MAYO CLINIC

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, MBBS¹, Rabbia Siddiqi, MBBS³, Noureen Asghar, MBBS³, M.Hassan Murad, MD, MPH⁴, and Hongfang Liu PhD²

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 metaanalysis (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a search strategy, screening for relevant citations, extracting and analyzing data and ironically the review can be outdated as soon as it published.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.

Study List Panel The imported studies are listed in a table, users Users can filter the list by PMID, author name, or 🖻 Adjuvant Renal Cell Carcinoma ₹Ξ Extract by paper Renal Cell Carcinoma can click the study row to check details. title to find the study to perform data extraction. 9 studies (included in SR) PMID, short #, first author name Outcomes embrolizumab after Nephrectomy in Renal-Cell Carcinoma Adjuvant Sunitinib for High-risk Renal Cell Carcinoma After Nephrectomy: Subgroup Analyses and Updated Overall Survival Results 530 28967554 Robert J Motzer et al 2017 Randomized Phase III Trial of Adjuvant Pazopanib Versus Placebo After Nephrectomy in Patients With Localized or Locally Advanced Renal Cell Carcinoma. 531 28902533 Robert J Motzer et al 2017 Adjuvant Sorafenib for Renal Cell Carcinoma at Intermediate or High Risk of Relapse: Results From the SORCE Randomized Phase III Intergroup Trial. 532 33052759 Tim Eisen et al 2020 Adjuvant Sunitinib in High-Risk Renal-Cell Carcinoma after Nephrectomy. By clicking the study in this list, the user start 533 27718781 Alain Ravaud et al 2016 Adjuvant sunitinib in patients with high-risk renal cell carcinoma: safety, therapy management, and patient-reported outcomes in the S-TRAC trial. 534 30412222 M Staehler et al 2018 data extraction for the selected study. The list of

User Interface Design for Data Extraction

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 semiautomating 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

