SWAR 26: Sensitivity and completeness of search strategies built using a text-mining word frequency tool (PubReMiner) compared to current best practice for building a search strategy

Objective of this SWAR
To determine the sensitivity and completeness of search strategies built using a text-mining word frequency tool (PubReMiner).

Study area: Statistical Analysis, Search Strategy Building
Sample type: Librarians
Estimated funding level needed: Low

Background
Systematic and rapid reviews inform evidence-based decision-making in health care and health policy. In both types of review, information must be synthesised in a systematic manner, and an accurate and comprehensive search strategy is critical to this. Several text mining and artificial intelligence (AI) modules are available, offering new ways to search and screen for relevant literature. These novel methods may be attractive for researchers, particularly healthcare professionals who may not have comprehensive skills in search strategy development or may not have access to expertise from an information specialist or librarian.[1] However, such tools would only be of use if acceptable sensitivity had been demonstrated and, currently, there is limited evidence on their implementation into routine research practices. Furthermore, researchers may be hesitant to deviate from 'conventional' search strategy building practices or best practice.

PubReMiner is a text-mining tool which can be used to identify frequently used words from key publications and to help build search strategies efficiently. It uses MEDLINE to determine the frequency of free-text and Medical Subject Headings (MeSH) terms in key publications by inputting their PubMed Identifiers (PMIDs). This differs from conventional search strategy building techniques, where the researcher is required to identify the MeSH terms and relevant synonyms themselves.[2]

This SWAR closely aligns with the eighth research priority identified in the Priority III study for rapid reviews: it is asking how to optimise the search process for rapid reviews.[3] It embeds a methodological research question across multiple reviews and it may provide a methodological framework for others to adopt if they wish to answer important methodological questions that cannot be answered within a single review alone.

Interventions and comparators
Intervention 1: Search strategies built using a text-mining word frequency tool (PubReMiner).
Intervention 2: Comparator: Conventional search strategies developed by a librarian.

Index Type:

Method for allocating to intervention or comparator
In addition to the conventional search strategy, which will be developed by one librarian, a second search strategy will be developed by a second independent librarian with equivalent experience and expertise using the PubReMiner tool. Both search strategies will be developed in accordance with a work instruction for building these search strategies, and the conventional search will be independently reviewed by a third librarian. Records retrieved by the two search strategies will be exported into two separate Endnote folders.

Outcome measures
Primary: Sensitivity: ratio of records that would have been included by the PubReMiner search strategy relative to the actual number included in the review where the conventional search was used.
Precision: number of relevant references identified by the database search relevant to the total number of relevant references found by a given search method.
Secondary: Number needed to read (NNR): number of references a researcher must screen/read to identify a relevant reference in each search strategy.

Number of unique references: number of included references retrieved from a database that were not retrieved from any other database.

Efficiency: time taken by each librarian to construct the search strategy.

Analysis plans

This SWAR will be hosted within reviews conducted in the Health Technology Assessment Directorate at the Health Information and Quality Authority (HIQA) in Ireland.[4] It is anticipated that 8-12 reviews will be conducted over 10 months. Meta-analytical techniques might be used to pool the estimates across the reviews and to generate an estimate of the sensitivity with accompanying confidence intervals for the PubReMiner search strategy. If meta-analysis is not appropriate, a narrative synthesis might be undertaken instead.

Possible problems in implementing this SWAR

Randomisation is not possible within the scope of this project; but performer biases will be mitigated by adherence to work instructions for the building of the conventional and PubReMiner search strategies. A priori assumptions used to inform sample size calculations may be inaccurate. It is currently estimated that 3430 records will be required, assuming a normal distribution, sensitivity of 95% and that only 10% of records are relevant. Following piloting of the SWAR and a review of the prevalence of relevant records, the power may be revised. The use of multiple reviews allows us to investigate this research question, which might not be answerable by a single review alone.

References

1. Hausner E, Guddat C, Hermanns T, Lampert U, Waffenschmidt S. Development of search strategies for systematic reviews: validation showed the noninferiority of the objective approach. Journal of Clinical Epidemiology 2015;68(2):191-9.
2. Slater L. PubMed PubReMiner. Journal of the Canadian Health Libraries Association / Journal de l'Association des bibliothèques de la santé du Canada 2014;33(2):106-7.
3. Beecher C, Toomey E, Maeso B, Whiting C, Stewart DC, Worrall A, et al. Priority III: top 10 rapid review methodology research priorities identified using a James Lind Alliance Priority Setting Partnership. Journal of Clinical Epidemiology 2022;151:151-60.
4. Health Information & Quality Authority (HIQA). Methods for Generic Justification of New Practices in Ionising Radiation. 2022.

Publications or presentations of this SWAR design

Examples of the implementation of this SWAR

People to show as the source of this idea: Andrew Dullea, Marie Carrigan, Susan M Smith, Lydia O'Sullivan, Martina Giusti, Isabelle Delaunois, Helen Clark, Kieran Walsh, Patricia Harrington, Máirín Ryan

Contact email address: dulleaa@tcd.ie

Date of idea: 1/MAR/2023

Revisions made by: N/A

Date of revisions: