Cochrane Evidence Synthesis and Methods

Cochrane Evidence Synthesis and Methods

BRIEF REPORT OPEN ACCESS

Data Extractions Using a Large Language Model (Elicit) and Human Reviewers in Randomized Controlled Trials: A Systematic Comparison

Joleen Bianchi[1,2] | Julian Hirt[1,3,4] | Magdalena Vogt[1] | Janine Vetsch[1]

[1]Department of Health, Eastern Switzerland University of Applied Sciences, St. Gallen, Switzerland | [2]Interdisciplinary Infant Unit, Eastern Switzerland Children's Hospital, St. Gallen, Switzerland | [3]Pragmatic Evidence Lab, Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), University Hospital Basel, University of Basel, Basel, Switzerland | [4]Institute of Health and Nursing Science, Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

Correspondence: Janine Vetsch (Janine.vetsch@ost.ch)

Received: 13 February 2025 | **Revised:** 20 May 2025 | **Accepted:** 21 May 2025

Funding: The authors received no specific funding for this work.

Keywords: artificial intelligence | data extraction | human reviewer | randomized controlled trial | systematic review

ABSTRACT

Aim: We aimed at comparing data extractions from randomized controlled trials by using Elicit and human reviewers.

Background: Elicit is an artificial intelligence tool which may automate specific steps in conducting systematic reviews. However, the tool's performance and accuracy have not been independently assessed.

Methods: For comparison, we sampled 20 randomized controlled trials of which data were extracted manually from a human reviewer. We assessed the variables study objectives, sample characteristics and size, study design, interventions, outcome measured, and intervention effects and classified the results into "more," "equal to," "partially equal," and "deviating" extractions. STROBE checklist was used to report the study.

Results: We analysed 20 randomized controlled trials from 11 countries. The studies covered diverse healthcare topics. Across all seven variables, Elicit extracted "more" data in 29.3% of cases, "equal" in 20.7%, "partially equal" in 45.7%, and "deviating" in 4.3%. Elicit provided "more" information for the variable study design (100%) and sample characteristics (45%). In contrast, for more nuanced variables, such as "intervention effects," Elicit's extractions were less detailed, with 95% rated as "partially equal."

Conclusions: Elicit was capable of extracting data partly correct for our predefined variables. Variables like "intervention effect" or "intervention" may require a human reviewer to complete the data extraction. Our results suggest that verification by human reviewers is necessary to ensure that all relevant information is captured completely and correctly by Elicit.

Implications: Systematic reviews are labor-intensive. Data extraction process may be facilitated by artificial intelligence tools. Use of Elicit may require a human reviewer to double-check the extracted data.

1 | Introduction

Systematic reviews are considered the most reliable method for synthesizing evidence, as they adhere to a structured, rigorous, and transparent research process. Due to their thoroughness, systematic reviews have long been pivotal in shaping health policy, clinical guidelines, and primary research [1]. The time required for a full systematic review, which is often more than 2 years after the publication of a protocol, represents a significant obstacle for both author teams and decision-makers [2]. Artificial intelligence (AI) tools have the potential to streamline the process of conducting systematic reviews, thereby reducing

the time required and the number of errors [3]. In the study by Affengruber et al. [1], the use of various tools, including Plot Digitizer, ChatGPT, ExaCT, Dextr, and DAA, was evaluated for their effectiveness in accelerating data extraction from studies. Manual data extraction by two reviewers, supplemented by the Plot Digitizer, exhibited comparable agreement with the original data, with slightly higher concordance achieved using the Plot Digitizer compared to manual extraction alone. In total, 87% of manually extracted data elements aligned with ExaCT, leading to altered outcomes in meta-analyses. ChatGPT demonstrated consistent agreement with human researchers across various parameters, including language, target disease, natural language processing model, sample size, and performance metrics, with moderate to fair agreement observed for clinical tasks and clinical implementation. User-friendliness was evaluated for DAA and Dextr, with both tools rated as highly user-friendly; however, DAA scored lower on feature-based assessments, while Dextr was noted for its flexible interface [1].

