Quality of Malnutrition Assessment Surveys Conducted During Famine in Ethiopia

Paul B. Spiegel, MD, MPH
Peter Salama, MBBS, MPH
Susan Maloney, MD, MSc
Albertien van der Veen, MSc

In humanitarian emergencies, timely and accurate data are needed to guide decision making by public health care professionals working in the field. These data help determine the affected populations’ needs and allow for monitoring and evaluation of their programs. Such information also enables governments and United Nations (UN) agencies to properly coordinate the overall response and allow donors to allocate funds appropriately and effectively.1 Cross-sectional, population-based surveys are often the most appropriate method for collecting certain essential data, such as prevalence of acute malnutrition, immunization coverage, and retrospective mortality rates of populations during complex emergencies.

Cluster surveys, first used to assess immunization coverage in developing countries,2 are commonly used in humanitarian emergencies because of their purported simplicity, reasonable validity, and precision. They require only approximate estimates of the relative sizes of the population units (eg, villages) sampled; no lists of individuals or households are necessary as with simple random or systematic sampling.3 Cluster sampling has been validated for estimating immunization coverage4 and nutritional status.5,6 Manuals that explain the methods step-by-step are available and widely used in the field.7,8 For nutrition cluster surveys, the recommended standardized method is the 30 × 30 cluster survey, which consists of 2-stage sampling. The first stage requires the grouping of the population into smaller geographical units, such as villages, and then choosing these units, or clusters, proportional to population size; the recommended number of clusters is 30. The second stage requires the selection of households and then children from whom to take anthropometric measurements within each cluster; the recommended number of children is 30.8,9 The choice of 30 clusters is based on statistical considerations for stability and distribution of means and proportion,10 while the choice of 30 children per cluster is based on the number of children necessary to have sufficient precision and the number who can reasonably be measured in 1 day.11

Despite the reported simplicity of conducting cluster surveys, there is evidence that errors in their implementation have likely occurred in humanitarian emergencies. Major methodological errors were identified among 30 cluster surveys designed to measure acute malnutrition prevalence in Ethiopia during the famine of 1999-2000. Donor agencies and NGOs should be educated about the need for improved quality of nutrition assessments and their essential role in directing allocation of scarce food resources.

Context During 1999 and 2000, approximately 10 million people were affected by famine in Ethiopia. Results of nutrition assessments and surveys conducted by humanitarian organizations were used by donors and government agencies to determine needs for food aid and to make other decisions on geographic allocation of limited resources; however, accurate results might have been hampered by methodological errors.

Objectives To identify common methodological errors in nutrition assessments and surveys and to provide practical recommendations for improvement.

Design and Setting Nutrition assessments and surveys (n=125) conducted by 14 nongovernmental organizations (NGOs) in 54 woredas (districts) in Ethiopia from May 1, 1999, through July 31, 2000. Surveys were ranked as valid and precise according to 5 criteria: use of population proportional to size sampling, sample size, number of clusters, number of children per cluster, and use of weight-for-height index.

Main Outcome Measures Number and proportion of surveys that used standard, internationally accepted methods and reported valid and precise results.

Results Fifty-eight of the 125 surveys (46%) were not intended to be standard 30 × 30 cluster surveys. Of the remaining 67 surveys, 6 (9%) met predetermined criteria for validity and precision. All 67 used the anthropometric index of weight-for-height, with 58 (87%) reporting z scores. Fifty-four (81%) used nonrandom sampling without consideration of population size and 6 (9%) had sample sizes of fewer than 500 persons.

Conclusions Major methodological errors were identified among 30 × 30 cluster surveys designed to measure acute malnutrition prevalence in Ethiopia during the famine of 1999-2000. Donor agencies and NGOs should be educated about the need for improved quality of nutrition assessments and their essential role in directing allocation of scarce food resources.

See also Patient Page.

©2004 American Medical Association. All rights reserved.

