Features and functioning of Data Abstraction Assistant, a software application for data abstraction during systematic reviews

Jens Jap, BA [Software Engineer],
Center for Evidence Synthesis in Health, Department of Health Services, Policy, and Practice, Brown School of Public Health

Ian J. Saldanha, MBBS, MPH, PhD [Systematic Review Data Repository (SRDR) Director, Assistant Professor],
Center for Evidence Synthesis in Health, Department of Health Services, Policy, and Practice, Brown School of Public Health, ian_saldanha@brown.edu

Bryant T. Smith, MPH, CPH [Systematic Review Data Repository (SRDR) Administrator],
Center for Evidence Synthesis in Health, Department of Health Services, Policy, and Practice, Brown School of Public Health, bryant_smith@brown.edu

Joseph Lau, MD [Professor Emeritus],
Center for Evidence Synthesis in Health, Department of Health Services, Policy, and Practice, Brown School of Public Health, joseph_lau@brown.edu

Christopher H. Schmid, PhD [Professor], and
Center for Evidence Synthesis in Health, Department of Biostatistics, Brown School of Public Health, christopher_schmid@brown.edu

Tianjing Li, MD, MHS, PhD [Associate Professor]
Center for Clinical Trials and Evidence Synthesis, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, tli19@jhu.edu

Abstract

INTRODUCTION—During systematic reviews, data abstraction is labor- and time-intensive and error-prone. Existing data abstraction systems do not track specific locations and contexts of abstracted information. To address this limitation, we developed a software application, the Data Abstraction Assistant (DAA), and surveyed early users about their experience using DAA.

FEATURES OF DAA—We designed DAA to encompass three essential features: (1) a platform for indicating the source of abstracted information; (2) compatibility with a variety of data abstraction systems; and (3) user-friendliness.

HOW DAA FUNCTIONS—DAA: (1) converts source documents from PDF to HTML format (to enable tracking of source of abstracted information); (2) transmits the HTML to the data

Corresponding Author Jens Jap, BA Software Engineer, Center for Evidence Synthesis in Health, Department of Health Services, Policy, and Practice, Brown School of Public Health, jens_jap@brown.edu.

abstraction system; and (3) displays the HTML in an area adjacent to the data abstraction form in the data abstraction system. The data abstractor can mark locations on the HTML that DAA associates with items on the data abstraction form.

EXPERIENCES OF EARLY USERS OF DAA—When we surveyed 52 early users of DAA, 83% reported that using DAA was either very or somewhat easy; 71% are very or somewhat likely to use DAA in the future; and 87% are very or somewhat likely to recommend that others use DAA in the future.

DISCUSSION—DAA, a user-friendly software for linking abstracted data with their exact source, is likely to be a very useful tool in the toolbox of systematic reviewers. DAA facilitates verification of abstracted data and provides an audit trail that is crucial for reproducible research.systematic reviews; data abstraction; software; data tracking; data exchange

Keywords

systematic reviews; data abstraction; software; data tracking; data exchange

INTRODUCTION

During the conduct of a systematic review, "data abstraction" (or "data extraction") refers to the key process of identifying relevant information about studies included in the review and transferring this information to a data collection form (either verbatim or after interpretation and manipulation). Information about included studies is typically obtained from journal articles and other source documents. Data abstraction is typically performed by trained researchers with varying degrees of content and methodological expertise.

Data abstraction is labor- and time-intensive and often error-prone.[1–7] The lack of an audit trail makes the verification of the abstracted information difficult. Errors made during data abstraction, which often remain undetected by peer reviewers, editors, and readers, can impact the validity of the results of systematic reviews.[1] With a surge in collaborative science and reuse of previously abstracted data, it is likely that decision-making by clinicians, policymakers, patients, and others may be compromised by data abstraction errors.

In recent years, several web-based data abstraction systems, such as the Systematic Review Data Repository (SRDR),[8,9] Covidence®,[10] EPPI-Reviewer®,[11] DistillerSR® (Evidence Partners, Ottawa, Canada), and Doctor Evidence® (www.drevidence.com), have been built to aid the development and population of data abstraction forms for organizing data in an efficient way for subsequent data analysis. These data abstraction systems have a major limitation, however. While they can record which source documents are used for data abstraction, they do not track the specific locations and contexts of relevant pieces of information in these often-lengthy documents. To enable accurate verification of abstracted data during a systematic review, the verifier often has to re-read the entire source document or large swaths of it, a task that can require as much time as abstracting the data in the first place. An ability to track the specific location and context of abstracted data in source documents would likely help to document initial data abstraction and facilitate data

verification and adjudication. This would likely promote the validity of the systematic review findings, save time, and advance the openness of the systematic review enterprise.

In this paper, we describe: (1) the features and functioning of Data Abstraction Assistant (DAA), a free, open-source, open-access software application to facilitate tracking the location of abstracted information in source documents, with the potential to reduce errors and time spent during data abstraction in systematic reviews; and (2) the results from a survey of early users of DAA.