2 | Background

There is an AI research tool called "Elicit" which, in comparison to manual data extraction by trained research staff, demonstrated an accuracy that was 13%–26% higher (Elicit). However, this accuracy data were conducted and reported by Elicit and there is—to the best of our knowledge—no full independent research report on Elicit's accuracy. Elicit leverages large language models, including GPT-3, to automate research workflows. As the accuracy data are derived from the developers of Elicit and no alternative data are available, there is a need to evaluate its data extraction capabilities against human reviewers for a more robust assessment of its performance. Therefore, the aim is to evaluate and compare the data extraction capabilities of Elicit with those of a human reviewer to assess the accuracy and completeness of the data.

2.1 | Design

To compare the data extraction from Elicit with human extractions, we compared 20 studies of which data were extracted manually by a human reviewer from FIT-Nursing Care versus Elicit. This procedure was undertaken to evaluate the comparative efficacy of the predefined variables (see below). FIT-Nursing Care is a nursing knowledge platform providing study summaries. These German summaries are developed by health researchers using predefined data extraction fields in an online platform [4].

2.2 | Methods and Materials

2.2.1 | Eligibility Criteria and Data Source

We considered individual and cluster randomized controlled studies (RCTs) indexed in FIT-Nursing Care. We used the platform-specific filter for "Intervention studies" which displayed 518 studies. We then purposively selected 20 studies which were published on the platform after 2015 as a convenience sample.

2.2.2 | Data Comparison

2.2.2.1 | Elicit Extractions. We uploaded the studies as PDFs into Elicit and extracted the following variables: study objectives, sample characteristics, study design, participant count, intervention, outcome measured, and intervention effects [5]. Two variables of Elicit were manually adjusted so that the results could be compared with those of FIT-Nursing Care. The commands behind the columns from Elicit are shown in the Appendix. To examine the demographic characteristics of the participants included in the study, the "sample characteristics" column was derived from the "Population Characteristics" column. We extended the command so that Elicit also extracts, total size of the final sample and a demographic description. Also, to extract the control intervention, we adapted the "Intervention" column so that Elicit also extracts information on the control intervention. We downloaded the data extracted from Elicit in a Comma-Separated Values (CSV) file and subsequently opened it in an Excel spreadsheet. We conducted the data extraction over a 2-week period between mid-August and September 2024.

2.2.2.2 | Human Extractions (Reference). FIT-Nursing Care follows a defined methodological approach. Two reviewers extract the relevant information for the variables described here and more. The first reviewer fills in all predefined study fields (such as background, study design, etc.) and a second reviewer double-checks all the information. All reviewers are trained nursing or health researchers [4]. We compared the variables listed in Table 1. The German variables from the RCTs extracted by a human reviewer in the FIT-Nursing Care platform were manually transferred to the same Excel spreadsheet.

2.2.2.3 | Data Analysis. We analyzed the combined data set from data extractions using Elicit and FIT-Nursing Care. One person (J. B.) assessed data completeness and accuracy and classified to: "more," "equal to," "partially equal," and "deviating." The category "more" meant that Elicit extracted more information than a human reviewer which the human had omitted but was correct. The second category "equal to" meant that Elicit extracted equal information to the human reviewer. "Partially equal" meant that Elicit extracted some equal information, but some important data is missing. The fourth category "deviating" meant that Elicit extracted different/wrong information than a human reviewer. The categories were defined within the study team in a preliminary assessment of five trials which were included in the final sample. The classification was verified by a second person (J. V.).

TABLE 1 | Variables of Elicit and FIT-Nursing Care.

