Conservation Biology

Conservation Methods

Predicting the time needed for environmental systematic reviews and systematic maps

Neal R. Haddaway [1,2]* and Martin J. Westgate [3]

[1]Mistra EviEM, Stockholm Environment Institute, Linnégatan 87D, Stockholm, Sweden
[2]Africa Centre for Evidence, University of Johannesburg, P.O. Box 524, 2006, Auckland Park, South Africa
[3]Fenner School of Environment and Society, The Australian National University, 2601, Acton, Australia

Abstract: *Systematic reviews (SRs) and systematic mapping aim to maximize transparency and comprehensiveness while minimizing subjectivity and bias. These are time-consuming and complex tasks, so SRs are considered resource intensive, but published estimates of systematic-review resource requirements are largely anecdotal. We analyzed all Collaboration for Environmental Evidence (CEE) SRs (*n = 66*) and maps (*n = 20*) published from 2012 to 2017 to estimate the average number of articles retained at each review stage. We also surveyed 33 experienced systematic reviewers to collate information on the rate at which those stages could be completed. In combination, these data showed that the average CEE SR takes an estimated 164 d (full-time equivalent) (SD 23), and the average CEE systematic map (SM) (excluding critical appraisal) takes 211 d (SD 53). While screening titles and abstracts is widely considered time-consuming, metadata extraction and critical appraisal took as long or longer to complete, especially for SMs. Given information about the planned methods and evidence base, we created a software tool that predicts time requirements of a SR or map with evidence-based defaults as a starting point. Our results shed light on the most time-consuming stages of the SR and mapping processes, will inform review planning, and can direct innovation to streamline processes. Future predictions of effort required to complete SRs and maps could be improved if authors provide more details on methods and results.*

Keywords: cost, efficiency, evidence synthesis, literature review, time commitment, workload

Pronóstico del Tiempo Necesario para las Revisiones Ambientales Sistemáticas y los Mapas Sistemáticos

Resumen: *El mapeo sistemático y las revisiones sistemáticas buscan maximizar la transparencia y la exhaustividad mientras minimizan la subjetividad y la parcialidad. Estas son labores complejas que consumen tiempo, por lo que las revisiones sistemáticas se consideran como intensivas en recursos, pero en el caso de los requerimientos de los recursos para las revisiones sistemáticas las estimaciones publicadas son en su mayoría anecdóticas. Analizamos todas las revisiones sistemáticas (*n = 66*) y todos los mapas (*n = 20*) de la Colaboración para la Evidencia Ambiental (CEE, en inglés) publicados entre 2012 y 2017 para estimar el número promedio de artículos retenidos en cada etapa de revisión. También encuestamos a 33 revisores sistemáticos experimentado para cotejar la información sobre la tasa a la cual se podrían completar esas etapas. La combinación de estos datos mostró que la revisión sistemática promedio del CEE tarda un estimado de 164 días (equivalente de tiempo completo) (SD 23), y que el mapa sistemático promedio del CEE (excluyendo la evaluación crítica) tarda 211 días (SD 53). Se considera ampliamente que el proceso de selección de títulos y resúmenes consume mucho tiempo, pero la extracción de meta-datos y la evaluación crítica tarda la misma cantidad de tiempo, o más, para completarse, especialmente en el caso de los mapas sistemáticos. Con la información sobre los métodos planeados y la base de evidencias creamos una herramienta de software que predice los requerimientos de tiempo para un mapa sistemático o una revisión sistemática con defaults basados en evidencias como puntos de partida. Nuestros resultados traen a la luz las etapas de la revisión*

*email neal_haddaway@hotmail.com
Article impact statement*: We provide data on the effort needed to complete systematic literature reviews and maps and new software to help users apply our findings.*
Paper submitted January 23, 2018; revised manuscript accepted July 18, 2018.

Conservation Biology, Volume 0, No. 0, 1–10

DOI: 10.1111/cobi.13231

sistemática o del mapeo sistemático que más tiempo consumen, informarán sobre la planeación de revisiones, y pueden dirigir la innovación en los procesos simplificados. Los pronósticos futuros del esfuerzo requerido para completar los mapas y las revisiones sistémicos podría mejorarse si los autores proporcionar más detalles sobre los métodos y los resultados.

Palabras Clave: carga laboral, compromiso de tiempo, costo, eficiencia, revisión de literatura, síntesis de evidencias

摘要: *系统评估和系统绘图的目的是最大限度地提高透明度和全面性, 尽量减少主观性和偏见。这样的任务复杂且耗时, 所以系统评估被认为是资源密集型的, 然而, 已发表的关于系统评估所需资源的估计却又没有充足的依据。我们分析了《环境证据》协作组织 (Collaboration for Environmental Evidence, CEE) 在 2012 至 2017 年间发表的所有的系统评估 (n = 66) 和绘图 (n = 20) , 来估计每个评估阶段保留的文章平均数量。我们还访问了33 名有经验的系统评估专家, 分析得出各个阶段的完成速度。总的来说, 这些数据表明, CEE 系统评估平均需要约164 天 (全日制) (标准偏差 23 天) , CEE 系统绘图 (不包括评读文献) 平均需要 211 天 (标准偏差 53 天) 。虽然人们普遍认为筛选标题和摘要很费时, 但宏数据提取和严格评价也需要一样长甚至更长的时间, 特别是对于系统绘图来说。利用所设计的方法和证据基础的信息, 我们开发了一款以基于证据的默认值为起点的软件来预测进行系统评估或绘图所需时间。我们的结果找出了系统评估和绘图过程中最耗时的阶段, 可以为评估规划提供信息, 并指导方法创新来简化流程。在作者提供更多方法和结果的详细信息的情况下, 未来可以改进对完成系统评估和绘图所需工作的预测。* **【翻译: 胡怡思; 审校: 聂永刚】**

