Methods Research Report

Creating Efficiencies in the Extraction of Data From Randomized Trials: A Prospective Evaluation of a Machine Learning and Text Mining Tool

AHRQ
Agency for Healthcare
Research and Quality

Methods Research Report

Creating Efficiencies in the Extraction of Data From Randomized Trials: A Prospective Evaluation of a Machine Learning and Text Mining Tool

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 290-2015-00001-I

Prepared by:
University of Alberta Evidence-based Practice Center
Edmonton, Alberta, Canada

Investigators:
Allison Gates, Ph.D.
Michelle Gates, Ph.D.
Shannon Sim, M.Sc.
Sarah A. Elliott, Ph.D.
Jennifer Pillay, M.Sc.
Lisa Hartling, Ph.D.

AHRQ Publication No. 21-EHC006
August 2021

Key Messages

Purpose of study

For a sample of 75 randomized trials, we prospectively evaluated an online machine learning and text mining tool's ability to (a) automatically extract 21 unique data elements, and (b) save time compared with manual extraction and verification.

Key messages

- The tool identified the reporting (reported or not reported) of data elements more than 90 percent of the time for 52 percent of data elements (n = 11/21). For three (14%) data elements (route of administration, early stopping, secondary outcome time point), the tool correctly identified their reporting (reported or not reported) ≤50 percent of the time.
- Among the top five sentences presented for each solution, for 81 percent (n = 17/21) of data elements at least one sentence was relevant more than 80 percent of the time.
- For 83 percent (n = 15/18) of data elements, relevant fragments were highlighted among the relevant sentences more than 80 percent of the time.
- Fully correct solutions were common (>80%) for some data elements (first author name, data of publication, DOI, funding number, registration number, early stopping) but performance varied greatly (from 0% for eligibility criteria to 93% for early stopping).
- Using ExaCT to assist the first reviewer in a pair resulted in a modest time savings compared with manual extraction by one reviewer (17.9 hours compared with 21.6 hours, 17.1%).

This report is based on research conducted by the University of Alberta Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2015-00001-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help healthcare decision makers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of healthcare services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

AHRQ appreciates appropriate acknowledgment and citation of its work. Suggested language for acknowledgment: This work was based on a report, Creating Efficiencies in the Extraction of Data From Randomized Trials: A Prospective Evaluation of a Machine Learning and Text Mining Tool, by the Evidence-based Practice Center Program at the Agency for Healthcare Research and Quality (AHRQ).

Suggested citation: Gates A, Gates M, Sim S, Elliott SA, Pillay J, Hartling L. Creating Efficiencies in the Extraction of Data From Randomized Trials: A Prospective Evaluation of a Machine Learning and Text Mining Tool. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2015-00001-I.) AHRQ Publication No. 21-EHC006. Rockville, MD: Agency for Healthcare Research and Quality; August 2021. Posted final reports are located on the Effective Health Care Program search page. DOI: https://doi.org/10.23970/AHRQEPCMETHODSCREATINGEFFICIENCIES.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

If you have comments on this Methods Research Project they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

David Meyers, M.D.
Acting Director
Agency for Healthcare Research and Quality

Arlene Bierman, M.D., M.S.
Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Craig Umscheid, M.D., M.S.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Jill Huppert, M.D., M.P.H.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Acknowledgments

We thank Bernadette Zakher (Alberta Research Centre for Health Evidence, University of Alberta, Canada) for contributing to study selection, and Dr. Alex Aregbesola and Amanda Coyle (Children's Hospital Research Institute of Manitoba, University of Manitoba, Canada) for extracting the characteristics of the included randomized trials.

Creating Efficiencies in the Extraction of Data From Randomized Trials: A Prospective Evaluation of a Machine Learning and Text Mining Tool

Structured Abstract

Background. Machine learning tools that semi-automate data extraction may create efficiencies in systematic review production. We prospectively evaluated an online machine learning and text mining tool's ability to (a) automatically extract data elements from randomized trials, and (b) save time compared with manual extraction and verification.

Methods. For 75 randomized trials published in 2017, we manually extracted and verified data for 21 unique data elements. We uploaded the randomized trials to ExaCT, an online machine learning and text mining tool, and quantified performance by evaluating the tool's ability to identify the reporting of data elements (reported or not reported), and the relevance of the extracted sentences, fragments, and overall solutions. For each randomized trial, we measured the time to complete manual extraction and verification, and to review and amend the data extracted by ExaCT (simulating semi-automated data extraction). We summarized the relevance of the extractions for each data element using counts and proportions, and calculated the median and interquartile range (IQR) across data elements. We calculated the median (IQR) time for manual and semiautomated data extraction, and overall time savings.

Results. The tool identified the reporting (reported or not reported) of data elements with median (IQR) 91 percent (75% to 99%) accuracy. Performance was perfect for four data elements: eligibility criteria, enrolment end date, control arm, and primary outcome(s). Among the top five sentences for each data element at least one sentence was relevant in a median (IQR) 88 percent (83% to 99%) of cases. Performance was perfect for four data elements: funding number, registration number, enrolment start date, and route of administration. Among a median (IQR) 90 percent (86% to 96%) of relevant sentences, pertinent fragments had been highlighted by the system; exact matches were unreliable (median (IQR) 52 percent [32% to 73%]). A median 48 percent of solutions were fully correct, but performance varied greatly across data elements (IQR 21% to 71%). Using ExaCT to assist the first reviewer resulted in a modest time savings compared with manual extraction by a single reviewer (17.9 vs. 21.6 hours total extraction time across 75 randomized trials).

Conclusions. Using ExaCT to assist with data extraction resulted in modest gains in efficiency compared with manual extraction. The tool was reliable for identifying the reporting of most data elements. The tool's ability to identify at least one relevant sentence and highlight pertinent fragments was generally good, but changes to sentence selection and/or highlighting were often required.

Contents

Tables

Appendixes

Introduction

Timely systematic reviews provide an indispensable resource for decision makers, many of whom lack the time and expertise to independently identify and evaluate new evidence. To be useful, systematic reviews must be conducted with a high degree of methodological rigor, and are therefore time and resource intensive. A typical systematic review will take a highly skilled team of clinician-experts, methodologists, and statisticians many months or even years to complete.[1] Especially in rapidly evolving fields, it is no longer feasible for traditional systematic review production to keep pace with the publication of new trial data,[2] seriously undermining the currency, validity, and utility of even the most recently published reviews.

As the number of newly registered randomized trials continues to grow,[3] the need to create efficiencies in the production of systematic reviews is increasingly pressing. Living systematic reviews, which are continually updated as new evidence becomes available,[4] represent a relatively new form of evidence synthesis aimed at addressing the heavy workload and fleeting currency associated with most traditional systematic reviews. Because living systematic reviews are updated in real time, the total workload for keeping them up to date is broken down into more manageable tasks.[4] Since living systematic reviews are held to the same methodological standards as traditional systematic reviews, the efficiency of their production will be critical to their feasibility and sustainability.[4]

To date, nearly 200 software tools aimed at facilitating systematic review processes have been developed, with machine learning and text mining being the driver behind the proposed efficiencies of many tools.[5] Most research investigating the use of machine learning tools in systematic reviews has focused on creating efficiencies during the study selection step.[6,7] The body of research investigating technologies designed to assist with data extraction, one of the most time- and resource-intensive steps of completing a systematic review,[8,9] is comparatively immature.[7,10] Machine learning tools that automatically identify relevant text may expedite data extraction in a number of ways: as a first check for manual data extraction performed in duplicate; to validate data extraction by a single reviewer; as the primary source for data extraction that would be validated by a human; and eventually to completely automate data extraction.[7]

Among the tools that have been developed to semiautomate data extraction, few[11-13] prototypes have been made accessible for review teams to evaluate in practice.[10] Of the tools that are available, relatively few support semiautomated data extraction from full texts,[7,10] and published evaluations of those that do are sparse.[7] Independent evaluations are needed to validate the relevance of automatically extracted data and potential for time and resource savings associated with using machine learning tools to assist with data extraction in systematic reviews.

Objectives

We aimed to: (1) prospectively evaluate an online machine learning and text mining tool's ability to automatically extract relevant data from randomized trials and (2) estimate the time savings associated with potential approaches to semiautomated data extraction compared with manual extraction and verification by two reviewers.

Methods

Machine Learning and Text Mining Tool: ExaCT

ExaCT (prototype available at https://exact.cluster.gctools.nrc.ca/ExactDemo/intro.php) is an online machine learning and text mining tool integrated within an automatic information extraction engine.[13] The tool assists human reviewers by automatically extracting study characteristics (hereafter referred to as "data elements") from publications of randomized trials.[13] ExaCT was the first tool (and remains one of few tools) to automatically extract data from full text publications; various other tools extract data from abstracts only.[7,13] Details of the design and development of ExaCT, and an early evaluation of its performance were reported in a 2010 publication by the tool's developers.[13]

After creating an account, full texts can be uploaded to ExaCT's user interface in HTML format. Nearly instantaneously, the tool extracts 21 unique data elements, as identified in sentences from each full text document. For each data element, the tool presents "solutions" consisting of five potentially relevant sentences presented in descending order of confidence. The top scoring sentence is termed the "system suggestion." Text fragments (a word or group of words) that the system identifies as containing target information are highlighted within the retrieved sentences when the confidence score of those sentences exceeds a certain threshold. For each data element, the tool provides any of four responses: not found (i.e., data not reported and no relevant sentences); exactly one answer provided by one instance of text; one answer repeated in several instances of text; or several distinct answers. The tool allows users to view, confirm, refute, and modify the extracted sentences and text fragments.

Using a sample of 50 randomized trials published across 25 scientific journals, ExaCT's developers reported 80 percent precision (i.e., the proportion of returned instances that are truly relevant) and recall (i.e., the proportion of relevant instances returned by the system) for extracted sentences.[13] Of the top five candidate sentences, the human reviewers considered at least one to be relevant 93 percent of the time.[13] With respect to the highlighted text fragments, on average the tool performed with 93 percent precision and 91 percent recall. It required the human reviewer a mean 7 minutes and 21 seconds per trial publication to review ExaCT's extracted data and make any necessary amendments. The authors did not measure time savings compared with extraction by human reviewers, acknowledging that a large-scale usability study is required to verify actual gains in efficiency.[13] Time savings attributed to the tool would result mainly from the reviewers being automatically directed to potentially relevant segments of text, expediting the identification and extraction of relevant information.

Sample of Randomized Trials

We leveraged a random sample of randomized trials originally identified for an ongoing surveillance study that is underway at our center.[14,15] On February 19, 2020, our research librarian undertook a search in the Cochrane Central Register of Controlled Trials (Wiley) for all child-relevant randomized trials of health interventions published in 2017. Details of the search are in Appendix A. The search retrieved 17,703 potentially relevant citations, which we randomly ordered using the random numbers generator in Microsoft Excel. From the randomly ordered list, two independent reviewers (either of AG, MG, and SS) screened the titles and

abstracts to identify the first 75 randomized trials that reported on outcomes for participants aged 21 years or younger (unrestricted by condition, intervention, comparator, or outcome type). Any record marked as "include" or "unsure" by either reviewer was eligible for scrutiny by full text. Two reviewers (either of AG, MG, and SS) independently screened the full texts and agreed upon the included randomized trials.

