Building a Systematic Online Living Evidence Summary of COVID-19 Research

The CAMARADES COVID-SOLES Group*
Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
Kaitlyn.Hair@ed.ac.uk

*The members of the CAMARADES COVID-SOLES Group are reported before the References

Abstract

Throughout the global coronavirus pandemic, we have seen an unprecedented volume of COVID-19 research publications. This vast body of evidence continues to grow, making it difficult for research users to keep up with the pace of evolving research findings. To enable the synthesis of this evidence for timely use by researchers, policymakers, and other stakeholders, we developed an automated workflow to collect, categorise, and visualise the evidence from primary COVID-19 research studies. We trained a crowd of volunteer reviewers to annotate studies by relevance to COVID-19, study objectives, and methodological approaches. Using these human decisions, we are training machine learning classifiers and applying text-mining tools to continually categorise the findings and evaluate the quality of COVID-19 evidence.

Key words: *COVID-19; evidence synthesis; machine learning; web application; database.*

Background

The COVID-19 pandemic continues to present a major challenge for health services and society worldwide. Since the emergence of the SARS-CoV-2 virus, the research community has shown an extraordinary response to the pandemic. This volume of information and rate of publication makes it exceedingly challenging for research stakeholders (including researchers, funders, and policymakers) to efficiently identify studies relevant to their interests, evaluate the quality of those studies, and utilise their findings for health benefit (1). This "infodemic", along with the dissemination of unsubstantiated claims in both lay and social media, risks fuelling a growing distrust in science and highlights the need for an accessible resource to support public understanding of, and access to, research findings.

Evidence is incremental, and new experimental findings offer the greatest value when considered in the context of other studies that have addressed the same or related research questions in different settings. Systematic reviews capture, summarise, and critically appraise the available evidence relevant to a pre-specified research question. They are considered the most effective method of reaching a rigorous understanding of the literature, and informing decision-making (2). Unfortunately, the time taken to perform traditional systematic reviews means that the findings are often outdated by the time of dissemination. The urgent need for evidence-based treatments for COVID-19 infection combined with a rapidly accumulating COVID-19 literature has made this an even greater challenge. Automation technologies (e.g. machine learning and text-mining) can be used to reduce the time and resources required. For example, we can train a machine to classify research as relevant or not relevant to our research question, or to extract structured information from publications, at greatly reduced human effort (3-5). Such technologies facilitate the development of "Living" systematic reviews, in which new evidence is incorporated into the review as and when it becomes available (6, 7). Further, by incorporating crowdsourcing approaches to recruit and train external reviewers, a much larger team can work together to extract information from publications at a faster pace.

Building upon existing living review methodologies, we have developed and integrated a series of automation tools and methodologies for the continual collection, categorisation, and quality assessment of COVID-19 evidence from primary research studies. We have built a Systematic Online Living Evidence Summary (SOLES) of all primary research relevant to COVID-

Address for correspondence: Kaitlyn Hair, the CAMARADES COVID-SOLES Group, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh BioQuarter, 49 Little France Crescent, Edinburgh, EH16 4SB, UK.
E-mail: Kaitlyn.Hair@ed.ac.uk

19; an interactive web application, which allows users to interact with a visual summary of the curated information, interrogate the dataset, and download relevant citations filtered by study characteristic of interest. This resource is intended for use by all stakeholders in COVID-19 research, including researchers working within the field or performing rapid or systematic reviews of COVID-19 literature.

METHODS

Identifying new research papers

To retrieve up-to-date research reports we retrieve citations weekly from PubMed (National Library of Medicine), Web of Science (all available databases: Web of Science Core Collection, BIOSIS Citation Index, Current Contents Connect, Data Citation Index, Derwent Innovations Index, KCI-Korean Journal Database, MEDLINE, Russian Science Citation Index, SciELO Citation Index, Zoological Record), EMBASE (OVID), and the World Health Organisation's COVID-19 database (8). Our search terms are described in our study protocol and have been updated over time to address changes in COVID-19 research terminology (9). To identify new research from PubMed programmatically, we use the pubmedTools R package (10) developed within our group to access the Entrez application programming interface, while other records are obtained through manual searching of the platforms/databases outlined above .

