THIS.Institute The Healthcare Improvement Studies Institute

Made possible by The Health Foundation

Citizen science: crowdsourcing for systematic reviews

Lucy Strang and Rebecca K. Simmons

UNIVERSITY OF CAMBRIDGE

Contents

1 Summary
2 Introduction
2 Methods
3 Findings
Case study 1: Cochrane Crowd
Case study 2: CrowdScreen SR
Case study 3: Using an online marketplace to crowdsource citation screening
10 Conclusion
11 References

Authors
Lucy Strang, RAND Europe
Rebecca K Simmons, The Healthcare Improvement Studies Institute

The Healthcare Improvement Studies Institute aims to strengthen the evidence base for improving the quality and safety of healthcare.
www.thisinstitute.cam.ac.uk

RAND Europe is a not-for-profit research organisation that helps to improve policy and decision making through research and analysis.
www.randeurope.org

Published by: The Healthcare Improvement Studies Institute, 2018

ISBN: 978-1-9996539-1-0

Summary

Systematic reviews often require thousands of hours from expert reviewers to search, screen, appraise and synthesise relevant literature. As the literature base continues to grow, researchers have begun to explore citizen science approaches to conduct systematic reviews, with the aim of generating evidence more quickly and efficiently.

One such approach is crowdsourcing, which draws on a large pool of people who individually make small contributions that add up to big efforts. We explore the promising, albeit limited, evidence of the benefits of this approach, which suggests that citizen science approaches like crowdsourcing can make the systematic review process more efficient, timely and relevant. With appropriate quality control mechanisms and participant training in place, the outputs from crowdsourced reviews may be of a high enough quality to meet the threshold of a traditional 'gold standard' systematic review.

Some challenges arise when involving a large group of participants with diverse backgrounds in crowdsourced systematic reviews. Participant drop-out rates can be high. To encourage participation and retention, crowd participants should be provided with clear goals and well-defined tasks, as well as feedback and rewards. As in other types of research, it is important to ensure that projects are conducted ethically and responsibly, particularly in relation to potential crowd participant exploitation.

1. Introduction

Systematic reviews are often described as one of the cornerstones of evidence-based medicine.[1] A systematic review is a type of literature review that takes the principles of scientific method to collect and critically analyse all current, relevant, empirical research to answer a pre-specified question.[2] Given the scale of such a task, completing a systematic review is labour-intensive and costly, often requiring thousands of hours of input from expert reviewers. Compounding these challenges is the exponential growth in the production of and access to scientific research in recent years.[3] This expanded literature base obliges reviewers to screen more citations to find all relevant studies, increasing costs, delaying completion and increasing the risk of missing eligible studies.[1]

As a result, researchers have begun to explore alternative approaches to conducting systematic reviews that may allow them to complete reviews more quickly at lower costs. Some of these approaches rely on automation, via machine learning or artificial intelligence.[4,*] Other researchers have used citizen science approaches like crowdsourcing,[5] which is the focus of this learning report.

Crowdsourcing draws on large pools of people who individually make small contributions that add up to big efforts. It can be used in citizen science research projects, where 'citizens' – usually members of the public – provide inputs and valuable contributions despite not being formally trained experts in the topic of study.[6] These contributions take the form of 'micro tasks', which are 'discrete, small units of work, which can be done independently of each other'.[2] While the use of crowdsourcing to conduct systematic reviews is still in its infancy, several research groups have successfully engaged crowds to conduct various stages of a systematic review.

The aim of this report is to provide a practical overview of crowdsourcing for systematic reviews. It covers useful platforms and tools, opportunities and challenges, and illustrative examples. While the report does not intend to provide an exhaustive overview of these areas, it should serve as a useful and up-to-date practical reference for the design of a research programme that makes use of crowdsourcing for systematic reviews.

