ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

TECHNICAL REPORT

Use and impact of new technologies for evidence synthesis

Literature review and qualitative data collection

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Use and impact of new technologies for evidence synthesis

Literature review and qualitative data collection

This report was commissioned by the European Centre for Disease Prevention and Control (ECDC), coordinated by Helena de Carvalho Gomes, and produced by RAND Europe under contract ECD.12594 implementing the ECDC framework contract for services ECDC/2019/028

Authors
Lucy Hocking (RAND Europe), Sarah Parkinson (RAND Europe), Avery Adams (RAND Europe), Emmanuel Nielsen (RAND Europe), Cecilia Ang (RAND Europe), Helena de Carvalho Gomes (ECDC).

Acknowledgements
We are grateful for the contributions of workshop, interview and survey participants:

Petra Apfalter (Institute for Hygiene, Microbiology and Tropical Medicine at Ordensklinikum Linz and BioLab GmbH, AT, ECDC Advisory Forum member), Andrew Booth (University of Sheffield, UK), Ana Maria Correia (National Institute of Health Ricardo Jorge (INSA Porto), PT, ECDC Advisory Forum alternate), Agnes Danielisz (National Public Health Centre (NNK), ECDC National Focal Point for scientific advice coordination, HU), Isabel de la Fuente (Centre Hospitalier de Luxembourg, LU, ECDC Advisory Forum member), Celine Dumas (Santé publique France, FR), Julian Elliott (Cochrane, AUS), María-José Faraldo-Vallés (Galician Health Servcie (SERGAS), ES), Lucía García San Miguel (Coordinating Centre of Sanitary Alerts and Emergencies, Ministry of Health, (CCAES), ECDC National Focal Point for scientific advice coordination, ES), Chantelle Garritty (Public Health Agency of Canada, CA), Candyce Hamel (Ottawa Hospital Research Institute, CA), Trish Harrington (Health Information and Quality Authority (HIQA), IE), Otto Helve (Finnish Institute for Health and Welfare (THL), ECDC National Focal Point for scientific advice coordination, FI), María Soledad Isern de Val (Aragon Health Sciences Institute (IACS), ES), Jari Jalava (Finnish Institute for Health and Welfare (THL), FI), Hanan Khalil (La Trobe University, AUS), Petra Klepac (National Institute of Public Health (NIJZ), SI), Mirjam Knol (National Institute for Public Health and the Environment (RIVM), NL), Natalija Kranjec (National Institute of Public Health (NIJZ), SI), Brian Kristensen (Statens Serum Institut (SSI), DK), Elaine Claire Lautier (Ministry of Health, MT), Tuija Leino (Finnish Institute for Health and Welfare (THL), FI), Malcolm Macleod (CAMARADES group, The University of Edinburgh, UK), Tanya Melillo (Ministry of Health, MT, ECDC Advisory Forum alternate, National Focal Point for scientific advice coordination), John Middleton (Wolverhampton University, UK, ECDC Advisory Forum observer representing the Association of Schools of Public Health in the European Region (ASPHER)), Ashley Muller (Norwegian Institute of Public Health (FHI), NO), Zachary Munn (University of Adelaide, AUS), Alison O'Mara-Eves (EPPI-Centre, UCL, UK), Georgios Panagiotakopoulos (National Public Health Organization (EODY), GR, ECDC Advisory Forum alternate National Focal Point for scientific advice coordination), Dimitrios Paraskevis (National Public Health Organization (EODY), GR), Uli Ronellenfitsch (Martin-Luther University, DE), Ana João Santos (National Institute of Health Ricardo Jorge (INSA), PT), Anne-Mae Scott (Institute for Evidence-Based Healthcare, Bond University, AUS), Nadja Sinkovec Zorko (National Institute of Public Health (NIJZ), SI), Carl Fredrik Sjöland (Swedish Public Health Institute, SE), Michaela Špačková (National Institute for Public Health (SZU), CZ), Britta Tendal Jeppesen (Danish Health Authority (SST), DK), James Thomas (EPPI-Centre, UCL, UK), Caroline van den Ende (National Institute for Public Health and the Environment (RIVM), NL), Sophie Vaux (Santé publique France, FR), Didrik Vestrheim (Norwegian Institute of Public Health (FHI), NO), Albert Vollaard (National Institute for Public Health and the Environment (RIVM), NL), ECDC participants: Barbara Albiger, Olivier Briet, Kim Brolin, Emma Cattermole, Stanislav Danchev, Rok Grah, Giorgos Kaltekis, John Kinsman, Spyros Ktenas, Kate Olsson, Ajibola Omokanye, Sari Palojoki, Diamantis Plachouras, Jeanine Pommier, Johanna Takkinen.

Suggested citation: European Centre for Disease Prevention and Control. Use and impact of new technologies for evidence synthesis. Stockholm: ECDC; 2022.

Stockholm, November 2022

ISBN 978-92-9498-600-9
doi: 10.2900/592449
Catalogue number TQ-03-22-169-EN-N

Contents

Abbreviations

AACODS	Authority, Accuracy, Coverage, Objectivity, Date, Significance
AI	Artificial Intelligence
CONSORT	Consolidated Standards of Reporting Trials
ECDC	European Centre for Disease Prevention and Control
EEA	European Economic Area
EU	European Union
ICASR	International Collaboration for the Automation of Systematic Reviews
RCT	Randomised Controlled Trial
SANRA	Scale for the assessment of narrative review articles
UK	United Kingdom
WHO	World Health Organization

Glossary

Citation searching, backward or forward (also called backward or forward snowballing)	The process of looking at the reference list of a publication to identify additional older references (backward). The process of identifying new references by tracking papers that cite a publication (forward).
Evidence synthesis/literature review	The process of bringing together data and evidence from multiple studies or sources of information to reach a conclusion on a topic. Here, we interpret this broadly to encompass a range of evidence synthesis approaches, including but not limited to, systematic reviews, meta-analyses, narrative reviews and living reviews.
Extraction	The process of identifying relevant information and/or data from articles that are included in a literature review. Often conducted using a standardised template.
Living review	A review which is updated periodically
Quality appraisal/risk of bias assessment	The process of assessing the quality or validity of a study, often using standardised criteria
Screening	The process of reviewing articles identified from a literature search to determine if they meet the review inclusion/exclusion criteria

Executive summary

Introduction

It is essential to have access to timely and reliable evidence syntheses, including literature reviews, in order to inform public health decisions, including decisions related to the prevention and control of infectious diseases. However, it is not always straightforward to identify, process and synthesise the vast amount of literature available, and doing so can be time and resource intensive. Technologies that help automate or semi-automate parts of the evidence synthesis have been proposed to help address these challenges.

Study aims and methods

This report studies the implementation and impact of new technologies to support the development of evidence syntheses and aimed to answer the following research questions:

- What types of automated technologies are available and have been used for evidence synthesis, and how have these been used?
- At what stage in the evidence synthesis pathway have these automated technologies been used (e.g. literature searching, screening, data extraction)?
- What are the experiences and lessons learned from using automated technologies in evidence synthesis?
- What is the impact of using automated technology for evidence synthesis?
- What are the needs of European Union (EU)/European Economic Area (EEA) competent health authorities in conducting technology-assisted evidence syntheses and how can these gaps be filled?

To answer these questions, we conducted a systematic review of the academic and grey literature on how new technologies have been used to automate or semi-automate parts of the evidence synthesis, and the experience of individuals and organisations that have used this technology. We searched Ovid MEDLINE, Embase, Cochrane Library, CORDIS and the websites of key public health authorities in EU/EEA countries, ECDC international partners and major evidence synthesis providers. Articles were screened against a set of inclusion/exclusion criteria and relevant ones were taken forward to data extraction. In total, we included 157 articles in this review.

We also conducted a mapping exercise in which we collected qualitative data around new technologies for evidence synthesis in the form of focus groups (n= 5, total of 46 participants), interviews (n= 3) and an online survey (n= 9). The aim of this was to map the current use of automated technologies for evidence synthesis by EU/EEA public health bodies, as well as to map the experiences of using these types of tools that could not be identified by reviewing literature alone. Participants included people that had experience using new technologies for evidence synthesis, organisations that regularly conduct systematic reviews and are active in initiatives around the use of technologies, representatives from public health competent authorities in EU/EEA Member States, representatives from ECDC, evidence synthesis experts and others working in the area of public health and infectious diseases.

Results

Based on the qualitative data collected from EU/EEA public health bodies participating in this study, we found that automated technology is used to a limited extent for conducting evidence synthesis, with most participants not currently using these types of tools. This may be because they do not conduct any type of review within their organisation (instead, relying on reviews produced by others) or because of different challenges and obstacles that can arise when attempting to use automated approaches (discussed below). However, there were a few public heath bodies that used automated tools, sometimes extensively. Among these public health bodies, technologies were used in different stages of the evidence synthesis process, although they seem to be particularly used to semi-automate the screening stage of reviews. For other public health bodies, the adoption of new technologies was being explored for the future.

Overall, this study shows that new technologies can facilitate the evidence synthesis process. Most notably, the literature review and qualitative work highlighted how greater use of technologies can allow for more efficient reviews in terms of time, effort and resources, which can help organisations working in public health and infectious diseases keep up with the ever-expanding body of evidence to inform (sometimes rapid) decision making. This has been particularly relevant during the COVID-19 pandemic which has seen thousands of publications on the disease. For example, one focus group participant reported that the use of an automated tool for conducting a literature search reduced the number of articles to be screened by over 90%. The literature review found that technologies also have the potential to improve reviews by supporting increased precision and accuracy (e.g. by reducing human error or double counting, or by facilitating systematic extraction and analysis of unstructured data) and by allowing human effort to be focused on synthesis and interpretation. Study participants and the literature described how automation can also allow for more living reviews to be produced, whereby the same review is updated with new evidence periodically, as well as conduct tasks that would usually be too time-intensive to conduct manually (such as translating texts into languages other than English). Study participants also noted the benefit automated tools can provide in supporting collaboration, as many can be used by multiple researchers across different organisations/countries at the same time.

Despite the potential of new technologies, there remain challenges in using these technologies for evidence synthesis. Both the literature review and qualitative work found that human input is still required to conduct evidence syntheses, even where new technologies are also used, for tasks such as developing training sets, checking the accuracy of the tool and adding credibility to the study. One participant noted that this challenge has created obstacles to procuring new tools. An example of another challenge was shared by representatives from one public health institute where the embedding of automated technologies in the evidence synthesis process took a large amount of time and effort from staff. The need for (sometimes extensive) human input to adopt and use automated technologies, particularly when first introducing them, can be a barrier to their use, particularly due to capacity constraints within an organisation (see below).

While evidence from the literature suggested that researchers generally find automated tools easy to use and user-friendly, study participants working in public health authorities felt more support was needed in adopting the new tools. Participants discussed how the use of automated tools for reviews requires staff to both be trained and have the capacity to appropriately use the technology. Staff training requires resources on top of what organisations will spend on the technologies or licenses themselves, which can be a barrier to adopting the technology, particularly for smaller organisations or public health bodies. Participants from two organisations described the challenge of having adequate capacity to effectively introduce an automated tool despite there being a willingness to introduce it. In addition to training and capacity issues, focus group participants highlighted how adoption of new technologies represents a change to usual ways of working and there can be concerns about job security, as well as a 'culture shock' for staff.

Both the literature review and qualitative work noted that even when technologies work well in one review or for a particular topic, this does not mean that they will perform the same for other reviews or topics, which points to the need for human input to help determine whether a technology is appropriate for particular tasks. There are also concerns around transparency, trust and reproducibility, particularly for machine learning technologies, sometimes viewed as 'black box', which can prevent them from being used. Additionally, although there is evidence that particular technologies can be accurate and precise (even in comparison to human reviewers), there is heterogeneity in the performance of new technologies, which can cause concerns around generalisability. This was raised by multiple participants who described issues such as reviews not being reproducible when using automated technologies (or tools are trained for one specific review topic and may not be effective for other topics without significant time spent re-training the algorithm) or, as mentioned, needing to check the accuracy of the decision-making of automated technologies. However, there are ways to improve how technologies perform and how results from automated reviews are reported, which can help to address issues around transparency and trust.

The literature review and qualitative work identified several ways in which the use of technologies for evidence synthesis can be supported and ensure that technologies create a positive impact in terms of the ability to stay up-to-date with evidence in the areas of public health and infectious diseases. These are in the areas summarised below:

- **Training and skills:** Many organisations that may benefit from using new technologies for evidence synthesis may lack the skills to incorporate such technology. Training and support, both in terms of conducting high quality reviews in general and in the use of technology, can be helpful. Being able to share knowledge across organisations trying to implement similar new processes was also flagged by participants as being of use.
- **Time and space:** Organisations working in public health may lack the capacity to conduct reviews themselves, or to incorporate technology into these reviews. Offering staff the time (and space) to learn how to use a new tool effectively and be able to adapt to a new way of working are important.

Participants noted that smaller, more confined teams with dedicated time to tasks that use automation may be the most effective approach.

- **Leadership support:** Leadership buy-in and support in the adoption of a new technology can also be beneficial, such as creating the dedicated time to work with the technology and making it more of a priority. Leadership support may also encourage the introduction of training for staff on using the technology. The introduction of a new technology also needs careful management by leadership to ensure there is staff buy-in and staff are made aware of why the change is being introduced and how it will change their role (if at all). A participant from a public health institute discussed how this was handled effectively during the introduction of a new machine learning tool, whereby a change communication strategy was implemented (see Chapter 4).
- **Financial resources:** While some automation tools are available for free, many must be purchased and/or are associated with licence fees. Financial support for public health bodies may improve their ability to keep up-to-date with developments in their field, especially for smaller and less well-resourced organisations. Organisations could also take advantage of freely available tools, should they prove to be robust. Collaboration and relationship building with organisations that already have a technology licence could be beneficial particularly to smaller public health organisations that do not have the financial resources for automation tools.
- **Collaboration:** There are a number of active collaborations between organisations and individuals that are working to improve how technologies are used in evidence syntheses and reviews (across public health authorities and beyond, e.g. academics). Participating in these collaborations can allow organisations working in public health to contribute to this area, and gain skills and confidence in using new technologies in reviews. Coordination between EU/EEA Member States public health bodies can also allow sharing recommendations for tools, lessons and best practice, as well as shared repositories, between national public health competent authorities. Setting up these collaborations now would be particularly useful for future emergency/crisis situations (as seen with COVID-19) to facilitate rapid sharing of learning and knowledge.
- **Standards to ensure technologies are used appropriately:** One concern that prevents reviewers from using new technologies is that reviews will be perceived as being of lower quality. New/updated standards could be established to help ensure that new technologies are used appropriately to support high quality reviews, and can help ensure robustness to relevant stakeholders such as those that use evidence from reviews and academic journal publishers. Relatedly, it may also be useful to include the use of automation in reporting guidelines, such as PRISMA (the Preferred Reporting Items for Systematic Reviews and Meta-Analyses), to improve the transparency of methodology. In addition, as we found little reflection in the literature on the actual use of automated tools, reviews that use these approaches could be encouraged by journals to report and reflect on the practical and performance aspects of the tool(s), to support the sharing of learning and experiences.
- **Interventions that improve the performance of new technologies:** The performance of some technologies depends on the availability of high-quality data sets for training algorithms. Creating these datasets can help improve the performance of new technologies in particular subjects relating to public health and infectious diseases. Reporting standards and improvements in how literature is indexed and stored can improve machine-readability of literature to allow for more automation. Increased interoperability between technologies can also support automation.

Conclusions

This study explored how new technologies have been used to automate or semi-automate parts of the evidence synthesis process, the impact of using these technologies, and how the use of technology could be supported within the EU and EEA for public health purposes. The systematic review for this study provided insight into the broad experiences of those using technology for evidence syntheses, while qualitative data provided in-depth insight into what lessons and support may be useful for other organisations looking to adopt similar types of new technologies.

Improving technologies for evidence synthesis and supporting organisations to make the best use of available technology, can improve public health decision making. The increasing pace of evidence production and the widespread impact of public health decisions during the COVID-19 pandemic points to the importance of making the evidence synthesis process as efficient as possible, while maintaining a high-quality and ensuring reviews are robust.

This study has revealed resources and support that can both help improve new technologies and ensure that these technologies are used in an appropriate way to support the evidence synthesis process. Upskilling, dedicated resources, leadership support and collaborative ways of working were just some of the ways in which public health bodies, particularly smaller ones, could be better supported to take advantage of the benefits automated technology could provide in conducting evidence synthesis.

Introduction

Timely, comprehensive, and reliable evidence synthesis is fundamental to the practice of evidence-based policy, practice and decision making, including in relation to public health [1]. Reviewing published evidence is a key tool for researchers and public health officials to understand the state of knowledge on a topic and inform robust and reliable public health decision making using transparent methods [2, 3].