关键词: 证据综合, 工作量, 成本, 效率, 投入时间, 文献综述

Introduction

Systematic review (SR) methods were developed in the field of healthcare as a means of collating, appraising, synthesizing, and reconciling broad bodies of primary research (Higgins & Green 2011). A suite of practices are applied that aim to maximize transparency and comprehensiveness and minimize subjectivity and bias (Pullin & Stewart 2006; Haddaway et al. 2015). SRs are now the gold standard of evidence synthesis across healthcare (Higgins & Green 2011), social welfare, education, international development, crime and justice (Shlonsky et al. 2011), and conservation and environmental management (Pullin & Stewart 2006). Nonprofit organizations have been established to develop SR methods and publish and endorse reviews meeting specific minimum standards (e.g., Collaboration for Environmental Evidence [CEE], Campbell Collaboration, Cochrane). Since their establishment, the number of SRs published by such bodies and across the research literature has increased considerably (Haddaway et al. 2015).

SRs involve several methodological steps that ensure the syntheses are reliable (CEE 2013). These steps are publication of a peer-reviewed a priori protocol describing the planned review methods, including detailed information regarding searching, screening, critical appraisal, and data synthesis; comprehensive, tried-and-tested searches across multiple resources of traditional academic publications and gray literature (Haddaway & Bayliss 2015); screening of studies at title, abstract, and full-text levels based on inclusion criteria tested for consistency among reviewers; careful, critical appraisal of all sources of uncertainty and bias (validity) in each study and assessment of the validity of all evidence collectively; consistent extraction of data (descriptive information, metadata, quantitative or qualitative study findings); accurate and reliable synthesis of study findings through appropriate quantitative (e.g., meta-analysis) or qualitative (e.g., meta-ethnography) methods; and fully transparent documentation of all activities to allow verification and repeatability. These are time-consuming and complex tasks, thus SRs are considered particularly resource intensive (Westgate & Lindenmayer 2017). Systematic maps (SMs) are similar to SRs but are used to catalogue an evidence base in a detailed database to identify knowledge gaps and clusters (James et al. 2016).

Although SRs are challenging, published estimates of precisely how long SRs take to complete are largely anecdotal (e.g., Collins et al. 2015). One exception is a study by Borah et al. (2017), who reported the average time from registry date to final report submission date in the PROSPERO database as 67 weeks. There is notable uncertainty in this estimate, however, because dates in PROSPERO do not necessarily reflect the time required to conduct the review, and there is no clear link between the total duration of a SR project and the actual time requirements in person days. No comparable analysis of SR effort has been completed in the environmental field, but results of an assessment of these data for the 86 reviews published by CEE from May 2012 to March 2017 suggest a mean time from protocol to review submission of 737 d (SD 364; range 48–1524 d). At the lower range, this likely represents an impossible speed and probably results from reviewers commencing a project before submitting their final review. At the upper end are cases where projects are known by CEE to have undergone numerous substantial hiatuses. Whatever the reason, this long period has implications for making review results available to the community at the earliest possible opportunity and may hamper evidence-informed policy and practice.

We sought to quantify the time requirements of CEE SRs and SMs by combining data from published SRs and SMs and their protocols and data from a survey of practitioners of environmental SRs. We focused on the CEE because it is a leading authority on the conduct of environmental-evidence syntheses and because CEE reviews should represent best practices in evidence synthesis. We focused on the time needed to conduct each review process, rather than the time needed to coordinate a project as a whole or on estimating the financial cost of completing a review. Although an estimate of the full length of a review project may be interesting, many reviews take additional time without additional financial costs to reviewers, and a long review project may not be an expensive one. Nonetheless, the time requirements of a review can inform decisions regarding budgets and staff availability.

Methods

Assessment of Published CEE SRs and SMs

An assessment of all CEE SRs and SMs published since May 2012 is available in the journal *Environmental Evidence* (https://environmentalevidencejournal.biomedcentral.com/) and the CEE Library (http://www.environmentalevidence.org/completed-reviews). Key metadata were extracted from all completed and in press SRs and SMs as of March 2017. Metadata included protocol and review submission dates; number of databases searched; number of gray-literature resources searched; number of search results identified from database searching; number of duplicates removed; number of titles included after screening; number of abstracts included after screening; number of titles and abstracts included (where screened together); number of full texts retrieved; number of full texts included after screening; number of studies included following critical appraisal (compulsory in a SR, optional in a SM); and number of studies with meta-analyzable data. Data were separated according to whether they came from an SM or SR and summary figures and calculations were undertaken independently for each type of review.

Survey of SR and SM Practitioners

A list of potential respondents ($n = 61$) was assembled from authorship lists of CEE SRs, maps, and protocols published from May 2012 to March 2017; no other relevant expert database exists. The list was supplemented with our personal contacts ($n = 34$). Thirteen email addresses did not work, so alternative authors from these reviews were selected. The final pool was 95 functional email addresses. An invitation to an online survey was emailed to each potential respondent (survey questions and data received are in Supporting Information). The survey was designed, conducted, and reported according to ethical guidelines in Kelley et al. (2003).

Thirty responses were received (32% response rate). Three responses were discarded because of incomplete information, resulting in a total of 27 valid responses. Data from 6 systematic reviewers at 1 organization were collated by their line manager and forwarded. Two separate reminders inviting potential respondents to take the survey were sent. We had a maximum of 33 data points for each question.

