

Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching

Neal R. Haddaway^{1,2}  | Matthew J. Grainger³  | Charles T. Gray⁴

¹Leibniz-Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany

²African Centre for Evidence, University of Johannesburg, Johannesburg, South Africa

³Norwegian Institute for Nature Research, Trondheim, Norway

⁴School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, UK

Correspondence

Neal R. Haddaway, Leibniz-Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany.

Email: neal_haddaway@hotmail.com; neal.haddaway@sei.org

Abstract

Systematic searching aims to find all possibly relevant research from multiple sources, the basis for an unbiased and comprehensive evidence base. Along with bibliographic databases, systematic reviewers use a variety of additional methods to minimise procedural bias. Citation chasing exploits connections between research articles to identify relevant records for a review by making use of explicit mentions of one article within another. Citation chasing is a popular supplementary search method because it helps to build on the work of primary research and review authors. It does so by identifying potentially relevant studies that might otherwise not be retrieved by other search methods; for example, because they did not use the review authors' search terms in the specified combinations in their titles, abstracts, or keywords. Here, we briefly provide an overview of citation chasing as a method for systematic reviews. Furthermore, given the challenges and high resource requirements associated with citation chasing, the limited application of citation chasing in otherwise rigorous systematic reviews, and the potential benefit of identifying terminologically disconnected but semantically linked research studies, we have developed and describe a free and open source tool that allows for rapid forward and backward citation chasing. We introduce *citationchaser*, an R package and Shiny app for conducting forward and backward citation chasing from a starting set of articles. We describe the sources of data, the backend code functionality, and the user interface provided in the Shiny app.

KEYWORDS

bibliographic checking, evidence synthesis tools, information retrieval, pearl growing, software development, systematic review tool, systematic searching

Highlights

- Supplementary methods of searching for articles are important for robust systematic reviews
- Tracking forwards and backwards through citation networks allows relevant articles to be identified that may be missed by bibliographic database searching

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Research Synthesis Methods* published by John Wiley & Sons Ltd.

- Here, we introduce a free and open source R-based tool (*citationchaser*) for rapidly performing forward and backward citation chasing to support comprehensive systematic review searches
- We believe *citationchaser* could substantially improve search comprehensiveness with minimal additional time and resources

1 | BACKGROUND

1.1 | Citation chasing as a systematic search technique

Systematic searching is the process by which researchers seek out all possibly relevant research records from a variety of sources in an attempt to collate an unbiased and comprehensive set of bibliographic records.¹ Along with searching bibliographic databases using complex search strings,² systematic review authors employ a suite of additional methods to minimise procedural bias,³ including but not limited to: targeted searches of organisational websites for grey literature; suggestions of articles from stakeholders; searches of institutional repositories for unpublished theses and data; and assessing records that are cited by and that cite a set of articles of known relevance (citation chasing).

Citation chasing (also referred to as ‘citation searching’, ‘citation tracking’, ‘snowballing’, ‘pearl growing’, ‘footnote chasing’, ‘reference scanning’, ‘reference checking’, ‘bibliographic checking’, ‘citation mining’, or ‘reference harvesting’) in systematic reviews exploits

connections between similar research articles to identify relevant records for consideration in a review by making use of explicit mentions of one article within another (*citations*). These connections from a single article travel back in time (a starting article references a set of previously published articles) and forwards in time (articles published after the starting article include it in their reference lists).

Citation chasing is a popular supplementary search method because it helps to build on the work of primary research and review authors. It does so by identifying potentially relevant studies that might otherwise not be retrieved by other search methods; for example, because they did not use the review authors' search terms in the specified combinations in their titles, abstracts or keywords.⁴

Backward citation chasing (Figure 1) is the process of obtaining and assessing the relevance of all records cited within the bibliographies or reference lists of a set of articles (cited articles). Backward citation chasing involves the collation of a finite and definable set of bibliographic records, defined by the number of records across multiple reference lists. For example, if articles cite on average 50 articles, then backward citation chasing of 100 starting records would yield ~5000 cited records (assuming

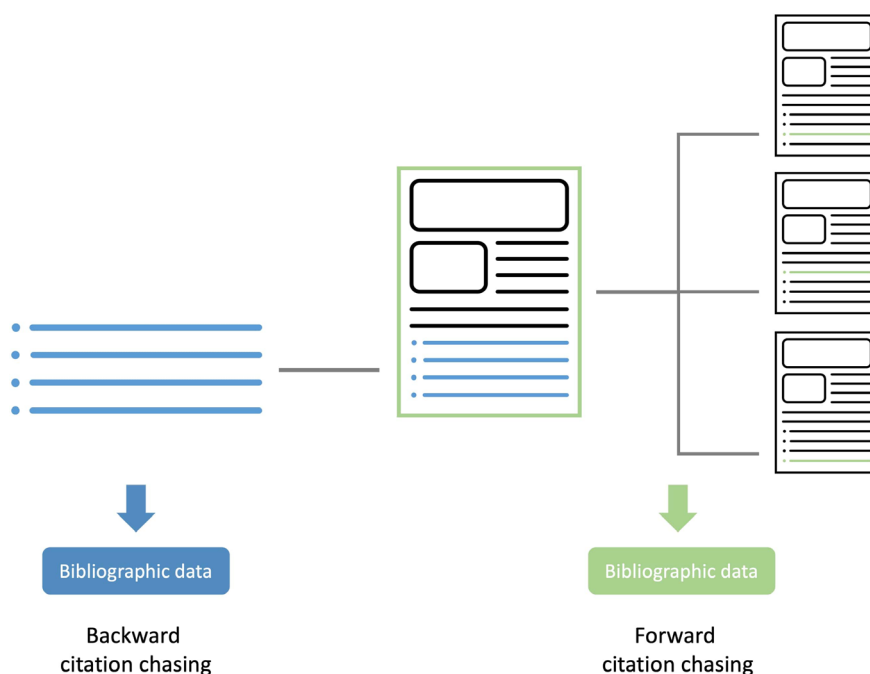


FIGURE 1 Schematic showing a graphical representation of backward and forward citation chasing, from a starting set of research articles [Colour figure can be viewed at wileyonlinelibrary.com]

minimal duplication). There is a quantifiable degree of accuracy to the obtained records, in that each included starting record has a ‘true’ number of references that can be obtained, and any attempt to retrieve these can be assessed for completeness. However, some databases only include references of a certain type (e.g., ignoring ‘grey literature’) or only those records that are already indexed in the database. This can lead to an incomplete set of references for a given article.