However, evidence synthesis can be an incredibly time-consuming and resource-intensive undertaking, in part due to the volume of evidence that it is necessary to sort through to conduct a review. Some estimates suggest that a systematic review requires an average of 67 weeks to complete, [4] and by the time they are published, they can be out-of-date. This issue is made worse by the increasing amount of scientific evidence [5, 6]. The COVID-19 pandemic has highlighted and exacerbated these issues in a public health context by presenting researchers and public health decision makers with the challenge of evaluating and synthesising emerging evidence relating to COVID-19 at an unprecedented speed and scale to influence decision-making [7]. These challenges have led to new solutions being applied to synthesise evidence, including living reviews [8] and international networks to help consolidate relevant emerging evidence from across countries [9].

While approaches to evidence synthesis such as rapid reviews and scoping studies have been shown to reduce the time required for evidence synthesis by up to half, these alternatives can be more limited in scope and may not be appropriate for all evidence needs [6]. Automating parts of the evidence synthesis process has the potential to deliver faster reviews with greater efficiency and at a lower cost without compromising the quality of the review [10]. Indeed, without an automated system for evidence synthesis, research organisations and public health decision makers may struggle to keep pace with key developments [5, 11]. Initial consultation from the International Collaboration for the Automation of Systematic Reviews (ICASR) suggests, however, that automating evidence synthesis comes with several difficulties [6, 12] which we have explored in this study. In fact, keeping track of automated evidence synthesis technologies is itself a challenge as these are subject to rapid, and continuous, updates and developments [12].

Study objectives

The study team completed a systematic review and collected qualitative data answering the research questions outlined in Box 1.

Box 1. Research questions on the use and impact of new technologies on evidence synthesis

What types of automated technologies are available and have been used for evidence synthesis, and how have these been used?

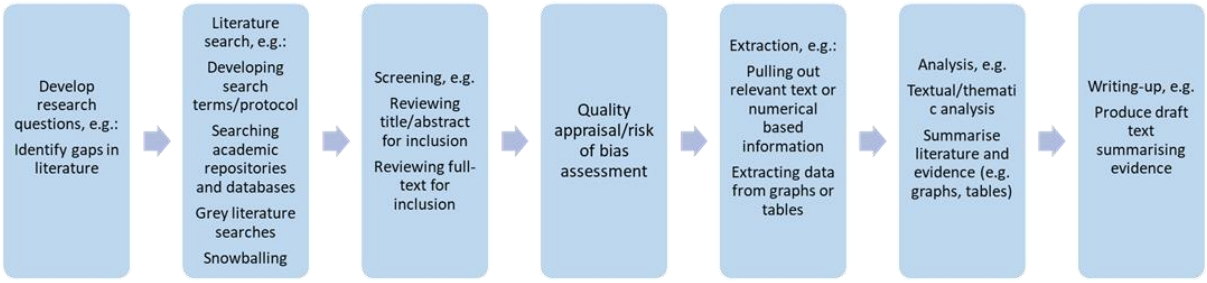
At what stage in the evidence synthesis process have these automated technologies been used (e.g., literature searching, screening, data extraction)?

What are the experiences and lessons learned from using automated technologies in evidence synthesis?

What are the needs of European Union (EU)/European Economic Area (EEA) competent health authorities in conducting technology-assisted evidence syntheses and how can these gaps be filled?

Throughout this report, we refer to the use of automated technologies at different stages of the evidence synthesis pathway. This pathway can be interpreted differently and the figure below outlines the definitions we use for this study.

Figure 1. Evidence synthesis pathway



Research methods

The study consisted of two tasks: A systematic review (Task 1) and qualitative data collection through a mapping exercise (Task 2). These tasks are described in more detail below.

Task 1: Systematic review

A systematic review was conducted to understand technologies that have been used to automate or semi-automate parts of the evidence synthesis process as well as the experience of individuals who have used them. This review included both academic and grey literature, as described in more detail in the following sections.

Task 1.1 Literature search

For the academic literature, we searched Ovid MEDLINE, Embase.com (Elsevier) and Cochrane Library (Wiley) for articles published since 2000. The search protocol (Annex 3 of the supplementary material) was registered in the Open Science Framework (OSF) before screening took place.[1] The grey literature search was done by searching CORDIS[2] and the websites of key public health authorities in the EU/EEA countries, ECDC international partners and major evidence synthesis providers. In addition, two sources were included that were identified by ECDC as being relevant but were not indexed in the literature repositories we searched.

Task 1.2 Screening

The academic literature search generated 8,606 articles, from which we then selected relevant studies based on title and abstract screening. The selection of the literature was based on the inclusion and exclusion criteria (see Annex 4 of the supplementary material). A pilot screening was performed by all screeners of the same 30 articles, and the results discussed. Where there was disagreement, this was discussed between reviewers, and the inclusion/exclusion criteria were refined if necessary.

After the pilot, full screening proceeded in two stages. First, a team of three reviewers (AA, EN, CA) conducted a single screening of titles and abstracts. For any articles that the reviewer was unsure of, this was discussed with the study team until an agreement was reached. For the second round of screening, two reviewers that did not conduct the initial screening (LH, SP) reviewed included articles and removed any articles that did not meet the inclusion criteria. Articles were categorised according to whether they would be suitable for 'light' or 'full'-touch extraction (see next section).

Task 1.3 Data extraction

The data extraction involved either 'full' or 'light' extraction, depending on the characteristics of the article. Any paper which was a review of new technologies for evidence synthesis which could provide more in-depth learning was fully extracted. Articles which were not reviews about the use of technology and thus offered less learning or reflection (e.g. studies using technology to conduct a review but with limited reflection on how well this worked) were included in the light extraction. This strategy allowed us to get in-depth insight into key lessons from using new technologies for evidence synthesis, while also collecting a broad array of relevant examples where this type of technology has been used. For more information, see Annex 4 of the supplementary material.

Articles that were extracted using the full template were appraised for quality using standardised guidelines and checklists. We used AMSTAR2 (a critical appraisal tool for systematic reviews of randomised and non-randomised studies) [13] for systematic reviews, SANRA (scale for the assessment of narrative review articles) for narrative reviews [14], and AACODS (Authority, Accuracy, Coverage, Objectivity, Date, Significance) for grey literature [15]. Quality appraisal was conducted independently by two reviewers for each article, with disagreements resolved by a third reviewer, whose ranking was taken as the final decision. See Annex 4 of the supplementary material for further detail.

1 The protocol was registered on 9 March 2022. Registration https://doi.org/10.17605/OSF.IO/K8P4C

2 CORDIS is the European Commission's primary public repository and portal to disseminate information on all EU-funded research projects and their results.

Task 1.4 Analysis and synthesis

Once data extraction and quality appraisal were complete, the study team met to discuss the evidence for the research question in order to identify key themes. The extracted information was then analysed, taking a narrative synthesis approach. An account of key messages and findings from this synthesis is presented in this report, backed by insights from the data extraction.

Task 2: Qualitative data collection (mapping exercise)

To get further insight into the experiences of those that have used new technologies to support evidence synthesis, and the needs of public health authorities in terms of evidence syntheses, we conducted a qualitative data collection exercise in the form of a mapping exercise. This included several different methods for data collection, described in more detail below. Prior to each data collection exercise, participants were provided a participant information sheet with information about the study. Invitations to participate in the qualitative data collection were sent to researchers identified through the literature review as well as the ECDC Advisory Forum members and National Focal Points for Scientific Advice Coordination. While Centres for Disease Control with whom ECDC has a Memorandum of Understanding were also invited to participate, none ultimately took part.

Task 2.1 Mapping relevant stakeholders

We reviewed initial results from the systematic review and mapped which individuals and organisations had experience with automated technologies for evidence synthesis. with experience of using automation in fields related to public health. A selection of these relevant individuals were invited to participate in data collection exercises.

Along with these stakeholders, ECDC also identified relevant stakeholders, which included competent public health authorities and National Focal Points for Scientific Advice Coordination in EU/EEA Member States, as well as organisations responsible for disease prevention and control in other countries, ECDC staff and members of the ECDC Advisory Forum. At least one representative from each EU/EEA Member State was invited to participate.

Task 2.2 Focus groups and interviews

Focus groups were the primary mechanism for data collection for the mapping exercise. The list of individuals identified in Task 2.1 were invited to participate in a focus group. All individuals invited were sent a Doodlepoll with dates for the focus groups and asked to select all dates they were available. Individuals were then sent an invitation for one of the focus groups. This supported attendance as participants were offered a selection of different dates and also allowed a more even spread of participants across the sessions.

A total of five focus groups were held between mid-May and mid-June 2022, each lasting approximately 90 minutes. The focus groups were guided by a semi-structured protocol covering the following topics: current approaches/tools for conducting evidence syntheses, uses and experiences of automation, needs and gaps in using automation, and interest in participating in future discussions/collaborations on this topic. Each focus group was co-facilitated by a member of the research team (LH or SP) and a representative from ECDC (HdCG), who helped ensure participants felt comfortable speaking and sharing their views.. The facilitator ensured that all topics were covered, circling back to relevant topics and progressing the conversation as necessary. The focus groups were audio- and video-recorded, with detailed notes also taken by a member of the study team.

If individuals declined or were not available to participate in the scheduled focus groups, one-to-one interviews were also offered to support the participation of as many individuals as possible. Three interviews were conducted overall, lasting up to one hour each. Interviews followed a semi-structured protocol, based on what was used for the focus groups, with detailed notes taken during the interview, along with audio- and video-recording.

Invited participants were also given the opportunity to provide written responses if they declined an interview. However, no one chose to participate in this way.

In total, 46 individuals participated in focus groups, and three individuals were interviewed. Both the focus group and interviews were conducted virtually via Microsoft Teams. For more information on the protocols used in the focus group and interviews, see Annex 5 and 6 of the supplementary material, respectively. For a list of participants in each data collection method, please see Table 16, Annex 1 within this document.

Task 2.3 Survey

A survey was designed based on the topics covered in the focus group, with a mixture of closed and open questions. The survey was sent to those who did not attend a focus group or an interview but was also distributed to others who did attend but wanted to provide additional information. The survey was sent out by ECDC and conducted via SmartSurvey [16]. It was kept open between 27th June 2022 and 11 July 2022. See Annex 7 of the supplementary material for the survey tool. Question routing was applied to ensure respondents only saw questions that were relevant to their experience.

In total, nine people participated in the survey, three of whom had previously attended a focus group. Given the limited sample size for this survey, results should be treated with caution.

Task 2.4 Analysis and reporting

The focus group and interview data were analysed using thematic analysis in Nvivo (see Annex 4 of the supplementary material for coding frame). The survey data were analysed using Excel and triangulated with the data from other data collection exercises. The mapping exercise originally intended to present mapping tables, to map which EU/EEA public health authorities are using automation. However, given the limited use of automated technologies for evidence synthesis across EU/EEA public health authorities, it was determined that these tables would provide limited insights and knowledge. Instead, we use a narrative format to describe a few examples of how automation has been used by public health authorities (and other organisations) to share learning of previous experiences.

Strengths and limitations

This study reviewed more than 150 articles, which provides an overview of the breadth of technologies that have been used for evidence synthesis. Furthermore, by bringing together reviewers with experience using new technologies and public health authorities that may benefit from this type of technology, this study sheds light on support that may help new technologies be used for evidence syntheses in the area of public health and infectious diseases. By providing multiple ways to participate in this study, we achieved good engagement from a wide range of stakeholders, across many different EU/EEA Member States and other countries.

There are several limitations to note regarding this study. Firstly, the systematic review was organised such that some articles (e.g. those *about* the use of new technologies in evidence synthesis) were subject to a more detailed extraction process than others (e.g. review articles that used a new technology as a method). While this allowed us to get in-depth insight while also capturing the breadth of where new technologies have been used, there may be some relevant insights that were missed. There are also limitations around the qualitative data collection – while every effort was made to accommodate participants and provide them with opportunities to engage, there may be relevant stakeholders that were not included in the study that have different views. For example, while we invited a representative from every EU/EEA Member State, there are some countries missing from the study. The survey results, as noted throughout this report, should be treated with particular caution given the small sample size.

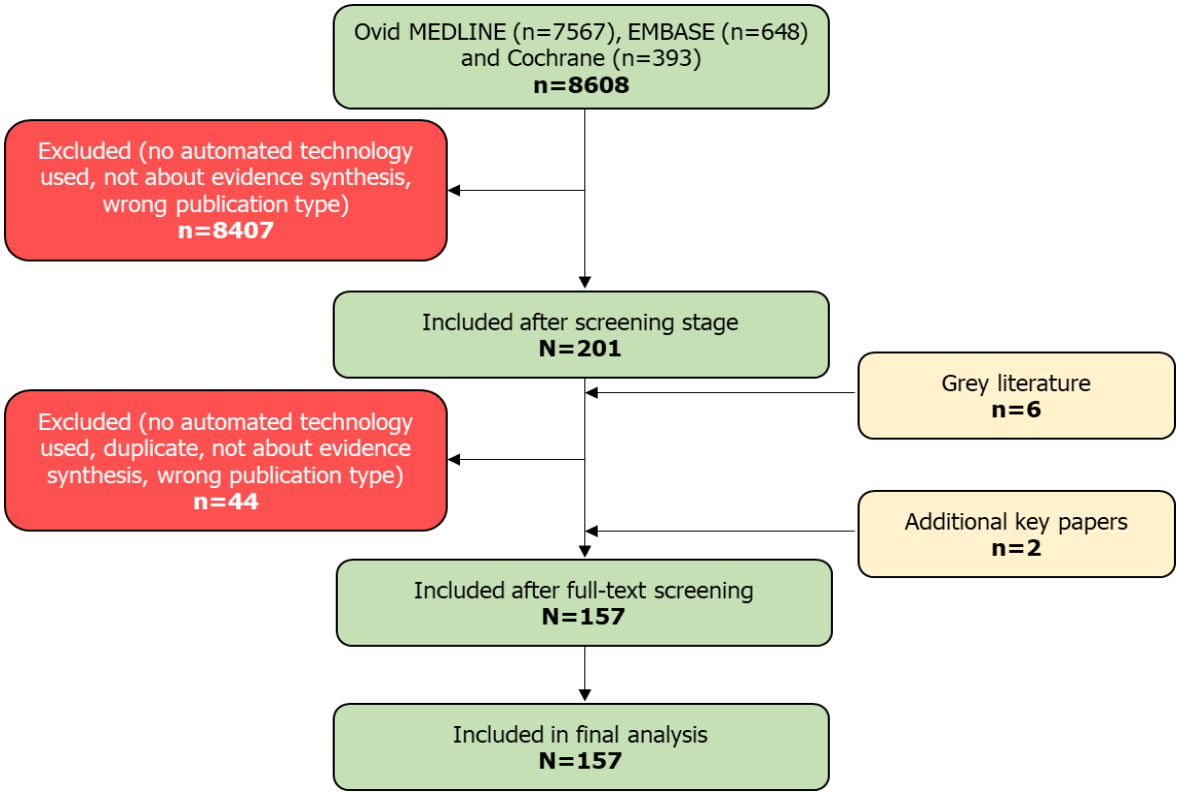
Results from the literature review

Here, we provide an overview of the reviewed literature and discuss the reasons for using automation, the types of automated technologies available for evidence syntheses, the impact of using these technologies, what has worked well (and what has not) and the gaps/needs in using automated technologies. Annex 1 includes a table with each type of technology discussed in the reviewed literature, a description, what did and did not work well, and any gaps/needs in using it.

Overview of the included articles

In total, 157 articles were included in the final analysis. Of these, 28 were analysed using the full extraction template and the other 129 using the lighter touch template.