Variables in Elicit	Variables in FIT-Nursing Care
Study objectives	Fragestellung/Zielsetzung
Sample characteristics	Stichprobenbeschreibung
Participant count	Stichprobengrösse
Study design	Design
Interventions	Intervention und Kontrolle
Outcome measured	Primäres Zielkriterium
Intervention effects	Ergebnisse

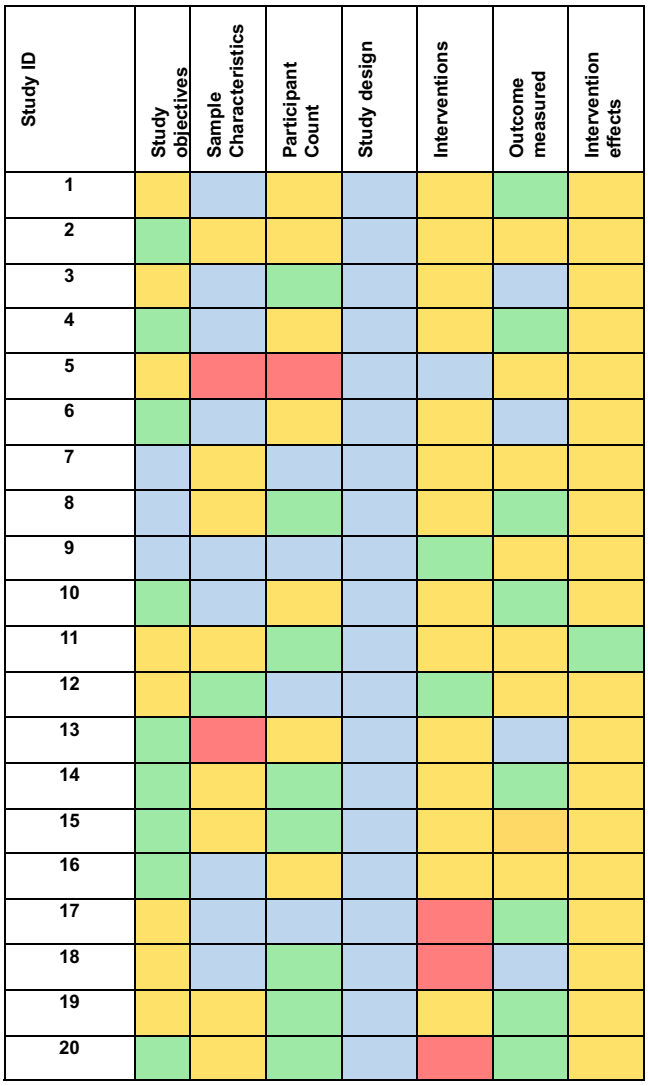

FIGURE 1 | Heatmap of the 7 variables across 20 RCTs compared with the data extraction of Elicit versus human reviewer. Blue = "more," green = "equal to," orange = "partially equal" with less information, and red = "deviating."

Finally, J. B. and J. V. discussed the "deviating" and "partially equal" categories to assess the extent of the deviation. During the peer review of our article, the fourth category, described as "more," was established. One author (J. B.) adapted the data extraction table and the results tables. These adjustments were then reviewed by another reviewer (J. V.). We narratively and visually summarized the result (Figure 1). To enhance the transparency of our results, we created a table (see Supporting Information S2: Appendix 2: Table 2 [6–18]), which presents the categories with one to two examples of each.

3 | Results

We compared data extractions from $N = 20$ RCTs from 14 different countries (Germany, Argentina, Brazil, South Africa, Austria, Australia, United States of America, England, Norway, Turkey, China, Switzerland, New Zealand, and Italy), published between 2016 and 2021. The topics of the included studies were diverse: maternal, infant and child health, vaccination and infection prevention, mental health and psychoeducation, nurse-led interventions and self-management, cancer care and survivorship, therapy and rehabilitation in older adults, symptom management and patient comfort, and health decision-making and screening. The details of the individual studies are shown in Supporting Information S1: Appendix 1.

Data extractions were "more" by Elicit compared to human extractions in 29.3% of the 7 variables over all 20 studies, "equal to" in 20.7%, "partially equal" in 45.7%, and "deviating" in 4.3% (see Figure 1 and Supporting Information S1: Appendix 1).

For the variable "study objectives," Elicit extracted "more" compared to human extractions in 15% (3/20 studies), "equal to" in 45% (9/20 studies), "partially equal" with less information being provided in 40% (8/20 studies).

Regarding the variable "sample characteristics," Elicit was found to have extracted "more" than a human reviewer in 45% (9/20 studies), "equal to" in 5% (1/20 studies), "partially equal" in 40% (8/20 studies), and in 10% of studies (2/20 studies) the data were "deviating" from that extracted by human reviewer.

