

Education for Information: Interdisciplinary Journal of Information Studies, 2018, 34(4):293-300. DOI: 10.3233/EFI-180222

## **Crowdsourcing a mixed systematic review on a complex topic and a heterogeneous population: Lessons learned**

Mathieu Bujold<sup>a</sup>, Vera Granikov<sup>b</sup>, Reem El Sherif<sup>a</sup> and Pierre Pluye<sup>a</sup>

a. Department of Family Medicine, McGill University, Montréal, QC, Canada

b. School of Information Studies, McGill University, Montréal, QC, Canada

### **ABSTRACT**

Systematic reviews (SR) typically focus on well-defined topics and homogeneous populations. In contrast, conducting a SR on patients with complex care needs (PCCN) presents two main challenges: a novel ill-defined topic and a heterogeneous population. This commentary summarizes the lessons learned from our experience using crowdsourcing to meet these challenges. PCCN often suffer from combinations of multiple issues (medical, interactional, sociocultural, psychological, related to healthcare services use), which can lead to difficult decision-making involving different stakeholders (themselves, caregivers, practitioners). The objective of our SR was to systematically assess decisional needs of PCCN reported in the literature. With the collaboration of librarians, we identified 8616 potentially relevant studies in five databases. A team of 20 crowdreviewers were trained and participated in the selection of 156 relevant studies, using specialised online software. The main benefit of crowdsourcing was the diversity of crowdreviewers' viewpoints which helped us establish an intersubjective knowledge classification of idiosyncratic concepts related to PCCN and their decisional needs. In line with other crowdsourced re-views, our experience confirms that crowdsourcing can be useful in SR with a large number of studies on ill-defined domains and with heterogeneous populations.

**Keywords:** Crowdsourcing, systematic review, studies selection, complex topic, heterogeneous population

### **1. INTRODUCTION**

Systematic reviews (SR) are commonly used to inform evidence-based decision- and policy-making in multiple areas. However, the selection of relevant studies is burdensome, given the continuously growing number of publications (Brown & Allison, 2014; Krivosheev et al., 2017;

Mortensen et al., 2017). Interestingly in the 19th century and the 20th century, the invention of abstracts and the implementation of bibliographic databases were aimed to deal with this very same issue, respectively. Crowdsourcing may be their 21st century counterpart. Specifically in SR on new research topics with no established classification of knowledge, this is even more challenging (Weiss, 2016). Librarians and information professionals are often involved in SR as expert searchers, knowledge managers and analysts (Cooper & Crum, 2013). Therefore, it is important for information professionals supporting or participating in SR to know about new SR methods such as crowdsourcing.

Crowdsourcing “draws on a large pool of people who individually make small contributions that add up to big efforts” (Strang & Simmons, 2018, p. 1). This innovative method of data acquisition is extensively utilized in many research fields (Samimi & Ravana, 2014). Specifically, crowdsourcing has been used as an “economical and time saving means to evaluate large bodies of published literature” (Brown & Allison, 2014, p. 6). In SR, crowdsourcing can overcome the challenges faced by small teams and potentially improve the overall efficiency (Ranard et al., 2014; Strang & Simmons, 2018). This commentary is aimed to summarize lessons learned from our experience using crowdsourcing to select relevant studies in the context of a SR on patients with complex care needs (PCCN) in primary health care (Bujold et al., 2017). Similar to other new research topics, PCCN-related studies are “broad, fragmented and growing quickly” which makes literature reviews more difficult (Weiss, 2016, p. 5). The following sections are presented using the Conceptual Foundations of Crowdsourcing proposed by Pedersen et al. (2013): problem, process, governance, people, technology, and outcome.

## **2. PROBLEM**

SR typically synthesize results of included studies on well-defined issues, homogeneous populations and simple programs. In contrast, conducting a SR on PCCN presents two main challenges: a new ill-defined topic and a heterogeneous population. PCCN often suffer from combinations of multiple chronic conditions, mental health problems, drug interactions and social vulnerability, which can lead to the overuse, underuse or misuse of healthcare services (Martello et al., 2014; Pluye et al., 2014). Moreover, the PCCN experience is exacerbated by interactional issues and difficulties in decisions-making process involving different stakeholders such as patients, caregivers and practitioners (Bujold et al., 2017). Our objective was to systematically assess decisional needs of PCCN from the literature, i.e., identify difficult decisions experienced by stakeholders.