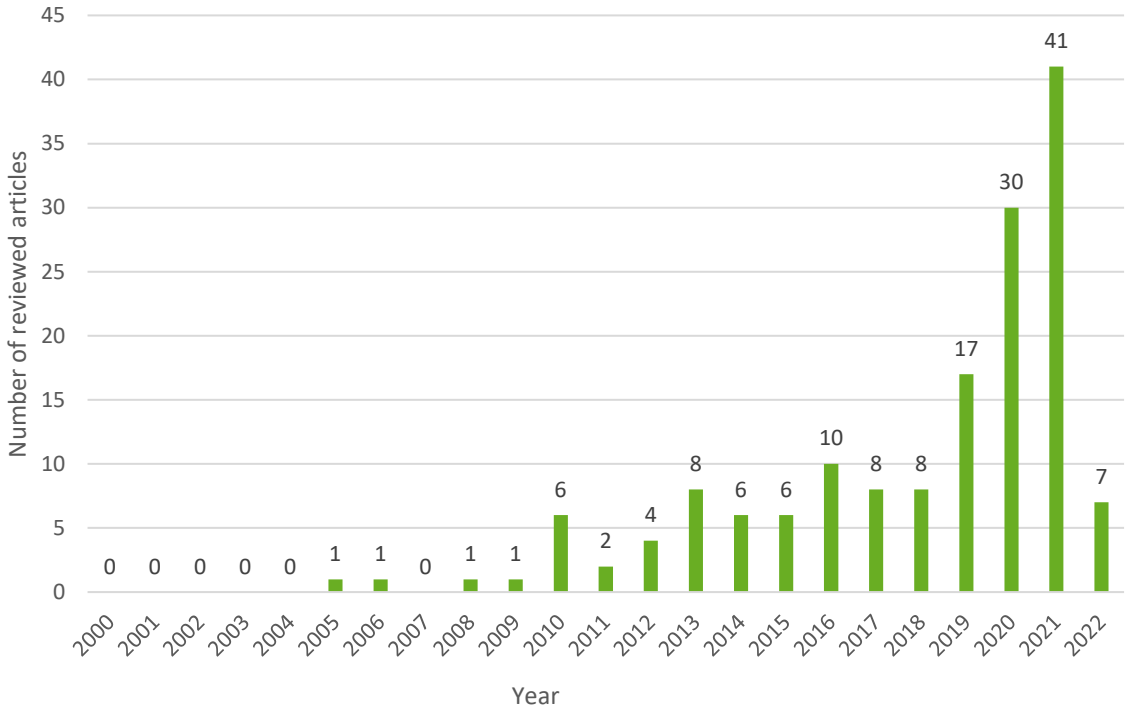
Figure 2. PRISMA diagram



The articles analysed using the full extraction template were also subject to a quality appraisal. Different appraisal tools were used depending on the type of article (see Chapter 2) which makes it difficult to compare across appraisals and two of the three appraisal criteria used do not encourage an overall quality rating to be applied to articles. However, to enable a high-level reflection of the quality of the articles that were extracted using the full template, each article was ranked as being low, medium or high quality (see the supplementary material for more information). Overall, 28 articles were of relatively good quality, with six ranked as high quality, 19 as medium and only three as low. While care needs to be taken in interpreting these rankings as some of the appraisal criteria did not recommend applying overall ranks to articles, it does suggest the evidence we have primarily drawn on for this review and focus on in greatest detail is of good quality.

While automated technologies for evidence synthesis have been around for decades, publications on this topic only started to increase since 2010 (Figure 3) and there has been a particularly rapid increase in publications since 2020. This coincides with the start of the pandemic, although it is not clear if these are linked. While our literature search included articles published since 2000, no studies published before 2005 were included in our final set.

Figure 3. Publication year of the included articles



Note: 2022 publications are low as the literature search was conducted in March 2022

Given the types of databases and websites searched to identify literature, it is unsurprising that most of the reviewed literature focused on topics relating to health, medicine and biomedicine (110 articles). However, there were very few articles (nine) focusing specifically on the use of automation to review evidence related to infectious diseases (including COVID-19).

Studies from the literature review used or reviewed the use of automated technology across the evidence synthesis pathway, from developing research questions through to reporting. It is challenging to produce exact numbers for how many technologies were cited as being used at specific stages of the review process, for example, authors often switched between discussing different technologies without making clear at what stage they are being used. However, automation seems to be most often used during the screening stage, with multiple technologies also used for literature searching, extraction and analysis. While examples of automation at the stages of developing research questions, quality assessment and writing-up were identified, these were the least common uses. Many technologies can be used at multiple stages of the evidence synthesis pathway (e.g. analysis and reporting). In addition, some technologies can be used in combination with other technologies to perform multiple tasks. Common combinations include 1) literature search and screening, 2) screening and quality assessment, 3) screening and analysis, and 4) extraction and analysis.

A large range of organisations were involved in producing the reviewed articles. A full list of organisations is provided in Annex 2 of the supplementary material and the organisations which contributed to five or more articles are outlined in the table below. The majority of articles were produced in collaboration with multiple organisations.

Table 1. Organisations contributing to five or more of the reviewed articles

Organisation	Number of articles
University of Manchester (UK)	10
Bond University (Australia)	9
University College London (UK)	9
Northeastern University (USA)	7
Oregon Health & Science University (USA)	6
University of Alberta (Canada)	6
University of Oxford (UK)	6
US National Library of Medicine (USA)	6
King's College London (UK)	5
University of London (UK)	5
University of Edinburgh (UK)	5
University of Ottawa (Canada)	5

Unsurprisingly given the varied organisations involved in producing the reviewed articles, researchers from a range of countries are exploring the area of evidence synthesis automation. The table below outlines the countries contributing to 10 or more of the reviewed publications, and a full list of countries is in Annex 2. Countries who most frequently collaborate are the UK and USA (11 articles), the USA and Australia (six articles) and the UK and Japan (five articles).

Table 2. Countries contributing to 10 or more of the reviewed articles

Country	Number of articles
USA	54
UK	46
Australia	25
Canada	23
Netherlands	12

Why are researchers using automated technologies for evidence synthesis?

There are two key reasons that automation is used for evidence synthesis: 1) to deal (quickly) with large volumes of published evidence and 2) to reduce the resources needed to conduct evidence reviews.

Large volumes of literature and evidence are published each year, and this is increasing over time [17-24]. Evidence synthesis can quickly become out of date (sometimes before they are even published, thus ending up publishing results that are no longer relevant) and it can be difficult to make sense of large bodies of evidence using manual techniques [17, 18, 25-29]. Therefore, automating aspects of evidence synthesis has the potential to enable a more robust and quicker exploration of large quantities of existing evidence [17-19, 30-33].

Reviews, especially systematic reviews, are very time, cost and labour intensive, and there are risks of human error [21, 22, 26-29, 34-37]. It is estimated that a systematic review takes one year between the literature search and publication, and a primary study takes 2.5 –6.5 years to be incorporated into a systematic review [29]. Systematic reviews can also cost up to $250 000 [29]. Automating parts of the evidence review process can help reduce this burden while maintaining (or improving) robustness, sustainability and efficiency [17, 18, 22, 23, 27, 29-33, 36, 38-46]. For example, automation can be used to organise articles by relevancy for the screening stage, to ensure those that are likely to be more relevant are prioritised [20], or used to distil and summarise information in a fast and accurate way [40].

Types of automated technologies used in evidence synthesis

While a large range of specific automated technologies (e.g. software, algorithms) were discussed in the literature, most of these took the form of machine learning. We briefly outline the overarching methodologies used in automation here, including examples of where these have been used to conduct evidence synthesis. A more in-depth overview of the impact, strengths and limitations of each is provided in Annex 1, along with a full list of the specific technologies (e.g. types of algorithms and software) discussed in the literature

Table 3. Description and examples of overarching automation methods

Technology	Description	Example of use
Machine learning	A type of artificial intelligence (AI) that uses computational methods [18, 47] to learn over time, improving the accuracy of a task without the need for explicit programming. There are two ways in which machine learning can operate, by using supervised or unsupervised learning. Supervised machine learning requires human input to initially train an algorithm on how to perform a task [48, 49]. Unsupervised learning does not need to be trained by a human. Rather, it learns using patterns in text and data and mimics these patterns [26]. For both types of learning, the algorithm continues to learn and improve over time. Machine learning can be used throughout the review process, at the literature search, screening, quality assessment, extraction and analysis stage [18, 29, 37, 39, 44, 47-56].	Bekhuis et al. tested a prototype of a machine learning tool on a systematic review on organ transplantation. The tool reduced the screening burden by automating the second screening phase (rather than having to manually re-screening articles). A set of articles were collated on organ transplantation to use as a training set. The researchers labelled each article as include or exclude, and half of this set was used to train the tool and the other half to test performance [57].
Natural language processing	A computational approach using AI (combined with linguistics and computer science) to analyse and interpret text. It can perform basic tasks, such as counting word frequencies, up to more complex activities, such as classifying and understanding text [21, 29, 58, 59]. Natural Language Processing can be used at the literature search, screening, quality assessment, extraction and analysis stages [21, 29, 44, 51, 55, 58, 59].	Zulkarnain et al. performed a systematic review on intelligent transportation systems using natural language processing. The tools were used to exclude irrelevant articles, explore research trends and identify gaps in evidence [60].
Text mining	A form of AI that uses linguistics, statistics, computer science, natural language processing, machine learning and data mining to convert unstructured text to a structured form that can be extracted and analysed [17, 18, 23, 61-63]. It can be used at the literature search, screening, quality assessment, extraction, analysis and reporting stage of a review [17, 21, 23, 38, 44, 61-68].	Lin et al. performed a literature review of urothelial cancer using text mining approaches to deal with large volumes of literature on this topic. Text mining was used to analyse the topics in title and abstracts of nearly 30 000 articles from three databases. Fifteen research topics relating to urothelial cancer were identified during the analysis. The authors also used text mining to analyse trends in research over time [69].
Neural networks	A group of computing algorithms that can identify similarities and relationships between words across articles [70] by mimicking the way neural networks within the human brain work. They can take the form of, for example, deep-learning (having more than three networks) or recurrent neural (enable temporal analysis) networks [22, 71]. Can be used at the screening stage [22, 70, 72].	Yamada et al. tested the performance of a deep neural network in a systematic review of clinical guidelines (diabetes, cardiology and stroke). Here, a set of relevant and irrelevant articles were used to train the algorithms. The tool learned from this set and was able to screen articles independently [73].

Impact of using the automated technology

There are three key impacts of using automated technology for evidence synthesis:

- 1) reducing the time and burden to complete a review
- 2) identifying (more) relevant articles
- 3) improved interpretation of results and precision.

Each of these describes the positive benefits of using automation. While we did not identify any negative impacts from the literature, there are many challenges of using automation and limitations of certain technologies which are described later in this chapter.

Reduce the burden of conducting reviews

As outlined earlier, a main reason for using automated technologies for evidence synthesis is to reduce the burden in terms of time, workload and cost of conducting a review [18, 20, 22-24, 27-33, 37-44, 74-77]. Examples of tasks that can be done quicker and more efficiently at different stages of the evidence synthesis pathway, and real-life examples, are outlined in the table below.

Overall estimates suggest that workload can be reduced by 40-50% across the life of a review [23]. This can make the use of automated technologies more cost-effective than traditional, manual methods (e.g. less expensive, as less senior researcher time is needed) [23, 34]. For particularly tedious tasks, such as backward citation searches and snowballing, automated and AI-based techniques can be especially helpful and cost-saving [27, 74]. Automation also allows greater energy to be put into the intellectual aspects of reviews and interpreting the findings [27]. However, automation does not remove human input altogether, time (and expertise) is often required to ensure the technology is able to accurately perform the required task [38].

Table 4. Descriptions with examples of how automation can be used to reduce the burden of evidence synthesis

Stage	Description of use	Example
Question development and literature search	Setting up literature searches to automatically search for new articles to add to a review. This can allow researchers to stay up-to-date with current evidence more easily, focus on the 'right' questions and prevent duplication of other reviews [27, 30]. This makes it easier and quicker to undertake living reviews[3] in particular [28] and can help inform funding decisions to focus effort on areas of need [24, 27, 78-81]. Similarly, automated technologies can help gather evidence around whether there is a need to update a systematic review by indicating the likelihood that new evidence will change conclusions, [82] which also improves the process of evidence synthesis and decisions to fund reviews [81].	John et al. have conducted a living review into the impact of COVID-19 on self-harm and suicide. To regularly update the review, automated daily searches are conducted via two databases as well as other public health websites (the code for which is freely available). [83, 84]. This enables new evidence to be identified regularly, without requiring significant manual input.
Screening	Automated technologies can screen articles automatically, or can prioritise articles by relevancy which reduces the burden of manual screening [22, 23, 44]. One study exploring the use of text mining reported findings that screening time for two reviewers was reduced by approximately 90% [20]. Others indicate a time saving of 50% [44] and a reduction of 30-50% of articles that need to be manually screened [27]. Allowing for faster screening can mean highly relevant papers are identified early, and extraction and analysis can begin in parallel to finishing screening, allowing the review to be completed faster [23, 44, 85].	Thiabaud et al. conducted a systematic review into the social, behavioural and cultural factors of HIV. The authors used a custom-built Python tool to automate article screening. The tool reduced the number of potentially relevant articles from nearly 17 000 to just over 500 [75].
Quality assessment	Automated approaches can be used to conduct quality assessments of reviewed articles. One study suggests that using machine learning can reduce the time it takes to perform quality assessments by around 25% [22].	Clark et al. conducted a systematic review on urinary tract infections. A variety of automated technologies were used, including the use of RobotReviewer to assess for any potential biases in the reviewed literature. It does this by highlighting phrases in the full text that could indicate bias. The full review was conducted in two weeks as a result of using automation [86].
Extraction and analysis	Automation can be used to identify and extract relevant information and data from	Guzik et al. used text mining approaches analyse the text in articles relating to polyhydroxyalkanoates (a natural polyester).

[3] A living review is a literature review that is updated at regular intervals to incorporate new evidence.

Stage	Description of use	Example
	full text articles. Automation can also support rapid analysis, for example, automatically producing summary graphs/tables, analysing and categorising research topics and creating PRISMA diagrams, without needing to do a full (manual) review,[19, 22, 23, 27, 38, 40, 63, 64, 74, 78, 80, 87-93] leading to efficiencies and improved flow of information. Using automation allows these tasks to be completed much faster than if done manually; one article suggests this could save years in the production of systematic reviews [27].	Automation was used to extract text to identify categories of research topics and trends in research over time [79]. The authors note that understanding current research areas can prevent duplication of reviews and focus research on outstanding gaps in knowledge.
Writing-up	Some authors report that automation can even be used to start drafting text, including abstracts and discussions (with some human input) or in different languages [38, 44]. This can save time in drafting more factual sections, such as an overview of study interventions or population sizes.	While no examples of where automation has been used to generate text for reporting were identified in the reviewed literature, one article provides a description of the software RevManHAL. This tool is able to automatically generate abstracts, results, discussion and acknowledgement sections in multiple languages. This can save time in conducting the review, allowing researchers to focus on interpreting the data [94].

Identify more relevant literature without needing additional time or cost

Using manual approaches to evidence synthesis creates challenges in adequately identifying all the relevant literature given the high volume of publications. As automated technologies can be used to develop more precise search strategies and automate parts of the screening and extraction stage (often the most resource-intensive points), they allow for larger and more ambitious reviews with a higher number of articles reviewed. This is possible without leading to further burden in terms of time or cost, or impacting on quality, which would not be feasible using traditional methods [18, 20-24, 27, 35, 38, 74-77]. For example, automation can overcome the limitations of using Boolean search[4] approaches [18] and can automatically translate search protocols to use in a wide range of literature databases [21]. Automation can also be used to support the identification of search terms that may not have been included using traditional methods [35].

Improve the interpretation of evidence with greater precision and transparency

Automated technologies allow for improved interpretation of results and greater precision [17, 21, 23, 25, 27, 28, 36, 38, 39, 43, 44, 49, 94]. For example through reduced human error; improved screening accuracy; no double counting; unstructured data can be analysed systematically; less subjective interpretation of evidence; help resolving disagreements among researchers; and improving study design, search strategies and question formulation [17, 21, 27, 34, 38, 74, 75, 89, 95].

While there are some concerns over transparency of algorithms (see also discussed later), some articles highlighted that automated approaches can improve reproducibility and transparency [27, 42, 47, 74, 78, 95, 96]. Technologies with AI/machine learning aspects are also able to learn over time, improving their precision and accuracy [39, 41]. Reproducibility and transparency can also be increased by using automated approaches as, for example, the same logic can be applied to other studies [21, 42].

What worked well when using automated technology?

As was expected, authors of the reviewed literature did not provide much reflection on what worked well and supports the use of automated technology for evidence synthesis. Most of the focus in this regard was on the positive impacts of using the technology. While some supporting factors were noted in the systematic review (good performance of the technology and user-friendliness), this area is more fully discussed in Chapter 4 where the results of the qualitative data collection exercises are discussed.

[4] A Boolean search strategy is that which uses key words separated by, for example, AND, OR, NOT to limit the search to more relevant articles.

