A modified cluster-sampling method for post-disaster rapid assessment of needs

J. Malilay,[1] W.D. Flanders,[2] & D. Brogan[3]

The cluster-sampling method can be used to conduct rapid assessment of health and other needs in communities affected by natural disasters. It is modelled on WHO's Expanded Programme on Immunization method of estimating immunization coverage, but has been modified to provide (1) estimates of the population remaining in an area, and (2) estimates of the number of people in the post-disaster area with specific needs. This approach differs from that used previously in other disasters where rapid needs assessments only estimated the proportion of the population with specific needs. We propose a modified n × k survey design to estimate the remaining population, severity of damage, the proportion and number of people with specific needs, the number of damaged or destroyed and remaining housing units, and the changes in these estimates over a period of time as part of the survey.

Introduction

Rapid assessment of needs, also known as rapid health or epidemiological assessment, refers to a collection of epidemiological, statistical, and anthropological techniques designed to provide, quickly and at low cost, accurate population-based information in a simple format to decision-makers (*1, 2*). It has been recognized as a useful and important method for determining the immediate needs of communities after acute events, such as disasters (*3, 4*). Past examples of applications include assessing nutritional status, mortality, morbidity, and access to camp and lifeline services in the aftermath of refugee and population emergencies (*5, 6*) and natural disasters (*7–9*).

The cluster sampling method can be used to conduct rapid assessment of needs in affected communities after natural disasters. For example, a modified cluster sampling approach was applied in three needs-assessment surveys after Hurricane Andrew struck Florida in 1992 (*9*). These surveys were modelled after the method of WHO's Expanded Programme on Immunization (EPI) which was designed to estimate immunization coverage.

They provided demographic and other information on household composition, trauma and post-trauma injuries and illnesses, availability of prescription medicines, water supplies and electricity, and the status of communications and transportation, and indicated the priority areas so that relief officials could focus on appropriate and effective responses. According to WHO, this cluster design is easy to implement in the field, requires few resources, and yields reasonably valid and precise estimates with relatively quick turnover for analysis and reporting. With this design, the width of the 95% confidence limits on a population percentage never exceeds 20% (point estimate plus or minus 10%) if 30 clusters of 7 subjects each are studied, as long as the design effect does not exceed 2.0 (*10*). Since its initial application in the 1980s, the method has been modified to yield greater accuracy for evaluating EPI programmes (*11*).

Experience in disaster settings, however, has shown that such assessments must provide estimates of the number of people, rather than only the percentage of people, needing specific assistance at the disaster site. Moreover, owing to changing needs as the disaster evolves, the surveys may have to be repeated.

This paper describes a modification of the cluster-sampling method, which can be used to provide the following information.

• Estimates of the *population remaining* in the area, which may differ from the pre-disaster population and may change with time, as the affected community moves to areas where public services become available.

• Estimates of the *number* of people with specific needs in the area after the disaster.

[1] Epidemiologist, Disaster Assessment and Epidemiology Section, Health Studies Branch, Division of Environmental Hazards and Health Effects, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Highway, NE (Mailstop F-46), Atlanta, GA 30341-3724, USA. Requests for reprints should be addressed to this author.

[2] Professor, Department of Epidemiology, The Rollins School of Public Health of Emory University, Atlanta, GA, USA.

[3] Professor, Department of Biostatistics, The Rollins School of Public Health of Emory University, Atlanta, GA, USA.

This approach differs from those used previously in disaster surveys, which usually estimated only the proportion of the population with specific needs. Other field issues related to rapid needs assessment are also discussed.

Methods

We propose a modified $n \times k$ survey design which estimates the remaining population, severity of damage, the proportion and number of people with specific needs, the number of damaged or destroyed and remaining housing units (HUs), and the changes in estimates over a period of time as part of the survey.

Procedure

• Divide the disaster site into a number of comprehensive, mutually exclusive blocks or clusters. The size of each block should be small enough so that the total number of HUs in each selected block can be counted. The division may be based on street grids, if available, or on natural geographical boundaries, such as rivers or hills as identified on topographic maps. Each block or cluster should have well-defined boundaries so that personnel can identify it in the field if it is chosen for inclusion in the sample.