Compilation of Data and Calculation of Metrics

Following collation of the data from published articles and survey respondents, data were summarized across replicates (review documents for the assessment of reviews and respondents for the survey) with means and SEs. Information regarding the volume of evidence at each stage of the review process was combined with data on processing speeds to yield mean times taken for each main stage of the review process, along with SEs (details in Supporting Information). Standard errors were propagated for each individual calculation with an online error-propagation tool (Laffers 2008). We built our model of survey effort following the main stages of the review process as outlined by the CEE guidelines on evidence synthesis (CEE 2013). Some data were arbitrarily set where CEE guidance exists (e.g., percentage of titles used as a subset for testing consistency before commencing screening) or where data depend heavily on the experience level and efficiency of the reviewer (e.g., time taken for meta-analysis). Details of the model construction process including SR stages, default values, summary data, and the calculations used to arrive at our conclusions are provided in Supporting Information.

Software for Estimating Effort

Following calculation of summary time and SEs for each stage of the review, we produced an interactive research effort estimation tool that allows users to replace the default data with specific values based on their own experiences or knowledge: PredicTER (Predicting Time requirements for Evidence Reviews). For example, if users know the likely number of search results or title-inclusion rate from scoping exercises, they can enter these data in place of the default values. The tool facilitates transparency by indicating the sources and evidence behind default values through the documentation provided herein; helps users understand the nature of each step in the review process; builds in details and instructions from published guidance on SRs; and is easy to use. The aim of this tool is to provide an indication of the minimum time requirements for an

SR or SM. We hope it will continue to develop as the data set on which it is based expands and the models are refined.

The tool is a web-based app, which is easily updated and refined as more data become available. The app was built in the R statistical environment (R Core Development Team 2017) with the R packages Shiny (Chang et al. 2017) and shinydashboard (Chang 2015) to construct the interactive framework and plotly (Sievert et al. 2017) to draw the diagrams.

The tool has several different types of user input. First, it requires an initial number of articles returned in the search stage. This is typically easy to estimate because it is simply the sum of hits from all databases searched. The tool then combines this total with estimates of the proportion of articles retained at each stage (i.e., title screening, abstract screening, etc.) and the rate at which articles can be processed during those stages. Finally, users can add estimates of the time taken to undertake specific tasks, such as conducting a meta-analysis or writing a report. These data are then combined into plots of the number of articles expected and the total time spent on each review stage.

Our final tool is published here along with detailed explanatory notes to guide users through its use and to ensure that reliable, contextualized data (i.e., through scoping) is provided where possible to increase estimate accuracy. The web app can be used online at http://www.predicter.org or downloaded for use in R with the source code on github (https://github.com/mjwestgate/PredicTER/).

Results

Published CEE SRs

The CEE produced 108 SRs from May 2012 to March 2017: 66 SRs and 20 SMs (86 in total). Thirty-five of these documents (41%) were protocols for incomplete projects. The majority of the data were from SRs, and of these data the majority related to unfinished SRs (review: 47 protocols, 19 reports; map: 8 protocols, 12 maps).

The variability around the mean number of records remaining after each key review stage was large, particularly for points in the review process where data were lacking (e.g., number of studies included at full-text screening [$n = 6$], inclusion rate following critical appraisal [$n = 7$], and number of studies included at meta-analysis [$n = 3$]) (Fig. 1 & Supporting Information). Some reviews could be perceived as outliers, for example, the SR on timing of mowing impacts on biodiversity in meadowland (Humbert et al. 2012) that resulted in a particularly small set of search results ($n = 367$) and a relatively high inclusion rate at title-screening stage (74.0%) and the SM of on-farm water-quality mitigation measures (Randall et al. 2015) that resulted in a very large set of search results ($n > 145{,}000$) and a relatively high percentage of duplicates (49.5%).

There was a lack of consistent reporting in published SRs and SMs. Despite the existence of published standards for the reporting of activities in SRs (e.g., PRISMA; Moher et al. 2009) and requirements for a high level of detail in reporting only 8 of the 32 completed SRs and SMs reported data for all stages of the review process (i.e., searching, duplicate removal, title, abstract, full-text screening, and full-text retrieval).

A typical CEE SR had a mean of almost 12,000 search results, which falls to approximately 8500 unique records following duplicate removal (Fig. 1 & Supporting Information). Approximately 1200 records remained following title screening and around 300 following abstract screening. With the addition of evidence from other sources, the total number of full texts obtained was on average 470. Screening of these full texts left around 100 relevant articles or studies. Critical appraisal retained approximately 75 articles or studies, and suitable data were present in about 45 of them. The average CEE SR, therefore, contained around 100 relevant studies, of which typically 3-quarters passed critical appraisal and approximately one-half were meta-analyzed.

The sample size for SMs ($n = 20$) was much smaller than for SRs ($n = 66$), but the volume of evidence was far greater for maps: almost 35,000 search results were obtained on average, for >22,000 unique records. Title screening returned >4000 relevant records, and abstract screening returned over 1000. Approximately 1200 full texts were retrieved, and over 400 were relevant at full-text screening. Where critical appraisal was performed for an SM, on average about 115 studies were retained in the final map (Fig. 1).

Survey of SR and SM Practitioners

Of the 33 included responses, 7 provided data for all 15 questions regarding respondents' experience with reviews. A further 12 provided data for the stages up to data or metadata extraction and beyond. On average, respondents had conducted a median of 2 SRs (range 0–18). Only 1 respondent had not previously conducted a review. Data from this respondent were relevant to full-text retrieval alone because the respondent had acted as an assistant for a larger group of reviewers. There were no clear patterns in the relationship between experience and variables relating to speed of review conduct (Supporting Information). We received fewer responses about later stages of the review than early stages (Table 1) and particularly few responses about the time taken to complete quantitative synthesis (effect-size calculation and meta-analysis; $n = 7$ and 8, respectively).

