No Limitations to Language, Date, Publication Type, and Publication Status in Search Step of Systematic Reviews

Ghazaleh Aali, BSc MSc
Systematic Reviewer at University of Nottingham, Nottingham, UK;
Ghazaleh.Aali@nottingham.ac.uk

Farhad Shokraneh, PhD
Systematic Reviewer and Information Scientist at King's College London, London, UK;
Farhad.Shokraneh@kcl.ac.uk

What is new

- Across bibliographic databases, the recording of publication date, language, publication types and publication status is neither standard nor accurate.
- Exclusion of non-English literature imposes limitation to the generalizability of findings of systematic reviews.
- The emergence of unorthodox publication types challenges the definition of 'publication' as well of publication dates.
- Those review teams who are thinking of applying limitations to their search should benefit from involvement of an information specialist.
- Limitation to language, publication date, publication types and publication status is more reasonable in title-abstract or full text screening steps rather than search step of the systematic reviews.

Pieper and Puljak[1] report a pragmatic rationale on why we should not apply language limitations in the search step of the systematic reviews. While we agree with their conclusion, there are more practical reasons to support this argument. Furthermore, we expand the discussion to why limitation by date, publication type and publication status at the search step is inadvisable.

A. Imperfect Data and Metadata Documentation and Processing

In common with any other data collection and processing efforts in research, data collection in bibliographic databases faces issues such as missing data, incomplete data, or errors in entered or curated data and/or metadata. As information practitioners, we have faced many occasions in which the bibliographic records lack the data for the language field. Our experience shows that most of such records are in English. Ironically, limitation to English language will exclude such English records during the search because the database will retrieve only the records that explicitly have defined English as a 'value' in their language data 'field', not those English records that are missing such value for the field. Leave alone that in some databases the data entry is manual and human error could be responsible for any unpredictable data entry mistakes such as assigning Chinese as a value to the language field of an English paper. Apart from that, some records have more than one recorded language.

Information specialists have found a way to circumvent the problem of missing English records using the so-called 'Not Not' command to cover the records with no language data (Figure 1 and

Figure 2). However, for three reasons it is not necessarily helpful: 1. This does not solve the problem with data entry errors; 2. While the studies link the inclusion of information specialists to the reporting quality and reproducibility of the systematic reviews[2,3], some systematic review teams still resist including an information specialist in their review team or do not have access to one; 3. It is a luxury to have an information specialist in the review team in low or middle-income settings.

