

Applying automation to maintain research registers in health promotion

Claire Stansfield and James Thomas

Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre, UCL Social Research Institute, University College London (UCL), London, UK

Abstract

Maintaining research registers and similar content can be resource intensive. Automation tools can help in identifying, assessing and describing content. This article describes some challenges and opportunities using automation to support coverage, keywording and improve sustainability for two research registers of health promotion effectiveness.

Key words: *health promotion; databases as topic; automation.*

Introduction

Health promotion research is challenging to identify owing to its breadth of topics and terminology. For over 20 years, the EPPI Centre has maintained two publicly available research registers that focus on research of the effectiveness of interventions in health promotion. Their longevity and focus make these unique resources that are useful for identifying such research, whether from brief enquiries or as a resource for conducting systematic reviews. They have largely been maintained through manual processes, which are challenging to maintain within the resources available. Although processes have been streamlined over the years and augmented using automation tools, a major change is using automation tools undertake the bulk of the processes, which will improve currency and sustainability. We describe here some of the challenges and opportunities from undertaking this shift across the processes of searching, screening, and keywording, and on forthcoming changes to the user-interface.

Context

The Trials Register of Promoting Health Interventions (TRoPHI) focuses on controlled trials and Database of Promoting Health Effectiveness Reviews (DoPHER) focuses on reviews (1). They contain over 20,000 and 9,000 records of research, respectively. They were developed following a methodological study of effectiveness reviews in health promotion (2) and undertaken

as part of the former Systematic Reviews Facility in Health Promotion and Public Health at the EPPI Centre (UCL). TRoPHI was the research register of the former Cochrane Health Promotion and Public Health field (1996-2008). Initially, they were compiled from searching and coding research for systematic reviews of all study designs of health promotion research within the Bibliomap database (now archived) (1), and included content from the Field's initiatives to identify trials not indexed in electronic databases, and conference abstracts (3). The registers are now maintained as part of the EPPI Centre's Policy Reviews Facility, which uses research from across health, public health social care, much of which goes beyond effectiveness research within health promotion (4), and which is funded by the UK's National Institute for Health and Care Research (NIHR). The Facility's current focus is much broader than the coverage of the two registers. The registers are also useful for investigating aspects of using automation tools to support reviews of research, a core activity within the EPPI Centre, and uses tools available within EPPI-Reviewer, a systematic review management tool developed in-house (5).

What is within the scope of health promotion research registers?

Unfortunately, there is no clear boundary of what falls within scope of health promotion intervention research, though it includes the promotion of changes of

Address for correspondence: Claire Stansfield, Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre, UCL Social Research Institute, University College London, 55-59 Gordon Square, London WC1H 0NU, UK. E-mail: c.stansfield@ucl.ac.uk

behaviours to improve health through education, communication or structural means, rather than from drug or surgical treatments. It also includes interventions to improve caregiver health, improve health-protecting factors or reduce health risk factors, and includes public policies to improve equity of health service delivery. Topics include drug use, obesity, mental health, sexually transmitted infections, uptake of medical care, such as vaccinations and screening, hygiene and accidents, among others. It does not include rehabilitation, management or treatment of people's existing health conditions. It also concerns effectiveness rather than efficacy, such as exercise programmes to prevent hypertension, rather than exercise for preventing hypertension. Whether such research meets these requirements is assessed by the topic, intervention and outcomes measured that are described within titles and abstracts of research records.

Identification – increasing content from a graph-based recommender

Since 2004, a core part of maintaining currency and breadth of content has been from routine searching using keywords for health promotion, rather than keywords targeted on topic areas of interest, with additional website scanning to find reviews. For example, the PubMed search uses text words for the phrases “public health” “health promotion”, “health education” “primary prevention” and MeSH terms for services in community health, child health, sanitation, preventive health, mass screening (for example), along with terms for randomised and non-randomised controlled trials. However, recently studies are also identified by a network graph “search” using OpenAlex within EPPI-Reviewer (5) (formerly using the discontinued Microsoft Academic Graph). The “search” is a recommender system which finds records that are similar to records that are within the register, Similarity is based on text in the titles and abstracts, citation connections, authorship, topics, and the set of records as whole. It was introduced based on previous work to produce a register on COVID-19 research (6-8). This method identifies more relevant records than the earlier approaches, partly owing to the large content coverage of OpenAlex, and is not solely dependent on terms for health promotion in the titles, abstracts and indexing. However, the great volume of content increases the workload of screening, whereby each record is checked for eligibility.

Screening – using machine classifiers

For many years, a machine learning classifier has been used in TRoPHI to rank records by relevance and automatically screen out the least relevant. The classifier was developed from using relevance decisions of humanly-screened records and checked against sets of gold standard data as described in Stansfield et al. (9). The classifier has subsequently been updated with more training data and applied to achieve 95% recall. A different classifier is used in DoPHER. However, given the increase in volume of records from research publishing generally, as well from using OpenAlex, we are investigating a threshold of automatic inclusion and exclusion without any human screening. While this approach inevitably introduces research records into the register that would ideally be screened out by a human screener, it helps make the maintenance of the registers more sustainable and increases coverage. Although such a threshold may compromise recall and is a significant shift from previous procedures, which aimed for high recall of all relevant records from the searches, it is appealing in favour of providing currency, breadth and sustainability.

Keywording – using a large language model

Historically, both registers contained manually-applied keywords based on titles and abstracts for TRoPHI, and the full-text for DoPHER. Around 2013, this was replaced by no keywording for DoPHER and streamlined keywords for TRoPHI to study design, topic focus, population focus and country setting. User data over three months during 2013 and 2023 both showed that users favoured free-text searching. However, with the introduction of greater content from OpenAlex, the application of keywording could be increasingly useful for navigate these resources. Furthermore, the use of keywording supports greater visualisation of the database (described further down).