• Preliminarily, estimate the number of HUs in each block by using census information, aerial maps, data from local officials, or other available sources. Denote this preliminary estimate of the number of HUs in cluster i by H_i' and the estimated total number of HUs in the area by H', where

$$H' = \sum_{i=1}^{N} H_i' \qquad (1)$$

and N is the total number of blocks or clusters in the disaster site, typically referred to as primary sampling units (PSUs).

• Select a sample of n blocks with probability proportional to the estimated number of HUs. The EPI method uses $n = 30$ clusters. In many instances, systematic probability proportional to size sampling is used on a sampling frame where the PSUs or blocks are ordered by geographical proximity.

• Within each sampled cluster, count (and list, if feasible) all HUs; denote the total number by H_i. In addition, count and indicate the number of destroyed HUs, denoted by D_i.

• Assuming that all of the HUs in a given block i are listed, choose an equal probability sample, without replacement, of k_i HUs. Systematic random sampling often is used when the HUs are listed by geographical proximity. The EPI method selects as many HUs as needed in order to identify seven subjects within a sample cluster, but our recommended strategy is to choose a fixed number, k_i, of HUs per sample cluster (11). However, if a selected cluster is too large to list all HUs, count or approximate the count of HUs in the cluster. Then use the segmenting procedure and select a segment from each cluster, as described by Brogan et al. (11).

• Count the number of people living in each selected HU; denote the number of people in household j of PSU or block i by C_{ij}. Administer the questionnaire to a member of each selected household who is capable of responding to the questions about household composition and individual needs; this person does not need to be selected at random. Similarly, denote the total number of people with a specific need in household j of block i by R_{ij}.

• If no one is at home, identify a neighbour to obtain information about the selected household or return later at a time when someone is likely to be at home. Efforts should be made to obtain information about the occupants of every randomly selected HU. We do not recommend substitution of new sample HUs for sample HUs in which no one is at home. Sample HUs that are vacant (unoccupied) are recorded as $C_{ij} = R_{ij} = 0$.

Population estimation. The total post-disaster population is defined as:

$$C = \sum_{i=1}^{N} \sum_{j=1}^{H_i} C_{ij} \qquad (2)$$

The point estimate of C is:

$$\hat{C} = \sum_{i=1}^{n} \sum_{j=1}^{k_i} w_i C_{ij} \qquad (3)$$

where $w_i = (1/n) \times (H'/H_i') \times (H_i/k_i)$, and k_i represents the number of HUs actually sampled in cluster i.

The approximate estimated variance (12) of this estimated population total is given by:

$$V\hat{a}r(\hat{C}) \cong \frac{n}{n-1} \sum_{i=1}^{n} \left(w_i \sum_{j=1}^{k_i} C_{ij} - \hat{C}/n \right)^2 \qquad (4)$$

Needs estimation. Similarly, the total number of people in the area with a particular need, R, is defined as

$$R = \sum_{i=1}^{N} \sum_{j=1}^{H_i} R_{ij} \qquad (5)$$

The point estimate of R is:

$$\hat{R} = \sum_{i=1}^{n} \sum_{j=1}^{k_i} w_i R_{ij} \qquad (6)$$

The estimated variance of $\hat{R}$ is given by equation (4), by substituting R_{ij} for C_{ij} and $\hat{R}$ for $\hat{C}$.

Estimating the number of destroyed HUs. The total number of destroyed HUs is defined as:

$$D = \sum_{i=1}^{N} D_i \qquad (7)$$

The point estimate of D is:

$$\hat{D} = \sum_{i=1}^{n} W_i D_i \qquad (8)$$

where $W_i = (1/n) \times (H'/H_i')$.

The approximate estimated variance of this total number of destroyed HUs is given by:

$$\hat{Var}(\hat{D}) \cong \frac{n}{n-1} \sum_{i=1}^{n} \left(W_i D_i - \hat{D}/n \right)^2 \qquad (9)$$

Assumptions

The assumptions involved in the point estimation and estimated variance formulas are outlined below.