We selected our sample size for feasibility with respect to time, resources, and available personnel. The sample used for this study should have zero overlap with the developers' test set, which included only randomized trials published in 2009.[13] There should also be no overlap with their training set, which included only randomized trials published before 2010.[13] One of two reviewers from a collaborating center extracted the study characteristics from each randomized trial.

Data Collection

Three reviewers completed the data extraction following a three stage process, summarized in Appendix B. First, using the random numbers generator in Microsoft Excel, each reviewer was randomized to manually extract data from one-third (n = 25) of the sample of randomized trials and to verify the extracted data for a different one-third (n = 25) of randomized trials. Next, for their original sample of randomized trials, each reviewer collected data about the relevance of ExaCT's automated extractions, as compared with their own verified extractions. Finally, for the remaining 25 randomized trials to which they were naïve (i.e., had not yet reviewed for the purpose of data extraction or verification), each reviewer prospectively simulated semi-automated data extraction in ExaCT to measure time savings. This three stage process allowed us to control for gains in efficiency that would result during the semiautomated extraction from being familiar with the randomized trials.

Prior to beginning formal extraction, all reviewers pilot tested the data extraction forms on three randomized trials and convened to ensure a mutual understanding of the form, data elements, and timing procedure.

A. Manual Extraction and Verification

For each randomized trial, the reviewers extracted ExaCT's standard 21 data elements to a Microsoft Excel spreadsheet: eligibility criteria, sample size (enrolled), start date of enrollment, end date of enrollment, name of experimental treatment(s), name of control treatment(s), dose (or number of sessions), frequency of treatment, route of treatment (or delivery method), duration of treatment, primary outcome name, primary outcome time point, secondary outcome name, secondary outcome time point, funding organization name, funding number, early stopping, registration number, author name, date of publication, and digital object identifier (DOI). A second reviewer verified the extraction. The reviewers used a digital chronograph to measure the amount of time required to extract the data and verify the extractions, to the nearest 5 seconds. The timing began when the reviewer started reading the full text to extract or verify the data elements, and ended when the final data element was extracted or verified.

For the purpose of this study, the data manually extracted by one reviewer and verified by another served as the reference standard. Although human reviewer judgment is imperfect,[16] dual independent extraction is recommended by leaders in evidence synthesis[17] and provided a reasonable standard for comparison.

B. Relevance of the Automated Extraction

For the same sample of randomized trials each reviewer reviewed the automatically extracted sentences and text fragments for each data element and judged the relevance of the sentences, highlighted text fragments, and overall solutions.

At the sentence level, for each data element the reviewers judged whether the top-ranked sentence was relevant (yes or no) and whether at least one sentence was relevant (even if it was not the top-ranked sentence; yes or no). At the fragment level, for each sentence that the reviewer considered relevant, they judged whether the highlighted text fragments were fully or at least partially relevant (yes or no).[13] Fully relevant fragments were those that encompassed all relevant information for the data element, without including additional irrelevant information or missing critical information. Partially relevant fragments were those that encompassed part of the relevant information, but either also included erroneous information or fell short of including all essential details. Appendix C shows examples of relevant and irrelevant sentences, and relevant, irrelevant, and partially relevant fragments.

To evaluate the relevance of the overall solutions, for each data element the reviewers recorded the number of fully relevant, partially relevant, and fully irrelevant solutions.[13] Solutions, which encompass both the extracted sentences and fragments, were considered fully relevant when the tool identified a sentence with the target information as its top sentence and extracted the relevant fragments, or the tool correctly reported the absence of a solution when the data element was not reported in the publication (i.e., returned a "not found" solution). Solutions were partially relevant when the correct solution was present among the five sentences, but not (only) in the top sentence and/or the fragment selection in the sentence(s) was not entirely relevant. Solutions were irrelevant when none of the five suggested sentences contained relevant information pertaining to the data element. Appendix C shows examples of fully relevant, partially relevant, and fully irrelevant solutions.

C. Time Savings

To measure the time saved by using ExaCT to assist with data extraction, the three reviewers examined the automatically extracted data elements and undertook necessary amendments, simulating a practical use of the tool. As with manual extraction, the reviewers used a digital chronograph to measure the time required to review and amend the automatically extracted data elements to the nearest 5 seconds. Timing began once the data extraction form was opened on the user interface and ended once all data elements were verified, revised, and downloaded.

Data Analysis

We synthesized the trial characteristics, the relevance of the extracted sentences, fragments, and overall solutions, and the timing data using descriptive statistics (counts, frequencies, median and interquartile range [IQR]). We compared the time to complete the manual data extraction and verification with the time to complete the semiautomated extraction and interpreted differences with respect to practical significance. We calculated the time savings for two potential uses of ExaCT: (a) to assist the first reviewer in a pair, and (b) to replace the first reviewer in a pair. We calculated time savings as follows:

If ExaCT were used to assist the first reviewer in a pair:

Time savings = (time the first reviewer spent manually extracting data from the randomized trials) – (time one reviewer spent reviewing and amending ExaCT's extractions).

Note that the time savings here applies only to the work of the first reviewer in a pair. For the purpose of this study, we have assumed that the work of the second reviewer (verification) would remain constant.

If ExaCT were used to replace the first reviewer in a pair:

Time savings = (time the two reviewers spent manually extracting and verifying data from the randomized trials) – (time one reviewer spent reviewing and amending ExaCT's extractions).

Results

Sample of Randomized Trials

The included randomized trials are listed in Appendix D and summary characteristics of the sample are in Table 1. Nearly all (n = 70/75, 93.3%) randomized trials were efficacy/superiority trials. Most randomized trials used either a parallel (n = 54/75, 72.0%) or cluster (n = 15/75, 20.0%) design. The most common interventions included drugs (n = 18/75, 24.0%), rehabilitation or psychosocial programs (n = 12/75, 16.0%), communication, organizational, or educational programs (n = 12/75, 16.0%), and medical devices (n = 11/75, 14.7%). Nearly half (n = 36/75, 48.0%) used an active control, 20.0 percent (n = 15/75) used a placebo, 20.0 percent (n = 15/75) used a no intervention control, and 12.0 percent used a wait-list control. The primary outcome was most commonly a measure of physiological (n = 22/75, 29.3%), behavioral (n = 16/75, 21.3%), or psychological (n = 13/75, 17.3%) health, or a biomarker (e.g., serum ferritin, glycosylated hemoglobin) (n = 12/75, 16.0%).

Table 1. Summary characteristics of the sample of trials (n = 75)

Characteristic	Category	n (%)
Study type	Efficacy/superiority	70 (93.3)
	Equivalence	4 (5.3)
	Noninferiority	1 (1.3)
Trial design	Parallel	54 (72.0)
	Cluster	14 (18.7)
	Crossover	3 (4.0)
	Split body	2 (2.7)
	Factorial	0 (0)
	Other	2 (2.7)
Intervention class	Drug	18 (24.0)
	Communication, organizational, or educational	12 (16.0)
	Rehabilitation or psychosocial	12 (16.0)
	Device	11 (14.7)
	Alternative therapeutic	7 (9.3)
	Prevention or screening	6 (8.0)
	Vaccine	3 (4.0)
	Surgery or radiotherapy	2 (2.7)
	Other	4 (5.3)
Control type	Active intervention	36 (48.0)
	No intervention	15 (20.0)
	Placebo	15 (20.0)
	Wait-list control	9 (12.0)
Primary outcome category	Physiological	22 (29.3)
	Behavioral	16 (21.3)
	Psychological	13 (17.3)
	Biomarker	12 (16.0)
	Techniques or training	5 (6.7)
	Quality of life	2 (2.7)
	Pain	1 (1.3)
	Other	4 (5.3)

A. Manual Extraction and Verification

On the basis of the human reviewers' manual extractions, the reporting of the 21 data elements varied across the randomized trials (Table 2). Eligibility criteria, sample size, the

experimental and control arms, and primary outcome(s) were reported in all 75 randomized trials. The primary outcome time point was reported in all but one randomized trial (n = 74/75, 98.7%). The funding source (n = 63/75, 84.0%), registration number (n = 52/75, 69.3%), enrolment start and end dates (n = 45/75, 60.0%), secondary outcome(s) (n = 55/75, 73.3%), and secondary outcome time point (n = 54/75, 72.0%) were reported in the majority of randomized trials. The funding number (n = 29/75, 38.7%) and early stopping (n = 4, 5.3%) were infrequently reported. Because of the nature of the interventions in this sample of randomized trials, the route of administration (n = 29/75, 38.7%) and dose (n = 37/75, 49.3%) were frequently irrelevant and not reported. The frequency (n = 43/75, 57.3%) and duration (n = 55/75, 73.3%) of the intervention were more frequently reported.

B. Relevance of the Automated Extraction

Relevance of the Extracted Sentences

The relevance of the automatically extracted sentences is in Table 2. For 19.0 percent (n = 4/21) of data elements (eligibility criteria, enrolment end date, control arm(s), and primary outcome(s)) ExaCT correctly identified a solution (i.e., returned that a reported data element was "found") for all randomized trials in which they were reported. For an additional 33.3 percent (n = 7/21) of data elements (first author name, date of publication, DOI, funding source, sample size, enrolment start date, and experimental arm[s]) solutions were identified for at least 90 percent of randomized trials in which they were reported. For an additional 23.8 percent (n = 5/21) of data elements (funding number, registration number, dose, duration of treatment, and secondary outcome[s]) solutions were identified for at least 75 percent of randomized trials in which they were reported. Solutions were less frequently correctly identified for the remaining 23.8 percent (n = 5/21) of data elements: early stopping (n = 2/4, 50.0%), route of administration (n = 14/29, 48.3%), frequency of administration (n = 28/43, 65.1%), primary outcome time point (n = 50/74, 67.6%), and secondary outcome time point (n = 23/54, 42.6%).

For data elements correctly identified as reported in the randomized trials, ExaCT provided five candidate sentences including a top sentence ("system suggestion"). The top sentence reported for the registration number and early stopping were relevant in all solutions, and for the funding number in 90.9 percent (n = 20/22) of solutions. For an additional 33.3 percent (n = 6/18) of data elements (the first author name, date of publication, DOI, enrolment start date, route of administration, and frequency of administration) the top sentence was relevant among at least 80 percent of solutions. For an additional 22.2 percent (n = 4/18) of data elements (funding source, enrolment end date, primary outcome[s], and secondary outcome[s]) the top sentence was relevant among at least 70 percent of solutions. The top sentence was less frequently relevant among the solutions for the remaining 44.4 percent (n = 8/18) of data elements: control arm(s) (n = 49/75, 65.3%), secondary outcome time point (n = 15/23, 65.2%), duration of treatment (n = 25/41, 61.0%), dose (n = 19/32, 59.5%), experimental arm(s) (n = 43/74, 58.1%), primary outcome time point (n = 27/50, 54.0%), eligibility criteria (n = 38/75, 50.7%), and sample size (n = 32/68, 47.1%).