Forward citation chasing (Figure 1) involves identifying articles that have cited a particular set of articles (*citing articles*). This process looks forward in time, identifying citation networks in the years since the starting articles were published. Forward citation chasing is non-definable in volume, in that there is continued growth in the number of citations of the starting set. Furthermore, databases of citations vary in their estimates of the number of citing articles depending on the type of citing records included. For example, Web of Science may state that a certain record has 50 citations since it only includes records already indexed within its databases, while Google Scholar may list 75 citations, since it counts citations from any web-based document classified as ‘scholarly’.

Previously cited benefits to citation chasing include:

- Even a few starting records can create a useful network⁵;
- Related but semantically disconnected (parallel) topics (i.e. those that cover the same topic but use entirely different vocabulary) can be identified that might otherwise be missed by targeted searching of bibliographic databases⁶;
- It is useful in topics where the terminology is not used consistently across the evidence base.⁷

In systematic reviews, the starting set of articles is often one of the following: the final list of included studies, for example, Reference 8; the set of identified search results, for example, Reference 9; a benchmark set of articles of known relevance assembled a priori by the review team, for example, Reference 10; a list of relevant review articles—sometimes those encountered during the reviewing process, for example, Reference 11.

1.2 | Practical considerations

Prior to automated methods, backward citation chasing may have often been conducted manually (i.e., searching for citations in reference lists one by one) and non-digitally (i.e., based on transcription and searching of each reference one-by-one). Traditionally, this might be done by individually screening the reference lists from

printouts of a relevant set of articles and noting relevance by hand. This is both time-consuming and very difficult to record transparently—ideally every record assessed should be documented, which would involve transcribing or copying thousands of records into a record-keeping tool, such as a spreadsheet. However, bibliographic databases and search engines typically provide lists of references that can be exported and integrated into reference or review management tools for digital transparency.

Forward citation chasing requires the use of tools that track citations as the citing records are indexed. According to Cooper et al.,¹² three resources have been most commonly used in citation chasing (both forward and backward) to date: Web of Science, Scopus and Google Scholar.

Numerous library guides exist to support researchers in conducting citation chasing (e.g., the University of Illinois Library; <https://guides.library.illinois.edu/c.php?g=563215&p=3877582>). These guides typically all advise users to manually examine the reference lists and lists of citing articles for each record one-by-one.

Several factors affect the efficacy and efficiency of citation chasing:

- *Currency* refers to how up to date a resource is for tracking citations of its included records; a significant lag in recording a citation would reduce the effectiveness of that resource for forward citation chasing in recent years and may omit important parts of the citation network.¹³
- *Network comprehensiveness* is affected both by the proportion of reference lists for included articles that are stored as digitised citations (if the database used is selective in its indexing of journals, then indexed article reference lists will only be partially complete), and by the efficiency at capturing new citations by latterly published articles.
- *Coverage* refers to the proportion of the research literature that is indexed by any bibliographic resource—poor coverage increases the chances that a starting record may be missing from the resource.
- The *functionality* of bibliographic resources may not allow the user to export reference lists and lists of citing articles for more than one starting record at a time, forcing reviewers to export these lists one-by-one, requiring substantial time and reducing efficiency of the overall process.

1.3 | Efficiency and time/resource cost of citation chasing

Here, we present several studies reporting on the time requirements for citation chasing, this is not intended to

be a comprehensive or systematic appraisal of the literature, but rather an informative example of some empirically quantified examples.

Wright et al.¹⁴ conducted forward citation chasing of 40 starting records using Web of Science (the precise databases were not specified), Medline via OVID, Google Scholar and Scopus on each record for comparison of citation coverage. This resulted in a total of 4161 results: Google Scholar, 1680; Scopus, 1173; Web of Science, 1095; and Medline via OVID, 213. The process yielded one relevant study that was omitted from bibliographic database searches. In addition, the authors estimated that the time to conduct forward citation chasing was a total of 4 days: 2 days downloading 1680 results from Google Scholar; 1 day downloading 2481 results from Web of Science (databases not specified), Scopus and Medline; 2 days screening the records identified.

Hinde et al.⁶ conducted backward and forward citation chasing from a starting set of 14 records. The authors performed four iterations of record retrieval in each direction until no new studies were identified (what they refer to as Bidirectional Citation Searching to Completion; BCSC). They used Web of Science (they did not specify which databases), screening a total of 4529 potentially relevant records and finally yielding an additional 76 relevant studies.

Levay et al.¹³ performed forward citation chasing for 46 starting records using Google Scholar and Web of Science (Science Citation Index and Social Science Citation Index) to identify citations for each record. The authors noted that the time required for this process was 75 hours (Google Scholar) and 4 h (Web of Science), including time to identify and deduplicate 783 records (Web of Science = 46 records and Google Scholar = 737 records).

It is clear from these examples that manual (but digital) citation chasing (i.e., copy-and-paste searching identifying information of individual references/citations) can be a particularly time-consuming process despite using web-based digitised resources.

1.4 | Current citation chasing practices and best practice

The Collaboration for Environmental Evidence (CEE) states that citation chasing is frequently used but provides no guidelines or set of minimum standards for how it should be conducted. Examination of 16 recent systematic reviews (and maps) published in 2020 and 2021 (to date, September 2021) in the CEE journal *Environmental Evidence* shows that citation chasing has not been common in recent reviews (Table 1), with six reviews not performing any backward citation chasing at all and no review performing forward citation chasing. In

five reviews, the set of starting records used in citation chasing was not provided or identifiable.