Duplicate removal

To maintain a database of unique citations, we identify and remove duplicate citations (bibliographic duplicates of work published in the same journal at the same time by the same authors) identified across different databases using an automated, R-based tool developed within our research group, the automated systematic search de-duplicator (11).

Retrieving full text publications

We retrieve full-text publications using custom R code (12) to access full-text portable document formats (PDFs) where we have institutional access (University of Edinburgh). The extraction code uses digital object identifiers (DOIs) to retrieve PDF links through CrossRef, PubMed Central, and doi.org, then downloads the PDF file using the retrieved link.

Crowdsourced study annotation

To adequately capture the broad spectrum of primary COVID-19 research, we developed a schema (*Figure 1*) to classify research by type, objective, methods, and patient population/ sample type, based on previously proposed definitions (13). Using these classifications, we

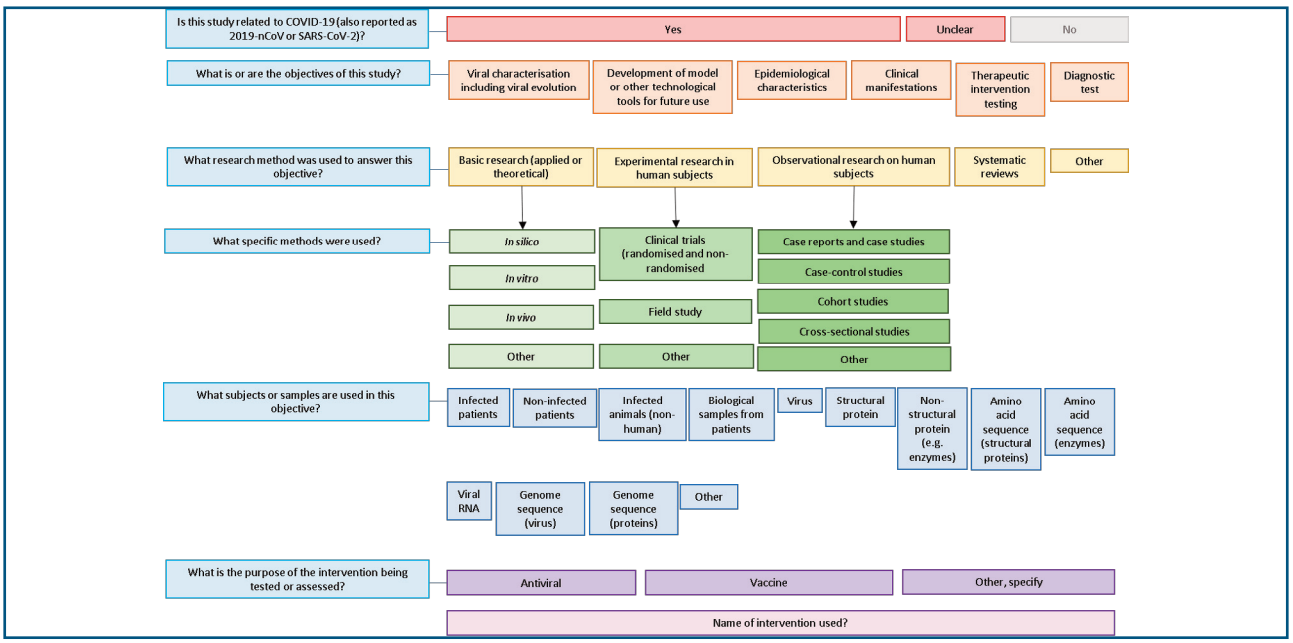

Fig. 1. *Research classification schema for primary COVID-19 studies. Arrows indicate a tree-like structure where reviews can only add subsequent annotations based on the previous annotation.*

designed a project on the Systematic Review Facility (SyRF; http://syrf.org.uk/), a widely used and freely available online platform developed within our research group (14). SyRF facilitates the conduct of large, collaborative systematic review projects and allows users to design structured annotation forms with custom questions. Once the project plan had been finalised, three independent researchers within our group annotated a test batch of 16 research papers. Through discussion, we arrived at a consensus on how each paper should be annotated. These annotations became our "gold-standard" annotated dataset used to train a crowdsourced team of human reviewers.

