New Methods and Technologies for KEEPING SYSTEMATIC REVIEWS UP TO DATE

EPPI Centre
Evidence for
Policy & Practice

James Thomas, PhD, Deputy Director, EPPI-Centre,
UCL Social Research Institute, University College London

Introduction

This Info Brief outlines innovations in bibliographic data searching that may be helpful in identifying studies for systematic reviews. James Thomas, deputy director of the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) at University College London, begins with an overview of methods currently used to search existing bibliographic databases. Next, he examines changes in the collection of bibliographic databases and recent advances in methods for retrieval of information, including a study using Microsoft Academic Graph (MAG) and OpenAlex as single-source databases. Finally, he illustrates the potential these new developments hold for the systematic review process with a case study of the EPPI-Centre's living systematic map of COVID-19 research.

LIVING SYSTEMATIC REVIEWS AND MAPS

Before the COVID-19 pandemic, "living" systematic reviews were infrequently used. A living systematic review (or map) is simply a systematic review that, from its outset, is updated as new research emerges (Elliott et al., 2017). There is not one specific approach for keeping living systematic reviews up to date; the method may vary. However, this usually necessitates re-running searches regularly. This process can yield innumerable results that require manual screening—a time-consuming task.

Conventional Approaches to Searching

Methods for searching in systematic reviews were developed during the 1990s, and they focused to a large extent on the MEDLINE database, published by the National Library of Medicine (NLM). Two types of data are usually searched: (a) the words (and parts of words) in the titles and abstracts of records; and (b) specific terms from a controlled vocabulary (Medical Subject Headings, or MeSH). Because specialists working for NLM usually assign MeSH indexing terms, there is often a delay between publication of a record in MEDLINE and assignment of its controlled vocabulary terms. Therefore, free-text searching of titles and abstracts is recommended, rather than relying on MeSH terms alone.

Searches are constructed using Boolean logic, which aims to maximize the sensitivity (recall) of the search, and precise keywords are used. Although Boolean searches are effective, especially when constructed by an information specialist, they are not infallible. Boolean searches rely on the searcher knowing the correct terminology to retrieve the records they seek, as well as anticipating variations that might be employed in unknown literature. Unless reviewers are willing to screen large volumes of mostly irrelevant records, Boolean searches are inherently fragile; if precise terminology is not used to retrieve a relevant record, it will be missed. There are no gray areas with Boolean searching: A record either meets the search criteria or it does not. Unlike Google search results, records are not ranked by relevance.

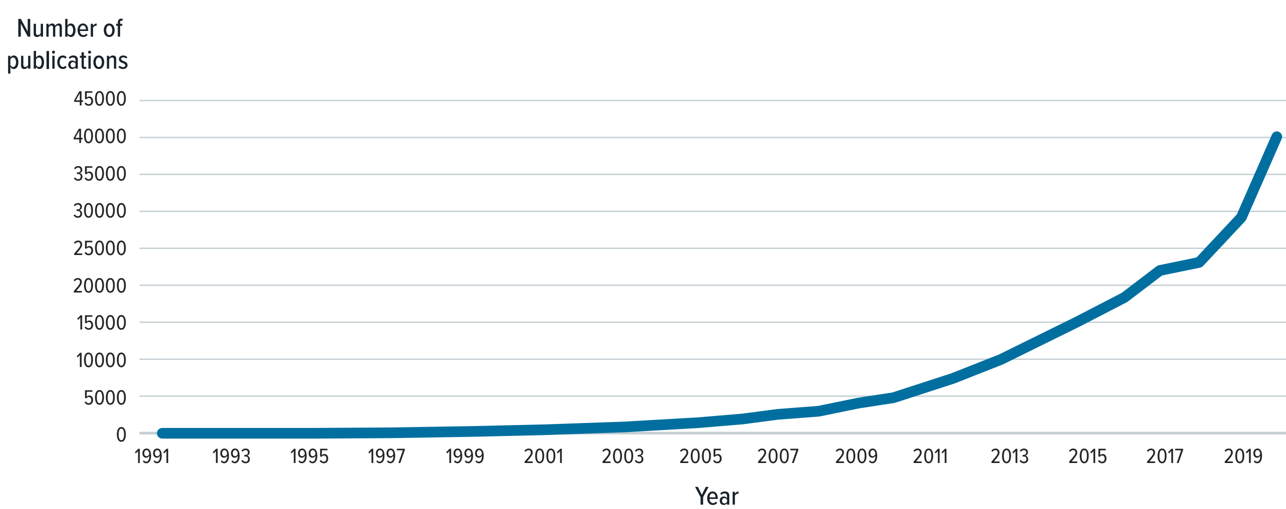
It may be convenient but not necessarily efficient to conduct most searches using online databases. Systematic reviewers may search many databases to locate records of interest. This fragmentation across database providers is the result of publishers indexing their work in different ways. For example, the American Psychological Association (APA) maintains the PsycInfo database, in which papers are indexed from many journals in the social sciences. Some of these journals—though not all—are also indexed by other providers, such as PubMed or Web of Science™ by Clarivate™.