At least one of the top five sentences was relevant among all solutions for 23.8 percent (n = 5/21) of data elements (funding number, registration number, enrolment start date, early stopping, and route of administration). For an additional 16.7 percent (n = 3/18) of data elements

(enrolment end date, frequency of administration, and secondary outcome[s]) at least one sentence was relevant across at least 90 percent of solutions. For an additional 27.8 percent (n = 5/18) of data elements (funding source, experimental arm(s), control arm(s), primary outcome(s), and secondary outcome time point) at least one sentence was relevant across at least 80 percent of solutions. For an additional 11.1 percent (n = 2/18) of data elements (duration of treatment and primary outcome time point) at least one sentence was relevant across at least 70 percent of solutions. At least one sentence was less frequently relevant among the solutions for the remaining 11.1 percent (n = 2/18) of data elements: eligibility criteria (n = 47/75, 62.7%) and sample size (n = 43/68, 63.2%).

Table 2. Relevance of the automatically extracted sentences

Report Section	Data Element	Reported in the Trial, n (%)[a]	Found by ExaCT, n (%)[b]	Relevance, Top Sentence, n (%)[c]	Relevance, Any Sentence, n (%)[c]	Relevant Sentences, n (%) of Total
Publication information	First author name	75 (100.0)	74 (98.7)	63 (85.1)	n/a	n/a
	Date of publication	75 (100.0)	74 (98.7)	64 (86.5)	n/a	n/a
	Digital object identifier	75 (100.0)	72 (96.0)	62 (82.7)	n/a	n/a
Meta information	Funding source	63 (84.0)	58 (92.1)	45 (77.6)	50 (86.2)	79/116 (68.1)
	Funding number	*29 (38.7)*	22 (75.9)	20 (90.9)	22 (100.0)	35/110 (31.8)
	Registration number	52 (69.3)	40 (76.9)	40 (100.0)	40 (100.0)	63/200 (31.5)
Enrollment	Eligibility criteria	75 (100.0)	75 (100.0)	*38 (50.7)*	*47 (62.7)*	*110/375 (29.3)*
	Sample size	75 (100.0)	68 (90.7)	*32 (47.1)*	*43 (63.2)*	125/340 (36.8)
	Enrolment start date	*45 (60.0)*	44 (97.8)	35 (79.5)	44 (100.0)	*55/220 (25.0)*
	Enrolment end date	*45 (60.0)*	45 (100.0)	35 (77.8)	44 (97.8)	*56/225 (24.9)*
	Early stopping	*4 (5.3)*	*2 (50.0)*	2 (100.0)	2 (100.0)	7/10 (70.0)
Intervention	Experimental arm(s)	75 (100.0)	74 (98.7)	*43 (58.1)*	65 (87.8)	123/370 (33.2)
	Control arm(s)	75 (100.0)	75 (100.0)	49 (65.3)	65 (86.7)	121/375 (32.3)
	Route of administration	*29 (38.7)*	*14 (48.3)*	12 (85.7)	14 (100.0)	32/70 (45.7)
	Dose	*37 (49.3)*	32 (86.5)	*19 (59.4)*	28 (87.5)	50/160 (31.3)
	Frequency of administration	*43 (57.3)*	*28 (65.1)*	23 (82.1)	27 (96.4)	45/140 (32.1)
	Duration of treatment	55 (73.3)	*41 (74.5)*	*25 (61.0)*	*30 (73.2)*	*57/205 (27.8)*
Outcome	Primary outcome(s)	75 (100.0)	75 (100.0)	53 (70.7)	62 (82.7)	*95/375 (25.3)*
	Primary outcome time point	74 (98.7)	*50 (67.6)*	*27 (54.0)*	*39 (78.0)*	76/250 (30.4)
	Secondary outcome(s)	55 (73.3)	44 (80.0)	33 (75.0)	40 (90.9)	75/220 (34.1)
	Secondary outcome time point	54 (72.0)	*23 (42.6)*	15 (65.2)	*19 (82.6)*	43/115 (37.4)
Summary measure	Median (IQR), n	**55 (45 to 75)**	**44 (23 to 68)**	**35 (23 to 45)**	**40 (23 to 44)**	**57 (37 to 91)**

Report Section	Data Element	Reported in the Trial, n (%)[a]	Found by ExaCT, n (%)[b]	Relevance, Top Sentence, n (%)[c]	Relevance, Any Sentence, n (%)[c]	Relevant Sentences, n (%) of Total
	Median (IQR), %	**73.3 (60.0 to 100.0)**	**90.7 (74.5 to 98.7)**	**77.6 (61.0 to 85.1)**	**87.7 (82.6 to 16.8**	**32.0 (29.3 to 36.1)**

IQR = interquartile range; n/a = not applicable (ExaCT presents only one solution for these elements). Values in *italics* typeface fall at or below the limit of the lowest quartile.
[a]As identified during manual data extraction and verification.
[b]Pertains to the studies where the data element was identified as reported in the study by the human reviewers (denominator, column 2).
[c]Pertains to the studies where the data element was correctly identified as reported in the study by ExaCT (denominator, column 3).

Relevance of the Highlighted Fragments

The relevance of the highlighted fragments within the relevant sentences is in Table 3. Seventy-nine percent (n = 124/157) of fragments for the funding source and 55.6 percent (n = 74/133) for the experimental arm(s) were considered relevant. For the remaining data elements, at least 81.5 percent of fragments were relevant.

For 16.7 percent (n = 3/18) of data elements (registration number and enrolment start and end date), more than 80 percent of fragments were exact matches. For an additional 22.2 percent (n = 4/18) of data elements (sample size, route of administration, frequency of administration, primary outcome[s]) more than 70 percent were exact matches. Exact matches were less frequent among the remaining 61.1 percent (n = 11/18) of data elements: duration of treatment (n = 35/56, 62.5%), control arm(s) (n = 34/62, 54.8%), funding number (n = 27/54, 50.0%), secondary outcome time point (n = 20/53, 37.7%), dose (n = 26/77, 33.8%), early stopping (n = 1/3, 33.3%), primary outcome time point (n = 28/76, 32.6%), secondary outcome(s) (n = 16/53, 30.2%), funding source (n = 24/157, 15.3%), and experimental arm(s) (n = 15/133, 11.3%). Partial matches were most common among fragments provided in relevant sentences for the funding source (n = 100/157, 63.7%), early stopping (n = 2/3, 66.7%), dose (n = 44/77, 57.1%), primary outcome time point (n = 50/86, 58.1%), secondary outcome(s) (n = 35/53, 66.0%), and secondary outcome time point (n = 31/53, 58.5%).

Table 3. Relevance of the highlighted text fragments among relevant sentences[a]

Report Section	Data Element	Relevant Sentences, n Total[b]	Fragments, n Total[c]	Relevant Fragments, n (%)[d]	Exact Matches, n (%)[d]	Partial Matches, n (%)[d]
Meta information	Funding source	79	157	*124 (79.0)*	*24 (15.3)*	100 (63.7)
	Funding number	35	54	*44 (81.5)*	27 (50.0)	17 (31.5)
	Registration number	63	104	104 (100.0)	103 (99.0)	*1 (1.0)*
Enrollment	Eligibility criteria	110	0	*0 (0.0)*	*0 (0.0)*	*0 (0.0)*
	Sample size	125	125	110 (88.0)	92 (73.6)	*18 (14.4)*
	Enrolment start date	55	51	51 (100.0)	43 (84.3)	8 (15.7)
	Enrolment end date	56	50	47 (94.0)	45 (90.0)	*2 (4.0)*
	Early stopping	7	3	3 (100.0)	1 (33.3)	2 (66.7)
Intervention	Experimental arm(s)	123	133	*74 (55.6)*	*15 (11.3)*	59 (44.4)
	Control arm(s)	121	62	55 (88.7)	34 (54.8)	21 (33.9)
	Route of administration	32	34	30 (88.2)	27 (79.4)	*3 (8.8)*
	Dose	50	77	70 (90.9)	26 (33.8)	44 (57.1)
	Frequency of administration	45	61	55 (90.1)	44 (72.1)	11 (18.0)

Report Section	Data Element	Relevant Sentences, n Total[b]	Fragments, n Total[c]	Relevant Fragments, n (%)[d]	Exact Matches, n (%)[d]	Partial Matches, n (%)[d]
	Duration of treatment	57	56	*48 (85.7)*	35 (62.5)	13 (23.2)
Outcome	Primary outcome(s)	95	78	74 (94.9)	55 (70.5)	19 (24.4)
	Primary outcome time point	76	86	78 (90.7)	*28 (32.6)*	50 (58.1)
	Secondary outcome(s)	75	53	51 (96.2)	*16 (30.2)*	35 (66.0)
	Secondary outcome time point	43	53	51 (96.2)	20 (37.7)	31 (58.5)
Summary measure	Median (IQR), n	**57 (54)**	**59 (33)**	**53 (47 to 74)**	**27 (15 to 41**	**18 (4 to 34)**
	Median (IQR), %	-	-	**90.4 (86.2 to 96.9)**	**52.4 (32.8 to 73.2)**	**28.0 (14.7 to 57.9)**

[a]ExaCT does not provide fragments for publication information. Data are shown for the remaining 18 data elements. Values in *italics* typeface fall at or below the limit of the lowest quartile.
[b]Across all 75 trials, the number of relevant sentences among the 5 sentences reported within the solution for each data element.
[c]Contained within sentences considered to be relevant by the human reviewers (column 2).
[d]Relevant fragments of those contained within sentences considered to be relevant by the human reviewers (denominator, column 3).

Overall Relevance of the Extracted Solutions

The overall relevance of the solutions is in Table 4. Across data elements a median (IQR) 36 (16 to 53) (48.0% [21.3% to 70.7%]) of all solutions (of a total 75 solutions for each data element across the randomized trials) were considered fully relevant, 22 (12 to 38) (29.3% [16.0% to 50.7%]) were considered partially relevant, and 13 (10 to 22) (17.3% [13.3% to 29.3%]) were considered fully irrelevant.

More than 80 percent of solutions were fully relevant for 29 percent (n = 6/21) of data elements: first author name, date of publication, DOI, registration number, and early stopping. The data elements for which the solutions were least frequently fully relevant included: control arm (n = 16/75, 21.3%), funding source (n = 16/75, 21.3%), secondary outcome time point (n = 15/75, 20.0%), experimental arm (n = 10/75, 13.3%), primary outcome time point (n = 7/75, 9.3%), and eligibility criteria (n = 0/75, 0.0%).

Accounting for both fully and partially relevant solutions, a median (IQR) 82.7 percent (70.7% to 86.7%) were at least partially relevant. More than 80 percent of solutions were at least partially relevant for 57.1 percent (n = 12/21) of data elements: first author name, date of publication, DOI, funding number, registration ID, eligibility criteria, sample size, early stopping, experimental arm(s), control arm(s), route of administration, and primary outcome(s). More than 70 percent of solutions were at least partially relevant for an additional 19.0 percent (n = 4/21) of data elements: funding source, dose, frequency of administration, and secondary outcome(s). For the remaining 23.8 percent (n =5/21) of data elements, solutions that were at least partially relevant were less frequent: enrolment end date (n = 50/75, 66.7%), enrolment start date (n = 49/75, 65.3%), primary outcome time point (n = 49/75, 65.3%), duration of treatment (n = 47/75, 62.7%), secondary outcome time point (n = 42/75, 56.0%).