Effectiveness of technologies to support automation

Many studies of machine learning, text mining and other approaches to support automation in evidence syntheses have found that these methods are effective compared to traditional literature review techniques [18, 20, 31, 43, 44, 50, 52, 53, 64, 66, 74, 82, 88, 96-119]. By effective, we mean that some automated methods have been shown to have high validity, accuracy and reliability. For example, some technologies were shown to have similar accuracy to humans in conducting risk of bias assessments and data extraction[41, 44]. Some studies found that these methods can be even more effective than traditional methods [27, 42, 47, 74, 75, 78, 89, 95, 96, 101]. For example, technologies have been shown to identify more relevant studies than traditional methods, to reduce human error in extraction and increase reliability of screening (meaning that the same text will be included or excluded regardless of who is conducting the review or when) [18, 21, 34]. Automated technologies can also improve search strings through repeated iteration of the search process, leading to search strategies that increase recall (the accurate identification of relevant articles) and coverage (the identification of all relevant articles for a review] [21, 23, 27, 35, 43]. Methods that combine human and machine efforts, such as where algorithms help humans to screen articles or assess quality can be particularly effective and help balance the time and cost savings with reliability concerns [65, 75, 88, 99, 109, 110, 120-124]. The box below outlines an example of the performance of one technology, RobotReviewer.

Box 2. Performance and efficacy overview of RobotReviewer

RobotReviewer is a machine learning system used to support evidence synthesis from RCTs. It works by identifying information about how studies have been conducted, and can identify information about the population, intervention, control and outcome of a study ('PICO'), along with determining study design and informing risk of bias assessments. It is open source and based on an natural language processing model [107].

Although there are some limitations to the performance of RobotReviewer (see Annex 1 of the supplementary material), one article reported that RobotReviewer has a level of accuracy similar to humans [41], and a cross-sectional evaluation of technologies for risk of bias assessment reported that its reliability between author groups is comparable to humans or better in some domains [101]. Similarly, other studies have reported that its bias classifications are 'reasonable', that it has moderate reliability[32] and that it achieves high accuracy compared to previous machine learning based approaches for quality assessment and extraction [107].

Beneficial features and user-friendliness of automated tools

Researchers who have used automated methods have highlighted that they are often easy to use and user friendly [38, 44, 74, 78, 96, 97, 125-128]. However, this was not consistent with the data collected from the qualitative work where those working in public health described the need for training and skill development to be able to use automated approaches effectively (see the following chapter for further information).

Specific technologies often have features noted by authors as being useful when undertaking a review. An overview of these is provided in Annex 1 and includes features such as use on multiple languages, [34, 88, 97] being free to use/open source, [101, 114, 120, 125, 126, 129] able to manually edit outputs (e.g. screening decisions), [89] ability to use offline and with multiple users at once, [45, 126] and ability to integrate with other tools [30, 80, 108, 125]. An example of these types of functions is provided in the box below.

Box 3. Examples of useful functions integrated into automation tools

Cohen et al. used a machine learning based document classifier which alerts researchers to articles that can be included in systematic review updates and identifies those that are likely to be most relevant using annotations. This was seen as a useful tool for planning, scheduling and allocating resources for updating systematic reviews as the recall rate[1] can be set depending on the resources available to conduct the review [81]. This enables more articles to be reviewed for systematic reviews with greater resource availability and vice versa for reviews with fewer resources.

What did not work well when using automated technology?

While information on what supports the use of automated technology was limited within the literature, there was greater reflection on aspects that made using automation for evidence synthesis more challenging. These factors are: 1) questions around bias, transparency and performance; 2) technical limitations; 3) availability of reference texts; and 4) continued need for human input. These often focus on the limitations of the technology itself, rather than the wider challenges of technology use, e.g. resource needs, staff training/expertise.

Here, we focus on high-level challenges identified from across the literature. A full list of the technology discussed in the literature, along with the specific challenges of using each one, can be found in Annex 1.

Questions around bias, transparency and performance

Text mining and similar methods to automate reviews are robust in many ways, but the choice of algorithm, the training set used and the interpretation of the results are not neutral and can be a source of bias [17, 30, 33, 40, 47, 63, 78, 96]. If initial training datasets, for example, are biased towards certain topics, countries or authors, algorithms may perform less well in identifying relevant articles outside of these areas. Alternatively, algorithms may be biased in that they reproduce the same errors of the researchers that trained them. Methods that can be considered a 'black box,' in that the exact algorithms are not designed by researchers, (such as those based on AI and machine learning rather than rule-based automation), have also been criticised due to a lack of transparency [17, 46, 47, 61, 74, 89, 127].

Although effectiveness of machine learning approaches is generally high, some machine learning approaches perform less well than other techniques compared to traditional methods [22, 23, 30-33, 40, 43, 49, 51, 52, 54, 63, 72, 74, 76, 87, 101, 113, 121, 130-132]. Many techniques that support automation struggle particularly in terms of recall, or correctly identifying relevant or high-quality articles, [18, 23, 27, 31, 43, 46, 48, 50, 90, 91, 110, 133, 134] assessing quality of articles, [107] and where the screening process is attempted entirely without human input [120, 121, 135]. Additionally, algorithms for screening may perform less well on ambiguous texts, [30, 89] as well as more complex subject matter which need more interpretation from experts and may have multiple interpretations, [17, 27, 40, 46, 47, 63, 69, 74, 76, 109, 119, 136], although some studies report good results even with complex subject matter [109]. The box below outlines an example of where there can be concerns around performance, using the tool Abstrackr as an example.

Box 4. Abstrackr performance concerns

Abstrackr is a free online tool to help facilitate reviews by prioritising abstracts for human researchers to review. It is a machine learning-based tool that learns from human decisions by observing inclusion and exclusion decision, and so its performance improves over time [22].

The precision of Abstrackr appears to vary [124], with studies identifying different performance results. Some studies report that Abstrackr can significantly reduce the workload of screening abstracts [98, 100, 102]. For example, one review that used the technology in the area of nutrition found that there was low risk of it excluding relevant studies [137] and another found it has a low false negative rates, at about 0.3% [102]. However, other studies note limitations in its performance. For example, one study found that using Abstrackr for automatic screening can reduce recall rate by approximately 5% [27] and may exclude 6-14% [120] of relevant records [98, 102]. Another study identified a median false positive rate of 12.6%, ([100)] indicating a potential need for human review of results to avoid overinclusion. This means that its use may be most appropriate where there is low risk associated with incorrectly excluding relevant articles [98].

One study found that Abstrackr performs better with mixed-methods or qualitative studies, as compared to observational studies and reviews, and that it also makes better choices with articles that were published more recently [98]. Similarly, it has been reported to miss fewer relevant studies when used in systematic reviews, rather than descriptive reviews, (100] perhaps reflecting variability in human decision making for descriptive reviews.

Piloting technologies that are meant to automate screening can help understand limitations of the technology and ensure it is suitable for specific evidence syntheses, [122, 125, 131] as can adjusting sensitivity and specificity levels on algorithms where relevant depending on the risks associated with under or over inclusion [119, 138]. In some cases, it may be appropriate to use automated techniques as a supplementary tool to reduce the time to produce an r evidence synthesis (e.g. in scoping or rapid reviews), being aware that these techniques may not always be as comprehensive as traditional techniques [63].

The perception that reviews conducted by people are more reliable than ones conducted by machines can also cause issues for studies that use this type of technology [22, 37] and can make researchers uncomfortable with handing over too much control to machines [23]. These concerns can cause extra tasks for researchers that use tools meant to support automation (e.g. in reviewing or confirming the work that machines have done), leading to additional workload and time, an issue which would need to be overcome if these methods are to produce time and cost savings [37].

Technical limitations in the type and format of information that can be processed

Each tool and method that supports automation will have its own limitations in terms of what type of information it is able to process. Many text mining techniques, for example, can only process certain formats of information and do not work well with certain document or citation formats (e.g. PDFs, RIS files), images, data in numerical formats, and tables [17, 27, 75, 102, 126, 133, 139]. Some tools and algorithms only work for texts in English, [17, 19, 27, 78, 82, 101] although some algorithms may be better able to handle non-English text [88]. Some algorithms may also struggle with short texts, as they do not provide enough information for the algorithm to use [72, 97, 140] or have missing abstracts where algorithms need to work with the title or key words only [102, 141]. Some databases may not be compatible with particular methods of automation, [27, 39, 64, 112, 129, 139], limiting the technical ability to automate and limiting the number of databases that can be searched. The box below provides two examples of where these limitations occurred when using automation.

Box 5. Examples of limitations in the information processed by automated tools

Yu et al. developed a method to automatically extract information from medical articles in Wikipedia. It was found that the tool was less accurate at identifying and extracting relevant information when the article was short. The accuracy of the tool was also influenced by the clarity of information within the article, for example, it confused human and animal information in some instances [140].

Millard et al. used text mining to automate risk of bias assessments. The authors found that using the tool on title/abstracts only was not as accurate as using full texts. However, accessing full texts requires greater effort and using full text to assess bias generates noise as the tool is reviewing lots of additional text that is not necessarily related to the quality of the article [65].

Availability of reference texts for training sets

For some automated technologies, there is a need to train the algorithm to make decisions using a set of articles (often called a training set). Training sets require a large volume of reliable literature, [17, 30] and the investment of time and effort from highly skilled researchers to identify and annotate relevant articles [18, 19, 22, 25, 40, 46, 49, 97, 104, 122]. The ability of text mining and machine learning methods to work properly depends on, and is limited by, the quality of training datasets [30, 32, 33, 40, 41, 46, 82, 93, 103, 117, 122, 124, 131, 142, 143]. This was highlighted in a systematic review of machine learning in biomedical literature reviews as the most commonly reported limitation of studies using these techniques,[18] and the box below highlights an example of how this created challenges in one review. If the initial sample is biased, algorithms (e.g. for automatic classification of documents to aid in screening) will only reproduce those biases [40]. Some subject areas will be more difficult to automate, given the paucity of reference texts in them, which has been shown in some studies to decrease performance of machine learning algorithms and automated techniques [30, 74, 81, 89, 115, 127, 143, 144].

Box 6. Example of how training sets can influence automation performance

van den Bulk et al. reported on the use of various automated approaches for the screening stage of evidence synthesis of food safety topics. The authors noted that the way in which training sets were set up influenced the performance of the tools they were using. The authors excluded articles classified as 'maybe' from the training set which could have reduced the performance of the tool, as not all relevant articles were used to train the algorithm. The size of the training set was also reported as being important, with algorithms trained on a smaller set by the authors demonstrating lower performance than those using larger sets. Finally, the authors also reflect that the training set needs to be developed for each individual review given the inclusion criteria will be different each time, requiring manual input and time [30].

The continued need for human input and skills

The need for human input, including from senior and highly skilled researchers and reviewers, is a key limitation of automated (or semi-automated) techniques [20, 22, 24, 35, 37-41, 43, 49, 63, 66, 88, 97, 103, 106, 121, 125, 133, 134, 145, 146]. Human input may be needed for more complex tasks that require human understanding, [12, 20, 35, 62, 76, 88, 97, 116, 145, 147] as well as more mundane tasks such as tracking down full text articles that are not available through automatic retrieval processes [41], or checking the accuracy of technology outputs.

There are some techniques that rely on machines and algorithms to rank or prioritise citations for screening, with a human then performing the screening in the order specified by the machine. In theory, the manual screening process could then be stopped once relevant citations are no longer being found. However, there are challenges with this technique in that it is not always clear where it is 'safe' for the process to be completely automated, as even those at the bottom of the list have a non-zero probability of being relevant [23, 33, 37, 129, 148]. Similarly, for techniques that use 'active learning' (humans teaching machines how to conduct reviews and training them to make correct decisions), it is not clear where it is 'safe' for machines to take over the process [73]. This uncertainty limits the ability to automate the screening process, meaning that human effort could still be needed to exclude studies with certainty.

In particular, using automated techniques for the extraction phase are difficult, as often, articles still need to be read in detail and because technology does not yet allow for reliable extraction of many types of information, [33, 38, 43, 130, 133] compared to earlier stages of evidence syntheses such as screening. Because of limitations in the current technology, full automation for extraction, synthesis, interpretation and reporting is not expected in the near future, according to reviews of available technology, [12, 74, 133, 145] although there are some prototypes attempting to automate text generation from evidence syntheses [94]. Some studies have also argued that other parts of the evidence synthesis process would be difficult to automate due to the need for human interpretation, namely the second stage of screening [74] and quality assessments [74, 107, 130].

Researchers that use technology to support automation also require time and effort to learn how to use these tools and become familiar with them, [28, 32, 35, 43, 63, 66, 86, 122, 142], with some technologies suffering from unclear instructions [102]. Some technologies can be difficult to use without specialist knowledge of machine learning, computer science or statistics, [23, 122, 133, 149-151] which can also be a barrier to their use.

Box 7. Example of manual input requirements when using automated tools

The Public Health Agency of Canada, in collaboration with Xtract AI, developed a natural language processing software used for data extraction which was tested in a review of immunisation-related articles. The authors noted that some involvement from experts was needed to check accuracy, including manual work, to review the predictions generated by the software [106].

Gaps and needs in using automated technology

Four key gaps to support the effective use of automated technology for evidence synthesis were noted from the literature.

- a need for additional evidence and standards around effectiveness of technology (e.g. its validity, accuracy, reliability and reproducibility)
- improvements in technology and the performance of current methods for automation
- how evidence is reported, stored and indexed
- support for the use of automated approaches.

While these are important areas that need further work to support the use of automation, these are largely outside the control of any one individual research team or organisation. Therefore, one of the main focuses of the qualitative data collection and mapping exercise (described in following chapter) was to explore what individual research teams need to support the use of existing automation tools (e.g. staff expertise/training, access to new equipment/facilities, collaborations etc.).

Need for additional evidence and standards around effectiveness

Text mining and other machine learning techniques are still fairly new in terms of evidence synthesis applications, and need to be tested to further understand their reproducibility, reliability, accuracy and validity [17, 19, 22-24, 37, 46, 61, 74, 107, 125, 133]. There is also a need for additional evidence to understand the human factors affecting their effectiveness (e.g. how researchers cognitively process text that has been automatically summarised and the experience of users of automated technology) [19, 107] and the topics that automated techniques can be applied to [127]. Additionally, the effectiveness achieved from combining different methods and tools to conduct automated reviews needs to be investigated, [25] although some studies have found that combining techniques can lead to better outcomes [25, 35]. Another study that attempted to reproduce the results of citation screening using text mining found that it was difficult if not impossible to assess the reproducibility of these methods based on current reporting practices in the published literature, [61] indicating a clear gap in the evidence.

There is also a need to clarify metrics and standards that can be applied to technologies to support automation in evidence syntheses. Studies point to the need for reference standards to ensure tools to support automation meet certain quality requirements that allow them to be used by other researchers, [19, 37], and the need for comparable metrics and methodological guidance to measure performance of semi-automated techniques [20, 23, 37, 46, 57, 74]. Reporting guidelines can also help standardise and improve how the use of automated techniques is reported, making methods more comparable and reproducible [18, 36, 61].

Improvements in technology and the performance of current methods for automation

The performance of available tools and the ability to further automate the evidence synthesis process are limited by the functionality of technologies. Some technologies are currently limited to title and abstract screening. and Expanding these tools to include full text of articles may aid in making them more effective [30, 47, 65, 133]. Better usability would also improve the use of automated technologies [74]. Improvements to natural language processing technology would allow for further automation and application to evidence syntheses, [27, 136, 139, 153], particularly for data extraction steps [74].

The lack of integration between different technologies and databases is a current limitation of automated techniques [27, 32, 40, 74, 86]. However, as text mining and other automated techniques become more routinely used, they will probably be combined to support a more integrated process by which the whole evidence synthesis process becomes more automated [22, 26, 28, 30, 33, 94, 127].

As mentioned above, the performance of automated techniques is often limited to the quality of training datasets and reference texts. Reviews of current machine learning techniques recommend that 'gold standard' training datasets covering an appropriate array of healthcare domains would be an important step towards improving the use of these techniques [18, 37, 111, 122]. Training sets in languages other than English would also be helpful in addressing gaps related to language in the use of automated reviews [19].

There are also other closely-related technologies whose improvement would help the ability to automate. For example, the ability to export large numbers of records (e.g. titles and abstracts) at a time from bibliographic databases, [20] web-crawlers to retrieve full text articles [111, 133] and improvements around machine-readable PDFs [28, 133] can improve the efficiency of some types of automated technologies for evidence syntheses.

How evidence is reported, stored and indexed

How research is reported can aid in the ability to automate or semi-automate parts of the review process. Reporting guidelines (e.g. Consolidated Standards of Reporting Trials [CONSORT]) and standardised data formats in published articles can improve algorithm performance, as they make data within reports more structured and machine readable [36, 38, 46, 53, 65, 74].

Several aspects relating to how research is stored and indexed can help make automation easier, including being able to use one tool to search multiple literature repositories/databases (which is a current challenge). Consistency in classifiers and improvements in indexing and ontologies make it easier to conduct automated reviews and to compare the performance of automated techniques [18, 39-41, 53, 70, 74, 111]. Using classifiers such as Medical Subject Headings (MeSH) terms or metadata associated with articles (e.g. referencing data] can improve the performance of algorithms, [18, 108, 133] particularly when these classifiers are general rather than specific [39]. Connections and interoperability across bibliographic databases [27, 36, 74] can also support automation.