FEATURES OF DAA

We designed DAA to encompass three desired essential features.

(1)   A platform for indicating the source of abstracted information.—The major impetus behind the development of DAA was to create a platform where data abstractors could indicate the source of information by pin-pointing specific locations in source documents, thereby creating a potentially permanent linkage (i.e., tracking) between abstracted information and its source.

(2)   Compatibility with a variety of data abstraction systems.—DAA's main purpose is to contain information that links individual abstracted data items to specific locations in source documents. To make DAA compatible with a variety of data abstraction systems (e.g., SRDR, Covidence®, DistillerSR®), we designed the DAA platform to be *distinct* from the data abstraction system. This distinction is attained by keeping separate the process of linkage with an item on the data abstraction form (in the data abstraction system) and the process of capturing and navigating to the location of information (in the source document). The source code for the implementation of this independence is available at https://bitbucket.org/cebmbrown/daa/src/master/). While distinct from the underlying data abstraction system, DAA is designed to appear to the end-user as part of the web-based data abstraction system. This is done to enable a seamless user experience. While we have developed DAA to be compatible across data abstraction systems, we describe in this paper the test case of DAA's compatibility with SRDR.

(3)   User-friendliness.—To make navigation easy and fast, we have developed DAA to be user-friendly and menu-driven. When abstracting data, the data abstractor visualizes DAA as integrated seamlessly into the data abstraction system.

For a technical description of how DAA achieves each of these desired features, see Table 1.

HOW DAA FUNCTIONS BEHIND THE SCENES

DAA works through three steps (Figure 1):

Step 1: Converting documents from Portable Document Format (PDF) to Hypertext Markup Language (HTML) format.—Most source documents that are used for data abstraction during systematic reviews (e.g., journal articles, conference proceedings, approval packages from regulatory authorities, clinical study reports [CSRs]) are accessed as PDF files. While documents are sometimes obtained in other formats, (e.g., websites or

word-processing documents), our experience has been that most systematic review teams convert those documents into PDF format. Tracking the location of abstracted information from a given document is best achieved through annotating (i.e., highlighting text in a different color, circling, or otherwise marking up) the specific source text, tables, and/or figures in that document.

Annotating PDFs is extremely challenging because of the lack of open-source solutions (i.e., software for which the original source code is made freely available and may be redistributed and modified) for manipulating and editing PDFs. Therefore, we elected to convert source documents from PDF to HTML, a format that has many open-source solutions for editing and location tagging. For converting PDFs to HTML format, we use pdf2htmlEX (an open-source tool available at https://github.com/coolwanglu/pdf2htmlEX). Another advantage of converting documents from PDF to HTML format is that distinct versions of a given document may originate from different publishers or platforms. Such distinctions arise because publishers or platforms might render PDFs of the same document using different tools (e.g., Adobe Reader®, Preview®), resulting in PDFs of inconsistent format. Inconsistencies in format are largely eliminated when DAA converts source documents into HTML format because during the conversion process, images and plain text are separated, creating a more consistent representation of the documents. The HTML format offers consistency because text and images are clearly marked as such, while, in PDF format, whether an item is text or an image is sometimes ambiguous. In some cases, PDFs might originate as scanned versions of paper copies of documents, which, instead of being rendered as text-based PDFs, are rendered as image-based PDFs. Image-based PDFs are particularly challenging to work with because not all imaging processing software can read text from images accurately. This problem is compounded when the scanned version of a printed document is blurry and/or of otherwise poor quality. DAA currently does not function with PDFs that are image-based instead of text-based.

Step 2: Transmitting the HTML version of the source document to the data abstraction system.—Once the source document is converted from PDF to HTML format, DAA uses encrypted communication to transmit the HTML to the data abstraction system (i.e., SRDR in this instance). We authorize communication between DAA and SRDR through the use of "security tokens", i.e., unique hash numbers that are required for access (see section "*How DAA works at the data abstractor end*"). Security tokens ensure that only authorized users have access to view and edit markers to relevant source documents.

Step 3: Displaying and allowing for annotating the HTML version of the source document in the data abstraction system.—Once DAA transmits the HTML version of the source document to SRDR, SRDR displays the document in HTML format on the screen adjacent to the data abstraction form ("Split Screen View", see online demonstration video at: https://goo.gl/ZhAkq4). To the data abstractor, the HTML format appears exactly like the PDF version of the source document. Using a mouse, the data abstractor can then drag a "flag" from any item on the data abstraction form to any desired location on the adjacent HTML (Figure 2). Upon doing so, DAA creates a link between the specific item on the data abstraction form and the selected location on the HTML. DAA allows for the

linkage of a given item on the data abstraction form with multiple locations on the HTML document, and also the linkage of a given location on the HTML document with multiple items on the data abstraction form. Linkages between abstracted information and corresponding locations on the HTML document are saved as markers on the DAA server, so that the next time that particular source document is requested, the data abstractor may view the HTML document and all previously-placed markers together. As a result, a running record of all markers added to any source document in the project is retained for future reference and access. By clicking on existing markers (for example, during data verification), DAA scrolls the screen to navigate to the exact location of the source text, with the pertinent text highlighted.

