CONFERENCE ARTICLE

EFSA Journal

APPROVED: 29 May 2019

doi: 10.2903/j.efsa.2019.e170715

Working with a new kind of team: harnessing the wisdom of the crowd in trial identification

Anna Noel-Storr

Cochrane Dementia and Cognitive Improvement Group, University of Oxford, United Kingdom

Abstract

Background: At a time when research output is expanding exponentially, citizen science, the process of engaging willing volunteers in scientific research activities, has an important role to play in helping to manage the information overload. It also creates a model of contribution that enables anyone with an interest in health to contribute meaningfully and in a way that is flexible. Citizen science models have been shown to be extremely effective in other domains such as astronomy and ecology. Methods: Cochrane Crowd (crowd.cochrane.org) is a citizen science platform that offers contributors a range of microtasks, designed to help identify and describe health research. The platform enables contributors to dive into needed tasks that capture and describe health evidence. Brief interactive training modules and agreement algorithms help to ensure accurate collective decision making. Contributors can work online or offline; they can view their activity and performance in detail. They can choose to work in topic areas of interest to them such dementia or diabetes, and as contributors progress, they unlock milestone rewards and new tasks. Cochrane Crowd was launched in May 2016. It now hosts a range of microtasks which help to identify health evidence and then describe it according to a PICO (Population; Intervention; Comparator; Outcome) ontology. The microtasks are either at 'citation level' in which a contributor is presented with a title and abstract to classify or annotate, or at the full-text level in which a whole or a portion of a full paper is displayed. Results: To date (March 2019), the Cochrane Crowd community comprises over 12,000 contributors from more than 180 countries. Almost 3 million individual classifications have been made, and around 70,000 reports of randomised trials have been identified for Cochrane's Central Register of Controlled Trials. Performance evaluations to assess crowd accuracy have shown crowd sensitivity is 99.1%, and crowd specificity is 99%. Main motivations for involvement are that people want to help Cochrane, and people want to learn. Conclusion: This model of contribution is now an established part of Cochrane's effort to manage the deluge of information produced in a way that offers contributors a chance to get involved, learn and play a crucial role in evidence production. Our experience has shown that people want to be involved and that, with little or no prior experience, can do certain tasks to a very high degree of collective accuracy. Using a citizen science approach effectively has enabled Cochrane to better support its expert community through better use of human effort. It has also generated large, high-quality data sets on a scale not carried out before which has provided training material for machine learning routines. Citizen science is not an easy option, but performed well it brings a wealth of advantages to both the citizen and the organisation.

Keywords: crowdsourcing, citizen science, meta-analysis, systematic review, randomised controlled trial, microtask

Correspondence: anna.noel-storr@rdm.ox.ac.uk

Suggested citation: Noel-Storr A, 2019. Working with a new kind of team: harnessing the wisdom of the crowd in trial identification. EFSA Journal 2019;17(S1):e170715, 8 pp. https://doi.org/10.2903/j.efsa.2019.e170715

ISSN: 1831-4732

The EFSA Journal is a publication of the European Food Safety Authority, an agency of the European Union.

Table of contents

1. Introduction

Cochrane is a global independent organisation with a mission to produce high-quality, relevant, accessible health evidence. As one of the world's leaders in producing systematic reviews assessing the effectiveness of health interventions across a broad range of healthcare domains, the timely identification of relevant evidence is critical. Once identified, the evidence can then be synthesised or summarised, either quantitatively or qualitatively, to determine whether a treatment is effective or not.

Over the last two decades the amount of research produced, in various formats including published articles in peer-reviewed journals, conference outputs or trial registry entries has grown exponentially, with global scientific output believed to be doubling every 9 years (Van Noorden, 2014; Bornmann and Mutz, 2015). Keeping up with the constant and increasing flow of information presents information brokering organisations like Cochrane with significant challenges. The longer it takes to produce systematic reviews, the longer uncertainty about treatments remains (Borah et al., 2017). Timely identification of relevant evidence is therefore critical in the process to turn information into knowledge about treatment effects.

Another challenge is around being able to provide willing contributors with meaningful ways to contribute to the research synthesis process. Cochrane has a long history or volunteer engagement but as reviews become increasingly complex, the opportunities for involvement have narrowed. A greater range of opportunities is therefore needed to ensure that people can play a meaningful and helpful role in producing robust health evidence.

In 2016, Cochrane launched Cochrane Crowd (http://crowd.cochrane.org), an online platform that offers potential contributors a range of 'microtasks' that help to identify and describe health research as it is published or produced. This paper describes how Cochrane Crowd works, looking in detail at the first microtask we launched, and how this model of contribution is transforming Cochrane's approach to producing systematic reviews.