In a methodological review of 96 systematic reviews published in the journal, *Campbell Systematic Reviews* a total of 84 reviews included citation chasing methods.¹⁵ The percentage of reviews including the method increased from 80% to 94% for January 2011 to September 2014 relative to October 2014 to February 2018. Forward citation chasing was not assessed.

Cochrane is a leading producer of systematic reviews for healthcare decision-making. Reference list checking (i.e., backward citation chasing), based on included studies or other relevant systematic reviews, is regarded as a mandatory requirement for authors conducting Cochrane reviews, under the current (2021) MECIR standards.¹⁶ The Cochrane Handbook¹⁷ suggests forward citation chasing as an option for reviews on complex and public health interventions.

Briscoe et al.¹⁸ reviewed the use of citation chasing for 198 Cochrane reviews, published between November 2016 and January 2017, after excluding withdrawn and empty reviews. Of these, 172 used backward citation chasing and 18 used forward citation chasing. Backward citation chasing was conducted by manually checking reference lists of all included studies, per the MECIR standards, in all cases. Three reviews did not describe the resource used for forward citation chasing. Web of Science was used in most cases (13/15). Scopus was used twice and Google Scholar once. Only one review made use of two named sources to identify forward citations (the remaining 14 of 15 named just one source).

1.5 | Existing tools for citation chasing

Some softwares already exist that support basic forms of citation chasing, but they typically have substantial limitations. For example, bibliographic data platforms such as Web of Science allow users to extract sets of references and citing articles for a given bibliographic record, but these must be selected and exported individually (rather than for multiple starting records). The co-citation exploration tool CoCites (<https://www.cocites.com/>) allows a starting set of records to be used to identify referenced and citing records (as well as records that are frequently cited along with the starting set). CoCites uses the iCite tool from the United States National Institutes of Health and is therefore limited in scope and content.

1.6 | Objectives

Given the challenges and high resource requirements associated with citation chasing, the limited application

TABLE 1 Assessment of citation chasing activities in 10 recent CEE reviews

Review	Backward citation chasing (<i>n</i>)	Forward citation chasing (<i>n</i>)	Starting records provided?	Relevant methods text
Tran et al. (2021) ³⁵	37	Not performed	Yes	'After having collected the literature from the different sources described above, we identified 37 relevant reviews (Additional file 3). We searched their bibliographic references manually, resulting in the identification of 30 articles of potential relevance to answer our research questions'.
O'Leary et al. (2021) ³⁶	49	Not performed	No	'In addition, bibliographic searches of all identified relevant tertiary review articles were undertaken'.
Knight et al. (2021) ³⁷	unclear	Not performed	No	'The third step involved examination of reference lists for any literature reviews found by the search in order to identify any new records'.
Lugassy et al. (2021) ³⁸	Not performed	Not performed	n/a	n/a
Häkkinen et al. (2021) ³⁹	41	Not performed	n/a	'To supplement the search, citation chasing was undertaken in the relevant review articles included at title/abstract stage but excluded at full text stage (the reviews, where citation chasing was conducted are identified in Additional file 6)'.
Savilaakso et al. (2021) ⁴⁰	25	Not performed	Yes	'Citation chasing was undertaken to supplement the search. Citations were checked in the following articles: Five most relevant review-articles that were excluded at full text stage (Additional file 4). Relevance was defined based on the proximity of their topic to this review's research question. Five most recent even-aged management articles. The most recent article within every group (see data coding and extraction) of uneven-aged management articles. Five most recent individual (not belonging to a group) uneven-aged management articles'.
Lorick et al. (2020) ⁴¹	4	Not performed	no	'Additionally, the bibliographies of all relevant review articles found were searched for relevant articles. Bibliographies of all relevant review articles were screened separately'.
Dick et al. (2020) ⁴²	Not performed	Not performed	n/a	n/a
Johannesdottir et al. (2020) ⁴³	Not performed	Not performed	n/a	n/a
Sordello et al. (2020) ⁴⁴	37	Not performed	Yes	'After having collected the literature from the different sources described above, we selected 37 relevant reviews from our corpus. Then, we extracted all their

(Continues)

TABLE 1 (Continued)

Review	Backward citation chasing (<i>n</i>)	Forward citation chasing (<i>n</i>)	Starting records provided?	Relevant methods text
				bibliographic references, resulting in 4025 citations (see the list of the 37 reviews and their corresponding number of extracted citations in Additional File 3). Among these citations we excluded all duplicates (intra-duplicates and duplicates between these bibliographies and our previous literature collection). We screened the titles of the remaining citations; we retrieved the pdf file of the selected titles and then we screened their full-texts'.
Duporge et al. (2020) ⁴⁵	Not performed	Not performed	n/a	n/a
Ouédraogo et al. (2020) ⁴⁶	Not performed	Not performed	n/a	n/a
Rytwinski et al. (2020) ⁴⁷	1287	Not performed	No	'In addition, we hand searched reference sections of accepted articles and 297 relevant reviews (see Additional file 2) for any relevant titles that were not found using the search strategy'.
Arton et al. (2020) ⁴⁸	Unclear	Not performed	No	'The references from other relevant reviews (e.g., [10,25,36] were included if they had not been found through the initial searches'.
Goulas et al. (2020) ⁴⁹	Not performed	Not performed	n/a	n/a
Algera et al. (2020) ⁵⁰	255	Not performed	Yes	'Reference sections of accepted articles and 168 relevant reviews were hand searched to evaluate relevant titles that were not found using the search strategy (see Additional file 2 for a list of relevant reviews). ... we screened bibliographies of: (1) a large number of relevant reviews identified at title and abstract (84 reviews) or full-text screening (30 reviews); (2) additional relevant reviews identified from within the bibliographies of the reviews (54 reviews); and (3) included articles. We searched these reference lists of papers until the reviewer deemed that the number of relevant returns had significantly decreased'.