Table 4. Relevance of the extracted solutions

Report Section	Data Element	Fully Relevant Solutions, n (%)[a]	Partially Relevant Solutions, n (%)[a]	Fully Irrelevant Solutions, n (%)[a]
Publication information	First author name	63 (84.0)	*0 (0.0)*	12 (16.0)
	Date of publication	64 (85.3)	*0 (0.0)*	11 (14.7)

Report Section	Data Element	Fully Relevant Solutions, n (%)[a]	Partially Relevant Solutions, n (%)[a]	Fully Irrelevant Solutions, n (%)[a]
	Digital object identifier	62 (82.7)	*0 (0.0)*	13 (17.3)
Meta information	Funding source	*16 (21.3)*	38 (50.7)	21 (28.0)
	Funding number	52 (69.3)	13 (17.3)	*10 (13.3)*
	Registration number	62 (82.7)	*5 (6.7)*	*8 (10.7)*
Enrollment	Eligibility criteria	*0 (0.0)*	66 (88.0)	*9 (12.0)*
	Sample size	31 (41.3)	38 (50.7)	*6 (8.0)*
	Enrolment start date	34 (45.3)	15 (20.0)	26 (34.7)
	Enrolment end date	37 (49.3)	13 (17.3)	25 (33.3)
	Early stopping	70 (93.3)	*3 (4.0)*	*2 (2.7)*
Intervention	Experimental arm(s)	*10 (13.3)*	56 (74.7)	*9 (12.0)*
	Control arm(s)	*16 (21.3)*	49 (65.3)	*10 (13.3)*
	Route of administration	53 (70.7)	*12 (16.0)*	*10 (13.3)*
	Dose	36 (48.0)	22 (29.3)	17 (22.7)
	Frequency of administration	39 (52.0)	14 (18.7)	22 (29.3)
	Duration of treatment	22 (29.3)	25 (33.3)	28 (37.3)
Outcome	Primary outcome(s)	38 (50.7)	24 (32.0)	13 (17.3)
	Primary outcome time point	*7 (9.3)*	42 (56.0)	26 (34.7)
	Secondary outcome(s)	23 (30.7)	31 (41.3)	21 (28.0)
	Secondary outcome time point	*15 (20.0)*	27 (36.0)	33 (44.0)
Summary measure	Median (IQR), n	**36 (16 to 53)**	**22 (12 to 38)**	**13 (10 to 22)**
	Median (IQR), %	**48.0 (21.3 to 70.7)**	**29.3 (16.0 to 50.7)**	**17.3 (13.3 to 29.3)**

Values in *italics* typeface fall at or below the limit of the lowest quartile.

[a]Out of a total 75 solutions per data element (i.e., one solution per data element, per trial). Partially correct solutions were those that included relevant information but either also included erroneous information or fell short of including all essential details.

C. Time Savings

It took the reviewers a median (IQR) 16.4 (14.3 to 19.8) minutes to manually extract the data from each randomized trial and an additional 8.0 (6.4 to 10.0) minutes for the second reviewer to complete the verification. The combined time to manually extract and verify the data from each randomized trial was a median (IQR) 24.7 (21.2 to 29.4) minutes. Overall, we spent 21.6 hours manually extracting and 10.7 hours verifying data from the 75 randomized trials, for a total workload of 32.3 hours.

It took the reviewers a median (IQR) 13.8 (11.0 to 17.6) minutes to review and amend the automated extractions. This equates to a median 2.6 minutes faster compared with manual extraction by a single reviewer. Overall, we spent a total of 17.9 hours extracting data from the 75 randomized trials with the assistance of ExaCT.

In the context of using the tool to assist the first reviewer in a pair (i.e., to expedite the first reviewer's extractions), this equates to a median 3.7 hours less time spent extracting data compared with manual extraction (17.9 hours versus 21.6 hours, 17.1% time savings across 75 randomized trials). The verification time (for the second reviewer, not measured) we assume, would remain constant. In the context of using the tool to replace the first reviewer in a pair (i.e., as a primary source for data extraction that would be validated by a human reviewer) this equates to a median 14.4 hours less time spent extracting and verifying data compared with manual extraction and verification (17.9 hours versus 32.3 hours, 44.6% time savings across 75 randomized trials).

Discussion, Limitations, and Conclusion

Across a sample of 75 randomized trials, ExaCT correctly identified the reporting (reported or not reported) of data elements more than 90 percent of the time for 52 percent of data elements (n = 11/21). For three (14%) data elements (route of administration, early stopping, secondary outcome time point), the tool correctly identified their reporting (reported or not reported) 50 percent of the time or less. Among the top five sentences presented for each solution, for 81 percent (n = 17/21) of data elements at least one sentence was relevant more than 80 percent of the time. For the remaining four data elements (eligibility criteria, sample size, duration of intervention, primary outcome time point) the relevance of the top five sentences was considerably less. For 83 percent (n = 15/18) of data elements, relevant fragments were highlighted among the relevant sentences more than 80 percent of the time. For the remaining three data elements (funding source, eligibility criteria, experimental arm) the highlighted fragments were more often irrelevant. Fully correct solutions were common (>80%) for some data elements (first author name, data of publication, DOI, funding number, registration number, early stopping) but performance varied greatly (from 0% for eligibility criteria to 93% for early stopping). Solutions were most frequently (>30%) fully irrelevant for enrolment start and end date, duration of treatment, and primary and secondary outcome time points. Using ExaCT to assist the first reviewer in a pair resulted in a modest time savings compared with manual extraction by one reviewer (17.9 hours compared with 21.6 hours, 17.1%). The time saved applies only to the small proportion of data elements that are typically extracted from randomized trials in the context of a systematic review, and only to the work of the first reviewer in a pair.

Our findings extend those published by Kiritchenko et al. in 2010.[13] We are not aware of any other published evaluations of the ExaCT prototype. For a sample of 50 drug trials, Kiritchenko et al. reported 80 percent precision (the proportion of returned instances that are truly relevant) and recall (the proportion of relevant instances returned by the system) for the system suggestion (top sentence); among 93 percent of solutions, at least one of the top five sentences was relevant.[13] Performance was substantially poorer only for the funding source, eligibility criteria, and primary outcome time point.[13] Precision and recall were more than 90 percent for extracted fragments. Across data elements, the solutions were fully correct 66 percent of the time.[13] We anticipated that performance in our evaluation would be poorer, given that the system was trained only on drug trials[13] and our sample consisted of randomized trials unrestricted by intervention (only 24% were drug trials). We presumed, then, that the tool would have greater difficulty correctly identifying the experimental arm and details of the intervention (e.g., frequency of administration, route of administration). Indeed, we found that the top sentence was relevant across a median 78 percent of solutions, but results varied greatly across data elements (from 47% for the sample size to 100% for registration number and early stopping). Remarkably, performance was relatively similar for the top five sentences (relevant across a median 88% of solutions) and extracted fragments (relevant across a median 90% of relevant sentences). Solutions were considered fully correct with lesser frequency, likely because the top sentence was less often correct (48% vs. 66%).

Our findings suggest that using ExaCT to assist the first reviewer may be slightly more efficient than manual extraction by a single reviewer; however, before adopting semi-automated approaches to data extraction, gains in efficiency must be weighed against usability and the

accuracy of the extractions. As we have demonstrated, substantially more time could be saved if the automated extractions could be used to fully replace the first reviewer; however, many review teams may not be comfortable adopting this approach. The majority of solutions required at least some editing (to sentence selection, the highlighted fragments, or both); thus, the automated extractions are likely not a suitable replacement for the first reviewer. Time was saved because the reviewers were often more quickly able to identify the location of relevant data in the full texts; however, the process otherwise often resembled manual extraction because the reviewers needed to add relevant data or make amendments based on what was found in the text. Reviewers must also account for the fact that the automated extractions were reflective only of information contained within the source document. Typically, reviewers would ensure the completeness of the extraction by using multiple sources, including the trial registry, associated publications, and supplementary files to complete the extraction.[17] As this is a common issue among automated data extraction systems,[12,18] to support their utility more sophisticated systems that can incorporate data from multiple sources per randomized trial will be required.

Strengths and Limitations

To our knowledge this is the first study to externally and prospectively evaluate the performance of the ExaCT tool. We tested the tool on a heterogeneous sample of randomized trials of pediatric health interventions published during a one year period. As all of the randomized trials in the sample were published relatively recently (2017), the performance of the tool on older randomized trials (which presumably would be less well reported) may be worse. The findings may also not be generalizable to randomized trials in specific clinical areas. Although time was saved when ExaCT was used to assist with data extraction, the efficiency gained applies only to a small proportion of the data typically extracted from randomized trials for the purpose of a systematic review. The automatically extracted data elements are also arguably those more quickly and easily manually identified and extracted (e.g., compared with outcome data, for which identification and extraction is often more complex).

We did not formally evaluate the accuracy and completeness of the semiautomated data extractions compared with those manually extracted by the reviewers. As the accuracy and completeness of the extracted data have important implications with respect to the results and conclusions of systematic reviews, evaluations directly comparing manually and semiautomatically extracted data will help to inform how ExaCT and similar tools may most reliably be used. Specifically, it may be interesting to know whether the accuracy and completeness of the semi-automated extractions are more similar to a single reviewer's manual extractions or to data manually extracted by one reviewer and verified by another. This would inform whether the tool may be better used to assist or fully replace the first reviewer in a pair.

Conclusions

In this prospective evaluation, using ExaCT to assist the first reviewer in a pair to extract data from randomized trials was slightly more efficient compared with manual extraction. The tool was reliable for identifying the reporting (reported or not reported) of most data elements; however, the relevance of the system suggestion (top sentence) varied substantially across data elements. Among the top five sentences presented for each solution, for 81 percent of data elements at least one sentence was relevant more than 80% of the time. For 83 percent of data

elements, relevant fragments were highlighted among the relevant sentences more than 80 percent of the time. Fully correct solutions were relatively infrequent for most data elements, with the exception of first author name, date of publication, DOI, funding number, registration number, early stopping. For other data elements, changes to sentence selection or the highlighted fragments were often required.