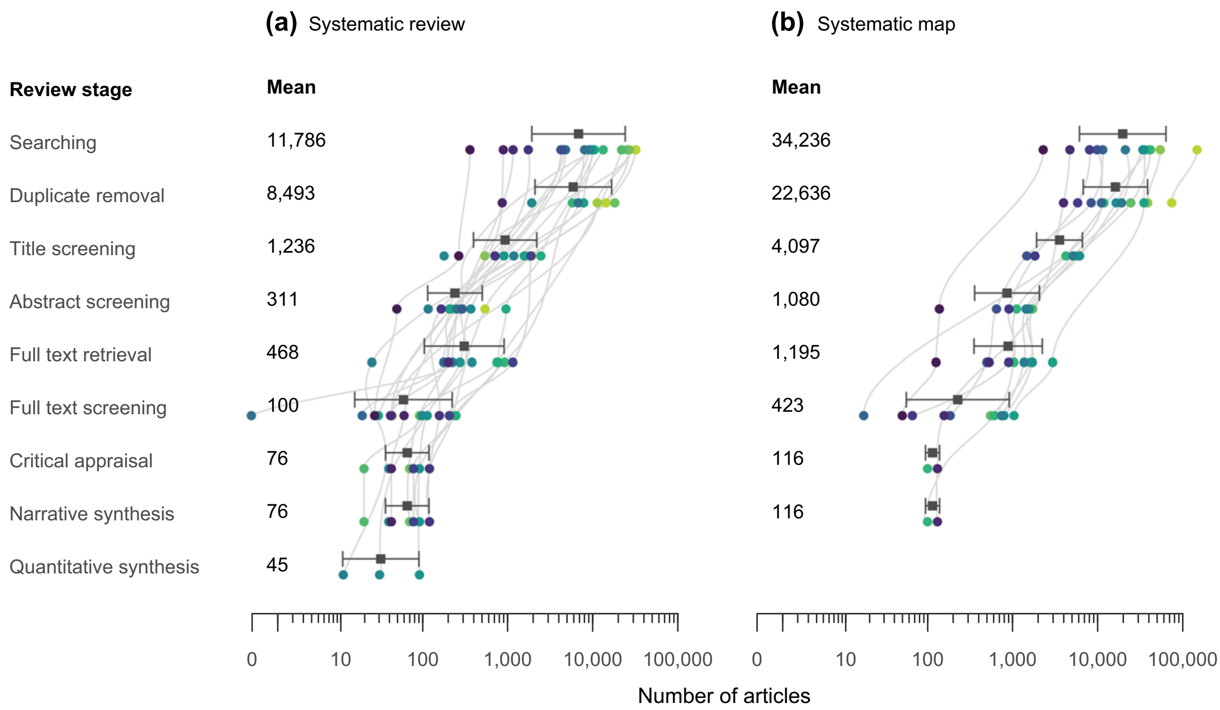

Figure 1. Number of articles remaining after each key stage of (a) systematic review and (b) systematic map processes (lines connecting dots, individual reviews; error bars, SD 1).

Table 1. Summary data of responses to the survey of experienced systematic reviewers.

Survey question	*Mean response*	*n*	*SD*
How many systematic reviews have you undertaken?	4.30	33	5.33
Time taken to download search results from each database (d)	0.25	20	0.22
Time taken to assemble library of results and remove duplicates (d)	1.37	18	1.40
Time taken to screen organization websites (each in days)	0.15	21	0.12
Number of titles screenable per day	854.35	23	533.62
Number of abstracts screenable per day	192.29	24	111.90
Number of titles and abstracts screenable per day (together)	468.14	22	128.22
Number of full texts retrievable per day	170.94	24	137.37
Number of full texts screenable per day	43.99	30	31.01
Number of articles for meta-data extraction/coding per day	16.69	21	11.57
Number of articles for critical appraisal per day	11.68	19	8.15
Number of articles for data extraction per day	6.87	19	5.09
Number of articles for effect size calculation per day	24.00	7	34.08
Time taken for meta-analysis (d)	6.75	8	5.09
Time taken for report writing (d)	15.53	20	10.23
Percentage of time required for administration	19.00	22	12.28

Model Outputs and Estimated Effort

The time taken for each stage in an SR was lower, on average, than for an SM (Fig. 2). The total time estimated for an average systematic was 164 person days at 1 full time equivalent (1.0 FTE) (SD 23), and the total time for an average SM was 211 person days (SD 53) when the optional critical appraisal step was excluded and 254 person days (SD 67) including critical appraisal. This estimate included a large amount of time allotted to planning and administration. In an effort to be conservative, we calculated an average percentage from our survey of reviewers and applied this to the total mean time from the models, resulting in 46 person days for SRs and 54 person days for SMs (excluding critical appraisal). In the PredicTER tool, stages calculated by the model include those from searching to effect-size calculation; other stages are set as arbitrary defaults that must be changed by the user. For these stages calculated in the model, the most time-consuming were title screening, full-text screening, and critical appraisal. Metadata and data extraction also required considerable time. Searching, assembling a library of evidence, full-text retrieval, and consistency checking required less time than most other stages. The uncertainty