Note: 'n' corresponds to the number of starting records used for citation chasing. Abbreviation: n/a, not applicable.

of citation chasing in otherwise rigorous systematic reviews, and the potential benefit of identifying terminologically disconnected but semantically linked research

studies, we sought out to develop and describe an open source tool that allows for rapid forward and backward citation chasing.

2 | SOFTWARE DEVELOPMENT

2.1 | Tool framework and functionality

We set out to develop an easy-to-use tool designed to collate references (backward citation chasing) and citations (forward citation chasing) from across an entire set of starting records in a single step. This thereby saves the time-consuming effort of searching for each record individually.

We also wanted to ensure that the records retrieved in both directions were provided in an easily usable way—as a research information Systems (RIS) bibliographic data file containing sufficient unique identifying information to allow record disambiguation without unreasonable effort. An RIS file is a standardised data format for a text file containing bibliographic data, named after the developer, RIS. This would allow users to perform automated deduplication of citation chasing results against bibliographic database search results to leave only those records missed by bibliographic searches. Such deduplication is a future aim for citationchaser.

Finally, we aimed to produce open source software, with the entire code base published and freely available online to support incremental development and refinement by anyone. To do this, the backend functionality was written as an R package¹⁹ dependent upon functions from the following packages: `express`²⁰; `httr`²¹; `jsonlite`²²; `tibble`²³ and `dplyr`.²⁴ The frontend functionality was provided as a web-based Shiny app hosted on a Shiny server and depends upon the following additional packages: `shiny`²⁵; `DT`²⁶; `shinycssloaders`²⁷; `data.table`²⁸; `scales`²⁹; `tidyr`³⁰; `networkD3`³¹ and `stringr`.³²

2.2 | Data sources

The tool makes use of The Lens.org bibliographic database aggregator, which collates content from across five bibliographic resources: Microsoft Academic Graph (MAG), CrossRef, CORE, PubMed and PubMedCentral (<https://www.lens.org/lens/search/scholar/structured>). At the time of writing, The Lens.org contains >245 million scholarly records (January 2022). Microsoft Academic Graph and CrossRef have been shown to have considerable overlap with traditional resources such as Scopus and Web of Science.^{33,34}

Lens.org is entirely free to use via its web-based interface, and an API (application programming interface) is provided for programmatic access of the database content. The API requires a token for personal identification, which can be requested free-of-charge for a 14 day trial, with extended access provided at the discretion of The

Lens.org, which is a not-for-profit organisation that benefits from institutional support.

The tool we developed as an R package requires the user to submit a token if used via R, but the web-based Shiny app has an inbuilt token that has been kindly provided free-of-charge by The Lens.org. This means that users of the web app have free access to the API without needing to apply for a token.

2.3 | Use cases

We planned for the following use cases:

1. Users wishing to assemble a set of forward and/or backward citation chasing results manually enter a set of starting records by copy-pasting/transcribing identifiers into the relevant input field, or by uploading a CSV file containing identifiers (and identifier type).
2. Users are referred to the site from another website (e.g., a review management tool) with the list of starting records embedded within the referring URL (e.g., <https://estech.shinyapps.io/citationchaser/test/?doi=10.1038/s41559-020-01295-x,10.1371/journal.pone.0138237&pmids=32652585,32706299>), bypassing the need to paste in or transcribe identifiers

2.4 | Platform for development

We have provided the code behind citation chaser via GitHub to allow for collaborative, incremental updates, bug fixes and functionality expansions. For details on planned developments, see ‘Potential and planned further developments’, below. Users can flag up problems and bugs that need to be resolved or suggest future development needs by raising a new ‘Issue’ (<https://github.com/nealhaddaway/citationchaser/issues>), creating a public discussion about the topic. Users can also make a copy of the code repository (a ‘fork’), modify the code, suggesting an improvement or extension of functionality, and suggest this be integrated into the main code (a ‘pull request’), becoming a co-developer on the project, if desired.

3 | SOFTWARE FUNCTIONALITY

3.1 | Citationchaser functionality

The free-to-use, web-based citationchaser app is available here: <https://estech.shinyapps.io/citationchaser>.

The app provides detailed instructions of how to use the tool on the landing page (Figure 2), including

Welcome to citationchaser!

In searching for research articles, we often want to obtain lists of references from across studies, and also obtain lists of articles that cite a particular study. In systematic reviews, this supplementary search technique is known as "citation chasing": forward citation chasing looks for all records citing one or more articles of known relevance; backward citation chasing looks for all records referenced in one or more articles.

Traditionally, this process would be done manually, and the resulting records would need to be checked one-by-one against included studies in a review to identify potentially relevant records that should be included in a review.

This package contains functions to automate this process by making use of the Lens.org API. An input article list can be used to return a list of all referenced records, and/or all citing records in the Lens.org database (consisting of PubMed, PubMed Central, CrossRef, Microsoft Academic Graph and CORE); [read more here](#).

Large searches may take several minutes to complete, so please be patient.

Consider asking your library to support Lens.org to continue to enable Open Discovery of research articles.

Follow these steps to start chasing!

- In the "Article input" tab, paste a list of article identifiers (e.g. DOIs)
- Check the articles returned are the ones you're interested in
- If you want to perform backward citation chasing (which articles did my articles reference?) then proceed to the "References" tab and click "Search for all referenced articles in Lens.org"
- If you want to perform forward citation chasing (which articles have cited my articles?) then proceed to the "Citations" tab and click "Search for all citing articles in Lens.org"
- You can download in RIS format a list of your input articles, referenced articles, and citing articles for easy integration with your reference/review management workflow
- Once you have finished citation chasing, why not check out the citation network visualisation in the "Network" tab? Please be aware that large citation networks (many or highly cited starting articles) may take considerable time to load.

Developers

Note that you can now refer users to citationchaser with a preloaded set of article identifiers, which will be initially run, ready for citationchasing. Simply concatenate the following URL stem with a comma-separated list of identifiers: [https://estech.shinyapps.io/citationchaser/?dois=\[doi1\],\[doi2\],\[doi3\]&pmids=9pmid1\],\[pmid2\],\[pmid3\]](https://estech.shinyapps.io/citationchaser/?dois=[doi1],[doi2],[doi3]&pmids=9pmid1],[pmid2],[pmid3]). Any identifier type can be used together for multiple concurrent searches.