Because each database may contain different records of interest, systematic reviewers often need to search these databases (and others) to ensure a search is comprehensive. This process involves the "translation" of a search from one database to another, since each database is built using different technologies and offers different search functionalities. It also involves the deduplication of records retrieved from different databases. In addition, this work requires paying for subscriptions to multiple database providers, thus making systematic reviews cost-prohibitive for many organizations. Database providers, such as APA and Clarivate, derive a significant amount of their income from multiple databases through which bibliographic data are scattered. To protect this income, they have resisted calls for citation data to be made openly accessible.

Increased Work Burden

The rates of publication of research findings have grown quickly over the last 10 to 20 years. During this period, interest in using research increased, along with a rise in the number of systematic reviews published. In 2010, Bastian and colleagues estimated that 75 randomized trials and 11 systematic reviews were published every day (Bastian et al., 2010). According to recent estimates, 142 randomized trials were published each day throughout 2020 (Marshall et al., 2020). In the same year, the number of systematic reviews increased dramatically, with more than 100 reviews published per day (Figure 1).[1]

Figure 1. Growth in publication of systematic reviews



Source: Data from Microsoft Academic, April 2021.

[1] These figures are likely to be considerable underestimates. They were generated from a simple search using the terms "systematic review" OR "meta-analysis" in Microsoft Academic in April 2021.

This increase in research publication makes it increasingly difficult to keep reviews up to date. For example, if each systematic review published in 2020 required the manual checking of 3,000 search results, then collectively, the authors of those reviews would have manually checked a staggering 120,000,000 references. Therefore, it is not surprising that systematic reviews are not kept up to date, and that some may be outdated as soon as they are published (Shojania et al., 2007).

The COVID-19 pandemic has heightened awareness of the need to stay up to date with evidence as soon as it has been published. However, with more than 1,000 new papers on COVID-19 being published every day, this task is time consuming and labor intensive. For example, the COVID-NMA initiative, supported by the World Health Organization and Cochrane, involves looking only at randomized trials to maintain the project's systematic review. By the end of 2020, COVID-NMA researchers had manually checked more than 45,000 records for eligibility. Researchers on another project, based at the EPPI-Centre in London, had screened more than 100,000 records. On yet another project, the Cochrane COVID-19 Study Register, researchers had screened over 200,000 studies at the time this paper was written.

With the need for thousands of other project teams to find relevant research amid the mass of COVID-19 papers (with fewer than one third of those teams reporting research results), the use and evaluation of novel tools and technologies has grown significantly in the last year. Both machine learning and crowdsourcing have been employed in different ways to ease the burden on review teams, and the *Journal of EAHIL* (*Journal of the European Association for Health Information and Libraries*) recently published a special issue describing some of these efforts (Wiley, 2021). For example, the project based at the EPPI-Centre in London uses machine learning and evaluated the utility of broader changes in database provision, which may have far-reaching effects on systematic review methodology (Wiley, 2021).