• The PSUs on the sampling frame are mutually exclusive and cover the geographical area of interest, i.e., the disaster area.

• First-stage sampling of the PSUs (or blocks) is *with* replacement or, if *without* replacement, n is a small percentage of N so that it is not relevant to incorporate the finite population correction factor into the estimated variance.

• First-stage sampling is with unequal probabilities, i.e., probability is proportional to size. However, the formulas can still be used if the first-stage sampling of blocks or clusters is with equal probability.

• Second-stage sampling of the k_i HUs within sampled PSU i is with equal probability.

• Information is obtained on all members of the sampled HU.

• The number of sampled PSUs is large enough for the approximate variance (equation (4)) to be

valid; a recommended sample is 30 sampled PSUs (13).

Other field issues

Stratified surveys. Surveys can be stratified on any relevant variable, such as severity of damage. Because disasters can differentially affect areas in the disaster zone, separate surveys may be obtained for different areas, such as those where damage is low, medium, or high. Such heterogeneity in needs can reflect variation in housing design and construction, existence of warning systems, or geographical location. Separate or stratified surveys, e.g., those examining the severity of damage, can provide a more comprehensive indication of needs in the affected communities.

In a stratified survey, if one desires to estimate population parameters for *each* stratum, then a sample size of 30 PSUs per stratum is required. Data from all (or some) strata can be combined, with appropriate weighting, to estimate population parameters for part of or the entire geographical area.

If one does not desire to estimate population parameters for each stratum, one may still wish to stratify the PSUs and use stratified random sampling to make sure that each stratum is represented in the sample. In this situation, the sample size does not need to be 30 PSUs per stratum, but a total of 30 PSUs or more. If stratification is used, then the formulas above are not valid. Appropriate modifications of these formulas for stratified surveys are given in Annex 1.

Repeat surveys over time. Because post-disaster needs will vary with time, surveys should be repeated to assess changes in needs. For instance, immediate post-impact needs focus on search and rescue, first aid, and the provision of food and water. Three days later, however, response activities may emphasize establishing temporary shelters and epidemiological surveillance systems among the encamped population. A week after the disaster, priorities during relief and recovery may shift to restoring communications, transportation, and other lifeline systems (14).

Ideally, a survey should be conducted immediately after the disaster, then repeated a few days later, and perhaps weekly thereafter throughout the recovery period, for roughly up to 1 month. One could conduct a repeat survey by resampling the PSUs and, if post-disaster populations change, also by changing the selection probabilities. Alternatively, one could return to the same households for each repeat survey. This option would provide direct

follow-up information about changing needs but would complicate the interpretation and analysis if, for example, some HUs that were occupied when the first survey was conducted were unoccupied at the time of the second survey. Conversely, interpretation would also be difficult if HUs that were not occupied when the first survey was performed were occupied when the second survey was conducted.

First-stage sampling without replacement and large sampling fraction. If sampling is done without replacement and the fraction of clusters sampled at the first stage is large (e.g., $n/N > 0.05$), then the finite population correction factor may not be negligible. In this situation, equation (4) will tend to overestimate the variance. For the simple situation in which clusters are sampled with equal probability at the first stage, the variance estimate with the finite population correction factor is given in Annex 2. (Formulas for the point estimates remain the same as those shown in the text.)

The variance estimate, which includes a finite population correction factor when one uses unequal probabilities for sampling at the first stage, is complicated (12). However, using this complicated equation can be avoided by sampling with replacement or by using a large number of clusters and sampling a small fraction of them (e.g., less than 5–10%).

Discussion

Our proposed modification differs from the EPI sampling method and from the modified cluster methods used previously in disaster settings. The EPI method calls for the random selection of a HU within a cluster or starting point. Selection of individuals begins from the starting point and continues until seven individuals (whose ages are of interest to immunization status) from the next nearest HUs are obtained. In the last HU, all members of the age group of interest are added so that a cluster may contain more than a minimum of seven individuals (10). In applying the cluster design after the Hurricane Andrew disaster in the USA, interviewers arrived near the centre of each of 30 clusters, walked in a randomly selected direction (indicated by a coin toss) to the nearest occupied HU, and interviewed an adult member of that HU. They then went consecutively to the next nearest HU until they had completed seven interviews with people in occupied HUs.