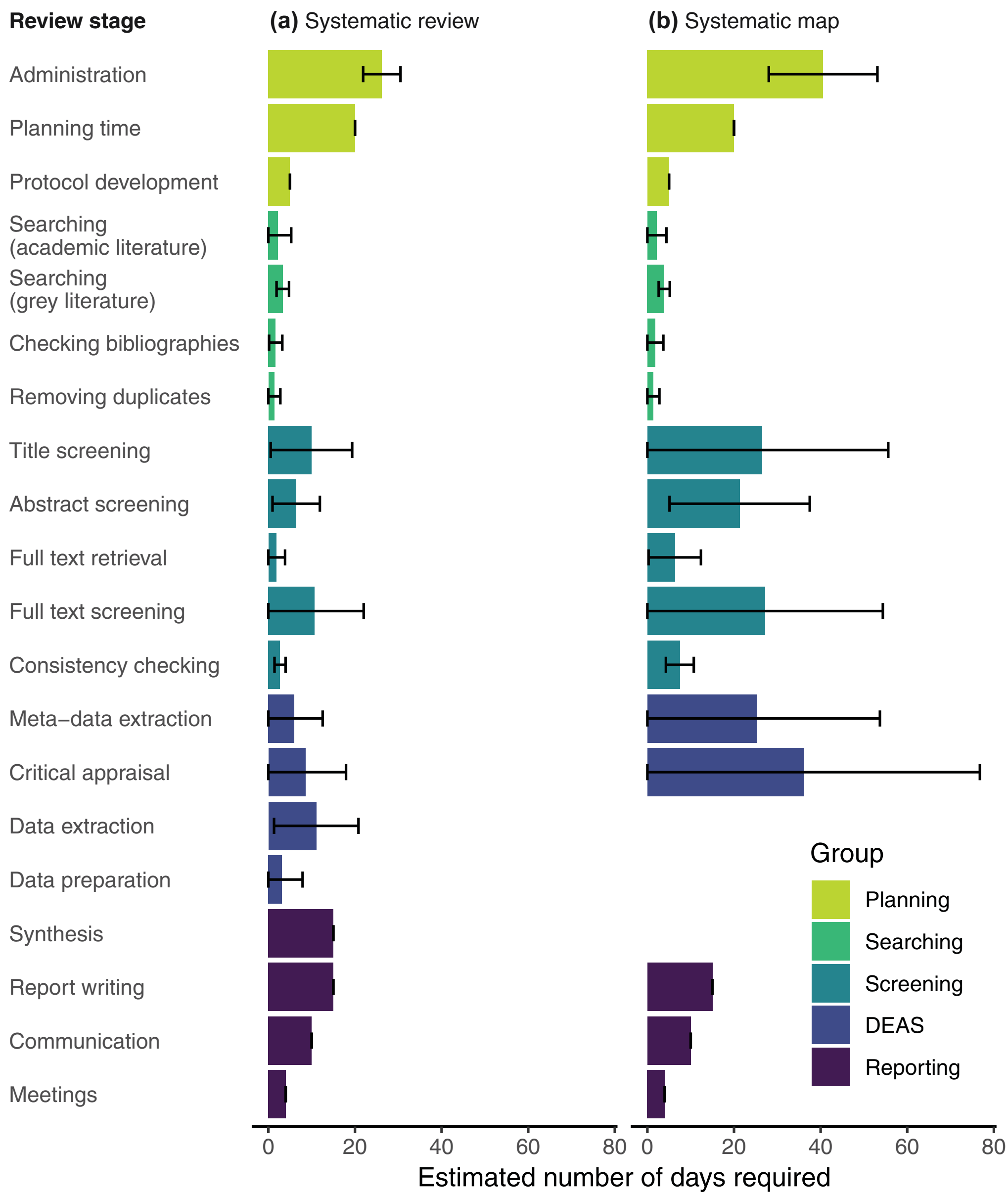

Figure 2. Estimated time taken for each stage of the systematic process for (a) systematic reviews and (b) systematic maps. Data are from a model combining data from a survey of experienced reviewers and published reviews or maps (error bars, SD 1; bars grouped into similar stages; DEAS, data extraction, appraisal, and synthesis).

around these data is substantial and resulted from the propagation of errors across the models and the variability in the underlying source data.

Discussion

The number of articles included in synthesis projects and the total time taken to complete them varied substantially, likely the result of broad diversity in the topics reviewed. However, there were key bottlenecks during early screening and at critical-appraisal and data-extraction stages of each review.

Emergent Patterns

Several trends did not match our expectations about which stages of the review would take the most time. A relatively small proportion of time was spent on searching and assembling a library of results: 7.0–7.4 d for reviews and maps, respectively. This result may reflect detailed preparation, given that searching should be preceded by in-depth building and testing of search strategies outlined in a priori protocols. Although we did not explicitly ask expert reviewers how long they spent designing and testing a search strategy, this part of the review process requires careful planning to ensure the review results are comprehensive and representative of an evidence base (Bayliss & Beyer 2015; Livoreil et al. 2017).

Also unexpected was our finding that respondents' reported time spent on administration was particularly large: on average 19% of their total time (26 d for reviews; 34 d for maps excluding critical appraisal; and 41 d for maps with critical appraisal). Reported administration time varied substantially (SD 12.3), perhaps indicating discrepancies in respondents' definitions of what should be included. However, this likely reflected that systematic reviewing often requires time spent coordinating a large, possibly international team and may also require substantial learning or relearning of particular skills, such as experimental design or statistics. We have not factored in training time in our analysis, but this is worth considering for new teams or those that will rely heavily on group tasks based on subject but not methodological expertise.

