space. Data extraction for systematic reviewing is a repetitive task. This opens opportunities for support through intelligent software. Tools and methods in this domain frequently focused on automatic

Rapid development in the field of systematic review (semi)automation

Language processing toolkits and machine learning libraries are well documented and available to use free of charge. At the same time, freely available training data make it easy to train classic machine-learning classifiers such as support vector machines, or even complex, deep neural networks such as long



Source: <https://ClinicalTrials.gov>

Figure 1. Study registrations on ClinicalTrials.gov show an increasing trend.

short-term memory (LSTM) neural networks. These are reasons why health data science, much like the rest of computer science and natural language processing, is a rapidly developing field. There is a need for fast publication, because trends and state-of-the-art methods are changing at a fast pace. Preprint repositories, such as the [arXiv](#), are offering near rapid publication after a short moderation process rather than full peer review. Consequently, publishing research is becoming easier.

Why this review is needed

An easily updatable review of available methods and tools is needed to inform systematic reviewers, data scientists or their funders on the status quo of (semi)automated data extraction methodology. For data scientists, it contributes to reducing waste and duplication in research. For reviewers, it contributes to highlighting the current possibilities for data extraction and empowering them to choose the right tools for their task. Currently, data extraction represents one of the most time consuming¹ and error-prone (Jones, Remington, Williamson, Ashby, & Smyth, 2005) elements of the systematic review process, particularly if a large number of primary studies meet the inclusion criteria. Data mining, paralleled by automatic data extraction of relevant data for any specific systematic review project, has the potential to disrupt the traditional systematic reviewing process. This systematic review workflow usually follows the steps of searching, screening, and extracting data. If high-quality and curated data mining results are available then the searching and screening process is likely to change in the future. This review will provide constant surveillance of emerging data extraction tools.

Many systematic reviewers are free to use any tool that is available to them and need sufficient information to make informed decisions about which tools are to be preferred. Our proposed continuous analysis of the available tools will include the final performance metrics that a model achieves, and will also assess dimensions such as transparency of methods, reproducibility, and how these items are reported. Reported pitfalls of applying health data science methods to systematic reviewing tasks will be summarised to highlight risks that current, as well as future, systems are facing. Reviewing the available literature on systematic review automation is one of many small steps towards supporting evidence synthesis of all available medical research data. If the evidence arising from a study is never reviewed, and as a result never noticed by policy makers and providers of care, then it counts towards waste in research.

Aims of this review

This review aims to:

1. Review published methods and tools aimed at automating or semi-automating the process of data extraction in the context of a systematic review of medical research studies.
2. Review this evidence in the scope of a living review, keeping information up to date and relevant to the challenges faced by systematic reviewers at any time.

Our objectives are three-fold. First we want to examine the methods and tools from the data science perspective, seeking to reduce duplicate efforts, summarise current knowledge, and encourage comparability of published methods. Second, we seek to highlight contributions of methods and tools from the perspective of systematic reviewers who wish to use (semi)automation for data extraction: what is the extent of automation?; is it reliable?; and can we identify important caveats discussed in the literature, as well as factors that facilitate the adoption of tools in practice?

Related research

We have identified three previous reviews of tools and methods, two documents providing overviews and guidelines relevant to our topic, and an ongoing effort to characterise published tools for different parts of the systematic reviewing process with respect to interoperability and workflow integration. In 2014, Tsafnat *et al.* provided a broad overview on automation technologies for different stages of authoring a systematic review². O'Mara-Eves *et al.* published a systematic review focusing on text-mining approaches in 2015. It includes a summary of methods for the evaluation of systems (such as recall, F1 and related scores). The reviewers focused on tasks related to PICO classification and supporting the screening process³. In the same year, Jonnalagadda *et al.* described methods for data extraction, focusing on PICO and related fields⁴.

These reviews present an overview of classical machine learning and NLP methods applied to tasks such as data mining in the field of evidence-based medicine. At the time of publication of these documents, methods such as topic modelling (Latent Dirichlet Allocation) and support vector machines constituted the state-of-the art for language models. The age of these documents means that the latest static or contextual embedding-based and neural methods are not included. These modern methods, however, are used in contemporary systematic review automation software⁵.

Marshall and Wallace (2019)⁶ present a more recent overview of automation technologies, with a focus on availability of tools and adoption into practice. They conclude that tools facilitating screening are widely accessible and usable, while data extraction tools are still at piloting stages or require higher amounts of human input.