2. Methods

We gathered evidence for this learning report by conducting a rapid review of the literature, desk research (including review of websites) and interviews with crowdsourcing experts. We searched academic databases and used snowballing to identify relevant academic and grey literature. We also interviewed six academic researchers who have used crowdsourcing to conduct systematic reviews:

1. Dayre McNally, Clinical Investigator at the Children's Hospital of Eastern Ontario (CHEO) Research Institute, Canada

2. James Thomas, Director of the EPPI-Centre's Reviews Facility for the Department of Health and Social Care, England, and Professor of Social Research and Policy at Institute of Education, University College London, UK

3. Byron Wallace, Assistant Professor at the College of Computer and Information Science, Northeastern University and adjunct at the Center for Evidence Synthesis in Health, Brown University, United States

4. Adam Dunn, Senior Research Fellow in the Centre for Health Informatics in the Australian Institute of Health Innovation at Macquarie University, Australia

5. Grant Miller, Project Manager and Communications Lead, Zooniverse

6. Elaine Beller, Associate Professor, Biostatistics, Centre for Research in Evidence-Based Practice, –Bond University, Australia

We limit our discussion in this report to main findings and broad conclusions about the use of crowdsourcing to conduct systematic reviews.

* For an overview of discussions on automation in systematic review projects, see O'Connor et al. (2018).

3. Findings

Crowdsourcing for systematic reviews

Through crowdsourcing, non-traditional researchers can contribute at various stages of a systematic review. Table 1 outlines the main systematic review stages and the micro tasks that may be completed by a crowd for each stage. So far crowds have performed a specific stage in the process, rather than conducting an entire systematic review.

Table 1. Main systematic review tasks and potential micro tasks.
Adapted from Thomas et al. (2017)[2]

Review tasks	Potential micro tasks
Team formation	• Using a wider range of personnel than traditional research teams
Searching	• Running a search on bibliographic databases
Screening/eligibility assessment	• Selecting studies for inclusion in the review
Data extraction or collection	• Extracting information on characteristics of the participants, interventions, outcomes • Assessing risk of bias
Synthesis	• Entering data into meta-analysis software • Conducting meta-analyses • Report writing and updating conclusions
Supportive systems that reduce duplication of effort	• Data sharing and reuse

Platforms and tools

In our review, we identified a number of tools and platforms that can be used for crowdsourcing different systematic review tasks (summarised in Table 2). These platforms are diverse in their design and purpose: some were expressly developed for use in systematic review projects, while others have much wider applications beyond research activities. Five of the platforms are open source or free for the user, five are available to users on a fee-for-service or subscription basis, and two are currently reserved for the use of local developers or research networks. A number of the platforms combine machine learning with crowdsourcing or human effort. Most have been used by researchers for a particular stage of the systematic review (citation screening, for example), but four of the platforms have the potential to be used at any stage of the process. In addition, eight of the platforms offer users additional services and/or resources such as training modules. New tools continue to be developed and refined.

We identified five published evaluations of the use of crowdsourcing methods to conduct systematic reviews.[1, 5, 7, 8, 9] These evaluations focused primarily on using crowds for citation screening (four studies) and on data extraction (one study). While many of the evaluations included in this report examined tools that were still in the beta testing stages of their development, all the evaluations identified a number of advantages to using crowdsourcing for systematic reviews.

Efficiency gains

Citizen science approaches like crowdsourcing have the potential to allow tasks within a systematic review to be accomplished more quickly and at a lower cost than the traditional expert-led approach. For example, one evaluation showed that using crowd participants to screen citations at both the abstract and full-text level led to a substantial reduction in investigator workload, with a total work saving of 73% across the entire abstract and full-text reviewing process[1] (see Case study 2). Researchers can decide how much time they wish to save by applying different algorithms when aggregating crowd participants' screening decisions. These algorithms can prioritise *sensitivity* – casting a wide net to ensure that all relevant citations are captured, but increasing the workload of the reviewers who have to sift through the results – or *specificity* – targeting only the most likely relevant citations and minimising researcher effort at the screening stage. Efficiency gains can also be realised through the increased speed of the process. In one study, crowd participants completed screening within four to 17 days per systematic review, and approximately 15,000 screening decisions were made within 100 hours.[5]

