Journal of Clinical Epidemiology 115 (2019) 77–89

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

# A randomized trial provided new evidence on the accuracy and efficiency of traditional vs. electronically annotated abstraction approaches in systematic reviews

Tianjing Li[a,*], Ian J. Saldanha[b,c], Jens Jap[d], Bryant T. Smith[d], Joseph Canner[e], Susan M. Hutfless[f], Vernal Branch[g], Simona Carini[h], Wiley Chan[i], Berry de Bruijn[j], Byron C. Wallace[k], Sandra A. Walsh[l], Elizabeth J. Whamond[m], M. Hassan Murad[n], Ida Sim[o], Jesse A. Berlin[p], Joseph Lau[d], Kay Dickersin[q], Christopher H. Schmid[r]

[a]*Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, E6140, Baltimore, MD 21205, USA*
[b]*Center for Evidence Synthesis in Health, Department of Health Services, Policy, and Practice, Brown University School of Public Health, 121 South Main Street, Box G-S121-8, Providence, RI 02912, USA*
[c]*Department of Epidemiology (Secondary), Brown University School of Public Health, 121 South Main Street, Box G-S121-8, Providence, RI 02912, USA*
[d]*Department of Health Services, Policy, and Practice, Center for Evidence Synthesis in Health, Brown University School of Public Health, 121 South Main Street, Box G-S121-8, Providence, RI 02912, USA*
[e]*Department of Surgery, Center for Outcomes Research, Johns Hopkins School of Medicine, 720 Rutland Avenue, Turner 74, Baltimore, MD 21205, USA*
[f]*Department of Surgery, Johns Hopkins School of Medicine, 600 N Wolfe St, Blalock 406, Baltimore, MD 21205, USA*
[g]*119 Landings Drive, Suite 403, Mooresville, NC 28117, USA*
[h]*Department of Medicine, School of Medicine, University of California San Francisco, 3333 California Street, San Francisco, CA 94143, USA*
[i]*Internal Medicine, Kaiser Permanente Northwest, 500 NE Multnomah Street, Suite 100, KPB-2, Portland, OR 97232, USA*
[j]*Internal Medicine, National Research Council Canada, Digital Technologies Research Centre, 1200 Montreal Road, Ottawa, Ontario K1A 0R6, Canada*
[k]*Internal Medicine, College of Computer and Information Science, Northeastern University, 440 Huntington Avenue, 476 West Village H, Boston, MA 02115, USA*
[l]*Internal Medicine, California Breast Cancer Organizations, 900 Kent Drive, Davis, CA 95616, USA*
[m]*Internal Medicine, Cochrane Consumer Network, 1421 Woodstock Road, Fredericton, New Brunswick E3C 1P2, Canada*
[n]*Internal Medicine, Evidence-based Practice Center, Mayo Clinic, 200 1st Street, Rochester, MN 55905, USA*
[o]*Department of Medicine, University of California San Francisco School of Medicine, San Francisco, CA, USA*
[p]*Internal Medicine, Epidemiology, Johnson & Johnson, 1125 Trenton-Harbourton Road, K3-Zone 1, Titusville, NJ 08560, USA*
[q]*Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, E6152, Baltimore, MD 21205, USA*
[r]*Department of Biostatistics, Brown University School of Public Health, 121 South Main Street, Box G-S121-7, Providence, RI 02912, USA*

Accepted 9 July 2019; Published online 11 July 2019

## Abstract

**Objectives:** Data Abstraction Assistant (DAA) is a software for linking items abstracted into a data collection form for a systematic review to their locations in a study report. We conducted a randomized cross-over trial that compared DAA-facilitated single-data abstraction plus verification (‘‘DAA verification’’), single data abstraction plus verification (‘‘regular verification’’), and independent dual data abstraction plus adjudication (‘‘independent abstraction’’).

**Study Design and Setting:** This study is an online randomized cross-over trial with 26 pairs of data abstractors. Each pair abstracted data from six articles, two per approach. Outcomes were the proportion of errors and time taken.

**Results:** Overall proportion of errors was 17% for DAA verification, 16% for regular verification, and 15% for independent abstraction. DAA verification was associated with higher odds of errors when compared with regular verification (adjusted odds ratio [OR] = 1.08; 95%

Conflict of interest statement: The authors declare that they designed the Data Abstraction Assistant (DAA), the software tool being evaluated in this article. However, the authors do not stand to gain, financially or otherwise, from the direction of the specific results that are presented in this article. The development and evaluation of DAA was funded through the Patient-Centered Outcomes Research Institute (contract number: ME-1310-07009; principal investigator: T.L.). The development and support of the SRDR, the data abstraction system within which the DAA was conducted, has been funded by the Agency for Healthcare Research and Quality (current Contract Number: HHSA290201500002I; principal investigator: I.S.).

Trial registration: This trial was registered with the National Information Center on Health Services Research and Health Care Technology (NICHSR) on November 9, 2015. The registration number is HSRP20152269, and can be accessed at https://wwwcf.nlm.nih.gov/hsr_project/view_hsrproj_record.cfm?NLMUNIQUE_ID=20152269&SEARCH_FOR=Tianjing%20Li (last accessed March 18, 2019).

* Corresponding author. Tel.: +1 410-502-4630.
*E-mail address:* tli19@jhu.edu (T. Li).

https://doi.org/10.1016/j.jclinepi.2019.07.005

confidence interval [CI]: 0.99−1.17) or independent abstraction (adjusted OR = 1.12; 95% CI: 1.03−1.22). For each article, DAA verification took 20 minutes (95% CI: 1−40) longer than regular verification, but 46 minutes (95% CI: 26 to 66) shorter than independent abstraction.

**Conclusion:** Independent abstraction may only be necessary for complex data items. DAA provides an audit trail that is crucial for reproducible research. 

*Keywords:* Data abstraction; Software application; Accuracy; Efficiency; Systematic review; Randomized cross-over trial

## 1. Introduction

Science needs to be robust, transparent, and reproducible. Systematic reviews (''reviews''), the pillar supporting clinical and health policy recommendations, must share these qualities [1]. The validity of a review's findings depends on collecting accurate and complete data from reports of the included studies (i.e., articles), a process known as *data abstraction* (or *data extraction*) [2].

As a predominantly manual process, data abstraction is labor intensive and error prone. Errors occur when abstractors either omit or incorrectly abstract information about a study [3−6]. To minimize errors, data abstraction is usually conducted independently by two abstractors (''independent dual abstraction'') with disagreements discussed between the two abstractors or by involving a third person (''adjudication''). Other approaches include a single abstractor abstracting with the second ''verifying'' what the first has abstracted or single abstraction without verification. Researchers have reported error proportions of 15% for independent dual abstraction plus adjudication, 18% for single abstraction plus verification, and 30% for single abstraction without verification [3].