(Reprinted) JAMA, August 4, 2004—Vol 292, No. 5 613
ian emergencies for years, often with unknown consequences. During the famine in Somalia in 1991-1992, Boss et al \(^{11}\) evaluated 16 nutrition surveys and found a lack of standardization in methods. Garfield \(^{12}\) described 27 nutrition surveys in Iraq during the 1990s and found them to be of uneven quality. For decades, the people of Ethiopia have experienced humanitarian emergencies consisting of recurrent drought, food shortages, and civil strife. \(^{13-19}\) Data from famine early warning systems in 1999 indicated that there was food insecurity and that the nutrition situation was rapidly worsening in many regions throughout Ethiopia. \(^{20}\) Among a population of approximately 63.5 million, more than 10 million people (15.7%) were estimated to need food assistance during the peak of this crisis in 2000. \(^{21}\)

We originally intended to collect and map all nutrition surveys undertaken by nongovernmental organizations (NGOs) and other agencies in Ethiopia for programmatic purposes as well as to compare them with quantities of food aid distributed to the regions during the same time period. However, after evaluating the differing methods used in the nutrition assessments and surveys, we concluded that many of the data were of insufficient quality to be used either in assessing program quality or in making decisions about resource allocation. Therefore, we decided instead to systematically collect and evaluate nutrition assessments and surveys undertaken in woredas (districts) in Ethiopia in 1999 and 2000, to evaluate survey methods and identify common methodological errors among the population-based surveys and to develop recommendations for NGOs, governments, UN agencies, and donors on how to improve the quality and interpretation of population-based nutrition surveys in humanitarian emergencies.

**METHODS**

We collected all accessible nutrition assessments and surveys undertaken by NGOs working in Ethiopia from May 1, 1999, through July 31, 2000. Using a list of registered NGOs working in Ethiopia in 2000 and a report of monthly nutrition coordination meetings between the United Nations Children’s Fund (UNICEF), the World Health Organization, and NGOs, we contacted the NGOs and asked them to provide hard copies of all available nutrition survey reports. These survey reports were collected, evaluated, categorized, and entered into a database, using 12 categories (TABLE): location of survey; date of survey; agency undertaking survey; sample size; number of clusters; sampling method; use of proportional to population size sampling, if appropriate; nutritional indicator; presentation of weight/height data (ie, $z$ score, percentage median); acute malnutrition results; method of reporting measles immunization; and percentage vaccinated.

Nutrition assessments and surveys were divided into 2 categories: rapid nutrition assessments, which used a convenience sample method and were not meant to be probabilistic, representative, or generalizable of the population in the locale being surveyed; and population-based surveys, which used probabilistic sampling methods and were meant to be representative and generalizable. This population-based category was subdivided into those surveys that intended to follow a version of the recommended nutrition cluster survey method \(^{5,6}\) (recommended) and those that did not (other) (Table); the latter refers to one NGO working directly with the Ethiopian government that intentionally did not follow the recommended survey method because their goal was to monitor changes in nutritional status among the same population over time using guidelines designed specifically for Ethiopia’s famine early warning system. \(^{22}\) The data were grouped into 4 broad categories (Table): survey characteristics, survey classification, nutritional status, and measles immunization status. Nutritional anthropometric measurement for children aged 6 through 59 months included mid upper arm circumference (MUAC) and weight-for-height index. The NGOs used 3 different methods for classifying acute malnutrition or wasting for children aged 6 through 59 months: (1) percentage of children having MUAC measurements less than a specified cutoff point; (2) weight-for-height $z$ score derived from the National Center for Health Statistics, US Centers for Disease Control and Prevention (CDC), and World Health Organization reference population; and (3) percentage median weight-for-height. Moderate acute malnutrition was defined as MUAC between 110 and 125 mm, weight-for-height $z$ score between −3.0 and −2.0, or percentage median between 70% and 80%. Severe acute malnutrition was defined as MUAC less than 110 mm, weight-for-height $z$ score less than −3.0, or percentage median less than 70%. Global acute malnutrition, a combination of moderate and severe acute malnutrition, was defined as MUAC less than 125 mm, weight-for-height $z$ score less than −2.0, or percentage median less than 80%. Whether or not edema was included as a case of severe acute malnutrition, as is recommended, \(^{5,6}\) depended on the NGO undertaking the survey. We classified nutrition assessments and surveys as valid (defined as the extent to which a variable measures what it is intended to measure) \(^{23}\) and precise (defined as the range in which the best estimates of a true value approximate the true value, often recorded as a confidence interval) \(^{23}\) if they met all of the following 5 criteria: (1) population-based sampling using probabilistic sampling, including proportional to population size sampling, in the first stage and all children or 1 randomly chosen child in each selected household in the second stage; (2) sample size of 500 or more children aged 6 through 59 months; (3) 25 or more clusters; (4) 20 or more children per cluster; and (5) weight-for-height used as the anthropometric index. Use of proportional to population size sampling is necessary unless all of the clusters are of the same size, which is rarely ever the case; if the report did not mention sampling proportional to population size, we assumed it had not been performed. The number of clusters sampled...
and children measured per cluster depends on numerous factors including logistics, security, and the availability of trained personnel. Thus, it is not always possible to meet the recommendations. Furthermore, the sample size needed depends on estimated prevalence, design effect, precision, and level of confidence desired. Other studies have shown how a reduction in the number of clusters and/or children per cluster can affect the validity and precision of results. Given the above field constraints, statistical factors, and results from previous studies, we decided that 500 or more children, 25 or more clusters, and 20 or more children per cluster were reasonable criteria for a survey to be sufficiently precise. Finally, weight-for-height index is the accepted anthropometric measurement for measuring acute malnutrition in nutrition surveys for children.

