The future of evidence-based medicine could be brighter: A meta-analysis of the time and workers needed to conduct systematic reviews of medical interventions

<table>
<thead>
<tr>
<th>Journal:</th>
<th>BMJ Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>bmjopen-2016-012545</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Research</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>05-May-2016</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Borah, Rohit; University of Alabama at Birmingham, Dean\'s Office, Office of Energetics  
Brown, Andrew; University of Alabama at Birmingham, Dean\'s Office, Office of Energetics  
Capers, Patrice; University of Alabama at Birmingham, Dean\'s Office, Office of Energetics  
Kaiser, Kathryn; University of Alabama at Birmingham, Dean\'s Office, Office of Energetics |
| Primary Subject Heading: | Health informatics |
| Secondary Subject Heading: | Evidence based practice |
| Keywords: | systematic reviews, metadata, PROSPERO registry, search methods |
The future of evidence-based medicine could be brighter: A meta-analysis of the time and workers needed to conduct systematic reviews of medical interventions

Rohit Borah¹, Andrew W. Brown²,³, Ph.D., Patrice L. Capers²,³, Ph.D., & Kathryn A. Kaiser²,³, Ph.D.

1. Science and Technology Honors Program, University of Alabama at Birmingham, USA
2. Office of Energetics, Dean’s Office, School of Public Health, University of Alabama at Birmingham, USA
3. Nutrition Obesity Research Center, University of Alabama at Birmingham, USA

Keywords: systematic review, registry, meta-analysis, qualitative review, metadata

Funding: Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Disease and the National Institute of General Medical Sciences of the National Institutes of Health under award numbers: P30DK056336 and K12GM088010. The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Alabama at Birmingham or the National Institutes of Health.

Corresponding Author: Kathryn A. Kaiser, Ph.D. - Office of Energetics, Dean’s Office, School of Public Health, 1720 2nd Ave. South, RPHB 140J, University of Alabama at Birmingham, Birmingham, AL 35294 USA, Email: kakaiser@uab.edu, Tel: +1-205-975-9212

Word count: 2965 (text)
Abstract -

Objectives: To summarize logistical aspects of recently completed systematic reviews that were registered in the PROSPERO Systematic Review Registry to quantify the time and resources required to complete such projects.

Design: Systematic review

Data Sources and Study selection: All of the 195 registered and completed reviews (status from the PROSPERO registry) with associated publications at the time of our search (1 July 2014).

Data extraction: All authors extracted data using registry entries and publication information related to the data sources used, the number of initially retrieved citations, the final number of included studies, the time between registration date to publication date, and number of authors involved for completion of each published. Information related to funding and geographical location was also recorded, when reported.

Results: The number of studies found in the literature searches ranged from 27 to 92,020; the mean yield rate of included studies was 2.96%; and the mean number of authors per review was 5, SD = 3. Average estimated time to complete the project and publish the review was 66.8 weeks. Funded reviews took significantly longer to complete and publish and involved more authors and team members than those that did not report funding, p<0.001.

Conclusions: Systematic reviews presently take much time and require large amounts of human resources. Appropriate application of existing computing and informatics technology could decrease this time and resource burden, and recently published guidelines provide a framework.
A vision of a future in which synthesizing a group of intervention studies may be accomplished with little to no human intervention is presented.

**Systematic review registration:** Not registered; no known registry currently accepts methodological systematic reviews.

What is already known on this subject:

- The number of published clinical trials is increasing exponentially
- Reviewing and synthesizing the ever-increasing literature on a given intervention is an important research activity to inform treatment guidelines and public policy

What this study adds:

- This study provides quantitative estimates on the time and human effort required to conduct a systematic review and publish it, which can be useful for grant applications and project planning.
- Recent data standards proposals and informatics technology can make the process of finding and synthesizing literature much more efficient if small additions can be added to the publication process.
Strengths and limitations of this study:

- This study provides an updated estimate of the time and effort required to conduct and publish a systematic review using a large sample of recently published reviews across a variety of topics of medical interventions.
- The study is limited by incomplete reporting in both sources of the data – the review registry and the published articles. These data sources were cross checked as data was available and conservative assumptions were stated and applied where necessary.
Introduction

The systematic review can be an effective and scientific method used across disciplines to consolidate vast amounts of research on a specific topic. Over the last 20 years, publishing of systematic reviews has increased exponentially, as has the primary literature. As the body of primary literature has increased, conducting systematic reviews and meta-analyses has become a necessity to more accurately and comprehensively present accumulated knowledge to scientific, clinical, and general audiences. In fact, new mandates are being formulated to require documented systematic reviews as part of research proposals\(^1\) and clinical trial data sharing,\(^2\) both of which have strong implications for how scientific literature and associated data are reported, managed, and curated.\(^3\)

The Cochrane Database of Systematic Reviews\(^4\) provides important guidelines and methods for systematic reviews through its *Cochrane Handbook*,\(^5\) which states four key considerations to define prior to beginning: the question, the inclusion and exclusion criteria, the search strategy, and the methods. In other words, systematic reviews should review the scientific literature in a scientific manner. The Cochrane Collaboration has built its reputation as the leading resource for systematic reviews by requiring that all Cochrane reviews be “…updated regularly in an effort to ensure that the most recent evidence is incorporated.”\(^6\) However, this onerous requirement may have the side effect of limiting the topics that are created and maintained up-to-date.

With ever-growing numbers of journals and publications, an immense amount of time and effort is needed to search the literature and summarize the findings. This is reflected in the Cochrane Collaboration statement that “[h]istorically, the aim was to update Cochrane reviews every two years, but recently there has been a move away from this policy in favour of prioritising the most clinically important reviews for updating.”\(^6\) When beginning a new systematic review, authors
are faced with the possibility of finding few to no studies that meet their criteria. While finding no studies that meet the criteria can be informative *per se*, such as by identifying directions for future research, the time and effort required to reach this conclusion may be great. Conversely, the scope of some reviews can be unpredictably large, and it may be difficult to plan the person-hours required to complete the research. The magnitude of this uncertainty has not been defined to date.

Recent efforts, such as the International Prospective Register of Systematic Reviews (PROSPERO)\(^7\) hosted through the University of York’s Centre for Reviews and Dissemination, have been deployed to prospectively and centrally register systematic reviews and meta-analyses. The PROSPERO registry “…aims to provide a comprehensive listing of systematic reviews registered at inception to help avoid unplanned duplication and enable comparison of reported review methods with what was planned in the protocol.”\(^7\) We found that such a database provides rich data for summarizing various logistical aspects of a sample of recently registered and completed systematic reviews. In the present meta-analysis, we aimed to use the PROSPERO registry to quantify the time and manpower needed to conduct a systematic review.

Our specific research questions were as follows:

1. How many people were involved in conducting or authoring the review?
2. How much time was required to complete and publish the review?
3. What was the average efficiency of the search strategy as indicated by the ratio of studies ultimately included in the review to the number of studies found in the database searches (i.e., yield rate)?
4. Did the number of people and time needed to complete a review differ between funded and unfunded reviews (regardless of funding source)?
Methods:

Study Selection

We searched the PROSPERO database hosted at the University of York’s Centre for Reviews and Dissemination. Per the website, “PROSPERO is an international database of prospectively registered systematic reviews in health and social care,” with an emphasis on intervention studies. We retrieved 3684 registered project records from the PROSPERO website on July 1, 2014, and 437 records were marked as having a completed status. Of the 437 records for completed projects, 195 records contained a link to one or more publications of a completed review (not simply a protocol). Data were extracted from 195 publications (reporting a total of 197 literature reviews) by two authors (RB, KAK) and were verified by at least one of two additional authors (AWB, PLC). Although an audit trail is available for each PROSPERO record, there is no way to verify the completeness or accuracy of the entries or whether the registrants created the files early in the timeline of the project as prescribed by the registry administrators. Additionally, this registry is not equipped to verify the completeness or accuracy of the entries as compared to the resultant publication. See the text box for clarification of the terms we used in our analysis.

<table>
<thead>
<tr>
<th>Term Used</th>
<th>Data Used in Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>The persons listed on the published article reporting the results of the review. Not all team members were listed as authors of resulting publications, so we counted and evaluated “team members” separately. A sum of unique names was generated to create a combined variable for</td>
</tr>
<tr>
<td>Team Members</td>
<td>The number of persons working on the review per the PROSPERO registry (see above for distinctions with “authors”).</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Reviews</td>
<td>The activities of searching the literature, selecting studies that meet inclusion criteria, and synthesis of study results. In our data, two governmental publications contained two review processes, thus resulting in our evaluation of 197 reviews reported in 195 publications.</td>
</tr>
<tr>
<td>Publications</td>
<td>The published results of the review process(es).</td>
</tr>
<tr>
<td>Studies</td>
<td>We counted studies and citations in PRISMA diagrams as the same thing, although some levels of PRISMA diagrams may contain some duplicate reports of studies in different citations.</td>
</tr>
</tbody>
</table>

**Data Extraction Process**

We extracted the following data from the published reviews: the dates a manuscript was received, accepted, and published (received and accepted dates were used only to compare to the registry date to validate the registered timeline); the number of authors; funding information; the data sources searched (total number of databases or websites searched and their names); and the number of studies in the literature filtering steps from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagrams\(^8\) (when present in the articles).
Companion data from the PROSPERO registry for each record included the number of team members, funding, registered start date, and geographic location of the team members.

Authors/Team Members - The authors listed on the publication were counted, as were the team members in the registry. In all but seven cases, the number of authors was larger than the registered number of team members. We added the number of unique authors and registered team members in a combined variable to reflect personal involvement. We also extracted country and institutional affiliations.

Time – We estimated the duration of the project as the time from the registered project start date to the publication date of the review. When the article text did not include complete dates, we sought additional information on the publishers’ websites. For articles indexed in PubMed (all but 12), we used the PubMed IDs to extract the dates provided by the publishers to PubMed. These dates were related to the publication timeline (e.g., accepted date, first available online, final publication available date) when reported.

Search sources - For the initial total number of studies found in the database searches (highest level of the PRISMA diagram), we recorded what was reported in the PRISMA diagram (or text if the article or supplemental material did not contain a PRISMA diagram). When authors reported searching less commonly used sources of information such as local clinical trial registries or country-specific websites, we counted those sources in the “other” category and did not include them in the “total databases searched” variable. Authors do not consistently indicate the same level of the PRISMA diagram for the PRISMA category “studies identified from other sources,” so we aggregated these sources in the variable we called “total N found” to compare across studies.
Search efficiency - We defined yield rate as a variable obtained by dividing the final number of included studies by the initial number of studies found (excluding duplicates) in the literature search. The final number of included studies was recorded from the PRISMA diagram or the text. If the PRISMA diagram or text separately reported the total number of studies included for quantitative (used in a statistical synthesis) and qualitative (used in a narrative summary only) synthesis, we used the highest number to calculate the overall merged yield rate. For example, if the PRISMA diagram noted that 20 unique studies were included in the review, and 8 were in the quantitative synthesis and 15 were in the qualitative summary, we used 20 for the overall merged yield rate owing to some overlap.