Changes in the Landscape of Bibliographic Database Provision

As mentioned earlier, systematic reviewers usually need to search multiple databases because no single database is sufficiently comprehensive. Google Scholar is the long-standing exception to this rule. This dataset, which contains all the bibliographic records that are likely to be included in any systematic review, is widely regarded as the most comprehensive in the world.[2] Unfortunately, however, Google Scholar's unsophisticated search engine and comparatively primitive user support for downloading references make it impossible for users to search Google Scholar systematically. Its lack of a Boolean search engine has been criticized as rendering it unsuitable for use in systematic reviews (Gusenbauer & Haddaway, 2020). The dataset is also "closed"—that is, it cannot be downloaded for use by third parties.

Between 2015 and 2021, a similar initiative by Microsoft, called Microsoft Academic Graph (MAG), challenged assumptions about the possibility of systematic reviewers being able to carry out a single search through a single database. While MAG's user interface was like that of Google Scholar and could not be used for systematic searching, its more than 250 million records were available for use under a Creative Commons license (Sinha et al., 2015). This was a potentially game-changing contribution because MAG's dataset identified research records through the Bing web crawler and, therefore, had the potential to be as comprehensive as Google Scholar. It was also a "graph" of data, illustrating citation, authorial, institutional, and semantic relationships among papers—valuable information for machine learning. Additionally, MAG was updated biweekly with new papers and catalyzed the creation of several new search engines and services[3] that claim to offer up-to-date, comprehensive views of evidence in ways that the conventional, "siloed" databases have never done. While Microsoft did not provide a standard Boolean search interface of the type that systematic reviewers use, the fact that it made the dataset available means that third-party providers potentially could do this—or use newer machine-learning technologies to facilitate searching.

[2] Specialist data, such as records from trial registries and clinical trial report databases, excepted.

[3] For example, Semantic Scholar, the Lens, and AMiner use the MAG dataset.

Unfortunately, despite its avowed commitment to supporting open data and the significant impact that this work has had, Microsoft discontinued its Microsoft Academic project at the end of 2021 (Microsoft, 2021). Fortunately, the OpenAlex database (https://openalex.org/), operated by OurResearch (https://ourresearch.org/), went live on January 3, 2022, picking up where Microsoft Academic left off. It took as its starting dataset the more than 200 million academic papers in Microsoft Academic and implemented its own system for keeping this database up to date. OpenAlex is firmly embedded in the "open science" movement and therefore is particularly strong in this area, and it makes all its data and source code openly available for anyone to use.

RESEARCH AND DEVELOPMENT:
Can MAG Be Used as a Single Source for Locating Studies for Systematic Reviews?

The MAG dataset (now updated by OpenAlex)[4] might have the potential to serve as a single source of references for systematic reviews, though research is needed to determine whether it *can* function as a single-source database. In addition, because the dataset contains more than 200 million records, it would be necessary to establish whether records could be identified with sufficient accuracy. We used a case study of a collaborative project between the EPPI-Centre and the University of York, to answer two parallel research questions:

