

1 **Predicting the time needed to conduct an environmental systematic review or systematic**
2 **map: analysis and decision support tool**

3 Neal R Haddaway^{1*}

4 Martin J Westgate²

5

6 1 – Mistra EviEM, Stockholm Environment Institute, Linnégatan 87D, Stockholm Sweden

7 2 – Fenner School of Environment and Society, The Australian National University, Acton
8 ACT 2601, Australia

9 * Corresponding author: neal_haddaway@hotmail.com

10

11 **Acknowledgements**

12 The authors thank Sif Johansson and Mistra EviEM for contributing time for the completion
13 of this manuscript.

14

15 **Abstract**

16 Systematic reviews aim to maximise transparency and comprehensiveness, whilst also
17 minimising subjectivity and sources of bias. Because of these time-consuming and complex
18 tasks, systematic reviews are perceived as being resource-intensive. To date, published
19 estimates of systematic review resource requirements have been largely anecdotal, being
20 imprecise and not based on evidence. However, it is valuable to provide reliable means of
21 estimating the resource and time requirements of systematic reviews and maps. We analysed
22 all CEE systematic reviews (n=66) and maps (n=20) published or registered between 2012
23 and 2017 to estimate the average time needed to complete a systematic review and map. We
24 then surveyed 33 experienced systematic reviewers to collate information on time needed for
25 each stage of the review process. Our results show that the average CEE systematic review
26 takes 157 days (SD; ± 22), whilst the average CEE systematic map takes 209 days (SD; ± 53).
27 While screening of titles and abstracts is widely accepted to be time-consuming, in practice
28 meta-data extraction and critical appraisal can take as long (or even longer) to complete,
29 especially when producing systematic maps. Finally, we present a tool that allows the user to
30 predict the time requirements of a review or map given information known about the planned
31 methods and evidence base likely to be identified. Our tool uses evidence-based defaults as a
32 useful starting point for those wishing to predict the time requirements for a particular
33 review. Our analyses shed light on the most time-consuming stages of the systematic review
34 and map process, and highlight key bottlenecks from the perspective of time requirements,
35 helping future reviewers to plan their time accordingly. Future predictions of effort required
36 to complete systematic reviews and maps could be improved if CEE and CEE review authors
37 provided more detailed reporting of the methods and results of their reviewing processes.

38

39 **Keywords:** evidence synthesis; workload; cost; efficiency; time commitment; literature
40 review

41 **Introduction**

42 Systematic review methods were developed in the field of healthcare in the 1990s as a means
43 of collating, appraising and synthesising broad (and sometimes contradictory) bodies of
44 primary research studies [1]. The methods revolve around a suite of practices during the
45 conduct of a literature review that aim to maximise transparency and comprehensiveness,
46 whilst also minimising subjectivity and sources of bias [2, 3]. Systematic reviews are now
47 viewed as a ‘gold standard’ in evidence synthesis across not only healthcare [1], but also
48 social welfare, education, international development, crime and justice [4], and conservation
49 and environmental management [2]. Within these fields, not-for-profit organisations have
50 been established to govern standards in systematic review and publish and endorse reviews
51 that meet specific minimum standards (e.g. the Collaboration for Environmental Evidence,
52 the Campbell Collaboration and Cochrane). Since their inception and development, the
53 number of systematic reviews published by these coordinating bodies and more broadly
54 across the research literature has increased considerably [3].

55 Systematic reviews should involve a number of important methodological steps to ensure the
56 syntheses are reliable [5]. These include: 1) the publication of a peer-reviewed *a priori*
57 protocol that sets out the planned methodology for the review, including detailed information
58 regarding the search, screening, critical appraisal and data synthesis strategies; 2)
59 comprehensive, tried-and-tested searches across a suite of resources for both traditional
60 academic research studies and grey literature [6]; 3) screening of identified studies at title,
61 abstract and full text levels using inclusion criteria that have been trialled and tested for
62 consistency amongst reviewers; 4) considered critical appraisal of all sources of uncertainty
63 and bias (validity) in each study, along with an assessment of the validity of all evidence
64 collectively; 5) consistent extraction of data (both descriptive information, or meta-data, and
65 quantitative or qualitative study findings); 6) accurate and reliable synthesis of study findings
66 through appropriate quantitative (e.g. meta-analysis) or qualitative (e.g. meta-ethnography)
67 methods; 7) throughout the process, full transparent documentation of all activities to allow
68 verification and repeatability. Because of these time-consuming and complex tasks,
69 systematic reviews are widely perceived as being particularly resource-intensive [7].

70 Although it is accepted that systematic reviews are challenging, the published estimates of
71 the resource requirements of systematic reviews have been largely anecdotal, and as such are
72 both imprecise and highly variable [e.g. 8]. One exception was a recent study by Borah et al.

73 [9] who found that the average time from the date of registry to date of submission of final
74 reports in the PROSPERO database was 67 weeks. There is notable uncertainty in this
75 estimate, however, since the dates held by the PROSPERO database do not necessarily
76 closely relate to the dates work commenced. Nor is there a clear link between the total
77 duration of a systematic review and the actual time requirements in person-days. No
78 comparable analysis of systematic review effort has been completed in the field of
79 conservation and environmental management; but any such estimate based on protocol and
80 final review report submission dates for the CEE journal *Environmental Evidence* (i.e.
81 duplicating the approach of Borah et al.) is unlikely to be reliable. Indeed, an assessment of
82 this data for the 86 reviews published by CEE between May 2012 and March 2017 suggest a
83 mean time from protocol to review submission of 737 days ($SD=\pm 364$) with a range of 48 to
84 1,524 days. At the lower range, this represents an impossible speed for review conduct, and at
85 the upper end we know these data represent projects that underwent numerous significant
86 hiatuses.

87 Since systematic reviews are known to be resource-intensive, and since current estimates of
88 their time requirements are largely based on anecdote or uncertain data linked to reviews in
89 other fields, there is a clear need to provide evidence-based estimates of the time needed to
90 conduct a systematic review. Here, we report the results of a project that aimed to collate data
91 from a variety of sources and summarise the time requirements of CEE systematic reviews.
92 We use a combination of data reported within published systematic reviews and protocols
93 along with data from a survey of systematic review practitioners in the environmental field.
94 We produce an estimate of the mean time required to conduct a CEE systematic review or
95 systematic map split by the key steps of the review process. We also describe a tool based on
96 this data that allows those planning a systematic review or map to predict the time needed for
97 their review based on their own scoping activities that reveal the likely volume of relevant
98 evidence and the working speeds of their team. To our knowledge this is the first evidence-
99 based tool for predicting workloads in a systematic review and has a broad applicability
100 across a range of disciplines.