Beller *et al.*⁷ present a brief overview of tools for systematic review automation. They discuss principles for systematic review automation from a meeting of the International Collaboration for the Automation of Systematic Reviews (ICASR). They highlight that low levels of funding, as well as the complexity of integrating tools for different systematic reviewing tasks have led to many small and isolated pieces of software. A working group formed at the ICASR 2019 Hackathon is compiling an overview of tools published on the Systematic Review Toolbox website⁸. This ongoing work is focused on assessing maintenance status, accessibility and supported reviewing

tasks of 120 tools that can be used in any part of the systematic reviewing process as of November 2019.

Protocol

Prospective registration of this review

We registered this protocol via OSF (<https://doi.org/10.17605/OSF.IO/ECB3T>). PROSPERO was initially considered as platform for registration, but it is limited to reviews with health related outcomes.

Choosing to maintain this review as a living review

The challenges highlighted in the previous section create several problems. A large variety of approaches and different means of expressing results creates uncertainty in the existing evidence. At the same time, new evidence is being published constantly. Rapid means of publications necessitate a structured, but at the same time easily updatable review of published methods and tools in the field. We therefore chose a living review approach as the updating strategy for this review.

Search and updates

For literature searches and updates we follow the living review recommendations published by Elliott *et al.*⁹ and Brooker *et al.*¹⁰, as well as F1000Research guidelines for projects that are included in their living evidence collection. We plan to run searches for new studies every second month. This will also include screening abstracts of the newly retrieved reports. The bi-monthly interval for screening was chosen because we expect no sudden rise in relevant publications that could justify daily, weekly or monthly screening. The review itself will be updated every six months, providing that a sufficient quantity of new records are identified for inclusion. As a threshold for updating, we plan to use 10 new records, but we will consider updating the review earlier if new impactful evidence is published. We define impactful evidence as, for example, the publication of a tool that is immediately accessible to systematic reviewers and offers substantial automation of the data extraction process, or a tool that aims to change the traditional SR workflow. **Figure 2** describes the anticipated reviewing process in more detail.

Our Medline search strategy was developed with the help of an information specialist. Due to the interdisciplinary topic of this review, we plan to search bibliographic databases related to both medicine and computer science. These include Medline via Ovid and Web of Science Core Collection, as well as the computer science arXiv and the DBLP computer science bibliography. We aim to retrieve publications related to two clusters of search terms. The first cluster includes computational aspects such as data mining, while the second cluster identifies publication related to systematic reviews. The Medline search strategy is provided as *Extended data*¹¹. We aim to adapt this search strategy for conducting searches in all mentioned databases. Previous reviews of data mining in systematic reviewing contexts identified the earliest text mining application in 2005^{3,4}. We therefore plan to search all databases from this year on. In a preliminary test our search

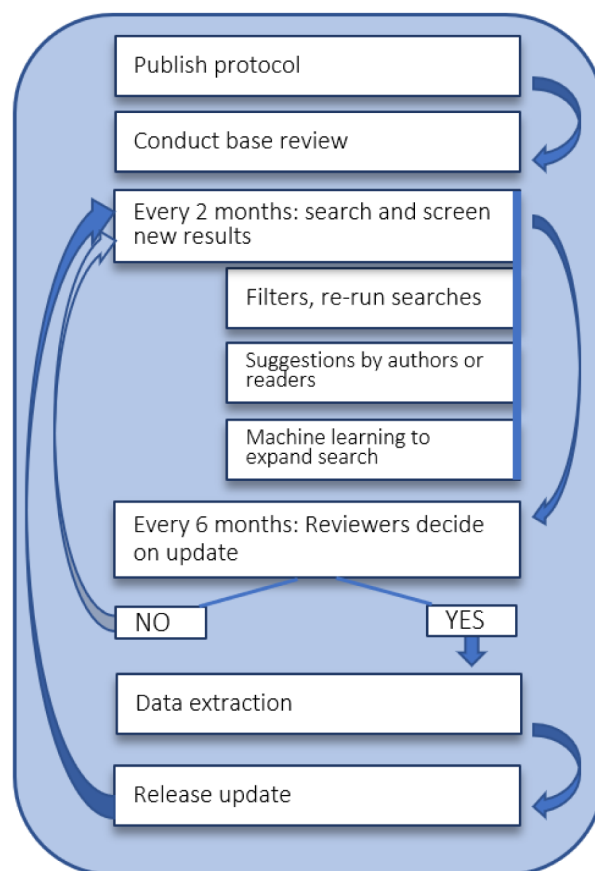


Figure 2. Continuous updating of the living review.

strategy was able to identify 4320 Medline records, including all Medline-indexed records included by O'Mara-Eves *et al.*⁵. We plan to search the Systematic Review Toolbox website for further information on any published or unpublished tools⁸.