More expected was the large amount of time spent on screening (including retrieval, consistency checking, and bibliographic checking): an average of 33 and 91 d for SRs and SMs, respectively. This was a large proportion of the time budget (20% and 45% for reviews and maps, respectively; 36% for SMs that include critical appraisal). These differences showed that resources were predominantly shifted toward identifying evidence in maps, whereas far more time was devoted to synthesis in reviews. In reviews, a similar amount of time was spent on extracting and analyzing the data as on screening data (35 d). For maps, however, the proportion of total time spent on extracting metadata and coding was relatively lower (25 d). Experience did not appear to improve efficiency, but our sample size for all variables was small, so the power of these correlations is low.

Implications for Optional Activities

We were able to estimate the impacts of optional activities on the total time requirements of a SR or SM. Current CEE guidance suggests a subset of articles be checked for consistency in the application of inclusion criteria between 2 reviewers prior to commencing screening (CEE 2013), and it suggests 10% of records be checked at minimum. However, in healthcare SRs, dual coding is commonly used to reduce subjectivity (e.g., Jones et al. 2016). By altering the level of consistency checking from the recommended minimum of 10% at each stage to 100% (i.e., complete dual screening), the total time required for a review changed from 164 to 193 d, an increase of 18%. Although regarded by some as best practice for SRs (Higgins & Green 2011), this increase in time is substantial and may prove too costly. However, it may be an important concession to maximize reliability and minimize human error. Similarly, by reducing consistency checking from 10% to 0%, the total time needed for a review is reduced by only 3 d from 161 person days. Thus, PredicTER can be used to justify important steps that may not have a significant impact on resource requirements.

Table 2. Previous estimates of the resource requirements for systematic reviews from a non-systematic search of the literature.

Financial cost	*Time requirement*	*Reference*
	0.5–3 years	Dicks et al. 2014
US$30,000–300,000	several years	CEE 2013
£80,000–120,000 GBP	10–18 months	Collins et al. 2015
≤US$250,000		McGowan & Sampson 2005
	9–24 months	CCACE 2013
	67.3 weeks	Borah et al. 2017

Similarly, CEE guidance suggests a selection of review bibliographies be screened to maximize comprehensiveness of the search (CEE 2013). Increasing this bibliographic checking, or citation chasing, can require considerable time if, for example, all identified reviews are screened in this way or even if all articles' bibliographies are screened. Assuming the inclusion rate at title, abstract, and full text (and retrieval rate) remain the same in bibliographic checking as for the core of the review, one can readily predict the additional time needed to screen a certain number of reviews or articles in this way. For an SM, a larger volume of reviews is likely to be found, and the user can specify this number. For example, in an SM of the impacts of vegetated strips within and around fields (Haddaway et al. 2016*b*), around 100 review bibliographies were checked for additional potentially relevant articles. Altering the number of bibliographies checked in our tool to 100 increased the time requirement for maps excluding critical appraisal from 211 to 230 d (9%).

Comparison with Existing Estimates

Previous estimates of the resource requirements of SRs are imprecise and vary substantially from 6–24 months or several years (Table 2). Anecdotally, we have heard estimates that are as long as 5 years by a leading institute that produces SRs in healthcare in Sweden (http://www.sbu.se/en/). Our analyses show that the time for an average CEE-style SR takes only 164 person days (SD 23). This estimate represents about 1 year FTE, including vacations, public holidays, and other regular disruptions. Therefore, we found a resource requirement in the lower end of the rough estimates provided in the literature. Our estimate is under half that of the only other evidence-based assessment of which we are aware, which is approximately 337 d (Borah et al. 2017). The time estimated by Borah et al. (2017) and the other time estimates in the literature are typically meant to reflect the total time it would take for an SR project to be completed, rather than the resource requirements. This compares with the 737 d (SD 364) needed to complete a CEE SR or SM we identified based on assessment of protocol and review submission dates in the *Environmental Evidence* (Supporting Information). The 337-d estimate does not represent time requirements in person days, and reflects our experience that few systematic reviewers conduct their reviews at 1.0 FTE.

We did not aim for full costing of a SR. However, our findings provide a greater understanding of the financial resources needed for an SR or SM by facilitating conversion of time requirements into local salary costs. The average total salary costs for a postdoctoral research at Bangor University (chosen arbitrarily due to our knowledge of the university, including insurance and pension contributions) for 12 months is £48,593 (https://www.bangor.ac.uk/finance/py/documents/pay-scales-en.pdf). After including other costs, such as support-staff time, travel, meeting attendance, software, and access to databases and articles, this sum is unlikely to rise above £100,000. This value is below the midpoint for the roughly estimated cost ranges in the literature. Users of PredicTER can convert these time requirements into their own local salary costs but should appropriately budget for other costs.

Estimates produced by our tool are realistic relative to the Swedish environmental reviews being conducted by Mistra EviEM (www.eviem.se/en). With approximately 20 years of 1.0 FTE (review project managers) and 2.2 years per review, the project aims to complete 17 SRs and SMs over 6.5 years. Lead staff are contributing approximately 0.3 to 0.4 FTE, exactly in-line with our estimates. However, our estimates for SMs are somewhat higher than those indicated for EviEM maps. This is almost certainly the result of a small and heterogeneous evidence base for completed SMs: fewer SMs have been completed to date and the variability around the volume of evidence is substantial (SD for SR total search results is 11,786 records, whereas it is 39,434 for SMs). SMs are more adaptable by definition (Haddaway et al. 2016*a*; James et al. 2016), but having a larger number of SMs to study would increase the precision of the data in our tool.