101 **Methods**

102 *Assessment of published CEE SRs/SMs*

103 An assessment was conducted of all CEE systematic reviews published since May 2012 in
104 both the journal *Environmental Evidence*
105 (<https://environmentalevidencejournal.biomedcentral.com/>) and the CEE Library
106 (<http://www.environmentalevidence.org/completed-reviews>). Key meta-data were extracted
107 from all completed systematic reviews and systematic maps, along with systematic review
108 and map protocols where no final review report had yet been published at the time of analysis
109 (March 2017). This meta-data included the following: protocol submission date and review
110 submission date (for all completed reviews and maps); the number of databases searched; the
111 number of grey literature resources searched; the number of search results identified from
112 database searching; the number of duplicates removed; the number of titles included after
113 screening; the number of abstracts included after screening; the number of titles and abstracts
114 included (where screened together); the number of full texts retrieved; the number of full
115 texts included after screening; the number of studies included following critical appraisal;
116 and, the number of studies with meta-analysable data. Data were separated according to
117 whether they came from systematic maps or systematic reviews and summary figures and
118 calculations were undertaken independently for these two types of review.

119

120 *Survey of systematic review practitioners*

121 A list of potential respondents (n=61) was assembled from authorship lists of CEE systematic
122 reviews, maps and protocols published between May 2012 and March 2017. The list was
123 supplemented with personal contacts from the systematic review community (n=34). A total
124 of 12 email addresses were no longer functioning and alternative authors from each EEJ
125 publication were selected as target respondents. One further email address failed to work and
126 so a tertiary alternative was selected and emailed. In total 95 respondents were targeted using
127 functioning email addresses. An email invitation to an online survey was sent to each
128 potential respondent (see Additional File 1 for survey questions). The key data collected are
129 outlined in Table 1.

130

131 Some 30 responses were received through the online survey, yielding a response rate of 32%.
132 Three responses were discarded because of incomplete information (only one page of
133 responses received), resulting in a total of 27 valid responses. In addition, data from 6
134 systematic reviewers at one organisation in Canada were collated by their line manager and
135 forwarded for use in the analysis. This resulted in a maximum of 33 data points or each
136 question.

137

138 *Compilation of data and calculation of metrics*

139 Following collation of the data from published articles and survey respondents, data were
140 summarised using means and standard errors. Data were then transformed into the same units
141 and information regarding the volume of evidence at each stage of the review process were
142 combined with data on processing speeds to yield a set of summary data on the mean time
143 taken for each main stage of the review process, along with a standard error. Standard errors
144 were propagated for each individual calculation using an online error propagation tool
145 (<https://www.eoas.ubc.ca/courses/eosc252/error-propagation-calculator-fj.htm>). The main stages
146 of the review process were identified as outlined in Table 2. These stages are based on the
147 CEE guidelines in systematic review [5]. Some data were arbitrarily set where CEE guidance
148 exists (e.g. the percentage of titles used as a subset for testing consistency before
149 commencing screening) or where data depend heavily on the experience level and efficiency
150 of the reviewer (e.g. time taken for meta-analysis). Details of the sources of each line of data
151 used in the calculation of times is provided in Table 3. Default values, the summary data and
152 the calculations used to arrive at the metrics in Table 2 are provided in Additional File 2.

153

154 *A software tool for estimating effort in future reviews*

155 Following calculation of summary time metrics for each stage of the review process (and
156 propagated standard errors), an interactive research effort estimation tool was produced that
157 builds on our framework by allowing end users to replace the default or mean data with
158 specific values based on their own experiences or knowledge. Key requirements for the tool
159 were: transparency in indicating the sources and evidence behind default values through
160 methodological documentation provided herein; helping the user understand the nature of
161 each step in the review process; building in details and instructions from published guidance

162 on systematic reviews; and, ease of use. The aim of this tool is to provide an indication of the
163 minimum time requirements for a systematic review. It is hoped that this tool will continue to
164 develop as the dataset upon which it is based expands and the models are refined. The
165 structure that provides caveats for missed steps is therefore intended as a conservative
166 warning where it is not based on evidence, and is informed by existing published systematic
167 review methodology and guidance.

168 The tool is provided in two formats. One format is based in an Excel spreadsheet, since this
169 format is downloadable and readily usable. The second format is a web-based app, which is
170 more easily updated and refined. The app was built in the R statistical environment [10] using
171 the R packages Shiny [11] and shinydashboard [12] to construct the interactive framework,
172 and plotly [13] to draw the diagrams. Both software tools make identical calculations, and
173 return identical results.

174 The software tool utilises several different types of user input. First, it requires an initial
175 number of articles that are returned by the 'search' stage of the systematic review or
176 systematic map. This is typically easy to estimate, as it is simply the sum of the number of
177 hits from all databases searched during the review. The software tool then combines this total
178 number of articles with estimates of the proportion of articles retained at each stage (i.e., title
179 screening, abstract screening, etc.), and the rate at which articles can be processed during
180 those stages. Typically, the proportion of articles retained increase as the review progresses,
181 while the number processed per day decreases. Finally, the user can add estimates of the time
182 taken to undertake specific tasks within the review process, such as conducting a meta-
183 analysis or writing a report. These data are then combined into plots of the number of articles
184 expected, and the total time spent, on each review stage.

185 The tool underwent substantial revisions and alterations during our analyses. Over time, we
186 developed an increasing level of detail to reflect the variability and nuance across the suite of
187 activities that make up a systematic review or map. Our final tool is published here along
188 with detailed explanatory notes to guide the user through its use and to ensure that reliable,
189 contextualised data (i.e. through scoping) is provided where possible to increase the accuracy
190 of predictions. The excel version of the app is available in the supplementary information
191 (Additional File 3), while the web app can be used online at
192 <https://mjwestgate.shinyapps.io/revtime/> or downloaded for use in R using the source code on
193 github (<https://github.com/mjwestgate/revtime/>).

194 **Results**

195 *Assessment of published CEE SRs*

196 A total of 108 systematic review publications were produced by CEE between May 2012 and
197 March 2017, of which 66 represented systematic reviews and 20 were systematic maps (86 in
198 total), though 35 of these documents (41%) were protocols for as-yet incomplete projects.
199 The majority of the data comes from systematic reviews, and of these data the majority relate
200 to as yet unfinished systematic reviews (Figure 1).

201 The mean number of records remaining after each key stage in the conduct of a systematic
202 review is outlined in Figure 2 and Table 4. The variability around the data is clearly large,
203 particularly for some points in the review process where data were lacking (e.g. meta-
204 analysis, n=3). Some notable reviews could be perceived as outliers: for example, the
205 systematic review on the timing of mowing impacts on biodiversity in meadowland that
206 resulted in a particularly small set of search results (n=367) and a relatively high inclusion
207 rate at title screening stage (74.0%) [14], and the systematic map of on-farm water quality
208 mitigation measures that resulted in a very large set of search results (n>145,000) and a
209 relatively high percentage of duplicates (49.5%) [15]. However, given the low sample size
210 these cases have been left in to reflect the real variability present.