To recruit a team of reviewers to annotate COVID-19 research, we advertised the project via our social media profiles, existing contacts, and university research networks. Trainee reviewers were required to annotate a minimum of eight papers which were then checked against the gold-standard annotations. Once complete, we provided feedback and either asked trainees to complete more training papers or allowed them to continue as a reviewer on the main project. To ensure quality, each article is annotated by two independent reviewers. To keep reviewers up to date, fortnightly progress reports are sent out via email. Reports are generated programmatically with R code which interacts with SyRF and published online on the RPubs server as a living RMarkdown document (15).

Integration with the Systematic Review Facility

Subsets from our dataset of unique COVID-19 records are selected based on the date they are retrieved, with older records uploaded first. Custom R scripts are scheduled (using the CronR package) to periodically interact with SyRF to obtain information on the number of reviewers working on the project, the number of studies annotated, and the annotations themselves. This allows us to keep an up-to-date record of progress.

Reconciliation of annotations

For each paper, annotations from two independent reviewers are compared using a custom R script. If reviewers agree on whether the paper describes primary research relevant to COVID-19, this study is immediately classified as "included" or "excluded" – irrespective of whether they agree on all classifications. If reviewers do agree across all classifications, the study is classed as "reconciled" and those classifications are final. If there are disagreements on one or more annotations, the paper is passed to a senior reviewer who will reconcile the disagreements before submitting a final set of classifications.

Machine-assisted classification of primary studies

We used the "included" or "excluded" decisions from reconciled annotations to train a machine learning algorithm hosted by collaborators at The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), University College London. The algorithm uses natural language processing to identify features within the Title and Abstract of citations. We aimed to train it to automatically classify non-annotated studies as either "primary COVID-19 research" or "other" research.

Web application and dataset availability

We built a user interface to access our entire COVID-19 dataset via an R Shiny web application. The application allows users to visualise the annotated evidence, search the citation database (using regular expressions), and download relevant citations. The COVID-SOLES application is freely available online (16).

RESULTS

COVID-SOLES citation database

At the time of writing (May 2021) we have identified a total of 812,261 potentially relevant citations since our COVID-19 searches began in March 2020. The distribution of records retrieved from each database is shown in *Figure 2*. We obtained the highest number of records from the WHO COVID-19 database (N= 246,299) and the lowest number from PubMed (N=129,973).

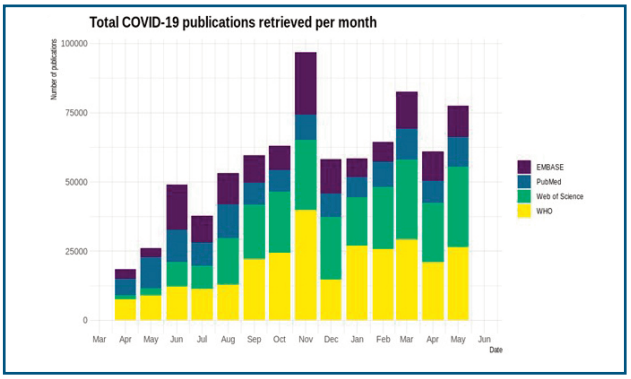

Fig. 2. *Total COVID-19 citations retrieved from each database per month.*

Following extensive de-duplication 349,726 unique citations have been identified. Over time, the number of unique publications retrieved per month has increased, with a brief levelling off period over the new year. In May 2021, we identified 50,095 publications, the largest monthly publication count yet.

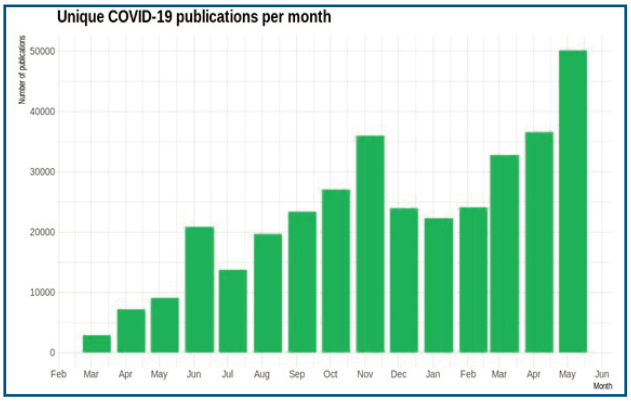

Fig. 3. *COVID-19 citations retrieved per month. Bars indicate unique citations retrieved across databases following the removal of duplicates.*

Crowdsourced annotation

We have recruited 88 trainee reviewers of which 78 have completed training and are able to annotate COVID-19 publications. The median number of papers annotated by each reviewer was 99 (interquartile range: 70.75 – 173.75). Two reviewers were particularly active, annotating 1,874 and 6,597 publications, respectively.