For the variable "participant count," Elicit extracted "more" than a human reviewer in 20% of studies (4/20 studies). In 40% (8/20), "equal to" information was provided by Elicit. In 35% of studies, the data extracted by Elicit was "partially equal" with less information and 5% (1/20 studies) was "deviating" from the data extracted by the human reviewer.

For the "study design" variable, Elicit extracted "more" in all studies (100%) than a human reviewer.

For the variable "Interventions," Elicit's data extraction was in 5% (1/20 studies) "more" than by the human reviewer, "equal to" in 10% (2/20 studies), "partially equal" with less information in 70% (14/20 studies), and in 15% (3/20 studies) "deviating."

For "outcome measured," the results demonstrated that in 20% of studies, Elicit extracted "more" data than that by the human reviewer. In 40%, the data were found to be "equal to" and in 40%, the data were found to be "partially equal" with less information available.

For the variable "intervention effects," 5% of the studies (1/20 studies) "more" data were extracted by Elicit than by human reviewer; the remaining 95% were "partially equal" with less information.

4 | Discussion

Our aim was to compare data extraction from Elicit versus a human reviewer from FIT-Nursing Care. Overall, in just below half of the variables which were compared Elicit extracted data that was "partially equal." In one-third of the data extraction, Elicit extracted "more" than the human reviewer. Only 4.3% of the extracted data from Elicit "deviated" from the human reviewer.

The process of extracting data from full texts is inherently labor-intensive. Furthermore, the inconsistent application of extraction

criteria across studies and human reviewers represents another source of variability. Additionally, the process of information retrieval is subject to variability in interpretation, which may also impact the accuracy of reported effects. As with any process involving human input, there is always the potential for human error to negatively impact the results [3]. If an AI tool could efficiently, accurately, and completely extract the data required it could facilitate the review process. Several scoping reviews have highlighted the potential of generative AI to assist in the data extraction process [19–22]. However, Elicit has not been independently evaluated before. Elicit has demonstrated that for certain variables, the information can be extracted accurately, for example, for the study design of RCTs. Elicit demonstrated proficiency, accurately extracting data and providing sufficient information in 100% of the studies. However, this needs to be tested for other study designs in further research. Overall, Elicit extracted 29.3% more than humans across all 7 variables and 20 trials. This may indicate a higher sensitivity and possibly consistency of the system in identifying information. However, this result should be interpreted with caution, as extracting "more" does not necessarily equate to extracting better or more accurate information. It is questionable if the additional information was always needed or indeed relevant. Therefore, further studies should focus on the qualitative evaluation of the extracted content and assess the precision and relevance of the additional data provided by Elicit.

For the variables sample characteristics, interventions, intervention effect, Elicit extracted "partially equal" with less information than human reviewer. However, it is questionable whether a less detailed extraction, as observed for the "intervention effect" variable with only 5% complete agreement, should be rated as "poor." It is important to note that discrepancies in the extraction process do not necessarily indicate inaccuracy in the results. Rather, they suggest that Elicit has likely prioritized different aspects than the human reviewer. It is essential to consider the context and the specific objective of Elicit's intended use to ascertain its suitability. It is also conceivable that the reviewer responsible for data extraction on FIT-Nursing Care may identify information of greater relevance than that extracted by Elicit. It is also the case that, even with systematic reviews, different authors extract the data differently, and the level of detail must therefore be determined beforehand.