Funding - We coded reviews as being funded if the registry or publication text included explicit statements of review-level funding. We did not code studies as being “funded” if there were only general statements of salary support or conflict disclosures for individual authors (e.g., “Dr. Smith is funded by an NIH grant”). Explicit statements of author salary support for the review were recorded separately for analysis.

Statistical Analysis

To answer our research questions, we calculated the average number of authors/team members, time from registered start date to publication in weeks, and study yield rates. We counted the number of reviews that reported funding for the project in general or review-related salary support of authors. We summarized frequency counts for each of the 12 most often reported databases used in literature searches in our sample. All summaries and analyses were calculated with SPSS version 22 (IBM, New York, USA) except where noted. Analysis of variance was used to compare means for time to complete and number of authors/team members between funded and unfunded reviews. The summary literature distillation process was generated with R.
(Figure 1). Because of the extreme skewness for the study count variables from the PRISMA diagrams, we calculated z-scores and generated means, standard deviations, and ranges based on those publications with complete data that were between -2.5 and +2.5 standard deviations in order to generate Figure 1. For reviews with incomplete data for any variable, we opted to summarize only reported data and did not write authors to request missing information.

**Results:**

Table 1 summarizes the results for number of authors/team members, time needed to complete the reviews, and yield rate. The mean project length (using the registered project start date to the review’s publication date) was 67.3 weeks (SD = 31.0; range, 6-186 weeks). We also calculated a merged yield rate from the number found without duplicates in the initial search to final included studies (with some small overlap between quantitative and qualitative studies, n=190). The mean merged yield rate was 2.94% (SD = 6.49; range, 0-64.71%). Of the countries listed for the locations of registered team members, the United Kingdom was most often listed: 62 of the 197 reviews included UK team members.

Figure 1 summarizes the literature filtering process across reviews, indicating the initial number of studies found and the interim steps until the final mean number of studies included in the sample we analyzed. The use of the “Büchner” plot (shaped like a Büchner funnel and not to be confused with a “funnel plot” used to evaluate potential publication bias) allows us to emphasize several key characteristics of the process. First, the Büchner plot allows for all data to be plotted in such a way that demonstrates the wide range of included studies at each stage as well as the distribution of the data (highly right-skewed). Second, the ordinality of included studies was not maintained from stage to stage. That is, the reviews with the greatest numbers at one stage were not necessarily the greatest at the next stage. This is visualized by lines crossing throughout the
plot. Third, the figure demonstrates that the filtration process can be dramatic, as is reflected in
the average yield rate being less than 3%.

Table 2 summarizes the differences in time to complete the review and number of authors/team
members per project stratified by reported review-level or salary funding. Reviews that reported
funding took longer to complete and included more authors/team members; this difference was
not seen when we considered only whether specific authors were funded. Table 3 summarizes
the most commonly used databases reported in the included reviews.

Discussion

Our aim was not to conduct a comprehensive review of all systematic reviews, but rather to
generate plausible estimates of the logistics of conducting reviews for medical interventions for
review teams that engaged in at least one best practice: prospective review registration. We
therefore chose to evaluate reviews that were registered in the PROSPERO registry and reported
to be completed and published. Aspects of performing systematic reviews that remain
unpublished, of performing systematic reviews in different domains, or of performing systematic
reviews by teams that neither register their projects nor update their registrations upon
completion may differ. Furthermore, the country affiliation reported in the registry for almost
one-third of the team members was the United Kingdom. We cannot determine whether our
results are representative of all systematic review teams or whether our results reflect an
experience more likely to be encountered by UK researchers.

An estimate of the review project start date was difficult to capture from the registry data. Our
calculations for time were anchored by the registered date of the start of the project, but it takes
some time to assemble the team, determine inclusion and exclusion criteria, conduct preliminary
searches to refine the search syntax, perform inter-rater reliability for literature screening, and
obtain funding (if the review is specifically funded). When a large number of studies are found in
a search, these early-stage tasks need to be refined to a standard operating procedure, especially
when the evaluation of the initial corpus of found studies cannot feasibly be performed in
duplicate. We therefore predict that the time-to-publication of some reviews may be substantially
longer than our estimates. As shown in Figure 1, at present, the full text of a large number of
papers (M = 63, upper point of range = 4385) may need to be screened to evaluate final inclusion
criteria.

As the scientific literature continues to grow, generating high-quality, comprehensive reviews in
a timely manner will become increasingly costly unless more automated resources are dedicated
or different methods to search and retrieve relevant papers can be developed.3 “A systematic
review attempts to collate all empirical evidence that fits pre-specified eligibility criteria to
answer a specific research question. It uses explicit, systematic methods that are selected with a
view to minimizing bias, thus providing reliable findings from which conclusions can be drawn
and decisions made.”10 Our results point to an ever-increasing logistical challenge to conducting
systematic reviews so that unbiased conclusions can be made about medical interventions. A
recent example that points to the emerging logistical challenges is illustrated in a study that
evaluated matched pairs of reviews published on the same research question within 5 years of
each other: one using Cochrane review methods and the other from other sources. These authors
found a 47% difference in the number of studies in the paired reviews (excluding individual
 trials included in both reviews) unaccounted for by order of publication.11 One potential source
of search differences may be the use of Medical Subject Headings (MeSH), which were
introduced in 1963.12 Although this system is still able to provide a list of articles on a topic of
interest, it is not designed to meet the needs of today’s clinicians or systematic reviewers. This
approach to indexing and searching the literature, along with author-assigned key words, is no
longer a serviceable approach because of the vast amount of irrelevant or missing articles
provided when using such methods, owing to imprecise and overlapping key words and MeSH
terms that provide little context. Clinicians have advised other clinicians who wish to search for
studies to inform their clinical practice in PubMed: “Warning: do not look at all the articles
found that although interesting are not pertinent to the present clinical question, or you will be
lost in the sea of PubMed!” A search for human clinical trials using the filters provided in
PubMed with or without additional index terms provides results with unacceptably low
sensitivity, precision, and specificity. Even something as simple as defining a study as a
randomized controlled trial has been observed to be incorrect 20% of the time.
Systematic reviews and meta-analyses can require large amounts of time and effort to complete,
with often (based on our own experiences) unpredictable uncertainty in the time and resources
required to complete a review. Although statistical methods have been developed to estimate
trends in large corpora of literature, and crowdsourcing may help to decrease the calendar time
needed to complete a review, we assert that the time is ripe to investigate metadata approaches to indexing publications so that more targeted yet comprehensive searches can be
performed efficiently with high specificity and precision. We suggest that human clinical trials
could be a starting point for testing the use of indexing systems that employ a Population,
Intervention, Comparisons, Outcomes, Study Design (PICOS) structure of coding metadata for
clinical trials in order to make the scientific literature become a truly searchable database. Our
preliminary testing has shown that a modestly trained worker can code a single paper in about 15
minutes. By application of the recently proposed FAIR (Findable, Accessible, Interoperable and
Reusable) metadata guidelines, clinical trial data (or other types of research data) could be
encoded upon publication so that the data are in fact findable, accessible, interoperable, and reusable. An added benefit of this approach is that it would make the recently proposed mandated data sharing\textsuperscript{2} essentially automatic. In a brighter future, evidence-based, up-to-date summaries could be produced on demand with little human effort.

**Acknowledgment:** The authors are grateful to Alison Booth and Jimmie Christy of the PROSPERO registry and to Madeline Jeansonne, MPH, and Eric Kim at the University of Alabama at Birmingham for assistance in data collection. Jennifer Holmes, ELS, performed the language editing of the manuscript.

**Competing Interests Statement:** All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that all authors have no financial or non-financial interests that may be relevant to the submitted work.

**Funding Statement:** Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Disease and the National Institute of General Medical Sciences of the National Institutes of Health under award numbers: P30DK056336 and K12GM088010 in the form of partial salary support for KAK, PLC and AWB. The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Alabama at Birmingham or the National Institutes of Health.

**Data Sharing:** The list of included review publications is in the online appendix. Analysis data may be obtained from the corresponding author at kakaiser@uab.edu.

**Authors’ contributions:** All authors contributed equally. KAK conceived the project. AWB and KAK retrieved and processed the data from the PROSPERO registry. RB retrieved articles and
organized the reference bibliography. All authors extracted data from included papers, analyzed
the data and contributed to the writing of the manuscript. All authors have full access to all the
data in the study and had final responsibility for the decision to submit for publication.

**Transparency Declaration**: Kathryn Kaiser affirms that the manuscript is an honest, accurate,
and transparent account of the study being reported; that no important aspects of the study have
been omitted; and that any discrepancies from the study as planned have been explained.

**Ethics Committee Approval**: No ethics committee approval is required for analyses of non-
human subjects research. This paper used data about publications from public sources.
References


7. PROSPERO - International Prospective Register for Systematic Reviews. (University of York, Centre for Reviews and Dissemination, United Kingdom, 2015). Available at: http://www.crd.york.ac.uk/prospero/ (accessed 1/4/2016).


Figure 1. Aggregated literature filtration process based on counts reported (n=195). Trimmed means are indicated in the boxes and trimmed ranges (+/- 2.5 standard deviations) are indicated at the right and left of each level. Some reviews were published reporting that 0 studies met the criteria for inclusion in the review.
Table 1. Descriptive statistics for number of authors, time for publication, and quantitative or qualitative yield rates for 195 records analyzed in the PROSPERO registry.

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean, SD</th>
<th>Median</th>
<th>Range</th>
<th>Interquartile Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors/Team Members (n = 195 publications)</td>
<td>5, 3</td>
<td>5</td>
<td>3</td>
<td>1 – 27</td>
</tr>
<tr>
<td>Time (in weeks; registered project start to publication, n = 192*)</td>
<td>67.3, 31.0</td>
<td>65.8</td>
<td>41.6</td>
<td>6 – 186</td>
</tr>
<tr>
<td>Quantitative Analysis Yield (n = 82), %</td>
<td>2.6, 4.7</td>
<td>1.0</td>
<td>2.7</td>
<td>0.03 – 32.43</td>
</tr>
<tr>
<td>Qualitative Analysis Yield (n = 80), %</td>
<td>2.7, 4.6</td>
<td>1.0</td>
<td>2.5</td>
<td>0.05 – 26.19</td>
</tr>
<tr>
<td>Merged Yield Rate (n=190), %</td>
<td>2.94, 6.49</td>
<td>0.93</td>
<td>2.5</td>
<td>0.0 – 64.71</td>
</tr>
</tbody>
</table>

* Three studies were excluded from this calculation because they were registered after the publication date.