The availability of other information, including access to data repositories associated with randomised trials and reviews, [27, 38, 74, 82] protocols, [65] preprints [75] and resources identifiers [28] can also improve the availability of information that automated technologies can then use to make decisions around screening. It may be more challenging to establish clear and consistent practices that make it easier for automated methods to be applied to screening for observational studies as compared to randomised trials, which can also have an impact in public health where more observational studies may be used to inform decision making. Lastly, open access publishing makes full texts of articles more available, which also facilitates automation [75].

Support and continued use of automated methods

As text mining and other automated techniques become more routinely used in the world of evidence synthesis, it is likely that their performance will improve [28, 38, 74]. To support routine use of these technologies, continued long-term funding and resources are needed to support tools that are available after funding or individual studies end, [28, 32, 33] as are communities around these technologies to support methods improvement [28] as well as support from funders to use these techniques [37]. These can help technologies move from a prototyping phase towards a more professionally maintained platform [33].

There can be challenges for researchers in accessing technology to support automation (e.g. cost of the technology and a lack of staff with sufficient expertise [22, 23, 32, 33]) and support from funders can be useful to address these challenges, as well as investment in infrastructure around automated technology, training for researchers and more access to freely available tools.

Potential future uses of automated technology

In the future, there will likely be applications of automated techniques along with human effort. For example, automated methods can be combined with crowd science (e.g. Cochrane Crowd), which can support mutual reinforcement of both techniques, with both the crowd and the algorithm training each other [28, 41, 74]. However, combing automated techniques with crowd science would require close supervision and quality control, [28] and some technical subjects may not be well suited to crowd science [22].

There is also potential for similar combinations of human and machine effort by performing a review with one reviewer being a researcher, and the second reviewer being an automated technology, [23, 27, 41] or by having an automated technique to resolve disagreements between two researchers [27]. Human and machine efforts can also be combined by using automated tools to prioritise or rank articles for screening, leading to reduced time and effort [74, 133] while still allowing human experts to retain control over the review process [23, 33, 133]. These combinations of human and machine effort are already being deployed, and the use of these types of blended methods may become more common in the future.

Additionally, existing technology may be applied to newer areas in the future, such as to policy development and analysis, [97] to problem formulation and search strategy tasks, [27, 74] or to living reviews that are automatically updated by technology [22, 28, 41, 45].

Results from the qualitative data collection and mapping exercise

An overview of the key findings of the qualitative data collection and mapping exercise are provided below.

- current approaches to evidence syntheses
- challenges with traditional approaches
- extent of the use of automation
- advantages of using automation
- challenges and needs in using automation
- factors that can support the use of automation.

Examples raised by participants are drawn on throughout this chapter to illustrate key themes.

Current approaches to evidence syntheses and changes during the pandemic

There were a range of responses when participants were asked about their current practices conducting evidence synthesis/literature reviews. Some teams made use of existing tools that automated multiple parts of the evidence synthesis process, like Covidence or DistillerSR, while others used several different tools in a customised workflow. Smaller organisations, and national public health bodies in smaller countries, reported having made it through COVID-19 using predominantly traditional (non-automated) review methods, but are now beginning to explore the possibility of using automation. While two survey respondents conducted their own evidence syntheses and literature reviews, neither of these respondents had used technology to automate or semi-automate parts of the evidence synthesis or literature review process.

Not all participants conduct evidence syntheses themselves. Some participants noted that they use reviews produced by other organisations (e.g. ECDC, WHO, universities, other public health authorities) or contract out other organisations to conduct reviews on their behalf. All nine survey participants stated that they made use of evidence syntheses and literature reviews produced by other organisations and groups.

The primary change noted by focus group and interview participants as a result of the pandemic in terms of being able to review evidence, was the volume of research produced in relation to the pandemic, as well as the greater number of living reviews (systematic reviews which are updated with relevant information as it becomes available). The volume of new research produced on COVID-19 often made it difficult to keep up with the production of a living review when using traditional approaches.

National public health bodies were also faced with the need for a fast turnaround on work such as public health guidelines. In many cases, the capacity of those bodies was insufficient to be able to apply systematic evidence synthesis methods and processes, or to learn to make use of automated technologies, as they were forced to adapt and produce outputs within a matter of days. This need for a quick turnaround makes it difficult to build the skills and workflows needed to support the use of new tools, even if those tools may eventually lead to time and cost savings.

Challenges of conventional literature review

There were several challenges raised in the focus groups, interviews and survey in relation to traditional approaches to reviewing literature, including volume, time and duplication.

The most significant challenge in the production of conventional systematic literature reviews was the large volume of literature, particularly in relation to health emergencies such as the COVID-19 pandemic and newly emerging monkeypox. It was described by focus group participants as 'impossible' to manage all of the incoming information, with automation being referred to as a necessity for some research purposes. In addition, eight of the nine survey participants agreed that a large volume of literature was a key challenge in keeping up to date on public health research (Figure 4).

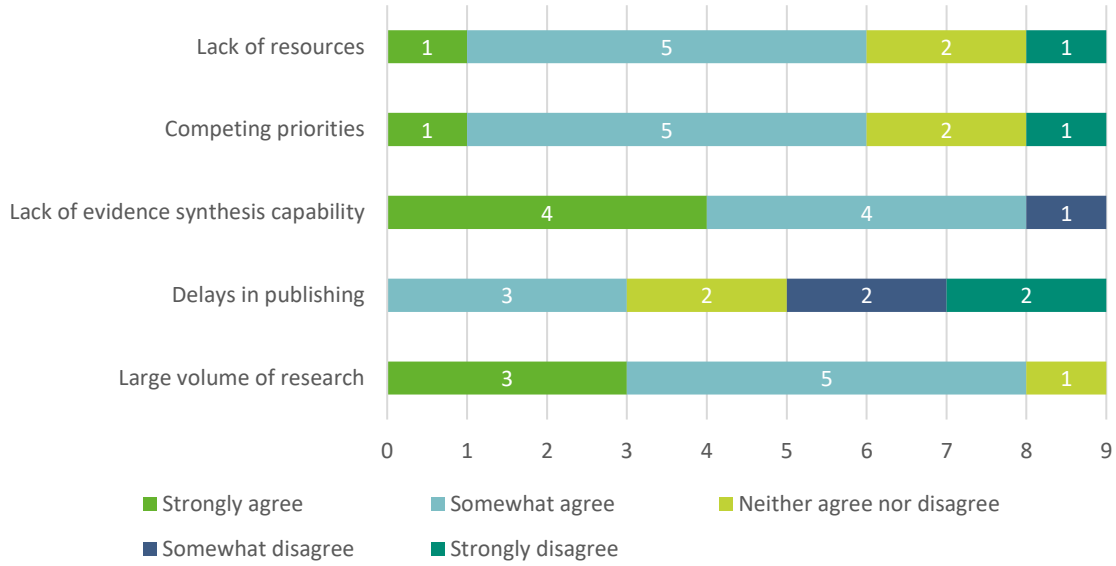
The time required to produce systematic reviews by conventional means was also described as a problem, especially regarding fields of research which are fast-moving, such literature related to the pandemic. This also means that evidence can be published after the point at which it is needed and potentially be outdated. For example, during the pandemic, some reviews became out of date within a matter of days. Similar concerns about outdated reviews were raised in relation to the emerging monkeypox situation. However, the findings from the

survey were more mixed when respondents were asked about delays to publishing being a challenge (Figure 4). While three of the nine respondents agreed that this was a challenge, four disagreed.

Participants also noted how duplication of efforts was limiting the efficiency of traditional systematic reviews, leading to multiple reviews being conducted of the same trials by different research groups simultaneously.

The survey provided insights to other challenges in conducting reviews. For example, survey respondents indicated that barriers to keeping up to date with evidence include a lack of resources (six of the nine respondents agreed this was a challenge), other competing priorities (six respondents) and a lack of evidence synthesis capability (eight).

Figure 4. Degree to which respondents agreed with 'reasons why it might be difficult to keep up to date with evidence' (n=9)



Extent of the use of automation

The extent to which participants used automation included not at all, assisting with specific tasks or prioritising records which had already been manually selected, the system making decisions about inclusion and exclusion and full automation in one case. Some of those who had not used automation reported that this was something their organisation was exploring.

Although focus group and interview participants had experience using technologies, of the nine survey participants, only two reported using public health evidence syntheses or literature reviews which were conducted using automated technology. The remainder of respondents reported either not knowing if the work they had utilised had been conducted using these technologies (four respondents) or not having made use of such work at all (three).

Those who did not use automation for evidence synthesis did not seem to be against its use to conduct reviews they would use for their own work. By some, it was felt that the reputation of the organisation/authors of the review was more important than the tools used to conduct the review. The Public Health Agency of Canada even goes so far as to mandate the use of software when contractors conduct systematic reviews on their behalf. Of the two survey respondents who conduct evidence syntheses but have not used automated technologies, one respondent reported that this was because of a lack of skills or knowledge within the organisation, lack of resources and lack of time and capacity to implement and learn how to use a new technology.

Software suites

A variety of different software packages were mentioned during the interviews and focus groups, with the most popular being comprehensive packages such as Covidence, DistillerSR and EPPI-Reviewer. The collaborative screening tool Rayyan was also mentioned frequently by participants. These pieces of software include functions and tools for many different stages in the production of a systematic review, although most participants reported using them primarily for screening.

Covidence was described as performing well and integrating well with other programmes like Review Manager although it was also described as being prone to over-inclusion and therefore requiring more human input to correct mistakes.

EPPI-Reviewer was described as a 'good' tool, but one designed for reviews in the social sciences or education and lacking the adaptability shown by other software. One participant described using EPPI-Reviewer to run automated searches of new literature but stated that it produced too many search results to be useful.

Rayyan, used primarily for collaborative screening, was described as 'a very good tool for teamwork and reviewing large amounts of abstracts' but was also suggested to not be as advanced as some of the more comprehensive tools.

While only one focus group participant had experience of using DistillerSR, they spoke positively of the range of tasks the programme could help with – abstract/title and full-text screening, reference management, risk-of-bias – and how it could be adapted to the needs of the project.

A huge range of other software packages were mentioned by participants during interviews and focus groups, highlighting the wide range of choice for such software. This includes packages like ASReview, RevMan and MAGICapp, and dedicated tools for performing specific parts of the review process such as Citation Chaser and the search-translation tool Polyglot.

Advantages of using automated approaches to conduct evidence syntheses

Across the focus groups and interviews, participants noted a range of benefits of using automation to (partially or fully) conduct evidence syntheses. This included a reduction in burden, ease of use and completion of new tasks. There are no survey findings to include in this section as none of the respondents who had used automated technologies for evidence synthesis reported any advantages to note here.

Reduction in burden

The most common advantage of automation tools mentioned in the interviews and focus groups was the reduction in the burden and cost of conducting literature reviews (see Box 8 below), as well as reduced error rate. Participants stated that automated tools had helped streamline the overall review process, in particular reducing the time needed for screening articles. This enables reviews to be conducted and published quicker, allowing evidence to inform (rapid) decision and policy making more effectively.

Box 8: Bond University's development of a two-week systematic review

An interviewee from Bond University, Australia, reported on the development of a two-week systematic review process which aims to disseminate evidence quickly by using automation tools. The project used a small team of four experienced clinicians and systematic review specialists, who utilised agile project management techniques and automation for protocol writing, searches, screening citation and write-up. While this technique did facilitate the completion of systematic reviews in a matter of weeks (e.g. 66 person-hours for a review of eight studies), researchers reported concerns about trust and performance. The interviewee noted the importance of transparency and evaluation, and the difficulty of engaging with funders to help them understand the value of automation tools.

One participant described how a COVID-19 research project had begun to integrate automation for different topic areas and classifiers until eventually, the process became fully automated and now collects research without the need for human input. Another participant reported how the use of an automated search refiner took one search from 11 500 papers down to 800, drastically reducing the associated workload in screening.

Ease of use

While automated software packages such as DistillerSR and Covidence were described as requiring time to train staff and to integrate into research workflows, they were also described as 'intuitive' and 'easy to use'. This may indicate a low barrier for entry into the use of automation in systematic reviews, even if specific aspects (such as reporting methods or carrying out specific tasks) need to be carefully considered (although staff knowledge and training was raised as a key barrier and is discussed elsewhere).

Enables completion of new tasks

Automation was also described as being used to perform a range of review tasks which would either be too taxing for a review team to complete, or which would otherwise take a significant amount of time to do manually. This includes backward-and-forward citation searching, the extraction of PDFs from research databases, and translating searches to extract information in multiple different languages.

It was also noted by several participants that software suites such as Covidence assisted with the less tangible aspects of the review process, such as thinking and discussion about future directions. It can also encourage collaboration, as a tool can be used by multiple (international) organisations for the same review.

Challenges and needs in using automation for evidence syntheses

While automation has the potential to offer several benefits in the mid- and long-term when conducting literature reviews, there are some notable barriers and challenges that need to be considered and understood. These include: continued need for human input, technical limitations, need for staff training/capacity, cost and equipment/resource requirements, and lack of trust and performance evaluations.

Continued need for human input

In almost all cases where participants had used automated approaches, it was stated that human input was still required. This was in part to ensure that the technology worked correctly, but also to provide research credibility. As one participant noted, journals may be concerned about researchers trying to publish an entirely automated review. Another participant noted that these concerns have been an obstacle in the procurement of software, and that it was important to 'be clear that you still need to use humans to do the job' rather than giving the impression that the software is intended to do everything.

Where full automation is possible, it was the result of a significant amount of work on the part of researchers (see Box 9 below). One participant described how their team had successfully implemented a fully automatic search and record-categoriser but noted that this was 'not necessarily reproducible elsewhere' and was the result of over a year of colleagues 'painstakingly labelling records' by hand.

Box 9. Example of use of automation that still requires human input

A participant from the Danish Health Authority reported on the use of Covidence for a range of tasks relating to evidence synthesis of public health topics, particularly screening and extraction. Covidence was used to make daily searches using a randomised control trial (RCT) classifier, which distinguishes between reports of RCTs and non-RCTs. While Covidence rarely missed records, it was noted that there is still noise produced and it tends to over-include and that the results produced have to be checked manually. Errors that were made by Covidence (and other tools) had to be identified and fixed by the Danish Health Authority.

Participants made clear that there is always a trade-off in using AI or automation tools which may complete a task more quickly, but not do it to as high a standard as humans. One participant stated that they 'might' use RobotReviewer as a second risk-of-bias reviewer when working to a tight deadline. Participants suggested that the need for continued human input meant the use of these tools was still outside the scope of some organisations.

Technical limitations

The technical limitations of automation software were mentioned by several participants. This includes crashes and troubleshooting, and the fact that automation tools cannot necessarily be adapted or used in different research topics or disciplines without being trained first (see Box 10 below).

Box 10. Example of difficulties using training sets across research topics

A participant from the EPPI-Centre reported on the development and evaluation of systematic review tools in response to the COVID-19 pandemic. Using Microsoft Academic and then OpenAlex, the team manually labelled studies relating to COVID-19 which are now used to train machine learning algorithms that automatically search for and categorise COVID-19 studies. For each new automation tool, the participant reported that the team conducted meetings to discuss and plan the next steps and evaluate discussion documents. Context was said to be paramount, as many of the evaluations that have been conducted only demonstrate if a tool was useful in a specific context. One tool the participant built worked effectively on MedLine and PubMed as it was trained on those databases, but did not work when applied to another field of research.

Another issue noted by participants is that no single tool is able to automate the entire review process, requiring different tools to be able to transfer data to each other in a legible format. Ensuring that the programmes can speak to each other also requires 'computer skills to manage interoperability', increasing the barrier for entry into using automation tools.

It was noted that while large sections of the literature review pathway could feasibly be automated, data extraction was a 'bottleneck point' in the process for which automation software had not yet provided an adequate solution. Automation was also reported to be of limited use in writing up aspects of a review, although some aspects can be semi-automated (e.g. writing up methods).

Other issues presented by the use of automation tools discussed in focus groups and interviews included data protection issues stemming from inputting sensitive healthcare data into online tools, and AI pushing users down a specific path of decision-making (perhaps while also not understanding how the tool made these decisions).

Need for staff training and capacity

While we noted earlier that researchers can often find automated tools user-friendly and easy to use, participants from public health authorities discussed how the supplementary knowledge required to use automation tools (e.g. IT skills, coding) or the lack of familiarity with the types of tools required staff to be trained on using the technology before they could work on literature reviews.