References

1. 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. PMID: 28242767.
2. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;**7**(9): e1000326. doi: 10.1371/journal.pmed.1000326. PMID: 20877712.
3. U.S. National Library of Medicine. Trends, charts, and maps [Internet]. 2020. Available from: https://clinicaltrials.gov/ct2/resources/trends. Accessed 27 July 2020.
4. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. J Clin Epidemiol. 2017;91:23-30. doi: 10.1016/j.jclinepi.2017.08.010. PMID: 28912002.
5. Marshall C. SR Tool Box [Internet]. 2020. Available from: http://systematicreviewtools.com/about.php. Accessed 1 March 2020.
6. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev. 2015;4(1):5. doi: 10.1186/2046-4053-4-5. PMID: 25588314.
7. Jonnalagadda SR, Gowal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Syst Rev. 2015;4(1):78. doi: 10.1186/s13643-015-0066-7. PMID: 26073888.
8. Tsertsvadze A, Chen Y-F, Moher D, Sutcliffe P, McCarthy N. How to conduct systematic reviews more expeditiously? Syst Rev. 2015;4(1):160. doi: 10.1186/s13643-015-0147-7. PMID: 26563648.
9. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev. 2014;**3**(1):74. doi: 10.1186/2046-4053-3-74. PMID: 25005128.
10. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev. 2019;8(1):163. doi: 10.1186/s13643-019-1074-9. PMID: 31296265.
11. Marshall IJ, Wallace BC. Automating biomedical evidence synthesis: RobotReviewer. Proc Conf Assoc Comput Linguist Meet. 2017:7-12. doi: 10.18653/v1/P17-4002. PMID: 29093610.
12. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc. 2015;23(1):193-201. doi: 10.1093/jamia/ocv044. PMID: 26104742.
13. Kiritchenko S, de Bruijn B, Carini S, Martin J, Sim I. ExaCT: automatic extraction of clinical trial characteristics from journal publications. BMC Med Inform Decis Mak. 2010;10(1):56. doi: 10.1186/1472-6947-10-56. PMID: 20920176.
14. Gates A, Hartling L, Vandermeer B, Caldwell P, Contopoulos-Ioannidis DG, Curtis S, et al. The conduct and reporting of child health research: an analysis of randomized controlled trials published in 2012 and evaluation of change over 5 years. J Pediatr. 2018;193:237-244.e37. doi: 10.1016/j.jpeds.2017.09.014. PMID: 29169611.
15. Hamm MP, Hartling L, Milne A, Tjosvold L, Vandermeer B, Thomson D, et al. A descriptive analysis of a representative sample of pediatric randomized controlled trials published in 2007. BMC Pediatr. 2010;10:96. doi: 10.1186/1471-2431-10-96. PMID: 21176224.

16. Mathes TP, 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. doi: 10.1186/s12874-017-0431-4. PMID: 29179685.

17. Li T, Higgins JPT, Deeks JJ (editors). Chapter 5: Collecting data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MF, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available at: www.training.cochrange.org/handbook.

18. Gates A, Vandermeer B, Hartling L. Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool. J Clin Epidemiol. 2018;96:54-62. doi: 10.1016/j.jclinepi.2017.12.015. PMID: 29289761.

Abbreviations and Acronyms

AHRQ	Agency for Healthcare Research and Quality
DOI	Digital object identifier
EPC	Evidence-based Practice Center
HTML	Hypertext markup language
ID	Identification number
IQR	Interquartile range

Appendix A. Search Strategy

Database	Date Searched	Number Retrieved
Cochrane Central Register of Controlled Trials (Wiley)	February 19, 2020	17703
Total		**17703**

Strategy

#1 (Infant* or infancy or Newborn* or Baby* or Babies or Neonat* or Preterm* or Prematur* or Postmatur* or Child* or Schoolchild* or School age* or Preschool* or Kid or kids or Toddler* or Teen* or Boy* or Girl* or Minors* or Pubert* or Pubescen* or Prepubescen* or Pediatric* or Paediatric* or Peadiatric* or Nursery school* or Kindergar* or Primary school* or Secondary school* or Elementary school* or High school* or Highschool*):ti,ab,kw

#2 Adolesc*:ti,ab

#3 (Infant or Child or Minors or Puberty or Pediatrics or Schools):kw

#4 #1 or #2 or #3

#5 adolescent*:kw

#6 (adolescent* and (adult* or elderly or "middle aged" or "aged, 80 and over")):kw

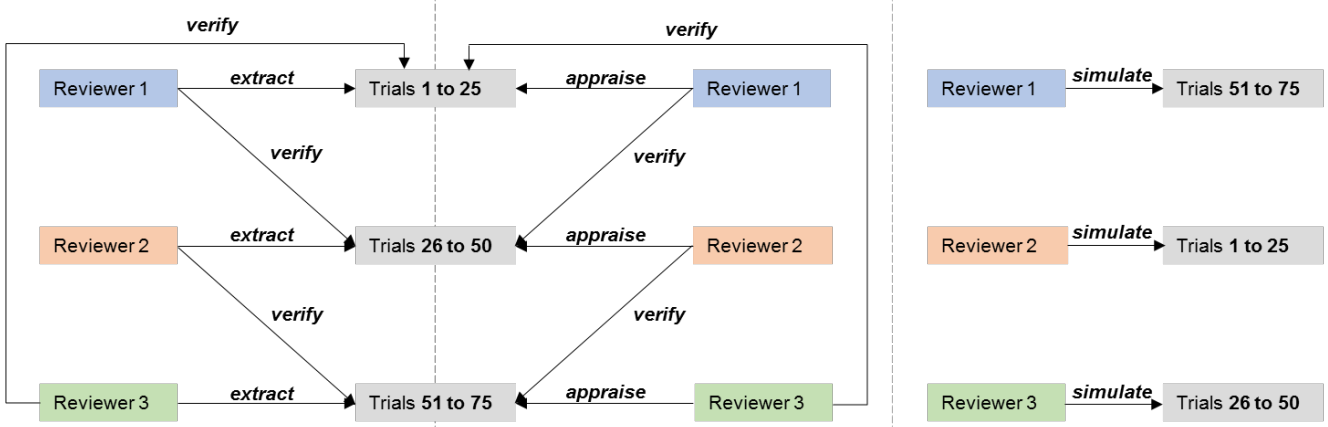
#7 #6 not #4

#8 #4 or #5

#9 #8 not #7 with Publication Year from 2017 to 2017, in Trials

Appendix B. Summary of the Data Extraction and Analysis Protocol

A. Manual extraction and verification | **B. Appraisal of the automated extractions** | **C. Simulated semi-automated extraction**



Protocol steps:

1. Each reviewer extracted data from 25 trials and verified another reviewer's extractions
2. The reviewers measured the time to extract and verify data from each trial to the nearest 5 seconds
3. For the same sample of trials from which they extracted data, each reviewer judged the sentence, fragment, and overall relevance of the automated extractions
4. A second reviewer verified the judgements
5. For the sample of trials to which they were naïve, each reviewer reviewed and revised the automated extractions
6. The reviewers measured the time to review and revise the extractions for each trial to the nearest 5 seconds

Time savings:

1. *If ExaCT were used to assist the first of two reviewers:*

Time savings = (time it took one reviewer to extract data from the trial manually) – (time it took one reviewer to review and revise the extractions in ExaCT)
Note: the time saved applies only to the work of the first reviewer

2. *If ExaCT were used to replace the second reviewer:*

Time savings = (time it took one reviewer to extract and one to verify data from the trial manually) – (time it took one reviewer to review and revise the extractions in ExaCT)

Appendix C. Examples of Relevant and Irrelevant Sentences, Fragments, and Solutions

Quality of the Extracted Sentences

At the sentence level, for each data element the reviewers judged whether the top-ranked sentence was relevant (yes or no) and whether at least one sentence was relevant (even if it was not the top-ranked sentence; yes or no).

Example of a relevant top-ranked sentence (Aboud 2017)

Human reviewer's verified extraction: "Iodized salt as early as it was available"

ExaCT's extraction:

Sentence Ranking	Sentence
1	The study used a cluster randomized design in which clusters were defined as districts randomly assigned to receive iodized salt as early as it was available (intervention) or when market forces brought it in (control).
2	Sixty district clusters were randomized to receive iodized salt early at their markets with assistance from regular salt distributors or later as introduced by market forces.
3	The hypothesis was that iodized salt would enhance mental development outcomes for intervention children in rural villages randomized to receive it early compared with control children.
4	The intervention group received iodized salt provided by private companies around Lake Afdera, north east of Amhara.
5	Consequently the intervention group received approximately 4 to 6 months more exposure to iodized salt than controls; intervention children had approximately 8 to 10 months of iodized salt and control children had 4 to 6 months at endline.

Explanation: the top-ranked sentence is relevant because it contains correct information about the experimental arm.

Example of an irrelevant top-ranked sentence (Guven 2017)

Human reviewer's verified extraction: "Various pulpotomy medicaments: (1) BD placed in pulp chamber and allowed to finish setting completely, followed by permanent restoration on the same session; (2) MTA-P + glass ionomer base place over the MTA; (3) PR-MTA placed in pulp chamber and condensed lightly with a moisted cotton pellet + glass ionomer base applied over the MTA; (4) 20% FS solution applied onto pulp stumps for 15s and after rinsing with water, ZOE base was placed."

ExaCT's extraction:

Sentence Ranking	Sentence
1	Success Rates of Pulpotomies in Primary Molars Using Calcium Silicate-Based Materials: A Randomized Control Trial
2	This randomized clinical trial was conducted to examine and compare the effectiveness of pulpotomy in primary molars treated with calcium silicate-based materials: two MTA products (i.e., ProRoot MTA and MTA-P) and BD and FS as the control material.
3	At 12 months, PCA was observed in two teeth treated with BD and one tooth treated with MTA-P; those teeth showed PCO at 24 months as well.
4	K.C. Huth, E. Paschos, N. Hajek-Al-Khatar et al. Effectiveness of 4 pulpotomy techniques – Randomized controlled trial. Journal of Dental Research, vol. 84, no. 12, pp. 1144-1148, 2005.
5	The aim of this study was to evaluate and compare, both clinically and radiographically, the effects of calcium silicate-based materials (i.e., ProRoot MTA (PR-MTA), MTA-Plus (MTA-P), and Biodentine (BD)) and ferric sulfate (FS) in pulpotomy of primary molars.

Explanation: the top-ranked sentence is irrelevant because it does not contain information about the nature of the experimental arms.

Example of at least one relevant sentence (Papathomas 2017)

Human reviewer's verified extraction: "Dialogic argumentation curriculum of four topics, including six electronic dialogs in peer pairs per topic, with an individual electronic dialog at the end of the school year. During topics 2, 3, and 4, at three of the six dialog sessions (1st, 3rd, and 5th), an adult substituted for an opposing peer pair."

ExaCT's extraction:

Sentence Ranking	Sentence
1	Participants were randomly assigned to an experimental group or a comparison group (24 studies each, balanced across gender and drawn equally from the two classrooms).
2	During two class periods per week throughout the school year, all sixth graders at the school participated in a dialogic argumentation curriculum very similar to the one reported on by Kuhn and Crowell (2011) and Kuhn, Hemberger, and Khait (2016).
3	As such, the role of the other (or others in discourse involving more than two people) assumes a prominent place.
4	Constructivists may claim that argumentation serves cognitive development by providing engagement and practice that allows both participants to develop their reasoning skills via shared exercise.
5	In this case, in addition to the exercise provided by participation alone, will this interaction also offer the less capable participant benefit in terms of skill development?

Explanation: the top-ranked sentence is irrelevant, but the second sentence contains correct information about the experimental arm.

Quality of the Extracted Fragments

At the fragment level, for each sentence that the reviewer considered relevant, they judged whether the highlighted text fragments were fully or at least partially relevant (yes or no). Fully relevant fragments were those that encompassed the full solution for the data element, without including additional irrelevant information or missing critical information. Partially relevant fragments were those that encompassed part of the solution, but either also included erroneous information or fell short of including all essential details.