+ Excludes small overlap between quantitative and qualitative studies when information was provided in the publication to differentiate the categories.
Table 2. Comparison (analysis of variance) of reported funding of author salaries (n=20 for both outcomes) or review projects (n=86 for time, n=88 for authors) to time and authors needed to complete and publish the reviews.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Did outcome differ from reviews reporting no funding?</th>
<th>Means (reported as funded vs. funding not reported)</th>
<th>F, p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to complete (n=191)</td>
<td>Review funding reported YES 42 weeks vs. 26 weeks</td>
<td>17.545, &lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Salary funding reported NO 68 weeks vs. 64 weeks</td>
<td>0.258, 0.612</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Authors/Team Members (n=195)</td>
<td>Review funding reported YES 6.8 persons vs. 4.8 persons</td>
<td>14.638, &lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Salary funding reported NO 6.0 versus 5.7 persons</td>
<td>0.08, 0.778</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 20)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Top 12 databases used in included reviews in descending order (N = 197 reviews).

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Frequency, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>162, 82.2%</td>
</tr>
<tr>
<td>Embase</td>
<td>160, 81.2%</td>
</tr>
<tr>
<td>Cochrane</td>
<td>148, 75.1%</td>
</tr>
<tr>
<td>CINAHL</td>
<td>87, 44.2%</td>
</tr>
<tr>
<td>PubMed</td>
<td>58, 29.4%</td>
</tr>
<tr>
<td>Web of Science</td>
<td>56, 28.4%</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>49, 24.9%</td>
</tr>
<tr>
<td>SCOPUS</td>
<td>29, 14.7%</td>
</tr>
<tr>
<td>AMED</td>
<td>30, 15.2%</td>
</tr>
<tr>
<td>LILACS</td>
<td>24, 12.2%</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>13, 6.6%</td>
</tr>
<tr>
<td>ProQuest</td>
<td>12, 6.1%</td>
</tr>
</tbody>
</table>

CINAHL – Cumulative Index to Nursing and Allied Health Literature

AMED – Allied and Complementary Medicine Database

LILACS – Literatura Latina-Americana e do Caribe em Ciências da Saúde
The future of evidence-based medicine could be brighter: A meta-analysis of the time and workers involved in conducting systematic reviews of medical interventions


98. Munn Z, Jordan Z. The effectiveness of interventions to reduce anxiety, claustrophobia, sedation and non-completion rates of patients undergoing high technology medical imaging. *The JBI Database of Systematic Reviews and Implementation Reports* 2012; *10*(19): 64.

99. Murray SF, Hunter BM, Bisht R, Ensor T, Bick D. Demand-side financing measures to increase maternal health service utilisation and improve health outcomes: a systematic review


132. Schneeberger AR, Dietl MF, Muenzenmaier KH, Huber CG, Lang UE. Stressful childhood experiences and health outcomes in sexual minority populations: a systematic review. *Social psychiatry and psychiatric epidemiology* 2014; 49(9): 1427-45.


159. Suchdev PS, Peña-Rosas JP, De-Regil LM. Multiple micronutrient powders for home (point-of-use) fortification of foods in pregnant women. *Cochrane Database of Systematic Reviews* 2014; (6).


## An analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry

<table>
<thead>
<tr>
<th><strong>Journal</strong></th>
<th><em>BMJ Open</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manuscript ID</strong></td>
<td>bmjopen-2016-012545.R1</td>
</tr>
<tr>
<td><strong>Article Type</strong></td>
<td>Research</td>
</tr>
<tr>
<td><strong>Date Submitted by the Author</strong></td>
<td>19-Sep-2016</td>
</tr>
</tbody>
</table>
| **Complete List of Authors** | Borah, Rohit; University of Alabama at Birmingham, Dean's Office, Office of Energetics  
Brown, Andrew; University of Alabama at Birmingham, Dean's Office, Office of Energetics  
Capers, Patrice; University of Alabama at Birmingham, Dean's Office, Office of Energetics  
Kaiser, Kathryn; University of Alabama at Birmingham, Dean's Office, Office of Energetics |
| **Primary Subject Heading** | Health informatics                                                     |
| **Secondary Subject Heading** | Evidence based practice                                                  |
| **Keywords**         | systematic reviews, metadata, PROSPERO registry, search methods          |
An analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry

Rohit Borah¹,², Andrew W. Brown²,³, Ph.D., Patrice L. Capers²,³, Ph.D., & Kathryn A. Kaiser²,³, Ph.D.

1. Science and Technology Honors Program, University of Alabama at Birmingham, USA
2. Office of Energetics, Dean’s Office, School of Public Health, University of Alabama at Birmingham, USA
3. Nutrition Obesity Research Center, University of Alabama at Birmingham, USA

Keywords: systematic review, registry, meta-analysis, qualitative review, metadata

Funding: Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Disease and the National Institute of General Medical Sciences of the National Institutes of Health under award numbers: P30DK056336 and K12GM088010. The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Alabama at Birmingham or the National Institutes of Health.

Corresponding Author: Kathryn A. Kaiser, Ph.D. - Office of Energetics, Dean’s Office, School of Public Health, 1720 2nd Ave. South, RPHB 140J, University of Alabama at Birmingham, Birmingham, AL 35294 USA, Email: kakaiser@uab.edu, Tel: +1-205-975-9212

Word count: 3707(text)
Abstract -

Objectives: To summarize logistical aspects of recently completed systematic reviews that were registered in the PROSPERO Systematic Review Registry to quantify the time and resources required to complete such projects.

Design: Meta-analysis

Data Sources and Study selection: All of the 195 registered and completed reviews (status from the PROSPERO registry) with associated publications at the time of our search (1 July 2014).

Data extraction: All authors extracted data using registry entries and publication information related to the data sources used, the number of initially retrieved citations, the final number of included studies, the time between registration date to publication date, and number of authors involved for completion of each publication. Information related to funding and geographical location was also recorded, when reported.

Results: The mean estimated time to complete the project and publish the review was 67.3 weeks (interquartile range = 42). The number of studies found in the literature searches ranged from 27 to 92,020; the mean yield rate of included studies was 2.94% (interquartile range = 2.5); and the mean number of authors per review was 5, SD = 3. Funded reviews took significantly longer to complete and publish (mean= 42 versus 26 weeks) and involved more authors and team members (mean= 6.8 versus 4.8 people) than those that did not report funding (both p<0.001).

Conclusions: Systematic reviews presently take much time and require large amounts of human resources. In light of the ever increasing volume of published studies, application of existing computing and informatics technology should be applied to decrease this time and resource
burden. We discuss recently published guidelines that provide a framework to make finding and accessing relevant literature less burdensome.

**Systematic review registration**: Not registered; no known registry currently accepts methodological systematic reviews.

**Strengths and limitations of this study:**

- This study provides an updated estimate of the time and effort required to conduct and publish a systematic review using a large sample of recently published reviews across a variety of topics of medical interventions.
- The study is limited by incomplete reporting in both sources of the data – the review registry and the published articles. These data sources were cross checked as data were available and conservative assumptions were stated and applied where necessary.
**Introduction**

The systematic review can be an effective and scientific method used across disciplines to consolidate vast amounts of research on a specific topic. Over the last 20 years, publishing of systematic reviews has increased exponentially, as has the primary literature. As the body of primary literature has increased, conducting systematic reviews and meta-analyses has become a necessity to more accurately and comprehensively present accumulated knowledge to scientific, clinical, and general audiences. In fact, new mandates are being formulated to require documented systematic reviews as part of research proposals and clinical trial data sharing, both of which have strong implications for how scientific literature and associated data are reported, managed, and curated.

The Cochrane Database of Systematic Reviews provides important guidelines and methods for systematic reviews through its *Cochrane Handbook*, which states four key considerations to define prior to beginning: the question, the inclusion and exclusion criteria, the search strategy, and the methods. In other words, systematic reviews should review the scientific literature in a scientific manner. The Cochrane Collaboration has built its reputation as the leading resource for systematic reviews by requiring that all Cochrane reviews be “…updated regularly in an effort to ensure that the most recent evidence is incorporated.” However, this onerous requirement may have the side effect of limiting the topics that are created and kept current.

With ever-growing numbers of journals and publications, an immense amount of time and effort is needed to search the literature and summarize the findings. This is reflected in the Cochrane Collaboration statement that “[h]istorically, the aim was to update Cochrane reviews every two years, but recently there has been a move away from this policy in favour of prioritising the most clinically important reviews for updating.” When beginning a new systematic review, authors
are faced with the possibility of finding few to no studies that meet their criteria. While finding no studies that meet the criteria can be informative *per se*, such as by identifying directions for future research, the time and effort required to reach this conclusion may be great. Conversely, the scope of some reviews can be unpredictably large, and it may be difficult to plan the person-hours required to complete the research. The magnitude of this uncertainty has not been defined to date.

Recent efforts, such as the International Prospective Register of Systematic Reviews (PROSPERO)\(^8\) hosted through the University of York’s Centre for Reviews and Dissemination, have been deployed to prospectively and centrally register systematic reviews and meta-analyses. The PROSPERO registry “…aims to provide a comprehensive listing of systematic reviews registered at inception to help avoid unplanned duplication and enable comparison of reported review methods with what was planned in the protocol.”\(^8\) We found that such a database provides rich data for summarizing various logistical aspects of a sample of recently registered and completed systematic reviews. In the present meta-analysis, we aimed to use the PROSPERO registry to quantify the time and person-hours needed to conduct a systematic review. Our specific research questions were as follows:

1. How many people were involved in conducting or authoring the review?
2. How much time was required to complete and publish the review?
3. What was the average efficiency of the search strategy as indicated by the ratio of studies ultimately included in the review to the number of studies found in the database searches (i.e., yield rate)?
4. Did the number of people and time needed to complete a review differ between funded and unfunded reviews (regardless of funding source)?
Methods:

Study Selection

We searched the PROSPERO database hosted at the University of York’s Centre for Reviews and Dissemination. According to the website, “PROSPERO is an international database of prospectively registered systematic reviews in health and social care,” with an emphasis on intervention studies, although other reviews concerning patient or clinical relevance are also accepted. We retrieved 3684 registered project records from the PROSPERO website on July 1, 2014, and 437 records were marked as having a completed status. Of the 437 records for completed projects, 195 records contained a link to one or more publications of a completed review (not simply a protocol). Data were extracted from 195 publications (reporting a total of 197 literature reviews) by two authors (RB, KAK) and were verified by at least one of two additional authors (AWB, PLC). Although an audit trail is available for each PROSPERO record, there is no way to verify the completeness or accuracy of the entries or whether the registrants created the files early in the timeline of the project as prescribed by the registry administrators. Additionally, this registry is not equipped to verify the completeness or accuracy of the entries as compared to the resultant publication. See the text box for clarification of the terms we used in our analysis.

