Protocol
Version March,30th 2016

Mapping of crowdsourcing in health: a systematic review

Ghislène Mansouri, Perrine Créquit, Ludovic Trinquart, Philippe Ravaud
INSERM U1153 Methods team

I. Background

Crowdsourcing is a recent approach that involves a group of general population; called the crowd workers to complete tasks of requesters(1). A "crowd" is a large group of independent people providing a large wide of activities with no formal training in a field specific to the topic of investigation, especially via the internet(2), by using specific platforms. Crowd workers can have an access on the crowdsourcing sites from anywhere at times convenient for them. They could be laypersons not performing any science-related activities themselves, and so not be considered as citizen scientists. Indeed, citizen scientists denote the conduct of science-related activities (3). Crowd workers carry out tasks posted by requesters, who accept or reject their work and pay them as a consequence(4).

By the way, the evolution of technology in the world, with 2.3 billion Internet users and 6 billion mobile phone subscriptions, allows crowdsourcing to grow rapidly(5). Crowdsourcing is not a new process and takes its origin in 1714 in England where the British Government proposed 20 000£ to anyone who could find a solution for calculating the longitudinal position of a ship. This concept has been utilized primarily by non-medical fields (6). Researchers from any domains use this process to get data and more and more, this process is becoming the center of attention of scientist's community.

A previous systematic review described the scope of crowdsourcing in health and medical research (5). In March 2013, from 440 references, Ranard et al. identified only 21 articles. Another narrative review described the use of crowdsourcing in health research studies through 2011 and developed crowdsourcing framework with participant organized (non-scientists asking non-scientists or scientists) and researcher organized (scientists asking non-scientists i.e. “open call” or asking other scientists)(3).

However, this field is moving fast and the process has known an explosion of its use in recent years. The number of studies using crowdsourcing has been multiplied by 17 since 2011 on Pubmed and by 42 on Embase (Appendix 1). Moreover, the concept of crowdsourced people was applied for the first time in surgery in 2014 to assess surgical skills(7). The following year, this approach was applied in 7 studies(8–14).

Crowdsourcing could be a great method to solve specific scientific mission that cannot be entirely automated and requires human intelligence. Therefore, a current mapping of crowdsourcing use in health is needed to describe the different applications using the framework. This update synthesis could be useful to scientists to transpose this concept in their research.

II. Objective

The aim of this systematic review is:

1. To describe the current different applications of crowdsourcing in health (mapping)
2. To detail characteristics of the tasks and the demographic of workers.

III. Method

Criteria for considering studies for this review

We will include studies concerning:

1) The three categories of health described by Prpic: health promotion, health research and health maintenance(15).

   Definitions:

   - Health promotion: activities such as disease detection and surveillance, behavioral interventions, health literacy and health education
   - Health research: activities such as pharmaceutical research, clinical trials and health experiment methodology and improving health care research knowledge
   - Health maintenance: activities such as patient-related or physicians-related, diagnostics, medical practice and treatment support

2) Study conducted on a crowdsourced population: workers are recruited with crowdsourcing (i.e., recruited online with a website or an open call to a large audience using internet-related technologies). Workers may or may not be acting as citizen scientists (i.e., conducting science-related activities).
3) Without any restriction of the type of study design.

Search method for identification of studies

We will systematically search the following electronic databases: Medline and Embase from inception to March 2016.

There will be no restriction on the language of publication when searching the electronic databases. All databases will be searched using both controlled vocabulary (namely Mesh in Medline and Emtree in Embase) and a wide range of free-text terms. Indeed, crowdsourced

health studies may be a blend of crowdsourcing and citizen science, these terms can be used interchangeably and so included in our search equation. We will use different terms referring to crowdsourcing, citizen science, some platforms used such as Mechanical Turk. The search strategy used to search Medline and Embase is listed in Appendix 2.

We will also screen to the reference lists of previous systematic reviews and of selected articles to identify additional studies.

Screening Google scholar seems not feasible because of the amount of records found (Appendix 1). It would be a long, challenging and arduous process for only two reviewers and could require the use of crowdsourcing.

Selection of studies

Two reviewers (GM, PC) will independently and in duplicate examine each title and abstract identified in the search to exclude obviously irrelevant reports. The two reviewers will then independently examine full-text articles to determine eligibility. The whole study selection process will be performed using the platform ''Resyweb''. Disagreements will be discussed with a third author to reach consensus (LT). We will list studies and document the primary reasons for exclusion.