Because ''*so little is known about how best to optimize accuracy and efficiency*,'' [1] major organizations have made different recommendations for approaches to data abstraction [1,2,6−8]. Cochrane recommended, ''*use (at least) two people working independently to extract study characteristics from reports of each study*,'' with a caveat that independent dual abstraction may be less important for study characteristics than for outcome data [2]. Because only one study has examined error proportions for the various methods of data abstraction [3], the Institute of Medicine (IOM) stopped short of recommending independent dual abstraction for all data elements, instead recommending, ''*at minimum, use two or more researchers, working independently, to extract quantitative and other critical data from each study*'' [1]. This recommendation includes an implicit acknowledgment that independent dual-data abstraction and adjudication is resource intensive; although it is a careful approach, little data exist to support it. An important gap in our current methodological understanding of how best to perform data abstraction remains.

We are aware of several Web-based systems that aid the process of data abstraction (e.g., the Systematic Review Data Repository [SRDR] [9,10], Covidence [11], EPPI-Reviewer, DistillerSR [Evidence Partners, Ottawa, Canada], and Doctor Evidence). Although some systems record which source documents (e.g., journal article and trial registration information) are used for abstraction, none track the specific locations and context of relevant pieces of information in these often-lengthy documents. To facilitate data verification and adjudication, we developed Data Abstraction Assistant (DAA), a software application to track the specific location of abstracted data in source documents. Through DAA, data abstractors mark the source of information by flagging (or marking) specific locations in source documents, thereby creating a potentially permanent linkage (i.e., tracking) between abstracted information and its source. By clicking on existing markers (e.g., during data verification), DAA navigates the screen to the exact location of the source document, with the pertinent text highlighted. We have described the technical details and functionality of DAA elsewhere [12].

We conducted the DAA trial, a randomized controlled trial (RCT) comparing the accuracy and efficiency of three data abstraction approaches: (A) DAA-facilitated single abstraction plus verification (''DAA verification''), (B) single abstraction plus verification (''regular verification''), and (C) independent dual abstraction plus adjudication (''independent abstraction'').

## 2. Methods

The DAA trial was approved by Institutional Review Boards at Johns Hopkins University (IRB number 00006521, July 13, 2015) and Brown University (August 21, 2015). The protocol for the DAA trial was published previously [13]. Appendix 1 includes a completed CONSORT checklist for the DAA trial.

### 2.1. Participants

Eligible participants were individuals who were aged at least 20 years with self-reported proficiency in reading scientific articles in English and had completed data abstraction from at least one journal article for a systematic review in any field. We formed pairs consisting of one ''less experienced'' (defined as coauthored fewer than three published systematic reviews) and one ''more experienced'' (defined as coauthored three or more published systematic

**What is new?**

**Key findings**

- Although the three data abstraction approaches evaluated during the Data Abstraction Assistant (DAA) Trial—DAA verification, regular verification, and independent abstraction—had similar proportions of errors in abstracted data (15−17%), DAA verification had more errors in data items related to study outcomes and results than did regular verification and independent abstraction.
- Regular verification and independent abstraction were associated with similar odds of errors in items related to study design, but regular verification was associated with marginally (although not statistically significantly) higher odds of errors in items related to outcomes and results than independent abstraction (odds ratio [OR] = 1.16; 95% confidence interval [CI]: 0.97−1.40).
- DAA verification took 20 minutes (95% CI: 1−40) longer than regular verification and 46 minutes (95% CI: 26−66) shorter than independent abstraction.

**What this adds to what was known?**

- Independent abstraction may be necessary for outcomes and results data although verification may be sufficient for other types of data.

**What is the implication and what should change now?**

- DAA provides an audit trail that can promote reproducible research. Systematic reviewers may not need to conduct independent dual abstraction for all types of data. We developed a set of considerations to guide systematic reviewers in their choice of data abstraction approach (see Box 1).

reviews) abstractor. We defined the level of experience with abstraction using the results of a pilot study and consensus among this study's investigators [13,14]. We developed DAA and tested it in the DAA trial using SRDR, a free, open-source, open-access Web-based data management system for systematic reviews [9,10].

### 2.2. Abstraction approaches

DAA verification (Approach A) required the less experienced abstractor in each pair to use DAA to first abstract data into the abstraction form in SRDR. SRDR displayed the article and the abstraction form side by side. The less experienced abstractor was required to place one or more markers (or flags) denoting the location(s) in the article that supported the response to each item. The more experienced abstractor then verified the abstracted data in SRDR by examining the data together with the flagged locations in the document, revising any data as appropriate (verification) and, if necessary, consulting with the less experienced abstractor.

Regular verification (Approach B), which did not involve the use of DAA, required the less experienced abstractor in the pair to first abstract data to the form in SRDR. The more experienced abstractor then verified the abstracted data in SRDR, as described previously.

Independent abstraction (Approach C) required each abstractor (less and more experienced abstractors) in a pair to abstract data independently using the abstraction form in SRDR; they did not use DAA. The two abstractors then compared their abstracted data and resolved any discrepancies (data adjudication).

### 2.3. Randomization process

We identified 48 journal articles reporting results of RCTs from four published reviews addressing the following topics: prevention of falls [15], treatment of hypercholesterolemia [16], promotion of physical activity [17], and treatment of depression [18] (12 articles per review). In cases where a review included more than 12 articles, we selected the 12 articles that reported the largest number of outcomes that we had selected for the DAA trial.

We randomized 48 abstractors in 24 pairs (of one less experienced and one more experienced abstractors) to a balanced cross-over design (see Appendix 2 and the protocol for the DAA trial for a detailed explanation of the trial design and sample size calculation [13]). Each pair abstracted data online for six articles, two articles using each of the three approaches (A, B, and C), in one of six possible sequences (AABBCC, AACCBB, BBCCAA, BBAACC, CCAABB, and CCBBAA; with a 1:1:1:1:1:1 allocation ratio). Three pairs abstracted data for each article.

The study statistician (C.H.S.) generated a simple randomization sequence using the *sample* command in R. To maintain allocation concealment, he provided the Project Director (I.J.S.) the sequence for each pair of abstractors via email only when I.J.S. notified C.H.S. that the pair was ready to be randomized.

### 2.4. Masking

The project director (I.J.S.) and abstractors were not masked because they needed to know the assigned approach to allocate articles and abstract data. Because both outcomes examined in this study (error proportions and time) were computer measured, the lack of masking is unlikely to have impacted our results.

### 2.5. Participant follow-up

The project director (I.J.S.) maintained regular email contact with abstractors until abstraction for all six articles had been completed. Each abstractor received $250 as compensation for participation in the trial once this had been achieved.

### 2.6. Data collection

We collected all data through Web sites, SRDR, and DAA. The data abstraction forms in SRDR are publicly available at https://bit.ly/2w7HAUK. Except for a few data fields, each form comprised entirely multiple-choice or numerical entry data items organized into the following sections or "tabs" (a term used in SRDR)—*design* (study design and risk of bias), *baseline* characteristics of participants, *outcomes* reported, and quantitative *results*. We combined data items from the outcomes and results tabs for analysis. The forms had a median of 145 multiple-choice or numerical entry data items (range 106–187).