<table>
<thead>
<tr>
<th>Term Used</th>
<th>Data Used in Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>The persons listed on the published article reporting the results of the review. Not all team members were listed as authors of resulting publications, so we counted and evaluated “team members” separately. A sum of unique</td>
</tr>
</tbody>
</table>
names was generated to create a combined variable for analysis of people involved (Authors/Team Members).

<table>
<thead>
<tr>
<th>Team Members</th>
<th>The number of persons working on the review per the PROSPERO registry (see above for distinctions with “authors”).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews</td>
<td>The activities of searching the literature, selecting studies that meet inclusion criteria, and synthesis of study results. In our data, two governmental publications contained two review processes, thus resulting in our evaluation of 197 reviews reported in 195 publications.</td>
</tr>
<tr>
<td>Publications</td>
<td>The published results of the review process(es) in a scholarly journal or government document.</td>
</tr>
<tr>
<td>Studies</td>
<td>We counted studies and citations in Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagrams as the same thing, although some levels of PRISMA diagrams may contain some duplicate reports of studies in different citations.</td>
</tr>
</tbody>
</table>

**Data Extraction Process**

We extracted the following data from the published reviews: the dates a manuscript was received, accepted, and published (received and accepted dates were used only to compare to the registry date to validate the registered timeline); the number of authors; funding information; the
data sources searched (total number of databases or websites searched and their names); and the number of studies in the literature filtering steps from the PRISMA diagrams9 (when present in the articles). Companion data from the PROSPERO registry for each record included the number of team members, funding, registered start date, and geographic location of the team members. Authors/Team Members - The authors listed on the publication were counted, as were the team members in the registry. In all but seven cases, the number of authors was larger than the registered number of team members. We added the number of unique authors and registered team members in a combined variable to reflect personnel involvement. We also extracted country and institutional affiliations. Duplicate entries were determined based on comparing titles and team members, with the most complete record being retained.

Time – We estimated the duration of the project as the time from the registered project start date to the publication date of the review. When the article text did not include complete dates, we sought additional information on the publishers’ websites. For articles indexed in PubMed (all but 12), we used the PubMed IDs to extract the dates provided by the publishers to PubMed. These dates were related to the publication timeline (e.g., accepted date, first available online, final publication available date) when reported.

Search sources - For the initial total number of studies found in the database searches (highest level of the PRISMA diagram), we recorded what was reported in the PRISMA diagram (or text if the article or supplemental material did not contain a PRISMA diagram). When authors reported searching less commonly used sources of information such as local clinical trial registries or country-specific websites, we counted those sources in the “other” category and did not include them in the “total databases searched” variable. Authors do not consistently indicate the same level of the PRISMA diagram for the PRISMA category “studies identified from other
sources,” so we aggregated these sources in the variable we called “total N found” to compare across studies.

**Search efficiency** - We used yield rate as a metric for search efficiency, calculated by dividing the final number of included studies by the initial number of studies found (excluding duplicates) in the literature search. The final number of included studies was recorded from the PRISMA diagram or the text. If the PRISMA diagram or text separately reported the total number of studies included for quantitative (used in a statistical synthesis) and qualitative (used in a narrative summary only) synthesis, we used the highest number to calculate the overall merged yield rate. For example, if the PRISMA diagram noted that 20 unique studies were included in the review, and 8 were in the quantitative synthesis and 15 were in the qualitative summary, we used 20 for the overall merged yield rate owing to some overlap.

**Funding** - We coded reviews as being funded if the registry or publication text included explicit statements of review-level funding. We did not code studies as being “funded” if there were only general statements of salary support or conflict disclosures for individual authors (e.g., “Dr. Smith is funded by an NIH grant”). Explicit statements of author salary support for the review were recorded separately for analysis.

**Statistical Analysis**

To answer our research questions, we calculated the average number of authors/team members, time from registered start date to publication in weeks, and study yield rates. We counted the number of reviews that reported general funding for the project or review-related salary support of authors. We summarized frequency counts for each of the 12 most often reported databases used in literature searches in our sample. All summaries and analyses were calculated with SPSS version 22 (IBM, New York, USA) except where noted. Analysis of variance was used to
compare means for time to complete and number of authors/team members between funded and unfunded reviews. The summary literature distillation process was generated with R\textsuperscript{10} (Figure 1). Because of the extreme skewness for the study count variables from the PRISMA diagrams, we calculated z-scores, means, standard deviations, quartiles, and ranges from publications with complete data, removing outliers that were beyond 2.5 standard deviations to generate Figure 1. For reviews with incomplete data for any variable, we summarized only reported data and did not write authors to request missing information.

**Results:**

Table 1 summarizes the results for number of authors/team members, time needed to complete the reviews, and yield rate. The mean project length (using the registered project start date to the review’s publication date) was 67.3 weeks (SD = 31.0; range, 6-186 weeks). We also calculated a merged yield rate from the number found without duplicates in the initial search to final included studies (with some small overlap between quantitative and qualitative studies, n=190). The mean merged yield rate was 2.94% (SD = 6.49; range, 0-64.71%). Of the countries listed for the locations of registered team members, the United Kingdom was most often listed: 62 of the 197 reviews included UK team members.

Figure 1 summarizes the literature filtering process across reviews, indicating the initial number of studies found and the interim steps until the final mean number of studies included in the sample we analyzed. The use of the “Büchner” plot (shaped like a Büchner funnel and not to be confused with a “funnel plot” used to evaluate potential publication bias) allows us to emphasize several key characteristics of the process. First, the Büchner plot allows for all data to be plotted in such a way that demonstrates the wide range of included studies at each stage as well as the distribution of the data (highly right-skewed). Second, the ordinality of included studies was not
maintained from stage to stage. That is, the reviews with the greatest numbers at one stage were not necessarily the greatest at the next stage. This is visualized by lines crossing throughout the plot. Third, the figure demonstrates that the filtration process can be dramatic, as is reflected in the average yield rate being less than 3%.

Table 2 summarizes the differences in time to complete the review and number of authors/team members per project stratified by reported review-level or salary funding. Reviews that reported funding took longer to complete and included more authors/team members; this difference was not seen when we considered only whether specific authors were funded. Table 3 summarizes the most commonly used databases reported in the included reviews, with Medline, Embase and Cochrane being the top three.

**Discussion**

Our aim was not to conduct a comprehensive review of all systematic reviews, but rather to generate plausible estimates of the logistics of conducting reviews for medical interventions for review teams that engaged in at least one best practice: prospective review registration. We therefore chose to evaluate reviews that were registered in the PROSPERO registry and reported to be completed and published. Aspects of performing systematic reviews that remain undocumented (e.g., true start date of work, total person hours required), in different domains, or by teams that neither register their projects nor update their registrations upon completion may differ. Furthermore, the country affiliation reported in the registry for almost one-third of the team members was the United Kingdom. We cannot determine whether our results are representative of all systematic review teams or whether our results reflect an experience more likely to be encountered by UK researchers.
Our convenience sample has several limitations that must be considered. An estimate of the review project start date was difficult to capture from the registry data. The instructions on the PROSPERO site request that registration be done no later than prior to the completion of the data extraction stage. Our calculations for time were anchored by the registered date of the start of the project, but it takes some time to assemble the team, determine inclusion and exclusion criteria, conduct preliminary searches to refine the search syntax, perform inter-rater reliability for literature screening, and obtain funding (if the review is specifically funded). When a large number of studies are found in a search, these early-stage tasks need to be refined to a standard operating procedure, especially when the evaluation of the initial corpus of found studies cannot feasibly be performed in duplicate. We therefore predict that the time-to-publication from the very first activities of some reviews may be substantially longer than our estimates. As shown in Figure 1, at present, the full text of a large number of papers (median = 63, maximum = 4385) may need to be screened to evaluate final inclusion criteria. One factor that may impact the timeline is the choice of which and how many databases to search. The varied and mixed sources of searched databases in our dataset prevent conclusions about the potential impact of these choices on time and work, but others who have done explicit comparisons have noted that using Web of Science is more efficient that Google Scholar, for example.

Another limitation, as can be seen from Figure 1, is that the data at each stage of literature search and selection are very skewed, making attempts to predict timelines from various factors statistically unsound. Other questions that may be better answered with other data, for example, is a comparison of types of reviews (e.g., interventions versus diagnostic utility), differences between quantitative and qualitative reviews, or whether higher AMSTAR ratings are
associated with completion time or number of team members. These questions may need to be answered by using systematic surveys or prospective data collection methods.

One study used time logs spent performing 37 reviews to develop a prediction equation on the time to complete a meta-analysis using the number of initial citations retrieved as a predictor, and found that the greatest proportion of time involved was in the preanalysis search, retrieval, and database development phase. Overall, using people specializing in this work at a private company, the median total time was 1110 hours, range = 216 to 2518 hours. In contrast, our data likely reflect the situation in which people who perform systematic reviews may interleave this work with many other job duties, particularly in academia. The average elapsed time in the present study is more than one year, while the median time reported by Allen & Olkin is a little over a person-year of work time when done by specialists.

Further constraining the present conclusions, the PROSPERO registry does not currently function as a day-to-day project diary or require answers to detailed questions about methods such as the use of automated text screening or data extraction approaches that may affect the timeline or people required. Others who have evaluated the use of text mining and automated data extraction report that these methods may have some utility in certain types of reviews (e.g. scoping), but more work is needed to provide significant reductions in work required by humans. Finally, we focused on reviews published in journals rather than those that may be published on websites, and thus the latter may not be subject to delays often encountered in the journal submission and peer-review process unrelated to literature search or synthesis factors. Indeed, the number of rounds of submission and peer-review a particular review went through could add considerably to the time between project initiation and publication.
As the scientific literature continues to grow, generating high-quality, comprehensive reviews in a timely manner will become increasingly costly unless more automated resources are dedicated or different methods to search and retrieve relevant papers can be developed.\(^4\) As noted by the authors of the PRISMA statement, “A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing reliable findings from which conclusions can be drawn and decisions made.”\(^16\) Our results point to an ever-increasing logistical challenge to conducting systematic reviews so that unbiased conclusions can be made about medical interventions. For example, at the transition between the full text review and the final inclusion stages, our data (shown in Figure 1) indicates that a ratio of about .76 of papers retrieved are not included (untrimmed data ratio = .68). It cannot be determined from our data how much of this is due to very stringent inclusion/exclusion criteria or incomplete/discrepant reporting, although authors who report the details of why each of the papers were excluded allow readers to know such information for a given review. In some cases, improved reporting and deposition of data in repositories\(^17\) may reduce bias from what otherwise may have resulted in excluded studies.