| 535 26969090 Naomi B Haas et al 2016 | 7 | Adjuvant sunitinib o trial. | r soraf | nib for high-risk, non-metastatic renal-cell carcinoma (ECOG-ACRIN E2805): a double-blind, placebo-controlled, randomised, ph | extraction tasks and the meta data of study (e.g., | | | | | |
|---|----------------|--------------------------------|----------|--|--|--|--|--|--|--|
| 536 33461782 Robert J Motzer et al 2021 3 Adjuvant Pazopanib Phase 3 PROTECT Tria | | | | Placebo After Nephrectomy in Patients With Localized or Locally Advanced Renal Cell Carcinoma: Final Overall Survival Analysis | for users to check details. | | | | | |
| 537 30346481 M Gross-Goupil et al 2018 | 13 | Axitinib versus place | ebo as a | n adjuvant treatment of renal cell carcinoma: results from the phase III, randomized ATLAS trial. | In addition, the PDF files could also be displayed if users uploaded. | | | | | |
| Interactive Extractor | | | | | | | | | | |
| | juvant Re | nal Cell Carcinoma | ¥∃ Ex | tract by paper | ý ³ 💄 | | | | | |
| 9 studies (included in SR) PMID, sh | ort #, first a | author name | × | Abstract PDFs | | | | | | |
| # P Naomi B Haas et al 2016 | | primary | | 2016-03-09 - Lancet | | | | | | |
| 503 3 is selected in 7 outcomes | de l | fault Disease free surviva | aba | Adjuvant sunitinib or sorafenib for high-risk, non-metastatic renal-cell carci | noma (ECOG-ACRIN E2805): a | | | | | |
| 530 2 Save O Toggle Outcome na | | Save 🗹 Fill 3 🗸 | ligh | double-blind, placebo-controlled, randomised, phase 3 trial. | | | | | | |
| 531 2 Interactive Table | Cor | mp 1 Comp 2 | rial | Naomi B Haas, Judith Manola, Robert G Uzzo, Keith T Flaherty, Christopher G Wood, Christopher Kane, Michael Jewett, J | ^{al} Kowwords are highlighted automatically to help | | | | | |
| 532 3 🗁 Network MA | Resu | lt | Ren | George Wilding, David Cella, Lynne Wagner, Surena Matin, Timothy M Kuzel, Wade J Sexton, Yu-Ning Wong, Toni K Choue | ir in the second s | | | | | |
| 533 2 D Primary Analysis | TE | | igh- | Walter Stadler, Michael Carducci, Robert Coomes, Robert S DiPaola | users identify potential words of interest. | | | | | |
| → Class | 1.0 |)2 | × | Clinical Trial ID: NCT00326898 PMID: 26969090 | | | | | | |
| □ All cause AEs (all grades) | Lowe | er Cl | itici | Abetre et | | | | | | |
| $35\ 2$ All cause AEs (grade \geq 3) | 0.8 | 35 | × | ADSTRACT | | | | | | |
| Disease free survival | | x Cl | rsus | Renal-cell carcinoma is highly vascular, and proliferates primarily through dysregulation of the vascular endothelial grow | vth factor (VEGF) pathway. We tested sunitinib | | | | | |
| Overall survival | | 23 | * | and sorafenib, two oral anti-angiogenic agents that are effective in advanced renal-cell carcinoma, in patients with resec | ted local disease at high risk for recurrence. In | | | | | |
| 537 3 Treatment related AEs (all grades) | | | asa | this double-blind, placebo-controlled, randomised, phase 3 trial, we enrolled patients at 226 study centres in the USA an | d Canada. Eligible patients had pathological | | | | | |
| □ Treatment related AEs (grade ≥ 3) | 🗹 Deta | il | | stage high-grade T1b or greater with completely resected non-metastatic renal-cell carcinoma and adequate cardiac, ren | al, and hepatic function. Patients were stratified | | | | | |
| 🗁 - default | Survi | ival in Control | | by recurrence risk, histology, Eastern Cooperative Oncology Group (ECOG) performance status, and surgical approach, and | nd computerised double-blind randomisation | | | | | |
| □ All cause AEs (all grades) | 6.6 | 5 | × | was done centrally with permuted blocks. Patients were randomly assigned (1:1:1) to receive 54 weeks of sunitinib 50 mg | g per day orally throughout the first 4 weeks of | | | | | |
| ✓ All cause AEs (grade ≥ 3) | I Ec | | | each 6 week cycle, sorafenib 400 mg twice per day orally throughout each cycle, or placebo. Placebo could be sunitinib p | lacebo given continuously for 4 weeks of every 6 | | | | | |
| ✓ Disease free survival | 287 | 1 | × | week cycle or sorafenib placebo given twice per day throughout the study. The primary objective was to compare diseas | e-free survival between each experimental | | | | | |
| Overall survival | 2 | | | group and placebo in the intention-to-treat population. All treated patients with at least one follow-up assessment were | included in the safe Whon right click on the highlighted | | | | | |
| Treatment related AEs (all grades) | | , | | registered with ClinicalTrials.gov. number NCT00326898. Between April 24, 2006, and Sept 1, 2010, 1943 patients from the | e National Clinical | | | | | |
| \Box Treatment related AEs (grade \geq 3) | 2 | | _ | assigned to sunitinib (n=647) sorafenib (n=649) or placebo (n=647). Following high rates of toxicity-related discontinuat | ion after 1323 natie text, a popup menu will be displayed | | | | | |
| Pairwise MA | | | | discontinued by 193 [44%] of 438 patients on sunitinib 199 [45%] of 441 patients on sorafenib) the starting dose of each | to show available extraction options | | | | | |
| Primary Analysis | | | | n Oct 16, 2014, because of low conditional newer for the primary endpoint, the Starting dose of each | N Data Safety Marie for the highlighted text | | | | | |
| B - default | | A list of pre- | defin | ed extraction tasks | in Data Salety Molin For the right on the ention to get | | | | | |
| All cause AEs (all grades) | 50 | is displayed | n ho | n user decide | osers can click on the option to set | | | | | |
| ✓ All cause AEs (grade ≥ 3) | ľ | | | (1.6-8.2) for sunitinib (nazard ratio [HR] 1.02 , 97.5% CI 0.85-1.23, p=0.8038), 6.1 years (IOP) | the extraction target. | | | | | |
| ✓ Disease free survival | | which task to | o ext | and 6-6 years (IQR 1-5-NE) for placebo. The x Highlighted 1-02 vents | For example, highlight "1.02" and | | | | | |
| ✓ Overall survival | | | | hts on sorafenib), hand-foot syndrome (94 [Result patier | select "TF" will extract the 1 02 for | | | | | |
| Treatment related AFs (all grades) | | | | sunitinib and 95 [15%] patients on sorafenib), and fatigue 110 [18%] pati∈ TE //e de | aths related to trea | | | | | |
| $\Box \text{ Treatment related AEs (grade \geq 3)}$ | | | | days of the end of treatment; one patient receiving sorafenib died from in due to each of neurological segueles, acqueles of patric performing to work CI | ients receiving sun The IE In current extraction task. | | | | | |
| | | | _ | due to each of neurological sequelae, sequelae of gastric perforation, put | ised dosing still res ulted in high concres. | | | | | |