Data extraction and management

The data will be extracted from reports by one author (GM) using a standardized from and checked for quality assurance by another author (PC). Disagreements will be discussed with a third author to reach consensus (LT).

We will extract the following characteristics from included studies:

1. General characteristics: identification number, journal and year of publication, study design (observational study, survey), funding source (public, private, both, unknown)
2. Type of crowdsourcing: crowdsourcing framework (researcher-organized, participant-organized(3)) and manners of crowdsourcing application (1. task divided in several parts and shared to each group of workers (16), 2. same task given to several groups of workers (17)).
3. Demographic and other characteristics of the crowd: size of the crowd, age, gender, status (researcher, physician, student, engineer, and general population), geographic location, motivations, skill set required, qualification test to recruit workers, training of workers.
4. Logistics of the crowdsourcing: category of health (health promotion, research and maintenance (15)), health field (public health, molecular biology, surgery…), length of time crowdsourcing was conducted, use of a web platform or a mobile platform, description of the task (using four types of crowdsourcing tasks: problem solving, data processing, surveillance/monitoring and surveying described by Ranard et al. (5)), use of individuals compared to teams or experts, time to perform the task, monetary incentives offered, data validation techniques.

Analysis

Summary statistics will be used. Descriptive statistics will be applied for categorical variables described with frequencies and percentages and quantitative variables with mean (SD) to characterize the data extracted from the selected studies.

References

1. Deal SB, Lendvay TS, Haque MI, Brand T, Comstock B, Warren J, et al. Crowd-sourced assessment of technical skills: an opportunity for improvement in the assessment of laparoscopic surgical skills. Am J Surg. 2016 Feb;211(2):398–404.

2. Khare R, Good BM, Leaman R, Su AI, Lu Z. Crowdsourcing in biomedicine: challenges and opportunities. Brief Bioinform. 2016 Jan;17(1):23–32.

3. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res. 2012;14(2):e46.

4. Brown AW, Allison DB. Using crowdsourcing to evaluate published scientific literature: methods and example. PloS One. 2014;9(7):e100647.

5. Ranard BL, Ha YP, Meisel ZF, Asch DA, Hill SS, Becker LB, et al. Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review. J Gen Intern Med. 2014 Jan;29(1):187–203.

6. Dawson R, Bynghall S. Getting results from crowds: [the definitive guide to using crowdsourcing to grow your business]. 2. ed. Sydney: Advanced Human Technologies; 2012. 229 p.

7. Chen C, White L, Kowalewski T, Aggarwal R, Lintott C, Comstock B, et al. Crowd-Sourced Assessment of Technical Skills: a novel method to evaluate surgical performance. J Surg Res. 2014 Mar;187(1):65–71.

8. Lendvay TS, White L, Kowalewski T. Crowdsourcing to Assess Surgical Skill. JAMA Surg. 2015 Nov 1;150(11):1086–7.

9. Powers MK, Boonjindasup A, Pinsky M, Dorsey P, Maddox M, Su L-M, et al. Crowdsourcing Assessment of Surgeon Dissection of Renal Artery and Vein During Robotic Partial Nephrectomy: A Novel Approach for Quantitative Assessment of Surgical Performance. J Endourol Endourol Soc. 2015 Dec 30;

10. Malpani A, Vedula SS, Chen CCG, Hager GD. A study of crowdsourced segment-level surgical skill assessment using pairwise rankings. Int J Comput Assist Radiol Surg. 2015 Sep;10(9):1435–47.

11. Maier-Hein L, Kondermann D, Roß T, Mersmann S, Heim E, Bodenstedt S, et al. Crowdtruth validation: a new paradigm for validating algorithms that rely on image correspondences. Int J Comput Assist Radiol Surg. 2015 Aug;10(8):1201–12.

12. Maier-Hein L, Mersmann S, Kondermann D, Stock C, Kenngott HG, Sanchez A, et al. Crowdsourcing for reference correspondence generation in endoscopic images. Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv. 2014;17(Pt 2):349–56.

13. Holst D, Kowalewski TM, White LW, Brand TC, Harper JD, Sorensen MD, et al. Crowd-Sourced Assessment of Technical Skills: Differentiating Animate Surgical Skill Through the Wisdom of Crowds. J Endourol Endourol Soc. 2015 Oct;29(10):1183–8.

14. Holst D, Kowalewski TM, White LW, Brand TC, Harper JD, Sorenson MD, et al. Crowd-sourced assessment of technical skills: an adjunct to urology resident surgical simulation training. J Endourol Endourol Soc. 2015 May;29(5):604–9.

