

RESEARCH ARTICLE

Using crowdsourcing to estimate populations in communities: Providing a key measurement for coverage assessment of community programmes

Adnan Ahmad Khan, Amjad Iqbal

Abstract

Introduction: Crowdsourcing pools together dispersed information that is considered public knowledge in an area, to form realistic estimates about the area, or to identify new ideas. The technique can be extremely helpful to develop estimates of public health indicators such as catchment area populations or healthcare providers; however, such uses must be scientifically validated.

Methods: We divided the community into 1040 discrete segments of similar lengths of streets (called spots) and then randomly selected 605 of these spots for crowdsourcing. Local respondents were asked to estimate the maximum and the minimum population residing in those spots. Five informants were interviewed per spot. Median values for the maximum and minimum were averaged to arrive at an estimate for the spot's population. Estimates for all spots were added together to arrive at the population of the community. One hundred spots from the 597 crowdsourced spots were revisited to conduct a household census as a "gold standard".

Results: Spots where both crowdsourcing and census estimates were computed had a crowdsource population estimate of 19,255 versus a census estimate of 18,119 - a variation of 5.9% ($p < 0.001$). However, within spot variation was a mean of 25%.

Conclusions: Crowdsourcing communities for public knowledge information can yield more accurate information about public health indicators such as populations. In turn these estimates can help to better understand public health programme coverage. Other applications to consider may be missed children for immunization or schooling, deaths or births in communities or to identify total formal or informal healthcare providers in a community.

Keywords: Crowdsourcing, Population estimates. (JPMA 71: S-67 [Suppl. 7]; 2021)

Introduction

Compared to the traditional model of funding and implementing development work without expectations of improved outcomes, the Millennium Development Goals (MDGs) succeeded in part by setting and achieving specific measurable targets. In turn, this focus on measurement helped improve the science of evaluations. Today, a number of institutions specialize in evaluating programmes or ideas, and these institutions generate considerable knowledge to guide measurements that aid evaluations.

One problem of measurement of the impact or outcomes of a programme is the ability to assess programme effects specifically for the populations where the programme was implemented. For example, if a family planning programme is implemented in a few communities, its effects must be measured in only those communities. However, government census or other surveys that are sometimes used to measure programme effects, often focus on larger domains such as provinces or districts.

.....
Research and Development Solutions, Islamabad, Pakistan.

Correspondence: Email: adnan@resdev.org

Using such surveys would underestimate the programme's effects. In these cases, "coverage" of programme effects would require an estimate of the specific population of localities where a programme was targeted (i.e., the catchment of a health facility) as the "denominator" against which programme achievements may be measured and reported.

Social phenomena are often emergent and knowledge about them is dispersed within the society so that no one person or group may have "perfect" knowledge. F. A. Hayek contended that such perfect knowledge in a society is impossible, impracticable and possibly even irrelevant to dealing with social events in a society; and that most planners must contend with "just accurate enough" information to make decisions.¹ One application of this insight may be that an aggregation of observations from many non-expert observers may be more accurate than estimates of the experts. This "crowdsourcing" has been used to estimate the weight of livestock at county fairs² as well as the locations of shipwrecks,³ to analyze healthcare data,^{4,5} identify gene structure,⁶ report disease outbreaks/events,⁷⁻⁹ predict public behaviours,^{8,9} and in surveys.¹⁰ Thus, crowdsourcing may potentially be used to solicit public health information from communities

directly, and this may be a relatively quick and inexpensive alternative to some surveys or surveillance.⁸⁻¹⁰

The key principles of the process are that a large number of unbiased "non-expert" observers are questioned about something they are expected to have at least some knowledge of. Care must be exercised to ensure that they do not confer with each other, in order to minimize "groupthink" or premature convergence onto commonly shared knowledge or a few dominant ideas.¹¹ On occasions a personal commitment by informants, such as placement of actual monetary bets, adds to the accuracy of predictions.¹²⁻¹⁹

We describe a study to estimate the population of specific neighbourhoods and compare these to a limited census as a "gold standard". The study was conducted in the informal urban settlement of Dhok Hassu in Rawalpindi, Pakistan, where our team had established an "urban laboratory" to study ideas in development and were about to initiate a family planning promotion intervention. Population estimates were to be used for planning of the intervention and to measure its outcomes.

