A preliminary assessment of sample size as a search strategy filter on Embase in targeted literature searches

Mangat G[1]; Saipriya, J[2]; Sharma, S[3]

[1]Parexel International, Mohali, India;
[2]Parexel International, Bengaluru, India;
[3]Parexel International, Chandigarh, India

Background

Studies with small sample sizes are generally unreliable and should be interpreted considering the disease context and available evidence. While these might be of interest for rare diseases and studies with limited evidence, it is an accepted practice to remove such studies during synthesis in reviews with substantial evidence. Therefore, we analyzed the use of a sample size filter to allow removing these studies during the search strategy phase.

Methods

The sample size is generally reported in the published abstract. A filter consisting of truncated numbers in proximity with keywords like patients, controls, adults, pediatrics, males, females, men, etc., and phrases like 'n:,' 'n=*' was developed. These filters were combined with disease terms of five oncology indications to establish reproducibility. The search was focused on epidemiology, as the sample size is an essential criterion for such reviews. Lastly, the filter was validated against published literature reviews on the same five indications that used sample size as a restriction but did not use this restriction in their search.

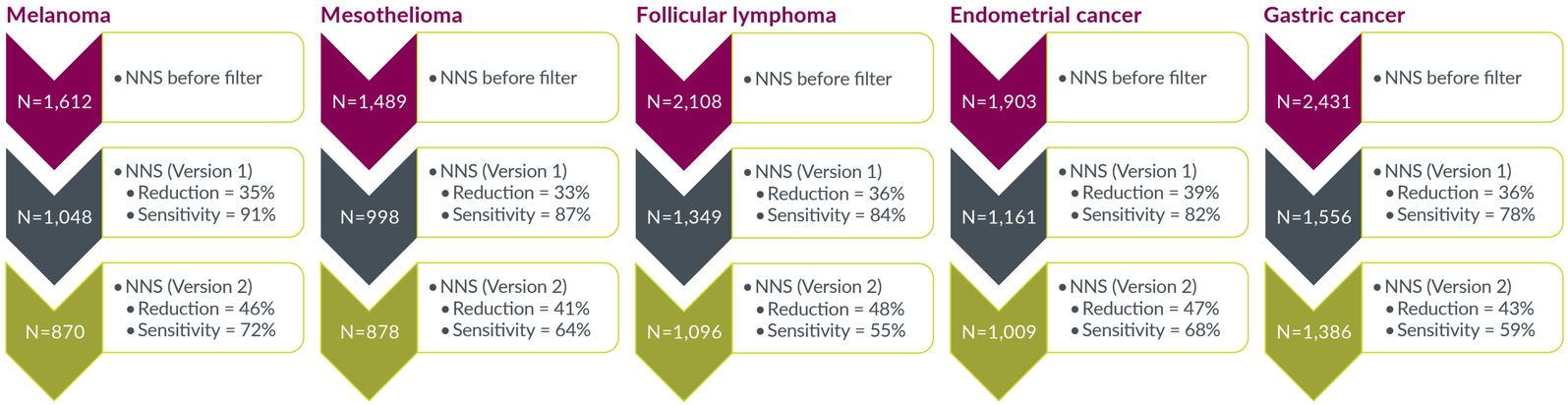
In our study, we employed two versions of search filters and a proximity sensor of "NEAR" to retrieve relevant articles included in evidence mapping. The filters were designed to capture numbers within the range of 10*-99*, where the asterisk (*) served as a wildcard for any number.

- Version 1: The filters encompassed the entire range of 10-99
- Version 2: We restricted the use of filters to a narrower range of 10-30

Example sample size filter:
((20* OR 21* OR 22* OR 23* OR 24* OR 25* OR 26* OR 27* OR 28* OR 29*) NEAR/2 (patient* OR case* OR adult* OR child* OR pediatric* OR geriatric* OR male* OR female*)):ab,ti

- These were preliminary search filters, and we developed two versions, i.e., a large-scale filter with truncated numbers ranging from 10 to 99 and a small-scale filter with truncated numbers ranging from 10 to 30. During the development phase, these filters were tested iteratively by starting from a small set of numbers. After the initial application of a few numbers, the retrieved output/citations were randomly checked for patient numbers. Based on the results, we repeated these test rounds. After several rounds, we finalized the two preliminary search filters
- We compared the impact on the number needed to screen (NNS), i.e., before and after applying the sample size filter. The reduction percentages were calculated for both versions. Secondly, we identified five separate published reviews in five indications and mapped their final included evidence with our retrieved output after applying the sample size criteria. The sensitivity of our filters (both versions) was evaluated based on the existing reviews, i.e., relevant studies identified with our sample size filters/total studies identified in the reviews without any filter

Results



The search resulted in 1489, 1612, 2108, 2431, and 1903 NNS for mesothelioma, melanoma, follicular lymphoma, gastric, and endometrial cancer, respectively. After applying the filter, the NNS was reduced by 33-48%. Despite the reduction of NNS, the search retrieved some non-relevant studies, such as studies mentioning years, percentages, specific numbered biomarkers, or the performance status of patients.

We observed that it was challenging to retrieve older studies as previous reporting standards didn't recommend mentioning the sample size in abstracts. However, this gap can be covered through manual searching or restricting this filter's use in reviews to retrieve the latest evidence. The filter's maximum sensitivity (on applying version 1 of search filters) across all disease indications ranged from 78-91% (melanoma: 91%, mesothelioma: 87%, follicular lymphoma: 84%, endometrial cancer: 82%, and gastric cancer: 78%).

Conclusions

The search filter would need further refinement and testing. However, it significantly reduces the NNS and can be considered in targeted reviews with sample size restriction.

Parexel International
2520 Meridian Pkwy
Durham, NC 27713, USA
+1 919 544-3170
www.parexel.com