211 It is also worth noting that there is a lack of consistent reporting in published systematic
212 reviews and maps. Despite the existence of published standards for the reporting of activities
213 in systematic reviews (e.g. PRISMA; [16]) and requirements for a high level of detail in
214 reporting within the journal *Environmental Evidence*, only 8 of the 32 completed systematic
215 reviews and maps reported data for all stages of the review process (i.e. searching, duplicate
216 removal, title, abstract and full text screening and full text retrieval).

217

218 *Survey of systematic review practitioners*

219 Of the 33 included responses, only 7 provided data for all 15 questions asked about their
220 experience with reviews, while a further 12 provided data for the stages up to data/meta-data
221 extraction and beyond. On average, respondents had conducted a median of 2 systematic
222 reviews (minimum=0, maximum=18). Only one respondent had not previously conducted a
223 review before: data from this respondent were in relation to full text retrieval alone, since
224 they had acted as an assistant for a larger group of reviewers. We received fewer responses

225 about later stages of the review than early stages (Table 5), and particularly few responses
226 about the time taken to complete quantitative synthesis (effect size calculation and meta-
227 analysis; n=7 and 8 respectively).

228 A typical CEE systematic review results in a mean of just over 11,000 search results, which
229 falls to approximately 8,500 unique records following duplicate removal (Table 4). Just over
230 1,200 records remain following title screening and around 300 following abstract screening.
231 With the addition of evidence from other sources, the total number of full texts obtained is on
232 average c. 400. Screening of these full texts leaves just over 90 relevant articles/studies.
233 Critical appraisal retains approximately 60 articles/studies, and suitable data are present in a
234 little over 40 of them.

235 The sample size for systematic maps was much smaller than for systematic reviews (n=20
236 versus n=66), but the volume of evidence was far greater for these maps: almost 35,000
237 search results were obtained on average, leaving over 20,000 unique records. Title screening
238 left over 4,000 relevant records and abstract screening left over 1,000. Just over 1,100 full
239 texts were retrieved, with over 400 being relevant at full text. Across the two cases where
240 critical appraisal was performed within a systematic map, on average of over 100 studies
241 were retained in the final map.

242

243 *Estimated effort*

244 The time taken for each stage of a systematic review were lower, on average, for the
245 corresponding stage of a systematic map (Figure 3). The total time estimated for an ‘average’
246 systematic review is 157 days (SD; ± 22), whilst the total time for an ‘average’ systematic
247 map is 252 days (SD; ± 67) when including an optional critical appraisal step, or 209 days
248 (SD; ± 53) excluding critical appraisal. This estimate includes a large amount of time allotted
249 to planning and administration, in an effort to be conservative (45 days for systematic
250 reviews and 60 days for systematic maps [including critical appraisal]). Stages that are
251 calculated by the model include those from searching to effect size calculation, whilst other
252 stages are set as arbitrary defaults that must be changed by the user (see ‘The tool’, below).
253 For these calculated stages, the most time consuming are title screening, full text screening
254 and critical appraisal, with meta-data and data extraction also requiring considerable time.
255 Searching (for traditional academic and grey literature), assembling a library of evidence, full

256 text retrieval, and consistency checking required less time than most other stages. The
257 uncertainty around this data is substantial, resulting from the propagation of errors across the
258 models and the variability in the underlying source data.

259

260 **Discussion**

261 In this paper, we have presented the most comprehensive estimate to date of the effort needed
262 to complete environmental systematic reviews and systematic maps. Our results revealed
263 substantial variability in the number of articles included in synthesis projects, and the total
264 time taken to complete them. However, we also found key bottlenecks during early screening
265 and at critical appraisal and data extraction stages of each review. Below we expand on these
266 findings and their implications for future research practice.

267

268 *Emergent patterns*

269 We found several trends during our analysis that did not match our expectations about which
270 stages of the review would take the most time. Particularly surprising was the observation
271 that a relatively small proportion of time is spent on performing searching activities: between
272 7 and 7.5 days for reviews and maps, respectively. This result may reflect detailed
273 preparation, given that searching should be preceded by in depth building and testing of
274 search strategies, that will be outlined in an *a priori* protocol. Whilst we did not explicitly ask
275 expert reviewers how long they spent designing and testing a search strategy, this part of the
276 review process requires careful planning to ensure the review results are comprehensive and
277 representative of the true evidence base for a particular topic [17, 18].

278 Equally unexpected was our finding that respondents' reported time spent on administration
279 was particularly large: on average 19% of their total time. For reviews this corresponded to
280 24 days, whilst for maps it was almost 40 for those including critical appraisal (35.5 for those
281 without critical appraisal). Reported administration time varied substantially ($SD=12.3$),
282 perhaps indicating discrepancies in respondents' definitions of what should be included.
283 However, this likely reflects the fact that systematic reviewing often requires time spent
284 coordinating a large, possibly international team, and may also require substantial learning or
285 'relearning' of particular skills, such as experimental design or statistics. We have not

286 factored in training time in our analysis, but this is worth considering for novel teams or those
287 that will rely heavily on group tasks using subject but not methodology expertise.

288 More expected was the large amount of time spent on screening (including retrieval); an
289 average of 80 and 27 days for systematic reviews and maps respectively. This is a large
290 proportion of the time budget (17% and 32% for reviews and maps, respectively [39% for
291 maps excluding critical appraisal]). These differences highlight the fact that resources are
292 predominantly shifted towards identifying evidence in maps, whereas far more time is
293 devoted to synthesis in reviews. In reviews a similar time is spent on extracting and analysing
294 the data as screening (25 days). In maps, however, the proportion of total time on extracting
295 meta-data and coding is relatively lower (also 25 days).

296

297 *The implications for 'optional' activities*

298 Our calculations allow us to estimate the impacts of various optional activities on the total
299 time requirements of a systematic review or map. Current CEE guidance suggests that a
300 subset of articles is checked for consistency in the application of inclusion criteria between
301 two reviewers prior to commencing screening in earnest [5], and it suggests that 10% of
302 records should be checked as a minimum. However, in the field of healthcare systematic
303 reviews, dual coding is common [e.g. 19]. By altering the level of consistency checking from
304 the recommended minimum of 10% at each stage to 100% (i.e. complete dual screening), the
305 total time required changes from 155 days to 183 days, an increase of 18%. Whilst regarded
306 by some as a gold standard for systematic review methodology [1], this increase in time
307 requirements is substantial and may prove too costly for some. However, it may be an
308 important concession to maximise reliability and minimise human error in some cases.

