Existing automated tools to assist evidence generation and better qualification of registries and real-world data: A systematic review from the More-EUROPA project

Pascal Godbillot, Pierre Tang, Jeverson Moreira, Julien Tanniou, Sonia Zebachi, Billy Amzal

Quinten Health, Paris, France

Introduction

- Real-world data (RWD), particularly **patient registries**, are useful source of information required at multiple stages of the drug lifecycle.
- The **identification and selection of fit-for-purpose registries** play a critical role in regulatory and/or Health Technology Assessment (HTA) decision-making processes.
- **Registries identification, selection, and assessment** are not optimized for regulatory and HTA procedures.
- **Objective:** Provide insights on existing machine learning (ML) tools to assist evidence generation and qualification of registries and RWD.
- **Project:** The More-EUROPA project, involving 15 public and private organizations from 7 EU countries, evaluates the effective and ethical use of registry data to support patient-centered decisions by drug regulators and HTA agencies in Europe.

Methods

- Conduct a **systematic literature search in MEDLINE** using free-text search terms (Figure 1).
- Retrieve articles published between **2013- 2023**.
- Undertake a **landscaping exercise** to provide an overview of the availability, accessibility, and quality of RWD assets.
- **The results are triangulated** with the stakeholders' survey, and interviews (poster MSR91), and the second literature review on fit-for-purpose frameworks (poster MSR30).

Figure 1. Search terms

((("NLP" OR "tool" OR "pipeline") AND ("RWD" OR "real world data") AND (identif OR collect OR screen OR assess OR evaluat)) OR (("systematic review" OR "systematic reviews")AND ("automated screening" OR ("automation" AND ("machine learning" OR "tool" OR "tools"))) OR (("data quality" OR "quality assessment") AND ("RWD" OR "real world data")) OR ("big data" AND ("RWD" OR "real world data") AND ("registry" OR "registries")) OR (("RWD" OR "real world data") AND outcome AND "identification" AND "assessment")) AND ("2013/01/01"[Date - Create] : "2023/03/14"[Date - Create])

Results

- **30 articles** selected after abstract and full text screening.
- **60+ automated tools** assist in evidence generation and the systematic review process.
- **10** are highly cited and up-to-date (Table 1).

Table 1. Most cited automated tools assisting evidence generation

TOOL NAME	CATEGORY	FUNCTIONALITIES	LICENSE STATUS	LATEST UPDATE
Rayyan	Web-applications	Screening	fee-based	2023
Nested Knowledge	Web-applications	Search, Screening, Data extraction	fee-based	2023
DistillerSR	Web-applications	Search, Screening, Data extraction	fee-based	2023
EPPI-Reviewer	Web-applications	Screening	fee-based	2023
Abstrackr	Web-applications	Screening	free	2019
Leximancer	Web-applications	Search	fee-based	2023
SR-Accelerator	Web-applications	Search, Screening	free	2023
ASReview	Desktop-applications	Screening	free	2023
RobotReviewer	Web-applications	Data extraction	free	2022
SWIFT-Review	Desktop-applications	Search, Screening	free	2019

- **Main functionalities**:
  - **Search:** Retrieving records from different sources
  - **Screening:** Selecting suitable records retrieved by the search
  - **Data extraction:** Extracting data from retrieved records

> No automated tools specifically designed for better qualification of registries.

- **RWD catalogues initiatives exist:**
  - **Institutional catalogues of registries**: ENCePP, EHDEN, and future EMA Metadata Catalogue.
  - **Institutional catalogues of registry-based studies**: ClinicalTrials.gov, and EU PAS Register.
  - **Private initiative**: IQVIA Health Data Catalog.
- **Key findings regarding automated tools**
  - The **lack of technical** skills is the most frequent barrier to tool adoption → intuitive and user-friendly web-based platform are more popular.
  - The **deficiency of tools** providing all functionalities, **most are task-specific**.
  - Tools with the most features and frequent updates have **fee-based license**.
- **Key findings regarding RWD catalogues**
  - **Data browsing and filtering** are the most common features.
  - **Traditional search** using keyword matching or pre-defined filters is the most common, **few provide advanced search capabilities** (e.g., semantic search, entity recognition).
  - Catalogues rely on the **willingness of data owners** to fill in information about their data sources. It provides availability and reliability of metadata associated with the RWD but limits coverage, and exhaustiveness.

Conclusion

- Insights from the state of the art provided an overview of the currently available ML tools to assist evidence generation and systematic review procedures.
- These findings will (Table 2):
  - **Inform the development** of a screening tool to identify suitable registries in the More-EUROPA project.
  - **Guide the design of the tool features**.
- A Minimal Viable Product (MVP) approach will be used – version of a tool with sufficient features to attract early-adopter with minimal effort.
- Iterations will then follow to integrate users' feedback at each step of the tool development.
- The tool will focus on user experience and user feedback to ensure its successful adoption.

Table 2. Potential features to integrate into the More-EUROPA tool

Search	Screening	Extraction	Actionability
Data aggregation from multiple sources	Ranking of results according to user query	Automated extraction of PICO terms	Data visualisations
Keyword search	Summarization of the data asset	Automated extraction of data assets' metadata	Comparison of results
Semantic search	Dynamic display of data assets	Risk of bias	Up-to-date information
Metadata search (pre-defined filters)	Keyword highlighting	Quality checks	Export of results

References:
Ramagopalan SV, Simpson A, Sammon C. Can real-world data really replace randomised clinical trials? BMC Med. 2020 Jan 15;18:13. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6961357/
Beaulieu-Jones BK, Finlayson SG, Yuan W, Altman RB, Kohane IS, Prasad V, et al. Examining the Use of Real-World Evidence in the Regulatory Process. Clin Pharmacol Ther. 2020 Apr;107(4):843–52. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7093234/
European Medicines Agency. List of metadata for Real World Data catalogues. EMA- heads of Medicines Agency, Amesterdam, Netherlands. 2022 May 31.

more-europa@umcg.nl

Website   LinkedIn