The study protocol was submitted through the process of ethical review at the CDC; approval was not needed because this was a retrospective study that did not involve contact with human participants. Data were analyzed using Excel 2000 (Microsoft Corp, Redmond, Wash), EpiInfo 6.04b (CDC, Atlanta, Ga), and ArcView 3.1 (ESRI, Redlands, Calif).

RESULTS

We systematically evaluated 125 nutrition assessments (n=16) and surveys (n=109) performed by 14 NGOs in 54 woredas, 17 zones, and 5 regions in Ethiopia over the 15-month study period (Table and Figure). One umbrella NGO had different teams from different NGOs and locations.

Table. Nutrition Assessment and Survey Results for Children Aged 6 to 59 Months From May 1, 1999-July 31, 2000, Ethiopia (N = 125)

<table>
<thead>
<tr>
<th>Survey characteristics</th>
<th>Rapid Nutrition Assessments (n = 16)</th>
<th>Recommended (n = 67)</th>
<th>Other (n = 42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGOs, No.</td>
<td>8</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Location, No.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Zone</td>
<td>9</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Woreda</td>
<td>13</td>
<td>33</td>
<td>14</td>
</tr>
<tr>
<td>Sampling method, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convenience</td>
<td>16 (100)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>First-stage PPS</td>
<td>NA</td>
<td>10 (14.9)</td>
<td>42 (100)</td>
</tr>
<tr>
<td>Clusters, % mean (median) [range]</td>
<td>2374 (127) [50-27,870] (n = 16)</td>
<td>918 (756) [261-3863] (n = 67)</td>
<td>305 (300) [220-400] (n = 42)</td>
</tr>
<tr>
<td>No. of children/cluster, % mean (median) [range]</td>
<td>NA</td>
<td>20 (22) [2-30] (n = 25)</td>
<td>6 (6) [4-6] (n = 36)</td>
</tr>
<tr>
<td>Survey classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid and precise, No. (%)</td>
<td>0</td>
<td>6 (9.0)</td>
<td>0</td>
</tr>
<tr>
<td>Not valid and precise, No. (%)</td>
<td>16 (100)</td>
<td>61 (91.0)</td>
<td>42 (100)</td>
</tr>
<tr>
<td>Nutritional status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nutritional indicator, No. (%)</td>
<td>14 (87.5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mid upper arm circumference</td>
<td>2 (12.5)</td>
<td>67 (100)</td>
<td>42 (100)</td>
</tr>
<tr>
<td>z Scores</td>
<td>2 (100)</td>
<td>58 (86.6)</td>
<td>0</td>
</tr>
<tr>
<td>% Median</td>
<td>0</td>
<td>9 (13.4)</td>
<td>42 (100)</td>
</tr>
<tr>
<td>Acute malnutrition results, % mean (median) [range]†</td>
<td>32.4 (32.2) [2.4-84.0] (n = 16)</td>
<td>15.9 (12.0) [4.0-62.0] (n = 67)</td>
<td>6.6 (6.8) [2.0-11.7] (n = 42)</td>
</tr>
<tr>
<td>Global</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>23.7 (24.2) [2.4-55.0] (n = 16)</td>
<td>13.5 (11.0) [3.0-42.0] (n = 50)</td>
<td>Not listed</td>
</tr>
<tr>
<td>Severe</td>
<td>8.7 (5.9) [0-29.0] (n = 16)</td>
<td>2.6 (1.0) [0-20.0] (n = 50)</td>
<td>Not listed</td>
</tr>
<tr>
<td>Measles immunization status</td>
<td>0</td>
<td>5 (7.5)</td>
<td>0</td>
</tr>
<tr>
<td>Recorded, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method of reporting, No. (%)</td>