Because there may be multiple source documents for a given study, DAA allows the data abstractor to "toggle" between multiple HTML files when abstracting data in SRDR. When using this feature, the data abstractor can create links between a given item on the data abstraction form and locations on two or more separate HTML files. A single item that is linked to multiple locations will have one flag representing each of these linkages. This individual flag carries both the identifier of the HTML file and the exact linked location in the HTML file. Having both of these pieces of information allows DAA to switch to the document when clicking on the existing marker and scroll automatically to the location as described above, even if it is on a separate HTML file.

HOW DAA FUNCTIONS AT THE DATA ABSTRACTOR END

DAA is designed to assist with data abstraction, which is a step that is carried out after the set of eligible studies for the systematic review is identified. Data abstractors can interface directly with DAA by logging into the password-protected DAA web application and uploading study documents (as PDFs). This uploading process can be centrally managed by the project lead if a more protected governance of the data abstraction and document management process is desired.

Once source documents are uploaded in PDF format, they are converted into HTML format and organized into "Document Stores", which are groups or collections of source documents. DAA assigns each Document Store a security token, allowing access to the HTML files from any systematic review project that the data abstractor is working on in SRDR. Upon logging into SRDR, SRDR requires the data abstractor to provide the security token in order to access the data abstractor's Document Stores. After the data abstractor selects the Document Store and, subsequently, a source document in HTML format, DAA transmits the HTML file to SRDR. Once DAA transmits the HTML to SRDR, SRDR displays the HTML in an area adjacent to the abstraction form (see section "*Step 3: Displaying and annotating the HTML version of the document in SRDR*").

EXPERIENCES OF EARLY USERS OF DAA

To evaluate rigorously the benefit of using DAA, we conducted an online randomized trial ("DAA Trial") comparing the accuracy (i.e., error rates) and efficiency of data abstraction (i.e., time taken) using DAA versus verification and adjudication approaches that do not use

DAA. As part of the DAA trial, data abstractors abstracted various kinds of information pertaining to published studies, including study design, risk of bias, characteristics of study participants, treatment arms, outcomes, within-arm results, and between-arm results. We previously reported the detailed protocol for this trial,[2] and will report its results separately. In this section, we describe the results of surveying the trial participants, i.e., the early users of DAA, with regards to the user-friendliness of DAA.

Methods of survey

We surveyed all 52 individual data abstractors enrolled in the DAA Trial. As part of the trial, we organized these individuals into 26 pairs, each pair comprising one less experienced and one more experienced data abstractor. During the DAA Trial, each pair abstracted data from six studies; for two of the six studies, the pair used DAA to abstract data (two other abstraction approaches without DAA were used for the other four studies that the pair abstracted). For both studies assigned to a given pair for abstraction using DAA, the junior data abstractor first abstracted data, and then the senior data abstractor verified the data, making changes to the abstraction as needed. Therefore, it was always the junior data abstractor who placed flags and the senior data abstractor who could view or remove existing flags or add new flags.

After completing data abstraction for the DAA Trial, we asked each data abstractor to a complete a brief survey designed using Qualtrics®. We asked questions pertaining to the data abstractor's self-reported ease with which each of the following tasks could be completed: (1) opening source documents in Split Screen view in SRDR; (2) scrolling between pages of a source document; (3) placing flags on a source document; and (4) clicking on existing flags to automatically navigate to the relevant location on the source document. We also asked data abstractors to assess the overall ease of using DAA, and to indicate the DAA feature that they liked the most. Finally, we asked data abstractors about their likelihood of using DAA in the future and of recommending that others use it in the future. The Appendix includes the entire survey instrument.

Results of survey

All 52 data abstractors who participated in the DAA Trial completed the survey (Table 2). Most data abstractors (43/52, 83%) found using DAA to be either very or somewhat easy overall. Opening source documents in Split Screen view and scrolling between pages of a source document were reported to be easy by 83% and 69% of data abstractors, respectively. Among those who placed flags initially, i.e., less experienced data abstractors, 62% agreed that doing so was easy. Among those who clicked on existing flags, i.e., more experienced abstractors, 73% agreed that doing so was easy.

When asked about use of DAA for data abstraction in the future, 65% of less experienced and 77% of more experienced data abstractors stated they are very or somewhat likely to use it. Similarly, 80% of less experienced and 93% of more experienced data abstractors stated that they are very or somewhat likely to recommend that others use it (see Table 2 for detailed breakdown of responses). When asked to name their favorite DAA feature, 54% of all data abstractors chose the ability to click on existing flags marking information sources

(73% of senior data abstractors named this feature); 19% of data abstractors chose the ability to open a document in Split Screen view; and 17% chose the ability to place flags on the PDF (23% of junior data abstractors named this feature) (Table 2).