Example of a fully relevant fragment in a relevant sentence (Freedman 2017)

Human reviewer's verified extraction: "In-clinic feeding of Impact(R) Peptide 1.5 (Nestle Health Science Inc) through digestive cartridge"

ExaCT's extraction:

Sentence Ranking	Sentence
1	After a 7-day washout period when participants received Peptamen 1.5 with their usual dose of PERT products but without digestive cartridge, participants crossed over and received EN through the opposite cartridge (placebo or digestive cartridge).
2	Participants were randomized to first receive EN through either digestive cartridge or placebo cartridge.
3	Increased Fat Absorption From Enteral Formula Through an In-line Digestive Cartridge in Patients With Cystic Fibrosis
4	Patients with CF receiving EN participated in a multicenter, randomized, double-blind, crossover trial with an open-label safety evaluation period.
5	Efficacy data were analyzed using differences in the plasma FA concentrations for 24 hours after a single EN feeding administered through either digestive cartridge or placebo cartridge.

Explanation: the first two sentences are relevant because they contain correct information about the experimental arm. The fragment in the first sentence is fully relevant because it contains all necessary information about the data element, without including additional irrelevant information or missing critical information.

Example of a partially relevant fragment in a relevant sentence (Freedman 2017)

Human reviewer's verified extraction: "In-clinic feeding of Impact(R) Peptide 1.5 (Nestle Health Science Inc) through digestive cartridge"

ExaCT's extraction:

Sentence Ranking	Sentence
1	After a 7-day washout period when participants received Peptamen 1.5 with their usual dose of PERT products but without digestive cartridge, participants crossed over and received EN through the opposite cartridge (placebo or digestive cartridge).
2	Participants were randomized to first receive EN through either digestive cartridge or placebo cartridge.
3	Increased Fat Absorption From Enteral Formula Through an In-line Digestive Cartridge in Patients With Cystic Fibrosis
4	Patients with CF receiving EN participated in a multicenter, randomized, double-blind, crossover trial with an open-label safety evaluation period.
5	Efficacy data were analyzed using differences in the plasma FA concentrations for 24 hours after a single EN feeding administered through either digestive cartridge or placebo cartridge.

Explanation: the first two sentences are relevant because they contain correct information about the experimental arm. The fragment in the second sentence is partially relevant because it contains relevant information about the experimental arm, but is missing important details (i.e., the name of the enteral nutrition product).

Example of a fully irrelevant fragment in a relevant sentence (Razi 2017)

Human reviewer's verified extraction: "Budesonide (1 mg/2 ml) with salbutamol nebules (0.15mg/kg/ dose, max. 5mg) driven by 100% oxygen at a flow of 6 L/min at 0, 20, and 40 min. Note: Both groups received one dose of intramuscularly methylprednisolone (1mg/kg/dose) at the onset of the treatment and salbutamol nebules at 80, 120, and 180 min"

ExaCT's extraction:

Sentence Ranking	Sentence
1	Children in the active treatment group were administered three doses of budesonide (1 mg/2 ml) with salbutamol nebules (0.15 mg/kg/dose, max. 5 mg) driven by 100% oxygen at a flow of 6 L/min at 0, 20, and 40 min and children in the control group received three doses of normal saline (2 ml) as a placebo as well as salbutamol.
2	In a study, which was conducted on 150 children who have moderate acute asthma exacerbation to examine the effects of different inhaled fluticasone doses, the authors did not demonstrate any improvement of SaO2 and PEF in the group who received three doses of salbutamol plus two doses of fluticasone 500 mcg/dose at 15 and 30 min after the first dose of salbutamol (accumulated dose of fluticasone = 1,000 mcg).
3	These beneficial effects have been reported only when patients received multiple ICS doses along with beta 2 agonists when compared with SCSs or placebo.
4	The object of this study was to determine whether high doses (total of 3 mg) of inhaled budesonide provide any additional benefits to a standardized treatment regimen that includes systemic steroids and salbutamol in preschool children who admitted to the ED with acute wheezing episodes.
5	The compared the effect of 1,500 g nebulized budesonide when added to standard acute asthma treatment (three doses of salbutamol, three doses of ipratropium bromide, and a single 2 mg/kg dose of prednisolone given at the beginning of therapy).

Explanation: the first and fourth sentences are relevant because they contain correct information about the experimental arm. The fragment in the first sentence is fully irrelevant because it does not contain any relevant information about the experimental arm.

Overall Quality of the Extracted Solutions

'Solutions', which encompass both the extracted sentences and fragments, were considered fully relevant when the system identified a sentence with the target information as its top sentence and extracted the relevant fragments, or

the system correctly reported the absence of the solution when it was not reported in the publication (i.e., returned a 'not found' solution). Solutions were partially relevant when the correct solution was present among the five sentences, but not (only) in the top sentence and/or the fragment selection in the sentence(s) was not entirely relevant. Solutions were irrelevant when none of the five suggested sentences contained relevant information pertaining to the data element.

Example of a fully relevant solution (Indrio 2017)

Human reviewer's verified extraction: "Freeze-dried L. reuteri DSM 17938 supplementation"

ExaCT's extraction:

Sentence Ranking	Sentence
1	Preterm newborns were randomly assigned to receive L. reuteri DSM 17938 supplementation or placebo by the use of a computer-generated randomization scheme.
2	Newborn were randomly allocated during the first 48 h of life to receive either daily probiotic (10.8 colony forming units (CFUs) of L. reuteri DSM 17938) or placebo for one month.
3	A total of 60 preterm newborns were randomly assigned to L. reuteri DSM 17938 or to the placebo group.
4	Efficacy and safety of available treatments for visceral leishmanlasis in Brazil: A multicenter, randomized, open label trial.
5	Women's education level amplifies the effects of a livelihoods-based intervention on household wealth, child diet, and child growth in rural Nepal

Explanation: the solution is fully correct because the top sentence contains correct information about the experimental arm, and the highlighted fragment contains all necessary information about the data element, without including additional irrelevant information or missing critical information.

Example of a fully relevant solution (Papathomas 2017)

Human reviewer's verified extraction: "Not reported"

ExaCT's extraction:

Sentence ranking	Sentence
0	**Funder:** not found
1	The transfer task data are, of course, critical to the current study.
2	The productivity of socially shared cognition is observed in the current study.
3	Thus, the study supports the concept of apprenticeship as a mechanism of development of higher-order and specifically argumentative reasoning skills, a conclusion of both theoretical and applied significance.
4	In this case, in addition to the exercise provided by participation alone, will this interaction also offer the less capable participant benefit in terms of skill development?
5	Much of what apprentices acquire through their participation is intellectual in nature knowledge and understanding.

Explanation: the solution is fully correct because the human reviewers did not identify the data element in the text, and ExaCT returned a 'not found' solution.

Example of a partially relevant solution (Laskin 2017)

Human reviewer's verified extraction: "More frequent, shorter hemodialysis (5 days per week for 2h25min, total=12 hours)"

ExaCT's extraction:

Sentence ranking	Sentence
1	Subjects were randomized to the treatment sequence: either 3 days per week conventional HD for four hours per treatment (12 hours total per week, control) followed by 5 days per week short, more frequent HD for two hours and 25 minutes per treatment (12 hours total per week, intervention), or vice versa.
2	Subjects were then randomized to the first treatment sequence (5 days versus 3 days per week HD, each for 12 hours total per week), during which three casual blood pressure measurements were obtained and averaged prior to each HD session, three days per week, regardless of which treatment frequency they were receiving.
3	In adults, the Frequent Hemodialysis Network (FHN) randomized trial demonstrated that an average of two extra in-center HD treatments was associated with improved survival and left ventricular hypertrophy.
4	Two subjects withdrew from the study early, both during 3 days per week treatment; one subject had completed 5 days per week HD but did not want to continue four hour treatments during the 3 days per week period and the second subject received a kidney transplant.
5	Short, frequent, 5 days per week in-center hemodialysis versus 3 days per week treatment: a randomized cross-over pilot trial through the Midwest Pediatric Nephrology Consortium

Explanation: the solution is partially correct because the top sentence contains correct information about the experimental arm, but the highlighted fragment is relevant to the control arm, not the experimental arm.

Example of a partially relevant solution (Gerceker 2017)

Human reviewer's verified extraction: "Care bundle with 3M™ Tegaderm™ Chlorhexidine Gluconate (CHG) I.V. Securement Dressings"

ExaCT's extraction:

Sentence Ranking	Sentence
1	This study included 27 patients; 14 patients were randomly assigned to intervention and 13 patients to the control group.
2	To compare the effects of the care bundles including chlorhexidine dressing and advanced dressings on the catheter-related bloodstream infection (CRBSI) rates in pediatric hematology-oncology patients with central venous catheters (CVCs).
3	Prevention of central venous catheter-associated bloodstream infections in paediatric oncology patients using 70% ethanol locks: A randomised controlled multi-centre trial
4	Impact of flushing with aseptic non-touch technique using pre-filled flush or manually prepared syringes on central venous catheter occlusion and bloodstream infections in pediatric hemato-oncology patients: A randomized controlled study
5	The effects of chlorhexidine dressing on health care-associated infection in hospitalized patients: A meta-analysis

Explanation: the solution is partially correct because the second sentence contains correct information about the experimental arm, but not the top sentence. Also, there were no highlighted fragments.

Example of a fully irrelevant solution (Hashi 2017)

Human reviewer's verified extraction: "Health education on key WASH messages and demonstration of hand washing with soap"

ExaCT's extraction:

Sentence Ranking	Sentence
1	Efficacy and Safety of Letibotulinum Toxin A for the Treatment of Dynamic Equinus Foot Deformity in Children with Cerebral Palsy: A Randomized Controlled Trial
2	J. 1964; 2 : 230233 [PMC free article] [PubMed] [Google Scholar] Lutry S.P., Agboatwalla M., Feikin D.R. Effect of handwashing on child health: a randomised controlled trial.
3	Lancet. 2005; 366 : 225233. [PubMed] [Google Scholar] Luby S.P., Agboatwalla M., Painter J. Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial.

Sentence Ranking	Sentence
4	Efficacy and Safety of Letibotulinum Toxin A for the Treatment of Dynamic Equinu…
5	Lancet. 2005; 366 : 225233. [PubMed] [Google Scholar] [Ref list]

Explanation: the solution is completely incorrect because none of the sentences contain any relevant information related to the experimental arm.

Appendix D. Sample of Trials

Aboud FE, Bougma K, Lemma T, Marquis GS. Evaluation of the effects of iodized salt on the mental development of preschool-aged children: a cluster randomized trial in northern Ethiopia. Matern Child Nutr. 2017;13(2):e12322. doi: 10.1111/mcn.12322.

Ammari WG, Obeidat N, Khater M, Sabouba A, Sanders M. Mastery of pMDI technique, asthma control and quality-of-life of children with asthma: a randomized controlled study comparing two inhaler technique training approaches. Pulm Pharmacol Ther. 2017;43:46-54. doi: 10.1016/j.pupt.2017.02.002.

Anantasit N, Cheeptinnakorntaworn P, Khositseth A, Lertbunrian R, Chantra M. Ultrasound versus traditional palpation to guide radial artery cannulation in critically ill children: a randomized trial. J Ultrasound Med. 2017;36(12):2495-501. doi: 10.1002/jum.14291.

Aras I, Pasaoglu A, Olmez S, Unal I, Tuncer AV, Aras A. Comparison of stepwise vs single-step advancement with the functional mandibular advancer in class II division 1 treatment. Angle Orthod. 2017;87(1):82-7. doi: 10.2319/032416-241.1.