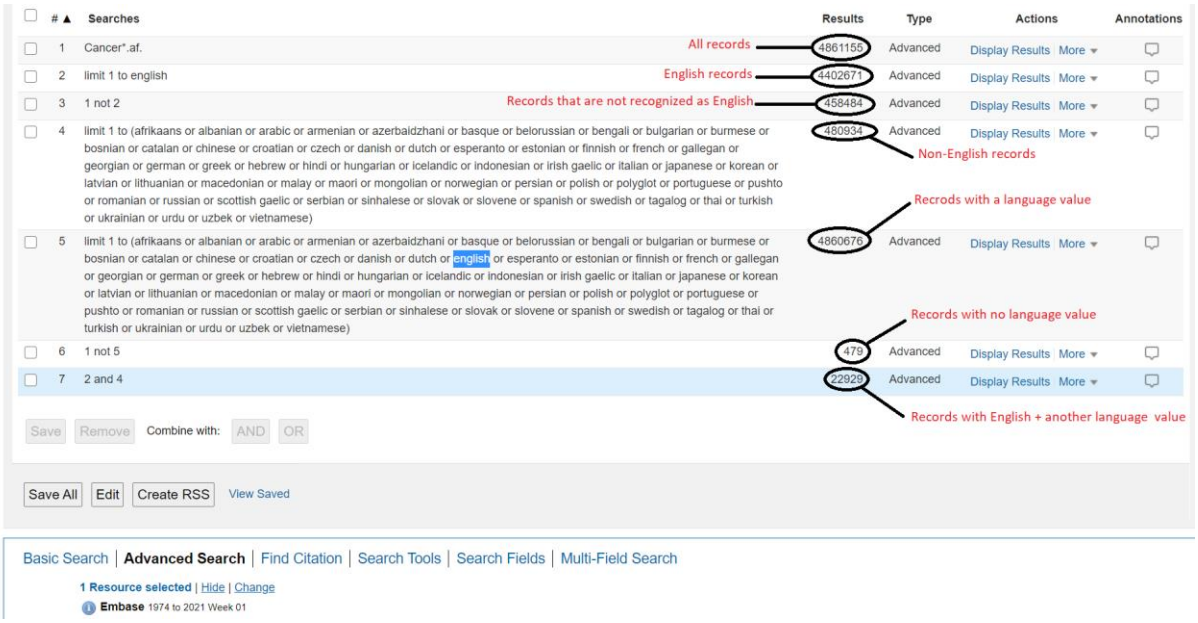

Figure 1: Language of records in the search results of a simple search in Embase

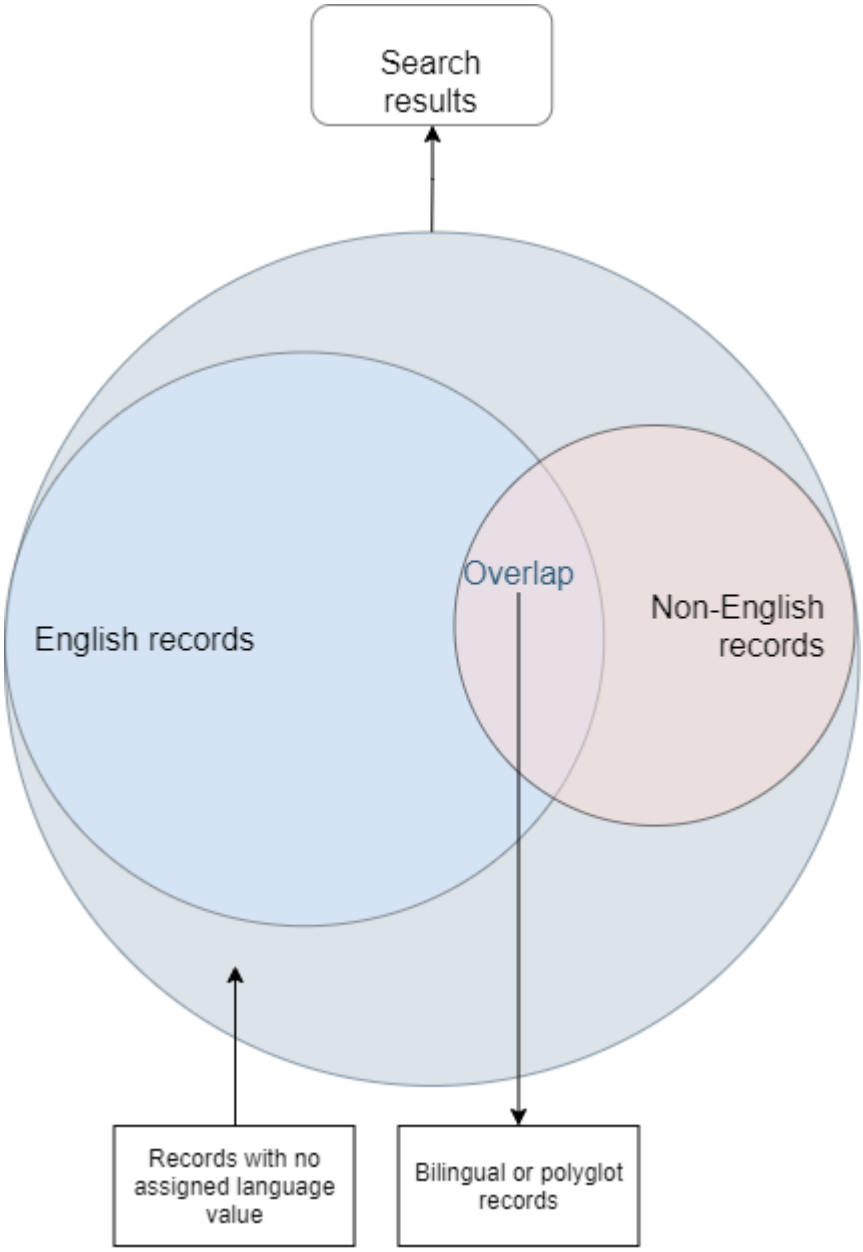

Figure 2: Venn diagram for language of all documents in search results in Embase

B. Not All Databases Support Language Limitation

Cochrane Library interface does not support language limitation so while searching Cochrane Central Register of Controlled Trials (CENTRAL) there will be non-English records among the results and since the language field is not documented for such records, they cannot be systematically identified and excluded in title-abstract screening stage.