It is important to note the possibility that the results could be improved by adjusting the command or configuration of Elicit. This is exemplified by the variable "Outcome measured," where in 20% "more" data were extracted and in 40% "equal to" and in 40% "partially equal" with less information. It is possible that targeted adjustments to the command underlying the variable "outcome measured" in Elicit could increase precision, thereby significantly enhancing the utility of the tool. However, this was not tested as part of our project. In their mapping review, Cierco Jimenez et al. [22] describe various AI-supported tools that can be used during the SR process. However, the evaluation of the accuracy and precision of these tools is missing. Our findings are similar to previous studies analyzing the performance of AI-supported tools. Lieberum et al. [19] show in their scoping review several tools, such as Generative Pretrained Transformer (GPT) and Claude, to support the data extraction process. Blaizot et al. [21] also demonstrated in their SR, several AI-supported tools that could facilitate the creation of an SR. For instance, the software SWIFT-Review is used for data extraction, but the software was not able to automate all aspects of data extraction. Consequently, individual variables, such as the sample size, had to be entered manually. A manual review of the automated processes was also necessary to identify any missing information in the data extraction from the software [21]. The results in the SR from Marshall and Wallace [20] suggest that AI-supported data extraction tools for systematic review—such as ExaCT and RobotReviewer—have made notable progress, but are still at an early stage of development. Although these tools have demonstrated encouraging accuracy rates, they are not yet sufficiently accurate to fully replace manual data extraction. The performance of these tools is constrained by limited and often imperfect training data, which can reduce accuracy and reliability [20].

Further, it is important to note that Elicit is currently unable to extract information from figures, such as flowcharts, which represents a significant limitation. Five of the involved studies used visual data in the form of charts, graphs, or tables [23–27]. Consequently, data extraction must be conducted manually if information from tables and figures needs to be extracted. It is also important to note that Elicit is subject to ongoing development, with new features being added frequently. Since conducting our study, Elicit has added the following features at the start of 2025: for example, extract data from tables in papers with high accuracy columns and is likely to amend their product continuously.

4.1 | Limitations

First, the studies from FIT-Nursing Care were used as a reference, having been extracted by different human reviewers. Even though these humans were educated, and data were double-checked, we cannot exclude that different reviewers extract data in different levels of detail. It must be acknowledged that human data extraction is not always flawless and needs to be performed independently in a rigorous approach. Second, the data from FIT-Nursing Care were in German and the data from Elicit were in English. So, there was a comparison between German and English data extraction which could have led to translation errors which, however, is unlikely as the native language of the authors is German and all are proficient in English. Furthermore, we assessed a convenience sample of 20 RCTs and further studies need to confirm our results within a larger sample size.

5 | Conclusion

Elicit was capable of extracting data (partly) correct for our predefined variables, particularly for variables such as the study design. However, it is important to consider the limitations of the tool, particularly in terms of its ability to process detailed data and visual information. In the extraction of more complex data, such as the variables "intervention effects" or "intervention," Elicit demonstrated to have certain deficits. Thus, human verification is necessary to ensure that all relevant information is captured completely and correctly by Elicit but future studies

are needed to confirm our results. The combination of automated data extraction and human control might enhance efficiency while maintaining scientific accuracy.

6 | Implications

Systematic reviews are labor-intensive. The data extraction process may be facilitated by AI tools. In 29.3%, Elicit extracted "more" data, and in 20.7%, "equal to" information compared to a human reviewer. However also half of the data showed "partially equal" data or "deviating" data. Use of Elicit may require a human reviewer to double-check the extracted data. AI tools may facilitate the data extraction process during synthesis. Elicit may be an important tool for other research fields than health.

Author Contributions

Joleen Bianchi: conceptualization, writing – original draft, writing – review and editing. **Julian Hirt:** conceptualization, writing – review and editing, writing – original draft, supervision. **Magdalena Vogt:** conceptualization, writing – original draft, writing – review and editing. **Janine Vetsch:** conceptualization, writing – review and editing, writing – original draft, supervision.

Acknowledgments

STROBE checklist was used to report the study.

Ethics Statement

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data supporting the results of this study are available in the article or in the Appendix.

Clinical Resources

- Elicit: The AI Research Assistant, https://elicit.com/ [5].
- FIT Nursing Care, Knowledge Plattform for Nurses in the German-speaking area, https://www.fit-care.ch.

References

1. L. Affengruber, M. M. van der Maten, I. Spiero, et al., "An Exploration of Available Methods and Tools to Improve the Efficiency of Systematic Review Production - A Scoping Review," ahead of print, (2024), https://doi.org/10.21203/rs.3.rs-4595777/v1.