Additional challenges encountered in this SR were related to identification and selection of potentially relevant studies. PCCN and PCCN-related difficult decision-making are idiosyncratic

concepts and can be conceived differently by doctors, patients, caregivers, and researchers due to different knowledge and lived experience. The absence of established classifications of knowledge makes it difficult to identify studies presenting relevant information and interpret it (Weiss, 2016). We worked with specialized librarians to establish a comprehensive search strategy with high recall to capture research studies looking at difficult decision-making in a population with at least one PCCN characteristic (interactional issues, comorbidities, mental health issues, drug interactions, social vulnerability, or healthcare services overuse/underuse/misuse). Establishing clear and comprehensive eligibility criteria was an iterative and time-consuming process.

### **3. PROCESS**

Two specialized librarians conducted a comprehensive search in MEDLINE (Ovid), Embase (Ovid), PsycINFO (Ovid), CINAHL (EBSCOhost) and Social Sciences Citation Index. All results were imported into EndNote, where duplicates were removed, resulting in 8616 potentially relevant unique studies. These studies were screened by at least two independent reviewers. To do this, all studies were imported into a specialised online software (DistillerSR) and two reviewers started the work (first and third author). However, it was too time consuming and we quickly realized that an alternative solution was needed. Thus, we turned to crowdsourcing which was applied at two levels. Our project is an example of “direct crowdsourcing” where one “requester” recruited participants, crowd members, to complete specific screening tasks (Weiss, 2016).

#### **3.1. Crowdsourcing level 1: Abstract screening**

We recruited and trained 15 crowdreviewers to help with Level 1 screening. They used a codebook with clear examples of included/excluded studies, and four eligibility criteria:

- (1) Is this an empirical study?
- (2) Does this abstract involve (deal with) a primary health care setting?
- (3) Is this abstract about a study involving PCCN?
- (4) Is this abstract useful for decisional needs assessment?

Each abstract was coded by two independent crowdreviewers randomly chosen by the software. The crowdreviewers had three answer options: “Yes”, “No”, “I cannot tell”. Studies that received two ‘No’ answers for at least one criterion were automatically excluded. Other studies were moved to the next level, with the exception of those that received a contradictory “Yes” vs. “No” answer for at least one criterion (automatically classified as a ‘conflict’ by the software). Disagreements between crowdreviewers were settled by the first author.

### **3.2. Crowdsourcing level 2: Full text screening**

At this level, 1293 full text were screened pairwise by 11 crowdreviewers (6 recruited at level 1, one new and the four co-authors). Level 2 answer options to eligibility criteria 3 and 4 were derived from a list of pre-established options. This helped crowdreviewers to quickly justify why the study was related to PCCN and decisional needs (see Table 1).

## **4. GOVERNANCE**

Motivation and engagement are key success factors in a crowdsourced project (Strang & Simmons, 2018). In our SR, weekly email updates were sent by one of the co-authors to encourage crowdreviewers. Crowdreviewers co-researchers were motivated (self-interest), and external crowdreviewers (not co-researchers on the project) were compensated for their time using gift cards (20\$ per 100 Level 1 abstracts, and 1\$ per one Level 2 full-text). Six crowdreviewers had previous experience using DistillerSR; those who did not, were trained.

The quality control of crowdsourced work is a common concern (Samimi & Ravana, 2014; Strang & Simmons, 2018; Weiss, 2016) and is of particular importance in SR. For each level, the software allowed the first author to verify the selection bias of crowdreviewers, specifically for two concepts due to the lack of established knowledge classification (PCCN and Decisional needs). For each crowdreviewer, sets of excluded/included studies were randomly selected and checked by the first author. Crowdreviewers who excluded potentially relevant studies or included irrelevant ones were met individually for conceptual clarification. While this process prevented some selection biases, it also contributed to the establishment of an inter-subjective knowledge classification.