### 2.7. Outcomes

The two primary outcomes for the DAA trial were the proportion of data items incorrectly abstracted or omitted (*error proportions*) and the *time* taken for the complete abstraction process (by both abstractors, including verification and, when appropriate, adjudication). Errors were ascertained by a computer program, which compared the verified or adjudicated final data from each senior abstractor with an answer key. In generating the answer key, two investigators with extensive experience abstracting data for systematic reviews (T.L. and I.J.S.) independently abstracted data and adjudicated all discrepancies. For any data item with an overall error proportion $\geq$50% across all abstractors, we double checked the answer key for correctness. To calculate the time for a given article (in minutes), we summed the time taken for initial abstraction(s) and subsequent verification/adjudication as appropriate. We used two methods to measure time: an automatic timer built into SRDR and a self-recorded method, in which participants recorded the time they spent conducting abstraction plus verification/adjudication. Our primary analysis uses the auto-recorded time. Appendix 3 includes further details about the measurement of the outcomes of error proportions and time.

### 2.8. Statistical approaches

We conducted all analyses using R (version 3.4.2) according to the intention-to-treat (ITT) principle. We computed the means of the error proportions and time for each approach, review topic, and type of question (design, baseline, or outcomes/results).

We used two-level mixed models to compare the mean error proportions and times of the three abstraction approaches. The unit of analysis was an article abstracted by a pair. Analyses for error proportions and times used a binomial generalized linear mixed and a linear mixed model, respectively. The first level described variation within pairs of abstractors; the second level described variation among pairs. Level 1 factors investigated included the approach and indicators for the approach used in the first and last article abstracted by each pair. Level 2 (pair) factors included the review and the sequence. We considered the pair as a random effect by including a random intercept in the Level 1 model. We also explored interactions of approach with sequence, review, and first and last articles reviewed. Because all participants completed all abstractions, the trial had no missing data.

Protocol violations occurred in two pairs (1): the project director assigned two incorrect studies to one pair, and (2) the first abstractor in another pair forgot to drop flags during data abstraction when using DAA. To address these violations, we enrolled two additional pairs. The results of our analyses did not differ based on whether or not the two additional pairs were included. We therefore analyzed

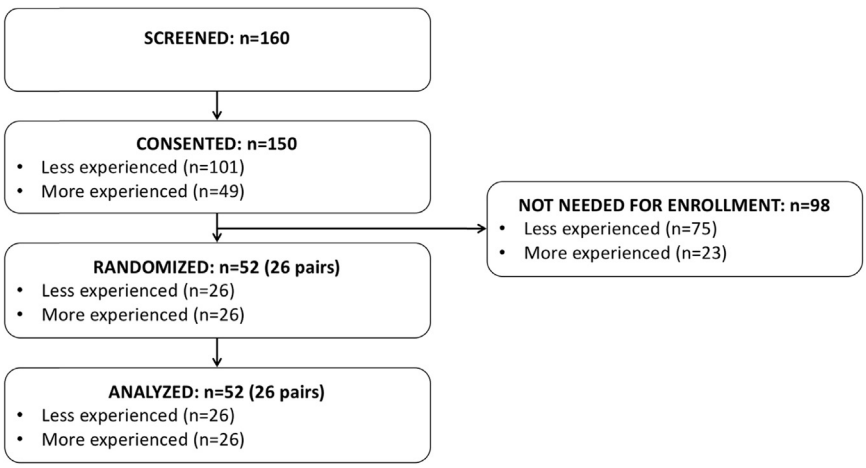


**Fig. 1.** Participant flow during the DAA trial. The 26 pairs were formed by pairing each of the first 26 less experienced abstractors with the next available participant from the first 26 more experienced abstractors.

**Table 1.** Baseline characteristics of all 52 participants in the DAA trial

| Characteristics | Random sequence | | | | | | |
|---|---|---|---|---|---|---|---|
| | AABBCC, *n* = 8 | BBCCAA, *n* = 8 | CCAABB, *n* = 8 | AACCBB, *n* = 10 | BBAACC, *n* = 10 | CCBBAA, *n* = 8 | All sequences (*N* = 52) |
| Age range (y) | | | | | | | |
| 20–29 | 3 (38) | 3 (37) | 7 (88) | 6 (60) | 5 (50) | 5 (63) | 29 (56) |
| 30–39 | 2 (25) | 5 (63) | — | 4 (40) | 4 (40) | 2 (25) | 17 (33) |
| 40–49 | 1 (13) | — | — | — | — | 1 (13) | 2 (4) |
| 50–59 | 2 (25) | — | — | — | 1 (10) | — | 3 (6) |
| 60–69 | — | — | — | — | — | — | — |
| ≥70 | — | — | 1 (13) | — | — | — | 1 (2) |
| Number of articles abstracted | | | | | | | |
| 1–9 | — | — | 2 (25) | 1 (10) | 1 (10) | 1 (13) | 5 (10) |
| 10–19 | — | 3 (38) | — | — | 2 (20) | 3 (38) | 8 (15) |
| ≥20 | 8 (100) | 5 (63) | 6 (75) | 9 (90) | 7 (70) | 4 (50) | 39 (75) |
| Number of systematic reviews published | | | | | | | |
| 0 | 1 (13) | 4 (50) | 3 (38) | 4 (40) | 4 (40) | 3 (38) | 19 (37) |
| 1–2 | — | — | 1 (13) | 1 (10) | 1 (10) | 1 (13) | 4 (8) |
| 3–5 | 3 (38) | 2 (25) | 2 (25) | 4 (40) | 2 (20) | 1 (13) | 14 (27) |
| ≥6 | 4 (50) | 2 (25) | 2 (25) | 1 (10) | 3 (30) | 3 (38) | 15 (29) |
| Last time abstracting data | | | | | | | |
| Within preceding 6 mo | 7 (88) | 7 (88) | 8 (100) | 10 (100) | 7 (88) | 7 (88) | 48 (92) |
| ≥6 mo earlier | 1 (13) | 1 (13) | — | — | 1 (13) | 1 (13) | 4 (8) |
| Training in systematic reviews[a] | | | | | | | |
| No training | — | — | — | — | — | — | — |
| Took an SR methods course | 5 (63) | 5 (63) | 3 (38) | 7 (70) | 7 (70) | 7 (88) | 34 (65) |
| Attended an SR workshop | 3 (38) | 3 (38) | 1 (13) | 2 (20) | 3 (38) | 2 (25) | 14 (27) |
| Received on-the-job training | 5 (63) | 4 (50) | 7 (88) | 5 (50) | 7 (70) | 5 (63) | 33 (64) |
| Received other forms of training | 2 (25) | 2 (25) | 2 (25) | 3 (38) | 1 (10) | 1 (13) | 11 (21) |
| Self-rated level of experience | | | | | | | |
| Slightly experienced | — | 1 (13) | — | 1 (10) | 1 (10) | 1 (13) | 4 (8) |
| Somewhat/moderately experienced | 4 (50) | 3 (38) | 2 (25) | 5 (50) | 6 (60) | 6 (75) | 26 (50) |
| Very experienced | 4 (50) | 4 (50) | 6 (75) | 4 (40) | 3 (30) | 1 (13) | 22 (42) |
| Primary professional status | | | | | | | |
| Faculty | 3 (38) | 1 (13) | 1 (13) | 2 (20) | 3 (38) | — | 10 (19) |
| Doctoral student | 1 (13) | 2 (25) | 2 (25) | 3 (30) | 2 (20) | 2 (25) | 12 (23) |
| Master's student | 2 (25) | 2 (25) | 1 (13) | 2 (20) | 2 (20) | 1 (13) | 10 (19) |
| Staff | 1 (13) | 3 (38) | 3 (38) | — | 2 (20) | 3 (38) | 12 (23) |
| Other | 1 (13) | — | 1 (13) | 3 (30) | 1 (10) | 2 (25) | 8 (15) |