309 Similarly, the CEE guidance suggests a selection of review bibliographies is screened to help
310 to maximise comprehensiveness of the search [5]. Increasing this bibliographic checking or
311 'citation chasing' can require considerable time if, for example, all identified reviews are
312 screened in this way, or even if all articles' bibliographies are screened. Assuming that the
313 inclusion rate at title, abstract and full text (and retrieval rate) remain the same in
314 bibliographic checking as for the core of the review, one can readily predict the additional
315 time needed to screen a certain number of reviews or articles in this way. Within a systematic
316 map, a larger volume of reviews is likely to be found, and the user can specify this number.

317 For example, in a systematic map of the impacts of vegetated strips within and around fields
318 [20], around 100 review bibliographies were checked for additional potentially relevant
319 articles. Altering the number of bibliographies checked in our tool to 100 increases the time
320 requirement from 255 to 271 days (6%).

321

322 *Comparison with existing estimates*

323 Previous estimates of the resource requirements of systematic reviews have been imprecise
324 and vary substantially, from between 6 months and 24 months or several years (Table 6).
325 Anecdotally, we have heard estimates that are as long as 5 years by a leading institute that
326 produces systematic reviews in healthcare in Sweden (SBU, <http://www.sbu.se/en/>). Our
327 analyses demonstrate that the time requirements for an ‘average’ CEE-style systematic
328 review need only take 157 days (FTE). This estimate represents just under 1 year FTE, taking
329 vacation, public holidays, and other regular disruptions to full time work into account.
330 Therefore, our analysis reveals a resource requirement in the lower end of the rough estimates
331 provided in the literature. Interestingly, our estimate is under half that of the only other
332 evidence-based assessment of which we are aware, which corresponds to approximately 337
333 days [9]. It is vital to remember that the time estimate by Borah et al. (2017) and the other
334 rough time estimates in the literature are typically meant to reflect the time required to
335 conduct a systematic review, rather than the resource requirements. However, the average
336 total salary costs for a postdoctoral research at Bangor University (chosen arbitrarily due to
337 our knowledge of the university, including National Insurance and USS pension
338 contributions) for 12 months is 48,593 GBP at the time of writing
339 (<https://www.bangor.ac.uk/finance/py/documents/pay-scales-en.pdf>). Including other
340 costs, such as support staff time and travel and subsistence for meetings, this sum is unlikely
341 to rise above 100,000 GBP. This value, again, sits below the mid points for the roughly
342 estimated cost ranges provided in the literature.

343 Our estimates do not attempt to predict a full costing of a systematic review. Furthermore,
344 our analyses are based on reported volumes of work and efficiency rates from the literature
345 and expert systematic reviewers. As such, the numbers are in need of validation using
346 detailed, accurate records of recently completed reviews. However, our estimates and tool are

347 a useful starting point for those wishing to better understand the demands and likely time
348 requirements of a systematic review or map.

349 The times estimated by our tool for the time required to undertake a systematic review are
350 realistic relative to the reviews conducted by Mistra EviEM. This 6 year project funded by
351 Mistra to undertake systematic reviews and maps in the field of environment relevant to the
352 Swedish environmental goals (www.eviem.se/en). The project will have completed 17
353 systematic reviews and maps over a 6.5 year period, with c. 20 years of full time equivalent
354 staff resources (review project managers); approximately 2.2 years of time per review.
355 However, our estimates for systematic maps are somewhat higher than those indicated for
356 EviEM maps in our experience. This is almost certainly the result of a small and
357 heterogeneous evidence base for completed systematic maps: fewer systematic maps have
358 been completed to date and the variability around the volume of evidence is substantial (SD
359 for systematic review total search results is 11,786 records, whilst it is 39,434 for systematic
360 maps). Systematic maps are more adaptable by their very nature [21, 22], but a larger
361 evidence base would be useful in increasing the precision of the data in our tool.

362

363 *Limitations of our analysis and the evidence base*

364 Our analysis presents the best available information on the number of articles, and the
365 amount of time, included in a typical environmental systematic review or map. However, it is
366 possible that a number of factors may adversely affect the reliability of our analyses of the
367 'average' time needed to complete a review or map. Below we outline some of these points
368 so as to avoid the risk of faulty interpretation of our results.

369 First, our systematic review calculations assume a quantitative synthesis will be performed,
370 which may be true for the majority of current systematic reviews in the field of environment.
371 However, qualitative synthesis is a valuable evidence synthesis method [23], and its use will
372 likely increase in CEE reviews in the future. However, qualitative systematic reviews often
373 do not hold the same regard for issues such as comprehensiveness that quantitative systematic
374 reviews do. For example, qualitative syntheses may stop screening after a certain point
375 because of information saturation. Accordingly, these reviews should be dealt with different
376 when performing an analysis relating to time requirements, and tools for predicting times
377 should be built specifically for these kinds of analysis. It may be the case, however, that

378 specific qualitative reviews could adapt our tools to fit the desired methods. Our intention,
379 however, is not to make a universal analysis or tool for all types of reviews. Instead, we focus
380 on traditional quantitative systematic reviews.

381 Second, all of the data in our analyses have a high level of variability, poor levels of
382 reporting, or both. This results both from a heterogeneous evidence base and a relatively
383 small sample size. For example, of the 19 completed systematic reviews, only 8 reported the
384 number of duplicates removed from total search results. Similarly, whilst 18 reviews reported
385 the total number of included articles, only 10-11 articles reported the number of articles
386 following title screening, abstract screening and full text retrieval. Future CEE reviews
387 should strive to report such methodological information consistently, and CEE should create
388 or adhere to accepted reporting standards for all published reviews (e.g. PRISMA [16] or
389 ROSES [24]). Future analyses should increase sample size, making the most of the rapidly
390 expanding body of reliable systematic reviews. Indeed, efforts to record descriptive summary
391 information regarding systematic review methods are underway (e.g. ROSES [24]).

392 Third, we were not able to provide evidence-based data for all parts of our analysis. A
393 number of key variables have been estimated from our personal experience, including the
394 time required for additional searching for literature, and the number of bibliographies
395 screened. Where possible, future analyses should attempt to examine the evidence base for
396 these data.

397 Fourth, there are clear exceptional circumstances that would affect the reliability of
398 predictions made from our tool. For example, a change in core staff midway through a project
399 would likely require a significant proportion of time to acquaint new staff with what has been
400 done to date. However, careful file management and clear record keeping could reduce this to
401 a minimum. Large review teams may require more resources to train and manage,
402 particularly if meeting remotely. Novice teams may require substantial training time and may
403 suffer from low efficiency in the earlier stages of a review. Finally, undertaking reviews over
404 an extended time period can result in particularly low efficiency if core staff must reacquaint
405 themselves with their own work after significant gaps.