DISCUSSION

In this paper, we described the features and functioning of DAA, a software application that assists the data abstraction process during systematic reviews by enabling viewing the source document (e.g., journal article) adjacent to the data abstraction form in the data abstraction system and enabling the tracking of the source of abstracted data. When we surveyed 52 early users of DAA, most found the software user-friendly, most would use it, and most would recommend that others use it for data abstraction in the future. The most popular feature of DAA appears to be the ability to click on existing flags to navigate to portions of text/figures/tables in the source document that contain relevant data, a feature that could be very useful when verifying abstracted data and when updating systematic reviews.

Potential utility of DAA for systematic reviews

DAA is likely to be a very useful tool in the toolbox of systematic reviewers. Systematic reviews take a median of 66 weeks from registration to publication (interquartile range 42 weeks, range 6 to 186 weeks).[12] With an ever-growing size of the body of relevant evidence in most topic areas, this duration is likely to get even longer. Data abstraction accounts for a large share of the time spent conducting systematic reviews, and tools such as DAA have the potential to reduce that time. The utility of DAA would likely be further enhanced if systematic reviewers choose to share their annotations publicly, thus allowing future systematic reviewers to capitalize on existing annotations in new systematic reviews. Similarly, in the case of review updates, access to existing annotations and exact data source location could greatly reduce the time spent on data abstraction.

DAA in the context of automated tools for systematic reviews

DAA can contribute to evaluating the performance of various automated or semi-automated tools that facilitate data abstraction during systematic reviews. These tools use natural language processing and machine learning approaches to assist with data abstraction. Most existing tools focus on automating the abstraction of data elements such as number of participants, their age, sex, country, recruiting centers, intervention groups, and outcomes.[13] A few tools are able to abstract information about study objective and certain aspects of study design (e.g., study duration, participant flow) and risk of bias.[14,15] However, to date, most of the data elements that are typically abstracted during systematic reviews have not been explored for automated abstraction. Before automated tools for text identification and highlighting can achieve their goals, their performance should be evaluated using a common dataset. The markers placed by DAA can facilitate this evaluation and provide lessons about how these tools can fit into existing systematic review workflows. Further, the manually annotated data collected by the tool could be used as training data for supervised machine learning approaches. Even in a future where the process of identification and highlighting of relevant locations of data elements in source documents is satisfactorily automated by these other tools, the features that DAA offers will be a much-needed complement by allowing

manual tracking and checking of data elements, entry of the data elements into a data abstraction system, and creation of a permanent linkage between abstracted data and their sources.

DAA can be particularly useful when extracting data from trial reports not traditionally used during systematic reviews, for example, CSRs and regulatory documents. A CSR contains an unabridged and comprehensive description of the clinical problem, design, conduct, and results of a clinical trial, following structure and content guidance prescribed by the International Conference on Harmonisation (ICH).[16] To obtain marketing approval of drugs or biologics for a specific indication, pharmaceutical companies submit CSRs and other required materials to regulatory authorities. CSRs differ from trial datasets (i.e., electronic individual patient data) in that they are paper (or mostly PDF) documents, and can be thousands of pages long. CSRs typically contain a wealth of information for evaluating the efficacy and safety of pharmacological treatments, including information that is often missing from the public domain.[17,18] Regulatory documents are summaries of CSRs and related files, prepared by the regulatory agency's staff as part of the process of approving the products. Regulatory documents are usually made available to the public in PDF format. Abstracting information and verifying and reconciling abstracted information from CSRs and regulatory documents can be particularly laborious;[17] DAA can greatly assist these processes by enabling tracking of the source of the information.

Current limitations of DAA

We expect to release DAA for use by the general public by September 2018. We are continuing to develop and refine DAA to address its current limitations and the feedback provided by the early users. For example, currently, the smallest unit of text that can be highlighted as source material for a given data item is an entire line in a paragraph in the source document (Figure 1). We are updating this feature to allow for the highlighting of more granular amounts of text, such as single words or partial words, by implementing a technique of creating ranges or groupings of characters. We do this by defining the beginning and the end of a character range and the character range's location in the source document. This improvement will also help overcome another limitation – that DAA does not currently allow for non-contiguous sections of text to be highlighted together. Once the ability to create ranges of text is in place, we can also group together these ranges of characters, allowing the grouping of non-contiguous sections of text. Another limitation is that DAA currently does not allow the highlighting of text that is in image-based tables and figures. DAA can, however, highlight text in tables and figures that are in text format. We plan on overcoming these limitations by incorporating into DAA general purpose annotation tools, such as Annotator.js, that have a more powerful set of annotation features. Finally, DAA is unable to read text from PDFs that are scanned documents of poor quality. In these instances, conversion of the scanned PDF to HTML format results in one large image (as opposed to text), which cannot be read or annotated. Addressing this challenge will likely require the use of commercially-available software packages that convert images to text.