Despite the stated importance of having someone with expertise supervise the introduction of automation tools, some teams reported not having access to such expertise at all. This is especially true of smaller organisations or public health authorities working in smaller countries with less developed research networks. The time and money to train new staff, and the number of staff required, also represents a significant obstacle to the adoption of automation tools, especially during emergencies. One participant reported that during COVID-19, their team was forced to use the 'lowest common denominator' technologies because they did not have time to learn the more complex functions. It was described as a long and sometimes difficult process to upskill researchers to the extent to which automation technology begins to save time.

Several participants also noted how the use of automated tools may present a 'culture shock' to researchers, and that staff need to be reassured that their work and skills are still going to be valued in a context where greater technology is going to be used.

Cost and equipment/resource requirements

The price of some automation tools was described as a 'big barrier' to their adoption to conduct systematic reviews, as most of the popular automation tools and packages are not free. One participant noted that collaboration between different organisations on licensing software may be one way to overcome this obstacle (where this is within the requirements of the licensing terms and condition).

Beyond the procurement of the software itself, it was noted that it was difficult to get funders to recognise the resources required to make systematic reviews 'more efficient' and that this was not being seen as a priority.

Two participants described examining the use of automation tools for their research but found that it was not workable due to the staff capacity required to implement and use the tools.

Lack of trust and performance evaluation

The need for greater trust in automation tools was another theme in focus groups and interviews. Participants stated that automation could at times be like a 'black box' to researchers, and discussed a lack of 'transparency of how articles are collected, and understanding [of] the scope of how many articles may be missed' in the review process.

The lack of communication between different organisations was also cited as a reason for the lack of trust in these new technologies, especially with so many different tools and pieces of software in use. One participant noted that 'everyone thinks everyone is doing it differently' and that 'more unity' between organisations might improve the situation.

The lack of quality evaluations of automated technologies was also noted which made it difficult to understand how well the tool perform and their accuracy and precision. It was also noted that evaluations of tools are often conducted in relation to specific topics/research questions, making it difficult to determine the generalisability of the technology.

The two survey respondents who had conducted evidence syntheses or literature reviews using automated technology agreed that they trusted the information produced by (semi-)automated reviews. By contrast, of the participants who did not report making use of such evidence syntheses, four agreed that they trusted information from (semi-)automated reviews, one neither agreed nor disagreed and the other two did not know if they trusted this information.

Factors that can support the use of automation

A number of gaps and needs were discussed by participants in relation to the use of automation for evidence synthesis. These primarily related to skills, knowledge and experience, resources and collaboration.

Skills, knowledge, and experience

Many participants noted how being able to share knowledge and experience around automated or semi-automated reviews could support their use in systematic reviews. Of particular importance was providing time and space for researchers to become familiarised with a new system, and to gain a 'sense of control' in utilising new tools. Some participants noted the value of embedding someone within the team with expertise or knowledge about such tools, who can 'mentor' or work closely alongside the researchers. Demonstrations of how to use the tool were also helpful. Both survey respondents who conduct evidence syntheses as part of their role said that training would enable them to make greater use of technologies.

Another theme arising from the interviews and focus groups was that researchers need training and a comprehensive grasp of the theory and methodology behind the conducting high-quality research, in order to be able to utilise automation tools successfully.

Resources

Participants in the qualitative data collection exercise noted that support from organisational leadership and human resources could be key in helping them use automated systematic review tools. For example, staff being familiarised with and having a say in how their work environment may change as a result of the implementation of automation, and having enough staff members with the capacity to implement the technology, can be helpful. Some participants noted that having smaller teams with more dedicated time to conduct (automated) reviews can help streamline the process and make it more manageable. One survey respondent cited the value in 'dedicated human resources to the purpose of searching for national and international scientific evidence'.

It was also noted how deliberate and careful discussion of new tools could facilitate and support their adoption, including clear and open communication with staff on the change and what it would mean for them and their role (see Box 11 below).

Box 11. Example of clear communication to staff about changes in processes

A participant from the Norwegian Institute of Public Health reported on efforts in the Machine Learning Implementation team to prepare for the implementation of machine learning in response to research waste and the overwhelming volume of research being published related to COVID-19. The team investigated published evaluations for two automated tools, considered if they would need to produce their own evaluation and used a 'superuser' to trial each piece of software. Outcome measures which could be used to evaluate performance and inform decisions about the use of technology in reviews were established by the research team, and the team began collecting data about resource use and time efficiency. Finally, the team consciously implemented a change communication strategy, highlighting what researchers could expect to change and affirming that their valuable skills would continue to be recognised. Machine learning is now an essential part of the work conducted by the Norwegian Institute of Public Health, amounting to around 80 reviews per year.

Several participants using different data collection methods noted the value of shared repositories of information on the use of different automated tools for systematic reviews. This includes the SR toolbox, a repository of information about what different tools can be used for, and a proposed database of different strategies and datasets for different areas of research.[5] Recommendations on the most useful technologies was raised by one survey respondent as a factor that would support their use of automation. In addition, repositories of papers organised by topics would be helpful with, e.g. labels, data already extracted, reasons for exclusion, to develop metadata and support the use of papers with automated tools. One survey respondent mentioned the value of having 'easy access' to these sources via a 'dedicated platform'.

Some tools require organisations to buy software packages, or licenses to use them, which can also be a barrier to using automated tools, especially for smaller organisations with fewer resources, which was reflected by some focus group participants. Using tools that are freely available can therefore also support implementation of automated tools across different types of organisations.

Collaboration

The value of coordination and collaboration across organisations was also a significant theme in focus groups and interviews. For example, participants discussed how 'fostering relationships' between policymakers/public health authorities and method specialists and sharing learnings about effective tools and pros/cons of certain tools could facilitate better work in an emergency. Collaboration with organisations who are experienced in using a particular tool can also be useful. The two survey respondents who conduct evidence synthesis as part of their role noted that they would value collaboration between ECDC and their organisations.

Examples of collaborations with automated evidence synthesis mentioned by participants from public health authorities included:

- Danish Health Authority working with the National Institute for Health and Care Excellence and Canadian researchers on post-COVID-19 illness using Covidence.
- The Health Technology Assessment arm of the Galician Health Service has been working with the Spanish Centre of Disease Control to deal with the pandemic. They are currently working on an expedited systematic review on non-pharmaceutical interventions for COVID-19.
- The Norwegian Health Authority working with the EPPI-Centre on how to introduce automation into their evidence synthesis work.

Finally, as noted above, it was discussed that collaboration between organisations on the licensing of software could be a way to provide smaller teams with access to tools which they might otherwise not be able to afford.

[5] Marshall, C., Sutton, A., O'Keefe, H., Johnson, E. 2022. The Systematic Review Toolbox. Available from: http://www.systematicreviewtools.com/

Discussion and conclusions

Based on the qualitative data collected from EU/EEA public health bodies participating in this study, we found that automated technology is used to a limited extent for conducting evidence synthesis, with most participants not currently using these types of tools. This may be because they do not conduct any type of review within their organisation (instead, relying on reviews produced by others) or because of different challenges and obstacles that can arise when attempting to use automated approaches (discussed below). However, there were a few public heath bodies who used automated tools, sometimes extensively. Among them, technologies were used in different stages of the evidence synthesis process, although they seem to be particularly used to semi-automate the screening stage of reviews. For other public health institutes, the adoption of new technologies was being explored for the future.

Overall, this study shows that new technologies have the potential to improve the evidence synthesis process. Most notably, the literature review and qualitative work highlighted how technologies have the potential to allow for more efficient reviews in terms of time, effort and resources, which can help organisations working in public health and infectious diseases keep up with the ever-expanding body of evidence to inform (sometimes rapid) decision making. This has been particularly relevant during the COVID-19 pandemic which has seen hundreds of thousands of publications on the topic. For example, one focus group participant reported that the use of an automated tool for conducting a literature search reduced the number of articles to be screened by over 90%. The literature review found that technologies also have the potential to improve reviews by supporting increased precision and accuracy (e.g. by reducing human error or double counting, or by facilitating systematic extraction and analysis of unstructured data), and by allowing human effort to be focused on synthesis and interpretation. Study participants and the literature described how automation can also allow for more living reviews to be produced, whereby the same review is updated with new evidence periodically, as well as conduct tasks that would usually be too time-intensive when conducted manually. Participants also noted the benefit automated tools can provide in supporting collaboration as many can be used by multiple researchers across different organisations/countries at the same time.

Despite the potential of new technologies, there remain challenges in using these technologies for evidence synthesis. Both the literature review and qualitative work found that human input is still required to conduct evidence syntheses, even where new technologies are also used, for various tasks such as developing training sets, checking the accuracy of the tool and adding credibility to the study. One participant noted that this challenge has created obstacles to procuring new tools as they do not have the staff capacity for this type of input. An example of another challenge was shared by representatives from the Danish Health Authority whereby automated technologies were embedded in the evidence synthesis process, but this was reported as having taken huge amounts of time and effort from staff to do so. The need for (sometimes extensive) human input to adopt and use automated technologies, particularly when first introducing them, can be a barrier to their use, particularly due to capacity constraints within an organisation (see below).

While evidence from the literature suggested that researchers generally find automated tools easy to use and user-friendly, the experiences described by study participants working in public health authorities was different and it was felt more support was needed in adopting new tools. Participants discussed how the use of automated tools for reviews requires staff to both be trained and have the capacity to appropriately use the technology. This training requires resources on top of what organisations will spend on the technologies or licenses themselves, which can be a barrier to adopting technology, particularly for smaller organisations or public health bodies. Ensuring staff receives the time and appropriate resources such as training material or tutoring to acquire the appropriate skills and get familiar with the new tools can also be a challenge if the organisation lacks some of the necessary knowledge in-house already. The time needed to train staff and implement new ways of working was highlighted as often requiring significantly more effort than anticipated before automated technologies would start to demonstrate the positive impacts described above, particularly saving time in conducting reviews. Participants also discussed that that capacity is also a challenge as public health staff are often very busy and lack the breathing space to learn about a new process. This has been particularly evident during the COVID-19 pandemic and the resulting pressures on public health bodies. Participants from two organisations described this challenge and how the capacity needed to effectively introduce the tool was not feasible at that time, despite there being a willingness to do it. In addition to training and capacity, focus group participants highlighted how adoption of new technologies represents a change to usual ways of working and participants flagged that there can be concerns about job security, as well as 'culture shock' for staff.

Both the literature review and qualitative work noted that even when technologies work well in one review or for a particular topic, this does not mean that they will perform the same for other reviews or topics, which points to the need for human input to help determine whether technology is appropriate for particular tasks. There are also concerns around transparency, trust and reproducibility, particular for machine learning technologies, which can prevent them from being used. Additionally, although there is evidence that particular technologies can be accurate and precise (even in comparison to human reviewers), there is heterogeneity in the performance of new technologies, which can cause concerns around generalisability. This was raised by multiple participants who described issues such as reviews not being reproducible when using automated technologies (or tools are trained for one specific review topic and may not be effective for other topics without significant time spent re-training the algorithm) or, as mentioned, needing to check the accuracy of the tools decision-making. However, there are ways to improve how automated technologies perform and how results are reported, by using diverse and representative training data sets to reduce bias in algorithms or by explaining decision-making processes (explainable AI, XAI), for example, in order to address issues around transparency and trust in results produced with the support of automated tools, and eventually increase the use of automated technologies .

The literature review and qualitative work identified several ways in which the use of technologies for evidence synthesis can be supported, and ensure that technologies create a positive impact in terms of the ability to stay up to date with evidence in the areas of public health and infectious diseases. These are in the areas summarised below:

- **Training and skills:** Many organisations that may benefit from using new technologies for evidence synthesis may lack the skills to incorporate such technology. Training and support, both in terms of conducting high quality reviews in general and in the use of technology, can be helpful. Being able to share knowledge across organisations trying to implement similar new processes was also flagged by participants as being of use.
- **Time and space,:** Organisations working in the area of public health may lack the capacity to conduct reviews themselves, or to incorporate technology into these reviews. Offering staff the time (and space) to learn how to use a new tool effectively and be able to adapt to a new way of working are important. Participants noted that smaller, more confined teams with dedicated time to tasks that use automation may be the most effective approach.
- **Leadership support:** Leadership buy-in and support in the adoption of a new technology can also be beneficial, such as by creating the dedicated time to work with the technology and making it more of a priority. Leadership support may also encourage the introduction of training for staff on using the technology. The introduction of a new technology also needs careful management by leadership to ensure there is staff buy-in and staff are made aware of why the change is being introduced and how it will change their role (if at all). A participant from one public health institute discussed how this was handled effectively during the introduction of a new machine learning tool, whereby a change communication strategy was implemented (see Chapter 4).
- **Financial resources:** While some automation tools are available for free, many must be purchased and/or are associated with licence fees. Financial support for public health bodies may improve their ability to keep up-to-date with developments in their field, especially for smaller and less well-resourced organisations. Organisations could also take advantage of freely available tools, should they prove to be robust. Collaboration and relationship building with organisations that already have a technology licence could be beneficial particularly to smaller public health organisations.
- **Collaboration:** There are a number of active collaborations that are working to improve how technologies are used in evidence syntheses and reviews (e.g. CAMARADES – The Collaborative Approach to Meta Analysis and Review of Animal Experimental Studies or ICASR - International Collaboration for the Automation of Systematic Reviews). Participating in these collaborations can allow organisations working in public health to contribute to this space, and gain skills and confidence in using new technologies in reviews. Coordination between EU/EEA Member States public health bodies can also allow for sharing of recommendations of tools, lessons and best practice, as well as shared repositories, between national public health competent authorities. Setting up these collaborations now would be particularly useful for future emergency/crisis situations (as seen with COVID-19) to facilitate rapid sharing of learning and knowledge.
- **Standards to ensure technologies are used appropriately:** One concern that prevents reviewers from using new technologies is that reviews will be perceived as lower quality. New/updated standards could be established around new technologies to help ensure that new technologies are used appropriately to support high quality reviews, and can help communicate robustness to relevant stakeholders such as those that use evidence from reviews and academic journal publishers. Relatedly, it may also be useful to include the use of automation in reporting guidelines, such as PRISMA (the Preferred Reporting Items for Systematic Reviews and Meta-Analyses), to improve the transparency of the methodology. In addition, as we found little reflection in the literature on the actual use of automated tools, reviews that use these approaches could be encouraged by journals to report and reflect on the practical and performance aspects of the tool(s) to support the sharing of learning and experiences.

- **Interventions that improve the performance of new technologies:** The performance of some technologies depends on the availability of high quality data sets for training algorithms. Creating these datasets can help improve the performance of new technologies in particular subjects relating to public health and infectious diseases. Reporting standards and improvements in how literature is indexed and stored can improve machine-readability of literature to allow for more automation. Increased interoperability between technologies can also support automation.

This study explored how new technologies have been used to automate or semi-automate parts of the evidence synthesis process, the impact of using these technologies, and how the use of technology could be supported within the EU and EEA for public health purposes. The systematic review for this study gave insight into the broad experiences of those using technology for evidence syntheses, while qualitative data provided in-depth insight into what lessons and support may be useful for other organisations looking to adopt similar types of new technologies.

Improving technologies for evidence synthesis, and supporting organisations to make the best use of available technology, can improve public health decision making. The increasing pace of evidence production and the widespread impact of public health decisions during the COVID-19 pandemic points to the importance of making the evidence synthesis process as efficient as possible, while maintaining a high quality standard.

This study has revealed resources and support that can both help improve new technologies and ensure that these technologies are used in an appropriate way to support the evidence synthesis process. Upskilling, dedicated resources, leadership support and collaborative ways of working were just some of the ways in which public health bodies, particularly smaller ones, could be better supported to take advantage of the benefits automated technology could provide in conducting evidence synthesis.