Bae DS, Valim C, Connell P, Brustowicz KA, Waters PM. Bivalved versus circumferential cast immobilization for displaced forearm fractures: a randomized clinical trial to assess efficacy and safety. J Pediatr Orthop. 2017;37(4):239-46. doi: 10.1097/BPO.0000000000000655.

Banupriya N, Bhat BV, Benet BD, Catherine C, Sridhar MG, Parija SC. Short term oral zinc supplementation among babies with neonatal sepsis for reducing mortality and improving outcome - a double-blind randomized controlled trial. Indian J Pediatr. 2018;85(1):5-9. doi: 10.1007/s12098-017-2444-8.

Ben-Pazi H, Cohen A, Kroyzer N, Lotem-Ophir R, Shvili Y, Winter G, et al. Clown-care reduces pain in children with cerebral palsy undergoing recurrent botulinum toxin injections- a quasi-randomized controlled crossover study. PloS One. 2017;12(4):e0175028. doi: 10.1371/journal.pone.0175028.

Blake MS, Waloszek JM, Raniti M, Simmons JG, Murray G, Blake L, et al. The SENSE Study: treatment mechanisms of a cognitive behavioral and mindfulness-based group sleep improvement intervention for at-risk adolescents. Sleep. 2017;40(6):1-11. doi: 10.1093/sleep/zsx061.

Bonnet D, Berger F, Jokinen E, Kantor PF, Daubeney PEF. Ivabradine in children with dilated cardiomyopathy and symptomatic chronic heart failure. J Am Coll Cardiol. 2017;70(10):1262-72. doi: 10.1016/j.jacc.2017.07.725.

Chang HJ, Hong BY, Lee SJ, Lee S, Park JH, Kwon JY. Efficacy and safety of letibotulinum toxin A for the treatment of dynamic equinus foot deformity in children with cerebral palsy: a randomized controlled trial. Toxins (Basel). 2017;9(8):252. doi: 10.3390/toxins9080252.

Darling A, McDonald CR, Urassa WS, Kain KC, Mwiru RS, Fawzi WW. Maternal dietary L-arginine and adverse birth outcomes in Dar es Salaam, Tanzania. Am J Epidemiol. 2017;186(5):603-11. doi: 10.1093/aje/kwx080.

DiVasta AD, Feldman HA, Rubin CT, Gallagher JS, Stokes N, Kiel DP, et al. The ability of low-magnitude mechanical signals to normalize bone turnover in adolescents hospitalized for anorexia nervosa. Osteoporos Int. 2017;28(4):1255-63. doi: 10.1007/s00198-016-3851-9.

Dray JB, Bowman, J, Campbell E, Freund M, Hodder R, Wolfenden L, et al. Effectiveness of a pragmatic school-based universal intervention targeting student resilience protective factors in reducing mental health problems in adolescents. J Adolesc. 2017;57:74-89. doi: 10.1016/j.adolescence.2017.03.009.

El-Chimi MS, Awad HA, El-Gammasy TM, El-Farghali OG, Sallam MT, Shinkar DM. Sustained versus intermittent lung inflation for resuscitation of preterm infants: a randomized controlled trial. J Matern Fetal Neonatal Med. 2017;30(11):1273-8. doi: 10.1080/14767058.2016.1210598.

Elkhayat HA, Aly RH, Elagouza IA, El-Kabarity RH, Galal YI. Role of P-glycoprotein inhibitors in children with drug-resistant epilepsy. Acta Neurol Scand. 2017;136(6):639-44.

Fjørtoft T, Ustad T, Follestad T, Kaaresen PI, Øberg GK. Does a parent-administrated early motor intervention influence general movements and movement character at 3months of age in infants born preterm? Early Hum Dev. 2017;112:20-4. doi: 10.1016/j.earlhumdev.2017.06.008

Freedman S, Orenstein D, Black P, Brown P, McCoy K, Stevens J, et al. Increased fat absorption from enteral formula through an in-line digestive cartridge in patients with cystic fibrosis. J Pediatr Gastroenterol Nutr. 2017;65(1):97-101. doi: 10.1097/MPG.0000000000001617.

Freira S, Lemos MS, Williams G, Ribeiro M, Pena F, Machado MDC. Effect of motivational interviewing on depression scale scores of adolescents with obesity and overweight. Psychiatry Res. 2017;252:340-5. doi: 10.1016/j.psychres.2017.03.020.

Fridenson-Hayo S, Berggren S, Lassalle A, Tal S, Pigat D, Meir-Goren N, et al. 'Emotiplay': a serious game for learning about emotions in children with autism: results of a cross-cultural evaluation. Eur Child Adolesc Psychiatry. 2017;26(8):979-992. doi: 10.1007/s00787-017-0968-0.

Gaesser AHK, O. C. A randomized controlled comparison of emotional freedom technique and cognitive-behavioral therapy to reduce adolescent anxiety: a pilot study. J Altern Complement Med. 2017;23(2):102-8. doi: 10.1089/acm.2015.0316.

Gal S, Ramirez JI, Maguina P. Autologous fat grafting does not improve burn scar appearance: a prospective, randomized, double-blinded, placebo-controlled, pilot study. Burns. 2017;43(3):486-9. doi: 10.1016/j.burns.2016.09.019.

Garnæs KK, Nyrnes SA, Salvesen K, Salvesen Ø, Mørkved S, Moholdt T. Effect of supervised exercise training during pregnancy on neonatal and maternal outcomes among overweight and obese women. Secondary analyses of the ETIP trial: a randomised controlled trial. PloS One. 2017;12(3):e0173937. doi: 10.1371/journal.pone.0173937.

Gerceker GO, Yardimci F, Aydinok Y. Randomized controlled trial of care bundles with chlorhexidine dressing and advanced dressings to prevent catheter-related bloodstream infections in pediatric hematology-oncology patients. Eur J Oncol Nurs. 2017;28:14-20. doi: 10.1016/j.ejon.2017.02.008

Giaccone A, Zuppa AF, Sood B, Cohen MS, O'Byrne ML, Moorthy G, et al. Milrinone pharmacokinetics and pharmacodynamics in neonates with persistent pulmonary hypertension of the newborn. Am J Perinatol. 2017;34(8):749-58. doi: 10.1055/s-0036-1597996.

Gottschlich MM, Mayes T, Khoury J, Kagan RJ. Clinical trial of vitamin D2 vs D3 supplementation in critically ill pediatric burn patients. JPEN J Parenter Enteral Nutr. 2017;41(3):412-21. doi: 10.1177/0148607115587948.

Grooten I, Koot M, Van Der Post J, Ris-Stalpers C, Naaktgeboren C, Mol BW, et al. Early enteral tube feeding in optimizing treatment for hyperemesis gravidarum: the Maternal and Offspring outcomes after Treatment of HyperEmesis by Refeeding (MOTHER) randomised controlled trial. Am J Clin Nutr. 2017;106(3):812-820. doi: 10.3945/ajcn.117.158931.

Guven Y, Aksakal SD, Avcu N, Unsal G, Tuna EB, Aktoren O. Success rates of pulpotomies in primary molars using calcium silicate-based materials: a randomized control trial. Biomed Res Int. 2017;2017:4059703. doi: 10.1155/2017/4059703.

Hamelmann E, Bernstein JA, Vandewalker M, Moroni-Zentgraf P, Verri D, Unseld A, et al. A randomised controlled trial of tiotropium in adolescents with severe symptomatic asthma. Eur Respir J. 2017;49(1):1601100. doi: 10.1183/13993003.01100-2016.

Han D, Liu YG, Pan S, Luo Y, Li J, Ou-Yang C. Comparison of hemodynamic effects of sevoflurane and ketamine as basal anesthesia by a new and direct monitoring during induction in children with ventricular septal defect: a prospective, randomized research. Medicine (Baltimore). 2017;96(50):e9039. doi: 10.1097/MD.0000000000009039.

Handeland K, Oyen J, Skotheim S, Graff IE, Baste V, Kjellevold M, et al. Fatty fish intake and attention performance in 14-15 year old adolescents: FINS-TEENS - a randomized controlled trial. Nutr J. 2017;16(1):64. doi: 10.1186/s12937-017-0287-9.

Hashi A, Kumie A, Gasana J. Hand washing with soap and WASH educational intervention reduces under-five childhood diarrhoea incidence in Jigjiga District, Eastern Ethiopia: a community-based cluster randomized controlled trial. Prev Med Rep. 2017;6:361-8. doi: 10.1016/j.pmedr.2017.04.011.

Indrio F, Riezzo G, Tafuri S, Ficarella M, Carlucci B, Bisceglia M, et al. Probiotic supplementation in preterm: feeding intolerance and hospital cost. Nutrients. 2017;9(9):965. doi: 10.3390/nu9090965.

Iserbyt P, Theys L, Ward P, Charlier N. The effect of a specialized content knowledge workshop on teaching and learning Basic Life Support in elementary school: a cluster randomized controlled trial. Resuscitation. 2017;112:17-21. doi: 10.1016/j.resuscitation.2016.11.023.

Karanja DMS, Awino EK, Wiegand RE, Okoth E, Abudho BO, Mwinzi PNM, et al. Cluster randomized trial comparing school-based mass drug administration schedules in areas of western Kenya with moderate initial prevalence of Schistosoma mansoni infections. PLoS Negl Trop Dis. 2017;11(10):e0006033. doi: 10.1371/journal.pntd.0006033

Kornmann MN, Christmann V, Gradussen CJW, Rodwell L, Gotthardt M, Van Goudoever JB, et al. Growth and bone mineralization of very preterm infants at term corrected age in relation to different nutritional intakes in the early postnatal period. Nutrients. 2017;9(12):1318. doi: 10.3390/nu9121318.

Lambrechts DA, de Kinderen RJ, Vles JS, de Louw AJ, Aldenkamp AP, Majoie HJ. A randomized controlled trial of the ketogenic diet in refractory childhood epilepsy. Acta Neurol Scand. 2017;135(2):231-9. doi: 10.1111/ane.12592.

Laskin BL, Huang G, King E, Geary DF, Licht C, Metlay JP, et al. Short, frequent, 5-days-per-week, in-center hemodialysis versus 3-days-per week treatment: a randomized crossover pilot trial through the Midwest Pediatric Nephrology Consortium. Pediatr Nephrol. 2017;32(8):1423-1432. doi: 10.1007/s00467-017-3656-x.

Lisante TA, Nuñez C, Zhang P. Efficacy and safety of an over-the-counter 1% colloidal oatmeal cream in the management of mild to moderate atopic dermatitis in children: a double-blind, randomized, active-controlled study. J Dermatolog Treat. 2017;28(7):659-67. doi: 10.1080/09546634.2017.1303569.

Locatelli F, Bernardo ME, Bertaina A, Rognoni C, Comoli P, Rovelli A, et al. Efficacy of two different doses of rabbit anti-T-lymphocyte globulin to prevent graft-versus-host disease in children with haematological malignancies transplanted from an unrelated donor: a multicentre, randomised, open-label, phase 3 trial. Lancet Oncol. 2017;18(8):1126-36. doi: 10.1016/S1470-2045(17)30417-5.