A recent example that points to the emerging logistical challenges is illustrated in a study that evaluated matched pairs of reviews published on the same research question within 5 years of each other: one using Cochrane review methods and the other from other sources. These authors found a 47% difference in the number of studies in the paired reviews (excluding individual trials included in both reviews) unaccounted for by order of publication.\(^18\) One potential source of search differences may be the use of Medical Subject Headings (MeSH), which were introduced in 1963.\(^19\) Although this system is still able to provide a list of articles on a topic of
interest, it is not designed to meet the needs of today’s clinicians or systematic reviewers. This approach to indexing and searching the literature, along with author-assigned keywords, is no longer a serviceable approach because of the vast amount of irrelevant or missing articles provided when using such methods, owing to imprecise and overlapping keywords and MeSH terms that provide little context, as well as the lag times between publication and MeSH heading assignment that can occur. Clinicians have advised other clinicians who wish to search for studies to inform their clinical practice in PubMed: “Warning: do not look at all the articles found that although interesting are not pertinent to the present clinical question, or you will be lost in the sea of PubMed!”

A search for human clinical trials using the filters provided in PubMed with or without additional index terms can provide results with low sensitivity, precision, and specificity. Even something as simple as defining a study as a randomized controlled trial has been observed to be incorrect 20% of the time, which may add non-trivial work to a large review if this leaves many full text articles to be reviewed before the true study design can be ascertained. CONSORT reporting guidelines have improved this problem by adding study design as a requirement in the title, but even top-tier journals still do not enforce compliance 100% of the time.

Systematic reviews and meta-analyses can require large amounts of time and effort to complete, with often (based on our own experiences) unpredictable uncertainty in the time and resources required to complete a review. Although statistical methods have been developed to estimate trends in large corpora of literature, and crowdsourcing may help to decrease the calendar time needed to complete a review, we assert that the time is ripe to investigate metadata approaches to indexing publications so that more targeted yet comprehensive searches can be performed efficiently with high specificity and precision. We suggest that human clinical trials...
could be a starting point for testing the use of indexing systems that employ a Population, Intervention, Comparisons, Outcomes, Study Design (PICOS)\textsuperscript{16} structure of coding metadata for clinical trials in order to make the scientific literature become a truly searchable database. Our preliminary testing has shown that a modestly trained worker can code a single paper in about 15 minutes. By application of the recently proposed FAIR (Findable, Accessible, Interoperable and Reusable) metadata guidelines,\textsuperscript{4} clinical trial data (or other types of research data) could be encoded upon publication so that the data are in fact findable, accessible, and thus potentially more interoperable and reusable. An added benefit of this approach is that it would make the recently proposed mandated data sharing\textsuperscript{2} essentially automatic, leaving only the critical appraisal to be done. In a brighter future, evidence-based, up-to-date summaries could be produced on demand with less human effort and may reduce the delay between the question and the evidence-based answer.

\textbf{Acknowledgment}: The authors are grateful to Alison Booth and Jimmie Christy of the PROSPERO registry and to Madeline Jeansonne, MPH, and Eric Kim at the University of Alabama at Birmingham for assistance in data collection. Jennifer Holmes, ELS, performed the language editing of a draft of the manuscript. We are also grateful to the reviewers for their helpful comments and suggestions.

\textbf{Competing Interests Statement}: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that all authors have no financial or non-financial interests that may be relevant to the submitted work.
**Funding Statement:** Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Disease and the National Institute of General Medical Sciences of the National Institutes of Health under award numbers: P30DK056336 and K12GM088010 in the form of partial salary support for KAK, PLC and AWB. The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Alabama at Birmingham or the National Institutes of Health.

**Data Sharing:** The list of included review publications is in the online appendix. Analysis data may be obtained from the corresponding author at kakaiser@uab.edu.

**Authors’ contributions:** All authors contributed equally. KAK conceived the project. AWB and KAK retrieved and processed the data from the PROSPERO registry. RB retrieved articles and organized the reference bibliography. All authors extracted the data from included papers, analyzed the data and contributed to the writing of the manuscript. All authors have full access to all the data in the study and had final responsibility for the decision to submit for publication.

**Transparency Declaration:** Kathryn Kaiser affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

**Ethics Committee Approval:** No ethics committee approval is required for analyses of non-human subjects research. This paper used data about publications from public sources.
References


8. PROSPERO - International Prospective Register for Systematic Reviews. (University of York, Centre for Reviews and Dissemination, United Kingdom, 2015). Available at: http://www.crd.york.ac.uk/prospero/ (accessed 1/4/2016).


Figure 1. Aggregated literature filtration process based on counts reported (n=195). Trimmed means are indicated in the boxes and trimmed ranges (+/- 2.5 standard deviations) are indicated at the right and left of each level. Some reviews were published reporting that 0 studies met the criteria for inclusion in the review.
Table 1. Descriptive statistics for number of authors, time for publication, and quantitative or qualitative yield rates for 195 records analyzed in the PROSPERO registry.

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean, SD</th>
<th>Median</th>
<th>Interquartile Range</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors/Team Members (n = 195 publications)</td>
<td>5, 3</td>
<td>5</td>
<td>3</td>
<td>1 – 27</td>
</tr>
<tr>
<td>Time (in weeks; registered project start to publication date, n = 192*)</td>
<td>67.3, 31.0</td>
<td>65.8</td>
<td>41.6</td>
<td>6 – 186</td>
</tr>
<tr>
<td>Quantitative Analysis Yield Rate (n = 82), %</td>
<td>2.6, 4.7</td>
<td>1.0</td>
<td>2.7</td>
<td>0.03 – 32.43</td>
</tr>
<tr>
<td>Qualitative Analysis Yield Rate (n = 80), %</td>
<td>2.7, 4.6</td>
<td>1.0</td>
<td>2.5</td>
<td>0.05 – 26.19</td>
</tr>
<tr>
<td>Merged Yield Rate (n=190), %</td>
<td>2.94, 6.49</td>
<td>0.93</td>
<td>2.5</td>
<td>0.0 – 64.71</td>
</tr>
</tbody>
</table>

* Three studies were excluded from this calculation because they were registered after the publication date.

+ Excludes small overlap between quantitative and qualitative studies when information was provided in the publication to differentiate the categories.
Table 2. Comparison (analysis of variance) of reported funding of author salaries (n=20 for both outcomes) or review projects (n=86 for time, n=88 for authors) to time and authors needed to complete and publish the reviews.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Did outcome differ from reviews reporting no funding?</th>
<th>Means (reported as funded vs. funding not reported)</th>
<th>F, p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to complete (n=191)</td>
<td></td>
<td></td>
<td>df (1, 189)</td>
</tr>
<tr>
<td>Review funding reported</td>
<td>YES</td>
<td>42 weeks vs. 26 weeks</td>
<td>17.545, &lt;0.001</td>
</tr>
<tr>
<td>(n = 86)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salary funding reported</td>
<td>NO</td>
<td>68 weeks vs. 64 weeks</td>
<td>0.258, 0.612</td>
</tr>
<tr>
<td>(n = 20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Authors/Team Members (n=195)</td>
<td></td>
<td></td>
<td>df (1,193)</td>
</tr>
<tr>
<td>Review funding reported</td>
<td>YES</td>
<td>6.8 persons vs. 4.8 persons</td>
<td>14.638, &lt;0.001</td>
</tr>
<tr>
<td>(n = 88)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salary funding reported</td>
<td>NO</td>
<td>6.0 versus 5.7 persons</td>
<td>0.08, 0.778</td>
</tr>
<tr>
<td>(n = 20)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Top 12 databases used in included reviews in descending order (N = 197 reviews).

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Frequency, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>162, 82.2%</td>
</tr>
<tr>
<td>Embase</td>
<td>160, 81.2%</td>
</tr>
<tr>
<td>Cochrane</td>
<td>148, 75.1%</td>
</tr>
<tr>
<td>CINAHL</td>
<td>87, 44.2%</td>
</tr>
<tr>
<td>PubMed</td>
<td>58, 29.4%</td>
</tr>
<tr>
<td>Web of Science</td>
<td>56, 28.4%</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>49, 24.9%</td>
</tr>
<tr>
<td>SCOPUS</td>
<td>29, 14.7%</td>
</tr>
<tr>
<td>AMED</td>
<td>30, 15.2%</td>
</tr>
<tr>
<td>LILACS</td>
<td>24, 12.2%</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>13, 6.6%</td>
</tr>
<tr>
<td>ProQuest</td>
<td>12, 6.1%</td>
</tr>
</tbody>
</table>

CINAHL – Cumulative Index to Nursing and Allied Health Literature

AMED – Allied and Complementary Medicine Database

LILACS – Literatura Latina-Americana e do Caribe em Ciências da Saúde
Figure 1. Aggregated literature filtration process based on counts reported (n=195). Trimmed means are indicated in the boxes and trimmed ranges (± 2.5 standard deviations) are indicated at the right and left of each level. Some reviews were published reporting that 0 studies met the criteria for inclusion in the review.

