Research on the Practical Application of Crowdsourcing in Evidence Synthesis: A Case Study of the Cochrane Crowd Citizen Science Project Postprint

Authors: Li Xiao, Qu Jiansheng, Kou Leilei

Date: 2023-05-08T00:00:00+00:00

Abstract

[Purpose/Significance] The timeliness of evidence generation is crucial for evidence-based decision-making, yet the current efficiency of evidence synthesis often fails to meet decision-makers' needs. Crowdsourcing is considered a potential approach to enhance the production efficiency of evidence synthesis. This study examines the crowdsourcing application in the Cochrane Crowd citizen science project as a case study to summarize the practical implementation of crowdsourcing in evidence synthesis. [Method/Process] Employing methods including literature research, web-based surveys, and case analysis, this study analyzes the application mechanism of crowdsourcing in the Cochrane Crowd citizen science project across five dimensions: crowdsourcers, volunteers, crowdsourcing tasks, the Cochrane Crowd platform, and quality assessment. [Results/Conclusion] By establishing clear objectives, incentive mechanisms, and well-defined tasks, coupled with comprehensive training and appropriate quality control mechanisms, crowdsourcing can be leveraged to deliver high-quality outputs for evidence synthesis. This provides a reference for future research on applying crowdsourcing in evidence synthesis across different domains and at various stages of the evidence synthesis process.

Full Text

Preamble

Crowdsourcing in Evidence Synthesis: A Practical Application Study—Taking the Crowdsourcing Application in the Cochrane Crowd Citizen Science Project as an Example

Authors: Li Xiao[12], Qu Jiansheng[123]*, Kou Leilei[4]
[1]Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000
[2]Department of Library Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100049
[3]Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041
[4]Qinghai-Tibet Plateau Human Environment Research Center, Lanzhou University, Lanzhou 730000

[**Purpose/Significance**] The timeliness of evidence generation is crucial for evidence-informed decision-making, yet the efficiency of evidence synthesis currently fails to meet decision-makers' needs. Crowdsourcing is considered a potential method to improve the productivity of evidence synthesis. This study examines the crowdsourcing application in the Cochrane Crowd citizen science project as a case study to summarize practical applications of crowdsourcing in evidence synthesis.

[**Method/Process**] Using literature review, web investigation, and case analysis methods, we analyzed the application mechanism of crowdsourcing in the Cochrane Crowd citizen science project across five dimensions: crowdsourcer, volunteers, crowdsourcing tasks, Cochrane Crowd platform, and quality assessment.

[**Results/Conclusions**] By setting clear objectives, implementing incentive measures, designing clear tasks, providing comprehensive training, and establishing appropriate quality control mechanisms, crowdsourcing can deliver high-quality results for evidence synthesis. This provides a reference for future research on applying crowdsourcing in evidence synthesis across different fields and at different stages of the synthesis process.

Keywords: evidence synthesis; crowdsourcing; Cochrane Crowd; evidence-based research

1 Introduction

Evidence-informed decision-making has become an important component of evidence-based decision-making in many fields, and evidence synthesis serves as a means to bridge the gap between research and policy. The accurate, concise, and unbiased synthesis of existing evidence is essential for informed decision-making. However, evidence synthesis is resource-intensive, particularly when synthesizing scientific knowledge on complex problems. With the breadth and depth of human research approximately doubling every nine years, solutions are often distributed across tens of thousands of individual studies. Studies indicate that a single evidence synthesis requires 18 months to 3 years of research team

time, which hinders the demand-driven cycle of policy-making.

The academic community continuously explores innovative technologies to improve evidence synthesis productivity. Some scholars have investigated machine learning, natural language processing, and text mining to search multiple bibliographic databases and gray literature sources. Others have examined citizen science methods like crowdsourcing to reduce costs and shorten timelines. Cochrane, an internationally renowned nonprofit organization dedicated to health evidence production, primarily develops systematic reviews based on randomized controlled trials. Although Cochrane focuses on medical evidence, its evidence synthesis methods and tools have been adopted by other fields. Cochrane was an early adopter of crowdsourcing in evidence synthesis, launching the Cochrane Crowd citizen science platform in 2016.