Machine classification of COVID-19 research

From a total of 226,417 citations in our dataset which had abstracts, 3,405 had been classified by humans as "primary COVID-19 research" (N=1312) or "other" (N=2093). This dataset was randomly split into a training set, validation set, and test set. We used a pre-set sensitivity threshold of at least >95% to ensure we captured the majority of relevant publications. On the test set (N=681), the classifier performed at a sensitivity (percentage of citations correctly included) of 95.2%, a specificity (percentage of citations correctly excluded) of 76.6%, and precision (percentage of correctly included citations from all included citations) of 71.9%. To date, the number of fully annotated primary studies is too low to train classifiers to identify specific objectives or study methodologies (N=1,174). A summary of the primary studies annotated by objective and methodology is shown in *Figure 4*. Due to our chronological approach to annotating studies, this summary reflects COVID-19 research conducted early in the pandemic, in March and April 2020.

Use of Web application

Since we developed the COVID-SOLES application, it has been accessed over 1,700 times by users from 45 countries.

LIMITATIONS AND FUTURE WORK

Optimising citation retrieval

Some retrieved citations lack useful meta-data, such as

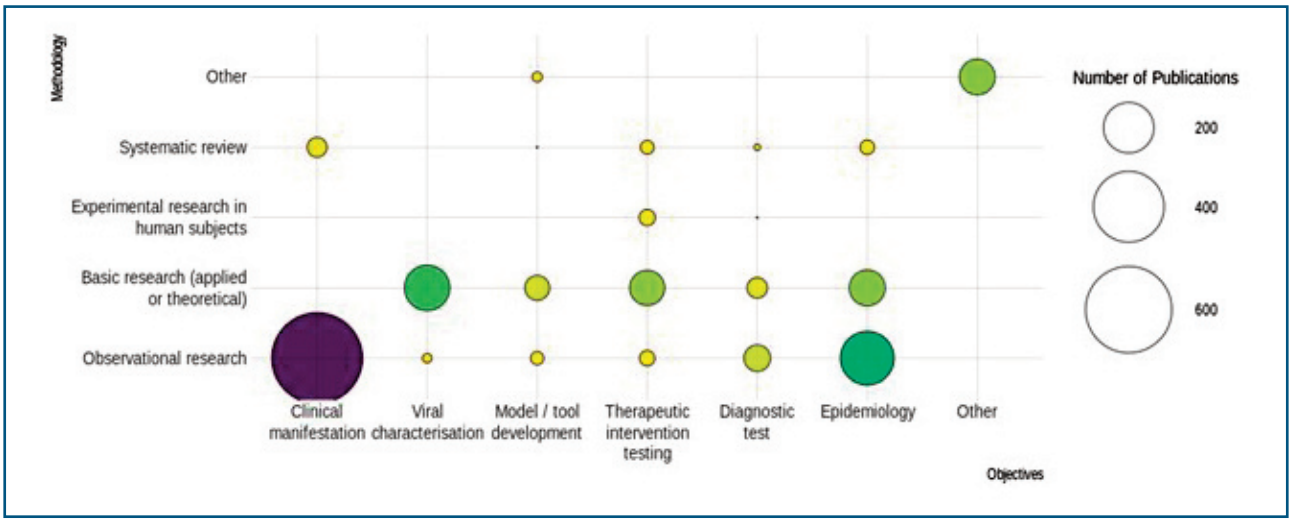

Fig. 4. *COVID-SOLES database citations annotated by objective and methodology (N=1,174). Darker colours and larger bubble sizes indicate a higher number of publications.*

DOI. This may be, in part, due to the uniquely challenging pace of COVID-19 research and our continual searching to retrieve newly published research. In some cases, we may be retrieving publications before they are fully indexed in biomedical databases. *Figure 5* indicates the percentage of unique citations retrieved from each database that lacks digital object identifiers (DOIs). Of unique records retrieved from the WHO COVID-19 database and Web of Science, 33.5% and 21.3% of citations are missing DOIs, respectively. To remedy this, we are now employing the rcrossref R package (17) to programmatically query the CrossRef database using titles and to identify the corresponding DOI information. Furthermore, we are refining our deduplication code to set a preference for retaining PubMed records over other databases, as 95.8% of citations we receive from PubMed have DOIs.