*Abbreviations:* DAA, Data Abstraction Assistant; SR, systematic review.
[a] Participants could select all options that apply; therefore, the percentages add up to more than 100%.

and reported data from all 26 pairs (52 participants) under the ITT principle.

We also performed a *post hoc* analysis of the potential impact of errors on meta-analyses using a continuous outcome and a binary outcome. This analysis was not intended to compare the three approaches tested in the DAA trial. We calculated the percentage bias in the estimate and standard error of the estimate. We describe our methods for this *post hoc* analysis more fully in Appendix 4 [19,20].

## 3. Results

Between March 18, 2016, and February 1, 2017, we screened 160 potential data abstractors and randomized 52 participants (26 pairs) of them (Fig. 1). All 52 participants completed the data abstraction required in the DAA trial by April 3, 2017.

Most participants were aged 20–29 years at the time of screening, and most had abstracted data within the preceding 6 months (Table 1). More than 90% of participants had

**Table 2.** Proportion of errors by data abstraction approach, type of error, type of data abstracted, and systematic review topic

| | DAA verification | | | | Regular verification | |
|---|---|---|---|---|---|---|
| | Type of error | | | | Type of error | |
| Type of data abstracted and systematic review topic | Total errors, mean % (range) | Errors of omission, mean % (range) | Incorrect abstractions, mean % (range) | Number of fields, mean (range) | Total errors, mean % (range) | Errors of omission, mean % (range) |
| Design tab | | | | | | |
| Topic 1 | 21 (7–49) | 0 (0–0) | 21 (7–49) | 42 (37–43) | 14 (7–19) | 0 (0–0) |
| Topic 2 | 18 (9–30) | 0 (0–0) | 18 (9–30) | 45 (42–46) | 13 (0–21) | 0 (0–0) |
| Topic 3 | 12 (5–20) | 0 (0–0) | 12 (5–20) | 45 (42–46) | 12 (2–20) | 0 (0–0) |
| Topic 4 | 17 (2–48) | 0 (0–0) | 17 (2–48) | 46 (42–48) | 15 (2–36) | 0 (0–0) |
| All topics | 17 (2–49) | 0 (0–0) | 17 (2–49) | 45 (37–48) | 13 (0–36) | 0 (0–0) |
| Baselines tab | | | | | | |
| Topic 1 | 11 (3–19) | 0 (0–0) | 11 (3–19) | 62 (59–65) | 10 (0–20) | 1 (0–9) |
| Topic 2 | 11 (0–35) | 0 (0–0) | 11 (0–35) | 76 (63–84) | 11 (0–34) | 0 (0–0) |
| Topic 3 | 9 (0–24) | 0 (0–0) | 9 (0–24) | 73 (64–81) | 9 (0–26) | 0 (0–0) |
| Topic 4 | 9 (0–33) | 0 (0–0) | 9 (0–33) | 52 (45–57) | 11 (0–27) | 0 (0–0) |
| All topics | 10 (0–35) | 0 (0–0) | 10 (0–35) | 65 (45–84) | 10 (0–34) | 0 (0–9) |
| Outcomes and results tabs | | | | | | |
| Topic 1 | 48 (9–95) | 44 (9–76) | 4 (0–19) | 24 (10–38) | 37 (10–65) | 35 (0–65) |
| Topic 2 | 42 (0–86) | 32 (0–86) | 9 (0–37) | 31 (7–52) | 41 (6–95) | 28 (0–95) |
| Topic 3 | 40 (0–100) | 36 (0–100) | 4 (0–23) | 22 (7–43) | 35 (7–100) | 32 (0–100) |
| Topic 4 | 35 (8–100) | 22 (0–100) | 13 (0–71) | 13 (3–25) | 31 (0–100) | 27 (0–100) |
| All topics | 41 (0–100) | 33 (0–100) | 8 (0–71) | 22 (3–52) | 36 (0–100) | 31 (0–100) |
| All tabs | | | | | | |
| Topic 1 | 20 (6–28) | 8 (2–17) | 11 (3–20) | 150 (131–162) | 16 (7–21) | 8 (0–13) |
| Topic 2 | 17 (6–24) | 6 (0–14) | 11 (6–16) | 166 (128–181) | 16 (7–33) | 6 (0–23) |
| Topic 3 | 14 (8–30) | 6 (0–15) | 8 (2–15) | 155 (133–187) | 14 (4–32) | 6 (1–15) |
| Topic 4 | 17 (9–33) | 7 (4–15) | 11 (1–24) | 130 (116–142) | 18 (6–25) | 7 (4–10) |
| All topics | 17 (6–33) | 7 (0–17) | 10 (1–24) | 150 (116–187) | 16 (4–33) | 6 (0–23) |

*Abbreviation:* DAA, Data Abstraction Assistant.
Topic 1: Multifactorial interventions to prevent falls in older adults [15].
Topic 2: Proprotein convertase subtilisin/kexin type 9 antibodies for adults with hypercholesterolemia [16].
Topic 3: Interventions to promote physical activity in cancer survivors [17].
Topic 4: Omega-3 fatty acids for adults with depression [18].

abstracted data from 20 or more studies, and all participants had received some systematic review methods training before participating in the trial. Nearly, all participants characterized their proficiency with data abstraction as "somewhat/moderately experienced" or "very experienced." Note that we used coauthoring fewer than three vs. three or more reviews to categorize abstractors' level of experience in the DAA trial, regardless of these self-reported proficiencies with data abstraction.

### 3.1. Errors

#### 3.1.1. Error proportions

Across all approaches, the mean proportion of errors by data abstractor pairs per article was 16% (range 2–33%) (Tables 2 and 3). These proportions were similar among abstraction approaches: 17% (range 6–33%) for DAA verification, 16% (range 4–33%) for regular verification, and 15% (range 2–30%) for independent abstraction. When we compared mean error proportions by types of data items, we found differences, however. Mean error proportions were higher for data items related to outcomes, and results (36%) compared with items related to design (15%) or participant baseline characteristics (10%). Errors in outcomes and results were related to abstraction approach: mean error proportions were higher for DAA verification (41%) than for regular verification (36%) or independent abstraction (31%). These between-approach differences were similar for data items related to design (error proportions ranged from 13% to 17%) and baseline characteristics (all 10%).