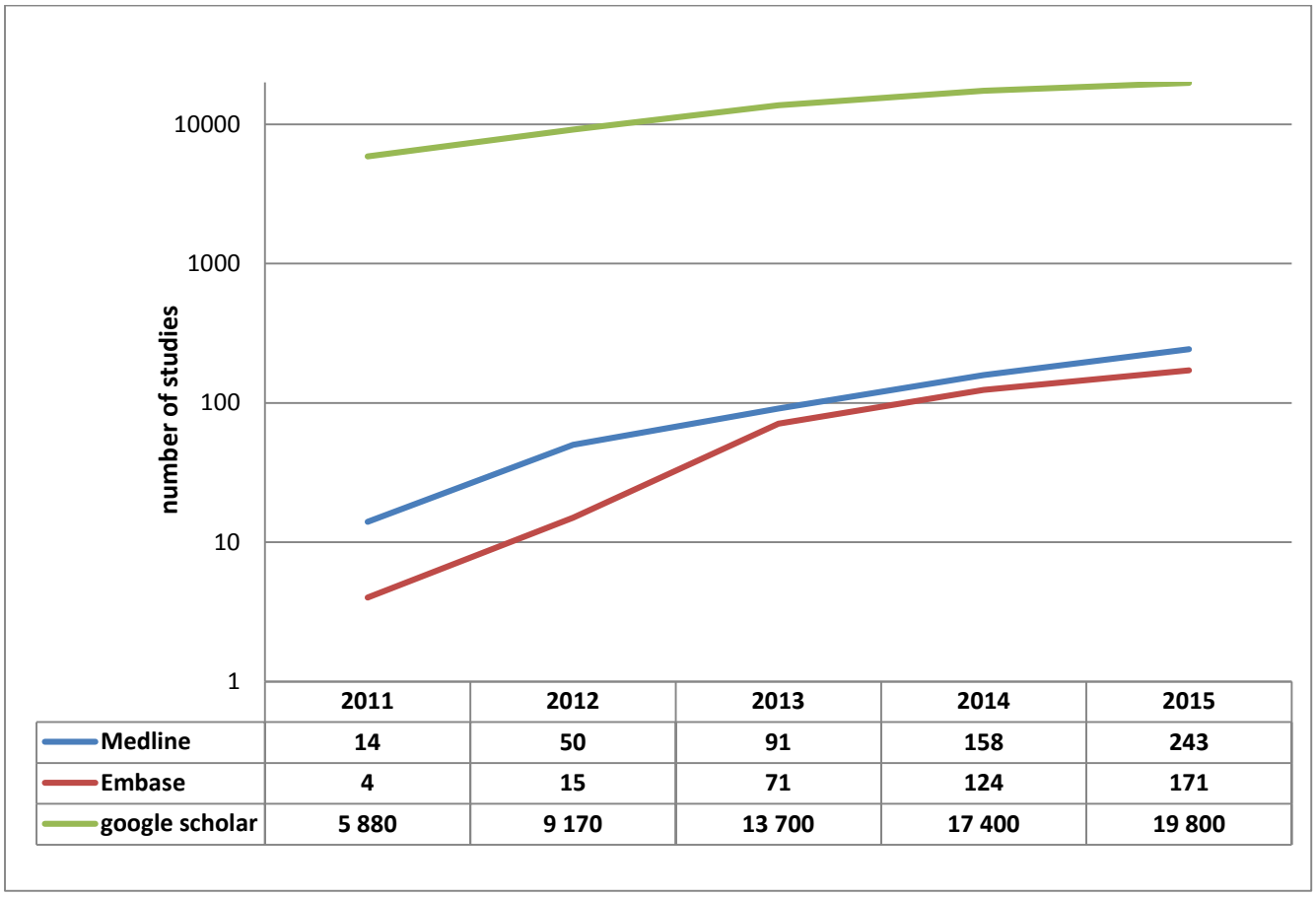
15. Health Care Crowds: Collective Intelligence in Public Health (2015). Prpić, J., (2015). Health Care Crowds: Collective I.

16. Michelucci P, Dickinson JL. The power of crowds. Science. 2016 Jan 1;351(6268):32–3.

17. Silberzahn R, Uhlmann EL. Crowdsourced research: Many hands make tight work. Nature. 2015 Oct 7;526(7572):189–91.

Appendix

Appendix 1: Evolution of the number of studies using crowdsourcing



	2011	2012	2013	2014	2015
Medline	14	50	91	158	243
Embase	4	15	71	124	171
google scholar	5 880	9 170	13 700	17 400	19 800

Search on Medline, Embase and Google scholar with the following search equation: "crowdsourcing" OR "Amazon Mechanical Turk" OR MTurk OR "Mechanical Turk".