Workflow and study design

All titles and abstracts will be screened independently by two reviewers. Any differences in judgement will be discussed, and resolved with the help of a third reviewer if necessary. The process for assessing full texts will be the same. Data extraction will be carried out by single reviewers, and random 10% samples from each reviewer will be checked independently. If needed, we plan to contact the authors of reports for clarification or further information. In the base review, as well as in every published update, we will present a cross-sectional analysis of the evidence from our searches. This analysis will include the characteristics of each reviewed method or tool, as well as a summary of our findings. In addition, we will assess the quality of reporting at publication level. This assessment will focus on transparency, reproducibility and both internal and external validity of the described data extraction algorithms. If we at any point deviate from this protocol, we will discuss this in the final publication.

All search results will be de-duplicated and managed with EndNote. The screening and data extraction process will be managed with the help of Abtrackr¹² and customised data extraction forms in Excel. All data, including bi-monthly screening results, will be continuously available on our Open Science Framework (OSF) repository, as discussed in the *Data availability* section.

Which systematic reviewing tasks are supported by the methods we review

Tsafnat *et al.*² categorised sub-tasks in the systematic reviewing process that contained published tools and methods for automation. In our overview, we follow this categorisation and focus on tasks related to data retrieval. More specifically, we will focus on software architectures that receive as input a set of full texts or abstracts of clinical trial reports. Report types of interest are randomised controlled trials, cohort, or case-control studies. As output, the tools of interest should produce structured data representing features or findings from the study described. A comprehensive list with data fields of interest can be found in the supplementary material for this protocol.

Eligibility criteria

Eligible papers

- We will include full text publications that describe an original natural language processing approach to extract data related to systematic reviewing tasks. Data fields of interest are adapted from the *Cochrane Handbook for Systematic Reviews of Interventions*¹³, and defined in the *Extended data*¹¹. We will include the full range of natural language processing (NLP) methods, including for example regular expressions, rule-based systems, machine learning, and deep neural networks.
- Papers must describe a full cycle of implementation and evaluation of a method.
- We will include reports published from 2005 until the present day, similar to O'Mara-Eves *et al.*³ and Jonnalagadda *et al.*⁴. We will translate non-English reports where feasible.
- The data that that included papers use for mining must be texts from randomised controlled trials, comparative cohort studies or case control studies in the form of abstracts, conference proceedings, full texts or part of the text body.

Ineligible papers

We will exclude papers reporting:

- methods and tools related solely to image processing and importing biomedical data from PDF files without any NLP approach, including data extraction from graphs;
- any research that focuses exclusively on protocol preparation, synthesis of already extracted data, write-up, pre-processing of text and dissemination will be excluded;
- methods or tools that provide no natural language processing approach and offer only organisational interfaces, document management, databases or version control; or

- any publications related to electronic health reports or mining genetic data will be excluded.

Key items of interest

Primary:

1. Machine learning approaches used
2. Reported performance metrics used for evaluation
3. Type of data
 - Scope: Abstract, conference proceeding, or full text
 - Target design: Randomised controlled trial, cohort, case-control
 - Type of input: The input data format, for example data imported as structured result of literature search (e.g. RIS), APIs, from PDF or text files.
 - Type of output: In which format are data exported after the extraction, for example as text file.

Secondary:

1. Granularity of data mining: Does the system extract specific entities, sentences, or larger parts of text?
2. Other reported metrics, such as impacts on systematic review processes (e.g. time saved during data extraction).

Assessment of the quality of reporting: We will extract information related to the quality of reporting and reproducibility of methods in text mining¹⁴. The domains of interest, adapted for our reviewing task, are listed in the following.

1. Reproducibility:
 - Are the sources for training/testing data reported?
 - If pre-processing techniques were applied to the data, are they described?
2. Transparency of methods:
 - Is there a description of the algorithms used?
 - Is there a description of the dataset used and of its characteristics?
 - Is there a description of the hardware used?
 - Is the source code available?
3. Testing:
 - Is there a justification/an explanation of the model assessment?
 - Are basic metrics reported (true/false positives and negatives)?
 - Does the assessment include any information about trade-offs between recall and precision (also known as sensitivity and positive predictive value)?
4. Availability of the final model or tool:
 - Can we obtain a runnable version of the software based on the information in the publication?
 - Persistence: is the dataset likely to be available for future use?