C. Missing Data During Export/Import Functions

Some databases provide a limited number of exportable fields per record and language field may not be one of them. Even if we export the language data, it is possible that import filters in citation managers such as EndNote ignore such data during the importing process.

D. Methodological Disadvantages

In a methodological context and in relation to pharmacogenomics and pharmacogenetics, language limitation sets barriers to generalisability of findings to non-English populations[4]. This is in addition to what Pieper and Puljak[1] list as language bias.

Limitation to Date

In databases, recording the date of publication is inconsistent for journal papers when the journals follow the 'ahead of print' publishing model when – for example – an accepted paper is being published online in 2018 but gets finally published and paginated within a journal issue in 2021. Some bibliographic databases miss updating the publication dates and some records in databases have no value in their publication date field.

Because of the emergence of unorthodox publication types with multiple 'date' fields, the traditional indexing processes in bibliographic databases should adapt adding date fields other than publication date, revision date or database entry date. These types of publications add to the problem when the publications are not standard journal papers but web pages or grey literature with publication, revision, and update dates. A recent indexing decision in CENTRAL, includes records from clinical trials registries. Such registry records have multiple date entries – registration, the last update, and completion dates – none of which equals to the publication date. With a publication date limit to after 2005 in a search, a trial started in 2003 but completed in 2007 may be excluded from the search if the database records the start date of the trial as the publication date.

Limitation to Publication Type

It is not just that Cochrane guidelines advise against limitation to publication type[4], for the aforementioned reasons, limiting the search to publication types is also unreliable. The definitions of publication types are not globally standard; sometimes the journal editors rightfully allow the length of editorials, letters and commentaries to exceed the publisher's assigned word count. Such decisions make it possible to find the full report of a randomised controlled trial as a letter to the editor. Excluding the letters from search could easily exclude a trial. As another instance, some of the databases index both conference abstract and conference papers as conference proceedings and do not discriminate conference papers from conference abstracts which makes the exclusion of conference abstracts challenging.

Limitation to Publication Status

In December 2019 and before the start of pandemic, we suggested to Elsevier that Embase should start indexing clinical trials registry records and records from preprint servers in a 'living indexing' manner[5]. As a result, not only Embase but also Clarivate Analytics – the owner of Web of Science – have started indexing preprint documents. Almost all the records in clinical trial registries and some of the preprint records are 'living' records at some point of their lives. It is challenging to assign the label such as published, unpublished, or ongoing to such records. Limitation of search to published or peer-reviewed or completed studies could excluded valuable borderline and non-binary records with unclear agreeable publication status.

Bottom line

It is better to focus the systematic review searches to the retrieval of records relevant to the research question and not to remove records during the search applying other limitations. Although such limitations seem fast and pragmatic solutions for rapid reviews and scoping reviews, removing records using the database facilities automatically and during the search step of

systematic reviews may withdraw some of the relevant records without human approval. However, exclusion of records in title-abstract or full text screening steps requires human check and approval that reduces the chance of exclusion of relevant records because of the abovementioned technical and/or technological barriers of the databases.

Funding

None.

Conflict of Interests

None

References

1. Pieper D, Puljak L. Language restrictions in systematic reviews should not be imposed in the search strategy but in the eligibility criteria if necessary. Journal of Clinical Epidemiology 2021; InPress. DOI: https://doi.org/10.1016/j.jclinepi.2020.12.027
2. Koffel, J. B., & Rethlefsen, M. L. (2016). Reproducibility of Search Strategies Is Poor in Systematic Reviews Published in High-Impact Pediatrics, Cardiology and Surgery Journals: A Cross-Sectional Study. PLoS One, 11(9), e0163309. doi:10.1371/journal.pone.0163309
3. Meert, D., Torabi, N., & Costella, J. (2016). Impact of librarians on reporting of the literature searching component of pediatric systematic reviews. J Med Libr Assoc, 104(4), 267-277. doi:10.3163/1536-5050.104.4.004
4. Lefebvre, C., Glanville, J., Briscoe, S., Littlewood, A., Marshall, C., Metzendorf, M.- I., Noel- Storr, A., Rader, T., Shokraneh, F., Thomas, J., Wieland, L.S. and (2019). Searching for and selecting studies. In Cochrane Handbook for Systematic Reviews of Interventions (eds J.P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M.J. Page and V.A. Welch). https://doi.org/10.1002/9781119536604.ch4
5. Shokraneh F. Random Reflections on Embase. Embase Strategy Meeting; 9 December 2019; Elsevier, Amsterdam, the Netherlands. DOI: 10.13140/RG.2.2.27112.29444