<td>0</td>
<td>5 (7.5)</td>
<td>0</td>
</tr>
<tr>
<td>Only verbal</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Only immunization card</td>
<td>NA</td>
<td>1 (20.0)</td>
<td>NA</td>
</tr>
<tr>
<td>Both</td>
<td>NA</td>
<td>3 (60.0)</td>
<td>NA</td>
</tr>
<tr>
<td>Unknown</td>
<td>NA</td>
<td>1 (20.0)</td>
<td>NA</td>
</tr>
<tr>
<td>Vaccinated, % mean (median) [range]</td>
<td>NA</td>
<td>56.6 (60.0) [25.0-71.0] (n = 5)</td>
<td>NA</td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable; NGO, nongovernmental organization; PPS, proportional to population size.

*NGOs and location (region, zone, woreda) may overlap between the rapid health assessments and the population-based surveys.

†See “Methods” section for definition of global, moderate, and severe acute malnutrition.

©2004 American Medical Association. All rights reserved.
The median percentage of global acute malnutrition reported was 12.0% (range, 4.0%-62.0%; n = 67). The median severe acute malnutrition was 1.0% (range, 0%-20.0%; n = 50).

COMMENT

Our analysis of 125 nutrition assessments and surveys conducted in various regions throughout Ethiopia during 1999 and 2000 revealed that the methods and reporting used by the majority of NGOs to measure the nutritional status of the population were nonstandardized. Furthermore, the design and methods of the surveys often failed to follow cluster sampling recommendations and guidelines, and thus yielded results that could not be considered sufficiently valid or precise. Despite these deficiencies, it is likely that the results of some of these studies influenced policy and decisions about resource allocation. In addition, due to the inadequacies found in the majority of the surveys, it is difficult, if not impossible, to use these data to evaluate how effectively the hundreds of millions of dollars of food aid and related resources provided primarily by North American and European donors to the Ethiopian government in 1999 and 2000 were used (eg, the US government provided $321 million in 2000).26

Personnel in NGOs often find themselves working in extreme and dangerous situations with insufficient training and supervision. Some regions in Ethiopia during 1999 and 2000, especially in the Somali region, were remote and insecure.20,21 Some of the woredas where rapid nutrition assessments occurred may have been too dan-

ent countries that worked independently from each other; they were categorized as different NGOs for the purpose of this study. The rapid nutrition assessments were conducted by 8 NGOs in 13 woredas, 9 zones, and 4 regions; none of these 16 assessments met the eligibility criteria to be valid and precise. Of the 16, 14 were among children aged 6 through 59 months while 2 were among elderly individuals (60 years and older). None recorded measles immunization status. Forty-two population-based nutrition cluster surveys (conducted by 1 NGO in 14 woredas, 5 zones, and 2 regions) intentionally used smaller sample sizes and fewer clusters and thus did not meet the study eligibility criteria to be valid and precise. None recorded measles immunization status.