Unoccupied HUs were not revisited. In the case of a multifamily dwelling, only the people in the first occupied HU were interviewed. If a cluster was nonresidential or destroyed, interviewers moved to the next closest cluster in a randomly chosen direction (9). In addition to estimating the proportion and number of people with specific needs, our proposed method also estimates the remaining population and the number of damaged or destroyed HUs.

These results show that simple modification of the EPI cluster-sampling method can provide information about the size of the post-disaster population and the magnitude of their needs. More than one survey or a stratified survey may be needed if the affected area is large or the damage is heterogeneous. Repeat surveys may be needed to assess changes in needs over a period of time.

To date, the cluster design has been used after disasters where the extent of damage is widespread, such as after a tropical cyclone (7, 9). After other disasters such as earthquakes, some areas of the disaster site may be more affected than others, such as those where older buildings were constructed before stringent earthquake-resistant codes were enacted. A complete survey (i.e., a 100% sample) of all affected areas may be more appropriate if the damage exists in certain areas only or in neighbourhoods located in the disaster zone.

Because the information must be timely and made available quickly to decision-makers, the logistics for implementing a rapid needs assessment merit careful consideration. A brief questionnaire should be structured so that it can be quickly completed; survey organizers should ensure that the number of people on needs assessment teams are adequate for conducting at least 210 interviews (30 blocks × 7 households). The number k_i of HUs to be selected for the sample can be increased beyond seven to account for unoccupied units. If the primary objective is to estimate C and R, then the unoccupied HUs do provide information, i.e., 0 people live there and 0 are in need. Thus, increasing the number k_i of HUs would be unnecessary. If, however, the primary objective is to estimate the proportion of people with needs, then it is important to have seven occupied HUs from each block. For example, if the researchers preliminarily estimate that 75% of HUs will be occupied, then k_i can be taken as 7/0.75. In this way, approximately seven occupied households should provide information about the needs in each cluster. Reasonable time estimates should be factored into the data collection process — i.e., shorter periods in urban areas where HUs are generally closer together, and longer periods in remote, rural settings where households may be spread apart.

Finally, other information can be included in the assessment. To assess the severity of damage, a needs assessment can provide estimates of 1) the number or percentage of destroyed HUs, and 2) the number or percentage of habitable HUs.

A rapid assessment can also provide estimates of disaster-related mortality and morbidity. This information would be especially helpful in areas where record-keeping is scanty or nonexistent. Rapid assessments of acute health conditions among high-risk subgroups, such as people with respiratory or diarrhoeal diseases that occur after geological and hydro-meteorological disasters, can be included to determine any departures from endemic levels. In addition, a rapid needs assessment may include environmental sampling to determine health outcomes and possible toxic exposures, such as pulmonary toxicity from airborne ash particles after volcanic eruptions, and gastrointestinal illness from biologically or chemically contaminated groundwater after floods.

Initial assessments of disaster-affected areas, which may be based on aerial photographs or verbal reports, can indicate the area to be surveyed and also the need to conduct more than one survey if the area is large and the damage is widespread. Alternatively, assessment of health needs could be a function of damage assessment teams and cluster selection could be restructured for the next survey (potentially independent) when the extent of the damage is known. Finally, the survey can only cover accessible areas since some clusters selected during the first stage will be inaccessible. Therefore, estimates for these areas would not represent the population of the entire area initially targeted for the survey.

In summary, the modified EPI cluster-sampling method can be applied to obtain reasonably reliable and valid estimates of post-disaster populations and the magnitude of their needs over a period of time.

Acknowledgement

Résumé

Une méthode modifiée d'échantillonnage par grappes pour l'évaluation des besoins à la suite d'une catastrophe

L'évaluation rapide des besoins au moyen d'un échantillonnage par grappes est reconnue en tant que méthode valable pour obtenir des informations sur les besoins immédiats des communautés à la suite d'événements graves comme une catastrophe naturelle. Cette évaluation fournit des informations rapides et d'une précision raisonnable qui permet de planifier les mesures appropriées et de mettre en place des programmes dans les communautés touchées.