References

1. Gough D, Davies P, Jamtvedt G, Langlois E, Littell J, Lotfi T, et al. Evidence Synthesis International (ESI): Position Statement. Systematic Reviews. 2020;9(1):155.
2. van Altena AJ, Spijker R, Olabarriaga SD. Usage of automation tools in systematic reviews. Research synthesis methods. 2019;10(1):72-82.
3. Cochrane Library. About Cochrane Reviews 2022. Available from: https://www.cochranelibrary.com/about/about-cochrane-reviews.
4. Scott AM, Forbes C, Clark J, Carter M, Glasziou P, Munn Z. Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey. Journal of clinical epidemiology. 2021;138:80-94.
5. Arno A, Elliott J, Wallace B, Turner T, Thomas J. The views of health guideline developers on the use of automation in health evidence synthesis. Systematic reviews. 2021;10(1):16.
6. Beller E, Clark J, Tsafnat G, Adams C, Diehl H, Lund H, et al. Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). Systematic reviews. 2018;7(1):77.
7. Khalil H, Tamara L, Rada G, Akl EA. Challenges of evidence synthesis during the 2020 COVID pandemic: a scoping review. Journal of clinical epidemiology. 2022;142:10-8.
8. Elliott J, Lawrence R, Minx JC, Oladapo OT, Ravaud P, Tendal Jeppesen B, et al. Decision makers need constantly updated evidence synthesis. Nature. 2021;600(7889):383-5.
9. Global Commission on Evidence to Address Societal Challenges. The Evidence Commission report: A wake-up call and path forward for decision-makers, evidence intermediaries, and impact-oriented evidence producers. 2022.
10. Norwegian Institute of Public Health. Aims and strategy for the implementation of machine learning in evidence synthesis in the Cluster for Reviews and Health Technology Assessments for 2021-2022. Oslo: Norwegian Institute of Public Health.; 2021.
11. O'Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Wolfe MS. Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR). Systematic reviews. 2018;7(1):3.
12. O'Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Shemilt I, Thomas J, et al. Still moving toward automation of the systematic review process: a summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR). Systematic reviews. 2019;8(1):57.
13. Shea BJ RB, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. . AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:J4008.
14. Baethge C, Goldbeck-Wood, S. & Mertens, S. . SANRA—a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019;4(5).
15. Tyndall J. AACODS Checklist. 2010.
16. SmartSurvey. Available from: https://www.smartsurvey.co.uk/.
17. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. International journal of methods in psychiatric research. 2016;25(2):86-100.
18. Abdelkader W, Navarro T, Parrish R, Cotoi C, Germini F, Iorio A, et al. Machine Learning Approaches to Retrieve High-Quality, Clinically Relevant Evidence From the Biomedical Literature: Systematic Review. JMIR medical informatics. 2021;9(9):e30401.
19. Mishra R, Bian J, Fiszman M, Weir CR, Jonnalagadda S, Mostafa J, et al. Text summarization in the biomedical domain: a systematic review of recent research. Journal of biomedical informatics. 2014;52:457-67.
20. Shemilt I, Simon A, Hollands GJ, Marteau TM, Ogilvie D, O'Mara-Eves A, et al. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. Research synthesis methods. 2014;5(1):31-49.
21. Wagner G, Lukyanenko R, Paré G. Artificial intelligence and the conduct of literature reviews. Journal of Information Technology. 2021;0(0):02683962211048201.
22. Marshall IJ, Johnson BT, Wang Z, Rajasekaran S, Wallace BC. Semi-Automated evidence synthesis in health psychology: current methods and future prospects. Health psychology review. 2020;14(1):145-58.
23. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic reviews. 2015;4:5.
24. Whelan J, Ghoniem M, Medoc N, Apicella M, Beck E. Applying a novel approach to scoping review incorporating artificial intelligence: mapping the natural history of gonorrhoea. BMC medical research methodology. 2021;21(1):183.

25. Xu M, Du J, Xue Z, Kou F, Xu X. A semi-supervised semantic-enhanced framework for scientific literature retrieval. Neurocomputing. 2021;461:450-61.

26. Amezcua-Prieto C, Fernandez-Luna JM, Huete-Guadix JF, Bueno-Cavanillas A, Khan KS. Artificial intelligence and automation of systematic reviews in women's health. Current opinion in obstetrics & gynecology. 2020;32(5):335-41.

27. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Systematic reviews. 2014;3:74.

28. Bannach-Brown A, Hair K, Bahor Z, Soliman N, Macleod M, Liao J. Technological advances in preclinical meta-research. BMJ open science. 2021;5(1):e100131.

29. van Dinter R, Tekinerdogan, B., & Catal, C. Automation of systematic literature reviews: A systematic literature review. ure review Information and Software Technology. 2021;136.

30. van den Bulk LM, Bouzembrak Y, Gavai A, Liu N, van den Heuvel LJ, Marvin HJP. Automatic classification of literature in systematic reviews on food safety using machine learning. Current research in food science. 2022;5:84-95.

31. Ferdinands G. AI-Assisted Systematic Reviewing: Selecting Studies to Compare Bayesian Versus Frequentist SEM for Small Sample Sizes. Multivariate behavioral research. 2021;56(1):153-4.

32. Khalil H, Ameen D, Zarnegar A. Tools to support the automation of systematic reviews: a scoping review. Journal of clinical epidemiology. 2021;144:22-42.

33. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Systematic reviews. 2019;8(1):163.

34. Alla K, Oprescu F, Hall WD, Whiteford HA, Head BW, Meurk CS. Can automated content analysis be used to assess and improve the use of evidence in mental health policy? A systematic review 16 Studies in Human Society 1605 Policy and Administration 11 Medical and Health Sciences 1117 Public Health and Health Services. Systematic Reviews. 2018;7(1).

35. Stansfield C, O'Mara-Eves A, Thomas J. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. Research synthesis methods. 2017;8(3):355-65.

36. Bashir R, Dunn AG. Software engineering principles address current problems in the systematic review ecosystem. Journal of clinical epidemiology. 2019;109:136-41.

37. O'Connor AM, Tsafnat G, Thomas J, Glasziou P, Gilbert SB, Hutton B. A question of trust: can we build an evidence base to gain trust in systematic review automation technologies? Systematic reviews. 2019;8(1):143.

38. Adams CE, Polzmacher S, Wolff A. Systematic reviews: work that needs to be done and not to be done. Journal of evidence-based medicine. 2013;6(4):232-5.

39. Vergara-Merino L, Verdejo C, Carrasco C, Vargas-Peirano M. Living systematic review: new inputs and challenges. La revision sistematica viva: nuevos aportes y desafios. 2020;20(11):e8092.

40. Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Research synthesis methods. 2011;2(1):1-14.

41. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. Journal of clinical epidemiology. 2017;91:31-7.

42. Buchter RB, Weise A, Pieper D. Reporting of methods to prepare, pilot and perform data extraction in systematic reviews: analysis of a sample of 152 Cochrane and non-Cochrane reviews. BMC medical research methodology. 2021;21(1):240.

43. Blaizot A, Veettil SK, Saidoung P, Moreno-Garcia CF, Wiratunga N, Aceves-Martins M, et al. Using artificial intelligence methods for systematic review in health sciences: A systematic review. Research synthesis methods. 2022.

44. Mohan K, Alam A, Patnaik R. A review on use of automation in systematic reviews for scientific evidence generation short title: An overview of automation in systematic reviews. European Journal of Molecular and Clinical Medicine. 2021;8(2):1736-55.

45. Park SE, Thomas J. Evidence synthesis software. BMJ evidence-based medicine. 2018;23(4):140-1.

46. Schmidt L, Olorisade BK, McGuinness LA, Thomas J, Higgins JPT. Data extraction methods for systematic review (semi)automation: A living systematic review. F1000Research. 2021;10:401.

47. Popoff E, Besada M, Jansen JP, Cope S, Kanters S. Aligning text mining and machine learning algorithms with best practices for study selection in systematic literature reviews. Systematic reviews. 2020;9(1):293.

48. Jonnalagadda S, Petitti D. A new iterative method to reduce workload in systematic review process. International journal of computational biology and drug design. 2013;6(1-2):5-17.

49. Tsafnat G, Glasziou P, Karystianis G, Coiera E. Automated screening of research studies for systematic reviews using study characteristics. Systematic reviews. 2018;7(1):64.

50. Norman CR, Gargon E, Leeflang MMG, Neveol A, Williamson PR. Evaluation of an automatic article selection method for timelier updates of the Comet Core Outcome Set database. Database : the journal of biological databases and curation. 2019;2019.

51. Dhrangadhariya A, Hilfiker R, Schaer R, Muller H. Machine Learning Assisted Citation Screening for Systematic Reviews. Studies in health technology and informatics. 2020;270:302-6.

52. Bekhuis T, Demner-Fushman D. Towards automating the initial screening phase of a systematic review. Studies in health technology and informatics. 2010;160(Pt 1):146-50.

53. Bekhuis T, Demner-Fushman D. Screening nonrandomized studies for medical systematic reviews: a comparative study of classifiers. Artificial intelligence in medicine. 2012;55(3):197-207.

54. Bekhuis T, Tseytlin E, Mitchell KJ, Demner-Fushman D. Feature engineering and a proposed decision-support system for systematic reviewers of medical evidence. PloS one. 2014;9(1):e86277.

55. Brassey J, Price C, Edwards J, Zlabinger M, Bampoulidis A, Hanbury A. Developing a fully automated evidence synthesis tool for identifying, assessing and collating the evidence. BMJ evidence-based medicine. 2021;26(1):24-7.

56. Almeida H, Meurs M-J, Kosseim L, Tsang A. Data Sampling and Supervised Learning for HIV Literature Screening. IEEE transactions on nanobioscience. 2016;15(4):354-61.

57. Bekhuis T, Tseytlin E, Mitchell KJ. A Prototype for a Hybrid System to Support Systematic Review Teams: A Case Study of Organ Transplantation. Proceedings IEEE International Conference on Bioinformatics and Biomedicine. 2015;2015:940-7.

58. Buchlak QD, Esmaili N, Bennett C, Farrokhi F. Natural Language Processing Applications in the Clinical Neurosciences: A Machine Learning Augmented Systematic Review. Acta neurochirurgica Supplement. 2022;134:277-89.

59. Buchlak QD, Esmaili N, Leveque J-C, Bennett C, Farrokhi F, Piccardi M. Machine learning applications to neuroimaging for glioma detection and classification: An artificial intelligence augmented systematic review. Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia. 2021;89:177-98.

60. Zulkarnain, Putri TD. Intelligent transportation systems (ITS): A systematic review using a Natural Language Processing (NLP) approach. Heliyon. 2021;7(12):e08615.

61. Olorisade BK, Brereton P, Andras P. Reproducibility of studies on text mining for citation screening in systematic reviews: Evaluation and checklist. Journal of biomedical informatics. 2017;73:1-13.

62. Dias AM, Mansur CG, Myczkowski M, Marcolin M. Whole field tendencies in transcranial magnetic stimulation: A systematic review with data and text mining. Asian journal of psychiatry. 2011;4(2):107-12.

63. Paynter RA, Featherstone R, Stoeger E, Fiordalisi C, Voisin C, Adam GP. A prospective comparison of evidence synthesis search strategies developed with and without text-mining tools. Journal of clinical epidemiology. 2021;139:350-60.

64. Li D, Wang Z, Wang L, Sohn S, Shen F, Murad MH, et al. A Text-Mining Framework for Supporting Systematic Reviews. American journal of information management. 2016;1(1):1-9.

65. Millard LAC, Flach PA, Higgins JPT. Machine learning to assist risk-of-bias assessments in systematic reviews. International journal of epidemiology. 2016;45(1):266-77.

66. Pham B, Jovanovic J, Bagheri E, Antony J, Ashoor H, Nguyen TT, et al. Text mining to support abstract screening for knowledge syntheses: a semi-automated workflow. Systematic reviews. 2021;10(1):156.

67. Baron JA, Senn S, Voelker M, Lanas A, Laurora I, Thielemann W, et al. Gastrointestinal adverse effects of short-term aspirin use: a meta-analysis of published randomized controlled trials. Drugs in R&D. 2013;13(1):9-16.

68. Davis AP, Wiegers TC, Johnson RJ, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, et al. Text mining effectively scores and ranks the literature for improving chemical-gene-disease curation at the comparative toxicogenomics database. PloS one. 2013;8(4):e58201.

69. Lin H-J, Sheu PCY, Tsai JJP, Wang CCN, Chou C-Y. Text mining in a literature review of urothelial cancer using topic model. BMC cancer. 2020;20(1):462.

70. Hashimoto K, Kontonatsios G, Miwa M, Ananiadou S. Topic detection using paragraph vectors to support active learning in systematic reviews. Journal of biomedical informatics. 2016;62:59-65.

71. Rost TB, Slaughter L, Nytro O, Muller AE, Vist GE. Using neural networks to support high-quality evidence mapping. BMC bioinformatics. 2021;22(Suppl 11):496.

72. Lange T, Schwarzer G, Datzmann T, Binder H. Machine learning for identifying relevant publications in updates of systematic reviews of diagnostic test studies. Research synthesis methods. 2021;12(4):506-15.

73. Yamada T, Yoneoka D, Hiraike Y, Hino K, Toyoshiba H, Shishido A, et al. Deep Neural Network for Reducing the Screening Workload in Systematic Reviews for Clinical Guidelines: Algorithm Validation Study. Journal of medical Internet research. 2020;22(12):e22422.

74. Wagner G, Lukyanenko R, Paré G. Artificial intelligence and the conduct of literature reviews. Journal of Information Technology.0(0):02683962211048201.

75. Thiabaud A, Triulzi I, Orel E, Tal K, Keiser O. Social, Behavioral, and Cultural factors of HIV in Malawi: Semi-Automated Systematic Review. Journal of medical Internet research. 2020;22(8):e18747.

76. van Lissa CJ. Mapping Phenomena Relevant to Adolescent Emotion Regulation: A Text-Mining Systematic Review. Adolescent research review. 2021:1-13.

77. Carrasco AJP, Volberg C, Pedrosa DJ, Berthold D. Patient Safety in Palliative and End-of-Life Care: A Text Mining Approach and Systematic Review of Definitions. The American journal of hospice & palliative care. 2021;38(8):1004-12.

78. Quang MN, Rogers T, Hofman J, Lanham AB. New framework for automated article selection applied to a literature review of Enhanced Biological Phosphorus Removal. PloS one. 2019;14(5):e0216126.

79. Guzik M, Witko T, Steinbuchel A, Wojnarowska M, Soltysik M, Wawak S. What Has Been Trending in the Research of Polyhydroxyalkanoates? A Systematic Review. Frontiers in bioengineering and biotechnology. 2020;8:959.

80. Blake C, Lucic A. Automatic endpoint detection to support the systematic review process. Journal of biomedical informatics. 2015;56:42-56.

81. Cohen AM, Ambert K, McDonagh M. Studying the potential impact of automated document classification on scheduling a systematic review update. BMC medical informatics and decision making. 2012;12:33.

82. Bashir R, Surian D, Dunn AG. The risk of conclusion change in systematic review updates can be estimated by learning from a database of published examples. Journal of clinical epidemiology. 2019;110:42-9.

83. John A, Eyles E, Webb RT, Okolie C, Schmidt L, Arensman E, et al. The impact of the COVID-19 pandemic on self-harm and suicidal behaviour: update of living systematic review. F1000Research. 2020;9:1097-.

84. John A, Okolie C, Eyles E, Webb RT, Schmidt L, McGuiness LA, et al. The impact of the COVID-19 pandemic on self-harm and suicidal behaviour: A living systematic review. F1000Research. 2020;9.

85. Norman CR, Leeflang MMG, Porcher R, Neveol A. Measuring the impact of screening automation on meta-analyses of diagnostic test accuracy. Systematic reviews. 2019;8(1):243.

86. Clark J, Glasziou P, Del Mar C, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. Journal of clinical epidemiology. 2020;121:81-90.

87. Scholten B, Simon LG, Krishnan S, Vermeulen R, Pronk A, Gyori BM, et al. Automated Network Assembly of Mechanistic Literature for Informed Evidence Identification to Support Cancer Risk Assessment. Environmental health perspectives. 2022;130(3):37002.

88. Muller AE, Ames HMR, Jardim PSJ, Rose CJ. Machine learning in systematic reviews: Comparing automated text clustering with Lingo3G and human researcher categorization in a rapid review. Research synthesis methods. 2021.

89. Stansfield C, Thomas J, Kavanagh J. 'Clustering' documents automatically to support scoping reviews of research: a case study. Research synthesis methods. 2013;4(3):230-41.

90. Shemilt I, Hollands GJ, Marteau TM, Nakamura R, Jebb SA, Kelly MP, et al. Economic instruments for population diet and physical activity behaviour change: a systematic scoping review. PloS one. 2013;8(9):e75070.

91. Shemilt I, Khan N, Park S, Thomas J. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews. Systematic reviews. 2016;5(1):140.

92. Helmick H, Nanda G, Ettestad S, Liceaga A, Kokini JL. Applying text mining to identify relevant literature in food science: Cold denaturation as a case study. Journal of food science. 2021;86(11):4851-64.

93. Simon C, Davidsen K, Hansen C, Seymour E, Barnkob MB, Olsen LR. BioReader: a text mining tool for performing classification of biomedical literature. BMC Bioinformatics. 2019;19(13):57.

94. Torres Torres M, Adams CE. RevManHAL: towards automatic text generation in systematic reviews. Systematic reviews. 2017;6(1):27.

95. Zingg W, Castro-Sanchez E, Secci FV, Edwards R, Drumright LN, Sevdalis N, et al. Innovative tools for quality assessment: integrated quality criteria for review of multiple study designs (ICROMS). Public health. 2016;133:19-37.

96. Adusumilli G, Pederson JM, Hardy N, Kallmes KM, Hutchison K, Kobeissi H, et al. Mechanical Thrombectomy With and Without Intravenous Tissue Plasminogen Activator for Acute Ischemic Stroke: A Systematic Review and Meta-Analysis Using Nested Knowledge. Frontiers in neurology. 2021;12:759759.