Methodology

Using Google Earth, we identified 1040 "spots". Each spot corresponds to an identifiable length street on Google Earth of approximately 50-70 meters. In reality this reflects the smallest street that is visible on Google Earth and can accommodate the passage of an automobile. There are smaller lanes that only allow passage of pedestrians/cycles and are not visible on Google Earth.

Out of the total of 1045 spots, 605 were randomly identified and visited for "crowdsourced mapping". In each spot, the team asked pedestrians or shopkeepers for their minimum and maximum estimates of the 1) Number of houses, 2) Number of households and 3) Total population of that spot. Between 5 and 7 informants were queried in each spot. Each informant had to be a local resident and informants were not allowed to confer with each other nor to have had a prior contact with our interviewers, and there was no prompting by our interviewers.

For each spot, we calculated the medians of the minimum and maximum estimates given by the informants and averaged these medians to arrive at the crowdsourced population of the spot. Estimates for all spots were then summed to get the estimated total population through crowdsourcing. These crowdsourced population estimates were then compared against the census data as below.

Within the 605 spots identified above, 100 spots were randomly assigned to receive a census. A separate team visited these spots and listed members from each of these households along with their ages. Difference between the crowd sourced population estimate and the census population was compared using a T-test statistic. The overall process took around 3 days to map the spots on Google Earth and around 5-7 days for the field work.

Results

Among the 100 spots randomly identified for verification, the estimated population from crowdsourcing came up to 19,255 compared to 18,119 that was computed from the household census, giving an error rate of 5.9% (SD: 25%, p value: <0.0001) due to large between spots variation (range: -862% to 88%).

Crowdsourced estimates from all 605 spots were summed and came to 137,857. Since 605 spots for crowdsourcing had been randomly selected from a total of 1045 spots, the sum of all crowdsourced estimates was multiplied with 1.727 (1045/605) to arrive at a total population for Dhok Hassu of 238,116 (range: 223,829 to 252,403). On census, there were 2026 households living in single-family dwellings and 1521 in multifamily dwellings. These extrapolate to a total of 36,888 households.

Discussion

We found that the accuracy of informant provided estimates of population of a community is approximately 6% when compared against a limited census of the same community. Given the costs of the exercise of just under USD 2000 and duration of around one week, our research demonstrates that crowdsourcing is an effective, accurate and rapid means for collecting certain public health relevant information, such as the number of households or population in a community.

Our findings supplement the existing research on pooling knowledge or cooperation across internet or social media to garner publicly known information.⁴⁻⁹ More specifically, we add to the relatively sparse research on crowdsourcing local information directly from communities.

The process of using crowdsourcing has evolved iteratively. Previous experience had shown that spots have to be of similar size as large variations in spot size add to error. Accuracy of estimates rises as more informants are queried per spot, levelling off at around 5-6 informants; little additional advantage is observed with 7 or more informants. It is also crucial to ensure that the informants reside in the area and are not just visitors. Additionally, to avoid "group think", interviewers must be

trained to ensure that the informants don't talk to each other. This can be made easier by keeping spot size small and, if possible, using linear length of a single street as one spot. Finally, we found that although informants can help estimate the number of households and populations, they are often unable to discern the proportion or number of men, women or children within that total. One additional lesson from this study was to limit extreme estimates by informants. e.g., some claimed that there were thousands living in the spot, which in turn increased the variation in that spot. We speculate that this could be resolved by training the interviewers to help informants limit their estimates to the spot rather than the larger neighbourhood.

In previous work, we have also identified the number and location of formal and informal healthcare providers and schools.

