SA29

ActiveSLR®: Optimizing Efficiency in Systematic Literature Reviews with Artificial Intelligence

Authors: Subrata Bhattacharyya[1], Md Sohail Aman[2], Kopal Dixit[2], Abhra Roy Choudhury[2],Deepti Rai[2], Anup Shaw[2]

[1]PharmaQuant International Ltd. [2]PharmaQuant Insights Pvt. Ltd.

INTRODUCTION

- Systematic literature reviews (SLRs) play a critical role in consolidating knowledge, identifying research gaps, and informing evidence-based medical practice.[1,2]
- The involvement of artificial intelligence (AI) to assist with outcomes research activities continues to increase; both in terms of the range of uses and the accuracy with which AI can perform time-intensive tasks.[3]
- While there are many AI-assisted tools available for conducting SLRs, each tool has inherent advantages and limitations. The efficiency and value of each tool should be quantified, giving potential users the opportunity to make a value-based decision.
- ActiveSLR®, a new application that supports literature review management and execution, includes large language model (LLM) software. However, given the recency of its launch, the potential efficiency of the AI-assisted functionalities has not yet been assessed.[4]
- There is a growing need to assess whether adding AI capabilities to these tools could improve efficiency.

OBJECTIVES

This initial research aimed to evaluate the efficiency of ActiveSLR® software for conducting SLRs compared with traditional practices involving Excel®.

METHODS

Two steps in the SLR process were identified where AI can be used:

Table 1: Comparison between traditional and AI-driven literature review steps

Methods Element	Traditional Process	AI functionality of ActiveSLR®
Deduplication	EndNote (version 21) deduplication (using default settings) with or without deduplication using Excel® conditional formatting	AI/Machine learning (ML) based deduplication to identify true and possible duplicates
Screening	Excel® based screening at title-abstract and full-text level	AI-driven population-intervention-comparator-outcome (AiPICO) identifier and keyword highlight function were used

We developed four oncology projects to assess the efficiency of AI-assisted SLR. Four independent reviewers conducted separate reviews, independently formulating review questions and developing population, intervention, comparator, and outcome (PICO) criteria before conducting searches on MEDLINE and EMBASE. The citations in these four projects ranged from 400 to 2,000 (Table 2).

Deduplication:

- All the four projects were used to test deduplication efficiency.
- Efficiency was assessed in two ways: researcher time spent on the deduplication process and also the accuracy of deduplication.
- One reviewer conducted deduplication on EndNote (version 21) and another reviewer conducted deduplication from same set of citations on ActiveSLR®. A third reviewer used conditional formatting on Excel® to manually validate the accuracy of identified duplicates on both ActiveSLR® and EndNote (version 21).

Screening

- Two of the four projects (renal cell carcinoma and pancreatic cancer) were used to assess screening efficiency.
- Non-duplicate studies were included for double-blind screening. The potential screening time that could have been saved was then estimated.
  - Screening time can depend on a range of factors, including complexity of the SLR PICOS and the clarity of the title and abstract.
  - Excel® was selected for this comparison as it is the most commonly-used platform apart from automated screening tools and could be the common comparator for benchmarking.
  - For each project one reviewer screened studies using Excel® at title-abstract level and full-text level and another reviewer screened same set of studies using ActiveSLR®.
  - Same reviewers were not involved in both the screening processes to avoid learning effect coming into the results.

Data Extraction

- The two projects used to assess screening efficiency were also used for data extraction.
- One reviewer used data extraction platform with the centralized, simple and menu driven system. Second reviewer used Excel® for the data extraction.

METHODS *continued..*

- Excel® was selected for this comparison as it is the most commonly-used platform apart from SLR tools and could be the common comparator for benchmarking.
- One reviewer extracted intention-to-treat (ITT) population data from identified randomized controlled trials (RCTs) on overall survival (OS), progression-free survival (PFS) and objective response rate (ORR) with Excel®. Another reviewer extracted same data from same RCTs on ActiveSLR®. Same reviewers were not involved in the data extraction on both the platforms to avoid learning effect coming into the results.

Table 2: List of test projects completed in ActiveSLR®

Disease	Description	Search Hits
Salivary gland neoplasm	To identify the randomized clinical trials (RCTs) of salivary gland neoplasm over the past 20 years	MEDLINE: 134 EMBASE: 272
Renal cell carcinoma	To determine the clinical effectiveness of treatments in adults with untreated advanced clear cell renal cell carcinoma	MEDLINE: 281 EMBASE: 240
Pancreatic cancer	To determine clinical effectiveness of systemic treatments in patients with previously untreated locally advanced/metastatic pancreatic cancer	MEDLINE: 291 EMBASE: 753
Hepatocellular cancer	To determine clinical effectiveness of systemic treatments in patients with hepatocellular cancer	MEDLINE: 1379 EMBASE: 756

RESULTS

Deduplication

- From the 4,106 references, across the four searches, EndNote (version 21) identified a total of 528 possible duplicate records, whereas ActiveSLR® identified 716. We ran manual checking with Excel® conditional formatting function and identified the same 716 duplicates in total.
  - Due to the higher number of possible duplicates to check, the mean time spent per project was approximately seven minutes more for ActiveSLR® versus EndNote (version 21).
  - However, had a reviewer deduplicated in EndNote (version 21) for all four searches, this would have resulted in a total of 188 additional references undergoing double-blind screening.
    - ➢ Based on our observed Excel®'s average screening rates—450 title-abstracts and 54 full-texts per 6 hours—screening these 188 studies would have required an additional 177 minutes for single screening and 354 minutes for double screening. This would have been followed by further time needed for reconciliation.