La méthode d'échantillonnage par grappes a été initialement mise au point pour évaluer la couverture vaccinale du programme élargi de vaccination (PEV) de l'Organisation mondiale de la Santé (30 grappes de 7 sujets). Les applications antérieures de cette méthode aux catastrophes naturelles montrent qu'outre l'estimation de la proportion de la population ayant des besoins spécifiques, l'évaluation des besoins doit également fournir une estimation raisonnablement fiable et valable du nombre de personnes concernées et de l'importance de leurs besoins sur une période donnée.

Nous proposons un protocole d'enquête $n \times k$ modifié pour évaluer l'effectif de la population restante, la gravité des dégâts, la proportion et le nombre de personnes ayant des besoins spécifiques, le nombre d'unités d'habitation endommagées, détruites et restantes, et les modifications de ces estimations au cours du temps. La zone sinistrée est partagée en un certain nombre de blocs ou grappes exhaustifs et mutuellement exclusifs. Pour chaque grappe, le nombre d'unités d'habitation est estimé d'après les données des recensements, des photographies aériennes ou d'autres sources disponibles. Un échantillon de blocs est ensuite choisi avec une probabilité proportionnelle au nombre estimé d'unités d'habitation. Ensuite, toutes les unités d'habitation et les unités détruites sont dénombrées à l'intérieur de chaque grappe. Un nombre fixe d'unités, choisi par échantillonnage à probabilité égale sans remplacement, est sélectionné dans chaque grappe. Le nombre de résidents et le nombre de résidents ayant des besoins spécifiques est déterminé pour chaque unité d'habitation retenue. Nous présentons ici des équations modifiées pour les estimations ponctuelles et la variance pour ces situations, et des dérivations similaires pour les estimations comportant une stratification des variables pertinentes, comme la localisation géographique et la gravité des dégâts, ainsi qu'un échantillonnage de premier stade sans remplacement et une vaste fraction échantillonnée utilisant un facteur de correction fini pour la population.

Cette méthode montre que le protocole d'échantillonnage par grappes du PEV, une fois modifié, peut être utilisé pour donner des estimations raisonnablement fiables et valables des populations à la suite d'une catastrophe, et de l'importance de leurs besoins sur une période déterminée.

References

1. **Anker M.** Epidemiological and statistical methods for rapid health assessment: introduction. *World health statistics quarterly*, 1991, **44**: 94–97.
2. **Smith GS.** Development of rapid epidemiologic assessment methods to evaluate health status and delivery of health services. *International journal of epidemiology*, 1989, **18**: S2–S15.
3. **Guha-Sapir D.** Rapid assessment of health needs in mass emergencies: review of current concepts and methods. *World health statistics quarterly*, 1991, **44**: 171–181.
4. **Lillibridge SR, Noji EK, Burkle FM.** Disaster assessment: the emergency health evaluation of a population affected by a disaster. *Annals of emergency medicine*, 1993, **22**: 1715–1720.
5. **de Ville de Goyet C, Seaman J, Geijer U.** *The management of nutritional emergencies in large populations.* Geneva, World Health Organization, 1978.
6. **Toole MJ.** The rapid assessment of health problems in refugee and displaced populations. *Medicine and global survival*, 1994, **1**: 200–207.
7. **Sommer A, Mosley WH.** East Bengal cyclone of November, 1970: epidemiological approach to disaster assessment. *Lancet*, 1972, **1**: 1029–1036.
8. **Centers for Disease Control and Prevention.** Rapid needs assessment following Hurricane Andrew — Florida and Louisiana, 1992. *Morbidity and mortality weekly report*, 1992, **41**: 685–688.
9. **Hlady WG et al.** Use of a modified cluster sampling method to perform rapid needs assessment after Hurricane Andrew. *Annals of emergency medicine*, 1994, **23**: 719–725.
10. **Henderson RH, Sundaresan T.** Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. *Bulletin of the World Health Organization*, 1982, **60**: 253–260.
11. **Brogan D et al.** Increasing the accuracy of the expanded programme on immunization's cluster survey design. *Annals of epidemiology*, 1994, **4**: 302–311.
12. **Shah BV et al.** *Statistical methods and mathematical algorithms used in SUDAAN.* Research Triangle Park, NC, Research Triangle Institute, 1993 (formula 2.1 on p. 4).
13. **Cochran WG.** *Sampling techniques*, 3rd ed. New York, Wiley, 1978.
14. *Assessing needs in the health sector after floods and hurricanes.* Washington, DC, Pan American Health Organization, 1987 (PAHO Technical Paper, No. 11).