89x63mm (300 x 300 DPI)
PROSPERO references of included reviews [1-195]


# STROBE Statement—Checklist of items that should be included in reports of cohort studies

<table>
<thead>
<tr>
<th>Item No</th>
<th>Recommendation</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title and abstract</strong></td>
<td>(a) Indicate the study’s design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found</td>
<td>1-2</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td>Explain the scientific background and rationale for the investigation being reported</td>
<td>3-5</td>
</tr>
<tr>
<td><strong>Objectives</strong></td>
<td>State specific objectives, including any prespecified hypotheses</td>
<td>5</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td>Present key elements of study design early in the paper</td>
<td>2, 6</td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td>Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection</td>
<td>6</td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (b) For matched studies, give matching criteria and number of exposed and unexposed</td>
<td>n/a</td>
</tr>
<tr>
<td><strong>Variables</strong></td>
<td>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable</td>
<td>6-9</td>
</tr>
<tr>
<td><strong>Data sources/measurement</strong></td>
<td>For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group</td>
<td>6-9</td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>Describe any efforts to address potential sources of bias</td>
<td>10</td>
</tr>
<tr>
<td><strong>Study size</strong></td>
<td>Explain how the study size was arrived at</td>
<td>6</td>
</tr>
<tr>
<td><strong>Quantitative variables</strong></td>
<td>Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why</td>
<td>6-10</td>
</tr>
<tr>
<td><strong>Statistical methods</strong></td>
<td>(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) If applicable, explain how loss to follow-up was addressed (e) Describe any sensitivity analyses</td>
<td>9-10</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td>(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage</td>
<td>6, n/a</td>
</tr>
</tbody>
</table>
(c) Consider use of a flow diagram

Descriptive data

14*

(a) Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders

(b) Indicate number of participants with missing data for each variable of interest

(c) Summarise follow-up time (e.g. average and total amount)

Outcome data

15*

Report numbers of outcome events or summary measures over time

n/a: Cross-sectional

Main results

16

(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g. 95% confidence interval). Make clear which confounders were adjusted for and why they were included

Fig. 1, Tables 1-3

(b) Report category boundaries when continuous variables were categorized

n/a

(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period

n/a

Other analyses

17

Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses

14

Discussion

18

Summarise key results with reference to study objectives

11-12

Limitations

19

Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias

12-13

Interpretation

20

Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence

14-15

Generalisability

21

Discuss the generalisability (external validity) of the study results

11-13

Other information

22

Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

16-17

*Give information separately for exposed and unexposed groups.

An analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry

Journal: BMJ Open

Manuscript ID: bmjopen-2016-012545.R2

Article Type: Research

Date Submitted by the Author: 21-Oct-2016

Complete List of Authors:
- Borah, Rohit; University of Alabama at Birmingham, Dean's Office, Office of Energetics
- Brown, Andrew; University of Alabama at Birmingham, Dean's Office, Office of Energetics
- Capers, Patrice; University of Alabama at Birmingham, Dean's Office, Office of Energetics
- Kaiser, Kathryn; University of Alabama at Birmingham, Dean's Office, Office of Energetics

Primary Subject Heading: Health informatics

Secondary Subject Heading: Evidence based practice

Keywords: systematic reviews, metadata, PROSPERO registry, search methods
An analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry

Rohit Borah¹,², Andrew W. Brown²,³, Ph.D., Patrice L. Capers²,³, Ph.D., & Kathryn A. Kaiser²,³, Ph.D.

1. Science and Technology Honors Program, University of Alabama at Birmingham, USA
2. Office of Energetics, Dean’s Office, School of Public Health, University of Alabama at Birmingham, USA
3. Nutrition Obesity Research Center, University of Alabama at Birmingham, USA

Keywords: systematic review, registry, meta-analysis, qualitative review, metadata

Funding: Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Disease and the National Institute of General Medical Sciences of the National Institutes of Health under award numbers: P30DK056336 and K12GM088010. The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Alabama at Birmingham or the National Institutes of Health.

Corresponding Author: Kathryn A. Kaiser, Ph.D. - Office of Energetics, Dean’s Office, School of Public Health, 1720 2nd Ave. South, WEBB 506, University of Alabama at Birmingham, Birmingham, AL 35294 USA, Email: kakaiser@uab.edu, Tel: +1-205-975-9212

Word count: 3735(text)
Abstract -

Objectives: To summarize logistical aspects of recently completed systematic reviews that were registered in the PROSPERO Systematic Review Registry to quantify the time and resources required to complete such projects.

Design: Meta-analysis

Data Sources and Study selection: All of the 195 registered and completed reviews (status from the PROSPERO registry) with associated publications at the time of our search (1 July 2014).

Data extraction: All authors extracted data using registry entries and publication information related to the data sources used, the number of initially retrieved citations, the final number of included studies, the time between registration date to publication date, and number of authors involved for completion of each publication. Information related to funding and geographical location was also recorded, when reported.

Results: The mean estimated time to complete the project and publish the review was 67.3 weeks (interquartile range = 42). The number of studies found in the literature searches ranged from 27 to 92,020; the mean yield rate of included studies was 2.94% (interquartile range = 2.5); and the mean number of authors per review was 5, SD = 3. Funded reviews took significantly longer to complete and publish (mean= 42 versus 26 weeks) and involved more authors and team members (mean= 6.8 versus 4.8 people) than those that did not report funding (both p<0.001).

Conclusions: Systematic reviews presently take much time and require large amounts of human resources. In light of the ever increasing volume of published studies, application of existing computing and informatics technology should be applied to decrease this time and resource
burden. We discuss recently published guidelines that provide a framework to make finding and accessing relevant literature less burdensome.

Systematic review registration: Not registered; no known registry currently accepts methodological systematic reviews of this type.

Strengths and limitations of this study:

- This study provides an updated estimate of the time and effort required to conduct and publish a systematic review using a large sample of recently published reviews across a variety of topics of medical interventions.

- The study is limited by incomplete reporting in both sources of the data – the review registry and the published articles. These data sources were cross checked as data were available and conservative assumptions were stated and applied where necessary.
Introduction

The systematic review can be an effective and scientific method used across disciplines to consolidate vast amounts of research on a specific topic. Over the last 20 years, publishing of systematic reviews has increased exponentially, as has the primary literature. As the body of primary literature has increased, conducting systematic reviews and meta-analyses has become a necessity to more accurately and comprehensively present accumulated knowledge to scientific, clinical, and general audiences. In fact, new mandates are being formulated to require documented systematic reviews as part of research proposals and clinical trial data sharing, both of which have strong implications for how scientific literature and associated data are reported, managed, and curated.

The Cochrane Database of Systematic Reviews provides important guidelines and methods for systematic reviews through its Cochrane Handbook, which states four key considerations to define prior to beginning: the question, the inclusion and exclusion criteria, the search strategy, and the methods. In other words, systematic reviews should review the scientific literature in a scientific manner. The Cochrane Collaboration has built its reputation as the leading resource for systematic reviews by requiring that all Cochrane reviews be “…updated regularly in an effort to ensure that the most recent evidence is incorporated.” However, this onerous requirement may have the side effect of limiting the topics that are created and kept current.

With ever-growing numbers of journals and publications, an immense amount of time and effort is needed to search the literature and summarize the findings. This is reflected in the Cochrane Collaboration statement that “[h]istorically, the aim was to update Cochrane reviews every two years, but recently there has been a move away from this policy in favour of prioritising the most clinically important reviews for updating.” When beginning a new systematic review, authors...
are faced with the possibility of finding few to no studies that meet their criteria. While finding no studies that meet the criteria can be informative per se, such as by identifying directions for future research, the time and effort required to reach this conclusion may be great. Conversely, the scope of some reviews can be unpredictably large, and it may be difficult to plan the person-hours required to complete the research. The magnitude of this uncertainty has not been defined to date.

Recent efforts, such as the International Prospective Register of Systematic Reviews (PROSPERO) hosted through the University of York’s Centre for Reviews and Dissemination, have been deployed to prospectively and centrally register systematic reviews and meta-analyses. The PROSPERO registry “…aims to provide a comprehensive listing of systematic reviews registered at inception to help avoid unplanned duplication and enable comparison of reported review methods with what was planned in the protocol.” We found that such a database provides rich data for summarizing various logistical aspects of a sample of recently registered and completed systematic reviews. In the present meta-analysis, we aimed to use the PROSPERO registry to quantify the time and person-hours needed to conduct a systematic review. Our specific research questions were as follows:

1. How many people were involved in conducting or authoring the review?
2. How much time was required to complete and publish the review?
3. What was the average efficiency of the search strategy as indicated by the ratio of studies ultimately included in the review to the number of studies found in the database searches (i.e., yield rate)?
4. Did the number of people and time needed to complete a review differ between funded and unfunded reviews (regardless of funding source)?
Methods:

Study Selection

We searched the PROSPERO database hosted at the University of York’s Centre for Reviews and Dissemination. According to the website, “PROSPERO is an international database of prospectively registered systematic reviews in health and social care,” with an emphasis on intervention studies, although other reviews concerning patient or clinical relevance are also accepted. We retrieved 3684 registered project records from the PROSPERO website on July 1, 2014, and 437 records were marked as having a completed status. Of the 437 records for completed projects, 195 records contained a link to one or more publications of a completed review (not simply a protocol). Data were extracted from 195 publications (reporting a total of 197 literature reviews) by two authors (RB, KAK) and were verified by at least one of two additional authors (AWB, PLC). Although an audit trail is available for each PROSPERO record, there is no way to verify the completeness or accuracy of the entries or whether the registrants created the files early in the timeline of the project as prescribed by the registry administrators. Additionally, this registry is not equipped to verify the completeness or accuracy of the entries as compared to the resultant publication. See the text box for clarification of the terms we used in our analysis.

<table>
<thead>
<tr>
<th>Term Used</th>
<th>Data Used in Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>The persons listed on the published article reporting the results of the review. Not all team members were listed as authors of resulting publications, so we counted and evaluated “team members” separately. A sum of unique</td>
</tr>
</tbody>
</table>

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml
names was generated to create a combined variable for analysis of people involved (Authors/Team Members).

<table>
<thead>
<tr>
<th>Team Members</th>
<th>The number of persons working on the review per the PROSPERO registry (see above for distinctions with “authors”).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews</td>
<td>The activities of searching the literature, selecting studies that meet inclusion criteria, and synthesis of study results. In our data, two governmental publications contained two review processes, thus resulting in our evaluation of 197 reviews reported in 195 publications.</td>
</tr>
<tr>
<td>Publications</td>
<td>The published results of the review process(es) in a scholarly journal or government document.</td>
</tr>
<tr>
<td>Studies</td>
<td>We counted studies and citations in Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagrams as the same thing, although some levels of PRISMA diagrams may contain some duplicate reports of studies in different citations.</td>
</tr>
</tbody>
</table>

**Data Extraction Process**

We extracted the following data from the published reviews: the dates a manuscript was received, accepted, and published (received and accepted dates were used only to compare to the registry date to validate the registered timeline); the number of authors; funding information; the
data sources searched (total number of databases or websites searched and their names); and the number of studies in the literature filtering steps from the PRISMA diagrams (when present in the articles). Companion data from the PROSPERO registry for each record included the number of team members, funding, registered start date, and geographic location of the team members.

**Authors/Team Members** - The authors listed on the publication were counted, as were the team members in the registry. In all but seven cases, the number of authors was larger than the registered number of team members. We added the number of unique authors and registered team members in a combined variable to reflect personnel involvement. We also extracted country and institutional affiliations. Duplicate entries were determined based on comparing titles and team members, with the most complete record being retained.

