Measuring anthropometric indicators through nutrition surveillance in humanitarian settings: Options, issues, and ways forward

Oleg Bilukha, Claudine Prudhon, Grainne Moloney, Peter Hailey, and David Doledec

Abstract

The technical discourse on nutrition surveillance started decades ago, and the first technical guidelines were proposed in mid-1970s. In spite of this long history, little evidence and consensus exists on the best methods for conducting nutrition surveillance, and on the validity of data produced by these approaches. Multiple nutrition surveillance systems exist in humanitarian settings; however, the validity and usefulness of data produced by these systems are often questionable. In this paper, we outline and define five major methodological approaches to collecting child anthropometric data through surveillance: repeated surveys, community-based sentinel sites, mass screenings, admission data from feeding centers, and data from health clinics. We discuss outstanding methodological and practical challenges with direct implications for quality, validity, and interpretability of collected data and highlight comparative advantages and disadvantages of different methods. We also propose ways forward to building a better evidence base by documenting the strengths and limitations of different approaches, with the eventual goal of achieving consensus on the best ways to collect anthropometric data through surveillance.

Key words: Anthropometry, methods, nutrition surveillance, validity

Introduction

A common definition of nutritional surveillance is “to watch over nutrition in order to make decisions that lead to improvement in nutrition in populations” [1, 2]. Surveillance may draw upon several types of data, but the essential features are that data are collected in a systematic manner over time, and that data collection and analysis are linked to decision-making. The main application areas of nutrition surveillance are policy and planning in the medium-to-long term, program management and evaluation, and timely warning [3]. Anthropometric data, especially prevalence or incidence of acute malnutrition, are crucial indicators to gauge the severity of a crisis, to track trends, and to inform funding and operational decisions during and after emergencies [4]. Anthropometric data generated by nutrition surveillance should allow estimation of levels of malnutrition, tracking trends over time, and differentiating high- and low-risk areas.

In 1974, the World Food Conference emphasized the need for a global nutrition surveillance encompassing all factors influencing food consumption patterns and nutritional status [5]. The World Health Organization (WHO), UNICEF, and the Food and Agriculture Organization (FAO) subsequently convened an expert meeting to define the scope, methodology, and principles of nutrition surveillance [5]. Health services, and especially mother and child health services, were recommended as a major source of anthropometric data on preschool children. During the 1980s and early 1990s, many national nutrition surveillance systems were implemented [6].

However, decades later, no common agreement exists on the best methods to implement and conduct nutrition surveillance, and stakeholders often lack confidence in surveillance data. Many methodological challenges remain, for example, weak evidence about the validity of estimates produced by surveillance, lack of clarity in defining, classifying, and describing different methodological approaches to nutrition surveillance, and few comparable data on costs of different
potentially effective systems [7].

Although linking information with action is of paramount importance, recent experience has demonstrated varying degrees of success for a variety of nutrition surveillance systems [6]. Sustainability and continued effectiveness of the systems are also challenging. Many information systems have “withered away” as donor interest has waned, either because the area served by the system has not experienced crisis for a number of years, or because the results produced by the systems were not seen as useful for decision-making or the validity of these results was questionable.

The current renewed interest in obtaining regular high-quality anthropometric data for informing decision-making and monitoring the effect of global events such as climate change and food price crises provides a good opportunity to revisit nutrition surveillance. This paper focuses on nutrition surveillance systems collecting anthropometric indicators in preschool children, with or without collection of additional indicators. Systems not collecting anthropometric data in children will not be discussed. The question of whether anthropometry (especially weight-for-height and mid-upper-arm-circumference [MUAC]) is an “early” or “late” indicator in nutrition emergencies and thus can or cannot provide useful information for early warning has not been convincingly resolved to date [8–10], and therefore the suitability of anthropometric surveillance for early warning will not be discussed further.

In this paper, we outline and define major methodological approaches to collecting anthropometric data through surveillance, discuss outstanding methodological and practical challenges with direct implications for the quality, validity, and interpretability of collected data, and highlight comparative advantages and disadvantages of different methods. We also discuss ways forward to building a better evidence base for different approaches through documenting their strengths and limitations, with the eventual goal of achieving consensus on the best ways to collect anthropometric data through surveillance depending on setting, available resources, and other contextual factors. Policy issues, such as achieving stakeholder buy-in, building systems and capacities, and using data for decision-making and sustainability of the systems, will not be discussed in depth but may be mentioned insofar as they relate to system design and properties.

Design of nutrition surveillance systems

The five types of nutrition surveillance designs discussed below are repeated nutrition surveys, community-based sentinel sites, periodic exhaustive screenings, admission data from feeding programs for moderately and severely malnourished children, and anthropometric data from health clinics.

Repeated nutrition surveys

In this surveillance design, population-representative cross-sectional nutrition surveys are conducted in the same area at specified time intervals, usually every 3, 6, or 12 months. Simple or systematic sampling can be used in small, well-defined areas such as refugee camps, whereas cluster sampling is used in larger dispersed populations. In cluster surveys, new clusters may be drawn for each round, or clusters drawn initially may be used in subsequent rounds of data collection. Anthropometric indicators may include prevalence of global acute malnutrition (GAM) (wasting and/or edema), stunting, underweight, and low MUAC. Additional indicators can be included in the survey questionnaire, such as morbidity, mortality, immunization coverage, access to water and sanitation facilities, feeding practices, food security indicators, etc. As a general rule, the number of additional variables should be kept to a necessary minimum in order not to compromise the quality of anthropometric measurements.