One solution is EPPI-Reviewer's beta-tool for automating data extraction using the large language model GPT-4 to apply keywords from text in the title and abstract (10). It is proving very promising and tests are ongoing to reduce some inaccuracies before finalising its use. Algorithmic keywording or indexing is challenging to achieve full accuracy and completeness, as highlighted recently by Amar-Zifkin et al. (11) in their commentary on its use in MEDLINE. However, our tests

with TRoPHI showed that the tool identifies relevant keywords that a human keyworder missed. Furthermore, the tool has promise for identifying certain traits present in some irrelevant records (such as study designs without a comparison group), and so could also be applied to support quality assurance of content and help remove these records. The consistency of GPT-4 in applying codes is also an aspect to investigate further. The model of humanly-applying codes requires interpretation of the abstract and applying to the most suitable keywords from a keyword tool. To use GPT-4, each keyword within the tool is translated into a yes or no question or “prompt”, so that if the answer is true the keyword is assigned. For example, to keyword a record as being about mental health, the prompt asks if the focus of the intervention or outcome is about mental health. However, further specification has been needed than was present in the human guidance to include caregiver burden, self-efficacy, isolation and others into the same prompt for “mental health”. This potentially introduces systematic bias to the keywords, whereas previously the bias would have been human interpretation on a case-by-case basis. Writing the prompts has required a tweaking of the health promotion keywording tool, largely unchanged since 1997. For example, a new keyword for diaspora and displaced populations is a useful addition.

Improved visualisation

The registers are being transferred from their old interface onto EPPI-Vis, which provides greater functionality and visualisation of content through graphical display of publication year and an interactive evidence maps. TRoPHI has functions for frequency and cross-tabulations based on available keywords. We expect the opportunities provided by automated keywording will enable this to be current for both registers, and will augment the basic search functions. A challenge is the expectations of users in utilising the keywords, and that they provide a greater functionality supporting exploration of registers rather than serve as definitive labels.

Conclusion: overall reflection and applicability

Maintaining research registers and similar content is resource intensive and there are many examples of specialist resources ceasing, most recently Social Care On-

line (12). Using automation tools is one way to improve sustainability though bring to the fore the trade-off decisions of coverage, and accuracy of keywords. In the case of TRoPHI and DoPHER, we consider these relatively low-risk for their purpose. The automation tools applied here generally perform better on records with abstracts than titles alone, and so there is a risk that title-only content, including grey literature without abstracts will become less findable. Automation can support identification and classification of research though there appears to be a lack of standards and transparency in what is acceptable in terms of system performance. Furthermore, there is a danger that automation influences how we think in a way that perpetuates hidden biases with unforeseen consequences that might be different to humanly-curated systems. As always, careful communication to the users is needed to support the use of resources. This work is part of a number of initiatives supporting collation of registers. Other examples include the FAIR database (13), and living maps of research, such as the COVID-19 living map (7) which draw on other types of automation tools and processes.

*Submitted on invitation.
Accepted on 7 June 2024.*

REFERENCES

1. EPPI Centre. Databases. Available from: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=185>
2. Peersman G, Harden A, Oliver S, Oakley A. Effectiveness reviews in health promotion. London: EPPI Centre, Social Science Research Unit, Institute of Education, University of London. 1999. Available from: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=253>
3. Howes F, Doyle J, Jackson N, Waters E. Evidence-based public health: The importance of finding ‘difficult to locate’ public health and health promotion intervention studies for systematic reviews. *J Public Health*. 2004;26(1):101-4.

4. EPPI Centre. The NIHR Policy Research Programme Reviews Facility. Available from: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=73>
5. Thomas J, Graziosi S, Brunton J, Ghouze Z, O'Driscoll P, Bond M. EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis. EPPI-Centre, UCL Social Research Institute, University College London. 2020. Available from: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2914>
6. Shemilt I, Arno A, Thomas J, Lorenc T, Khouja C, Raine G, Sutcliffe K, D'Souza P, Wright K, Sowden A. Using automation to produce a "living map" of the COVID-19 research literature. JEAHIL [Internet]. 2021;17(2):11-5. Available from: <https://ojs.eahil.eu/JEAHIL/article/view/469/414>
7. Shemilt I, Arno A, Thomas J, Lorenc T, Khouja C, Raine G, Sutcliffe K, D'Souza P, Wright K, Sowden A. Cost-effectiveness of Microsoft Academic Graph with machine learning for automated study identification in a living map of coronavirus disease 2019 (COVID-19) research [version 2; peer review: 2 approved]. Wellcome Open Research, 2024; 6:210 <https://doi.org/10.12688/wellcomeopenres.17141.2>
8. Thomas J. New methods and technologies for keeping systematic reviews up to date. Center on Knowledge Translation for Disability and Rehabilitation Research, American Institutes for Research. 2022. Available from: <https://ktddr.org/>
9. Stansfield C, Stokes G, Thomas J. Applying machine classifiers to update searches: Analysis from two case studies. Res Synth Methods. 2022Jan;13(1):121-33. doi: 10.1002/jrsm.1537
10. EPPI Centre Automated data extraction using GPT-4. Available from: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3921>
11. Amar-Zifkin A, Ekmekjian T, Paquet V, Landry T. Brief communication concerning algorithmic indexing in MEDLINE. JEAHIL [Internet]. 17Mar.2024;20(1):18-1. Available from: <https://ojs.eahil.eu/JEAHIL/article/view/604/527>
12. Social Care Online <https://www.scie.org.uk/social-care-online/> [accessed 29 April 2024]
13. EPPI Centre. Automation for discovery of inequalities research. Available from: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3830>