406 Fifth, our tool allows the end user to estimate the time required to complete a systematic
407 review or systematic map, and our analyses of the evidence base provide useful default
408 values should any information be unknown to the user. These default values, however are

409 based on an ‘average’ systematic review or map. It is important to note that the heterogeneity
410 across CEE reviews means that this ‘average’ review, whilst helpful as a starting point, is
411 perhaps not a meaningful entity. Context is highly important for each review, and knowing
412 something about the volume or the nature of the evidence (e.g. proportional relevance of a
413 subset) will allow end users to estimate time requirements much more accurately. We should
414 not assume that all reviews are alike and that the times calculated in our analysis are a
415 reliable estimate alone when planning a review. We encourage users to undertake good
416 quality scoping, as suggested in the CEE Guidelines [5] so as to provide reliable predictions
417 of the volume of evidence, the proportional relevance of articles and studies, and the time
418 required by the user’s team to undertake specific tasks.

419 Finally, we have calculated mean volumes of evidence at each stage of the review process
420 and have used inclusion rates and working speeds to calculate an independent mean time
421 requirement for each stage based on available evidence. However, many reviews do not
422 report all data for each stage of the review, and the results of one stage are dependent upon
423 the nature of the stages preceding it. In ideal circumstances, we would have full data from all
424 reviews that would allow us to model the time requirement based on various contextual
425 variables, for example the inclusion rate of the preceding stage. This is not possible with our
426 limited dataset, however, and our methods represent a necessary compromise.

427

428 *Future work needed*

429 As described above, there is a need for a greater number of data points in future analyses,
430 both for published systematic reviews and maps and for survey data relating to processing
431 speeds. This itself would be aided by better reporting of methods used and records found at
432 all stages of the review process in CEE reviews. Some efforts are underway to record this
433 data more consistently (e.g. ROSES [24]).

434 We also highlight the need for evidence-based estimates of the financial costs associated with
435 systematic reviews, taking into account the price of necessary software, consultancy support
436 (e.g. from an informatician), registration and publication fees, communication materials, and
437 physical meetings. Although there will be considerable local and regional variability in the
438 real world prices of these services, an itemised list of recommended activities is a vital point
439 of departure for those planning an efficient and successful review.

440 Finally, results from our analyses and predictions using our tool should be continually tested
441 and the tool refined in order to match developments in systematic review methodology (e.g.
442 machine learning and prioritised screening [25]).

443

444 **Conclusions**

445 Our analyses shed light on the most time-consuming stages of the systematic review and map
446 process. We have highlighted key bottlenecks from the perspective of time requirements, and
447 our results allow future reviewers to plan their time accordingly. Our tool uses evidence-
448 based defaults as a useful starting point for those wishing to predict the time requirements for
449 a particular review. We also call on CEE and CEE review authors to improve the reporting of
450 the methods and results of their reviewing processes.

451 **References**

- 452 1. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions:
453 John Wiley & Sons; 2011.
- 454 2. Pullin AS, Stewart GB. Guidelines for systematic review in conservation and
455 environmental management. *Conservation biology*. 2006;20(6):1647-56.
- 456 3. Haddaway N, Woodcock P, Macura B, Collins A. Making literature reviews more
457 reliable through application of lessons from systematic reviews. *Conservation Biology*.
458 2015;29(6):1596-605.
- 459 4. Shlonsky A, Noonan E, Littell JH, Montgomery P. The role of systematic reviews and
460 the Campbell Collaboration in the realization of evidence-informed practice. *Clinical Social*
461 *Work Journal*. 2011;39(4):362-8.
- 462 5. CEE. Guidelines for Systematic Review and Evidence Synthesis in Environmental
463 Management. Version 4.2. Environmental Evidence;

464 [http://environmentalevidence.org/wp-content/uploads/2014/06/Review-guidelinesversion-](http://environmentalevidence.org/wp-content/uploads/2014/06/Review-guidelinesversion-4.2-finalPRINT.pdf)
465 [4.2-finalPRINT.pdf](http://environmentalevidence.org/wp-content/uploads/2014/06/Review-guidelinesversion-4.2-finalPRINT.pdf)2013.
- 466 6. Haddaway NR, Bayliss HR. Shades of grey: two forms of grey literature important for
467 reviews in conservation. *Biological Conservation*. 2015;191:827-9.
- 468 7. Westgate MJ, Lindenmayer DB. The difficulties of systematic reviews. *Conservation*
469 *Biology*. 2017.
- 470 8. Collins A, Coughlin D, Miller J, Kirk S. The production of quick scoping reviews and
471 rapid evidence assessments: a how to guide. 2015.
- 472 9. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers
473 needed to conduct systematic reviews of medical interventions using data from the
474 PROSPERO registry. *BMJ open*. 2017;7(2):e012545.
- 475 10. R Core Development Team. R: a language and environment for statistical computing.
476 Version 3.4.1. R Foundation for Statistical Computing, Vienna, Austria. 2017.
- 477 11. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. shiny: Web Application
478 Framework for R. R package version 1.0. 3. 2017.
- 479 12. Chang W. shinydashboard: Create Dashboards with 'Shiny'. R package version 05.
480 2015;1.
- 481 13. Sievert C, Parmer C, Hocking T, Chamberlain S, Ram K, Corvellec M, et al. plotly:
482 Create Interactive Web Graphics via plotly. js. R package version 4.7.1. [https://CRAN.R-](https://CRAN.R-project.org/package=plotly)
483 [project.org/package=plotly](https://CRAN.R-project.org/package=plotly). 2017.
- 484 14. Humbert J-Y, Pellet J, Buri P, Arlettaz R. Does delaying the first mowing date benefit
485 biodiversity in meadowland? *Environmental Evidence*. 2012;1(1):9.
- 486 15. Randall NP, Donnison LM, Lewis PJ, James KL. How effective are on-farm
487 mitigation measures for delivering an improved water environment? A systematic map.
488 *Environmental Evidence*. 2015;4(1):18.
- 489 16. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for
490 systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*.
491 2009;6(7):e1000097.
- 492 17. Bayliss HR, Beyer FR. Information retrieval for ecological syntheses. *Research*
493 *synthesis methods*. 2015;6(2):136-48.
- 494 18. Livoreil B, Glanville J, Haddaway NR, Bayliss H, Bethel A, Lachapelle FF, et al.
495 Systematic searching for environmental evidence using multiple tools and sources.
496 *Environmental Evidence*. 2017;6(1):23.