## **5. PEOPLE**

A strength of the crowdsourcing process is that crowdworkers “give varying meaning to a single fact or question” (Lebraty & Lobre, 2013, p. 20). As Surowiecki stated in his book on the “Wisdom of the crowd”, collective intelligence may result in better decision-making compared to individual decisions (Surowiecki, 2007). Our team was composed of 20 crowdreviewers who had different levels of (a) experience and familiarity with reviewing scientific articles, (b) using a specialised SR software, and (c) knowledge of complex care needs. They were researchers, graduate students and clinicians from different disciplines (family medicine, nursing, occupational therapy, epidemiology, and social sciences). This diversity was advantageous, leveraging different viewpoints on the complexity of care needs: medical, nursing, occupational, populational, social, and combined viewpoints (some crowdreviewers had multidisciplinary training and working experience). Moreover, some crowdreviewers personally knew

a PCCN and were aware of the patient's viewpoint.

## 6. TECHNOLOGY

Online technology is necessary to enable and facilitate crowdsourcing (Samimi & Ravana, 2014; Strang & Simmons, 2018; Thomas et al., 2017; Wazny, 2017). In our experience, the DistillerSR software was very helpful, user-friendly and available online 24/7. The technical help was accessible and effective in a timely manner. The software was essential to manage the crowdsourcing in three main aspects: inclusions, exclusions and conflicts. At Level 1, the crowdreviewers appreciated the software's highlighting function of predefined keywords, a function also described as time-saving in other crowdsourcing projects (Strang & Simmons, 2018).

## 7. OUTCOMES

The main benefit of crowdsourcing was the diversity of crowdreviewers' viewpoints. Crowdsourcing resulted in the selection of 156 potentially relevant studies to be included in the synthesis. The crowdsourced selection took place over a period of six weeks (June 19 to July 31, 2017). In total, crowdworkers spent 360 hours in Level 1 and Level 2 selection. Level 1 selection was done during the last two weeks of June, and required about 145 hours of crowdwork ( $8616 \text{ abstracts} \times 2 \text{ crowdreviewers} \times 30 \text{ seconds}$ ; the average time to include/exclude an abstract being about 30 seconds). Level 1 crowdworking time was about four hours/week/crowdreviewer on average. Level 2 took place over the month of July, and required about 215 hours of crowdwork ( $1293 \text{ full-texts} \times 2 \text{ crowdreviewers} \times 5 \text{ min}$ ; the average time to include/exclude a full-text being about 5 min). Level 2 crowdworking time was about 5 hours/week/crowdreviewer on average. Indeed, two experienced researchers working almost full-time could review about 8700 abstracts (e.g., 100/hour) and 1300 full-text papers over six weeks ( $6 \text{ weeks} \times 30 \text{ hours/week} \times 2 \text{ crowdreviewers} = 360 \text{ hours}$ ); this would have been feasible but burdensome.

## 8. CONCLUSION

In line with other crowdsourced literature reviews, our experience confirms that crowdsourcing can be useful in SR with a large number of studies such as mixed studies reviews (Hong et al., 2017), and reviews of studies on ill-defined domains and involving heterogeneous populations. Similar to what is reported in the literature (Lebraty & Lobre, 2013), we experienced that crowdsourcing adds value to the creation and production of knowledge via the diversity and independence of the crowd's members. Even though some items proposed by Cox et al. (2015) to assess the outcomes of crowdsourcing are not appropriate for assessing our work (e.g., the dimension "data value"), we addressed most of them in this

commentary (see Table 2). Furthermore, this framework can be used by information professionals to assess and improve the crowdsourcing activities they are involved in.

In our SR, crowdsourcing faced challenges and led to benefits. For example, while building training material and training crowdreviewers took time, the diversity of their viewpoints, comments and questions helped to build a clear manual for selecting abstracts and full-texts on the PCCN complex issue. By way of another illustration, crowdreviewers had different levels of expertise, which was challenging, while the diversity of crowdreviewers contributed to establish an intersubjective knowledge classification of idiosyncratic concepts related to PCCN and their decisional needs. Given the heterogeneity of the population of PCCN and the newness of related concepts, the interweaving of the crowdreviewers' viewpoints allowed us to develop comprehensive and clear definitions.