Conclusion

Crowdsourcing has been used in a number of scientific applications to collate and use publicly held knowledge. We describe a particular application to public health, the measurement of denominator populations in a public health intervention. This technique can help during planning an intervention by identifying the number of beneficiaries or during measurement of effects by providing a denominator from which effect size can be calculated. It would be interesting to explore if there could be more complex applications of this technique such as estimating the number of deaths – all or of specific groups such as children or women, missed children for vaccination, children attending school, children that are working for a living etc. Such estimations can be initiating points for more in-depth inquiries and may help reduce the requirement for more extensive community surveys. Complemented with geospatial placement techniques, crowdsourcing can also help design and manage programmes better by locating social phenomena on maps and better understand the measurements of programme outcomes against intended target populations.

References

- Hayek FA. The Use of Knowledge in Society. *Am Econ Rev* 1945;35:519-30.
- Galton F. Vox Populi. *Nature* 1907;75:450-1. doi: 10.1038/075450a0
- Surowiecki J. *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York, USA: Doubleday & Co; 2004.
- 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;29:187-203. doi: 10.1007/s11606-013-2536-8.
- Nickoloff S. Capsule commentary on Ranard et al., crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. *J Gen Intern Med* 2014;29:186. doi: 10.1007/s11606-013-2620-0.
- Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. *Nature* 2010;466:756-60. doi: 10.1038/nature09304.
- Brownstein JS, Freifeld CC, Chan EH, Keller M, Sonricker AL, Mekaru SR, et al. Information technology and global surveillance of cases of 2009 H1N1 influenza. *N Engl J Med* 2010;362:1731-5. doi: 10.1056/NEJMs1002707.
- Chunara R, Freifeld CC, Brownstein JS. New technologies for reporting real-time emergent infections. *Parasitology* 2012;139:1843-51. doi: 10.1017/S0031182012000923.
- Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am J Trop Med Hyg* 2012;86:39-45. doi: 10.4269/ajtmh.2012.11-0597.
- Behrend TS, Sharek DJ, Meade AW, Wiebe EN. The viability of crowdsourcing for survey research. *Behav Res Methods* 2011;43:800-13. doi: 10.3758/s13428-011-0081-0.
- Janis IL. *Victims of groupthink; a psychological study of foreign-policy decisions and fiascoes*. Boston, Massachusetts: Houghton Mifflin Harcourt; 1972.
- Munafo MR, Pfeiffer T, Altmeld J, Heikensten E, Almenberg J, Bird A, et al. Using prediction markets to forecast research evaluations. *R Soc Open Sci* 2015;2:e150287. doi: 10.1098/rsos.150287.
- Dreber A, Pfeiffer T, Almenberg J, Isaksson S, Wilson B, Chen Y, et al. Using prediction markets to estimate the reproducibility of scientific research. *Proc Natl Acad Sci U S A* 2015;112:15343-7. doi: 10.1073/pnas.1516179112.
- Tung CY, Chou TC, Lin JW. Using prediction markets of market scoring rule to forecast infectious diseases: a case study in Taiwan. *BMC Public Health* 2015;15:766. doi: 10.1186/s12889-015-2121-7.
- Pfeiffer T, Almenberg J. Prediction markets and their potential role in biomedical research—a review. *Biosystems* 2010;102:71-6. doi: 10.1016/j.biosystems.2010.09.005.
- Rothschild D. Forecasting elections: Comparing prediction markets, polls, and their biases. *Public Opin Q* 2009;73:895-916. doi: 10.1093/poq/nfp082
- Almenberg J, Kittlitz K, Pfeiffer T. An experiment on prediction markets in science. *PLoS One* 2009;4:e8500. doi: 10.1371/journal.pone.0008500.
- De Vries A, Feleke S. Prediction of future uniform milk prices in Florida federal milk marketing order 6 from milk futures markets. *J Dairy Sci* 2008;91:4871-80. doi: 10.3168/jds.2008-1138.
- Arrow KJ, Forsythe R, Gorham M, Hahn R, Hanson R, Ledyard JO, et al. Economics. The promise of prediction markets. *Science* 2008;320:877-8. doi: 10.1126/science.1157679.