Overall, the mean proportion of errors arising due to omission was lower than errors arising due to incorrect abstractions (6% vs. 10%). However, among the outcomes and results items, omissions constituted a much larger share

| Regular verification | | Independent abstraction | | | |
|---|---|---|---|---|---|
| Type of error | | Type of error | | | |
| Incorrect abstractions, mean % (range) | Number of fields, mean (range) | Total errors, mean % (range) | Errors of omission, mean % (range) | Incorrect abstractions, mean % (range) | Number of fields, mean (range) |
| | | | | | |
| 14 (7–19) | 42 (37–43) | 18 (7–35) | 0 (0–0) | 18 (7–35) | 42 (37–43) |
| 13 (0–21) | 45 (42–46) | 10 (2–23) | 0 (0–0) | 10 (2–23) | 45 (42–46) |
| 12 (2–20) | 45 (42–46) | 10 (4–21) | 0 (0–0) | 10 (4–21) | 45 (42–46) |
| 15 (2–36) | 46 (42–48) | 17 (4–36) | 0 (0–0) | 17 (4–36) | 46 (42–48) |
| 13 (0–36) | 45 (37–48) | 14 (2–36) | 0 (0–0) | 14 (2–36) | 45 (37–48) |
| | | | | | |
| 9 (0–20) | 62 (59–65) | 7 (0–14) | 0 (0–0) | 7 (0–14) | 62 (59–65) |
| 11 (0–34) | 76 (63–84) | 15 (0–41) | 0 (0–0) | 15 (0–41) | 76 (63–84) |
| 9 (0–26) | 72 (64–81) | 10 (1–24) | 0 (0–0) | 10 (1–24) | 72 (64–81) |
| 11 (0–27) | 51 (45–57) | 8 (0–20) | 0 (0–0) | 8 (0–20) | 52 (45–57) |
| 10 (0–34) | 65 (45–84) | 10 (0–41) | 0 (0–0) | 10 (0–41) | 65 (45–84) |
| | | | | | |
| 2 (0–17) | 24 (10–38) | 43 (4–100) | 40 (0–100) | 3 (0–14) | 24 (10–38) |
| 13 (0–86) | 31 (7–52) | 35 (7–86) | 33 (0–86) | 2 (0–14) | 31 (7–52) |
| 3 (0–27) | 23 (7–43) | 21 (0–60) | 17 (0–53) | 5 (0–57) | 23 (7–43) |
| 4 (0–40) | 11 (3–25) | 29 (0–100) | 28 (0–100) | 1 (0–8) | 11 (3–25) |
| 5 (0–86) | 22 (3–52) | 31 (0–100) | 29 (0–100) | 2 (0–57) | 22 (3–52) |
| | | | | | |
| 8 (4–13) | 144 (123–168) | 16 (8–27) | 8 (0–18) | 9 (2–16) | 144 (123–156) |
| 10 (3–20) | 161 (123–181) | 16 (8–30) | 6 (0–12) | 10 (4–21) | 161 (128–179) |
| 8 (1–18) | 148 (123–177) | 12 (2–19) | 4 (0–13) | 9 (2–15) | 149 (123–177) |
| 11 (2–21) | 122 (106–137) | 17 (6–25) | 7 (4–15) | 10 (2–20) | 123 (106–142) |
| 9 (1–21) | 143 (106–181) | 15 (2–30) | 6 (0–18) | 9 (2–21) | 144 (106–179) |

of the errors (31%) than incorrect abstractions (5%). Omissions in outcomes and results occurred mostly because abstractors did not name one or more of the outcomes. For example, if incidence of falls by 12 months was not named as an outcome, results data for this outcome (number of falls and person-time in each group, as well as incidence rate and precision measures between groups) would have been missed, constituting errors of omission. Similar patterns were observed for each approach and review topic.

### *3.1.2. Between-approach comparisons of errors*

Overall, across all types of data items, although the crude error proportions were similar (17% for DAA verification, 16% for regular verification, and 15% for independent abstraction), DAA verification was associated with a higher odds of errors than regular verification (adjusted odds ratio [OR] = 1.08; 95% confidence interval [CI]: 0.99–1.17) and independent abstraction (adjusted OR = 1.12; 95% CI: 1.03–1.22) (Table 4). The majority of these between-approach differences arose from data items related to outcomes and results, where DAA verification was associated with higher odds of errors than regular verification (adjusted OR 1.30; 95% CI: 1.11–1.53) or independent abstraction (adjusted OR = 1.52; 95% CI: 1.27–1.82), respectively.

Regular verification and independent abstraction were associated with similar odds of errors in items related to study design, but regular verification was associated with marginally (although not statistically significantly) higher odds of errors in items related to outcomes and results than independent abstraction (adjusted OR = 1.16; 95% CI: 0.97–1.40). No notable between-approach differences were observed in items related to baseline characteristics.

Error proportions were generally higher for the first and lower for the last article abstracted. The interaction between approach and sequence is hard to interpret.

**Table 3.** Proportion of errors across all approaches, by type of error, type of data abstracted, and systematic review topic

| Type of data abstracted and systematic review topic | All approaches | | | |
|---|---|---|---|---|
| | Type of error | | | |
| | Total errors, mean % (range) | Errors of omission, mean % (range) | Incorrect abstractions, mean % (range) | Number of fields, mean (range) |
| Design tab | | | | |
| Topic 1 | 18 (7–49) | 0 (0–0) | 18 (7–49) | 42 (37–43) |
| Topic 2 | 14 (0–30) | 0 (0–0) | 14 (0–30) | 45 (42–46) |
| Topic 3 | 11 (2–21) | 0 (0–0) | 11 (2–21) | 45 (42–46) |
| Topic 4 | 16 (2–48) | 0 (0–0) | 16 (2–48) | 46 (42–48) |
| All topics | 15 (0–49) | 0 (0–0) | 15 (0–49) | 45 (37–48) |
| Baseline tab | | | | |
| Topic 1 | 9 (0–20) | 0 (0–9) | 9 (0–20) | 62 (59–65) |
| Topic 2 | 12 (0–41) | 0 (0–0) | 12 (0–41) | 76 (63–84) |
| Topic 3 | 9 (0–26) | 0 (0–0) | 9 (0–26) | 72 (64–81) |
| Topic 4 | 9 (0–33) | 0 (0–0) | 9 (0–33) | 52 (45–57) |
| All topics | 10 (0–41) | 0 (0–9) | 10 (0–41) | 65 (45–84) |
| Outcomes and results tabs | | | | |
| Topic 1 | 43 (4–100) | 40 (0–100) | 3 (0–19) | 24 (10–38) |
| Topic 2 | 39 (0–95) | 31 (0–95) | 8 (0–86) | 31 (7–52) |
| Topic 3 | 32 (0–100) | 28 (0–100) | 4 (0–57) | 23 (7–43) |
| Topic 4 | 32 (0–100) | 26 (0–100) | 6 (0–71) | 12 (3–25) |
| All topics | 36 (0–100) | 31 (0–100) | 5 (0–86) | 22 (3–52) |
| All tabs | | | | |
| Topic 1 | 17 (6–28) | 8 (0–18) | 9 (2–20) | 146 (123–168) |
| Topic 2 | 17 (6–33) | 6 (0–23) | 11 (3–21) | 163 (123–181) |
| Topic 3 | 14 (2–32) | 5 (0–15) | 8 (1–18) | 151 (123–187) |
| Topic 4 | 17 (6–33) | 7 (4–15) | 10 (1–24) | 125 (106–142) |
| All topics | 16 (2–33) | 6 (0–23) | 10 (1–24) | 145 (106–187) |

Topic 1: Multifactorial interventions to prevent falls in older adults [15].
Topic 2: Proprotein convertase subtilisin/kexin type 9 antibodies for adults with hypercholesterolemia [16].
Topic 3: Interventions to promote physical activity in cancer survivors [17].
Topic 4: Omega-3 fatty acids for adults with depression [18].