The time lag between protocol submission and review report submission to *Environmental Evidence* was on average 737 d (SD 364). This is review-completion time and includes hiatuses and work conducted below 1.0 FTE. Our results reflect staff time requirements, and as such are more useful for budget estimation than timeframe estimates because many external factors affect the timing of review projects. The much longer times suggest reviews could perhaps be conducted faster if resources were used allocated differently, for example, employing multiple people simultaneously at key stages that require great amounts of time. These times also demonstrate that it can take considerable time to provide decision makers with synthesis outputs, which could reduce the usability and impact of the projects.

Analysis and Evidence-Base Limitations

We used the best available information on the number of articles and the amount of time associated with a typical environmental SR or SM. However, a number of factors may adversely affect the reliability of our calculations.

We assumed a quantitative synthesis is performed, which may be true for the majority of current SRs in environmental fields. Qualitative synthesis is a valuable method (Flemming 2007), and its use will likely increase in CEE reviews. However, qualitative SRs often focus on other cornerstones of rigor that quantitative reviews ignore. For example, qualitative syntheses may stop screening after a certain point because of information saturation (Dixon-Woods et al. 2005) (i.e., no new themes, concepts, or theories are identified after a certain number of articles have been read). These reviews should be dealt with differently when performing an analysis relating to time requirements, and tools for predicting times should be built specifically for them. Specific qualitative reviews may be able to adapt our tools to fit the desired methods.

All our data had a high level of variability, low levels of reporting, or both. This resulted from a highly heterogeneous evidence base and a relatively small sample size. Thus, a detailed investigation of differences in time requirements for different subtopics was not possible. Limited reporting precluded detailed analyses. Of the 19 completed SRs, only 8 reported the number of duplicates removed from total search results. Similarly, 18 reviews reported the total number of included articles, but only 10–11 articles reported the number of articles following title screening, abstract screening, and full-text retrieval. Future CEE reviews should strive to report such methodological information consistently. To that effect, CEE and *Environmental Evidence* now enforce reporting standards for all published review and map protocols and reports in the form of ROSES forms and flow diagrams (see www.roses-reporting.com). Future analyses should increase sample size and refine the model and its estimations based on new evidence. Efforts to record descriptive summary information regarding SR methods are underway (e.g., ROSES; Haddaway et al. 2018).

We could not provide evidence-based data for all parts of our analysis. A number of key variables were estimated based on personal experience, including time required for additional searching and number of bibliographies screened. Where possible, future analyses should examine the evidence base for these data.

Particular circumstances would affect the reliability of predictions made with our tool. For example, a change in core staff midway through a project would likely require a substantial proportion of time to acquaint a new person with what has been accomplished. However, careful file management and clear record keeping could reduce this. Large review teams may require more resources to train and manage, particularly if meeting remotely. Novice teams may require substantial training time and may be inefficient in earlier review stages. Finally, undertaking reviews over an extended period can result in particularly low efficiency if core staff must reacquaint themselves with their own work after significant gaps.

Our tool allows estimation of the time required to complete an SR or SM, and our analyses of the evidence base provide useful default values should any information related to the likely volume or nature of the evidence that might be encountered during a review be unknown. These default values, however, are based on an average SR or SM. Heterogeneity across CEE reviews means this average review, although helpful as a starting point, is perhaps not meaningful. Context is highly important for each review, and knowing something about the volume or the nature of the evidence (e.g., proportional relevance of a subset) allows users to estimate time requirements more accurately. One should not assume that all reviews are alike and that times we calculated are a reliable estimate alone when planning a review. We encourage users to undertake reliable scoping, as suggested in the CEE Guidelines (CEE 2013), to provide reliable predictions of the volume of evidence, proportional relevance of articles, and time required by the team to undertake specific tasks.

We calculated mean volumes of evidence at each stage of the review process and used inclusion rates and working speeds to calculate an independent mean time requirement for each stage based on available evidence. However, many reviews do not report all data for each review stage, and the results of 1 stage depend on the nature of the preceding stages. If we would have had complete data from all reviews, we would have been able to model time requirements based on contextual variables, for example, the inclusion rate of the preceding stage. This was not possible with our limited data set, however.

Suggestions for Future Work

Increasing the number of data points in future analyses itself would be aided by better reporting of methods used and records found at all stages of the review process in CEE reviews. Some efforts are underway to record these data more consistently (e.g., ROSES; Haddaway et al. 2018).

Although there will be considerable local and regional variability in the real-world prices of the services required to conduct SRs and SMs (see above), an itemized list of recommended activities is a vital starting point for those planning a review. Furthermore, the procedures we identified as the most time-consuming should be seen as important areas for methodological and technological development to increase efficiency (e.g., using machine learning [Westgate et al. 2018]).

We also suggest research be undertaken to improve understanding of why SRs and SMs can take so long to complete (i.e., mean review report time of 737 d). There is a need for qualitative research into the reasons behind the timing of review activities in cases where time taken to complete the review project is different from the total number of person days needed for the tasks involved. This would also highlight practices that increase efficiency, and outputs would be of great use to those commissioning evidence syntheses for direct use in decision making. Finally, results from our analyses and predictions using our tool should be continually tested and the tool refined to match developments in SR methods (e.g., machine learning and prioritised screening) (Shekelle et al. 2017).

Acknowledgment

We thank S. Johansson and Mistra EviEM for contributing time to the completion of this manuscript.

Supporting Information

The email survey (Appendix S1), data and calculations used to arrive at metrics and standard errors (Appendix S2), and additional methods and results (Appendix S3) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

Bayliss HR, Beyer FR. 2015. Information retrieval for ecological syntheses. Research Synthesis Methods **6**:136–148.