Table 2. Tools and platforms for use in crowdsourcing systematic reviews

	Platform											
	Abstrackr[1]	Amazon Mechanical Turk[2]	Cochrane Crowd[3]	CrowdFlower[4]	CrowdScreen SR[5]	DistillerSR[6]	EPPI-Reviewer[7]	Mark2Cure[8]	RobotSearch[9]	Systematic Review Data Repository[10]	Upwork[11]	Weka[12]
Open source/free	✓							✓	✓	✓		✓
Fee-for-service/subscription		✓		✓		✓	✓				✓	
Reserved for developers' use			✓		✓							
Combined machine learning with human effort	✓		✓	✓					✓			✓
Used in evaluations of crowdsourcing	✓		✓		✓							
Connects participants with researchers		✓		✓	✓						✓	
Used in screening	✓	✓			✓	✓						
Used in identification of concepts/RCTs			✓					✓	✓			
Has potential to be used at any stage of process		✓									✓	✓
Offer users additional services and/or resources	✓	✓	✓			✓	✓	✓	✓	✓		

1 http://abstrackr.cebm.brown.edu
2 https://www.mturk.com
3 http://crowd.cochrane.org
4 https://www.crowdflower.com
5 http://www.cheori.org/en/crowdscreenOverview
6 https://www.evidencepartners.com/products/distillersr-systematic-review-software
7 https://eppi.ioe.ac.uk/CMS/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4&
8 https://mark2cure.org
9 https://github.com/ijmarshall/robotsearch
10 https://srdr.ahrq.gov
11 https://www.upwork.com
12 https://www.cs.waikato.ac.nz/ml/weka

When using a commercial platform to recruit for and crowdsource a systematic review (see Case study 3) crowdsourced screening decisions were much less expensive than using expert reviewers. Screening costs for crowd participants ranged from $460 to $2,220 USD for each review. This represents a cost reduction of up to 88% compared to using expert screeners.[5]

Data quality

All the studies we reviewed found that the quality of crowdsourced work was high. In an evaluation of Cochrane Crowd (see Case study 1), the crowd's sensitivity and specificity were both over 99% compared to the performance of an information specialist and a systematic reviewer.[7, 14] One evaluation used a number of different algorithms to aggregate crowd participant citation decisions. The research team's most conservative approach (considering a citation relevant where any of the five participating crowd participants chose to include it) identified 95–99% of the citations included in the expert review and excluded 68–82% of irrelevant citations.[5]

Challenges of using crowdsourcing for systematic reviews

Involving a large group of participants with diverse backgrounds creates a number of challenges. Attracting and retaining the right participants is crucial to the success of crowdsourced research, but can be difficult to achieve and is resource-intensive. Participants typically lack formal training in systematic reviewing and will have variable knowledge and competence. Researchers must also carefully consider intellectual property rights, copyright agreements and ethical issues.

Participant retention

Motivated and engaged participants are critical to the success of any citizen science project. However, evaluations of crowdsourced systematic review projects show that participant drop-out rates can be high. While some studies we identified attracted more volunteers than they needed,[1] others had difficulty retaining participants. A pilot study examining a citation screening task found that only 20 out of 100 people who expressed an interest in participating actually completed the task.[10] In a different study where the research team utilised a crowd to extract structured data from abstracts of articles describing randomised controlled trials (RCTs),[8] only six out of 20 participants passed the pre-qualification tests and were able to proceed with completing the task.

All our interviewees stressed the importance of providing participants with clear instructions, engaging consistently with them over the course of the project, setting and keeping to (ideally short) project timelines, and making tasks as easy to follow as possible. For a more detailed discussion of crowd motivation and participation, see the first learning report in our citizen science series.[11]

Quality control mechanisms

Researchers must put appropriate processes in place to ensure that crowdsourced screening and data extraction activities meet the high quality standards required in systematic reviews. One approach is to apply a threshold of decision consensus among crowd participants to manage participant error. Crowd participants should also be provided with training on the systematic review task. For example, in one evaluation, participants received mandatory training. They were asked to screen 20 practice citations and received feedback on their performance.[12]

Other researchers used two standard quality control mechanisms: hidden gold-standard control tests, known as honeypots, and qualification tests.[5] Qualification tests evaluated crowd participants' ability to correctly classify the citation they were screening. The research team provided crowd participants with four citations to screen, ranging in difficulty, and asked them to answer a series of questions on whether the citation met the review's inclusion criteria. If the crowd participant passed the qualification test, they could then work on the systematic review micro tasks. The next quality control step, early on in the screening process, was to insert 'honeypots' among the citations, in order to identify and eliminate unconscientious participants. For these tests, the participant's performance for a particular citation was evaluated against answers supplied by domain experts. If the participant answered the questions correctly, they continued screening uninterrupted. If they answered the questions incorrectly, they were informed of the correct answer and warned that more failed tests might exclude them from the project. The authors concluded that using these mechanisms produced relatively high accuracy at comparatively low cost.

