Validation of artificial intelligence for performing systematic literature review searching for clinical efficacy and safety data

Vikte Lionikaite[1], Alistair S Curry[1], Audrey E Brown[1]

[1]Genesis Research, Newcastle Upon Tyne, United Kingdom; email: solutions@genesisrg.com

POSC3000

Introduction and study objective

- Performing systematic literature reviews (SLRs) on clinical data is a time-consuming process, partly due to the sheer volume of published data available.
- Artificial intelligence (AI) technology can be used to conduct literature searches that are more effective and efficient.
- The objective of this study was to evaluate and validate the use of artificial intelligence and machine learning in clinical evidence gathering.

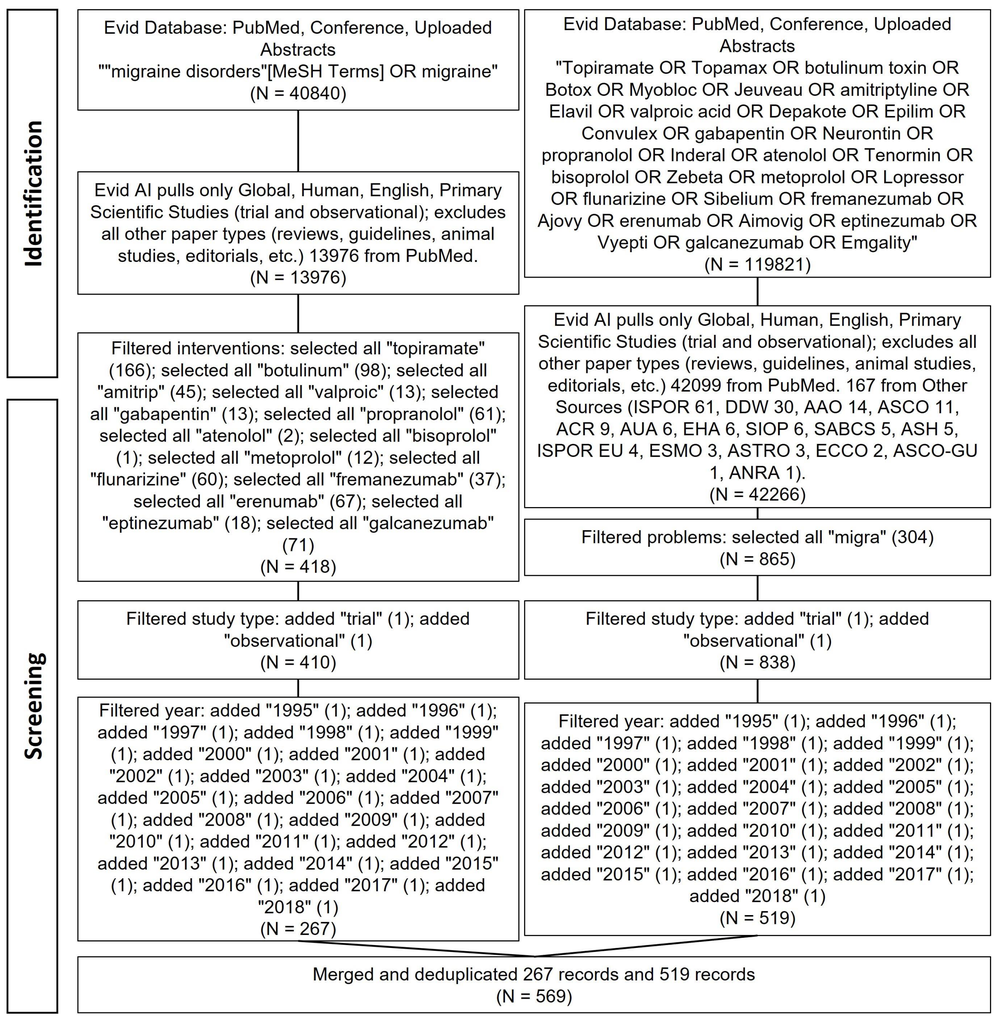
Methods

- A retrospective analysis was conducted based on the clinical effectiveness searches from health technology assessments (HTAs).
- The most recent HTAs from the National Institute for Health and Care Excellence (NICE) were randomly selected if they included sufficient detail on the efficacy and safety SLR search criteria and provided the reference list of included studies.
- An AI platform (Evid Science) capable of extracting disease terms, interventions, and outcomes from abstracts in PubMed and selected conferences was used to perform searches corresponding to the search criteria for each SLR[1].
- The AI platform employs filters based on the information available in each abstract allowing the user to filter references according to the interventions, outcomes and/ or study types of interest.