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.

Adjuvant treatment with the VEGF receptor tyrosine kinase inhibitors sor Furthermore, substantial treatment discontinuation occurred because of Customizable Extraction Task Panel use of these drugs for high-risk kidney cancer in the adjuvant setting and National Cancer Institute and ECOG-ACRIN Cancer Research Group, Pfizer Ec Class | Disease free survival 🕑 Fill 2 🗸 default | All cause AEs (grade ≥ 3 Save 🗹 Fill 4 🗸 primary Comp 2 Comp 3 default | All cause AEs (all grad Fill 3 Save reatment Arm Measure Placebo 394 Lower CI 625 Upper Cl Control Arm 625 Event 159 Survival Total 626 159 626 Ec & Et Fc T1 Each extraction task may have different items for extraction for various purpose. As shown in this figure, we select three tasks Ec T2 with different items. Users can input those items manually or use right-click popup Et T2 menu to set them.

efit relative to placebo in a definitive phase 3 study. . These results provide a strong rationale against the Survival in Contro nce might be independent of angiogenesis. US All rights reserved.

| | Integrate | d Intera | ctive PDF Viewer | | By clicking th | By clicking the "PDFs" tab, users can uploa | | | | | | |
|-----|-------------------|----------------|---------------------------|------------------------|----------------|---|-------|--|---|--|--|--|
| es) | Abstract | 🕒 PDFs | | | VIEW PDFS III | | | | | | | |
| | LNMAExt-2021 | 0706-1.0.3-fin | al (1).pdf Choose Files | LNMAl (1).pdf 🚨 Upload | | | | | | | | |
| _ | | | 1 of 1 | — + Automa | atic Zoom 🗸 | | 🛱 🖶 🖆 | | > | | | |