Research and practice on crowdsourcing applications in evidence synthesis remain in their infancy. Most studies focus on applying crowdsourcing to specific stages rather than the complete process, particularly the resource-intensive screening and quality assessment phases. The most successful implementation is the Cochrane Crowd citizen science project initiated by Cochrane. Since 2014, Cochrane has used crowdsourcing to efficiently identify studies for its CENTRAL database, launching the Cochrane Crowd platform in May 2016 to recruit global volunteers to help classify studies for health decision-making. As of the survey, 29,004 contributors from 180 countries have classified nearly 8 million records.

This study employs literature review and case analysis methods to examine the crowdsourcing application mechanism in Cochrane Crowd, providing reference for Chinese researchers across disciplines to apply crowdsourcing in evidence synthesis and improve evidence generation efficiency.

---

2 Crowdsourcing Application in Cochrane Crowd

Crowdsourcing operations can be analyzed through multiple dimensions. Based on HOSSEINI et al.'s four-dimensional framework for crowdsourcing operations, this paper examines the Cochrane Crowd application mechanism from five dimensions: crowdsourcer, volunteers, crowdsourcing tasks, Cochrane Crowd platform, and quality assessment.

2.1 Crowdsourcer

The crowdsourcer is the initiator of the crowdsourcing project. In Cochrane Crowd, there are two types: Cochrane organization and Cochrane systematic review author teams. The former is the primary initiator and main service object, while the latter emerged after the Screen4Me service launch.

Cochrane is a global nonprofit organization comprising researchers, healthcare professionals, and others interested in health research, dedicated to collecting

and analyzing the best available health evidence and producing systematic reviews. Its work is recognized as the international gold standard for reliable information, helping people make informed decisions about health and healthcare. Cochrane reviews enjoy high reputation due to their rigor. To improve systematic review productivity, Cochrane built the Cochrane Central Register of Controlled Trials (CENTRAL), which specifically collects reports of randomized and quasi-randomized controlled trials (RCTs and qRCTs) from databases including Embase.com, ClinicalTrials.gov, CINAHL, and WHO ICTRP.

The purpose of Cochrane's crowdsourcing initiative is to identify RCTs and qRCTs for the CENTRAL database. In 2019, Cochrane launched the Screen4Me service, which has been used by over 60 systematic review teams from 15 Cochrane review groups. The service comprises three components: Known Assessment Service, RCT Classifier Service, and Cochrane Crowd crowdsourcing platform. The workflow involves uploading records for a specific systematic review, matching them against previously crowd-screened records through the Known Assessment Service, then using the RCT Classifier Service to identify potential RCTs from remaining records, and finally having volunteers further screen these potential RCTs through the crowdsourcing platform. Pilot studies show volunteers can not only identify potential RCTs but also assess relevance to review topics based on inclusion criteria, though these tasks require customized training modules.

2.2 Volunteers

Cochrane Crowd is open to all without restrictions on qualifications, allowing individuals without health research experience to participate. To facilitate volunteer participation, the platform employs strategies to improve cognitive levels while reducing task difficulty through carefully designed learning activities, task access methods, and feedback mechanisms.

Survey data shows volunteer motivations include altruism (helping Cochrane), learning, and self-actualization. Cochrane Crowd provides five color-coded milestone badges representing completion of training tasks, 100 classifications, 1,000 classifications with high accuracy, and outstanding performance. These appear on the task interface to motivate volunteers.

2.3 Crowdsourcing Tasks

2.3.1 Learning Activities The platform provides learning activities covering key concepts, research design understanding, and CONSORT (Consolidated Standards of Reporting Trials) reporting standards. The key concepts module comprises seven sub-modules covering fair trial concepts. The research design module introduces major study designs used in health research. The CONSORT module teaches information reported in RCTs. These activities help volunteers learn relevant knowledge.

Each volunteer receives a crowdsourcing activity summary showing their learning and task history. Participants are classified into three levels based on performance: regular screeners, expert screeners, and arbiters. All volunteers start as regular screeners and can upgrade to expert screeners after completing 1,000 classifications with very high accuracy. Expert screeners' classifications carry greater weight, and records screened by them require fewer decisions to confirm. Arbiters are volunteers with outstanding screening performance who make final decisions when other screeners disagree or classify records as uncertain.