- Is the use of third-party frameworks reported and are they accessible?

5. Internal and external validity of the model:

- Does the dataset or assessment measure provide a possibility to compare to other tools in same domain?
- Are explanations for the influence of both visible and hidden variables in the dataset given?
- Is the process of avoiding over- or underfitting described?
- Is the process of splitting training from validation data described?
- Is the model's adaptability to different formats and/or environments beyond training and testing data described?

6. Other:

- Does the paper describe caveats for using the method?
- Are sources of funding described?
- Are conflicts of interest reported?

Dissemination of information

We plan to publish the finished review, along with future updates, via F1000Research.

All data will be available via a project on Open Science Framework (OSF): <https://osf.io/4sgfz/> (see *Data availability*).

Study status

Protocol published. We did a preliminary Medline search as described in this protocol and the supplementary material.

The final search, including all additional databases, will be conducted as part of the full review.

Data availability

Underlying data

No underlying data are associated with this article.

Extended data

Open Science Framework: Data Extraction Methods for Systematic Review (semi)Automation: A Living Review / Protocol. <https://doi.org/10.17605/OSF.IO/ECB3T>¹¹

This project contains the following extended data:

- Additional_Fields.docx (overview of data fields of interest for text mining in clinical trials)
- Search.docx (additional information about the searches, including full search strategies)

Reporting guidelines

Open Science Framework: Data Extraction Methods for Systematic Review (semi)Automation: A Living Review / Protocol. <https://doi.org/10.17605/OSF.IO/ECB3T>¹¹

Data are available under the terms of the [Creative Commons Attribution 4.0 International](#) (CC BY 4.0) data waiver.

Acknowledgements

We thank Sarah Dawson for developing and evaluating the search strategy, and providing advice on databases to search for this review. Many thanks also to Alexandra McAleenan and Vincent Cheng for providing valuable feedback on this protocol.

References

1. Borah R, Brown AW, Capers PL, *et al.*: **Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry.** *BMJ Open.* 2017; 7(2): e012545. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Tsafnat G, Glasziou P, Choong MK, *et al.*: **Systematic review automation technologies.** *Syst Rev.* 2014; 3(1): 74. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. O'Mara-Eves A, Thomas J, McNaught J, *et al.*: **Using text mining for study identification in systematic reviews: a systematic review of current approaches.** *Syst Rev.* 2015; 4: 5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Jonnalagadda SR, Goyal P, Huffman MD: **Automating data extraction in systematic reviews: a systematic review.** *Syst Rev.* 2015; 4: 78. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Marshall I, Kuiper J, Wallace B: **Robotreviewer on github.** 2020; Last accessed 14 Jan 2020. [Reference Source](#)
6. Marshall IJ, Wallace BC: **Toward systematic review automation: a practical guide to using machine learning tools in research synthesis.** *Syst Rev.* 2019; 8(1): 163. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Beller B, Clark J, Tsafnat G, *et al.*: **Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR).** *Syst Rev.* 2018; 7(1): 77. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Marshall C: **The systematic review toolbox.** 2019; Last accessed 11 Nov 2019. [Reference Source](#)
9. Elliott JH, Synnot A, Turner T, *et al.*: **Living systematic review: 1. Introduction-the why, what, when, and how.** *J Clin Epidemiol.* 2017; 91: 23–30. [PubMed Abstract](#) | [Publisher Full Text](#)
10. Brooker A, Synnot A, McDonald S, *et al.*: **Guidance for the production and publication of Cochrane living systematic reviews: Cochrane reviews in living mode.** 2019; Last accessed 06 Mar 2020. [Reference Source](#)
11. Schmidt L, McGuinness LA, Olorisade BK, *et al.*: **Protocol.** 2020. <http://www.doi.org/10.17605/OSF.IO/ECB3T>
12. Wallace BC, Small K, Brodley CE, *et al.*: **Deploying an interactive machine learning system in an evidence-based practice center: Abstrackr.** In: *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. IHI'12*, New York, NY USA, Association for Computing Machinery. 2012; 819–824. [Publisher Full Text](#)
13. Higgins J, Thomas J, Chandler J, *et al.*: **Cochrane Handbook for Systematic Reviews of Interventions.** John Wiley & Sons, Chichester (UK), 2nd edition, 2019. [Reference Source](#)
14. Olorisade BK, Brereton P, Andras P: **Reproducibility of studies on text mining for citation screening in systematic reviews: Evaluation and checklist.** *J Biomed Inform.* 2017; 73: 1–13. [PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 06 July 2020