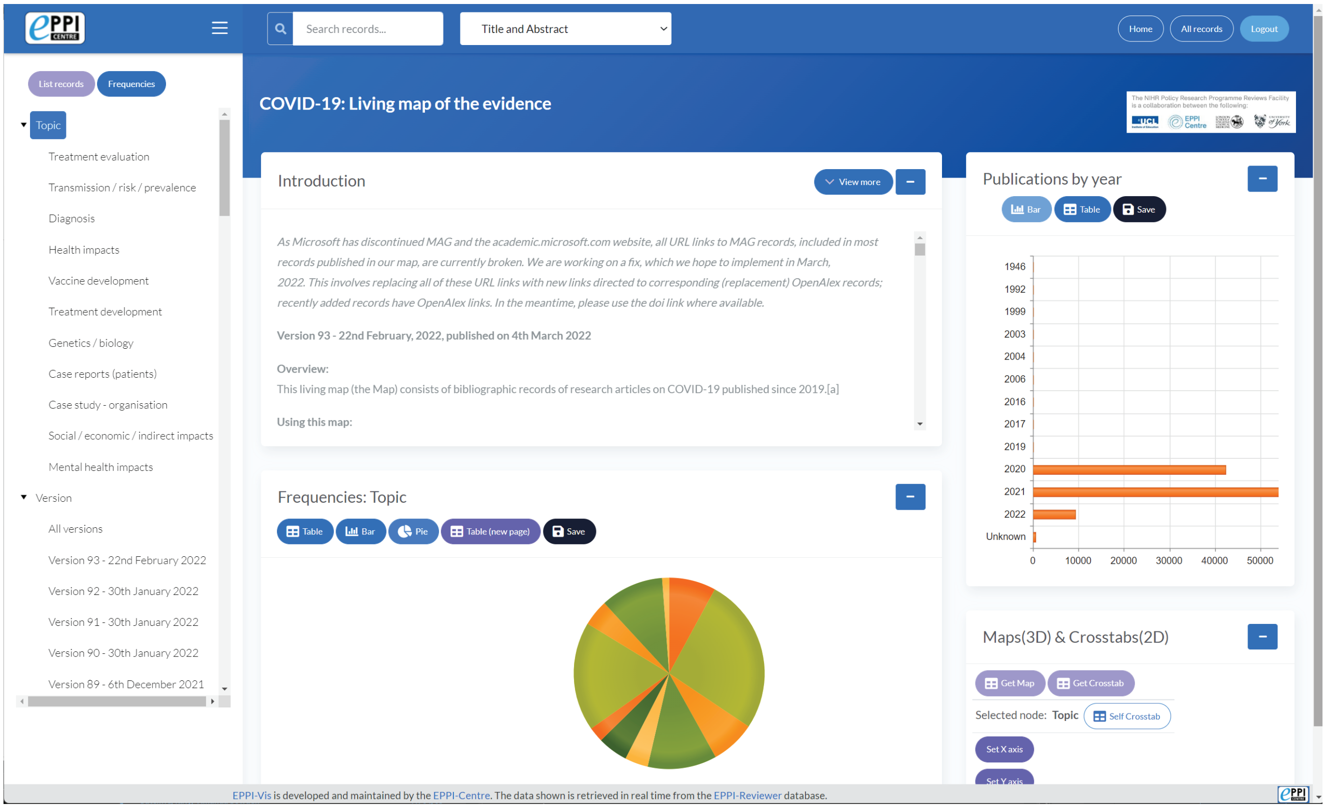
1. Does the MAG dataset contain all the studies that we need? *(sensitivity)*
2. Can relevant studies be identified efficiently? *(specificity)*

This project involved ongoing surveillance of the emerging COVID-19 research literature in a living systematic map. Since the end of February 2020, we have performed weekly searches in the MEDLINE and Embase® databases, deduplicated the results, and then manually screened the results in EPPI-Reviewer (Thomas et al., 2020). Records that meet the inclusion criteria for the map are then allocated to a specific category (for example, *diagnosis* or *treatment*).

[4] This evaluation was carried out while Microsoft was still maintaining Microsoft Academic, so it will be referred to as the "MAG dataset" throughout this section.

The map is updated and published online each week. The user interface is shown in Figure 2.