2.3.2 Task Access Methods Three task access pathways accommodate different experience levels: direct task access, newcomer pathway, and student pathway. Direct access suits volunteers with domain knowledge and professional experience. The newcomer pathway suits those unfamiliar with health research and evidence-based medicine. The student pathway suits those learning health-related knowledge and wanting to understand evidence-based medicine.

The direct access interface presents all learning activities and task types for voluntary selection. The newcomer and student pathways combine learning activities and tasks in sequential order based on progressive principles, as shown in . Volunteers must complete them sequentially.

2.3.3 Training Model and Feedback Mechanism Each task type has a brief, standardized training module comprising interactive training records. For example, RCT and ICTRP identification tasks include 7 training records, while CT identification includes 20. Volunteers receive immediate feedback and guidance after each classification selection. These records reflect real-time task scenarios, helping volunteers understand task requirements and learn proper classification through feedback.

For COVID study identification tasks, training includes both learning and assessment components. After completing 10 training records, volunteers must pass an assessment (requiring 7 correct classifications) before task execution, repeating training until qualified.

2.3.4 Task Types Cochrane Crowd tasks fall into three categories: mainstream tasks, extended tasks, and pilot tasks.

Mainstream tasks involve identifying RCTs from literature titles and abstracts. Record sources include four databases: Embase.com, CINAHL, ClinicalTrials.gov, and WHO ICTRP, with varying formats and difficulty levels. Mainstream tasks are subdivided into three types based on record format: RCT identification, CT identification, and ICTRP identification. CT and ICTRP identification tasks display structured field groups, making information clearer and more beginner-friendly. [Figure 1: see original paper] and [Figure 2: see original paper] show examples of RCT and ICTRP identification tasks.

Extended tasks include Screen4Me, COVID Quest, and COVID Quest Lite.

In response to COVID-19, Cochrane created the Cochrane COVID-19 Study Register, launching COVID Quest and COVID Quest Lite to rapidly identify COVID-19-related studies and assign tags.

Pilot tasks include PICO extraction, DTA identification, and table retrieval, serving two purposes: enabling volunteers to become Cochrane systematic review authors or conducting feasibility tests. Task details are shown in .

2.4 Cochrane Crowd Platform

Crowdsourcing platforms typically feature interaction functions with crowdsourcers and crowd workers, task-related facilities, and platform-related infrastructure. Task-related facilities concern crowdsourcing tasks, while platform-related facilities concern the platform itself. Cochrane Crowd has distinctive features in both areas.

2.4.1 Task-related Facilities The platform employs a crowd agreement algorithm requiring four consecutive identical classifications for final decisions. When the chain breaks or uncertainty arises, experienced arbiters make final decisions. The platform also uses highlighting functions to mark key phrases or words in titles/abstracts, guiding volunteer attention and aiding classification decisions. Blue highlighting marks method sections in abstracts, while other colors mark keywords in potentially relevant or irrelevant records.

2.4.2 Platform-related Facilities A distinctive platform facility is the machine learning-based RCT Classifier. Trained on Cochrane Crowd's crowd dataset, it predicts the probability of a record being an RCT, enabling deletion of obvious non-RCT records to reduce volunteer workload and improve identification efficiency. Studies show the classifier can exclude 60-80% of irrelevant records while maintaining over 99% recall. [Figure 3: see original paper] illustrates the classifier's role in evidence identification, showing how it processes records from multiple databases.

2.5 Quality Assessment

Cochrane Crowd has successfully organized an online community of nearly 30,000 volunteers who have completed substantial screening tasks sustainably. Evaluation shows high accuracy, with crowd sensitivity at 99.1% and specificity at 99%. Less than 20% of records require arbiter identification. While individual volunteers take longer per abstract than experts, parallel group work offsets this, making total completion time far shorter than author teams.

Several studies have assessed volunteers' ability to identify literature relevant to specific systematic review topics based on inclusion criteria, demonstrating good sensitivity and specificity. This shows that while Cochrane Crowd's mainstream tasks focus on identifying specific study designs, the platform can also support topic-specific screening.

3 Implications of Cochrane Crowd for Evidence Synthesis Crowdsourcing

Although research on crowdsourcing in evidence synthesis remains nascent, experimental studies and Cochrane Crowd's successful operation demonstrate crowdsourcing's potential for improving evidence synthesis productivity. Cochrane Crowd's mechanisms offer several insights.