In summary, we described the features and functioning of Data Abstraction Assistant (DAA), a software application to facilitate tracking of the location of abstracted information

in source documents, with the potential to reduce errors and time spent during data abstraction in systematic reviews. When we surveyed 52 early users of DAA, 83% stated that they found using DAA to be either very or somewhat easy; 71% stated they are very or somewhat likely to use DAA in the future; and 87% stated that they are very or somewhat likely to recommend that others use DAA in the future.

ACKNOWLEDGEMENTS

APPENDIX:: Survey instrument

Data Abstraction Assistant (DAA) Trial Exit Survey for DAA Trial Participants (Data Abstractors)

Purposes of Survey

This is a paper copy of a survey that is being administered online using Qualtrics®. The survey will be administered to each individual participant (data abstractor) in the DAA Trial. The purposes of the survey are to:

1. Obtain the opinion of the Data Abstraction Assistant (DAA) Trial Participants regarding their use of the DAA software; and

2. Obtain suggestions from the Participants for ways in which the software can be improved.

Instructions for Survey Respondents

• Please answer all 10 questions in this survey.

• We estimate that completing this survey will you no more than 3 minutes.

• If you have any questions during the completion of this survey, please email Ian Saldanha at isaldan1@jhmi.edu.

Q1. What is your **FIRST** name?

Your response will be kept confidential. Please note that we only share data from this survey in aggregate.

☐

Q2. What is your **LAST** name?

Your response will be kept confidential. Please note that we only share data from this survey in aggregate.

☐

Q3. Have you EVER worked with any of the following systematic review tools **BEFORE** the DAA Trial? (Select all that apply)

[ ]1	Abstrackr	
[ ]2	Cochrane Author Support Tool (CAST)	
[ ]3	Covidence	
[ ]4	Distiller SR	
[ ]5	DOC™ Data	
[ ]6	Early Review Organizing Software (EROS)	
[ ]7	EPPI-Reviewer	
[ ]8	OpenMeta[Analyst]	
[ ]9	Rayyan	
[ ]10	Review Manager (RevMan)	
[ ]11	System for the Unified Management, Assessment, and Review of Information (SUMARI)	
[ ]12	Other (please specify) - _____13	
[ ]14	I used some tool, but don't remember which one	
[ ]15	None of the above	

Q4. Do you agree or disagree with each of the following statements?

	Agree	Neutral	Disagree
a. It was easy to open documents in Split Screen view in SRDR.	( )1	( )2	( )3
b. It was easy to scroll between pages on the document.	( )1	( )2	( )3
c. It was easy to click on existing flags to automatically navigate to a relevant location on the document.	( )1	( )2	( )3

Q5. Do you agree or disagree with the following statement?

	Agree	Neutral	Disagree	Not applicable (I was a second data abstractor and did not place flags myself.)
a. It was easy to place flags on the document.	( )1	( )2	( )3	

Q6. How would you characterize the **OVERALL EASE** of using DAA? (Select one)

( )1	Very easy
( )2	Somewhat easy
( )3	Neutral
( )4	Somewhat difficult
( )5	Very difficult
( )6	No opinion/don't know

Q7. If DAA is available at no cost the next time you conduct a systematic review, how likely would you be to **USE** DAA for data abstraction? (Select one)

( )$_1$ Very likely

( )$_2$ Somewhat likely

( )$_3$ Neutral

( )$_4$ Somewhat unlikely. If checked, please specify why: _____5

( )$_6$ Very unlikely. If checked, please specify why: _____7

( )$_8$ No opinion/don't know

Q8. If DAA is available at no cost, how likely would you be to **RECOMMEND** that others conducting systematic reviews use DAA for data abstraction? (Select one)

( )$_1$ Very likely

( )$_2$ Somewhat likely

( )$_3$ Neutral

( )$_4$ Somewhat unlikely

( )$_5$ Very unlikely

( )$_6$ No opinion/don't know

Q9. Even if you did not perform each of the following tasks yourself, overall what ability did you like the **MOST** about the Data Abstraction Assistant (DAA)? (Select one)

( )$_1$ The ability to open a document in Split Screen view in SRDR

( )$_2$ The ability to place flags on the document

( )$_3$ The ability to click on existing flags to automatically navigate to a relevant location on the document