97. Alla K, Oprescu F, Hall WD, Whiteford HA, Head BW, Meurk CS. Can automated content analysis be used to assess and improve the use of evidence in mental health policy? A systematic review. Systematic reviews. 2018;7(1):194.

98. Gates A, Gates M, DaRosa D, Elliott SA, Pillay J, Rahman S, et al. Decoding semi-automated title-abstract screening: findings from a convenience sample of reviews. Systematic reviews. 2020;9(1):272.

99. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, et al. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Systematic reviews. 2019;8(1):278.

100. Gates A, Johnson C, Hartling L. Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool. Systematic reviews. 2018;7(1):45.

101. Gates A, Vandermeer B, Hartling L. Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool. Journal of clinical epidemiology. 2018;96:54-62.

102. Giummarra MJ, Lau G, Gabbe BJ. Evaluation of text mining to reduce screening workload for injury-focused systematic reviews. Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention. 2020;26(1):55-60.

103. Eun M-Y, Jeon E-T, Seo K-D, Lee D, Jung J-M. Reperfusion Therapy in Acute Ischemic Stroke with Active Cancer: A Meta-Analysis Aided by Machine Learning. Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association. 2021;30(6):105742.

104. Guo Y, Silins I, Stenius U, Korhonen A. Active learning-based information structure analysis of full scientific articles and two applications for biomedical literature review. Bioinformatics (Oxford, England). 2013;29(11):1440-7.

105. Shekelle PG, Shetty K, Newberry S, Maglione M, Motala A. Machine Learning Versus Standard Techniques for Updating Searches for Systematic Reviews: A Diagnostic Accuracy Study. Annals of internal medicine. 2017;167(3):213-5.

106. Begert D GJ, Irwin B, Brogly C. Towards automating systematic reviews on immunization using an advanced natural language processing–based extraction system. Can Commun Dis Rep. 2020;36(6):174-9.

107. Marshall IJ, Kuiper J, Banner E, Wallace BC. Automating Biomedical Evidence Synthesis: RobotReviewer. Proceedings of the conference Association for Computational Linguistics Meeting. 2017;2017:7-12.

108. Matwin S, Kouznetsov A, Inkpen D, Frunza O, O'Blenis P. A new algorithm for reducing the workload of experts in performing systematic reviews. Journal of the American Medical Informatics Association : JAMIA. 2010;17(4):446-53.

109. Miwa M, Thomas J, O'Mara-Eves A, Ananiadou S. Reducing systematic review workload through certainty-based screening. Journal of biomedical informatics. 2014;51:242-53.

110. Orgeolet L, Foulquier N, Misery L, Redou P, Pers J-O, Devauchelle-Pensec V, et al. Can artificial intelligence replace manual search for systematic literature? Review on cutaneous manifestations in primary Sjogren's syndrome. Rheumatology (Oxford, England). 2020;59(4):811-9.

111. Shao W, Adams CE, Cohen AM, Davis JM, McDonagh MS, Thakurta S, et al. Aggregator: a machine learning approach to identifying MEDLINE articles that derive from the same underlying clinical trial. Methods (San Diego, Calif). 2015;74:65-70.

112. Chapman AL, Morgan LC, Gartlehner G. Semi-automating the manual literature search for systematic reviews increases efficiency. Health information and libraries journal. 2010;27(1):22-7.

113. Buchlak QD, Esmaili N, Leveque JC, Bennett C, Farrokhi F, Piccardi M. Machine learning applications to neuroimaging for glioma detection and classification: An artificial intelligence augmented systematic review. J Clin Neurosci. 2021;89:177-98.

114. Choong MK, Galgani F, Dunn AG, Tsafnat G. Automatic evidence retrieval for systematic reviews. Journal of medical Internet research. 2014;16(10):e223.

115. Cohen AM, Ambert K, McDonagh M. Cross-topic learning for work prioritization in systematic review creation and update. Journal of the American Medical Informatics Association : JAMIA. 2009;16(5):690-704.

116. Zimmerman J, Soler RE, Lavinder J, Murphy S, Atkins C, Hulbert L, et al. Iterative guided machine learning-assisted systematic literature reviews: a diabetes case study. Systematic reviews. 2021;10(1):97.

117. Aum S, Choe S. srBERT: automatic article classification model for systematic review using BERT. Systematic reviews. 2021;10(1):285.

118. Bahor Z, Liao J, Macleod MR, Bannach-Brown A, McCann SK, Wever KE, et al. Risk of bias reporting in the recent animal focal cerebral ischaemia literature. Clinical science (London, England : 1979). 2017;131(20):2525-32.

119. Burns JK, Etherington C, Cheng-Boivin O, Boet S. Using an artificial intelligence tool can be as accurate as human assessors in level one screening for a systematic review. Health information and libraries journal. 2021.

120. Gates A, Gates M, Sebastianski M, Guitard S, Elliott SA, Hartling L. The semi-automation of title and abstract screening: a retrospective exploration of ways to leverage Abstrackr's relevance predictions in systematic and rapid reviews. BMC medical research methodology. 2020;20(1):139.

121. Gartlehner G, Wagner G, Lux L, Affengruber L, Dobrescu A, Kaminski-Hartenthaler A, et al. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: a user study. Systematic reviews. 2019;8(1):277.

122. Hamel C, Hersi M, Kelly SE, Tricco AC, Straus S, Wells G, et al. Guidance for using artificial intelligence for title and abstract screening while conducting knowledge syntheses. BMC medical research methodology. 2021;21(1):285.

123. Przybyla P, Brockmeier AJ, Kontonatsios G, Le Pogam M-A, McNaught J, von Elm E, et al. Prioritising references for systematic reviews with RobotAnalyst: A user study. Research synthesis methods. 2018;9(3):470-88.

124. Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. Systematic reviews. 2020;9(1):73.

125. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nature Machine Intelligence. 2021;3(2):125-33.

126. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environmental Evidence. 2018;7(1):8.

127. Cohen AM, Hersh WR, Peterson K, Yen P-Y. Reducing workload in systematic review preparation using automated citation classification. Journal of the American Medical Informatics Association : JAMIA. 2006;13(2):206-19.

128. Cleo G, Scott AM, Islam F, Julien B, Beller E. Usability and acceptability of four systematic review automation software packages: a mixed method design. Systematic reviews. 2019;8(1):145.

129. Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, et al. SWIFT-Review: a text-mining workbench for systematic review. Systematic reviews. 2016;5:87.

130. Halamoda-Kenzaoui B, Rolland E, Piovesan J, Puertas Gallardo A, Bremer-Hoffmann S. Toxic effects of nanomaterials for health applications: How automation can support a systematic review of the literature? Journal of applied toxicology : JAT. 2022;42(1):41-51.

131. Hamel C, Kelly SE, Thavorn K, Rice DB, Wells GA, Hutton B. An evaluation of DistillerSR's machine learning-based prioritization tool for title/abstract screening - impact on reviewer-relevant outcomes. BMC medical research methodology. 2020;20(1):256.

132. Choong MK, Tsafnat G. Role of citation tracking in updating of systematic reviews. AMIA Joint Summits on Translational Science proceedings AMIA Joint Summits on Translational Science. 2014;2014:18.

133. van Dinter R, Tekinerdogan B, Catal C. Automation of systematic literature reviews: A systematic literature review. Information and Software Technology. 2021;136:106589.

134. Frunza O, Inkpen D, Matwin S, Klement W, O'Blenis P. Exploiting the systematic review protocol for classification of medical abstracts. Artificial intelligence in medicine. 2011;51(1):17-25.

135. Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. PloS one. 2019;14(12):e0225980.

136. Ng V, Rees EE, Niu J, Zaghool A, Ghiasbeglou H, Verster A. Application of natural language processing algorithms for extracting information from news articles in event-based surveillance. Can Commun Dis Rep. 2020;46(6):186-91.

137. Aucoin M, LaChance L, Naidoo U, Remy D, Shekdar T, Sayar N, et al. Diet and Anxiety: A Scoping Review. Nutrients. 2021;13(12).

138. Wallace BC, Small K, Brodley CE, Lau J, Schmid CH, Bertram L, et al. Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining. Genetics in medicine : official journal of the American College of Medical Genetics. 2012;14(7):663-9.

139. Bashir R, Dunn AG, Surian D. A rule-based approach for automatically extracting data from systematic reviews and their updates to model the risk of conclusion change. Research synthesis methods. 2021;12(2):216-25.

140. Yu L, Yu S. Developing an automated mechanism to identify medical articles from wikipedia for knowledge extraction. International journal of medical informatics. 2020;141:104234.

141. Badgett RG, Dylla DP, Megison SD, Harmon EG. An experimental search strategy retrieves more precise results than PubMed and Google for questions about medical interventions. PeerJ. 2015;3:e913.

142. Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. BMC bioinformatics. 2010;11:55.

143. Casey A, Davidson E, Poon M, Dong H, Duma D, Grivas A, et al. A systematic review of natural language processing applied to radiology reports. BMC medical informatics and decision making. 2021;21(1):179.

144. Cohen AM, Ambert K, McDonagh M. A Prospective Evaluation of an Automated Classification System to Support Evidence-based Medicine and Systematic Review. AMIA Annual Symposium proceedings AMIA Symposium. 2010;2010:121-5.

145. Bui DDA, Del Fiol G, Hurdle JF, Jonnalagadda S. Extractive text summarization system to aid data extraction from full text in systematic review development. Journal of biomedical informatics. 2016;64:265-72.

146. Zerva C, Batista-Navarro R, Day P, Ananiadou S. Using uncertainty to link and rank evidence from biomedical literature for model curation. Bioinformatics (Oxford, England). 2017;33(23):3784-92.

147. Clark J, McFarlane C, Cleo G, Ishikawa Ramos C, Marshall S. The Impact of Systematic Review Automation Tools on Methodological Quality and Time Taken to Complete Systematic Review Tasks: Case Study. JMIR medical education. 2021;7(2):e24418.

148. Chai KEK, Lines RLJ, Gucciardi DF, Ng L. Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews. Systematic reviews. 2021;10(1):93.

149. Gates M, Elliott SA, Gates A, Sebastianski M, Pillay J, Bialy L, et al. LOCATE: a prospective evaluation of the value of Leveraging Ongoing Citation Acquisition Techniques for living Evidence syntheses. Systematic reviews. 2021;10(1):116.

150. Afzal M, Hussain M, Ali T, Hussain J, Khan WA, Lee S, et al. Knowledge-Based Query Construction Using the CDSS Knowledge Base for Efficient Evidence Retrieval. Sensors (Basel, Switzerland). 2015;15(9):21294-314.

151. Aphinyanaphongs Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis CF. Text categorization models for high-quality article retrieval in internal medicine. Journal of the American Medical Informatics Association : JAMIA. 2005;12(2):207-16.

152. Buchlak QD, Clair J, Esmaili N, Barmare A, Chandrasekaran S. Clinical outcomes associated with robotic and computer-navigated total knee arthroplasty: a machine learning-augmented systematic review. European journal of orthopaedic surgery & traumatology : orthopedie traumatologie. 2021.

Annex 1 Participants in focus groups, interviews and survey

Table 5. Participants in focus groups, interviews and survey

Name	Organisation	Round
Emma Cattermole	ECDC	Focus group 1 23 May 2022
Giorgos Kaltekis	ECDC	
Brian Kristensen	Statens Serum Institut (SSI), DK	
Georgios Panagiotakopoulos[vi]	National Public Health Organisation (EODY), GR ECDC Advisory Forum alternate and National Focal Point for scientific advice coordination	
Johanna Takkinen	ECDC	
Britta Tendal Jeppesen	Danish Health Authority (SST), DK	
Raymon van Dinter	Wageningen University, NL	
Andrew Booth	University of Sheffield, UK	Focus group 2 26 May 2022
Lucía García San Miguel	Coordinating Centre of Sanitary Alerts and Emergencies (CCAES), ES ECDC National Focal Point for scientific advice coordination	
Chantelle Garritty	Public Health Agency of Canada	
Petra Klepac	National Institute of Public Health (NIJZ), SI	
Natalija Kranjec	National Institute of Public Health (NIJZ), SI	
Nadja Sinkovec Zorko	National Institute of Public Health (NIJZ), SI	
Dimitrios Paraskevis	National Public Health Organisation (EODY), GR	
Barbara Albiger	ECDC	Focus group 3 31 May 2022
Stanislav Danchev	ECDC	
Celine Dumas	Santé Publique France	
Hanan Khalil	La Trobe University, AUS	
Elaine Claire Lautier	Ministry of Health, MT	
Tuija Leino	Finnish Institute for Health and Welfare (THL), FI	
Kate Olsson	ECDC	
Ajibola Omokanye	ECDC	
Diamantis Plachouras	ECDC	
Jeanine Pommier	ECDC	
Uli Ronellenfitsch	Martin-Luther University, DE	
Michaela Špačková	National Institute for Public Health, CZ	
Caroline van den Ende	National Institute for Public Health and the Environment (RIVM), NL	
Sophie Vaux	Santé Publique France	
Albert Vollaard	National Institute for Public Health and the Environment (RIVM), NL	
Kim Brolin	ECDC	Focus group 4 08 June 2022
Julian Elliott	Cochrane, AUS	
Rok Grah	ECDC	
Candyce Hamel	Ottawa Hospital Research Institute, CA	
Trish Harrington	Health Information and Quality Authority (HIQA), IR	
Otto Helve	Finnish Institute for Health and Welfare (THL), FI ECDC National Focal Point for scientific advice coordination	
Jari Jalava	Finnish Institute for Health and Welfare (THL), FI	
John Kinsman	ECDC	
Spyros Ktenas	ECDC	

[vi] Also completed a survey response

Name	Organisation	Round
Tanya Melillo[vii]	Ministry of Health, MT ECDC Advisory Forum alternate and National Focal Point for scientific advice coordination	
Alison O'Mara-Eves	EPPI-Centre, UCL, UK	
Ana João Santos	National Institute of Health Ricardo Jorge (INSA), PT	
Carl Fredrik Sjöland	Swedish Public Health Institute	
Petra Apfalter	Institute for Hygiene, Microbiology and Tropical Medicine at Ordensklinikum Linz and BioLab GmbH, AT ECDC Advisory Forum member	Focus group 5 14 June 2022
Oliver Briet	ECDC	
María-José Faraldo-Vallés[viii]	Galician Health Service (SERGAS), ES	
María Soledad Isern de Val	Aragon Health Sciences Institute (IACS), ES	
Mirjam Knol	National Institute for Public Health and the Environment (RIVM), NL	
Ashley Muller	Norwegian Institute of Public Health (FHI)	
Sari Palojoki	ECDC	
James Thomas	EPPI-Centre, UCL, UK	
Malcolm Macleod	CAMARADES group, The University of Edinburgh, UK	Interview
Zachary Munn	University of Adelaide, AUS	
Anne-Mae Scott	Institute for Evidence-Based Healthcare, Bond University, AUS	
Ana Maria Correia	National Institute of Health Ricardo Jorge (INSA Porto), PT ECDC Advisory Forum alternate	Survey[ix]
Agnes Danielisz	National Public Health Centre, HU ECDC National Focal Point for scientific advice coordination	
Isabel de la Fuente	Centre Hospitalier de Luxembourg ECDC Advisory Forum member	
John Middleton	Wolverhampton University, UK ECDC Advisory Forum observer representing the Association of Schools of Public Health in the European Region (ASPHER)	
Didrik Vestrheim	Norwegian Institute of Public Health (FHI)	

[vii] Also provided a survey response

[viii] Also provided a survey response

[ix] One survey respondent chose not to provide their name and affiliation.

PDF TQ-03-22-169-EN-N

European Centre for Disease Prevention and Control (ECDC)

Gustav III:s Boulevard 40, 16973 Solna, Sweden

Tel. +46 858601000
Fax +46 858601001
www.ecdc.europa.eu

An agency of the European Union
www.europa.eu

Subscribe to our publications
www.ecdc.europa.eu/en/publications

Contact us
publications@ecdc.europa.eu

Follow us on Twitter
@ECDC_EU

Like our Facebook page
www.facebook.com/ECDC.EU

ECDC is committed to ensuring the transparency and independence of its work

In accordance with the *Staff Regulations for Officials and Conditions of Employment of Other Servants of the European Union* and the *ECDC Independence Policy*, ECDC staff members shall not, in the performance of their duties, deal with matters in which they may, directly or indirectly, have a personal interest that could impair their independence. Declarations of interest must be received from any prospective contractor before a contract can be awarded.
www.ecdc.europa.eu/en/aboutus/transparency

Publications Office of the European Union

PDF ISBN 978-92-9498-600-9