## ACKNOWLEDGEMENTS

The four authors contributed to the design of the SR, the data collection/analysis (including as crowdreviewers), and approved the final version of the manuscript. We thank warmly all other crowdreviewers (alphabetical order): Meshack Achore, Thomas Bertrand, Paula Bush, Vinita D'Souza, Julie Fréchette, Justin Gagnon, Quan Nha Hong, Emilie Hudson, Jenny Jing, Navdeep Kaur, Irina Kudrina, Stephanie Long, Manon Pluye, Marie-eve Poitras, Christian Ruchon, and Eva Suarathana.

## REFERENCES

- Brown, A. W., & Allison, D. B. (2014). Using crowdsourcing to evaluate published scientific literature: Methods and example. *PLoS ONE*, 9(7), e100647. doi: 10.1371/journal.pone.0100647.
- Bujold, M., Pluye, P., Légaré, F., Haggerty, J., Gore, G. C., Sherif, R. E., Wensing, M., et al. (2017). Decisional needs assessment of patients with complex care needs in primary care: A participatory systematic mixed studies review protocol. *BMJ Open*, 7(11), e016400. doi: 10.1136/bmjopen-2017-016400.
- Cooper, I. D., & Crum, J. A. (2013). New activities and changing roles of health sciences librarians: A systematic review, 1990–2012. *Journal of Medical Library Association*, 101(4), 268–277. doi: 10.3163/1536-5050.101.4.008.
- Cox, J., Oh, E. Y., Simmons, B., Lintott, C., Masters, K., Greenhill, A., Holmes, K., et al. (2015). Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects. *Computing in Science & Engineering*, 17(4), 28–41. doi: 10.1109/MCSE.2015.65.

- Hong, Q. N., Pluye, P., Bujold, M., & Wassef, M. (2017). Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. *Systematic Reviews*, 6(1), 1–14. doi: 10.1186/s13643-017-0454-2.
- Krivosheev, E., Casati, F., Caforio, V., & Benatallah, B. (2017). *Crowdsourcing paper screening in systematic literature reviews*. Paper presented at the Fifth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2017).
- Lebraty, J.-F., & Lobre, K. (2013). *Crowdsourcing: One step beyond*. London: John Wiley & Sons.
- Martello, C., Bessière, G., Bigras, M., Boulet, A., Brenner, K., Crist, E., Pluye, P., et al. (2014). *What do we mean when we say “this patient is complex”?* Poster presented at the NAPCRG Annual Conference (North American Primary Care Research Group), New York.
- Mortensen, M. L., Adam, G. P., Trikalinos, T. A., Kraska, T., & Wallace, B. C. (2017). An exploration of crowdsourcing citation screening for systematic reviews. *Research Synthesis Methods*, 8(3), 366–386. doi: 10.1002/jrsm.1252.
- Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Vreede, G.-J., et al. (2013). Conceptual foundations of crowdsourcing: A review of IS research. *46th Hawaii International Conference on System Sciences*, pp. 579–588.
- Pluye, P., Bessière, G., Bigras, M., Boulet, A., Brenner, K., Crist, E., Grad, R., et al. (2014). *Characteristics of complex care needs and interventions suited for patients with such needs: A participatory scoping review*. Poster presented at the NAPCRG Annual Conference (North American Primary Care Research Group), New York.
- Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., Merchant, R. M., et al. (2014). Crowdsourcing – harnessing the masses to advance health and medicine, a systematic review. *Journal of General Internal Medicine*, 29(1), 187–203. doi: 10.1007/s11606-013-2536-8.
- Samimi, P., & Ravana, S. D. (2014). Creation of reliable relevance judgments in information retrieval systems evaluation experimentation through crowdsourcing: A review. *The Scientific World Journal*, 2014, 1–13. doi: 10.1155/2014/135641.
- Strang, L., & Simmons, R. K. (2018). *Citizen science: Crowdsourcing for systematic reviews*. A Report of The Healthcare Improvement Studies Institute (Ed.) Retrieved from <https://www.thisinstitute.cam.ac.uk/wp-content/uploads/2018/06/THIS-Institute-Citizen-Science-Crowdsourcing-for-systematic-reviews-978-1-9996539-1-0.pdf> (Date accessed 2018-11-25).