Results

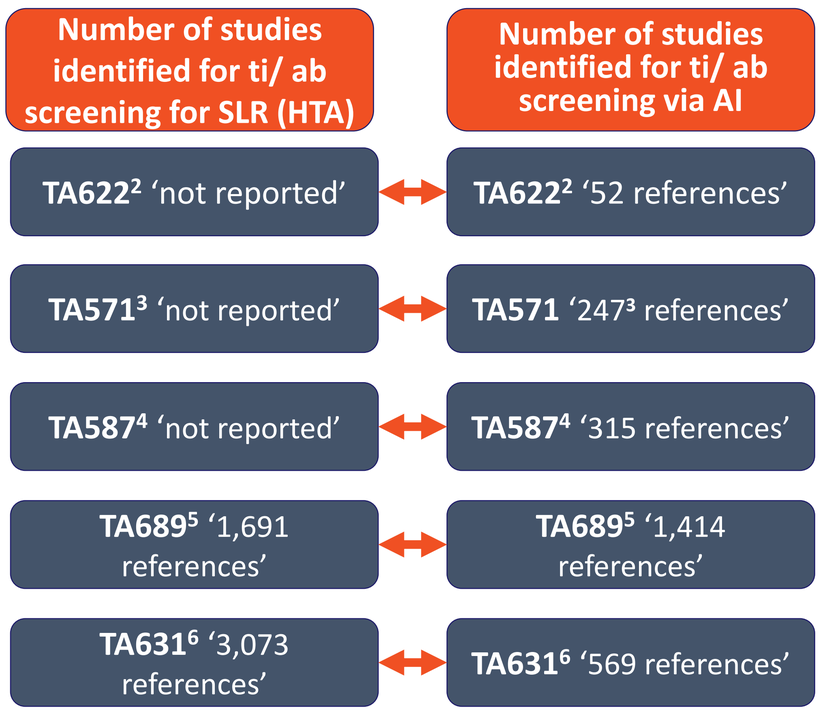
- Five HTAs containing sufficient detail on the efficacy and safety SLRs were identified across a range of indications: type 1 diabetes (TA622)[2], non-small-cell lung cancer (TA571)[3], multiple myeloma (TA587)[4], chronic lymphocytic leukaemia (TA689)[5], and migraine (TA631)[6].
- For each clinical SLR identified, a dual search was performed on the AI platform. The first search contained disease terms filtered by interventions, study type, and year. The second search contained terms for the interventions of interest, filtered by disease, study type, and year. The searches were merged and deduplicated (**Figure 1**).

Figure 1: Example of TA631[6] search and AI-generated PRISMA diagram



- Across all five HTA SLRs, 42 relevant clinical evidence publications were identified.
- The AI was able to identify 40 of the 42 publications.
- Of the two publications not identified, one was from a conference not searched on the AI platform. The AI was unable to read and extract data from the second publication which may be related to the irregular formatting of the abstract.
- Information on the total number of studies screened for inclusion at the title/ abstract (ti/ ab) stage was available from two SLRs (**Figure 2**).

Figure 2: Number of references screened for SLR vs AI platform



- In TA689[5], a total of 1,691 studies were screened for inclusion across 3 databases searched while 3,073 studies were screened in TA631[6].
- In contrast, the searches developed using the AI platform identified 1,414 and 569 studies to be screened, respectively.
- These AI platform searches represent 83.6% and 18.5% of studies screened by the HTA SLRs, and enabled identification of 18 out of the 19 studies included in the 2 SLRs.
- The Evid Science AI platform enhances the search precision enabling relevant studies to be identified while likely reducing screening time. The extent of the time savings are likely to vary according to indication.

Conclusion

- In this study, the error rate from inputs to the machine learning was 2.4% (1 of 41).
- HTA-compliant SLRs require multiple databases to be searched, as well as searches of the grey literature, databases of conference abstracts, and review of reference lists of included publications and reviews. Therefore, the error from inputs to the machine learning may be mitigated by performing additional searches as per SLR protocol.
- The use of AI enhances search precision enabling relevant studies to be identified while likely reducing screening time for irrelevant abstracts. Further research is required to understand the typical magnitude of time savings.

References

1. Michelson M, Chow T, Martin N, Ross M, Tee Qiao Ying A, Minton S. Artificial Intelligence for Rapid Meta-Analysis: Case Study on Ocular Toxicity of Hydroxychloroquine. J Med Internet Res 2020;22(8):e20007.
2. NICE [TA622] (2020) Sotagliflozin with insulin for treating type 1 diabetes
3. NICE [TA571] (2019) Brigatinib for treating ALK-positive advanced non-small-cell lung cancer after crizotinib
4. NICE [TA587] (2019) Lenalidomide plus dexamethasone for previously untreated multiple myeloma
5. NICE [TA689] (2021) Acalabrutinib for treating chronic lymphocytic leukaemia
6. NICE [TA631] (2020) Fremanezumab for preventing migraine