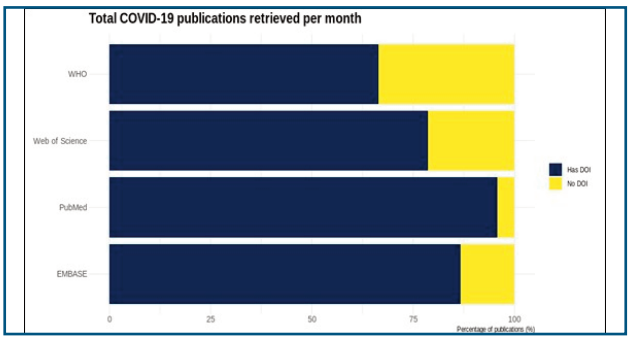

Fig. 5. *DOI status across databases searched to obtain COVID-19 citations.*

Supplementing our study type annotations

A major limitation is that we are not yet able to classify research automatically. The ability to do this as new research emerges would provide us with insights into research trends over time and identify gaps where more research may be needed. To obtain more study type annotations to drive automatic study type detection, we aim to recruit more volunteers by launching a new campaign across social media and other research networks. We are also exploring the possibility of exploiting annotation data from other openly available systematic evidence summaries of the COVID-19 literature and from published systematic reviews with accessible data. Past reviews have focused primarily on the clinical literature, so we will aim to make use of the existing data to classify human research and focus our crowd towards areas where there has been comparatively less attention e.g. in vivo research and in vitro research.

Improving our user interface

At present, some elements of the R Shiny user interface load slowly and it does not support full text searching of PDFs or Boolean searching of our database. We are currently building a new web interface to support these functionalities and sustain the growing COVID-SOLES database going forward.

Conclusion

We have developed a living workflow to synthesise COVID-19 research which enables research users to make rapid use of the currently available evidence. The SOLES workflow is sustainable, requiring minimal human effort to maintain – except the efforts of crowdsourced volunteers – and is transferrable to other research areas. We will continue to improve upon this workflow, enable more automated categorisation tools, and upgrade the user interface to enable features most useful to the evidence synthesis community.

Acknowledgements

This paper originates from a presentation at the International Collaboration for the Automation of Systematic Reviews (ICASR) meeting held in April 2021.

Members of The COVID-SOLES Group (authorship)

Project administration: Emily Sena (University of Edinburgh). **Study design:** Gillian Currie, Zsanett Bahor, Emily Sena, Malcolm Macleod, Emma Wilson, Kaitlyn Hair. **Software and architecture:** Kaitlyn Hair, Emma Wilson, Chris Sena, Can Ayder, Jing Liao, Ezgi Tanriver Ayder. **Annotation:** Joly Ghanawi, Anthony Tsang, Anne Collins, Malcolm Macleod, Alice Carstairs, Sarah Antar, Katie Drax, Kleber Neves, Thomas Ottavi, Yoke Yue Chow, David Henry, Cigdem Selli, Mariam Fofana, Martina Rudnicki, Brendan Gabriel, Esther J. Pearl, Simran S. Kapoor, Julija Baginskaite, Santosh Shevade, Alexandria Chung, Marianna Antonia Przybylska, David E. Henshall, Karina Lôbo Hajdu, Sarah McCann, Catherine Sutherland, Tiago Lubiana Alves, Rachel Blacow, Rebecca J. Hood, Nadia Soliman, Alison Harris, Stephanie L. Swift, Torsten Rackoll, Nathalie Percie du Sert, Fergal Waldron, Magnus Macleod, Ruth Moulson, Juin W. Low, Kristiina Rannikmae, Kirsten Miller, Alexandra Bannach-Brown, Fiona Kerr, Harry L. Hebert, Sarah Gregory, Isaac William Shaw,

Alexander Christides, Mohammed Alawady, Robert Hillary, Alex Clark, Natasha Jayasuriya, Samantha Sives, Ahmed Nazzal, Nimesh Jayasuriya, Michael Sewell, Rita Bertani, Helen Fielding, Broc Drury. **Senior reviewers (Reconciliation):** Joly Ghanawi, Anthony Tsang, Gillian Currie, Zsanett Bahor, Malcolm Macleod, Emily Sena.