Table 3: Comparing deduplication in ActiveSLR® vs Excel®

Tool	Duplicates identified	Time/project (mean)	False positive	Missed
ActiveSLR®	716	~14 minutes	0	0%
EndNote (version 21)	528	~7 minutes	2	26.2%
Excel®	716	~2.7 hours	0	0%

Screening

1. **Renal cell carcinoma**
   - Two reviewers separately screened each of the total 462 citations using Excel® and ActiveSLR® at title-abstract level.
   - The reviewer on ActiveSLR® took 4 hours to screen using AiPICO and keyword highlight features, while the Excel® process took 6 hours to screen the same studies.
   - The full text screening of 60 full texts took four hours using ActiveSLR® and eight hours using Excel®.
2. **Pancreatic cancer**
   - Two reviewers performed double-blind screening for all 885 citations using both ActiveSLR® and Excel®.
   - The reviewer on ActiveSLR® took 5 hours 45 minutes hours to screen whereas using Excel®, it took 12 hours.
   - A total of 85 full texts were reviewed. The process took five hours using ActiveSLR® and eight hours using Excel®.

RESULTS *continued..*

With ActiveSLR®, title-abstract screening and full text screening efficiency gain were 50-109% and 60-100% respectively vs Excel®. A reviewer in 6 hours could screen 828 title-abstracts and 96 full-text articles while with MS-Excel it was 450 title-abstracts and 54 full-text articles. ActiveSLR® also auto-fetched free full-texts in the full-text screening and enabled reviewer identify 234 available full-text in 6-hour vs manual identification of 120 full-text.

Table 4: Comparing screening efficiency in ActiveSLR® vs Excel®

Project	Screening stage	Time taken in ActiveSLR®	Time taken in Excel®.
Renal cell cancer	Title-abstract screening-462 citation	4 hours	6 hours
Pancreatic cancer citation	Title-abstract screening-885 citation	5 hours 45 minutes	12 hours
Renal cell cancer	Full-text screening-60 citation	4 hours	8 hours
Pancreatic cancer citation	Full-text screening-85 citation	5 hours	8 hours

Data extraction

Using ActiveSLR®, a reviewer could extract ITT data on OS, PFS, and ORR from over four randomized controlled trials (RCTs) in six hours. In comparison, using Excel®, the same reviewer could only extract the same data from fewer than three of those RCTs in the same timeframe."

Table 5: Comparing data extraction efficiency in ActiveSLR® vs Excel®

Project	Screening stage	Time taken in ActiveSLR®	Time taken in Excel®
Renal cell cancer	5 RCTs [only ITT data on OS, PFS, ORR were extracted]	6 hours 30 minutes	11 hours
Pancreatic cancer	5 RCTs [only ITT data on OS, PFS, ORR were extracted]	7 hours 47 minutes	11 hours 51 minutes

Discussion and Conclusion

- AI-assisted functionality continues to be developed for improving the efficiency of evidence generation and synthesis.
- However, the absence of standardized approaches to quantifying these potential benefits presents an industry-wide challenge when comparing benefits/time savings.
- This research suggests the AI assistance tools that have been assessed do indeed result in meaningful and consistent time-savings for users.
- However, further research is needed in a number of different areas:
  - The potential cost of a false positive duplicate has not been considered in this analysis. To do so would require:
    - Estimating the chances of an incorrectly-removed reference being one that should have been included.
    - Understanding the true impact of missed identification of duplicate references (which is subjective and dependent on the project).
  - To avoid the risk of a learning effect bias, future research could assess the time savings of AI functionality within the same software (rather than reverting to traditional methods of Excel® and EndNote (version 21)).
- The authors encourage industry-wide efforts to standardize how to assess and quantify the value of AI in literature review software, to ensure user selection is made with visibility and clarity as to potential benefits.

REFERENCES

1. Critical appraisal skill programme. Glossary.https://casp-uk.net/news/what-is-a-systematic-review/
2. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024). Cochrane, 2024. Available from www.training.cochrane.org/handbook.
3. Ge L, Agrawal R, Singer M, Kannapiran P, De Castro Molina JA, Teow KL, Yap CW, Abisheganaden JA. Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. Syst Rev. 2024 Oct 25;13(1):269. doi: 10.1186/s13643-024-02682-2. PMID: 39456077; PMCID: PMC11504244.
4. https://www.ActiveSLR®.io/

This study was funded by PharmaQuant Insights Pvt. Ltd., India

PharmaQuant www.pharmaquant.org linkedin.com/company/pharmaquant ActiveSLR