### 3.2. Time

Mean times for abstracting each article during the DAA trial, as captured by the automatic timer, were generally greater than those captured by self-recorded time. The mean times per article were 136 minutes (range 39–399) and 107 minutes (range 30–285), as captured by the auto- and self-recorded clocks, respectively. Our primary analysis uses the auto-recorded time; Appendix 3 includes results for self-recorded time. The patterns observed and conclusions drawn are comparable regardless of which time is used.

**Table 4.** Between-approach comparisons of error proportions by type of data abstracted

| Tab | DAA verification vs. independent abstraction | | Regular verification vs. independent abstraction | | DAA verification vs. regular verification | |
|---|---|---|---|---|---|---|
| | Adjusted OR (95% CI) | *P* value | Adjusted OR (95% CI) | *P* value | Adjusted OR (95% CI) | *P* value |
| Design tab | **1.30** (1.11–1.53) | 0.002 | 0.99 (0.83–1.17) | 0.90 | **1.32** (1.12–1.55) | 0.001 |
| Baseline tab | 1.02 (0.87–1.20) | 0.83 | 1.05 (0.89–1.23) | 0.58 | 0.97 (0.83–1.14) | 0.74 |
| Outcomes and results tabs | **1.52** (1.27–1.82) | <0.0001 | 1.16 (0.97–1.40) | 0.10 | **1.30** (1.09–1.56) | 0.004 |
| All tabs | **1.12** (1.03–1.22) | 0.01 | 1.04 (0.95–1.13) | 0.41 | 1.08 (0.99–1.17) | 0.09 |

Abbreviations: DAA, Data Abstraction Assistant; CI, confidence interval; OR, odds ratio.
Bolded items are significant at a 0.05 level. The model adjusted for sequence, systematic review topic, and indicators for the approach used on the first and last article abstracted by each pair. The model that did not include indicators for the approach used on the first and last article abstracted by each pair rendered similar findings.

**Table 5.** Autorecorded time spent (in minutes) by data abstraction approach, type of data abstracted, and systematic review topic

| Type of data abstracted and systematic review topic | DAA verification, mean % (range) | Regular verification, mean % (range) | Independent abstraction mean, % (range) | All approaches, mean % (range) |
|---|---|---|---|---|
| Design tab | | | | |
| Topic 1 | 46 (21–107) | 36 (19–59) | 51 (22–70) | 44 (19–107) |
| Topic 2 | 61 (17–111) | 58 (10–232) | 50 (36–85) | 56 (10–232) |
| Topic 3 | 54 (17–148) | 37 (9–82) | 84 (43–145) | 58 (9–148) |
| Topic 4 | 63 (16–199) | 41 (16–81) | 63 (24–166) | 55 (16–199) |
| All topics | 56 (16–199) | 43 (9–232) | 63 (22–166) | 54 (9–232) |
| Baselines tab | | | | |
| Topic 1 | 9 (5–18) | 7 (4–15) | 15 (5–32) | 10 (4–32) |
| Topic 2 | 27 (8–66) | 14 (3–24) | 29 (16–78) | 24 (3–78) |
| Topic 3 | 19 (6–52) | 11 (3–20) | 28 (15–47) | 19 (3–52) |
| Topic 4 | 23 (4–155) | 9 (5–19) | 23 (8–75) | 18 (4–155) |
| All topics | 20 (4–155) | 10 (3–24) | 24 (5–78) | 18 (3–155) |
| Outcomes and results tabs | | | | |
| Topic 1 | 29 (5–68) | 27 (10–82) | 75 (16–165) | 44 (5–165) |
| Topic 2 | 44 (11–111) | 46 (9–96) | 55 (27–138) | 48 (9–138) |
| Topic 3 | 43 (10–128) | 58 (8–244) | 69 (20–140) | 57 (8–244) |
| Topic 4 | 27 (5–69) | 33 (8–97) | 72 (13–245) | 44 (5–245) |
| All topics | 36 (5–128) | 41 (8–244) | 68 (13–245) | 48 (5–245) |
| All tabs | | | | |
| Topic 1 | 98 (44–170) | 84 (39–194) | 162 (48–243) | 114 (39–243) |
| Topic 2 | 146 (50–290) | 132 (44–341) | 150 (93–267) | 143 (44–341) |
| Topic 3 | 134 (46–350) | 118 (40–311) | 199 (113–310) | 151 (40–350) |
| Topic 4 | 132 (41–326) | 96 (42–172) | 174 (51–399) | 134 (41–399) |
| All topics | 128 (41–350) | 107 (39–341) | 172 (48–399) | 136 (39–399) |

*Abbreviations:* DAA, Data Abstraction Assistant.

Time for all tabs is greater than the sum of the design, baselines, and outcomes and results tabs because all tabs also incorporates time spent on the other tabs in SRDR, that is, key questions, publications, arms, and finalize tabs.

Topic 1: Multifactorial interventions to prevent falls in older adults [15].

Topic 2: Proprotein convertase subtilisin/kexin type 9 antibodies for adults with hypercholesterolemia [16].

Topic 3: Interventions to promote physical activity in cancer survivors [17].

Topic 4: Omega-3 fatty acids for adults with depression [18].

Mean auto-recorded times were higher for independent abstraction (172 minutes [range 48–399]) than for DAA verification (128 minutes [range 41–350]) or regular verification (107 minutes [range 39–341]; Table 5). Some review topics took longer to abstract than others. Regardless of abstraction approach, abstractors spent between two and three times more time on data items related to design, outcomes, and results than on items related to baseline characteristics.

### 3.2.1. Between-approach comparisons of time

When considering the total time spent on each article, DAA verification took 20 minutes (95% CI: 1–40) longer than regular verification, but 46 minutes (95% CI: 26–66) shorter than independent abstraction (Table 6). When considering the time spent by type of data items, DAA verification took longer than regular verification for items related to design and for baseline characteristics, but not for items related to outcomes and results. DAA verification took shorter than independent abstractions for all types of items.