As in other types of research, it is important to ensure that projects are conducted ethically and responsibly, particularly in relation to crowd participant exploitation. In the literature we reviewed, no feedback from crowd participants was included. However, these issues are explored in depth in our related report that explores crowdsourcing in research more broadly.[11]

4. Case studies

Case study 1: Cochrane Crowd

Cochrane is a network of researchers, healthcare professionals and patients who collaborate to produce independent, comprehensive reviews of health information.[13] The organisation has developed Cochrane Crowd†, an online platform where members of the public can contribute to its systematic reviews by carrying out micro tasks, such as assessing whether or not a study is a randomised controlled trial (RCT). Crowdsourcing, combined with algorithms to safeguard accuracy and improve efficiency, has brought a significant, and welcome boost to Cochrane's systematic review activity.

The crowd's tasks

Cochrane Crowd reviewers do not conduct all steps of a systematic review, but rather focus on binary classification – determining whether or not an article describes, for example, an RCT. The Cochrane team chose this approach after it found that engaging a crowd in the full systematic review process was very challenging. The team decided to break the process down into manageable tasks where accuracy can be assessed to ensure that results are reliable.

In addition to article classification, the Cochrane Crowd team is developing additional tasks for the crowd related to the identification of basic research study information such as Patient, Population or Problem, Intervention, Comparison and Outcome (PICO).

Platform and tools

The Cochrane team developed its own platform, which integrates with the other parts of the organisation's IT infrastructure. An important part of the system is an algorithm that assesses whether a sufficient level of agreement has been reached within the crowd about a particular record, or whether a further check to resolve uncertainty or disagreement is needed by a more experienced screener (a 'resolver' screener). The algorithm currently requires that an article receive four consecutive agreements by screeners (who are blind to the decisions made by others) before it is classified. The algorithm previously required three successive agreements,[14] but this was increased to four to safeguard accuracy when the tool was opened up to a wider group of screeners.

The platform includes a training feature, which allows participants to learn and practice how to screen, and a text highlighting function that helps screeners quickly identify the relevant parts of a record.[14] This has been shown to help screeners work more quickly and more accurately.[14]

The crowd's activity also generated a large dataset, which has been used to train machine-learning algorithms to automatically identify RCTs. These algorithms work better than search filters for identifying RCTs[15] and significantly reduce the number of articles that require human screening.[9]

The crowd

While the platform is open to anyone, most contributors do have medical experience or expertise. Volunteers are required to train on 20 practice records before they begin.[9] Screeners are ranked at different levels (novice, expert, resolver) and unlock rewards and tasks as they complete more classifications. As of June 2017, the Cochrane Crowd platform had nearly 6,000 contributors, who had screened more than a million articles.

Lessons learned

Evaluations have shown that the crowd performs very accurately, indicating that 'non-traditional' reviewers can provide valuable contributions to the systematic review process. The crowd's sensitivity and specificity were both above 99% compared to the performance of an information specialist and a systematic reviewer.[7, 14]

The Cochrane Crowd team has discovered that crowdsourcing requires significant investment to ensure it runs effectively. Challenges with participant drop-out are reported. Time is required to monitor the crowd and its activity, to engage with participants (e.g. through social media), to encourage their continued involvement, and to run evaluations.

The Cochrane Crowd team recommends making it as easy as possible for participants to carry out tasks, and suggests that tasks can be made easier through:

(i) the use of user-friendly interfaces;
(ii) accessible training and support;
(iii) time-saving aids, such as a text highlighting function.