3.1 Quality Control Mechanisms

Cochrane Crowd's task-centered design includes rich learning activities, interactive customized training modules, diverse task access methods, and feedback mechanisms, improving correct task execution likelihood. Platform-level agreement algorithms aggregate crowd classifications, further improving accuracy. When applying crowdsourcing in evidence synthesis, crowdsourcers must establish appropriate processes and provide comprehensive training to ensure quality standards for screening, data extraction, and other activities.

3.2 Platform Usability

Cochrane Crowd features good interaction functions, task-related and platform-related facilities. From a design perspective, the platform follows simplicity principles with user-friendly navigation. Highlighting functions direct volunteer attention to key phrases, reducing participation burden. Currently, besides Cochrane Crowd, the evidence synthesis field has few independently developed crowdsourcing platforms, such as the CHEO Research Institute's InsightScope. Both platforms currently limit task posting to development teams and specific personnel, restricting applicability. There is an urgent need to develop universal evidence synthesis crowdsourcing platforms accessible to all evidence synthesis author teams across fields.

3.3 Human-Machine Collaboration

Cochrane Crowd generates large, high-quality datasets through collective effort, providing training material for machine learning classifiers. As the crowd screens more records, the RCT classifier becomes more accurate and efficient at deleting non-RCT records, allowing the crowd to focus on tasks requiring human intelligence. This exemplifies complementary human-machine collaboration. In evidence synthesis, machine learning classifiers typically require gold-standard classifications from professionals, which crowdsourcing can potentially provide as a data generation method.

4 Conclusion

Crowdsourcing is an effective method to improve evidence synthesis efficiency and shorten production cycles. Through comprehensive participant training, appropriate quality control mechanisms, clear objectives, clear tasks, and timely feedback or rewards, high-quality crowdsourcing results meeting evidence synthesis standards are achievable.

Interest and activity in introducing crowdsourcing to evidence synthesis are growing rapidly. As researchers from different disciplines apply crowdsourcing in evidence synthesis projects, new tools and platforms facilitating crowdsourcing need further development. Future research should explore crowdsourcing applications in evidence synthesis across different fields and at different synthesis stages.

References

[1] LITTELL J H. Conceptual and practical classification of research reviews and other evidence synthesis products[J]. Campbell systematic reviews, 2018, 14(1): 1-21.

[2] DONNELLY C A, BOYD I, CAMPBELL P, et al. Four principles to make evidence synthesis more useful for policy[Z]. Nature Publishing Group, 2018.

[3] BORNMANN L, MUTZ R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references[J]. Journal of the association for information science and technology, 2015, 66(11): 2215-2222.

[4] HADDAWAY N R, WESTGATE M J. Predicting the time needed for environmental systematic reviews and systematic maps [J]. Conservation biology, 2019, 33(2): 434-443.

[5] CHALMERS I, BRACKEN M B, DJULBEGOVIC B, et al. How to increase value and reduce waste when research priorities are set[J]. The lancet, 2014, 383(9912): 156-165.

[6] TSAFNAT G, GLASZIOU P, CHOONG M K, et al. Systematic review automation technologies[J]. Systematic reviews, 2014, 3(1): 1-5.

[7] BROWN A W, ALLISON D B. Using crowdsourcing to evaluate published scientific literature: Methods and example[J]. PLoS one, 2014, 9(7): e100647.

[8] SUN Y, CHENG P, WANG S, et al. Crowdsourcing information extraction for biomedical systematic reviews [J]. arXiv preprint arXiv: 1609.01017, 2016.

[9] NAMA N, ILIRIANI K, XIA M Y, et al. A pilot validation study of crowdsourcing systematic reviews: Update of a searchable database of pediatric clinical trials of high-dose vitamin D[J]. Translational pediatrics, 2017, 6(1): 18.

[10] NAMA N, SAMPSON M, BARROWMAN N, et al. Crowdsourcing the citation screening process for systematic reviews: Validation study[J]. Journal

of medical Internet research, 2019, 21(4): e12953.

[11] MORTENSEN M L, ADAM G P, TRIKALINOS T A, et al. An exploration of crowdsourcing citation screening for systematic reviews[J]. Research synthesis methods, 2017, 8(3): 366-386.