**Time** – We estimated the duration of the project as the time from the registered project start date to the publication date of the review. When the article text did not include needed date information, we sought additional information on the publishers’ websites. For articles indexed in PubMed (all but 12), we used the PubMed IDs to extract the dates provided by the publishers to PubMed. These dates were related to the publication timeline (e.g., accepted date, first available online, final publication available date) when reported.

**Search sources** - For the initial total number of studies found in the database searches (highest level of the PRISMA diagram), we recorded what was reported in the PRISMA diagram (or text if the article or supplemental material did not contain a PRISMA diagram). When authors reported searching less commonly used sources of information such as local clinical trial registries or country-specific websites, we counted those sources in the “other” category and did not include them in the “total databases searched” variable. Authors do not consistently indicate the same level of the PRISMA diagram for the PRISMA category “studies identified from other
sources,” so we aggregated these sources in the variable we called “total N found” to compare across studies.

Search efficiency - We used yield rate as a metric for search efficiency, calculated by dividing the final number of included studies by the initial number of studies found (excluding duplicates) in the literature search. The final number of included studies was recorded from the PRISMA diagram or the text. If the PRISMA diagram or text separately reported the total number of studies included for quantitative (used in a statistical synthesis) and qualitative (used in a narrative summary only) synthesis, we used the highest number to calculate the overall merged yield rate. For example, if the PRISMA diagram noted that 20 unique studies were included in the review, and 8 were in the quantitative synthesis and 15 were in the qualitative summary, we used 20 for the overall merged yield rate owing to some overlap.

Funding - We coded reviews as being funded if the registry or publication text included explicit statements of review-level funding. We did not code studies as being “funded” if there were only general statements of salary support or conflict disclosures for individual authors (e.g., “Dr. Smith is funded by an NIH grant”). Explicit statements of author salary support for the review were recorded separately for analysis.

Statistical Analysis

To answer our research questions, we calculated the average number of authors/team members, time from registered start date to publication in weeks, and study yield rates. We counted the number of reviews that reported general funding for the project or review-related salary support of authors. We summarized frequency counts for each of the 12 most often reported databases used in literature searches in our sample. All summaries and analyses were calculated with SPSS version 22 (IBM, New York, USA) except where noted. Analysis of variance was used to
compare means for time to complete and number of authors/team members between funded and unfunded reviews. The summary literature distillation process was generated with R\textsuperscript{10} (Figure 1).

Because of the extreme skewness for the study count variables from the PRISMA diagrams, we calculated z-scores, means, standard deviations, quartiles, and ranges from publications with complete data, removing outliers that were beyond 2.5 standard deviations to generate Figure 1. For reviews with incomplete data for any variable, we summarized only reported data and did not write authors to request missing information.

**Results:**

Table 1 summarizes the results for number of authors/team members, time needed to complete the reviews, and yield rate. The mean project length (using the registered project start date to the review’s publication date) was 67.3 weeks (SD = 31.0; range, 6-186 weeks). We also calculated a merged yield rate from the number found without duplicates in the initial search to final included studies (with some small overlap between quantitative and qualitative studies, n=190). The mean merged yield rate was 2.94% (SD = 6.49; range, 0-64.71%). Of the countries listed for the locations of registered team members, the United Kingdom was most often listed: 62 of the 197 reviews included UK team members.

Figure 1 summarizes the literature filtering process across reviews, indicating the initial number of studies found and the interim steps until the final mean number of studies included in the sample we analyzed. The use of the “Büchner” plot (shaped like a Büchner funnel and not to be confused with a “funnel plot” used to evaluate potential publication bias) allows us to emphasize several key characteristics of the process. First, the Büchner plot allows for all data to be plotted in such a way that demonstrates the wide range of included studies at each stage as well as the distribution of the data (highly right-skewed). Second, the ordinality of included studies was not
maintained from stage to stage. That is, the reviews with the greatest numbers at one stage were
not necessarily the greatest at the next stage. This is visualized by lines crossing throughout the
plot. Third, the figure demonstrates that the filtration process can be dramatic, as is reflected in
the average yield rate being less than 3%.

Table 2 summarizes the differences in time to complete the review and number of authors/team
members per project stratified by reported review-level or salary funding. Reviews that reported
funding took longer to complete and included more authors/team members; this difference was
not seen when we considered only whether specific authors were funded. Table 3 summarizes
the most commonly used databases reported in the included reviews, with Medline, Embase and
Cochrane being the top three.

Discussion

Our aim was not to conduct a comprehensive review of all systematic reviews, but rather to
generate plausible estimates of the logistics of conducting reviews for medical interventions for
review teams that engaged in at least one best practice: prospective review registration. We
therefore chose to evaluate reviews that were registered in the PROSPERO registry and reported
to be completed and published. Aspects of performing systematic reviews that remain
undocumented (e.g., true start date of work, total person hours required), in different domains, or
by teams that neither register their projects nor update their registrations upon completion may
differ. Furthermore, the country affiliation reported in the registry for almost one-third of the
team members was the United Kingdom. We cannot determine whether our results are
representative of all systematic review teams or whether our results reflect an experience more
likely to be encountered by UK researchers.
Our convenience sample has several limitations that must be considered. An estimate of the review project start date was difficult to capture from the registry data. The instructions on the PROSPERO site request that registration be done no later than prior to the completion of the data extraction stage. Our calculations for time were anchored by the registered date of the start of the project, but it takes some time to assemble the team, determine inclusion and exclusion criteria, conduct preliminary searches to refine the search syntax, perform inter-rater reliability for literature screening, and obtain funding (if the review is specifically funded). When a large number of studies are found in a search, these early-stage tasks need to be refined to a standard operating procedure, especially when the evaluation of the initial corpus of found studies cannot feasibly be performed in duplicate. We therefore predict that the time-to-publication from the very first activities of some reviews may be substantially longer than our estimates. As shown in Figure 1, at present, the full text of a large number of papers (median = 63, maximum = 4385) may need to be screened to evaluate final inclusion criteria. One factor that may impact the timeline is the choice of which and how many databases to search. The varied and mixed sources of searched databases in our dataset prevent conclusions about the potential impact of these choices on time and work, but others who have done explicit comparisons have noted that using Web of Science is more efficient that Google Scholar, for example.

Another limitation, as can be seen from Figure 1, is that the data at each stage of literature search and selection are very skewed, making attempts to predict timelines from various factors statistically unsound. Other questions that may be better answered with other data, for example, is a comparison of types of reviews (e.g., interventions versus diagnostic utility), differences between quantitative and qualitative reviews, or whether higher AMSTAR ratings are
associated with completion time or number of team members. These questions may need to be answered by using systematic surveys or prospective data collection methods.

One study used time logs spent performing 37 reviews to develop a prediction equation on the time to complete a meta-analysis using the number of initial citations retrieved as a predictor, and found that the greatest proportion of time involved was in the preanalysis search, retrieval, and database development phase. Overall, using people specializing in this work at a private company, the median total time was 1110 hours, range = 216 to 2518 hours. In contrast, our data likely reflect the situation in which people who perform systematic reviews may interleave this work with many other job duties, particularly in academia. The average elapsed time in the present study is more than one year, while the median time reported by Allen & Olkin is a little over a person-year of work time when done by specialists.

Further constraining the present conclusions, the PROSPERO registry does not currently function as a day-to-day project diary or require answers to detailed questions about methods such as the use of automated text screening or data extraction approaches that may affect the timeline or people required. Others who have evaluated the use of text mining and automated data extraction report that these methods may have some utility in certain types of reviews (e.g. scoping), but more work is needed to provide significant reductions in work required by humans. Finally, we focused on reviews published in journals rather than those that may be published on websites, and thus the latter may not be subject to delays often encountered in the journal submission and peer-review process unrelated to literature search or synthesis factors. Indeed, the number of rounds of submission and peer-review a particular review went through could add considerably to the time between project initiation and publication.
As the scientific literature continues to grow, generating high-quality, comprehensive reviews in a timely manner will become increasingly costly unless more automated resources are dedicated or different methods to search and retrieve relevant papers can be developed. As noted by the authors of the PRISMA statement, “A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing reliable findings from which conclusions can be drawn and decisions made.”

Our results point to an ever-increasing logistical challenge to conducting systematic reviews so that unbiased conclusions can be made about medical interventions. For example, at the transition between the full text review and the final inclusion stages, our data (shown in Figure 1) indicates that a ratio of about .76 of papers retrieved are not included (untrimmed data ratio = .68). It cannot be determined from our data how much of this is due to very stringent inclusion/exclusion criteria or incomplete/discrepant reporting, although authors who report the details of why each of the papers were excluded allow readers to know such information for a given review. In some cases, improved reporting and deposition of data in repositories may reduce bias from what otherwise may have resulted in excluded studies.

A recent example that points to the emerging logistical challenges is illustrated in a study that evaluated matched pairs of reviews published on the same research question within 5 years of each other: one using Cochrane review methods and the other from other sources. These authors found a 47% difference in the number of studies in the paired reviews (excluding individual trials included in both reviews) unaccounted for by order of publication. One potential source of search differences may be the use of Medical Subject Headings (MeSH), which were introduced in 1963. Although this system is still able to provide a list of articles on a topic of
interest, it is not designed to meet the needs of today’s clinicians or systematic reviewers. This approach to indexing and searching the literature, along with author-assigned keywords, is no longer a serviceable approach because of the vast amount of irrelevant or missing articles provided when using such methods, owing to imprecise and overlapping keywords and MeSH terms that provide little context, as well as the lag times between publication and MeSH heading assignment that can occur. An additional potential explanation for the observed differences in reviews reported in this study\(^\text{18}\) may be some differences in study selection criteria or databases searched. Clinicians have advised other clinicians who wish to search for studies to inform their clinical practice in PubMed: “Warning: do not look at all the articles found that although interesting are not pertinent to the present clinical question, or you will be lost in the sea of PubMed!”\(^\text{20}\) A search for human clinical trials using the filters provided in PubMed with or without additional index terms can provide results with low sensitivity, precision, and specificity.\(^\text{21}\) Even something as simple as defining a study as a randomized controlled trial has been observed to be incorrect 20% of the time, which may add non-trivial work to a large review if this leaves many full text articles to be reviewed before the true study design can be ascertained.\(^\text{22}\) CONSORT reporting guidelines have improved this problem by adding study design as a requirement in the title, but even top-tier journals still do not enforce compliance 100% of the time\(^\text{23}\).