† http://crowd.cochrane.org

Case study 2: CrowdScreen SR

CrowdScreenSR[‡] is a prototype crowdsourcing platform where individuals can contribute to systematic review projects by screening citations. It was set up by a team in the Children's Hospital of Eastern Ontario (CHEO) Research Institute who recognised that students at the University of Ottawa Medical School were interested in participating in research projects and were capable of conducting work that was being done by busy clinicians and investigators. The CHEO team has completed one pilot validation study, where it compared the results of a paediatric Vitamin D systematic review performed through crowdsourcing to those generated using the traditional gold-standard, trained expert approach.[1]

The crowd's tasks

Participants completed a 'demo' module to enable them to practice on 16 abstracts and nine full-text citations from the original systematic review. They were given feedback on the accuracy of their assessment. Following this informal training stage, participants were asked to assess the eligibility of both abstract and full-text citations for inclusion in the systematic review and to classify the citation as either 'retain', 'exclude' or 'unclear – I cannot assess this citation'. When a citation was excluded, the participant was asked to indicate which eligibility criteria were not met. Citations were randomly allocated and participants could screen as many citations as they liked. Each citation was evaluated a minimum of four times by participants. A threshold of consensus among the participants on the classification of the citation was required before the citation was ultimately either retained, reviewed further by the investigative team, or excluded.

Platform and tools

The research team originally designed the CrowdScreenSR platform using third-party software called Knack, which requires some technical expertise to operate. As a consequence, researchers need to spend a few hours on the back end of the platform to set up every new review. While the software has the capacity to manage hundreds of thousands of records at a time, the Research Institute had to pay per citation to use it, which limited the project's scalability.

Through the team's contacts with the Hacking Health[§] initiative, researchers are currently in the process of improving the platform to make it more intuitive to use. It will also become more scalable and allow investigators to set up projects themselves and upload their own citations. For this process, researchers have moved to a new software called InsightScope, and is using REDCap, a secure online web application for building and managing online databases, as its model. Researchers are presently testing the "beta" version of this software using systematic reviews at CHEO, with the expectation that the platform will be released to the broader scientific community later in 2018. They hope to develop the platform so that people can log on to the website, find projects of interest to them, and contribute to a systematic review.

The crowd

For the pilot project, the researchers sought to recruit individuals with post-secondary education and medical training, who had not been involved in the design of the systematic review protocol and had not received training from the investigators on how to screen citations.

Researchers were initially concerned about recruitment, as it did not have the financial resources to pay for participants' contribution to the systematic review. Nor were researchers in a position to offer authorship to participants. In subsequent projects, the researchers offered $100 gift cards to the top three reviewers in each review. However, moving forward, they will no longer offer financial incentives. Instead they will link students with investigators and suggest opportunities to qualify for full authorship by taking on further responsibilities within the systematic review. Feedback suggested that these opportunities are highly valued by participants.

Lessons learned

The evaluation showed that the crowd performed very accurately. Participants retained 100% of eligible citations through the screening process. Utilising the crowd reduced the potential workload of the research team to 27%, calculated as the percentage of citations that the crowd either excluded or retained without the investigative team needing to evaluate (for conflicts or confirming eligibility at full text). When the algorithm was adjusted to prioritise sensitivity (citation excluded if all four reviewers agree), the workload reduction at the abstract stage was 73% while sensitivity was maintained at 100%. When the algorithm prioritised specificity (citations not obtaining three out of four 'retain' assessments excluded), the work required by the research team decreased to 8%, while reducing sensitivity to 85%. While this is a significant reduction in sensitivity, there is increasing evidence that traditional systematic review approaches may actually achieve sensitivities well below 85%. Another benefit of this project was the dozens of meaningful connections that were forged between students

‡ http://www.cheori.org/en/crowdscreenOverview

§ Hacking Health is a global non-profit organization that seeks to bring together technology creators, medical professionals and other stakeholders to collaborate on digital healthcare solutions. http://hacking-health.org/

and investigators conducting systematic reviews in areas of interest to them.

