Generalizing from Survey Data – How Big is Big Enough in a Disaster?

Jared P. Lander, JP Lander Consulting

Richard Garfield, CDC and Professor, Columbia University

Introduction

Needs assessments must be carried out rapidly to be of use in responding to humanitarian emergencies. They also should accurately reflect the needs and opinions of those affected by emergencies. Larger samples can reach a wider selection of those affected, and provide more precise estimates. But larger samples cost more money and more time.

To date most large scale multi-cluster needs assessments have been unable to produce a report in time to contribute to decision-making about the amount and distribution of humanitarian aide. The pursuit of large samples, then, frequently dooms such assessments to fruitlessness.

Speed of data collection requires that the most efficient sample be taken. Efficiency here is defined as the minimum effort needed to generate the most accurate information. This effort is a combination of the distribution of the sample as well as its size.

Criteria to determine the needed size of a random sample are well established. When randomness cannot be assured, and when even the effected population is not well defined, criteria for sample size selection are not as clear. In practice, major disaster needs assessments have taken as large a sample as funding and logistics made possible. This included 3000 household surveys in Myanmar[[1]](#endnote-1) and 2000 in Haiti[[2]](#endnote-2).

Here we compare the results of \_\_ quantitative indicators from the entire sample of that household surveys to smaller samples to determine an optimal sampling strategy for similar events in the future.

Methods

Data used here are taken from the multi-cluster rapid humanitarian needs assessment carried out as large scale monsoon-caused floods inundated one-fifth of the country’s landmass, affecting 20 million people[[3]](#endnote-3). The survey was conducted in 4 of the 5 affected provinces, reaching 2442 households, in 53 tehsils, in 27 of what were believed to be 47 affected districts[[4]](#footnote-1).

Survey sites were chosen from a preliminary list of affected but accessible tehsils (districts) provided by the WFP 3 weeks after the rains began. When the survey team reached the tehsil, they chose 1-3 villages which could be reached within 2 hours to sample. Upon reaching the village, 5-10 households were to be chosen at random to interview. The rapidly changing nature of the disaster meant that many people were “on the move” during the survey. People then were sampled in the selected areas in or near their homes, living with host families, in camps or spontaneous settlements along the road. Because the floods evolved in different time periods in different provinces, the distribution and conditions of affected people varied a great deal. Survey fieldwork was carried out over 5 of the 20 days utilized for the assessment.

We summarize data from 5 critical questions asked of heads of households in all 2442 sample forms. We then looked at a randomly selected subsample of tehsils included in the survey. The level of variation between the entire sample and subsamples of 3-15 tehsils per province were examined. Because only 5 tehsils were reached in the Gilgit Baltistan province, these were excluded from analysis.

Results

How much variation from the total sample in those 3 provinces.

Discussion

Cannot say the total sample was true measure of the population but it was adequate to characterize population conditions and needs.

Could a similar result have been reached with a smaller sample?

Defining logistical question was distance to reach the tehsil. Once reached, how many villages to sample in a day was set at 3.

Conclusions

Rapidly changing situation, total number of affected areas unknown. Need operative information quickly. Survey process in total was 20 days; prior large scale surveys have been 35-40 days from conception to reporting of results. It was this short because experienced teams were already in the field. With a smaller sample, fieldwork could have been cut from 5 to 2 or 3 days. With salary costs about \_\_% of the total, the cost of carrying out the survey would have been about \_\_% less, with no notable loss in information generated.

So go to more places, identify various key subgroups of the population better, and take a smaller overall sample to optimize the efficiency and reduce the time and cost of such a survey.

1. Tripartite Core Group. Post-Nargis Joint Assessment, Annex 1. July 2008, Jakarta: Indonesia. accessed http://www.scribd.com/doc/51527522/PONJA. [↑](#endnote-ref-1)
2. http://www.acaps.org/img/documents/rinah-report-final-rinah-report-feb2010.pdf [↑](#endnote-ref-2)
3. http://www.acaps.org/img/documents/mcram-report-121010-mcram-assessment-report.pdf [↑](#endnote-ref-3)
4. A month later, government registered a total of 78 affected districts [↑](#footnote-ref-1)