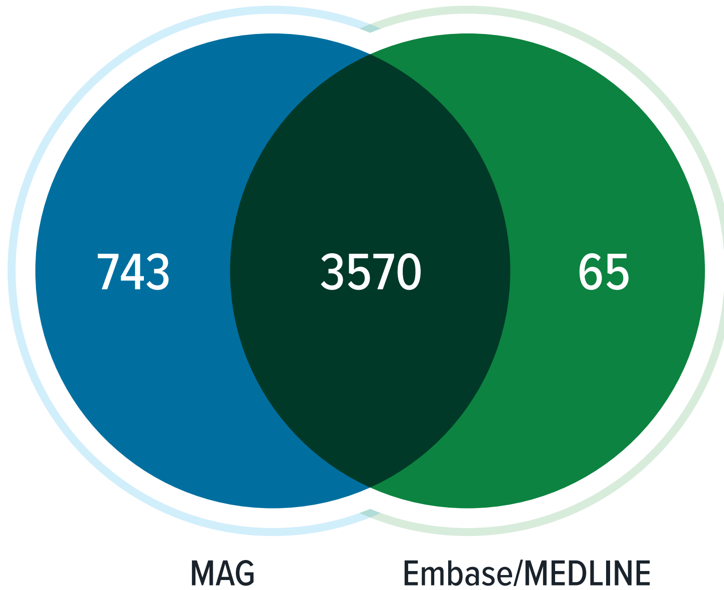
Figure 2. COVID-19 living systematic map



Source: Lorenc et al., 2020.

We answered our first research question (sensitivity) by comparing the records we retrieved in our MEDLINE/Embase searches in June 2020 with those available in MAG. We found that nearly all the records in MEDLINE/Embase (99%) also were available in MAG (99%). In addition, we found that MAG contained 743 records that we did not find in our MEDLINE/Embase searches. This discovery was concerning. Although we searched what are widely considered to be the main sources of research in health, we achieved a recall of only 83% of the records that we sought. The overlap (and non-overlap) of records is depicted in Figure 3.

Figure 3. Coverage of MEDLINE/Embase compared with MAG

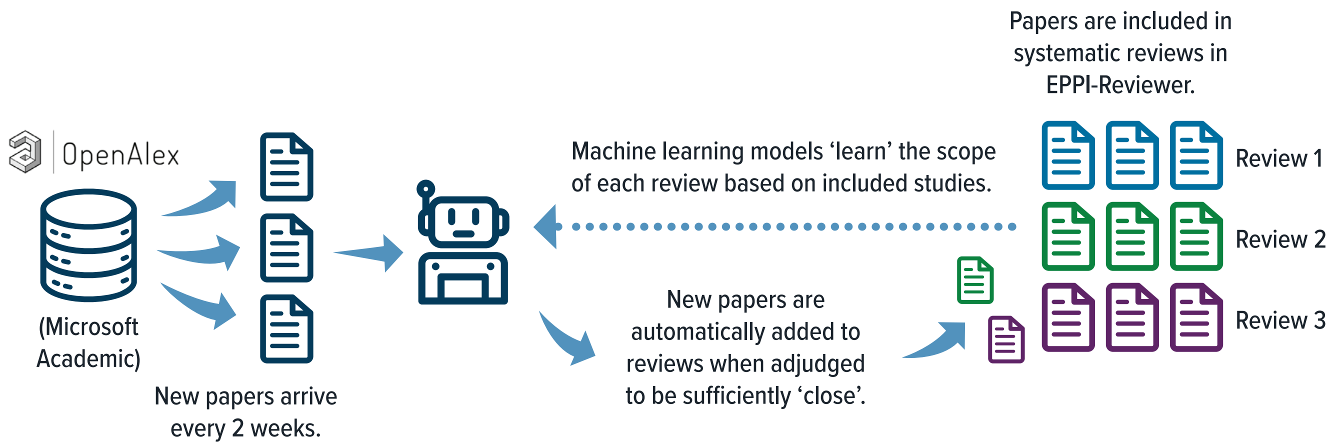


Source: Diagram generated by author.