Figure 1

Click here to access/download;Figure;Figure 1.png

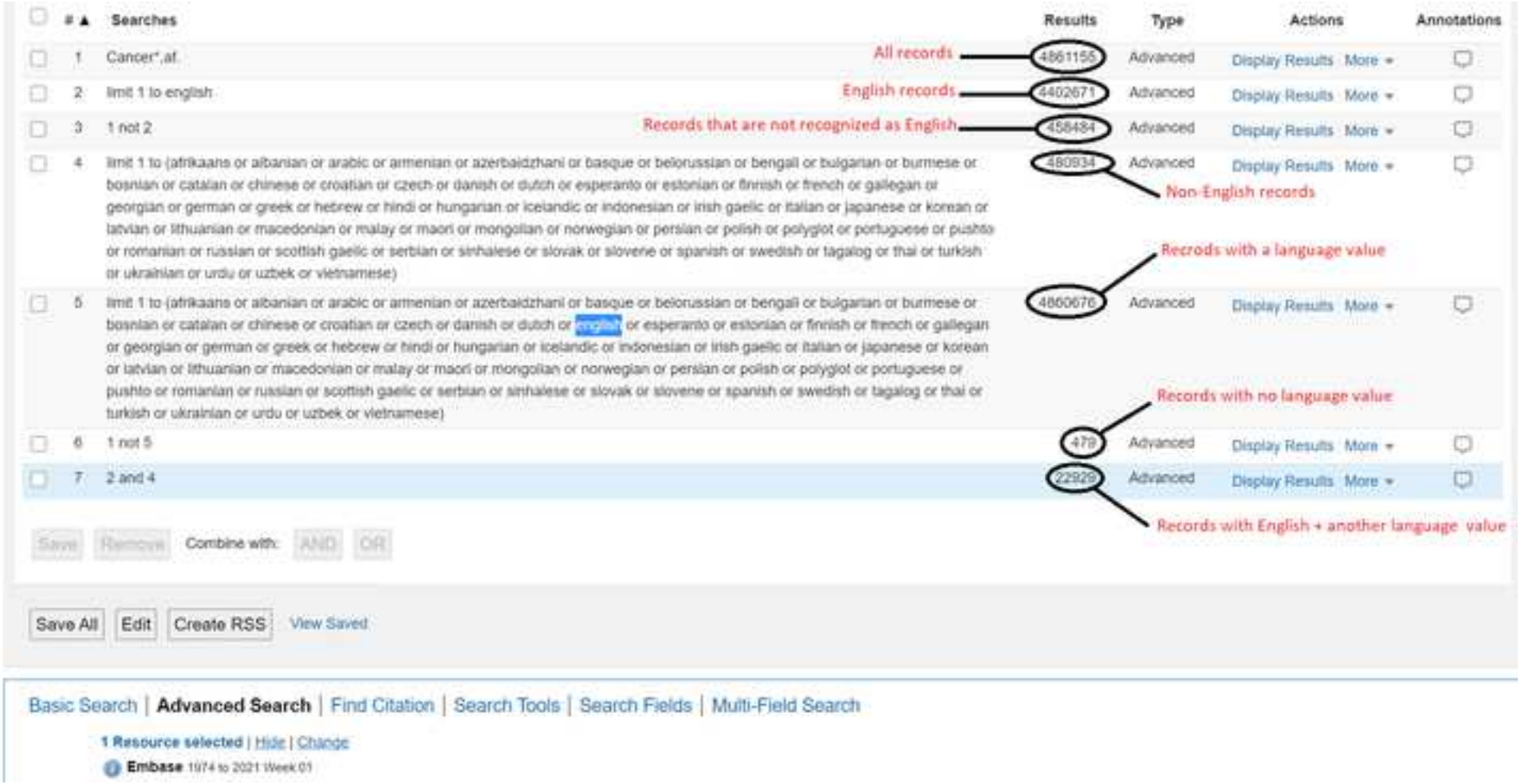

Figure 2