( )$_4$ The ability to copy text from the document to the data abstraction form

( )$_5$ Other, (please specify):_____6

( )$_7$ None, i.e., no particular ability stood out

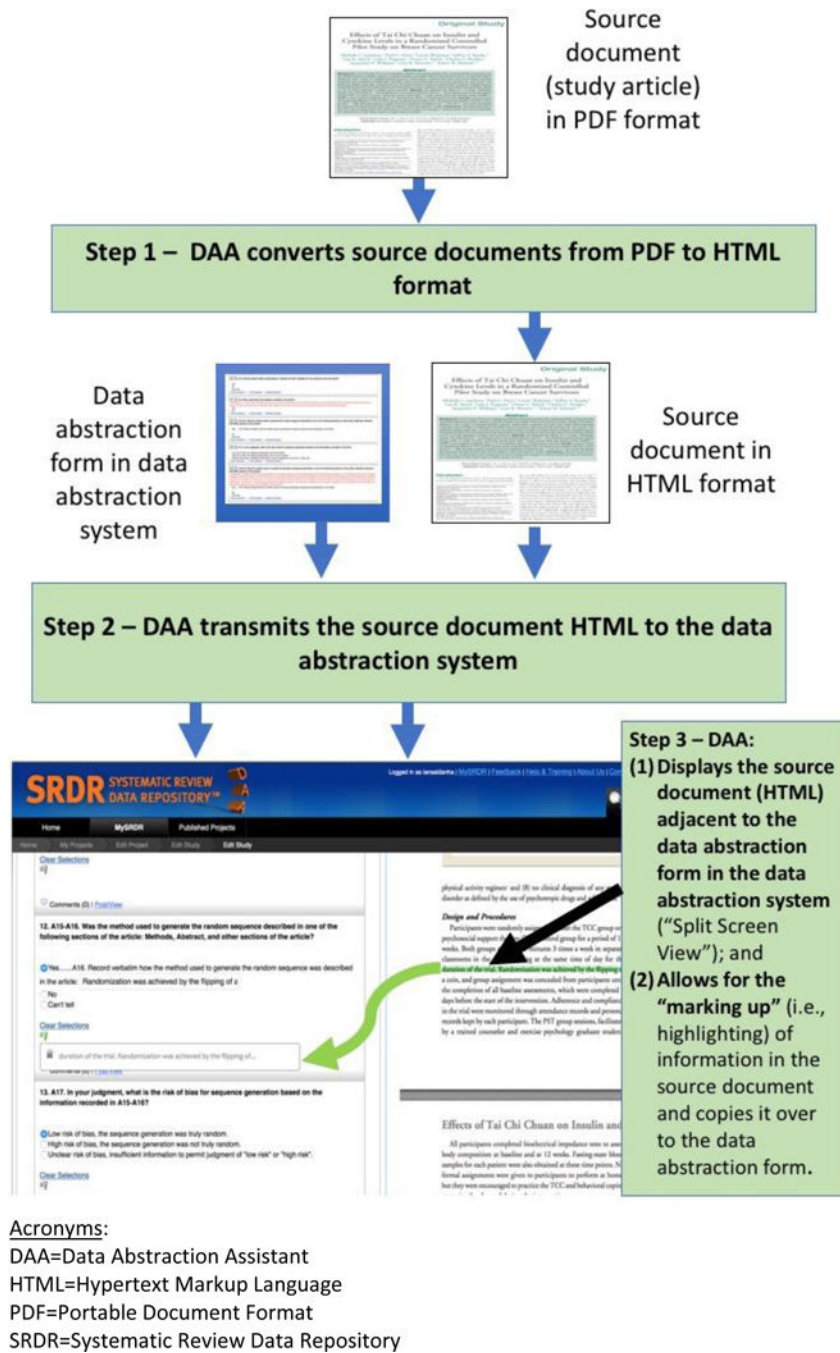Q10. Do you have any suggestions that might help improve the DAA software? (Select one)

( )$_1$ No

( )$_2$ Yes, please specify: _____5

ADMINSTRATIVE DETAILS
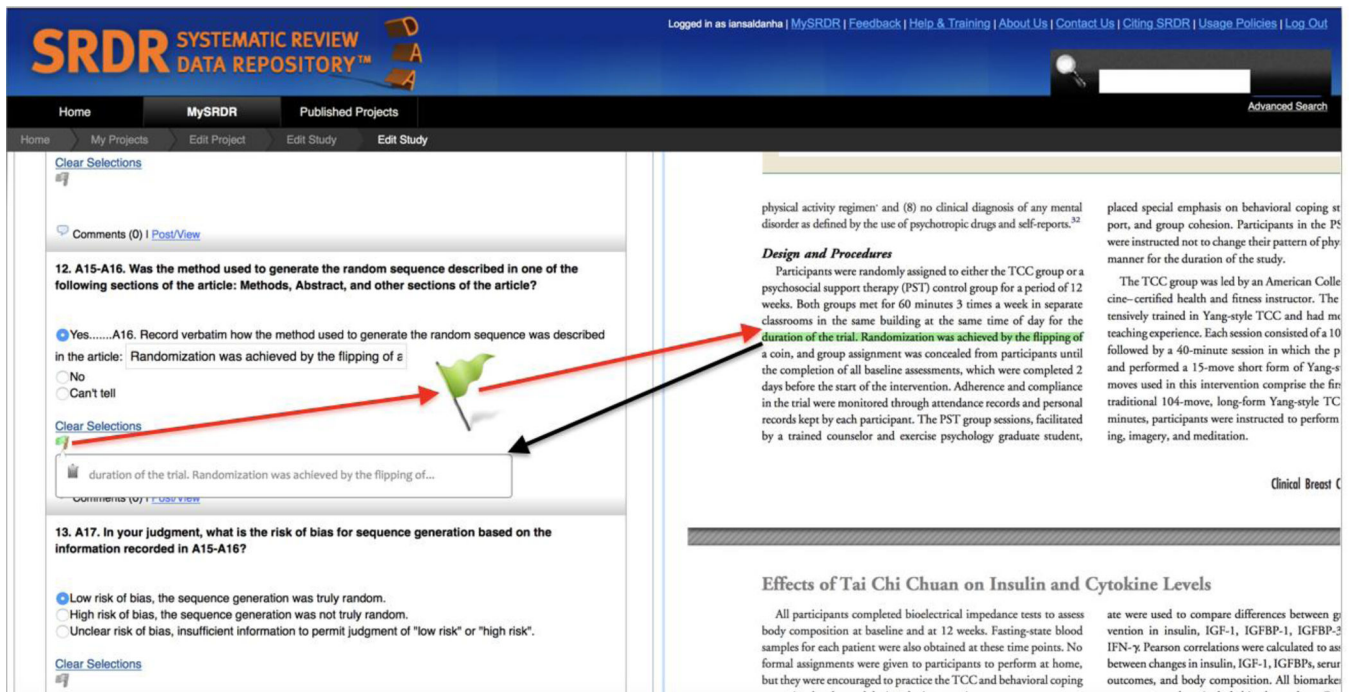
Date survey completed (MM/DD/YYYY): __ __ / __ __ / __ __ __ __ *(Auto-filled by Qualitrics®)*