This works for DOIs ('dois='), PubMed IDs ('pmids='), PMC IDs ('pmcids='), CORE IDs ('coreids='), and Microsoft Academic IDs ('magids='), as per the following worked example:

FIGURE 2 Screenshot of the *citationchaser* landing page, showing detailed instructions and information on data sources and citation [Colour figure can be viewed at wileyonlinelibrary.com]

instructions for developers on how to structure referral URLs to pre-populate the article input field with identifiers.

Users can directly paste identifiers into the article input fields in the 'Article input' tab (Figure 3). *Citationchaser* accepts digital object identifiers (DOIs), COREIDs (the CORE repository identifiers), MAGIDs (Microsoft Academic Graph identifiers), PMIDs (PubMed identifiers), and PMCIDs (PubMed Central identifiers). A recent update allows multiple IDs to be searched concurrently by pasting them into the relevant box on the 'Article input' tab.

As a first step, the users must confirm the records they entered have been correctly identified in The Lens.org database. The table of retrieved starting records is downloadable as an RIS file.

Once this is completed, users can then proceed to the references (backward citation chasing) and/or citations (forward citation chasing) tabs to retrieve all of the corresponding records obtainable. For each, a textual summary is provided below the action button summarising the number of records obtained, the set of unique (i.e., deduplicated) records retrieved from citation chasing, and the number of records shown in the table and downloadable as an RIS file.

In an additional step, users can then visualise their citation network using the 'Network' tab (Figure 4),

which provides an interactive visualisation of which retrieved articles are cited by (red dots) and cite (blue dots) the starting articles (black dots). Users can zoom in and out, and by clicking on an article, they can navigate to the relevant record in The Lens.org in a new browser tab.

Developers wishing to direct users from a reference/review management tool or other evidence synthesis technology can refer directly to *citationchaser* by including a list of identifiers in the suffix of the app URL using the following syntax:

```
https://estech.shinyapps.io/citationchaser/?dois=[doi1],[doi2],[doi3]&pmids=[pmid1],[pmid2].
```

Where the [doiX/pmidX] is replaced with a string of identification codes (e.g., <https://estech.shinyapps.io/citationchaser/?dois=10.1038/s41559-020-01295-x,10.1371/journal.pone.0138237&pmids=32652585,32706299>). Other identifiers can also be used together for a simultaneous search across multiple identifier types: PubMed IDs ('pmids='); PMC IDs ('pmcids='); CORE IDs ('coreids='); and Microsoft Academic IDs ('magids='). Each new set of identifiers should be prefixed with '&'. The resultant landing page from a referral URL is the set of retrieved starting articles. Users can then verify these records, and retrieve backward and forward citation chasing results, finally also visualising their network if so desired.

The figure consists of three screenshots from the citationchaser web application. The top screenshot shows the 'Article input' tab, where users can enter article titles and search for them in the Lens.org database. It includes fields for 'Input article title', 'Public location (PMID)', and 'Publication identifier (PMID)'. Below these are sections for 'Input article identifiers (PMID)', 'Cross-references (DOI/URL)', and 'Lens.org identifiers (EMBASE)'. The middle screenshot shows the 'References' tab, displaying a table of references with columns for authors, year, title, source title, volume, issue, start page, end page, and doi. The bottom screenshot shows the 'Citations' tab, displaying a table of citations with the same columns as the references tab.

FIGURE 3 Screenshots of the *citationchaser* article input (top), references (bottom left) and citations (bottom right) tabs [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]

3.2 | R package (and local Shiny app)

The backend functionality behind *citationchaser* that checks records with The Lens.org API and formulates the RIS outputs is also provided as an R package (see the GitHub repository; <https://github.com/nealhaddaway/citationchaser>). This R package has the same core functionality to the *citationchaser* web-based Shiny app, but it also provides the search results as a dataframe converted from the complete Lens.org API query response. This provides far more fields than are encoded into the RIS files (e.g., author identifiers and affiliations, for full details see <https://docs.api.lens.org/response-scholar.html>). Users wishing to retain all citing and reference article identifiers before internal deduplication should make use of the R package, which provides this information in the dataframe outputs.

The R package also requires the user to have obtained an API token from The Lens.org (<https://www.lens.org/lens/user/subscriptions#scholar>). The GitHub version of the R package also includes a locally deployable Shiny app, providing a user interface that is perhaps easier to use than running the code in the R console. This Shiny app will not appear in any version of the R package that is archived on the R package repository CRAN (<https://cran.r-project.org/>) but will be accessible via the GitHub repository.

Currently, users can load the GitHub version of the *citationchaser* R package with the following code:

```
library(devtools)
install_github("nealhaddaway/citationchaser")
library(citationchaser)
```

4 | DISCUSSION

4.1 | Integrating citationchaser into systematic searching

We believe that *citationchaser* can fit well into existing common practice in systematic searching (and systematic reviews) in the following pipeline of activities:

1. Users identify a set of relevant articles they wish to use as starting records for forward and/or backward citation chasing to supplement bibliographic database searching (e.g., this may be a final set of included articles in a review, a list of reviews encountered during screening, or a list of articles of known relevance submitted by experts and stakeholders).
2. Unique identifiers (e.g., DOIs) for these articles are entered as comma-separated lists into the relevant fields of the 'Article input' tab of *citationchaser*. The

Visualise the citation network

Once you have loaded your input articles, you can visualise your network.



The network visualisation may take a few moments to generate. Zoom in and out using your mouse wheel or two fingers on a trackpad. Move around the network by clicking and dragging. Click on a node to see the record on Lens.org.