The remaining 67 population-based nutrition cluster surveys were conducted by 9 NGOs in 33 woredas, 15 zones, and 5 regions; all 67 intended to undertake a version of the 30 × 30 cluster survey and so were analyzed in more detail. All 67 evaluated children aged 6 through 59 months; 2 also included elderly individuals, and 1 included adults (18-59 years) and elderly individuals. Of the 67 surveys examining children, all used weight-for-height as the anthropometric calculation and 58 (86.6%) reported the prevalence of acute malnutrition using z scores, while 9 (13.4%) used percentage median. Measles immunization status was included in 5 (7.5%) of the surveys.

Of the 67 survey reports, 6 (9.0%) met all 5 eligibility criteria to be valid and precise. Ten (14.9%) stated that they had used proportional to population size sampling during first-stage sampling and 54 (80.6%) that they had used nonrandom sampling to choose the clusters; 3 (4.5%) did not state the sampling method and thus were counted as not having used proportional to population size sampling. The nonrandom sampling identified consisted primarily of choosing clusters without consideration of population size. All 67 surveys reported sample sizes, of which 6 surveys (9.0%) had fewer than 500 children, 36 (53.7%) had between 500 to 899 children, and 25 (37.3%) had more than 900 children. Of the 25 (37.3%) that stated the number of clusters sampled and number of children per cluster, 10 (40.0%) had fewer than 25 clusters, 2 (8.0%) had 25 to 30 clusters, and 8 (32.0%) had more than 30 clusters; no surveys had fewer than 20 as the mean number of children measured per cluster, 3 (12.0%) had 20 to 30 children, and 22 (88.0%) had more than 30 children.
dangerous to implement a proper survey. These assessments of convenience samples may provide some evidence that there is a nutritional crisis and that a more in-depth survey is needed. It is not our intention to demean humanitarian NGOs' motives or quality of work, but rather to document the inadequacies and suggest practical means to rectify the situation in the future.

The 30×30 nutrition cluster survey, despite documented methodological limitations,27-30 is the recommended method for measuring acute malnutrition in humanitarian emergencies whenever simple random or systematic sampling is not possible. In Ethiopia, a number of errors in design and methods were documented that reduced the usefulness and generalizability of the survey data. First, proportional to population size sampling was not used in the majority of surveys for first-stage sampling; therefore, clusters with large populations were undersampled, introducing bias that reduced the validity of the estimate. Second, sample sizes, the number of clusters, and the number of persons per cluster sampled were insufficient in many of these surveys, decreasing the precision of the overall estimates. Third, although the methods for second-stage sampling were not recorded or stated in most survey reports and not systematically recorded in our study, some surveys chose only 1 child per household and at least 1 survey chose the youngest child to be sampled. While the former could be adjusted during analysis, the latter is not possible to correct. These errors again introduced bias that reduced the validity of the estimate.

Measles is one of the leading causes of death among children in humanitarian emergencies.27-30 It is also preventable through vaccination. Of the 109 population-based nutrition surveys listed in the Table, only 5 (4.8%) recorded children’s measles immunization coverage. Children’s measles immunization status is easy to record from the same children being assessed for nutritional status and should always be included in any nutrition survey. Such data are critical to program decision making and when not available may contribute to organizational complacency and inaction in implementing a measles immunization campaign, as occurred in Gode, Ethiopia, in 2000, with fatal consequences.19

Overall, the presentation of nutritional data in Ethiopia in 1999 and 2000 improved compared with that reported previously for assessments conducted during the famine in Somalia in 1991 and 1992.11 Sixty (48%) of the 125 nutrition assessments and surveys conducted in Ethiopia presented the data as weight-for-height in z scores, the recommended method,28,31 compared with 4 (25%) of the 16 nutrition studies assessed in Somalia. Expression of acute malnutrition results in weight-for-height z scores is recommended because they are more statistically meaningful than percentage median.