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

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

3

| Extraction Task Def | inition | 6 Ext | 6 Extraction Results | | | | | | Users can select which items to be displayed in this | | | | | | |
|--|---|--|--|----------------------------|------------------|------------|------------|-------------------------|--|-------------------|--------------------|--------------------|--------------|-------------------------------|------------------|
| ■ Renal Cell Carcinoma | Adjuvant Renal Cell Carcinor | ma 🔳 | Renal Cell | Carcinoma 📑 Adjuvant F | Renal Cell Carci | noma | 📝 Extra | ct by outcor | ne | result ta | ible to to | cus on t | ne intere | ested Item | ۱S |
| Save Interactive Table Design | Extract Information for Interactive Table | □ ■ [Intera | active Tal | ole Interactive Table | default Int | eractiv | e Table] | PMID, shor | t #, first au | thor 1 🗙 | Toggle all s | tudies 🕞 | Save 🕼 E | dit this outcom | e |
| Table Columns ≡ Add | Category | | CHARACTE | RISTICS 🗹 Trial Name 🗹 | Full Pub or Abs | tract 🗹 🛛 | Phase 🗹 N | umber of Ar | ms 🗹 Tre | eatment Arn | n 1 🗹 Treatm | ent Arm 2 | Control Ar | m | |
| | | 🗹 Type o | of Agent in T | Treatment Arm 🗹 Treatmen | t Arm | | | | | | | | | | |
| | + Add Attribute | | LATION CH | ARACTERISTICS ON Number | er of Participan | ts [🗆 ITT | Population | n 🗆 Treatme | ent Arm | Control A | rm] 🗆 Plan | ned Treatm | ent Duratior | n (years) | • |
| - Trial Name | Add Sub | | | | | | | | | | | | | | |
| Full Pub or Abstract | + Add Sub | 🗆 Risk Category Included in Trial 🗆 Definition of "highest-risk" group 🗆 % of Patients with Clear-Cell Histology | | | | | | | | | | | | | |
| - Phase | + Add Sub | RESULTS FOR OVERALL POPULATION Name of Primary Endpoint Name of Secondary Endpoints Overall Survival (years) [Rx Ontrol] | | | | | | | | | | | | | |
| Number of Arms | + Add Sub | Diseas | Disease Free Survival (years) [Rx Control] % of All Grade TrAEs [Rx Control] % of Grade 3 or Higher TrAEs [Rx Control] | | | | | | | | | | | | |
| - Treatment Arm 1 | + Add Sub | □% of A | ○ % of All Grade All-Cause AEs [Rx Control] ○ % of Grade 3 or Higher All-Cause AEs [Rx Control] □ Treatment discontinuation at 1 year (%) [Rx Control] | | | | | | | | | | | | |
| - Treatment Arm 2 | + Add Sub | | | | | | | | | | | | | | |
| - Control Arm | + Add Sub | SYS | □ Is study | included in Meta-analysis | | | | | | | | | | | |
| Type of Agent in Treatment | + Add Sub | # | PID | (9 selected) | Checked | Arms | Trial Name | Full Pub or Abstract | Phase | Number of Arms | Treatment Arm 1 | Treatment Arm 2 | Control Arm | Type of Agent in Treatment | Treatment Arm |
| Treatment Arm | + Add Sub | 503 | 34407342 | Toni K Choueiri et al 2021 | ✓ | 2 🗸 | KEYNOTE | Full Publ | 3 | 2 | Pembrolizu | | Placebo | Arm | Pembrolizu |
| | | 530 | 28967554 | Robert I Motzer et al 2017 | | 2 2 | S-TRAC | Full Publ | 3 | 2 | Sunitinib | | Placebo | ткт | Sunitinib |
| - POPULATION CHARACT | + Add Attribute | 531 | 28902533 | Robert J Motzer et al 2017 | | 2 ¥ | PROTECT | Full Publ | 3 | 2 | Pazopanib | | Placebo | TKI | Pazopanib |
| - Number of Participants | + Add Sub | 532 | 33052759 | Tim Eisen et al 2020 | | 2 🗸 | SORCE | Full Publ | 3 | 3 | Sorafenib | Sorafenib | Placebo | TKI | Arm 1: Sor |
| ITT Population | | 533 | 27718781 | Alain Ravaud et al 2016 | | 2 🗸 | S-TRAC | Full Publ | 3 | 2 | Sunitinib | | Placebo | TKI | Sunitinib |
| Treatment Arm | | 534 | 30412222 | M Staehler et al 2018 | - 2 | 2 🗸 | S-TRAC | Full Publ | 3 | 2 | Sunitinib | | Placebo | TKI | Sunitinib |
| - Control Arm | | 535 | 26969090 | Naomi B Haas et al 2016 | | 2 🗸 | ASSURE | Full Publ | 3 | 3 | Sunitinib | Sorafenib | Placebo | TKI | Arm 1: Sun |
| Planned Treatment Dui | + Add Sub | 536 | 33461782 | Robert J Motzer et al 2021 | | 2 ~ | PROTECT | Full Publ | 3 | 2 | Pazopanib | | Placebo | TKI | Pazopanib |
| - Median On-Treatment (| + Add Sub | | | | | 2 🗸 | ATLAS | Full Publ | | | | | 11 | | |
| - Madian Colleve Lin Durr | | The items in | the ext | raction task can | | | | | The | extraction | on results | s of all st | udies ar | e listed ir | ı a table |
| - Median Follow Up Dura | T Add Sub | be defined i | n a tree- | like view | | | | | for u | users to | orowse tł | ne overa | ll status | | |
| Median Age (years) | + Add Sub | | | | | | | | | | | | | | |

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.