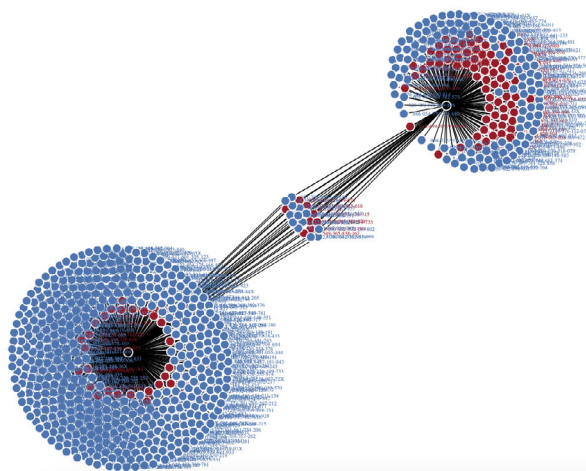


FIGURE 4 Screenshot of the *citationchaser* network diagram tab, showing the visualisation of four example starting records [Colour figure can be viewed at wileyonlinelibrary.com]

user checks that all articles have been found within The Lens.org database.

3. Backward and forward citation chasing results are obtained from the ‘References’ and ‘Citations’ tabs: the results are downloaded as RIS files.
4. Users use another reference/review management software to deduplicate the citation chasing results against bibliographic database search results, retaining only unique records in the citation chasing results. This step is optional and may instead be replaced with full screening of all *citationchaser* results by one or more reviewers if required.
5. The unique citation chasing results represent potentially relevant records missed by bibliographic database searching and can be screened for relevance alongside other digitised bibliographic records.

4.2 | Potential and planned further developments

4.2.1 | Allow multiple levels of citation chasing

It is implicitly accepted that most citation chasing goes only one level deep (articles that cite or are cited

by a starting record). However, and as demonstrated by one methods paper,⁶ by going multiple levels forward and backward, making use of automated deduplication within citation chasing results and against bibliographic search results, a comprehensive body of potentially missed records could be obtained. At present, *citationchaser* goes only one level deep, but this could be expanded. Prior to deduplication, the body of potentially relevant records for backward citation chasing multiplies at each level by the average number of references per paper (perhaps around 50), and for forward citation chasing it would depend on the field and age of the articles, but could multiply by substantially more—at present, this may be too computationally complex, but this is likely to change in the coming years. Therefore, it becomes increasingly important to build in efficiency to *citationchaser*, and to make use of efficient deduplication and screening tools (e.g., machine learning screening algorithms) for each additional level searched. In addition, building in analysis of co-citation and other indirect citation chasing options could help to deal with the exponential increase in the volume of records beyond one level deep. Other tools exist for this independent co-citation analysis, but as yet no tool allows for

both direct citation chasing and indirect co-citation analysis.

4.2.2 | Upload an RIS file

In addition, it would be useful to allow users to upload an RIS file that could be used as a basis for extracting identifiers for the starting set. This has not been prioritised in *citationchaser* because there is an on-going project to fill gaps in RIS and other bibliographic files, and a core function of this package (*bibfix*; <https://www.eshackathon.org/software/bibfix.html>) will be to obtain record identifiers for those lacking them. This would drastically increase the efficiency of an RIS upload function in *citationchaser*, and so this functionality will be developed once *bibfix* is available.

4.2.3 | Deduplicate against an RIS file

the process of deduplicating a large set of bibliographic (and other) search results against the citation chasing output RIS files is somewhat complex, since it is ONLY the unique citation chasing results not found in the larger set that should be retained (see Figure 5). Traditional deduplication tools would only remove the overlap, retaining all unique records, so users must go one extra step to tag and remove all records from the larger comparison step (the bibliographic search results in Figure 5). We hope to be able to provide a powerful deduplication tool that would automate this process with a high degree of accuracy, integrating tools like CrossRef for better record disambiguation where non-exact duplicates exist.

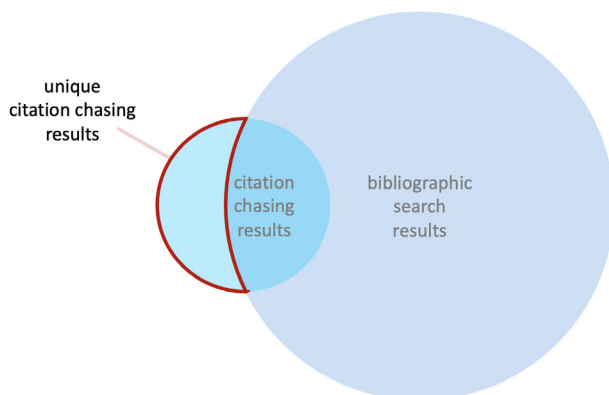


FIGURE 5 Schematic showing the body of results required from citation chasing deduplication against bibliographic search results in a systematic search [Colour figure can be viewed at wileyonlinelibrary.com]

4.2.4 | Integrating other data sources

The Lens.org database aggregates data from five bibliographic sources, but Microsoft Academic Graph (MAG) is the largest component (193 million records of the 236 m in the entire database, with 124 m from CrossRef, 32 m from PubMed and PubMed Central and 16 m from CORE—the degree of overlap is not reported). Microsoft has recently announced that it will be retiring MAG from the end of 2021, but several organisations have begun to provide replacements aiming for minimal disruption (e.g., OpenAlex, <https://blog.ourresearch.org/openalex-update-june/>). It is anticipated that this will not have any effect on backward citation chasing, although comprehensiveness of the database and the completeness of forward citation chasing from the end of 2021 onwards may be temporarily affected. The Lens.org already has plans in place to supplement current data sources until a replacement is integrated (most likely, OpenAlex; <https://about.lens.org/the-lens-scholarly-metarecord-strategy-beyond-microsoft-academic-graph/>). In the future, *citationchaser* could integrate other APIs that support bulk exports from their APIs as they become available, should they not be already integrated into The Lens.org as an internal replacement for MAG. Building in source redundancy would reduce the risk of problems in data supply from any one source.