2. Methods

2.1. Crowdsourcing

Crowdsourcing is the practice of enlisting help or input from a large number of people into a project or task, usually via the internet. The type of input sought from the crowd can vary hugely, and the type of input sought will usually inform the way the task or project is put to the crowd. For example, an organisation seeking crowd input to help determine whether a new product will sell well will require a very different crowd approach than a citizen science project wanting members of the general public to record the number and types of birds that visit their back gardens in the winter months. Choosing the right form of crowdsourcing depends on the problem the organisation is trying to solve (Brabham, 2008; Ikediego et al., 2018).

In Cochrane, our problem was similar to that of many other research organisations – that of keeping up with the ever increasing flow of information, much of it of dubious quality. We observed how other domain areas and organisations were tackling this problem, particularly through citizen science projects in areas such as astronomy, ecology and the humanities in which large amounts of data either needed collecting, processing or classifying. In these projects, citizens would be asked to perform relatively simple, often visual, tasks. Multiple decisions would be collected and, using agreement algorithms, help generate a collective truth.

2.2. The microtask

The tasks available on Cochrane Crowd are called microtasks because they represent tasks broken down into smaller, usually single-question-based tasks. This paper will focus on Cochrane Crowd's main microtask: the identification of reports of randomised controlled trials, from here referred to as randomised controlled trials (RCT) ID. For this task, the contributor is presented with a title and abstract or a journal or conference publication, and asked to classify the record in one of three ways: RCT/CCT, Reject or Unsure. A classification of RCT/CCT means that the contributor feels confident that the record is indeed describing or reporting a randomised or quasi-randomised trial. Reject means the contributor is confident the record is not a randomised or quasi-randomised trial, and Unsure is used where the contributor is unable to make a decision (see Figure 1).

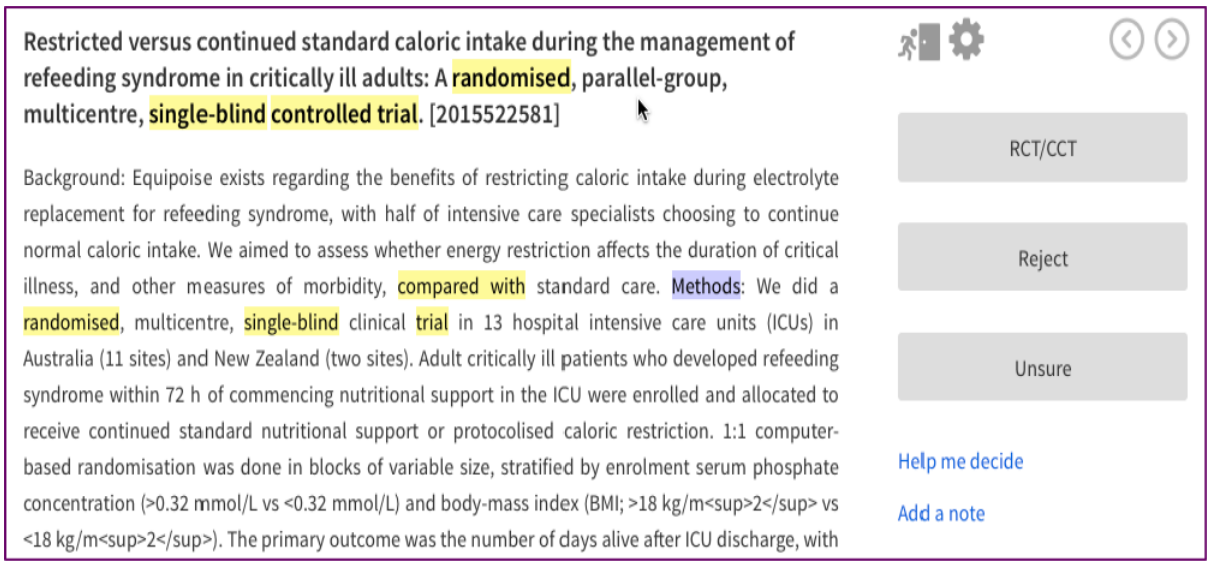

Figure 1: A screen shot from the Cochrane Crowd RCT ID task

2.3. Task training and guidance

All microtasks on Cochrane Crowd are supported by brief, mandatory, interactive training modules. The RCT ID training module is made up of 20 practice records, each one designed to introduce the contributor to key aspects of randomised trial design and how those aspects might be described in a journal or conference abstract. Once compete, a contributor can progress to the live task and start classifying 'live' records.

In addition to the training module, each task has a 'Quick Reference Guide' – a brief look-up table summarising the main points from the training module, accompanied by examples. And, if the contributor get stuck on a particular record, the RCT ID microtask also has an on-screen *Help Me Decide* feature. This feature guides a contributor through a series of questions to help them make a correct decision on a record.