Most of the methodological errors occurring during the survey design and data analysis phases can be prevented by organizations having properly trained technical personnel in the field at the time of the surveys. However, due to the high turnover of health care personnel among NGOs, many of whom have clinical and not public health or epidemiology backgrounds, finding and retaining people with such skills is difficult. Furthermore, training NGO personnel in survey methods and analysis is time-consuming and expensive. Unfortunately, many donors—the main source of funds for most NGOs—either do not consider training to be a priority or they consider training and the conduct of surveys to be indirect or administrative costs that divert funds from the beneficiaries. Since these surveys generally need to occur every 3 to 6 months depending on the severity of the humanitarian emergency, one possibility is for a group of NGOs to develop a small cadre of shared competent technical personnel who can be sent to the field to train local staff and oversee the implementation of these surveys. Another solution may be for some NGOs to form an alliance with certain other specialized NGOs whose main focus is nutrition and who have trained personnel specifically for this objective. However, both solutions require a level of coordination and cooperation that is not common in the field and requires funds specifically designated by donors for this purpose.

We believe that the 30×30 cluster survey method is sufficiently well-documented in many commonly used manuals in the field and probably cannot be further simplified without sacrificing its validity and precision. As was eventually done in Ethiopia, a sample survey questionnaire, translated into local languages with a training manual that includes a sampling method that has been modified according to the situation in the country and the context of the crisis, should be developed and shared among the various NGOs, UN agencies, and government agencies. Furthermore, EpiInfo files for data entry as well as programs for analysis and a manual for interpretation should also be developed and shared. Survey questionnaires should aim not only to measure the prevalence of malnutrition but, unlike many of the surveys in Ethiopia, should also examine potential underlying causes that may then be addressed through future programming.

Besides the need for technical expertise among the NGOs, coordinating bodies of the humanitarian emergency, such as governments and UN agencies, should have staff with the appropriate expertise to evaluate, interpret, and use survey results to effect program management and planning. Donor agencies do not necessarily need to have staff in-country during the crisis, but they should at least have access to and use this expertise to inform their decision making. For such an arrangement to be successful, survey reports will need to be composed in a more comprehensive and systematic way than was done in Ethiopia. We propose a centralized coordinating body that can enforce the standards, coordinate where nutrition surveys take place, review protocols, peer-review survey reports and possibly the data, and act as a repository for the surveys.
MALNUTRITION SURVEYS DURING FAMINE

Limitations
There are some limitations to this study. We may not have had access to all nutrition surveys undertaken during this time period. However, we believe this list to be comprehensive because each of us worked as a nutrition technical advisor for a UN agency at various times in Ethiopia from June through October 2000, thus chairing or having access to regular nutrition coordination meetings as well as access to complete lists of NGOs working in Ethiopia during this time. Misclassification of surveys as not valid and precise may have occurred due to insufficient documentation by the NGOs of survey and analysis methods. Furthermore, some surveys may have been sufficiently precise with a sample size of fewer than 500 children and/or fewer than 25 clusters and/or fewer than 20 children per cluster, depending on the estimated prevalence, homogeneity of malnutrition among and within clusters, and level of precision and confidence desired. Finally, we were not able to evaluate other important factors affecting the validity or precision of survey results, such as properly applying proportional to population size sampling, choosing and replacing households appropriately, and accurately weighing, measuring, and discerning edema. If these essential elements of undertaking a survey were not done correctly, then proper sampling and analysis of a survey is meaningless.

Epilogue
As a consequence of the 1999-2000 crisis, an Emergency Nutrition Coordination Unit (ENCU) was established within the Ethiopian government. New national guidelines for 30 × 30 nutrition cluster surveys were developed and officially released in 2002. All NGOs active in the recent food crisis in Ethiopia have agreed to use them. The ENCU is the centralized coordinating body for organizations undertaking nutrition surveys in Ethiopia and, in conjunction with some NGOs, provides training in nutrition survey methods to other NGOs. Because there were no WHO guidelines for nutrition surveys at this time and because data collection methods and precision in probability proportionate to size clustering were not available to the public, this study was the first to use WHO guidelines.

REFERENCES
3. Bennett S, Woods T, Liyanage WM, Smith DL. A simplified general method for cluster-sample surveys of malnutrition among and within clusters, and level of precision and confidence desired. Finally, we were not able