However, the process requires a significant degree of supervision. The research team noted that potential participants' assessment of their ability to commit to a project is typically based on their current capacity. With systematic reviews often stretching over months or even years, crowdsourced projects run the risk of losing participants over time. Researchers need to ensure that projects move quickly in order to ensure crowd retention. Researchers also need a degree of caution in assuming that participants who have performed well in a particular systematic review can perform equally well on a different review. They have found that some participants who enjoyed their first project have struggled on subsequent reviews. It is important to include training and an assessment at the start of a project to inform whether a participant is suitable.

Finally, a committed team is needed to run a platform such as CrowdScreenSR. It requires multiple people managing the front end, for example answering questions from researchers about methodology issues, and at the back end, for example maintaining the system and integrating new records.

Case study 3: Using an online marketplace to crowdsource citation screening

A research team – based at Brown University in the US – was motivated by the expense and long timelines of traditional systematic reviews to explore alternative methods to modernise aspects of the review process. They tested the use of paid crowd participants to complete the citation screening stage of a systematic review. The team used convenience samples of citation data and screening decisions, which had previously been gathered in four systematic reviews conducted by their teams at Tufts Medical Center and the Center for Evidence-based Medicine at Brown University. These data were compared to the results generated from crowd participants to make screening decisions on the same reviews.

The crowd's tasks

The crowd was asked to make screening decisions on citations by answering sets of simple successive pattern-matching and information-extraction questions regarding study eligibility sub-criteria. These questions were designed to be answered with minimal understanding of the context or methods of the study. Available answers were 'yes', 'no', and 'I cannot tell'. 'Yes' and 'I cannot tell' indicated possible inclusion, while 'no' indicated definite exclusion. Participants were given definitions of technical terms and illustrative positive and negative examples of sub-criteria to help in making their decisions. The team bundled three citation decisions into each task, with crowd participants paid $0.15 to $0.21 USD per task.

Platform and tools

The research team used the MTurk platform, an online marketplace owned by Amazon, where potential 'employers' may post tasks known as Human Intelligence Tasks (HITs) for crowd participants. It is the most widely used crowdsourcing platform and provides easy access to a large pool of available participants, with built-in payment and participant systems. It has an extensive application programming interface, which enables researchers to incorporate tools such as quality controls and qualification tests.[5] Participants can search for and accept HITs, some of which, as with this systematic review study, require qualifying tests. Qualified crowd participants are then given a set of HITs, which they complete sequentially. When HITs are submitted, their work is sent to the employer who may either accept it as meeting the instructions, or reject it. When the work is accepted, the crowd participant receives payment.

The crowd

Crowd participants were recruited through MTurk. Crowd participants on MTurk are from increasingly diverse backgrounds, a mix of moderate-income, US-based participants and a young and well-trained cohort from developing economies.[16] They are not likely to have any background in evidence-based medicine, especially given the diversity of tasks posted on MTurk. For this study, crowd participants were required to pass a qualification test. In addition, after the participants were selected, the study team used hidden control tests, or 'honeypots', to identify and remove underperforming crowd participants.

Lessons learned

The crowd performed very accurately across the four reviews, showing a high sensitivity for relevant citations compared to manual screening, although at some cost in specificity. The team used nine different algorithms in aggregating crowd participants' decisions on including citations. The study team's most conservative approach (considering a citation relevant where any of the five crowd participants chose to include it) identified 95–99% of the citations that were ultimately included in the reviews, while excluding 68–82% of irrelevant citations.

The study found that using crowdsourcing to screen citations was inexpensive compared with using expert reviewers. Screening was completed within four to 17 days per review and cost between $460 and $2,220 USD, representing a cost reduction of up to 88% compared to using expert screeners (Case study 1). Crowdsourcing also produced efficiency gains through the speed of the process: approximately 15,000 screening decisions were made within 100 hours for one of the reviews.

However, the research team found that there was no easy way to test or recruit suitable crowd participants using MTurk. Other platforms may be more suitable for targeting appropriate crowd participants. In addition, the team built its own citation screening interface for the project as the MTurk interface was inflexible and not particularly user-friendly. They suggest that the traditional approach to conducting systematic reviews is changing and the future will involve a hybrid of automation, distributed work (through crowd participants) and domain expertise.