2.4. Task agreement algorithm

The task-specific training is an important component in helping to ensure the accuracy of each individual contributor. However, in a crowdsourced model like Cochrane Crowd, it is collective accuracy that is critical in ensuring reliable and consistent high-quality data output. To ensure high collective accuracy, an agreement algorithm is needed. This algorithm determines how many, and in what order, individual classifications are warranted. The agreement algorithm in place for the RCT ID microtask is that for each record, four independently made consecutive agreeing classifications are needed for no further human intervention to be required (Figure 2). For example, if a record receives four *RCT* classifications made by four contributors, the record will be deemed to be reporting or describing an RCT. Conversely, if a record receives four *Reject* classifications consecutively from four contributors, the record will deemed to be a reject (i.e. not describing or reporting an RCT). When the consecutive chain is broken, records go to a Resolver-level screener within Cochrane Crowd, as do records that receive an *Unsure* classification.

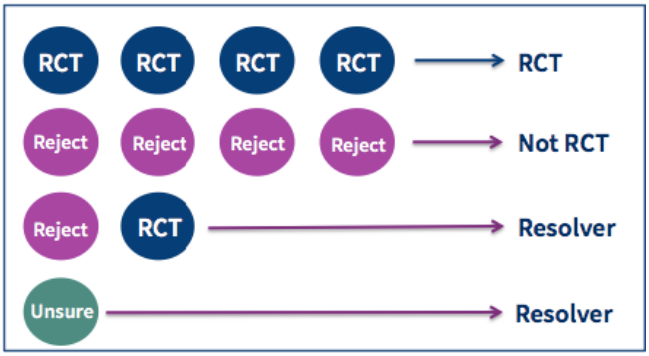

Figure 2: The current RCT ID agreement algorithm

2.5. The Crowd

Anyone can join Cochrane Crowd. Sign up is required but it requires only a name and email address. Everyone begins Cochrane Crowd as a 'normal' contributor. However, contributors can progress to 'expert' status within each of the available Cochrane Crowd microtasks. For the RCT ID task, a contributor needs to complete 1,000 classifications to a very high degree of accuracy in combination with a very low proportion of *Unsure* classifications to be eligible to become an expert for that particular microtask. Once an expert, their decisions carry more weight for that task bringing increased system efficiency as fewer decisions are needed on records to derive a reliable collective answer.

2.6. Machine learning

The primary purpose of the RCT ID microtask on Cochrane Crowd is to identify reports of RCTs or quasi-RCTs for Cochrane's Central Repository of Controlled Trials (https://www.cochranelibrary.com/) so that researchers and systematic reviewers around the world can find the evidence they need, when they need it. However, an additional output of this crowdsourced task has been the generation of high-quality training data for the machine. Machine learning gives 'computers the ability to learn without being explicitly programmed'. In the context of Cochrane, this is about building classifiers that provide likelihood scores on records. For the RCT ID microtask, the identification of tens of thousands of RCTs, and hundreds of thousands of Reject records has been used to build, train, test and validate an RCT classifier (Thomas et al., 2017; Wallace et al., 2017; Marshall et al., 2018).

3. Results

3.1. High-level performance metrics

Cochrane Crowd was launched in May 2016. To date (March 2019), over 12,000 people have signed up. The platform hosts five long-term microtasks. Almost 3 million individual classifications have been made by contributors on over 500,000 records that have so far been through the Cochrane Crowd platform.

3.2. The Crowd

The Cochrane Crowd community is a truly global community with contributors based in over 180 countries of the world with most aged between 26 and 35 years old. In a survey sent to a sample of the community (1,000 contributors who had joined up during a 6-month period), we learnt that a significant proportion of respondents either work, teach or study in a health-related area indicating an educated and possibly health research savvy crowd. Main motivations for contributing were the altruistic desire to help, to participate in something useful, and to learn. In addition, many had previously heard of Cochrane and wanted to contribute to Cochrane Crowd hoping that it might lead to further opportunities for involvement with the organisation.

3.3. Crowd accuracy

We measure Crowd accuracy for each task in Cochrane Crowd by comparing the Crowd's collective decisions against a gold standard. For the RCT ID microtask, we took a month's worth of records and got them classified outside of the Cochrane Crowd system by experts with extensive experience of screening citations for RCTs. We then compared the Crowd's collective decisions on the same set of records to the experts' decisions. Of the 6,041 records, the Crowd collectively misclassified four records as not reporting an RCT equating to four false-negative decisions; the Crowd collectively misclassified 58 records as reporting an RCT when in fact the records should have been rejected (the false positives). Overall, therefore Crowd's sensitivity (the Crowd's collective ability to accurately identify the records that should be classified as RCTs) and the Crowd's specificity (the Crowd's collective ability to accurately identify the records that should be classified as Reject records) was 99.1% and 99.0%, respectively.

