An Interactive Data Extraction System to Create the Living Systematic Reviews and Meta-Analysis

MAYO CLINIC

Irbaz Bin Riaz, MD, MS [1,4]*, Huan He, PhD [2]*, Syed Arsalan Ahmed Naqvi, MBBS [1], Rabbia Siddiqi, MBBS [3], Noureen Asghar, MBBS [3], M.Hassan Murad, MD, MPH [4], and Hongfang Liu PhD [2]

[1] Department of Oncology, Mayo Clinic, Phoenix, AZ, USA
[3] Dow University of Health Sciences, Karachi, Pakistan
[2] Department of AI and Informatics Research, Mayo Clinic, Rochester, MN, USA
[4] Mayo Clinic Evidence Based Practice Center, Mayo Clinic, Rochester, MN, USA

* Those authors have contributed equally as co-first authors
riaz.dr@mayo.edu
He.Huan@mayo.edu
† Liu.Hongfang@mayo.edu

Background

It takes months to years for conducting rigorous systematic review (SR) and meta-analysis (MA), and it involves developing a search strategy, screening for relevant citations, extracting and analyzing data and ironically the review can be outdated as soon as it published. The current process of creating SRMAs is inadequate to keep pace with rapid influx of evidence as seen in COVID-19 pandemic and dynamic fields like Oncology.

Thus, living SRs (LSRs)—which are updated as soon as new evidence becomes available— are necessary to overcome the limitations of conventional reviews.

Previously, we have described semi-automating the screening process and analyses, and here we describe our innovative method for **data extraction** to decrease the effort for creating and maintain LSRs.

Methods

Data extraction is a time-consuming task, which is highly dependent on manual operation and cannot be processed by machines automatically due to its complexity and expertise requirement. Therefore, we adopted a user-centered approach to design this tool to facilitate the data extraction:

The main module is the interactive extractor, which is shown on the right. It contains multiple panels and features to support the data extraction process. For example, the study list panel (1) helps users to select which study to extract. The customized popup menu help users to set the highlighted text as extraction item (2). The integrated PDF viewer (4) also supports the customized popup menu to show extraction items. The extraction task panel can show extraction items based on pre-defined extraction task (3).

Other modules provide extraction task definition and result table overview for users to manage the extraction in multiple ways.

User Interface Design for Data Extraction

Users can filter the list by PMID, author name, or title to find the study to perform data extraction.

The imported studies are listed in a table, users can click the study row to check details.

By clicking the study in this list, the user start data extraction for the selected study. The list of extraction tasks and the meta data of study (e.g., title, author list, and abstract, etc.) is displayed for users to check details. In addition, the PDF files could also be displayed if users uploaded.

Keywords are highlighted automatically to help users identify potential words of interest.

A list of pre-defined extraction tasks is displayed to help user decide which task to extract.

When right click on the highlighted text, a popup menu will be displayed to show available extraction options for the highlighted text. Users can click on the option to set the extraction target. For example, highlight "1.02" and select "TE" will extract the 1.02 for the TE in current extraction task.

Each extraction task may have different items for extraction for various purpose. As shown in this figure, we select three tasks with different items. Users can input those items manually or use right-click popup menu to set them.

By clicking the "PDFs" tab, users can upload and view PDFs files for data extraction

An embedded web-based PDF viewer is integrated to show user-uploaded PDF files for data extraction. It also support the popup menu for extraction.

User Interface Design for Extraction Task Definition and Result Table

Users can select which items to be displayed in this result table to focus on the interested items.

The items in the extraction task can be defined in a tree-like view.

The extraction results of all studies are listed in a table for users to browse the overall status.

Discussion

Extracting data for LSRs is particularly challenging as it requires management of the key information from meta data and unstructured free texts such as PDF files and web pages. Although existing tools, such Rayyan and Covidence, work well on independent tasks (e.g., screening, data retravel) and generate data files, linking those isolated data files correctly is challenging and requires tedious manual operations in multiple tools. Thus, to ease the burden of using we created a user-friendly interface which brings together data from multiple sources and facilitates data extraction and creation of summary tables and data tables for analysis.

Next steps include integrating this data extraction system with our previously created modules for screening and analyzing the data thereby creating a pipeline for true, living systematic reviews which will be updated in "almost" real time.