Table 3. Costs and duration of each crowdsourcing experiment
Adapted from Mortensen et al. (2017)[5]

Dataset	Participant salary (with Amazon fee[a])	Approximate cost of experts' screening (with Fringe[b])	Experiment running time (after task setup)
Proton beam	$1,187.25 ($1,305.98)	$6,859.67 ($8,917.57)	4d, 21h, +36m
Appendicitis	$416.00 ($457.60)	$3,034.23 ($3,944.50)	5d, 10h, +58m
DST	$2,017.75 ($2,219.53)	$6,173.75 ($8,025.88)	16d, 20h, +11m

[a] At the time the reseach team ran the experiments, Amazon Mechanical Turk charged a 10% commission fee on HIT, with a minimum payment of $0.005 per HIT; this has since been increased to 20%. (http://requestersandbox.mturk.com/pricing).

[b] Fringe benefits costs are estimated here to be 30% of salary, reflecting (roughly) the true costs at the institutes at which this work was performed (Tufts and Brown).

5. Conclusion

While research on the use of crowdsourcing to conduct systematic reviews is still in its infancy, there is already promising evidence of the benefits of this approach as well as the risks that need to be mitigated. Crowdsourcing offers researchers a way to make systematic reviews more efficient, timely and relevant. It also has the potential to allow tasks within a review to be accomplished at a lower cost to the research team than the traditional expert-led approach. With appropriate quality control mechanisms and participant training in place, the outputs from crowdsourced reviews may be of a high enough quality to meet the threshold of a traditional 'gold standard' systematic review.

We have highlighted a number of challenges to consider when involving a large group of participants with diverse backgrounds in research projects. Since participant drop-out rates can be high, individuals should be provided with clear goals and short, well-defined tasks, as well as feedback and rewards. As in other types of research, it is important to ensure that projects are conducted ethically and responsibly, particularly in relation to crowd participant exploitation.

Interest in and activity around the crowdsourcing of systematic reviews is rapidly growing. New tools and platforms that facilitate crowdsourcing continue to be developed, as researchers in diverse disciplines – including healthcare improvement – expand the use of citizen science approaches in systematic reviews projects. Future research should continue to develop the underlying methodology at the same time as producing the reviews themselves.

