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

Irbaz Bin Riaz, MD, MS [1, 4*], Huan He, PhD [2*], Syed Arsalan Ahmed Naqvi, MBBS [3]
Rabbia Siddiqi, MBBS [3], Noureen Asghar MBBS [3],
M. Hassan Murad, MD, MPH [4], Hongfang Liu, PhD [2]
[1]Department of Oncology, Mayo Clinic, Phoenix, AZ, USA
[2] Department of AI and Informatics Research, Mayo Clinic, Phoenix, AZ, USA
[3] Dow University of Heath Sciences, Karachi, Pakistan
[4] Mayo Clinic Evidence Based Practice Center, Mayo Clinic, Rochester, MN, USA

Introduction: 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a search strategy, screening for relevant citations, extracting and analyzing data and ironically the review can be outdated as soon as it published.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: We designed three major modules to facilitate the data extraction: the outline module (Fig. 1(A)), the table module (Fig. 1(B)), and the interactive extractor (Fig. 1(C)). The purpose of outline module is to create a skeleton of tables for data extraction. In the outline module, the users specify the outcomes of interest as well as the structure of the data format required to analyze each outcome. Users are provided with a range of standard options to create summary and data tables and are also provided with the flexibility to generate new variables as necessary. The tables in the table module are automatically populated with studies to be included in the review with the meta-data such as year of publication, author and journal name and rest of tables are completed with the help of extractor module. The studies are listed row by row. Users could select the "check" option and select each particular study for extraction. When a study row in this table is clicked, the interactive extractor would show its text details, including the abstract and full-text PDF files. The extractor module allows user to extract data from associated text. The user simply needs to highlight the text, right-click on the highlighted text, select which attribute the highlighted text belongs to. Then the highlighted text would be saved as the value of the selected attribute for this study. Thus, the tables generated from these modules can be presented to summarize the results of the systematic review or perform further analyses such as pairwise or network meta-analyses, create summary of findings tables and evidence maps.

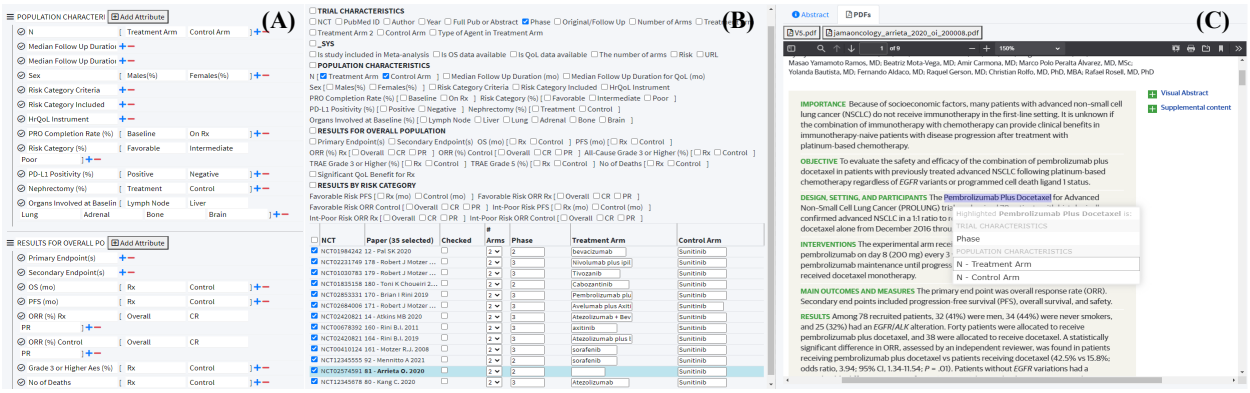

Figure 1. The screenshots of our extraction tool. (A) the outline module; (B) the table module; (C) the interactive extractor.

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.

* Those authors have contributed equally as co-first authors