Systematic reviews and meta-analyses can require large amounts of time and effort to complete, with often (based on our own experiences) unpredictable uncertainty in the time and resources required to complete a review. Although statistical methods have been developed to estimate trends in large corpora of literature,\(^\text{22}\) and crowdsourcing may help to decrease the calendar time needed to complete a review,\(^\text{24}\) we assert that the time is ripe to investigate metadata\(^\text{4, 25}\).
approaches to indexing publications so that more targeted yet comprehensive searches can be performed efficiently with high specificity and precision. We suggest that human clinical trials could be a starting point for testing the use of indexing systems that employ a Population, Intervention, Comparisons, Outcomes, Study Design (PICOS) structure of coding metadata for clinical trials in order to make the scientific literature become a truly searchable database. Our preliminary testing has shown that a modestly trained worker can code a single paper in about 15 minutes. By application of the recently proposed FAIR (Findable, Accessible, Interoperable and Reusable) metadata guidelines, clinical trial data (or other types of research data) could be encoded upon publication so that the data are in fact findable, accessible, and thus potentially more interoperable and reusable. An added benefit of this approach is that it would make the recently proposed mandated data sharing essentially automatic, leaving the critical appraisal, synthesis and interpretation to be done. In a brighter future, evidence-based, up-to-date summaries could be produced on demand with less human effort and may reduce the delay between the question and the evidence-based answer.

Acknowledgment: The authors are grateful to Alison Booth and Jimmie Christy of the PROSPERO registry and to Madeline Jeansonne, MPH, and Eric Kim at the University of Alabama at Birmingham for assistance in data collection. Jennifer Holmes, ELS, performed the language editing of a draft of the manuscript. We are also grateful to the reviewers for their helpful comments and suggestions.

Competing Interests Statement: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author)
and declare that all authors have no financial or non-financial interests that may be relevant to the submitted work.

**Funding Statement:** Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Disease and the National Institute of General Medical Sciences of the National Institutes of Health under award numbers: P30DK056336 and K12GM088010 in the form of partial salary support for KAK, PLC and AWB. The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Alabama at Birmingham or the National Institutes of Health.

**Data Sharing:** The list of included review publications is in the online appendix. Analysis data may be obtained from the corresponding author at kakaiser@uab.edu.

**Authors’ contributions:** All authors contributed equally. KAK conceived the project. AWB and KAK retrieved and processed the data from the PROSPERO registry. RB retrieved articles and organized the reference bibliography. All authors extracted the data from included papers, analyzed the data and contributed to the writing of the manuscript. All authors have full access to all the data in the study and had final responsibility for the decision to submit for publication.

**Transparency Declaration:** Kathryn Kaiser affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

**Ethics Committee Approval:** No ethics committee approval is required for analyses of non-human subjects research. This paper used data about publications from public sources.
References


8. PROSPERO - International Prospective Register for Systematic Reviews. (University of York, Centre for Reviews and Dissemination, United Kingdom, 2015). Available at: http://www.crd.york.ac.uk/prospero/ (accessed 1/4/2016).


Figure 1. Aggregated literature filtration process based on counts reported (n=195). Trimmed means are indicated in the boxes and trimmed ranges (+/- 2.5 standard deviations) are indicated at the right and left of each level. Some reviews were published reporting that 0 studies met the criteria for inclusion in the review.
Table 1. Descriptive statistics for number of authors, time for publication, and quantitative or qualitative yield rates for 195 records analyzed in the PROSPERO registry.

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean, SD</th>
<th>Median</th>
<th>Interquartile Range</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors/Team Members (n = 195 publications)</td>
<td>5, 3</td>
<td>5</td>
<td>3</td>
<td>1 – 27</td>
</tr>
<tr>
<td>Time (in weeks; registered project start to publication date, n = 192*)</td>
<td>67.3, 31.0</td>
<td>65.8</td>
<td>41.6</td>
<td>6 – 186</td>
</tr>
<tr>
<td>Quantitative Analysis Yield Rate (n = 82), %</td>
<td>2.6, 4.7</td>
<td>1.0</td>
<td>2.7</td>
<td>0.03 – 32.43</td>
</tr>
<tr>
<td>Qualitative Analysis Yield Rate (n = 80), %</td>
<td>2.7, 4.6</td>
<td>1.0</td>
<td>2.5</td>
<td>0.05 – 26.19</td>
</tr>
<tr>
<td>Merged Yield Rate (n=190), %</td>
<td>2.94, 6.49</td>
<td>0.93</td>
<td>2.5</td>
<td>0.0 – 64.71</td>
</tr>
</tbody>
</table>

* Three studies were excluded from this calculation because they were registered after the publication date.

+ Excludes small overlap between quantitative and qualitative studies when information was provided in the publication to differentiate the categories.
Table 2. Comparison (analysis of variance) of reported funding of author salaries (n=20 for both outcomes) or review projects (n=86 for time, n=88 for authors) to time and authors needed to complete and publish the reviews.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Did outcome differ from reviews reporting no funding?</th>
<th>Means (reported as funded vs. funding not reported)</th>
<th>F, p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to complete (n=191)</td>
<td></td>
<td></td>
<td>df (1, 189)</td>
</tr>
<tr>
<td>Review funding reported</td>
<td>YES</td>
<td>42 weeks vs. 26 weeks</td>
<td>17.545, &lt;0.001</td>
</tr>
<tr>
<td>(n = 86)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salary funding reported</td>
<td>NO</td>
<td>68 weeks vs. 64 weeks</td>
<td>0.258, 0.612</td>
</tr>
<tr>
<td>(n = 20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Authors/Team Members (n=195)</td>
<td></td>
<td></td>
<td>df (1,193)</td>
</tr>
<tr>
<td>Review funding reported</td>
<td>YES</td>
<td>6.8 persons vs. 4.8 persons</td>
<td>14.638, &lt;0.001</td>
</tr>
<tr>
<td>(n = 88)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salary funding reported</td>
<td>NO</td>
<td>6.0 versus 5.7 persons</td>
<td>0.08, 0.778</td>
</tr>
<tr>
<td>(n = 20)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Top 12 databases used in included reviews in descending order (N = 197 reviews).

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Frequency, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>162, 82.2%</td>
</tr>
<tr>
<td>Embase</td>
<td>160, 81.2%</td>
</tr>
<tr>
<td>Cochrane</td>
<td>148, 75.1%</td>
</tr>
<tr>
<td>CINAHL</td>
<td>87, 44.2%</td>
</tr>
<tr>
<td>PubMed</td>
<td>58, 29.4%</td>
</tr>
<tr>
<td>Web of Science</td>
<td>56, 28.4%</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>49, 24.9%</td>
</tr>
<tr>
<td>SCOPUS</td>
<td>29, 14.7%</td>
</tr>
<tr>
<td>AMED</td>
<td>30, 15.2%</td>
</tr>
<tr>
<td>LILACS</td>
<td>24, 12.2%</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>13, 6.6%</td>
</tr>
<tr>
<td>ProQuest</td>
<td>12, 6.1%</td>
</tr>
</tbody>
</table>

CINAHL – Cumulative Index to Nursing and Allied Health Literature

AMED – Allied and Complementary Medicine Database

LILACS – Literatura Latina-Americana e do Caribe em Ciências da Saúde
Figure 1. Aggregated literature filtration process based on counts reported (n=195). Trimmed means are indicated in the boxes and trimmed ranges (+/- 2.5 standard deviations) are indicated at the right and left of each level. Some reviews were published reporting that 0 studies met the criteria for inclusion in the review.

89x63mm (300 x 300 DPI)
PROSPERO references of included reviews [1-195]


### STROBE Statement—Checklist of items that should be included in reports of cohort studies

<table>
<thead>
<tr>
<th>Item</th>
<th>Recommendation</th>
<th>Page Number</th>
</tr>
</thead>
</table>
| **Title and abstract** | (a) Indicate the study’s design with a commonly used term in the title or the abstract  
(b) Provide in the abstract an informative and balanced summary of what was done and what was found | 1-2 |
| **Introduction** | | |
| Background/rationale | Explain the scientific background and rationale for the investigation being reported | 3-5 |
| Objectives | State specific objectives, including any prespecified hypotheses | 5 |
| **Methods** | | |
| Study design | Present key elements of study design early in the paper | 2, 6 |
| Setting | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 6 |
| Participants | (a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up  
(b) For matched studies, give matching criteria and number of exposed and unexposed | n/a |
| Variables | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 6-9 |
| Data sources/measurement | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | 6-9 |
| Bias | Describe any efforts to address potential sources of bias | 10 |
| Study size | Explain how the study size was arrived at | 6 |
| Quantitative variables | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | 6-10 |
| Statistical methods | (a) Describe all statistical methods, including those used to control for confounding  
(b) Describe any methods used to examine subgroups and interactions  
(c) Explain how missing data were addressed  
(d) If applicable, explain how loss to follow-up was addressed  
(e) Describe any sensitivity analyses | 9-10 |
| **Results** | | |
| Participants | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed  
(b) Give reasons for non-participation at each stage | 6, n/a |
(c) Consider use of a flow diagram Figure 1

<table>
<thead>
<tr>
<th>Section</th>
<th>Code</th>
<th>Item Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive data</td>
<td>14*</td>
<td>(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders</td>
<td>Table 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Indicate number of participants with missing data for each variable of interest</td>
<td>Table 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) Summarise follow-up time (eg, average and total amount)</td>
<td>n/a</td>
</tr>
<tr>
<td>Outcome data</td>
<td>15*</td>
<td>Report numbers of outcome events or summary measures over time</td>
<td>n/a: Cross-sectional</td>
</tr>
<tr>
<td>Main results</td>
<td>16</td>
<td>(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included</td>
<td>Fig. 1, Tables 1-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Report category boundaries when continuous variables were categorized</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period</td>
<td>n/a</td>
</tr>
<tr>
<td>Other analyses</td>
<td>17</td>
<td>Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses</td>
<td>14</td>
</tr>
<tr>
<td>Discussion</td>
<td>18</td>
<td>Summarise key results with reference to study objectives</td>
<td>11-12</td>
</tr>
<tr>
<td>Limitations</td>
<td>19</td>
<td>Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias</td>
<td>12-13</td>
</tr>
<tr>
<td>Interpretation</td>
<td>20</td>
<td>Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence</td>
<td>14-15</td>
</tr>
<tr>
<td>Generalisability</td>
<td>21</td>
<td>Discuss the generalisability (external validity) of the study results</td>
<td>11-13</td>
</tr>
<tr>
<td>Other information</td>
<td>22</td>
<td>Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based</td>
<td>16-17</td>
</tr>
</tbody>
</table>

*Give information separately for exposed and unexposed groups.