Development and validation of a database filter for study size – preliminary results

Sabrina T.G. Gunput and Wichor M. Bramer
Medical Library, Erasmus MC, University Medical Center, Rotterdam, The Netherlands

Abstract
Researchers performing systematic reviews often express the desire to limit the search results to a certain study size. The aim of our study was to develop a filter in embase.com and Ovid to retrieve references above a specified threshold of study size. The filter that was developed consists mainly of truncated numbers in proximity with words such as patients, cases, adults, females and phrases like "n=". Preliminary results showed that the sensitivity of the filter, as evaluated on existing systematic reviews, was at least 94%. The burden of screening for systematic reviews can be greatly reduced with the study size filter.

Key words: *sample size; bibliographic databases; information storage and retrieval.*

Introduction

Researchers performing systematic reviews (SRs) often express the desire to limit the search results to their inclusion criteria, such as articles that included a certain study size: "I want to include only studies of more than 50 patients". While we of course can discuss about the validity of such a request, limiting the search results to match the inclusion criteria can reduce the burden of screening. Usually, the sample size of a study is mentioned in the abstract or title of an article. Instead of screening for the numbers in title or abstract manually, we aimed to develop a filter in embase.com and Ovid to retrieve references using a certain threshold of study size.

Methods

We developed the filter for embase.com and Ovid MEDLINE as these platforms have the ability to use proximity operators. Together with researchers who expressed the desire to limit search results to a certain number of patients, we constructed preliminary filters. These were tested iteratively by evaluating the patient numbers of relevant references that had not been retrieved. If the patient numbers matched the inclusion criteria, the filter was adapted to retrieve the missed articles and used for a new test round, until all relevant references were retrieved. After several iterative rounds of improvement, the filter was validated against existing SRs that used study sizes as inclusion criteria but did not limit their search to a study size.

Preliminary results

The filter that was developed consists mainly of truncated numbers in proximity with words such as patients, cases, adults, females, and phrase like "n=". The filter can and should be adapted to the research topic by combining these truncated numbers with specific terms for diseases, interventions or body parts of interest such as melanomas, surgeries, eyes or knees. The sensitivity of the filter as evaluated on existing SRs was at least 94% and the number of total references found with the filter was reduced to 40-75% (*Figure 1*).

Discussion

Preliminary results of the filter have been demonstrated so far. Our study should be expanded by validating the filter with more systematic reviews, although the first results seem promising. We already encountered some challenges and limitations:

- the filter searches for numbers in title or abstract. Large numbers that are often in abstracts, and could be in proximity of words such as "patients", are years or dates. This means the filter can obtain a little bit of noise retrieved by years, such as the year 1998 or 2021, mentioned in text. Also, phrases such as "80% of patients" are found with the filter;

Address for correspondence: Sabrina T.G. Gunput, Medical Library, Erasmus MC, University Medical Center, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. E-mail: s.gunput@erasmusmc.nl.

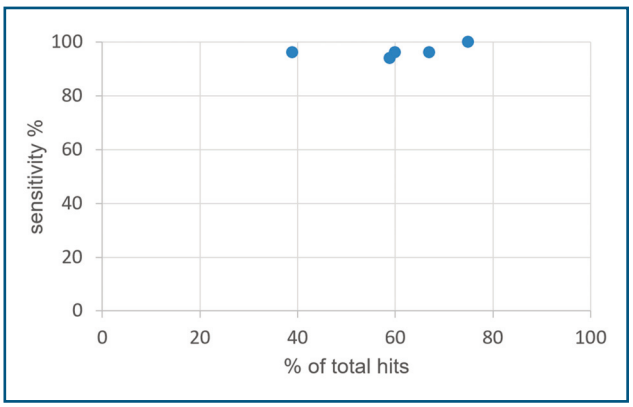

Fig. 1. *The sensitivity and percentage of hits found with the filter was evaluated on existing systematic reviews (SRs). Each dot represents the results of a test with an SR. Sensitivity was 94% or higher. The total number of hits retrieved with the filter compared to without the filter ranged from 40 to 75%.*

- the filter works with a threshold and will find all numbers higher/larger than specified number. Unfortunately, the maximum threshold of the filter is 100 in the current version;
- the filter works with proximity operators and therefore is only available in Ovid and embase.com, but not for PubMed. However, for PubMed it could be considered to use the Study size Search Tool of Baladrón *et al.* (1), although this tool is currently unavailable and does not have the option to adapt the filter to terms specific for the research question;
- the filter is not 100% sensitive, which is rarely the case with filters. Some references that were not retrieved were older articles that did not include the study size in their abstract. Since the late 1980's and early 1990's, more guidelines for structured abstracts have been available that also recommend mentioning of study sample size (2). Searching for older studies could be done without the filter and screening full texts of articles manually, whereas the newer studies have a high chance to be found with the filter.

When a threshold of study size is used as an inclusion criterion for a review, the screening time can be greatly reduced by using the filter for study size. We will test the sensitivity of the filter on several more SRs and adapt the filter where necessary. We will further investigate the cut-off point where the filter generates (near) 100% sensitivity to advise researchers on which publication dates should be searched without the study size filter.

Acknowledgements

The authors of this paper received the award for best oral presentation by a first-timer at the EAHIL 2022 Conference "Broaden the horizons - diversity, partnership and innovation with a human touch" Rotterdam, The Netherlands, 1-3 June 2022.

Submitted on invitation.
Accepted on 10 August 2022.

REFERENCES

1. Baladrón C, Santos-Lozano A, Aguiar JM, Lucia A, Martín-Hernández J. Tool for filtering PubMed search results by study size. J Am Med Inform Assoc. 2018;25(7):774-9.
2. Vrijhoef HJM, Steuten LMG. How to write an abstract. Eur Diabetes Nursing. 2007;4(3):124-7.

This paper is published under a CC BY license