- 497 19. Jones E, Taylor B, MacArthur C, Pritchett R, Cummins C. The effect of early
498 postnatal discharge from hospital for women and infants: a systematic review protocol.
499 Systematic reviews. 2016;5(1):24.
- 500 20. Haddaway NR, Brown C, Eggers S, Josefsson J, Kronvang B, Randall N, et al. The
501 multifunctional roles of vegetated strips around and within agricultural fields. A systematic
502 map protocol. Environmental Evidence. 2016;5(1):18.
- 503 21. Haddaway NR, Bernes C, Jonsson B-G, Hedlund K. The benefits of systematic
504 mapping to evidence-based environmental management. Ambio. 2016;45(5):613-20.
- 505 22. James KL, Randall NP, Haddaway NR. A methodology for systematic mapping in
506 environmental sciences. Environmental Evidence. 2016;5(1):7.
- 507 23. Flemming K. The synthesis of qualitative research and evidence-based nursing.
508 Evidence-based nursing. 2007;10(3):68-71.
- 509 24. Haddaway NR, Macura B, Whaley P, Pullin AS. ROSES RepOrting standards for
510 Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the
511 plan and conduct of environmental systematic reviews and systematic maps. Environmental
512 Evidence. in press.
- 513 25. Shekelle PG, Shetty K, Newberry S, Maglione M, Motala A. Machine learning versus
514 standard techniques for updating searches for systematic reviews: a diagnostic accuracy
515 study. Annals of Internal Medicine. 2017;167(3):213-5.

516

517 **Additional Files**

518 Additional File 1. Email survey sent to systematic review practitioners.

519 Additional File 2. Data and calculations used to arrive at metrics and standard errors
520 compiled to produce time requirements for the various steps of a systematic review.

521 Additional File 3. Systematic review and map time planner tool (Excel version).

522 **Figure legends**

523 **Figure 1.** The number of publications from CEE between May 2012 and March 2017.

524 **Figure 2.** The mean number of records remaining after each key stage of the a) systematic
525 review and b) systematic map processes. Error bars are ± 1 standard deviation. Labels are data
526 values.

527 **Figure 3.** Time taken for each stage of the systematic map process in days. Error bars are ± 1
528 standard deviation.

529 **Tables**

530 **Table 1.** Key data collated from the survey of systematic review practitioners.

How many SRs have you undertaken?
Time taken to download search results from each database (days)
Time taken to assemble library of results and remove duplicates
Time taken to screen organisation websites (each in days)
Number of titles screenable per day
Number of abstracts screenable per day
Number of titles and abstracts screenable per day (together)
Number of full texts retrievable per day
Number of full texts screenable per day
Number of articles for meta-data extraction/coding per day
Number of articles for critical appraisal per day
Number of articles for data extraction per day
Number of articles for effect size calculation per day
Time taken for meta-analysis
Time taken for report writing
Percentage of time required for administration

531

532 **Table 2.** The main stages of a systematic review used to predict time requirements.

Activity	Source of data
Planning time (stakeholder engagement, Q formulation)	Semi-fixed values to be inputted by reviewer (arbitrary defaults)
Protocol development	Semi-fixed values to be inputted by reviewer (arbitrary defaults)
Searching (academic literature)	Values calculated by tool
Searching (grey literature)	Values calculated by tool
Assembling library and removing duplicates	V Data from survey
Title screening	Values calculated by tool
Abstract screening	Values calculated by tool
Full text retrieval	Values calculated by tool
Full text screening	Values calculated by tool
Consistency checking (screening)	Values calculated by tool
Meta-data extraction	Values calculated by tool
Critical appraisal	Values calculated by tool
Data extraction	Values calculated by tool
Effect size calculation	Values calculated by tool
Meta-analysis	Data from published reviews/protocols
Report writing	Data from published reviews/protocols
Communication materials	Data from published reviews/protocols
Administration	Data from survey

533

534 **Table 3.** Summary metrics used in intermediate steps of time calculation and the sources of
 535 data used to calculate metrics in Table 2.

Activity	Source of data
Number of searching databases	Data from published reviews/protocols
Time taken to download one database's search results (days)	Data from survey
Number of grey literature website sources	Data from published reviews/protocols
Time required for additional grey literature sources (other languages, etc.) (days)	Semi-fixed values to be inputted by reviewer (arbitrary defaults)
Number of grey literature sources (website) screenable per day	Data from survey
Number of search results (total)	Data from published reviews/protocols
Estimated duplicate rate (percent)	Data from published reviews/protocols
Number of search results after duplicate removal	Data from published reviews/protocols
Estimated inclusion rate (title) (percent)	Data from published reviews/protocols
Screening speed (titles per day)	Data from survey
Estimated inclusion rate (abstract) (percent)	Data from published reviews/protocols
Screening speed (abstracts per day)	Data from survey
Retrieval rate (percent retrievable)	Data from published reviews/protocols
Retrieval speed (articles per day)	Data from survey
Estimated inclusion rate (full text) (percent)	Data from published reviews/protocols
Screening speed (full texts per day)	Data from survey
Percentage of results screened for consistency (title)	Semi-fixed values to be inputted by reviewer (arbitrary defaults)
Percentage of results screened for consistency (abstract)	Semi-fixed values to be inputted by reviewer (arbitrary defaults)
Percentage of results screened for consistency (full text)	Semi-fixed values to be inputted by reviewer (arbitrary defaults)
Number of consistency checkers at each stage	Semi-fixed values to be inputted by reviewer (arbitrary defaults)
Meta-data extraction rate (articles per day)	Data from survey
Critical appraisal rate (articles per day)	Data from survey
Percentage of studies with meta-analysable data	Data from published reviews/protocols
Data extraction rate (articles per day)	Data from survey
Effect size calculation rate (articles per day)	Data from survey
Estimated inclusion rate following critical appraisal	Data from published reviews/protocols
Percentage of time needed for administration	Data from survey

536

537

538

539 **Table 4.** Mean number of records remaining in each stage of the systematic review and map
 540 process, along with minima, maxima, standard errors (SE) and sample sizes (n=number of
 541 reviews/protocols providing data).

542 *Systematic reviews*

Number of items remaining after:	Mean	SD	SE	n⁵⁴³
Searching	11,786	10,230	2347	19
Duplicate removal	8,493	6,040	2135	8
Title screening	1,236	766	231	11
Abstract screening	309	269	85	10
Full text retrieval	405*	347	105	11
Full text screening	93	85	20	18
Critical appraisal	60	41	15	8
Narrative synthesis	60	41	15	8
Quantitative synthesis	45	43	25	3

544

545 *Systematic maps*

Number of items remaining after:	Mean	SD	SE	n⁵⁴⁶
Searching	34,165	39,434	11,384	12
Duplicate removal	22,584	20,578	6,204	11
Title screening	4,118	2,002	817	6
Abstract screening	1,027	575	217	7
Full text retrieval	1,126*	859	286	9
Full text screening	423	371	117	10
Critical appraisal	116	22	16	2
Narrative synthesis	116	22	16	2

547

548 * The number of full texts is greater than the number of screened abstracts due to the inclusion of grey
 549 literature at the full text stage that have been externally screened for relevance from the searching of
 550 websites, web-based search engines, etc.