### 3.3. Potential impact of errors on meta-analyses

Appendix 4 presents the results of the *post hoc* analysis. In summary, we found that for both the analyzed outcomes, that is, one continuous and one binary outcome, any errors in the data abstracted during the DAA trial would not make a sizable impact on meta-analytic summaries. For the continuous outcome, the percentage biases in the estimate and standard error of the mean difference were −8.0% and +28.1%, respectively (Appendix 4 explains how percentage bias was calculated). Similarly, for the binary outcome, the percentage biases in the estimate and standard error of the relative risk were −2.3% and +19.3%, respectively.

## 4. Discussion

We found that the overall proportions of errors were high among the three abstraction approaches compared in this randomized cross-over trial. The errors were similar

**Table 6.** Between-approach comparisons of autorecorded time by type of data abstracted

| | DAA verification—independent abstraction | | Regular verification—independent abstraction | | DAA verification—regular verification | |
|---|---|---|---|---|---|---|
| **Tab** | **Adjusted MD (95% CI)** | ***P* value** | **Adj. MD (95% CI)** | ***P* value** | **Adjusted MD (95% CI)** | ***P* value** |
| Design tab | −7.2 (−17.7, 3.3) | 0.18 | **−20.6** (−31.1, −10.2) | 0.0001 | **13.4** (3.0, 23.9) | 0.01 |
| Baseline tab | −4.2 (−9.8, 1.5) | 0.15 | **−13.8** (−19.5, −8.2) | <0.0001 | **9.6** (4.0, 15.3) | 0.0008 |
| Outcomes and results tabs | **−33.3** (−45.7, −20.9) | <0.0001 | **−27.4** (−39.8, −15.0) | <0.0001 | −5.9 (−18.3, 6.5) | 0.35 |
| All tabs | **−45.9** (−65.5, −26.3) | <0.0001 | **−66.1** (−85.7, −46.5) | <0.0001 | **20.2** (0.6, 39.8) | 0.04 |

*Abbreviations:* DAA, Data Abstraction Assistant; MD, mean difference; CI, confidence interval.

Bolded items are significant at a 0.05 level. The model adjusted for sequence, systematic review topic, and indicators for the approach used on the first and last article abstracted by each pair. The model that did not include indicators for the approach used on the first and last article abstracted by each pair rendered similar findings, except that the comparison between approaches A and B for all data items was not statistically significant (when those indicators were not included in the model).

across approaches. DAA verification was associated with more errors than each of the other two approaches, especially for data items related to outcomes and results. Not surprisingly, DAA verification took substantially shorter than independent abstraction, and longer than regular verification, likely because of the requirement for placing flags by the junior abstractor in DAA verification.

Although somewhat high, the error proportions observed in this trial (15–17%) are consistent with other studies [21]. In this trial, the highest proportions of errors were observed for data items related to outcomes and results, and most errors (in outcomes and results) arose because of omissions, either of entire outcomes or of some results fields within outcomes. When outcomes and results were not missed, data were generally abstracted accurately. As discussed in our *post hoc* analysis (Appendix 4), the errors we noted did not have a sizable impact on meta-analytic effect estimates. Regardless, quality assurance procedures, including detailed protocols and instructions for data abstraction, and regular and ongoing abstractor training, should focus on outcomes and results. In a separate paper, we will explore whether more experienced abstractors made fewer errors than less experienced abstractors.

Reasons for the slightly higher error proportions observed using DAA verification compared with the other approaches are worth exploring. First, DAA is a new software application that was tested among abstractors who were naïve to using it. Although we provided them training videos for using DAA, some errors likely arose from the learning curve and unfamiliarity with a new technology. Second, we did not monitor whether DAA was being used as intended. When dropping and reviewing flags, participants may have flagged only the first instance of relevant information for a given data item in an article and missed other relevant locations in the article. This may be related partly to the fact that only the less experienced abstractors dropped flags; it is possible that the more experienced abstractors might have been more comprehensive in placing flags (not tested in this trial). It is also possible that verifiers were anchored to what had already been flagged and therefore prone to missing unflagged information. Because there generally is a delay in peak human performance with a new tool [22], we expect that the error proportions we observed may reduce over time.

Not surprisingly, the two verification approaches required considerably less time (1 hour less per article) than independent abstraction. In addition, the adjudication process in independent abstraction (i.e., after initial independent abstraction) took two-thirds as long as the initial abstraction. When adjudicating, abstractors had to reorient themselves to the article and identify and discuss discrepant data to arrive at consensus.

Our findings fill a critical methodological gap in current understanding of data abstraction best practices. It appears that IOM's recommendation that independent abstraction be used for quantitative data is on target. We found that taking both time and errors into account, independent abstraction is most important for data needed for meta-analysis (i.e., outcomes and results), but may not be necessary for other items. We developed a set of considerations to guide systematic reviewers in their choice of data abstraction approach (Box 1).

### 4.1. Limitations and strengths

We acknowledge several limitations to our study. First, we evaluated DAA's compatibility with only one data abstraction system (SRDR). Second, we did not require data abstractors to be familiar with the topics of the reviews. It is possible that the abstractors in the DAA trial were less familiar with the data items to be abstracted than might be expected of real-life data abstractors, who often participate in developing protocols, screening studies, and designing data abstraction forms. Third, although we tried to make the questions and instructions on the data abstraction forms clear and pilot-tested the forms among the DAA trial investigators, we did not pilot test the forms with trial participants. Finally, we did not intervene during the trial to improve the quality of abstraction, such as through ongoing training and group discussions.

This study has several strengths and presents some opportunities. We used a rigorous randomized trial

**Box 1 Considerations when selecting data abstraction approaches during systematic reviews**

| Tasks | Guidance |
|---|---|
| Data abstraction system | • Use electronic data abstraction systems where possible. The system chosen should be able to implement best practices of form development, enhance open science and reproducibility, and reduce research waste. |
| Form development | • Pilot test form.<br>• Provide clear instructions. Provide definitions to clarify terms.<br>• Minimize open-ended questions.<br>• Use existing templates and existing (and common) data items, tailoring questions to specific topics as needed. |
| Training and composition of data abstractor team | • Conduct regular and ongoing training to reinforce methods and prevent inconsistencies in interpretation of data items. |
| Data abstraction approach (directly informed by findings of the DAA trial) | • Avoid single-data abstraction to minimize errors.<br>• Regular verification leads to a similar amount of errors overall as independent abstraction, but takes a substantially shorter amount of time. Regular verification may lead to more errors than independent abstraction for data items related to outcomes and results.<br>• DAA verification could be considered because it:<br>1. has similar overall error proportions to independent abstraction;<br>2. takes a substantially shorter amount of time than independent abstraction;<br>3. has the potential to promote reproducible science through creation of permanent linkages between abstracted data and their sources (something that regular verification does not). This can facilitate updating of systematic reviews and sharing of previously abstracted data for other purposes; and<br>4. can contribute to evaluating and advancing the use of various automated and semiautomated natural language processing and machine learning tools for systematic review production.<br>• Pay careful attention to data items that are more prone to errors (such as outcomes and numerical results) and those that are subjective and require judgment (such as risk of bias). These types of data items may benefit from independent abstraction. |
| Managing abstracted data | • Anticipate challenges associated with the complexities of data management, especially for large systematic reviews, and plan accordingly.<br>• Decide whether calculation-type questions should be dealt with by abstractors during data abstraction or centrally during data management. |