In addition to these quantitative findings, we observed qualitatively that the MAG records contained more non-English language results than those from our MEDLINE/Embase search. Because MAG data are generated through a web crawl, it may be that MAG is not only more comprehensive than conventional databases but also is less geographically biased than the data sources traditionally used in systematic reviews.

To find relevant studies to update our living map in the MAG dataset, we used machine learning rather than the conventional Boolean search. It is possible to build a full-featured Boolean search on a dataset the size of Microsoft Academic, but doing so would require substantial effort and computational time. Boolean searching also would be restricted to text from titles and abstracts, because no controlled vocabulary (such as MeSH) has been applied consistently across such a large and varied collection of documents. In this process (depicted in Figure 4), new papers are automatically fed into the system biweekly when they arrive in our copy of Microsoft Academic. Every 2 weeks, the papers already included in reviews that are "subscribed" to the auto-update process are used to build machine-learning models against which every new paper is "scored." A "distance" metric is calculated between each new paper and each review. If a review is sufficiently "close" to a new record, then the new record is automatically added to a list of records that require manual checking for that review.

Figure 4. EPPI-Reviewer auto-update workflow



Source: Diagram created by author.

Rather than giving us a simple *found/not found* result, as is the case in a Boolean search, use of machine learning enables us to rank items in order of relevance. We conducted a detailed analysis that compared using a Boolean search of MEDLINE/Embase to employing machine learning in MAG to identify the most cost-effective way of identifying studies to include in our living map (Wiley, 2021). We found that the use of a fixed screening target (in which the team screens a fixed number of records each week—in our case, 1,500), combined with the machine-learning process and MAG as the data source, was the most cost-effective solution. The team adopted this workflow in November 2020.[5]

Discussion

For maintaining our living systematic map of COVID-19, we found that using Microsoft Academic as a single-source database is more cost-effective than searching MEDLINE and Embase and then following conventional methods for deduplicating and screening results. The literature search tool yields higher numbers of relevant records at a lower cost. We use OpenAlex (now that Microsoft Academic has closed) every week to maintain a map of the health-related COVID-19 research literature (Lorenc et al., 2020). In addition, we now use the same workflow to maintain a second map that covers social scientific research on

[5] For full details of our cost-effectiveness analysis, please see Shemilt and colleagues (2021).

COVID-19 (Shemilt et al., 2022). We have integrated the auto-update workflow within EPPI-Reviewer, and several organizations are trialing this new tool to maintain their own systematic reviews. We are also building a larger model that will identify randomized trials as they are published and associate them with relevant Cochrane reviews. Work is ongoing, but current evaluation suggests that most Cochrane reviews can be kept up to date, with only a handful of records to examine each week.

While still in its infancy, and despite the need to replace the MAG data feed, this work suggests that we may be on the cusp of a new paradigm of searching for studies to include in systematic reviews. Rather than having to search numerous overlapping databases, it is possible to obtain equally good—if not better—results from a single source. Searching a single source is not only more efficient than searching numerous smaller databases but is also more equitable in two ways. First, it opens the possibility of running systematic searches for individuals and organizations that cannot afford subscriptions to all the databases they need. Second, it potentially overcomes a geographical and linguistic bias that affects many conventional sources. This approach looks promising for review updates, where there is an existing pool of studies from which a machine learning algorithm can "learn." If the machine learning can be further refined to maximize specificity without sacrificing sensitivity, it may be possible to keep any review up to date more easily, increasing decision makers' ability to make decisions informed by the most current evidence.