3.4. System efficiency

In terms of system efficiency, the crowd model only (i.e. without the use of machine learning) is able to handle around four times more records than the previous expert-only model that used to be in operation. In 2016, when we used just the Crowd (so no machine learning at that point), the Crowd handled around 150,000. When we added the machine learning classifier into the workflow, we were able to handle double that amount. In fact, the machine classifier we built for the RCT identification task now handles around 70% of the records that we import into our system. The machine acts as a very effective 'noise reducer' leaving the trickier records for the humans.

4. Discussion

These results indicated the huge potential of using crowdsourcing in this way for certain tasks. We have shown that a crowd can successfully be engaged in helping to identify reports of randomised trials.

The success of such a model comes down to three components: the crowd, the task and the agreement algorithm. Our crowd is clearly a reasonably 'expert' crowd. While we hope to attract a broad contributor base, we have, to date, appealed largely to those who are already familiar with health research. This of course takes some of the pressure off the training and support aspects of the task, but it does not lessen the need for the task itself it be well designed and, critically, doable. Microtasking in this way will only work if the question or questions being asked are answerable. Questions that are subjective will not work well. Having variation and complexity in the content being classified is workable but there must be a right and a wrong answer. The agreement algorithm is the final critical component. A poor performing algorithm will enable individual errors to carry too much weight in the system and lead to inconsistent output in terms of accuracy. The current algorithm in place for RCT ID works well in terms of ensuring Crowd sensitivity. Specificity is good too but with around 15% of records needing to be resolved we recognise that there is room for improvement here.

The RCT classifier has had a significant impact in enabling us to scale the quantity of records that can go through the system. For the RCT ID microtask, we have seen how a once human-only task has evolved to a task shared between the human and the machine (Wallace et al., 2017; Marshall et al., 2018). A virtuous cycle has been created in which human generated data have helped to build and to train a machine to perform a portion of the task. This cycle demonstrated by the RCT ID use case is one that we hope to follow for all Cochrane Crowd microtasks, and for the RCT ID we anticipate that the machine will continue to handle more and more records as the classifier is trained on more and more records fed to it from the Crowd.

However, crowdsourcing is not an 'easy option'. It takes time, resource and skill to ensure that the required components are in place. While many view crowdsourcing as a technical solution to information overload, at the heart of it lies people. It truly represents a new kind of team and like all successfully functioning teams, there needs to be a clear and strong sense of shared mission and a good understanding of motivations and expectations from both contributors and the host organisation alike.

5. Conclusion

At a time when research output is growing exponentially, new methods and processes are needed to help researchers and others keep up with the flood of information. Cochrane Crowd is helping to do just that. In the 3 years since launch, Cochrane Crowd has attracted thousands of people from around the world who want to play a valuable role in the research process, by helping to turn information into knowledge about treatments. This model of contribution has not only enabled us to identify thousands of reports of trials and other studies, it has enabled us to train machine classifiers to carry out the tasks, so that human effort can be directed where it is needed most.

References

Borah R, Brown AW, Capers PL and Kaiser KA, 2017. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. British Medical Journal Open, 7, e012545.

Bornmann L and Mutz R, 2015. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology, 66, 2215–2222. https://doi.org/10.1002/asi.23329

Brabham DC, 2008. Crowdsourcing as a model for problem solving: an introduction and cases. Convergence: The International Journal of Research into New Media Technologies, 14, 75–90.

Ikediego HO, Ilkan M, Abubakar AM and Bekun FV, 2018. Crowd-sourcing (who, why and what). International Journal of Crowd Science, 2, 27–41. https://doi.org/10.1108/IJCS-07-2017-0005

Marshall IJ, Noel-Storr AH, Kuiper J, Thomas J and Wallace B, 2018. Machine learning for identifying Randomized Controlled Trials: an evaluation and practitioner's guide. Research Synthesis Methods, 9, 602–614.

Thomas J, Noel-Storr AH, Marshall I, Wallace B, McDonald S, Mavergames C, Glasziou P, Shemilt I, Synnot A, Turner T, Elliott J and the Living Systematic Review Network, 2017. Living systematic reviews: 2. Combining human and machine effort. Journal of Clinical Epidemiology, 91, 31–37. https://doi.org/10.1016/j.jclinepi.2017.08.011

Van Noorden R, 2014. Global scientific output doubles every nine years. Nature News Blog. http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.htmlne-years.html#

Wallace BC, Noel-Storr A, Marshall IJ, Cohen AM, Smalheiser NR and Thomas J, 2017. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. Journal of the American Medical Informatics Association, 24, 1165–1168. https://doi.org/10.1093/jamia/ocx053

Abbreviations

PICO Population; Intervention; Comparator; Outcome
RCT randomised controlled trial