Appendix 2: Search terms for Medline, accessed via Pubmed, and Embase

PUBMED

March 21st 2016

#1	**"crowdsourcing"[MeSH Terms] OR crowdsource[tiab] OR crowdsourced[tiab] OR crowdsourcers[tiab] OR crowdsources[tiab] OR crowdsourcing[tiab] OR crowd-source[tiab] OR crowd-sourced[tiab] OR crowd-sourcing[tiab] OR crowdworker[tiab] OR crowdworkers[tiab] OR "crowd science"[tiab] OR "crowd-based"[tiab] OR crowds[tiab]**	**N= 946**
#2	**"Mechanical Turk"[tiab] OR "Mturk"[tiab] OR "crowdflower"[tiab] OR "foldit"[tiab]**	**N= 228**
#3	**"citizen science"[tiab] OR "citizen scientist"[tiab] OR "citizen scientists"[tiab]**	**N= 269**
#4	**Microtask[tiab] OR "online task"[tiab]**	**N= 19**
#5	**#1 OR #2 OR #3 OR #4**	**N=1 359**

EMBASE

March 21st 2016

#1	**crowdsource OR crowdsourced OR crowdsourcers OR crowdsources OR 'crowdsourcing'/exp OR 'crowd source' OR 'crowd sourced' OR 'crowd sourcing'/exp OR crowdworker OR crowdworkers OR 'crowd science' OR 'crowd-based' OR crowds AND [embase]/lim**	**N = 703**
#2	**'mechanical turk' OR 'mturk' OR 'crowdflower' OR 'foldit' AND [embase]/lim**	**N = 164**
#3	**'citizen science' OR 'citizen scientist' OR 'citizen scientists' AND [embase]/lim**	**N = 174**
#4	**microtask OR 'online task' AND [embase]/lim**	**N= 12**
#5	**#1 OR #2 OR #3 OR #4**	**N= 986**

Data extraction form: Crowdsourcing

General characteristics

Identification number: ……

Journal: ……

Publication year: ……

Study design: observational study ☐ survey ☐ other ☐ ……

Funding source: public ☐ private ☐ both ☐ unknown ☐

Type of crowdsourcing

Crowdsourcing framework:

Researcher-organized ☐ Participants-organized ☐

Manners of crowdsourcing application:

Task divided in several parts and shared to each group of workers ☐

Same task given to several groups of workers ☐

Demographic and other characteristics of the crowd

Size of the crowd: ……

Age: Mean or median: ……

Gender: Female: ……. %

Status:
Researchers ☐…….% Physicians ☐…….% Engineers ☐……. % Students ☐ ……. %

Patients ☐……. % General population ☐……. % Other☐……

Socioeconomic status: ……

Geographic location: National ☐ International ☐

Motivations: Social ☐ Money Compensation ☐ Fun ☐

Contribution to an important cause ☐ Other ☐……

Skill set required: Yes ☐ No ☐
If yes, detail: ……

Qualification test to recruit workers: Yes ☐ No ☐
If yes, detail: ……

Training of workers: Yes ☐ No ☐
If yes, detail: ……

Logistics of the crowdsourcing

Category of health: Health promotion ☐ Health research ☐ Health Maintenance ☐

Type of activities: ……

Research field: Surgery ☐ Dermatology ☐ Psychology ☐ Neurology ☐ Pathology/hematology ☐ Genomic ☐ Radiology ☐ Public Health ☐ Molecular biology ☐ Oncology ☐ Nutrition ☐ Other ☐……..

Length of time crowdsourcing was conducted: ……

Use of a web platform or a mobile platform:

Amazon Technical Turk ☐ Crowdflower ☐ Quora ☐ Yahoo Answer☐ Genomera ☐ Web based Game ☐ Web-questionnaire ☐ CureTogether ☐ Other☐……..

Description of the task:

Problem solving ☐ Data processing ☐ Surveillance/monitoring ☐ Surveying ☐ Other ☐……

Detail:
……
…….

Use of individuals compared to teams or experts: Yes ☐ No ☐

Time to perform the task: ……

Monetary incentives offered: Yes ☐ No ☐
If yes, detail: ……

Data validation techniques: Yes ☐ No ☐
If yes, detail: ……