REFERENCES

1. Mathes T, Klaßen P, Pieper D. Frequency of data extraction errors and methods to increase data extraction quality: a methodological review. BMC Med Res Methodol 2017;17(1):152. [PubMed: 29179685]

2. Saldanha IJ, Schmid CH, Lau J, et al. Evaluating Data Abstraction Assistant, a novel software application for data abstraction during systematic reviews: Protocol for a randomized controlled trial. Syst Rev 2016;5(1):196. [PubMed: 27876082]

3. Carroll C, Scope A, Kaltenthaler E. A case study of binary outcome data extraction across three systematic reviews of hip arthroplasty: errors and differences of selection. BMC research notes 2013;6:539. [PubMed: 24344873]

4. Horton J, Vandermeer B, Hartling L, Tjosvold L, Klassen TP, Buscemi N. Systematic review data extraction: cross-sectional study showed that experience did not increase accuracy. J Clin Epidemiol 2010;63(3):289–98. [PubMed: 19683413]

5. Gøtzsche PC, Hróbjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses that use standardized mean differences. JAMA 2007; 298(4):430–7. [PubMed: 17652297]

6. Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. J Clin Epidemiol 2006;59(7):697–703. [PubMed: 16765272]

7. Jones AP, Remmington T, Williamson PR, Ashby D, Smyth RL. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. J Clin Epidemiol 2005;58(7):741–2. [PubMed: 15939227]

8. Li T, Vedula SS, Hadar N, Parkin C, Lau J, Dickersin K. Innovations in data collection, management, and archiving for systematic reviews. Ann Intern Med 2015;162(4):287– 94. [PubMed: 25686168]

9. Ip S, Hadar N, Keefe S, et al. A Web-based archive of systematic review data. Syst Rev 2012;1:15. doi: 10.1186/2046-4053-1-15. [PubMed: 22588052]

10. Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia Available at www.covidence.org. Last accessed June 18, 2018.

11. Thomas J, Brunton J, Graziosi S (2010) EPPI-Reviewer 4.0: software for research synthesis. EPPI-Centre Software London: Social Science Research Unit, Institute of Education, University of London Available at https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2967. Last accessed June 18, 2018.

12. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7(2):e012545. doi: 10.1136/bmjopen-2016-012545.

13. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Syst Rev 2015; 4:78. [PubMed: 26073888]

14. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc 2016; 23: 193–201. [PubMed: 26104742]

15. Millard LAC, Flach PA, Higgins JPT. Machine learning to assist risk-of-bias assessments in systematic reviews. Int J Epidemiol 2016; 45:266–277. [PubMed: 26659355]

16. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Structure and Content of Clinical Study Reports E3 1995 Available at www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf. Last accessed June 18, 2018.

17. Mayo-Wilson E, Li T, Fusco N, Dickersin K; MUDS investigators. Practical guidance for using multiple data sources in systematic reviews and meta-analyses (with examples from the MUDS study). Res Synth Methods 2018 3;9(1):2–12. [PubMed: 29057573]

18. Doshi P, Jefferson T. Clinical study reports of randomised controlled trials: an exploratory review of previously confidential industry reports. BMJ Open 2013;3(2). pii: e002496. doi: 10.1136/bmjopen-2012-002496.