551

552

553

554

555 **Table 5.** Summary data for respondents to the survey: n, sample size; SD, standard deviation.

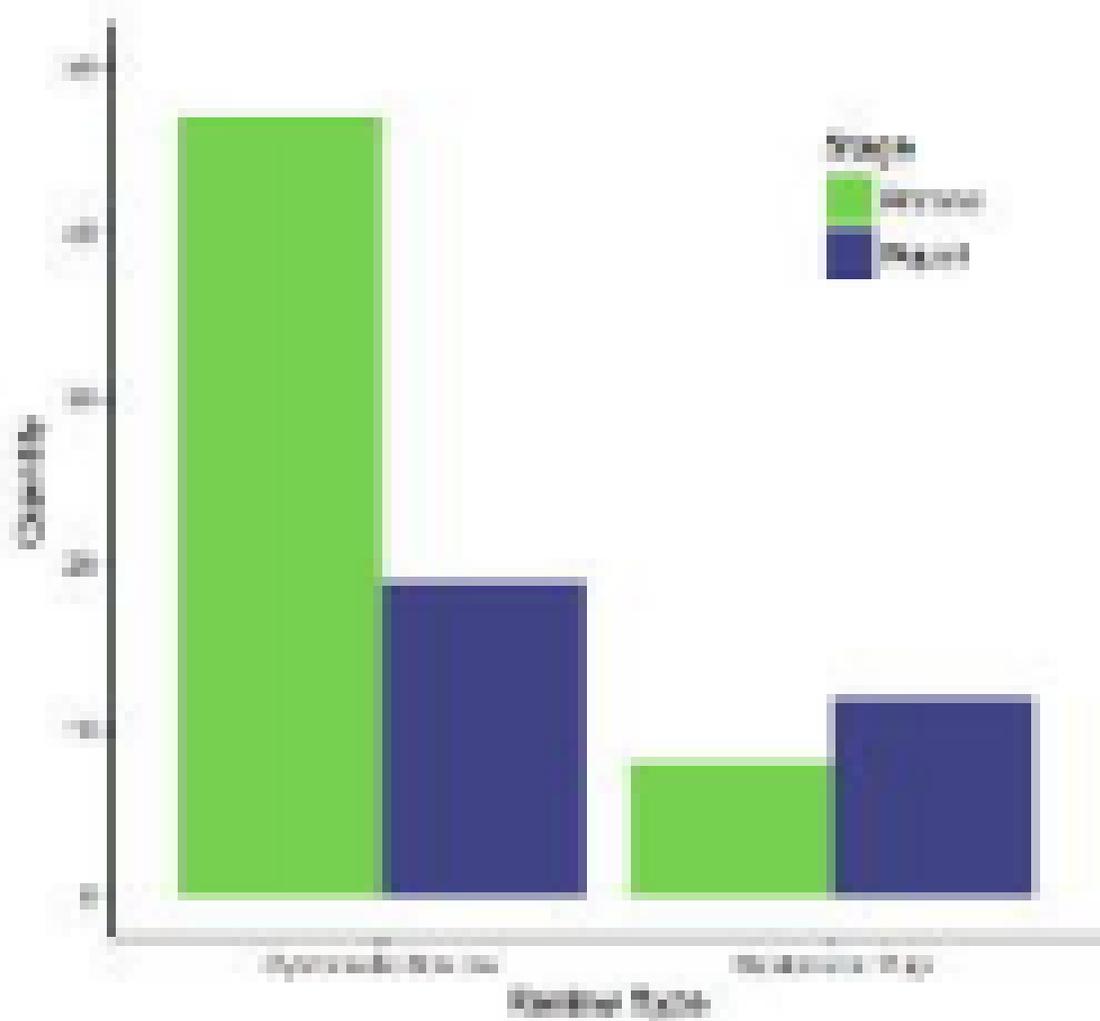
Survey question	Mean response	n	SD
Time taken to download search results from each database (days)	0.25	20	0.220
Time taken to assemble library of results and remove duplicates	1.37	18	1.403
Time taken to screen organisation websites (each in days)	0.15	21	0.122
Number of titles screenable per day	854.35	23	533.622
Number of abstracts screenable per day	192.29	24	111.900
Number of titles and abstracts screenable per day (together)	468.14	22	128.216
Number of full texts retrievable per day	170.94	24	137.369
Number of full texts screenable per day	43.99	30	31.014
Number of articles for meta-data extraction/coding per day	16.69	21	11.574
Number of articles for critical appraisal per day	11.68	19	8.145
Number of articles for data extraction per day	6.87	19	5.093
Number of articles for effect size calculation per day	24.00	7	34.083
Time taken for meta-analysis	6.75	8	5.092
Time taken for report writing	15.53	20	10.228
Percentage of time required for administration	19.00	22	12.276

556

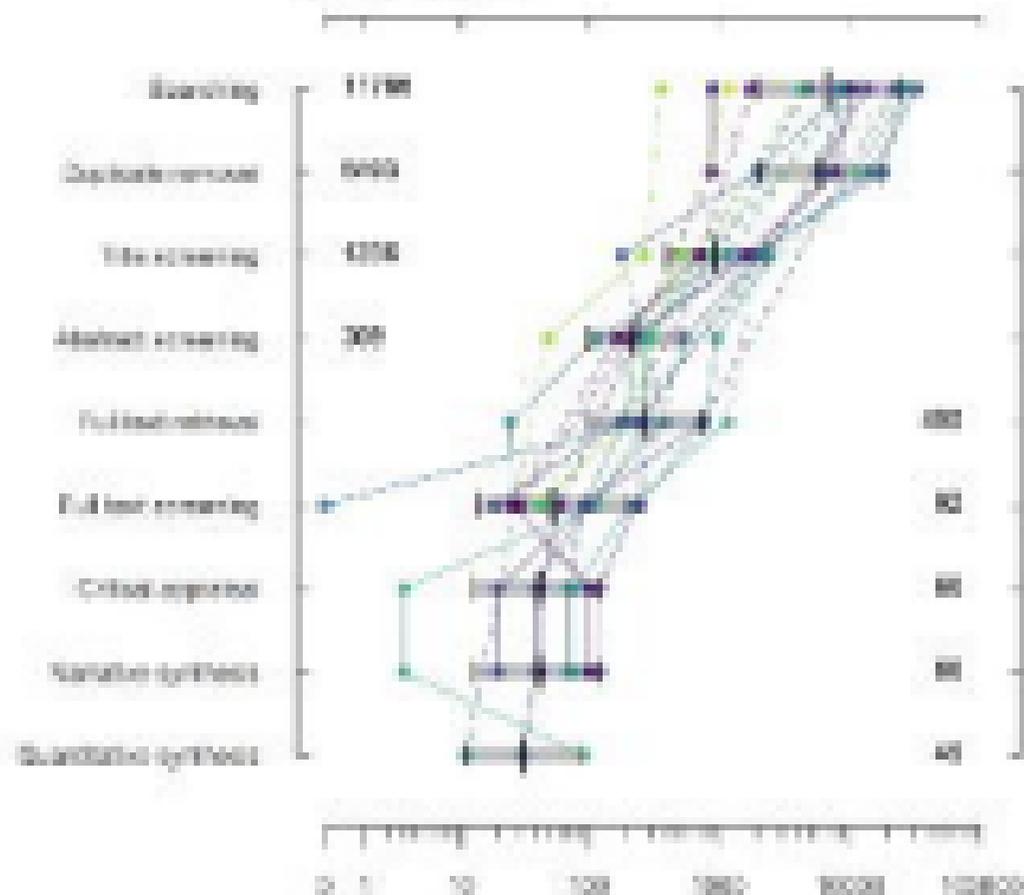
557 **Table 6.** Estimates of the resource requirements for systematic reviews from a non-systematic search of the literature.