methodology, comparing DAA-assisted single abstraction plus verification to two predominant recommended approaches to data abstraction. In addition, compared with previous studies, we included both less and more experienced abstractors who abstracted data on a range of topics. We believe that, to the extent that DAA is used as intended, it should promote reproducible science through the creation of permanent linkages between abstracted data and their sources. Such links may facilitate the updating of reviews and sharing of previously abstracted data for other purposes. These linkages also can contribute to evaluating various automated or semiautomated tools for abstraction [23–25]. DAA is one example of a tool that could help achieve the overarching goal of improving the transparency and reproducibility of systematic reviews. We encourage researchers to test other user cases and approaches for incorporating DAA in systematic review workflows (e.g., having more experienced abstractors use DAA first for outcomes and results data).

## 5. Conclusions

Considering accuracy and efficiency together, our findings suggest that independent dual abstraction may be necessary for outcomes and results data during systematic

reviews, and a verification approach may be sufficient for other types of data. By linking abstracted data with their exact source, DAA can provide an audit trail that is crucial for reproducible research; this is not achieved using the other two approaches. Systematic reviewers should choose their data abstraction approach based on the inevitable trade-off between saving time and minimizing errors.

## CRediT authorship contribution statement

**Tianjing Li:** Conceptualization, Funding acquisition, Investigation, Methodology, Data curation, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing. **Ian J. Saldanha:** Methodology, Data curation, Project administration, Validation, Writing - original draft, Writing - review & editing. **Jens Jap:** Conceptualization, Data curation, Formal analysis, Software, Writing - original draft, Writing - review & editing. **Bryant T. Smith:** Conceptualization, Data curation, Software, Writing - original draft, Writing - review & editing. **Joseph Canner:** Formal analysis, Writing - original draft, Writing - review & editing. **Susan M. Hutfless:** Conceptualization, Methodology, Writing - review & editing. **Vernal Branch:** Conceptualization, Investigation, Writing - review & editing. **Simona Carini:** Writing - original draft. **Wiley Chan:** Conceptualization, Investigation, Writing - review & editing. **Berry de Bruijn:** Conceptualization, Investigation, Methodology, Software, Writing - review & editing. **Byron C. Wallace:** Investigation, Software, Writing - review & editing. **Sandra A. Walsh:** Conceptualization, Investigation, Writing - review & editing. **Elizabeth J. Whamond:** Conceptualization, Investigation, Writing - review & editing. **M. Hassan Murad:** Conceptualization, Investigation, Methodology, Writing - review & editing. **Ida Sim:** Conceptualization, Investigation, Methodology, Writing - review & editing. **Jesse A. Berlin:** Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Joseph Lau:** Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Kay Dickersin:** Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Christopher H. Schmid:** Conceptualization, Funding acquisition, Investigation, Methodology, Data curation, Formal analysis, Supervision, Validation, Writing - original draft, Writing - review & editing.

## Acknowledgments

The authors are grateful to all the 52 abstractors who participated in the DAA trial as well as the 98 individuals who provided consent but whose participation was not needed. The Patient-Centered Outcomes Research Institute (PCORI) funded this work under contract number ME-1310-07009. The funders played no role in the design, analysis, and reporting of this trial.

## Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.07.005.

## References

[1] Committee on Standards for Systematic Reviews of Comparative Effectiveness Research, Board on Health Care Services. In: Eden J, Levit L, Berg A, Morton S, editors. Finding what works in health care: standards for systematic reviews. Washington, DC: National Academies Press; 2011.

[2] Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological expectations of Cochrane intervention reviews. Version 1.05, January 2018. Available at http://community.cochrane.org/mecir-manual. Accessed March 18, 2019.

[3] 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;59:697−703.

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

[5] 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;58:741−2.

[6] Patient-Centered Outcomes Research Institute Methodology Committee. Draft methodology report: our questions, our decisions: standards for Patient-Centered Outcomes Research. July 23, 2012 version. Available at http://pcori.org/assets/MethodologyReport-Comment.pdf. Accessed March 18, 2019.

[7] Center for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York, UK: York Publishing Services, Ltd. Available at https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. Accessed March 18, 2019.

[8] Methods guide for effectiveness and comparative effectiveness reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD. Available at https://effectivehealthcare.ahrq.gov/topics/cer-methods-guide/overview. Accessed March 18, 2019.

[9] Ip S, Hadar N, Keefe S, Parkin C, Iovin R, Balk EM, et al. A Web-based archive of systematic review data. Syst Rev 2012;1:15.

[10] 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;162:287−94.

[11] Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org. Accessed March 18, 2019.

[12] Jap J, Saldanha IJ, Smith BT, Lau J, Schmid C, Li T, On Behalf of the Data Abstraction Assistant Investigators. Features and functioning of Data Abstraction Assistant, a software application for data abstraction during systematic reviews. Res Synth Methods 2019;10:2−14.

[13] Saldanha IJ, Schmid CH, Lau J, Dickersin K, Berlin JA, Jap 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.

[14] Saldanha IJ, Wen J, Schmid CH, Li T. Data Abstraction Assistant (DAA): What characteristics classify ''experience'' with data abstraction? 2016 Cochrane Colloquium, October 23-27, Seoul, South Korea.

[15] Choi M, Hector M. Effectiveness of intervention programs in preventing falls: systematic review of recent 10 years and meta-analysis. J Am Med Dir Assoc 2012;13(2):188.e13−21.

[16] Navarese EP, Kolodziejczak M, Schulze V, Gurbel PA, Tantry U, Lin Y, et al. Effects of proprotein convertase subtilisin/kexin type 9 antibodies in adults with hypercholesterolemia: a systematic review and meta-analysis. Ann Intern Med 2015;163:40−51.

[17] Fong DY, Ho JW, Hui BP, Lee AM, Macfarlane DJ, Leung SS, et al. Physical activity for cancer survivors: meta-analysis of randomised controlled trials. BMJ 2012;344:e70.

[18] Appleton KM, Sallis HM, Perry R, Ness AR, Churchill R. Omega-3 fatty acids for depression in adults. Cochrane Database Syst Rev 2015;11:CD004692.

[19] Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. J Am Stat Assoc 1977;72:320−38.

[20] Jennrich RI, Sampson PF. Newton-Raphson and related algorithms for maximum likelihood variance component estimation. Technometrics 1976;18:11−7.

[21] 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;17:152.

[22] Dearing JW. Applying diffusion of innovation theory to intervention development. Res Soc Work Pract 2009;19(5):503−18.

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

[24] 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.

[25] Millard LAC, Flach PA, Higgins JPT. Machine learning to assist risk-of-bias assessments in systematic reviews. Int J Epidemiol 2016;45:266−77.