References

Bastian, H., Glasziou, P., & Chalmers, I. (2010). Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? PLoS *Medicine, 7*(9), e1000326. https://doi.org/10.1371/journal.pmed.1000326

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akle, E. A., McDonald, S., Salantih, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., & Thomas, J. (2017). Living systematic review: 1. Introduction—the why, what, when, and how. Series: Living Systematic Reviews. *Journal of Clinical Epidemiology, 91,* 23–30. https://www.jclinepi.com/article/S0895-4356(17)30636-4/fulltext

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other sources. *Research Synthesis Methods, 11,* 181–217. https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrsm.1378

Lorenc, T., Khouja, C., Raine, G., Shemilt, I., Sutcliffe, K., D'Souza, P., Burchett, H., Hinds, K., Khatwa, M., Macdowall, W., Melton, H., Richardson, M., South, E., Stansfield, C., Thomas, S., Kwan, I., Wright, K., Sowden, A., & Thomas, J. (2020). COVID-19: A living map of the evidence. EPPI-Centre, Social Science Research Unit, UCL Social Research Institute, University College London. https://eppi.ioe.ac.uk/cms/Projects/DepartmentofHealthandSocialCare/Publishedreviews/COVID-19Livingsystematicmapoftheevidence/tabid/3765/Default.aspx

Marshall, I. J., Nye, B., Kuiper, J., Noel-Storr, A., Marshall, R., Maclean, R., Soboczenski, F., Nenkova, A., Thomas, J., & Wallace, B. C. (2020). Trialstreamer: A living, automatically updated database of clinical trial reports. *Journal of the American Medical Informatics Association, 27*(12), 1903–1912. https://academic.oup.com/jamia/article/27/12/1903/5907063

Microsoft. (2021, May 4). Next steps for Microsoft Academic – Expanding into new horizons. https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach/

Shemilt, I., Arno, A., Thomas, J., Lorenc, T., Khouja, C., Raine, G., Sutcliffe, K., D'Souza, P., Kwan, I., Wright, K., & Sowden, A. (2021). Cost-effectiveness of Microsoft Academic Graph with machine learning for automated study identification in a living map of coronavirus disease 2019 (COVID-19) research. *Wellcome Open Research, 6,* 210. https://wellcomeopenresearch.org/articles/6-210

Shemilt, I., Gough, D., Thomas, J., Stansfield, C., Bangpan, M., Brunton, J., Dickson, K., Graziosi, S., Hull, P., Kneale, D., Larsson, C., Mendizabal-Espinosa, R., Muraki, S., Ramadani, F., Vigurs, C., & Wang, P. (2022). Living map of systematic reviews of social sciences research evidence on COVID-19. EPPI-Centre, UCL Social Research Institute, University College London. https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3806

Shojania, K. G., Sampson, M., Ansari, M. T., Doucette, S., & Moher, D. (2007). How quickly do systematic reviews go out of date? A survival analysis. *Annals of Internal Medicine*. https://www.acpjournals.org/doi/full/10.7326/0003-4819-147-4-200708210-00179?rfr_dat=cr_pub++0pubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). *An overview of Microsoft Academic Service (MA) and applications. Proceedings of the 24th International Conference on World Wide Web* (WWW '15 Companion), 243–246. https://dl.acm.org/doi/10.1145/2740908.2742839

Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., & Bond, M. (2020). EPPI-Reviewer: Advanced software for systematic reviews, maps and evidence synthesis. EPPI-Centre, UCL Social Research Institute, University College London. https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2914

Wiley. (2021). *Journal of the European Association for Health Information and Libraries (EAHIL), 17*(2), 11–15. http://ojs.eahil.eu/ojs/index.php/JEAHIL/issue/view/187/17_2

The contents of this document were developed under grant number 90DPKT0001 from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this document do not necessarily represent the policy of NIDILRR, ACL, HHS, and you should not assume endorsement by the Federal Government.

About AIR

Established in 1946, with headquarters in Arlington, Virginia, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance to solve some of the most urgent challenges in the U.S. and around the world. We advance evidence in the areas of education, health, the workforce, human services, and international development to create a better, more equitable world. For more information, visit AIR.ORG.

AIR®

AIR® Headquarters
1400 Crystal Drive, 10th Floor | Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

17255 v5_5/22