2. M. Cumpston, E. Flemyng, J. Thomas, J. Higgins, J. Deeks, and M. Clarke, "Chapter I: Introduction," in *Cochrane Handbook for Systematic Reviews of Interventions version 6.5*, ed. J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, and V. Welch (Cochrane, 2024), training.cochrane.org/handbook.

3. B. Degen, S. Vogel, and D. Rzejak, "Leveraging Artificial Intelligence for Systematic Reviews: The FRAISR Reporting Framework and Guidance for Researchers," preprint, (2024), https://doi.org/10.31219/osf.io/ju8dk.

4. J. Vetsch, S. Haug, and B. Vosseler, *Methodenpapier FIT-Nursing Care - Version 2.1 Stand April 2022* (2022).

5. Elicit, "*AI to Speed up HTA & SLR*," https://elicit.com/solutions/heor.

6. L. R. Baden, H. M. El Sahly, B. Essink, et al., "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine," *New England Journal of Medicine* 384, no. 5 (2021): 403–416, https://doi.org/10.1056/NEJMoa2035389.

7. D. A. Forster, A. M. Moorhead, S. E. Jacobs, et al., "Advising Women With Diabetes in Pregnancy to Express Breastmilk in Late Pregnancy (Diabetes and Antenatal Milk Expressing [DAME]): A Multicentre, Unblinded, Randomised Controlled Trial," *Lancet* 389, no. 10085 (2017): 2204–2213, https://doi.org/10.1016/S0140-6736(17)31373-9.

8. K. K. Garnæs, S. Mørkved, Ø. Salvesen, and T. Moholdt, "Exercise Training and Weight Gain in Obese Pregnant Women: A Randomized Controlled Trial (ETIP Trial)," *PLoS Medicine* 13, no. 7 (2016): e1002079, https://doi.org/10.1371/journal.pmed.1002079.

9. G. Ö. Gerçeker, S. A. Sevgili, and F. Yardımcı, "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," *European Journal of Oncology Nursing* 33 (2018): 78–84, https://doi.org/10.1016/j.ejon.2018.02.002.

10. N. E. Glass, N. A. Perrin, G. C. Hanson, et al., "The Longitudinal Impact of an Internet Safety Decision Aid for Abused Women," *American Journal of Preventive Medicine* 52, no. 5 (2017): 606–615, https://doi.org/10.1016/j.amepre.2016.12.014.

11. B. Lay, W. Kawohl, and W. Rössler, "Outcomes of a Psycho-Education and Monitoring Programme to Prevent Compulsory Admission to Psychiatric Inpatient Care: A Randomised Controlled Trial," *Psychological Medicine* 48, no. 5 (2018): 849–860, https://doi.org/10.1017/S0033291717002239.

12. N. Moadad, K. Kozman, R. Shahine, S. Ohanian, and L. K. Badr, "Distraction Using the BUZZY for Children During an IV Insertion," *Journal of Pediatric Nursing* 31, no. 1 (2016): 64–72, https://doi.org/10.1016/j.pedn.2015.07.010.

13. F. P. Polack, S. J. Thomas, N. Kitchin, et al., "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine," *New England Journal of Medicine* 383, no. 27 (2020): 2603–2615, https://doi.org/10.1056/NEJMoa2034577.

14. C. M. Rickard, N. M. Marsh, E. N. Larsen, et al., "Effect of Infusion Set Replacement Intervals on Catheter-Related Bloodstream Infections (RSVP): A Randomised, Controlled, Equivalence (Central Venous Access Device)-Non-Inferiority (Peripheral Arterial Catheter) Trial," *Lancet* 397, no. 10283 (2021): 1447–1458, https://doi.org/10.1016/S0140-6736(21)00351-2.

15. C. L. Rock, S. W. Flatt, T. E. Byers, et al., "Results of the Exercise and Nutrition to Enhance Recovery and Good Health for You (ENERGY) Trial: A Behavioral Weight Loss Intervention in Overweight or Obese Breast Cancer Survivors," *Journal of Clinical Oncology* 33, no. 28 (2015): 3169–3176, https://doi.org/10.1200/JCO.2015.61.1095.