<https://doi.org/10.5256/f1000research.27003.r64379>

© 2020 Carter M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Matt Carter 

Bond University Centre for Research in Evidence-Based Practice, Bond University, Robina, Australia

No further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Automation of Systematic Reviews

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 18 May 2020

<https://doi.org/10.5256/f1000research.25153.r62963>

© 2020 Carter M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Matt Carter 

Bond University Centre for Research in Evidence-Based Practice, Bond University, Robina, Australia

I believe that the research proposal clearly lays out its objectives and aims. Although there are some minor edits as per the below list:

1. The abstract should specifically mention PICO data extraction rather than data extraction generally. The "Aims" section outlines this but seems to contradict with the more general "Introduction / Objective" section.
2. While the Tsafnat *et al.* workflow is considered the first, other papers such as Clark *et al.* (2019) or Macleod *et al.* for animal testing are a little more current and also fix some of the gaps with Tsafnat. Since there are a few in this area now a reason should be given for the preference.
3. "We plan to run searches for new studies every second month" - There is currently an effort to get this adjusted to bi-monthly as you suggest but the guidelines specify that this should be monthly. Perhaps just a quick sentence acknowledging this and saying that it is excessive for the subject area?
4. The generated search strategy has a specific filter for English / German papers which seems to contradict the more general phrasing of "We will translate non-English reports where feasible".
5. Data sets are obviously not provided as this is an ongoing paper.

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Automation of Systematic Reviews

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 May 2020

Lena Schmidt, University of Bristol, Bristol, UK

"I believe that the research proposal clearly lays out its objectives and aims. Although there are some minor edits as per the below list:

1. *The abstract should specifically mention PICO data extraction rather than data extraction generally. The "Aims" section outlines this but seems to contradict with the more general "Introduction / Objective" section."*

Thank you for pointing this out, we have made some changes to the abstract and to the aims. Firstly, in the abstract we clarified that the data extraction goal is wider (not limited to PICO, it included fields that are generally of interest in systematic reviews as defined in the Cochrane

Handbook). We have made changes in the aims section to reflect that, and to make it more consistent throughout. In the method section, we deleted the section “Objectives” and summarised everything under “Aims of this review” in the introduction.

1. *“While the Tsafnat et al. workflow is considered the first, other papers such as Clark et al. (2019) or Macleod et al. for animal testing are a little more current and also fix some of the gaps with Tsafnat. Since there are a few in this area now a reason should be given for the preference.”*

In response to this comment we added a more recent paper review paper to our summary of related research (Marshall and Wallace, 2019). The related research was chosen because it was either in the form of a systematic review or it was an overview that is directly related to our topic of interest. We also added an additional explanation for our preference to focus on the data extraction stage of the systematic review process in response to this, and also a previous peer review comment. To summarise this quickly, data extraction is one of the most time-consuming and error-prone tasks in the systematic reviewing process. By reviewing automation of data extraction, we aim to summarise the current knowledge. Furthermore, the area of data extraction has future potential to disrupt the “traditional” systematic review process – if data are extracted and well classified centrally then the searching and screening workflow can change as well.

1. *““We plan to run searches for new studies every second month” - There is currently an effort to get this adjusted to bi-monthly as you suggest but the guidelines specify that this should be monthly. Perhaps just a quick sentence acknowledging this and saying that it is excessive for the subject area?”*

Thank you, yes. We have added a statement: “The bi-monthly interval for screening was chosen because we expect no sudden rise in relevant publications that could justify daily, weekly or monthly screening”. Furthermore, we added a statement defining impactful research that would lead to a review update even if the threshold of new studies is not met (please see reply to the first peer review for reference)

1. *“The generated search strategy has a specific filter for English / German papers which seems to contradict the more general phrasing of “We will translate non-English reports where feasible”.*

Thank you, this item was unclear. This initial draft of the search strategy was created for protocol publication and will be minimally altered for the full search and when the remaining databases are searched. Then it will be impossible to use a language filter on most databases. The initial Medline search specifically included German studies because they are feasible to assess, and for any other database search we have no language restrictions. The process of searching all databases will be described in detail in the full review.