[12] PIANTA M J, MAKRAI E, VERSPOOR K M, et al. Crowdsourcing critical appraisal of research evidence (CrowdCARE) was found to be a valid approach to assessing clinical research quality[J]. Journal of clinical epidemiology, 2018, 104: 8-14.

[13] BUJOLD M, GRANIKOV V, SHERIF R E, et al. Crowdsourcing a mixed systematic review on a complex topic and a heterogeneous population: Lessons learned[J]. Education for information, 2018, 34(4): 293-300.

[14] Cochrane. About us[EB/OL]. [2022-09-04]. https://www.cochrane.org/about-us.

[15] Cochrane Collaboration. Cochrane Crowd [EB/OL]. [2022-06-18]. https://community.cochrane.org/help/tools-and-software/cochrane-crowd.

[16] SHANG H L, ZHANG S J, WEI Z P, et.al. Evidence integration framework of evidence-based digital humanities[J]. Journal of library and information science in agriculture, 2022, 34(11): 38-47.

[17] PHAM B, BAGHERI E, RIOS P, et al. Improving the conduct of systematic reviews: A process mining perspective [J]. Journal of clinical epidemiology, 2018, 103: 101-111.

[18] NOEL-STORR A, DOOLEY G, ELLIOTT J, et al. An evaluation of Cochrane Crowd found that crowdsourcing produced accurate results in identifying randomized trials[J]. Journal of clinical epidemiology, 2021, 133: 130-139.

[19] NOEL-STORR A, DOOLEY G, AFFENGRUBER L, et al. Citation screening using crowdsourcing and machine learning produced accurate results: Evaluation of Cochrane's modified Screen4Me service[J]. Journal of clinical epidemiology, 2021, 130: 23-31.

[20] NOEL-STORR A, GARTLEHNER G, DOOLEY G, et al. Crowdsourcing the identification of studies for COVID-19-related Cochrane Rapid Reviews[J]. Research synthesis methods, 2022, 13(5): 585-594.

[21] NOEL-STORR A H, REDMOND P, LAMÉ G, et al. Crowdsourcing citation-screening in a mixed-studies systematic review: A feasibility study[J]. BMC medical research methodology, 2021, 21(1): 1-10.

[22] THOMAS J, NOEL-STORR A, MARSHALL I, et al. Living systematic reviews: 2. Combining human and machine effort[J]. Journal of clinical epidemiology, 2017, 91: 31-37.

[23] NAKATSU R T, GROSSMAN E B, IACOVOU C L. A taxonomy of crowdsourcing based on task complexity[J]. Journal information science, 2014, 40(6): 823-834.

[24] PRPIC J, SHUKLA P P, KIETZMANN J H, et al. How to work a crowd: Developing crowd capital through crowdsourcing[J]. Business horizons, 2015, 58(1): 77-85.

[25] ESTELLÉS-AROLAS E, GONZÁLEZ-LADRÓN-DE-GUEVARA F. Towards an integrated crowdsourcing definition[J]. Journal information science, 2012, 38(2): 189-200.

[26] HOSSEINI M, PHALP K, TAYLOR J, et al. The four Pillars of crowdsourcing: A reference model [C]//2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS). Piscataway, NJ, USA: IEEE, 2014: 1-12.

[27] ZHAO Y X, ZHU Q H. Evaluation on crowdsourcing research: Current status and future direction [J]. Information systems frontiers, 2014, 16(3): 417-434.

[28] PEDERSEN J, KOCSIS D, TRIPATHI A, et al. Conceptual foundations of crowdsourcing: A review of IS research[C]//2013 46th Hawaii International Conference on System Sciences. Piscataway, NJ, USA: IEEE, 2013: 579-588.

[29] The Cochrane Library. Cochrane COVID-19 register of studies[EB/OL]. [2022-09-05]. https://covid-19.cochrane.org/.

[30] WALLACE B C, NOEL-STORR A, MARSHALL I J, et al. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach[J]. Journal of the American medical informatics association, 2017, 24(6): 1165-1168.

Author Contributions: Li Xiao: conceptualized the study, designed the research framework, and wrote the manuscript; Qu Jiansheng: revised the manuscript; Kou Leilei: revised the manuscript.

Note: Figure translations are in progress. See original paper for figures.