Submitted on invitation.
Accepted on 10 June 2021.

REFERENCES

1. Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus JV. Pandemic publishing poses a new COVID-19 challenge. Nature Human Behaviour. 2020;4(7):666-9. doi: 10.1038/s41562-020-0911-0.
2. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, Al-Shahi Salman R, Chan AW, Glasziou P. Biomedical research: increasing value, reducing waste. Lancet. 2014;383(9912):101-4. Epub 2014/01/15. doi: 10.1016/s0140-6736(13)62329-6.
3. Bannach-Brown A, Przybyła P, Thomas J, Rice ASC, Ananiadou S, Liao J, Macleod MR. Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. Systematic Reviews. 2019;8(1):23. doi: 10.1186/s13643-019-0942-7.
4. Liao J, Ananiadou S, Currie GL, Howard BE, Rice A, Sena ES, Thomas J, Varghese A, Macleod MR. Automation of citation screening in pre-clinical systematic reviews. bioRxiv. 2018. doi: 10.1101/280131.
5. Bahor Z, Liao J, Macleod MR, Bannach-Brown A, McCann SK, Wever KE, Thomas J, Ottavi T, Howells DW, Rice A, Ananiadou S, Sena E. Risk of bias reporting in the recent animal focal cerebral ischaemia literature. Clinical science (London, England : 1979). 2017;131(20):2525-32. Epub 2017/10/14. doi: 10.1042/cs20160722. PubMed PMID: 29026002.
6. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, Salanti G, Meerpohl J, MacLehose H, Hilton J, Tovey D, Shemilt I, Thomas J. Living systematic review: 1. Introduction—the why, what, when, and how. Journal of Clinical Epidemiology. 2017;91:23-30. doi: https://doi.org/10.1016/j.jclinepi.2017.08.010.
7. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, Glasziou P, Shemilt I, Synnot A, Turner T, Elliott J. Living systematic reviews: 2. Combining human and machine effort. Journal of Clinical Epidemiology. 2017;91:31-7. doi: https://doi.org/10.1016/j.jclinepi.2017.08.011.
8. WHO. Global literature on coronavirus disease. Available from: https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/.
9. Currie G, Bahor Z, Liao J, Sena C, Hair K, Wilson E, Wang Q, Bannach-Brown A, Tanriver-Ayder E, Ayder C. Protocol for a "living" evidence summary of primary research related to Covid-19.
10. Liao J. pubmedTools 2021.
11. Hair K, Bahor Z, Liao J, Macleod MR, Sena ES. The Automated Systematic Search Deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. bioRxiv. 2021:2021.05.04.442412. doi: 10.1101/2021.05.04.442412.
12. Liao J. fulltext_extractor 2020. Available from: https://github.com/shihikoo/fulltext_extractor.
13. Röhrig B, du Prel J-B, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2009;106(15):262-8. doi: 10.3238/arztebl.2009.0262. PubMed PMID: 19547627.
14. Bahor Z, Liao J, Currie G, Ayder C, Macleod M, McCann SK, Bannach-Brown A, Wever K, Soliman N, Wang Q, Doran-Constant L, Young L, Sena ES, Sena C. Development and uptake of an online systematic review platform: the early years of the CAMARADES Systematic Review Facility (SyRF). BMJ Open Science. 2021;5(1):e100103. doi: 10.1136/bmjos-2020-100103.
15. CAMARADES. COVID-SOLES Progress Report. Available from: https://rpubs.com/CAMARADES/covid-soles-progress-report.
16. CAMARADES. COVID-SOLES. Available from: https://camarades.shinyapps.io/COVID-19-SOLES/.
17. Chamberlain S, Zhu H, Jahn N, Boettiger C, Ram K. rcrossref. Available from: https://github.com/ropensci/rcrossref.