5 | CONCLUSIONS

Citation chasing is a useful supplementary source of information in systematic searches inherent to rigorous evidence syntheses.³ However, currently there is little evidence that it is being used in rigorous systematic reviews conducted according to the standards of leading evidence synthesis coordinating bodies like Cochrane. Given the current systems and frameworks for backward and forward citation chasing, recommendations to integrate comprehensive citation chasing into robust systematic reviews (e.g., via review guidance published by the Collaboration for Environmental Evidence and the Campbell Collaboration) are likely to be ignored due to the method's time demands.

Here, we present a free-to-use, open source tool for conducting rapid backward and forward citation chasing. Furthermore, we show that it produces standardised output files that can be readily and efficiently merged with bibliographic database search results to minimise duplication while maximising the comprehensiveness of the body of potentially relevant records that can be screened within an evidence synthesis.

Relative to reviewers using fully manual, non-digitised citation chasing, *citationchaser* could present a

significant time saving. We believe that the ease-of-use and efficiency of this tool will facilitate the integration of comprehensive citation chasing as a supplementary method in future evidence syntheses. We hope that evidence synthesis coordinating bodies (e.g., Cochrane, the Campbell Collaboration and the Collaboration for Environmental Evidence) will recommend the use of such open source technologies to improve and standardise rigorous information retrieval methods.

ACKNOWLEDGEMENTS

The authors wish to thank Tom Hudson from the National Institutes for Health Care and Excellence for support and advice in preparing the manuscript.

AUTHOR CONTRIBUTIONS


Neal R. Haddaway: Conceptualisation, methodology, software, validation, investigation, writing - original draft, writing - review and editing and project administration. **Matthew J. Grainger and Charles T. Gray:** Software and writing - review and editing.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Neal R. Haddaway  <https://orcid.org/0000-0003-3902-2234>

Matthew J. Grainger  <https://orcid.org/0000-0001-8426-6495>

REFERENCES

1. Gough D, Oliver S, Thomas J. An Introduction to Systematic Reviews. Sage; 2017.
2. Livoreil B, Glanville J, Haddaway NR, et al. Systematic searching for environmental evidence using multiple tools and sources. *Environ Evid*. 2017;6(1):1-14.
3. Cooper C, Lovell R, Husk K, Booth A, Garside R. Supplementary search methods were more effective and offered better value than bibliographic database searching: a case study from public health and environmental enhancement. *Res Synth Methods*. 2018;9(2):195-223.
4. Cribbin TF. Citation Chain Aggregation: An Interaction Model to Support Citation Cycling 2011:2149-2152.
5. Robinson KA, Dunn AG, Tsafnat G, Glasziou P. Citation networks of related trials are often disconnected: implications for bidirectional citation searches. *J Clin Epidemiol*. 2014;67(7):793-799.
6. Hinde S, Spackman E. Bidirectional citation searching to completion: an exploration of literature searching methods. *PharmacoEconomics*. 2015;33(1):5-11.
7. Janssens ACJ, Gwinn M. Novel citation-based search method for scientific literature: application to meta-analyses. *BMC Med Res Methodol*. 2015;15(1):84.
8. Ogilvie D, Hamilton V, Egan M, Petticrew M. Systematic reviews of health effects of social interventions: 1. Finding the evidence: how far should you go? *J Epidemiol Community Health*. 2005;59(9):804-808.
9. Gregorio GV, Gonzales MLM, Dans LF, Martinez EG. Polymer-based oral rehydration solution for treating acute watery diarrhoea. *Cochrane Database Syst Rev*. 2016;12:CD006519.
10. Ng E, Taddio A, Ohlsson A. Intravenous midazolam infusion for sedation of infants in the neonatal intensive care unit. *Cochrane Database Syst Rev*. 2017;1:CD002052.
11. Eales J, Haddaway NR, Bernes C, et al. What is the effect of prescribed burning in temperate and boreal forest on biodiversity, beyond pyrophilous and saproxylic species? A systematic review. *Environ Evid*. 2018;7(1):1-33.
12. Cooper C, Booth A, Britten N, Garside R. A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. *Syst Rev*. 2017;6(1):1-16.
13. Levay P, Ainsworth N, Kettle R, Morgan A. Identifying evidence for public health guidance: a comparison of citation searching with web of science and Google scholar. *Res Synth Methods*. 2016;7(1):34-45.
14. Wright K, Golder S, Rodriguez-Lopez R. Citation searching: a systematic review case study of multiple risk behaviour interventions. *BMC Med Res Methodol*. 2014;14(1):1-8.
15. Wang X, Welch V, Li M, et al. The methodological and reporting characteristics of Campbell reviews: a systematic review. *Campbell Syst Rev*. 2021;17(1):e1134.
16. Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. *Methodological Standards for the Conduct of New Cochrane Intervention Reviews*. Vol 3. Sl: Cochrane Collaboration; 2013: 1-14.
17. Lefebvre C, Pritchard S, Weightman A. Cochrane Information Retrieval Methods Group. *The Cochrane collaboration methods groups newsletter*.30.
18. Briscoe S, Bethel A, Rogers M. Conduct and reporting of citation searching in Cochrane systematic reviews: a cross-sectional study. *Res Synth Methods*. 2020;11(2):169-180.
19. Haddaway N, Grainger M, Gray C. Citationchaser: an R package and shiny app for forward and backward citations chasing in academic searching. 2021.
20. Demin G. expss: tables with labels and some useful functions from spreadsheets and 'SPSS'Statistics. R package version 0.8.7. 2018.
21. Wickham H. httr: Tools for Working with URLs and HTTP. 2018. R package version. 1(0):360. <http://CRANR-project.org/package=httr>
22. Ooms J. The jsonlite package: a practical and consistent mapping between json data and r objects. *arXiv preprint arXiv:14032805*. 2014;
23. Müller K, Wickham H. tibble: Simple Data Frames. R package version 2.1. 3. 2019.
24. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation 2020. R package version 08. 2020
25. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. shiny: Web Application Framework for R. R package version 1.4. 0.2. R Found Stat Comput, Vienna. <https://CRANR-project.org/package=shiny> (2017) Accessed 12 Feb 2018.
26. Xie Y, Cheng J, Tan X. DT: a wrapper of the JavaScript library 'DataTables'. R package version 0.2. 2016.
27. Sali A, Attali D. Shinycssloaders: add loading animations to a shiny output while it's recalculating. R package version. <https://github.com/daattali/shinycssloaders> 2020.