Borah R, Brown AW, Capers PL, Kaiser KA. 2017. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open **7**:e012545.

CEE (Collaboration for Environmental Evidence). 2013. Guidelines for systematic review and evidence synthesis in environmental management. Version 4.2. Environmental Evidence. Available from http://environmentalevidence.org/wp-content/uploads/2014/06/Review-guidelinesversion-4.2-finalPRINT.pdf (accessed January 2018).

CCACE (Centre for Cognitive Ageing and Cognitive Epidemiology). 2013. Systematic reviews and meta-analyses: a step-by-step guide. Available from http://www.ccace.ed.ac.uk/research/software-resources/systematic-reviews-and-meta-analyses (accessed September 2018).

Chang W. 2015. shinydashboard: create dashboards with 'Shiny'. R package version 0.5 **1**. Available from https://cran.r-project.org/web/packages/shinydashboard/index.html (accessed January 2018).

Chang W, Cheng J, Allaire J, Xie Y, McPherson J. 2017. Shiny: web application framework for R. R package version 1.0.3. Available from https://cran.r-project.org/web/packages/shiny/index.html (accessed January 2018).

Collins A, Coughlin D, Miller J, Kirk S. 2015. The production of quick scoping reviews and rapid evidence assessments: a how to guide. Joint Water Evidence Group, United Kingdom.

Dicks LV, Walsh JC, Sutherland WJ. 2014. Organising evidence for environmental management decisions: a '4S' hierarchy. Trends in Ecology & Evolution **29:**607–613.

Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. 2005. Synthesising qualitative and quantitative evidence: a review of possible methods. Journal of Health Services Research & Policy **10:** 45–53.

Flemming K. 2007. The synthesis of qualitative research and evidence-based nursing. Evidence-based Nursing **10:**68–71.

Haddaway N, Woodcock P, Macura B, Collins A. 2015. Making literature reviews more reliable through application of lessons from systematic reviews. Conservation Biology **29:**1596–1605.

Haddaway NR, Bayliss HR. 2015. Shades of grey: two forms of grey literature important for reviews in conservation. Biological Conservation **191:**827–829.

Haddaway NR, Bernes C, Jonsson B-G, Hedlund K. 2016*a*. The benefits of systematic mapping to evidence-based environmental management. Ambio **45:**613–620.

Haddaway NR, Brown C, Eggers S, Josefsson J, Kronvang B, Randall N, Uusi-Kämppä J. 2016*b*. The multifunctional roles of vegetated strips around and within agricultural fields. A systematic map protocol. Environmental Evidence **5:**18.

Haddaway NR, Macura B, Whaley P, Pullin AS. 2018. ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. Environmental Evidence **7:**7.

Higgins JP, Green S. 2011. Cochrane handbook for systematic reviews of interventions. John Wiley & Sons, Hoboken, New Jersey.

Humbert J-Y, Pellet J, Buri P, Arlettaz R. 2012. Does delaying the first mowing date benefit biodiversity in meadowland? Environmental Evidence **1:**9.

James KL, Randall NP, Haddaway NR. 2016. A methodology for systematic mapping in environmental sciences. Environmental Evidence **5:**7.

Jones E, Taylor B, MacArthur C, Pritchett R, Cummins C. 2016. The effect of early postnatal discharge from hospital for women and infants: a systematic review protocol. Systematic Reviews **5:**24.

Kelley K, et al. 2003. Good practice in the conduct and reporting of survey research. International Journal for Quality in Health Care **15:**261–266.

Laffers R. 2008. Error propogation calculator. Available from https://www.eoas.ubc.ca/courses/eosc252/error-propagation-calculator-fj.htm (accessed September 2018).

Livoreil B, Glanville J, Haddaway NR, Bayliss H, Bethel A, Lachapelle FF, Robalino S, Savilaakso S, Zhou W, Petrokofsky G. 2017. Systematic searching for environmental evidence using multiple tools and sources. Environmental Evidence **6:**23.

McGowan J, Sampson M. 2005. Systematic reviews need systematic searchers. Journal of the Medical Library Association **93:**74.

Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLOS Medicine **6:**e1000097.

Pullin AS, Stewart GB. 2006. Guidelines for systematic review in conservation and environmental management. Conservation Biology **20:**1647–1656.

R Core Development Team. 2017. R: a language and environment for statistical computing, Version 3.4.1. R Foundation for Statistical Computing, Vienna, Austria.

Randall NP, Donnison LM, Lewis PJ, James KL. 2015. How effective are on-farm mitigation measures for delivering an improved water environment? A systematic map. Environmental Evidence **4:**18.

Shekelle PG, Shetty K, Newberry S, Maglione M, Motala A. 2017. Machine learning versus standard techniques for updating searches for systematic reviews: a diagnostic accuracy study. Annals of Internal Medicine **167:**213–215.

Shlonsky A, Noonan E, Littell JH, Montgomery P. 2011. The role of systematic reviews and the Campbell Collaboration in the realization of evidence-informed practice. Clinical Social Work Journal **39:**362–368.

Sievert C, Parmer C, Hocking T, Chamberlain S, Ram K, Corvellec M, Despouy P. 2017. plotly: create interactive web graphics via plotly.js. R package version 4.7.1. Available from https://CRAN.R-project.org/package=plotly (accessed January 2018).

Westgate MJ, Haddaway NR, Cheng SH, McIntosh EJ, Marshall C, Lindenmayer DB. 2018. Software support for environmental evidence synthesis. Nature Ecology & Evolution **2:**588–590.

Westgate MJ, Lindenmayer DB. 2017. The difficulties of systematic reviews. Conservation Biology **31:**1002–1007.