Financial cost	Time requirement	Review type	Citation
	0.5-3 years	Systematic review	Dicks, L.V., Walsh, J.C. and Sutherland, W.J., 2014. Organising evidence for environmental management decisions: a '4S' hierarchy. <i>Trends in ecology & evolution</i> , 29(11), pp.607-613.
30,000-300,000 USD	Several years	Systematic review	Collaboration for Environmental Evidence. 2013. Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. Version 4.2. Environmental Evidence: www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf
80,000-120,000 GBP	10-18 months	Systematic review	Collins, A., Coughlin, D., Miller, J. and Kirk, S., 2015. The production of quick scoping reviews and rapid evidence assessments: a how to guide.
<=250,000 USD		Systematic review	McGowan, J. and Sampson, M., 2005. Systematic reviews need systematic searchers. <i>Journal of the Medical Library Association</i> , 93(1), p.74.
	9-24 months	Systematic review	Centre for Cognitive Ageing and Cognitive Epidemiology (http://www.ccace.ed.ac.uk/research/software-resources/systematic-reviews-and-meta-analyses)
	67.3 weeks	Systematic review	Borah, R., Brown, A.W., Capers, P.L. and Kaiser, K.A., 2017. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. <i>BMJ open</i> , 7(2), p.e012545.

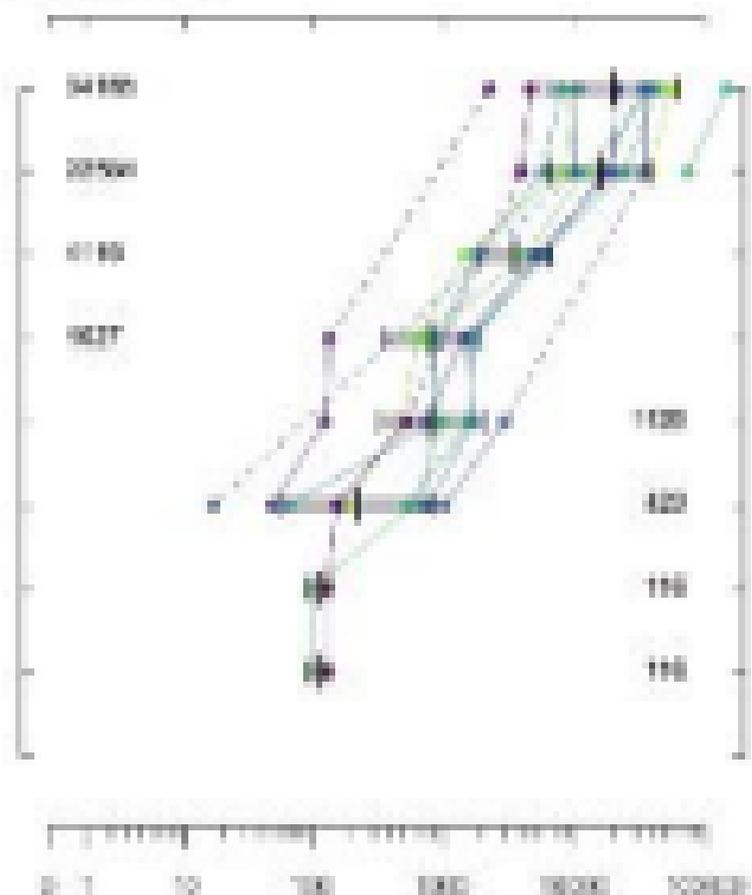
558



a) Systematic Review



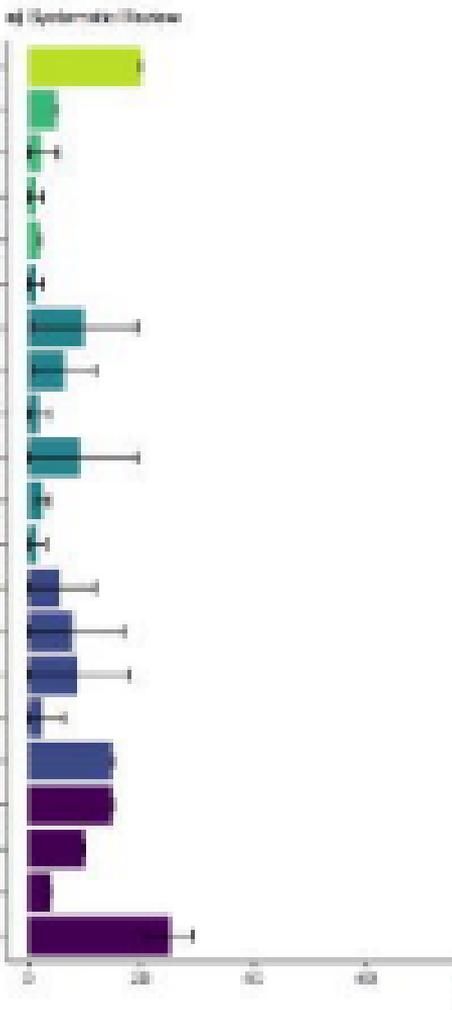
b) Systematic Map



Number of articles

a) Systematic Review

- Planning (e.g. search/selector engagement, Q formulation)
- Protocol development
- Searching (academic literature)
- Searching (grey literature)
- Additional searching (e.g. stakeholder, hand searching etc.)
- Assessing fit/relevance and removing dupes
- Title screening
- Abstract screening
- Full-text relevance
- Full-text screening
- Consistency checking
- Blinding/report checking
- Meta-data extractor
- Critical appraisal
- Data extractor
- Effect size calculator
- Meta-analysis
- Report writing
- Communication materials
- Meetings
- Archival work



b) Systematic Map