6. References

1. Nama, N., K. Iliriani, M.Y. Xia, B.P. Chen, L.L. Zhou, S. Pojsupap, C. Kappel, K. O'Hearn, M. Sampson, K. Menon, & J.D. McNally. 2017. A Pilot Validation Study of Crowdsourcing Systematic Reviews: Update of a Searchable Database of Pediatric Clinical Trials of High-Dose Vitamin D. *Translational Pediatrics* 6 (1): 18–26.
2. Thomas, J., A. Noel-Storr, I. Marshall, B. Wallace, S. McDonald, C. Mavergames, P. Glasziou, I. Shemilt, A. Synnot, T. Turner, & J. Elliot. 2017. Living Systematic Reviews: 2. Combining Human and Machine Effort. *Journal of Clinical Epidemiology* 91: 31–37.
3. Kim, S. & J. Choi. 2014. An SVM-Based High-Quality Article Classifier for Systematic Reviews. *Journal of Biomedical Informatics* 47: 153–159.
4. O'Connor, A.M., G. Tsafnat, S.B. Gilbert, K.A. Thayer, & M.S. Wolfe. 2018. Moving Toward the Automation of the Systematic Review Process: A Summary of Discussions at the Second Meeting of International Collaboration for the Automation of *Systematics Reviews* (ICASR). *Systematic Reviews* 7: 3.
5. Mortensen, M., G.P. Adam, T.A. Trikalinos, T. Kraska, & B.C. Wallace. 2017. An Exploration of Crowdsourcing Citation Screening for Systematic Reviews. *Research Synthesis Methods* 8: 366–386.
6. Swan, M. 2012. Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem. *Journal of Medical Internet research* 14 (2): e46.
7. Noel-Storr A., C. Struthers, S. Cullum, R. McShane, S. Creavin, D. Davis, & K. Huckvale. 2013. 'Many Hands Make Light Work – Or Do They? Results of Two Pilot Studies Looking at the Effects of Crowdsourcing.' In Better Knowledge for Better Health | *Un meilleur savoir pour une meilleure santé*. Abstracts of the 21st Cochrane Colloquium. Québec City, Canada: John Wiley & Sons. As of: 27 February 2018: http://abstracts.cochrane.org/2013-qu%C3%A9bec-city/many-hands-make-light-work%E2%80%94or-do-they-results-two-pilot-studies-looking-effects
8. Sun Y., P. Cheng, S. Wang, H. Lyu, M. Lease, I. Marshall, & B.C. Wallace. 2016. Crowdsourcing Information Extraction for Biomedical Systematic Reviews. arXiv:1609.01017 [cs.HC].
9. Wallace, B,C., A. Noel-Storr, I.J. Marshall, A.M. Cohen, N.R. Smalheiser, & J. Thomas. 2017. Identifying Reports of Randomized Controlled Trials (RCTs) via a Hybrid Machine Learning and Crowdsourcing Approach. *Journal of the American Medical Informatics Association* 24 (6): 1165–1168. doi:10.1093/jamia/ocx053
10. Noel-Storr A. (2013) Many Hands Make Light Work: Results of Two Pilot Studies Looking at the Effects of Crowdsourcing. Presentation to Cochrane Colloquium, September 2013. As of 27 February 2018: https://www.researchgate.net/publication/263084731_Many_hands_make_light_work_Results_of_two_pilot_studies_looking_at_the_effects_of_crowdsourcing
11. Lichten, C., R. Iopollo, C. D'Angelo, RK. Simmons & M. Morgan Jones. 2018. 'Citizen science: crowdsourcing for research.' THIS Institute (The Healthcare Improvement Studies Institute).
12. Wallace B.C., I.J. Dahabreh, C.H. Schmid, J. Lau, & T.A. Trikalinos. 2013. Modernizing the Systematic Review Process to Inform Comparative Effectiveness: Tools and Methods. *Journal of Comparative Effectiveness Research* 2 (3): 273–282.
13. Cochrane. 2017. 'Our Vision, Mission, and Principles.' As of 27 February 2018: http:// cochrane.org/uk/about-us/our-vision-mission-and-principles
14. Thomas, J., A. Noel-Storr, & J. Elliott. 2015. 'Human and Machine Effort in Project Transform: How Intersecting Technologies Will Help Us to Identify Studies Reliably, Efficiently and at Scale.' In Cochrane Methods edited by J. Chandler, J. McKenzie, I. Boutron, & V. Welch. *Cochrane Database of Systematic Reviews* Suppl. 1: 37–41.
15. Marshall I., A.H. Noel-Storr, J. Kuiper, J. Thomas, & B.C. Wallace. 2017. 'Machine Learning for Identifying Randomized Controlled Trials: An Evaluation and Practitioner's Guide.' *Research Synthesis Methods* 2018: 1–12.
16. Ross J., L. Irani, M. Silberman, A. Zaldivar, & B. Tomlinson. 2010. 'Who Are the Crowdworkers?: Shifting Demographics in Mechanical Turk.' CHI'10 Extended Abstracts on Human Factors in Computing Systems. ACM, 2863–2872.

At THIS Institute we aim to strengthen the evidence base for improving the quality and safety of healthcare.

Co-created by two exceptional organisations – the University of Cambridge and the Health Foundation – THIS Institute is founded on the guiding principle that efforts to improve care should be based on the highest quality evidence.

Our work is defined by a highly inclusive approach that combines academic rigour with the real concerns of patients and staff. We're open, transparent, and we do not shy away from difficult subjects.

Together we're creating an evidence base to improve healthcare.

THIS Institute is made possible by the Health Foundation.

ISBN: 978-1-9996539-1-0

THIS.Institute The Healthcare Improvement Studies Institute

THIS Institute
School of Clinical Medicine
University of Cambridge
Cambridge Biomedical Campus
Hills Road
Cambridge CB2 0AH

thisinstitute.cam.ac.uk
@THIS_Institute

UNIVERSITY OF CAMBRIDGE

Made possible by
The Health Foundation