Annex 1

Stratified cluster sample

Define the clusters as before. Form G groups (or strata) of clusters with N_g clusters in group g, $g = 1, 2, \ldots, G$. One might form these groups after preliminary assessment of damage in the area so that the damage and needs within each group are relatively homogeneous. For example, one group might have predominantly mild to moderate damage, a second moderate to heavy damage, and a third the most severe damage. Alternatively, the strata may be defined by geography, e.g., by county, quarter.

Denote the preliminary estimate of the number of HUs in cluster i of group g by H_{gi}', and the corresponding preliminary estimate of the total number of HUs in group g by H_g'. Select n_g clusters from each group or stratum g, $g = 1, 2, \ldots, G$ with probability proportional to the estimated post-disaster number of HUs (with replacement or without replacement where n_g/N_g is small). As before, list all HUs in each selected cluster, and denote the total number by H_{gi}. Then, randomly sample (equal probability without replacement) from the list, k_{gi} HUs from each cluster selected at the first stage and interview a person living in that HU (or a neighbour) to determine the number of people living there and the needs of those people. Denote the number of people in HU j of cluster i of group g by C_{gij}, for $j = 1, \ldots, k_{gi}$, $i = 1, \ldots, n_g$, and $g = 1, \ldots, G$, where k_{gi} is the number of HUs actually sampled from cluster i of group g. Then the total population and its estimate are given by:

$$C = \sum_{g=1}^{G} \sum_{i=1}^{N_g} \sum_{j=1}^{H_{gi}} C_{gij} \qquad \hat{C} = \sum_{g=1}^{G} \sum_{i=1}^{n_g} \sum_{j=1}^{k_{gi}} w_{gi} * C_{gij}$$

where $w_{gi} = (1/n_g)\,(H_g'/H_{gi}')\,(H_{gi}/k_{gi})$. The estimated variance is given by:

$$\widehat{Var}(\hat{C}) \cong \sum_{g=1}^{G} \left(\frac{n_g}{n_g - 1} \right) \sum_{i=1}^{n_g} \left(w_{gi} \sum_{j=1}^{k_{gi}} C_{gij} - \hat{C}_g / n_g \right)^2$$

where

$$\hat{C}_g = \sum_{i=1}^{n_g} \sum_{j=1}^{k_{gi}} w_{gi} C_{gij}.$$

Annex 2

Sampling without replacement at Stage 1 and *n/N* is large

We assume that we have conducted a cluster survey as described in the text, sampling without replacement, with equal probability at Stage 1 and *without* stratification. The point estimate of C is as given in equation (3). Using the notation as defined in the text, the variance of the estimated post-disaster population is estimated by:

$$\widehat{Var}_2(\hat{C}) \cong \left(1-\frac{n}{N}\right)\frac{n}{n-1}\sum_{i=1}^{n}\left[w_i\sum_{j=1}^{k_i}C_{ij}-\hat{C}/n\right]^2$$

$$+\frac{n}{N}\sum_{i=1}^{n}\left(\frac{k_i}{k_i-1}\right)\left(1-\frac{k_i}{H_i}\right)\sum_{j=1}^{k_i}w_i^2\left[C_{ij}-C_i/k_i\right]^2$$

where

$$C_i = \sum_{j=1}^{k_i} C_{ij}.$$

If *n/N* is small, i.e., $\doteq 0$, then $\widehat{Var}_2(\hat{C})$ reduces to that given by equation (4).