Surowiecki, J. (2007). *The wisdom of crowds*. New York: Random House.

Thomas, J., Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C., et al.,  
Living Systematic Review Network. (2017). Living systematic reviews: 2. Combining human  
and machine effort. *Journal of Clinical Epidemiology*, 91, 31–37.  
doi:10.1016/j.jclinepi.2017.08.011.

Wazny, K. (2017). “Crowdsourcing” ten years in: A review. *Journal of Global Health*, 7(2), 1–15. doi:  
10.7189/jogh.07.020602.

Weiss, M. (2016). Crowdsourcing literature reviews in new domains. *Technology Innovation  
Management Review*, 6(2), 5–14.



## TABLES AND FIGURES

**Table 1** Level 2 answer options to eligibility criteria 3 and 4

Eligibility criteria	Answer options – check all that apply (at least one)
3. Is this study is related to PATIENTS WITH COMPLEX CARE NEEDS?	<ul style="list-style-type: none"> <li>– Direct mention of a study on “Patients with complex care needs”.</li> <li>– Combinations of (complex* OR complicat* OR high-effort* OR burden OR demanding OR difficult*) AND (patient* OR decision OR need* OR care OR case* OR situation* OR condition* OR population OR problem* OR issue* OR existence* OR experience* OR live* OR realit* OR health seeking* OR intervention*).</li> <li>– Combinations of at least two of the following characteristics: multiple chronic conditions, mental health problems, drug interactions, vulnerable population or health care services overuse, underuse or misuse.</li> <li>– The authors of this study or the participants, qualifies as complex a situation of care needs presented.</li> <li>– Include (for other raison).</li> <li>– None of the above (EXCLUDE).</li> <li>– I cannot tell.</li> </ul>
4. Is this study is useful for DECISIONAL NEEDS ASSESSMENT?	<ul style="list-style-type: none"> <li>– Primary care decisions are taken (or needed to be taken).</li> <li>– Interactional issues related to uncertainty or disagreement regarding possible options or multiple decisions among stakeholders (decisional conflict).</li> <li>– There are factors (barriers/facilitators) affecting the decision-making process. – There are needs to better support decision-making.</li> <li>– There is decision-making support such as decision support tool or intervention. – Include (for other raison).</li> <li>– None of the above (EXCLUDE).</li> <li>– I cannot tell.</li> </ul>

**Table 2** Conceptual framework for assessing “outcomes of crowdsourcing in sciences”

Concepts	Dimensions	Items
1. Contribution to science	1.1. Data value	<ul style="list-style-type: none"> <li>– Number of publications produced.</li> <li>– Academic impact (citations).</li> <li>– The analysis is scientifically and statistically valid.</li> </ul>
	1.2. Project design and resource allocation	<ul style="list-style-type: none"> <li>– Resource savings realised (considering the time that would be required for a professional researcher to carry out the project tasks).</li> <li>– Equity of the distribution of effort across participants.</li> <li>– The proportion of volunteers who were successfully trained (as indicated by the volunteers going on to complete at least one task after doing the tutorial).</li> </ul>
2. Public engagement	2.1. Dissemination and feedback	<ul style="list-style-type: none"> <li>– Collaboration – the number of papers that include at least one citizen scientist as an author.</li> <li>– Communication – the amount of communication activity that took place (e.g. blog posts, tweets).</li> <li>– Interaction – the number of interaction events (e.g. blog</li> </ul>

post comments and replies) that took place between the researchers and citizen scientists.

## 2.2. Participation and opportunities for learning

– Project appeal – the number of contributors.

– Sustained engagement – the median time period over which each participant engaged with the project.

– Public contribution – the median number of classifications completed by each.

---

Source: Cox, J., E.Y. Oh, B. Simmons, C. Lintott, K. Masters, A. Greenhill, G. Graham, & K. Holmes. 2015. 'Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects.' *Computing in Science Engineering* 17: 28–41.