Acronyms:
DAA=Data Abstraction Assistant
HTML=Hypertext Markup Language
PDF=Portable Document Format
SRDR=Systematic Review Data Repository

Figure 1:
Pictorial representation of DAA's functioning

Figure 2:
Screenshot of how DAA displays the source document in HTML format (*right*) adjacent to the data abstraction form in the data abstraction system (SRDR, *left*). DAA allows for (1) placing of "flags" which links the data item to a specific location on the source document (red arrows); and (2) automatically records the content of the highlighted text in the data abstraction system (black arrow).

Table 1:

Technical description of how Data Abstraction Assistant (DAA) achieves our three desired essential features

#	Desired essential feature	Technical description
1	*A platform for indicating the source of abstracted information*	We built a RESTful application programming interface (API) server, which exposes end-points and returns the following information (HTTP Verb): • List of documents with document titles and unique identifiers (GET) • HTML of document (GET) • List of markers (GET) • Add a new marker to a document (POST) • Remove a markexr to a document (DELETE)

2	*Compatibility with a variety of data abstraction systems*	For DAA to work with a given data abstraction system, the data abstraction system must: (1) have sufficient screen real-restate to load the HTML document into view; and (2) have a mechanism that lets the data abstractor isolate a section of the HTML document and send a web request to the DAA server with the following three pieces of information:
		  1. The text to highlight (STRING);
		  2. The position of the text identified by the unique class identifier of the text (STRING); and
		  3. The identification number of the document (NUMERIC).
		DAA will receive this information and save it in its database. The data abstraction system's user interface then needs to reload the document or partially update it to reflect the flag placement by fetching the information from DAA. DAA also makes available a list of existing flags. The integrated system must have a mechanism to display this information as well, so that the data abstractor can choose from the list and select a marker to display.

3	*User-friendliness of DAA*	We built DAA's interface to have a simple to navigate, menu-driven design.The data abstractor receives immediate feedback upon each step and menus are updated in real time as the data abstractor makes selections. The update to the selection menu is done using Google's AngularJS components. Whenever the data abstractor makes a selection, a JavaScript on-update listener triggers an API call to the DAA server and updates each component's content.

Table 2.

Survey responses, by level of experience with data abstraction

Survey item	Less experienced data abstractors (N=26) n (%)		More experienced data abstractors (N=26) n (%)		All data abstractors (N=52) n (%)	
Ease of use of DAA						
It was easy to open documents in Split Screen view in SRDR.						
Agree	20	(77)	3	(88)	43	(83)
Disagree	2	(8)	0	(0)	7	(4)
Neutral	4	(15)	3	(12)	2	(13)
It was easy to scroll between pages on the document.						
Agree	17	(65)	19	(73)	36	(69)
Disagree	3	(12)	3	(12)	6	(12)
Neutral	6	(23)	4	(15)	10	(19)
It was easy to place flags on the document.						
Agree	16	(62)		N/A		N/A
Disagree	3	(12)				
Neutral	7	(27)				
It was easy to click on existing flags to automatically navigate to a relevant location on the document.						
Agree	16	(61)	21	(85)	37	(71)
Disagree	2	(8)	1	(4)	3	(3)
Neutral	8	(31)	4	(15)	12	(23)
Overall ease of using DAA						
Very easy	5	(19)	9	(35)	14	(27)
Somewhat easy	17	(65)	12	(46)	29	(56)
Somewhat difficult	2	(8)	3	(12)	5	(10)
Very difficult	0	(0)	0	(0)	0	(0)
Neutral	2	(8)	2	(8)	4	(8)
Future use of DAA						
Likelihood of using DAA for data abstraction in the future.						
Very likely	10	(38)	12	(46)	22	(42)
Somewhat likely	7	(27)	8	(31)	15	(29)
Somewhat unlikely	2	(8)	4	(15)	6	(12)
Very unlikely	1	(4)	0	(0)	1	(2
Neutral	6	(23)	2	(8)	8	(15)
Likelihood of recommending that others use DAA for data abstraction in the future.						
Very likely	11	(42)	15	(58)	26	(50)
Somewhat likely	10	(38)	9	(35)	19	(37)

Survey item	Less experienced data abstractors (N=26) n (%)		More experienced data abstractors (N=26) n (%)		All data abstractors (N=52) n (%)	
Somewhat unlikely	0	(0)	1	(4)	1	(2)
Very unlikely	2	(8)	0	(0)	2	(4)
Neutral	3	(12)	1	(4)	4	(8)
Favorite DAA feature (even if abstractors themselves did not use the feature)						
Ability to open a document in Split Screen View	7	(27)	3	(12)	10	(19)
Ability to place flags on the document	6	(23)	3	12	9	(17)
Ability to click on existing flags to navigate	9	(35)	19	(73)	28	(54)
Ability to copy text to SRDR	3	(12)	0	(0)	3	(6)
Other	0	(0)	1	(4)	1	(2)
None (i.e., no particular ability stood out)	1	(4)	0	(0)	1	(2)

Note: Percentages are calculated using the column totals as the denominator.