16. G. Sorrentino, M. Fumagalli, S. Milani, et al., "The Impact of Automatic Devices for Capillary Blood Collection on Efficiency and Pain Response in Newborns: A Randomized Controlled Trial," *International Journal of Nursing Studies* 72 (2017): 24–29, https://doi.org/10.1016/j.ijnurstu.2017.04.001.

17. A. Toots, H. Littbrand, N. Lindelöf, et al., "Effects of a High-Intensity Functional Exercise Program on Dependence in Activities of Daily Living and Balance in Older Adults With Dementia," *Journal of the American Geriatrics Society* 64, no. 1 (2016): 55–64, https://doi.org/10.1111/jgs.13880.

18. D. Yıldırım, G. Can, and G. Köknel Talu, "The Efficacy of Abdominal Massage in Managing Opioid-Induced Constipation," *European Journal of Oncology Nursing* 41 (2019): 110–119, https://doi.org/10.1016/j.ejon.2019.05.013.

19. J. L. Lieberum, M. Toews, M. I. Metzendorf, et al., "Large Language Models for Conducting Systematic Reviews: On the Rise, but Not yet Ready for Use-A Scoping Review," *Journal of Clinical Epidemiology* 181 (2025): 111746, https://doi.org/10.1016/j.jclinepi.2025.111746.

20. I. J. Marshall and B. C. Wallace, "Toward Systematic Review Automation: A Practical Guide to Using Machine Learning Tools in Research Synthesis," *Systematic Reviews* 8, no. 1 (2019): 163, https://doi.org/10.1186/s13643-019-1074-9.

21. A. Blaizot, S. K. Veettil, P. Saidoung, et al., "Using Artificial Intelligence Methods for Systematic Review in Health Sciences: A Systematic Review," *Research Synthesis Methods* 13, no. 3 (2022): 353–362, https://doi.org/10.1002/jrsm.1553.

22. R. Cierco Jimenez, T. Lee, N. Rosillo, et al., "Machine Learning Computational Tools to Assist the Performance of Systematic Reviews: A Mapping Review," *BMC Medical Research Methodology* 22, no. 1 (2022): 322, https://doi.org/10.1186/s12874-022-01805-4.

23. B. Leininger, C. McDonough, R. Evans, T. Tosteson, A. N. A. Tosteson, and G. Bronfort, "Cost-Effectiveness of Spinal Manipulative Therapy, Supervised Exercise, and Home Exercise for Older Adults With Chronic Neck Pain," *Spine Journal* 16, no. 11 (2016): 1292–1304, https://doi.org/10.1016/j.spinee.2016.06.014.

24. S. Raphaelis, F. Frommlet, H. Mayer, and A. Koller, "Implementation of a Nurse-Led Self-Management Support Intervention for Patients With Cancer-Related Pain: A Cluster Randomized Phase-IV Study With a Stepped Wedge Design (EvANtiPain)," *BMC Cancer* 20, no. 1 (2020): 559, https://doi.org/10.1186/s12885-020-06729-0.

25. B. J. Taylor, A. R. Gray, B. C. Galland, et al., "Targeting Sleep, Food, and Activity in Infants for Obesity Prevention: An RCT," *Pediatrics* 139, no. 3 (2017): e20162037, https://doi.org/10.1542/peds.2016-2037.

26. J. Turner, P. Yates, L. Kenny, et al., "The ENHANCES Study: A Randomised Controlled Trial of a Nurse-Led Survivorship Intervention for Patients Treated for Head and Neck Cancer," *Supportive Care in Cancer* 27, no. 12 (2019): 4627–4637, https://doi.org/10.1007/s00520-019-04748-7.

27. R. J. Volk, L. M. Lowenstein, V. B. Leal, et al., "Effect of a Patient Decision Aid on Lung Cancer Screening Decision-Making by Persons Who Smoke: A Randomized Clinical Trial," *JAMA Network Open* 3, no. 1 (2020): e1920362, https://doi.org/10.1001/jamanetworkopen.2019.20362.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.