1. *“Data sets are obviously not provided as this is an ongoing paper.”*

Thank you for your peer review and for helping us to improve the quality of our paper.

Competing Interests: No competing interests.

Reviewer Report 01 May 2020

<https://doi.org/10.5256/f1000research.25153.r61722>

© 2020 McFarlane E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Emma McFarlane 

National Institute for Health and Care Excellence, Manchester, UK

This protocol outlines a project to review methods and tools for data extraction to help automate a step in the systematic reviewing process. This will be done in the context of a living systematic review with the aim of providing guidance to reviewers who may want to semi-automate their work.

The objectives of the study are described clearly and are set within the current context of increasing publications. However, it would be helpful, as part of the aim and purpose of the study, to note why the focus is around the data extraction stage specifically.

The methods of the study are described in enough detail to be replicated. In terms of the methods, can the authors double check the searching approach for accuracy? The abstract notes searches will be conducted monthly whereas the body of the protocol states every two months. The authors indicated they will update their systematic review when evidence expected to impact is identified however, it would be helpful to include some detail to note how impactful evidence will be defined for people doing similar work.

A comment from the authors about their choice of study design for the included papers would be helpful. Identifying RCTs in the context of data science is likely to be challenging so it would be interesting to understand the expectations about the evidence base and how study design could link to the point about impact on the results of the review and need to update.

The outcomes listed in the protocol appear to be comprehensive. However, it is not clear if consideration was given to accuracy of the tools identified as that could link to the objective of identifying whether automation is reliable.

Overall, this appears to be an interesting study straddling the fields of systematic reviewing and data science.

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I am currently conducting a systematic review of automation in systematic reviewing or guideline development. I am also working on a research project on machine learning within the context of guideline recommendations.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 May 2020

Lena Schmidt, University of Bristol, Bristol, UK

Thank you for providing this very helpful peer review. We tried to address the concerns below:

“This protocol outlines a project to review methods and tools for data extraction to help automate a step in the systematic reviewing process. This will be done in the context of a living systematic review with the aim of providing guidance to reviewers who may want to semi-automate their work.”

“The objectives of the study are described clearly and are set within the current context of increasing publications. However, it would be helpful, as part of the aim and purpose of the study, to note why the focus is around the data extraction stage specifically.”

Thank you for pointing this out. In the current revision we added more details about why we chose to focus on data extraction. To summarise this quickly, data extraction is one of the most time-consuming and error-prone tasks in the systematic reviewing process. By reviewing automation of data extraction, we aim to summarise the current knowledge. Furthermore, the area of data extraction has future potential to disrupt the “traditional” systematic review process – if data are extracted and well classified centrally then the searching and screening workflow can change as well.

“The methods of the study are described in enough detail to be replicated. In terms of the methods, can the authors double check the searching approach for accuracy? The abstract notes searches will be conducted monthly whereas the body of the protocol states every two months. “

We addressed this, thank you. New articles will be screened every two months.

“The authors indicated they will update their systematic review when evidence expected to impact is identified however, it would be helpful to include some detail to note how impactful evidence will be defined for people doing similar work.”

Thank you, this was not clear previously, and we added some further explanation: “We define impactful evidence as, for example, the publication of a tool that is immediately accessible to systematic reviewers and offers substantial automation of the data extraction process, or a tool that aims to change the traditional SR workflow”

“A comment from the authors about their choice of study design for the included papers would be helpful. Identifying RCTs in the context of data science is likely to be challenging so it would be interesting to understand the expectations about the evidence base and how study design could link to the point about impact on the results of the review and need to update.”

We clarified that we are open to include any paper, as long as the paper reports an automation technology for systematic reviewing that processes texts from clinical studies (not electronic health reports). To clarify further, the texts that the technologies process are likely to be RCT reports, but we are not looking to identify RCTs that compare data extraction tools.

“The outcomes listed in the protocol appear to be comprehensive. However, it is not clear if consideration was given to accuracy of the tools identified as that could link to the objective of identifying whether automation is reliable.”

Thank you for this comment. In response we clarified that we will extract the reported performance metrics. We do not plan to rank tools based on these metrics because results can vary throughout different datasets. Instead, we aim to assess quality of reporting (reproducibility, transparency...) in detail.

“Overall, this appears to be an interesting study straddling the fields of systematic reviewing and data science.”

Thank you very much for your review, it was very helpful in improving the protocol.

Competing Interests: None

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research