Lotfi Y, Rezazadeh N, Moossavi A, Haghgoo HA, Rostami R, Bakhshi E, et al. Preliminary evidence of improved cognitive performance following vestibular rehabilitation in children with combined ADHD (cADHD) and concurrent vestibular impairment. Auris Nasus Larynx. 2017;44(6):700-707. doi: 10.1016/j.anl.2017.01.011.

Lundbye-Jensen J, Skriver K, Nielsen JB, Roig M. Acute exercise improves motor memory consolidation in preadolescent children. Front Hum Neurosci. 2017;11:182. doi: 10.3389/fnhum.2017.00182.

Manoj M, Satya Prakash MVS, Swaminathan S, Kamaladevi RK. Comparison of ease of administration of intranasal midazolam spray and oral midazolam syrup by parents as premedication to children undergoing elective surgery. J Anesth. 2017;31(3):351-357. doi: 10.1007/s00540-017-2330-6.

Martinon-Torres F, Safadi MAP, Martinez AC, Marquez PI, Torres JCT, Weckx LY, et al. Reduced schedules of 4CMenB vaccine in infants and catch-up series in children: immunogenicity and safety results from a randomised open-label phase 3b trial. Vaccine. 2017;35(28):3548-3557. doi: 10.1016/j.vaccine.2017.05.023.

Mayfield CA, Child S, Weaver RG, Zarrett N, Beets MW, Moore JB. Effectiveness of a playground intervention for antisocial, prosocial, and physical activity behaviors. J Sch Health. 2017;87(5):338-45. doi: 10.1111/josh.12506.

McConnachie A, Haig C, Sinclair L, Bauld L, Tappin DM. Birth weight differences between those offered financial voucher incentives for verified smoking cessation and control participants enrolled in the Cessation in Pregnancy Incentives Trial (CPIT), employing an intuitive approach and a Complier Average Causal Effects (CACE) analysis. Trials. 2017;18(1):337. doi: 10.1186/s13063-017-2053-x.

Miklowitz DJ, Schneck CD, Walshaw PD, Garrett AS, Singh MK, Sugar CA, et al. Early intervention for youth at high risk for bipolar disorder: a multisite randomized trial of family-focused treatment. Early Interv Psychiatry. 2017;13(2):208-216. doi: 10.1111/eip.12463.

Miller LC, Joshi N, Lohani M, Rogers B, Mahato S, Ghosh S, et al. Women's education level amplifies the effects of a livelihoods-based intervention on household wealth, child diet, and child growth in rural Nepal. Int J Equity Health. 2017;16(1):183. doi: 10.1186/s12939-017-0681-0.

Moody KM, Baker RA, Santizo RO, Olmez I, Spies JM, Buthmann A, et al. A randomized trial of the effectiveness of the neutropenic diet versus food safety guidelines on infection rate in pediatric oncology patients. Pediatr Blood Cancer. 2018;65(1). doi: 10.1002/pbc.26711.

Muratori P, Bertacchi I, Giuli C, Nocentini A, Lochman JE. Implementing coping power adapted as a universal prevention program in Italian primary schools: a randomized control trial. Prev Sci. 2017;18(7):754-61. doi: 10.1007/s11121-016-0715-7.

O'Sullivan A, Fitzpatrick N, Doyle O. Effects of early intervention on dietary intake and its mediating role on cognitive functioning: a randomised controlled trial. Public Health Nutr. 2017;20(1):154-64. doi: 10.1017/S1368980016001877.

Papathomas L, Kuhn D. Learning to argue via apprenticeship. J Exp Child Psychol. 2017;159:129-39. doi: 10.1016/j.jecp.2017.01.013.

Pastor-Villaescusa B, Canete MD, Caballero-Villarraso J, Hoyos R, Latorre M, Vazquez-Cobela R, et al. Metformin for obesity in prepubertal and pubertal children: a randomized controlled trial. Pediatrics. 2017;140(1): e20164285. doi: 10.1542/peds.2016-4285.

Paz Castro R, Haug S, Kowatsch T, Filler A, Schaub MP. Moderators of outcome in a technology-based intervention to prevent and reduce problem drinking among adolescents. Addict Behav. 2017;72:64-71. doi: 10.1016/j.addbeh.2017.03.013.

Porter S, McConnell T, McLaughlin K, Lynn F, Cardwell C, Braiden HJ, et al. Music therapy for children and adolescents with behavioural and emotional problems: a randomised controlled trial. J Child Psychol Psychiatry. 2017;58(5):586-94. doi: 10.1111/jcpp.12656.

Rajavi Z, Feizi M, Naderi A, Sabbaghi H, Behradfar N, Yaseri M, et al. Graded versus ungraded inferior oblique anterior transposition in patients with asymmetric dissociated vertical deviation. J AAPOS. 2017;21(6):476-9.e1. doi: 10.1016/j.jaapos.2017.07.213.

Razi CH, Cörüt N, Andiran N. Budesonide reduces hospital admission rates in preschool children with acute wheezing. Pediatr Pulmonol. 2017;52(6):720-8. doi: 10.1002/ppul.23667.

Romanzini LP, Dos Santos AA Nunes ML. Characteristics of sleep in socially vulnerable adolescents. Eur J Paediatr Neurol. 2017;21(4):627-634. doi: 10.1016/j.ejpn.2016.12.013.

Romero GAS, Costa DL, Costa DL, Costa CHN, de Almeida RP, de Melo EV, et al. Efficacy and safety of available treatments for visceral leishmaniasis in Brazil: a multicenter, randomized, open label trial. PLoS Negl Trop Dis. 2017;11(6):e0005706.

Rowe SM, Daines C, Ringshausen FC, Kerem E, Wilson J, Tullis E, et al. Tezacaftor-ivacaftor in residual-function heterozygotes with cystic fibrosis. N Engl J Med. 2017;377(21):2024-35. doi: 10.1056/NEJMoa1709847.

Senders SD, Bundick ND, Li J, Zecca C, Helmond FA. Evaluation of immunogenicity and safety of VARIVAX™ New Seed Process (NSP) in children. Hum Vaccin Immunother. 2018:14(2):442-449. doi: 10.1080/21645515.2017.1388479.

Sharpe HP, Patalay P, Vostanis P, Belsky J, Humphrey N, Wolpert M. Use, acceptability and impact of booklets designed to support mental health self-management and help seeking in schools: results of a large randomised controlled trial in England. Eur Child Adolesc Psychiatry. 2017;26(3):315-24. doi: 10.007/s00787-016-0889-3.

Spektor Z, Pumarola F, Ismail K, Lanier B, Hussain I, Ansley J, et al. Efficacy and safety of ciprofloxacin plus fluocinolone in otitis media with tympanostomy tubes in pediatric patients a randomized clinical trial. JAMA Otolaryngol Head Neck Surg. 2017;143(4):341-9. doi: 10.1001/jamaoto.2016.3537.

Thurman TR, Nice J, Taylor TM, Luckett B. Mitigating depression among orphaned and vulnerable adolescents: a randomized controlled trial of interpersonal psychotherapy for groups in South Africa. Child Adolesc Ment Health. 2017;22(4):224-31. doi:10.1111/camh.12241.

Tonguet-Papucci A, Houngbe F, Huybregts L, Ait-Aissa M, Altare C, Kolsteren P, Huneau J. Unconditional seasonal cash transfer increases intake of high-nutritional-value foods in young burkinabe children: results of 24-hour dietary recall surveys within the Moderate Acute Malnutrition Out (MAM'Out) randomized controlled trial. J Nutr. 2017;147(7):1418-25. doi: 10.3945/jn.116.244517.

Urbancikova I, Prymula R, Goldblatt D, Roalfe L, Prymulova K, Kosina P. Immunogenicity and safety of a booster dose of the 13-valent pneumococcal conjugate vaccine in children primed with the 10-valent or 13-valent pneumococcal conjugate vaccine in the Czech Republic and Slovakia. Vaccine. 2017;35(38):5186-5193. doi: 10.1016/j.vaccine.2017.07.103.

Wei C, Allen RJ, Tallis PM, Ryan FJ, Hunt LP, Shield JP, et al. Cognitive behavioural therapy stabilises glycaemic control in adolescents with type 1 diabetes-Outcomes from a randomised control trial. Pediatr Diabetes. 2018;19(1):106-113. doi: 10.1111/pedi.12519.

Wei X, Zhang Z, Walley JD, Hicks JP, Zeng J, Deng S, et al. Effect of a training and educational intervention for physicians and caregivers on antibiotic prescribing for upper respiratory tract infections in children at primary care facilities in rural China: a cluster-randomised controlled trial. Lancet Glob Health. 2017;5(12):e1258-e67. doi: 10.1016/S2214-109X(17)30383-2.

Whittaker R, Stasiak K, McDowell H, Doherty I, Shepherd M, Chua S, et al. MEMO: an mHealth intervention to prevent the onset of depression in adolescents: a double-blind, randomised, placebo-controlled trial. J Child Psychol Psychiatry. 2017;58(9):1014-22. doi: 10.1111/jcpp.12753.

Winkler P, Janoušková M, Kozeny J, Pasz J, Mlada K,Weissova A. Short video interventions to reduce mental health stigma: a multi-centre randomised controlled trial in nursing high schools. Soc Psychiatry Psychiatr Epidemiol. 2017;52(12):1549-57. doi: 10.1007/s00127-017-1449-y

Wong JMW, Ebbeling CB, Robinson L, Feldman HA, Ludwig DS. Effects of advice to drink 8 cups of water per day in adolescents with overweight or obesity a randomized clinical trial. JAMA pediatr. 2017;171(5):e170012. doi: 10.1001/jamapediatrics.2017.0012.

Wu YJ, Wu WF, Hung CW, Ku MS, Liao PF, Sun HL, et al. Evaluation of efficacy and safety of Lactobacillus rhamnosus in children aged 4-48 months with atopic dermatitis: an 8-week, double-blind, randomized, placebo-controlled study. J Microbiol Immunol Infect. 2017;50(5):684-92. doi: 10.1016/j.jmii.2015.10.003.

Yuen VM, Li BL, Cheuk DK, Leung MKM, Hui TWC, Wong IC, et al. A randomised controlled trial of oral chloral hydrate vs. intranasal dexmedetomidine before computerised tomography in children. Anaesthesia. 2017;72(10):1191-5. doi: 10.1111/anae.13981.

Zambrano LD, Priest JW, Ivan E, Rusine J, Nagel C, Kirby M, et al. Use of serologic responses against enteropathogens to assess the impact of a point-of-use water filter: a randomized controlled trial in western province, Rwanda. Am J Trop Med Hyg. 2017;97(3):876-87. doi: 10.4269/ajtmh.16-1006.

Zhou Z, Chen T, Jin L, Zheng D, Chen S, He M, et al. Self-refraction, ready-made glasses and quality of life among rural myopic Chinese children: a non-inferiority randomized trial. Acta Ophthalmol. 2017;95(6):567-75. doi: 10.1111/aos.13149.

Zhu Y, Lin J, Long H, Ye N, Huang R, Yang X, et al. Comparison of survival time and comfort between 2 clear overlay retainers with different thicknesses: a pilot randomized controlled trial. Am J Orthod Dentofacial Orthop. 2017;151(3):433-9. doi: 10.1016/j.ajodo.2016.10.019.