28. Dowle M, Srinivasan A. data. table: Extension of data. frame'. R package version 1.10. 4–3 (2017). 2020.
29. Wickham H, Seidel D. Scales: scale functions for visualization. R package version 1.1. 1. 2020.
30. Wickham H, Henry L. Tidy: tidy messy data. R package version 1.2. 397 2020.
31. Allaire J, Ellis P, Gandrud C, et al. Package 'networkD3'. *D3 JavaScript network graphs from R*. 2017;
32. Wickham H. Stringr: simple, consistent wrappers for common string operations (R package version 1.4. 0) [computer software]. 2018.
33. Martín-Martín A, Thelwall M, Orduna-Malea E, López-Cózar ED. Google scholar, Microsoft academic, Scopus, dimensions, web of science, and OpenCitations' COCI: a multi-disciplinary comparison of coverage via citations. *Scientometrics*. 2021;126(1):871-906.
34. Visser M, van Eck NJ, Waltman L. Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, Crossref, and Microsoft academic. *Quant Sci Stud*. 2021;2(1):20-41.
35. Tran F, Holland JE, Quesada N, et al. What evidence exists on the effectiveness of the techniques and management approaches used to improve the productivity of field-grown tomatoes under conditions of water-, nitrogen- and/or phosphorus-deficit? A systematic map. *Environ Evid*. 2021;10(1):14. doi:10.1186/s13750-021-00229-9
36. O'Leary BC, Copping JP, Mukherjee N, et al. The nature and extent of evidence on methodologies for monitoring and evaluating marine spatial management measures in the UK and similar coastal waters: a systematic map. *Environ Evid*. 2021;10(1):13. doi:10.1186/s13750-021-00227-x
37. Knight T, Price S, Bowler D, et al. How effective is 'greening' of urban areas in reducing human exposure to ground-level ozone concentrations, UV exposure and the 'urban heat Island effect'? An updated systematic review. *Environ Evid*. 2021;10(1):12. doi:10.1186/s13750-021-00226-y
38. Lugassy L, Amdouni-Boursier L, Alout H, et al. What evidence exists on the impact of specific ecosystem components and functions on infectious diseases? A systematic map. *Environ Evid*. 2021;10(1):11. doi:10.1186/s13750-021-00220-4
39. Häkkinen M, Johansson A, Sandgren T, et al. Are small protected habitat patches within boreal production forests effective in conserving species richness, abundance and community composition? A systematic review. *Environ Evid*. 2021;10(1):2. doi:10.1186/s13750-020-00216-6
40. Savilaakso S, Johansson A, Häkkinen M, et al. What are the effects of even-aged and uneven-aged forest management on boreal forest biodiversity in Fennoscandia and European Russia? A systematic review. *Environ Evid*. 2021;10(1):1. doi:10.1186/s13750-020-00215-7
41. Lorick D, Macura B, Ahlström M, Grimvall A, Harder R. Effectiveness of struvite precipitation and ammonia stripping for recovery of phosphorus and nitrogen from anaerobic digestate: a systematic review. *Environ Evid*. 2020;9(1):27. doi:10.1186/s13750-020-00211-x
42. Dick J, Carruthers-Jones J, Carver S, Dobel AJ, Miller JD. How are nature-based solutions contributing to priority societal challenges surrounding human well-being in the United Kingdom: a systematic map. *Environ Evid*. 2020;9(1):25. doi:10.1186/s13750-020-00208-6
43. Johannesdottir SL, Macura B, McConville J, et al. What evidence exists on ecotechnologies for recycling carbon and nutrients from domestic wastewater? A systematic map. *Environ Evid*. 2020;9(1):24. doi:10.1186/s13750-020-00207-7
44. Sordello R, Ratel O, Flamerie De Lachapelle F, Leger C, Dambry A, Vanpeene S. Evidence of the impact of noise pollution on biodiversity: a systematic map. *Environ Evid*. 2020;9(1):20. doi:10.1186/s13750-020-00202-y
45. Duporge I, Hodgetts T, Wang T, Macdonald DW. The spatial distribution of illegal hunting of terrestrial mammals in sub-Saharan Africa: a systematic map. *Environ Evid*. 2020;9(1):15. doi:10.1186/s13750-020-00195-8
46. Ouédraogo D-Y, Villemey A, Vanpeene S, et al. Can linear transportation infrastructure verges constitute a habitat and/or a corridor for vertebrates in temperate ecosystems? A systematic review. *Environ Evid*. 2020;9(1):13. doi:10.1186/s13750-020-00196-7
47. Rytwinski T, Harper M, Taylor JJ, et al. What are the effects of flow-regime changes on fish productivity in temperate regions? A systematic map. *Environ Evid*. 2020;9(1):7. doi:10.1186/s13750-020-00190-z
48. Arton A, Leiman A, Petrokofsky G, Toonen H, Longo CS. What do we know about the impacts of the marine stewardship council seafood ecolabelling program? A systematic map. *Environ Evid*. 2020;9(1):6. doi:10.1186/s13750-020-0188-9
49. Goulas A, Belhadi D, Descamps A, et al. How effective are strategies to control the dissemination of antibiotic resistance in the environment? A systematic review. *Environ Evid*. 2020;9(1):4. doi:10.1186/s13750-020-0187-x
50. Algera DA, Rytwinski T, Taylor JJ, et al. What are the relative risks of mortality and injury for fish during downstream passage at hydroelectric dams in temperate regions? A systematic review. *Environ Evid*. 2020;9(1):3. doi:10.1186/s13750-020-0184-0

How to cite this article: Haddaway NR, Grainger MJ, Gray CT. Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. *Res Syn Meth*. 2022;1-13. doi:10.1002/jrsm.1563