This approach has several disadvantages. First, it does not allow for continuous monitoring of trends, since cross-sectional data collection rounds are relatively infrequent (every 3 to 12 months). Second, it requires advanced technical expertise for survey design, data collection, and data analysis and substantial logistical support (vehicles, personnel) during collection of field data. Lastly, the estimate produced in the survey is representative of the whole survey area and does not allow for disaggregation to smaller areas or settlements. If estimates representative of several small areas are needed, then surveys need to be conducted in each area, multiplying resource requirements.

On the other hand, the advantages of this approach include the use of standard survey methodology, widely recognized by the stakeholders as a reliable method of data collection generating population-representative estimates, and the possibility of collecting a number of additional indicators.

Examples of the current nutrition surveillance systems using this approach include surveillance implemented by Action Contre la Faim USA (ACF-USA) in famine-affected areas in northern Kenya (Garissa and Mandera) [11] and Uganda (Karamoja) [12], and biannual rounds of surveys conducted in Somalia by the Food Security and Nutrition Analysis Unit-Somalia (FSNAU) of FAO [13–15]. ACF-USA conducts quarterly surveys (four per year) generating district-level estimates. In Uganda, estimates representative of livelihood zones are also generated through weighted analysis of clusters from different districts that fall in a given livelihood zone. Some of the indicators measured, in addition to anthropometric features, include morbidity, mortality, measles immunization and bednet coverage, access to safe water sources, excreta disposal practices, and food security indicators such as types
of food consumed and sources of food. FSNAU conducts biannual nutrition surveys (in April–June and October–December) throughout Somalia. Surveys are representative of the livelihood zone within a region (for example, if the region has three livelihood zones, three surveys representative of each of these zones are conducted every 6 months). In addition to anthropometric indicators, these surveys collect information on household demographics, infant and young child feeding practices (IYCF), morbidity, maternal health, and food security indicators, such as coping strategies and dietary diversity.

**Community-based sentinel sites**

We use this term specifically to denote a nutrition surveillance design whereby periodic cross-sectional data collection rounds take place in preselected settlements (not to be confused, for example, with systems collecting data from sentinel health clinics or sentinel feeding centers). In essence, these are periodic small random surveys conducted in each of the sentinel sites. Sample size usually varies between 30 and 50 children per site, and periodicity between every month and every 3 months. Sites are usually villages or internally displaced persons (IDP) camps, and children within each site are selected through some type of random sampling. New children may be selected in each round, or the same cohort of children may be followed for longer time periods. In the latter case, only a few children at the lowest end of the age spectrum may need to be added each round to replace those “aging up” in order to maintain the sample size and representative age distribution of the sample; in the former case, the whole sample is selected anew each round. Anthropometric indicators may include GAM, stunting, underweight, and low MUAC. As with repeated surveys, additional indicators can be included in the questionnaire used during data collection. Often a certain number of sites (usually 5 to 10) are selected in each area of interest (such as the district, province, or livelihood zone), and the results from these sites are combined at the analysis stage to produce “area-representative” estimates.

Although this type of surveillance design has been used in recent emergencies, there are multiple outstanding methodological issues that need to be resolved before this method can be recommended for wider use. First, selection of sites can be random or purposive. If the sites are selected purposively (for example, often the most “vulnerable” settlements are targeted), then combining the results from individual sites to produce “area-representative” estimates will probably produce a biased (in this example, too pessimistic) picture of the situation. Second, the sites within a larger area may stay the same or may be “rotated” or selected anew for each round of data collection. If new sites are selected for each round, it is not possible to continuously monitor trends at the individual sites. Third, even if sites within a larger area are selected randomly rather than purposively, combining site-specific estimates to produce “area-representative” results is problematic, especially if there is a small number of sites (3 to 10) per area. It is not clear what statistical techniques (if any) should be used to produce combined estimates and confidence intervals around these estimates. Fourth, because of the small sample sizes per site (normally 30 to 50 children), interpretation of the estimates produced in each round of data collection may be very problematic. Small sample sizes result in very wide confidence intervals and large random fluctuations of estimates from round to round, often masking real trends and producing spurious ones, which may lead to erroneous programmatic action. Comparing estimates among individual sites in order to identify the most vulnerable areas would be equally problematic due to the same problem of large random variation in estimates. Lastly, using the same children or households for several consecutive rounds may present additional challenges, including loss to follow-up, children “aging up,” and responder fatigue.

This approach may also require substantial logistical resources, especially if the same teams travel from site to site to conduct these minisurveys. Advantages of this approach include the possibility of disaggregating data to each individual site and the possibility of collecting information on additional indicators.

Examples of sentinel site systems include surveillance currently operated by the State Ministries of Health and UNICEF in the Darfur region of Sudan [16] and the recent sentinel system that was operated by FSNAU in Somalia until mid-2007 [14, 17, 18]. In the Darfur region, data are collected every month from approximately 20 sites (10 IDP settlements and 10 resident settlements) in each state (West, North, and South Darfur), although all sites may not be accessed in each round, depending on the security situation. Thirty children are randomly selected every round at each site, and the same sites are surveyed at every round (no site rotation). Sentinel sites were purposively selected according to predefined criteria, e.g., affected by conflict, frequent population movements, substantial loss of livelihoods, secure access to the site, etc. Data collected include anthropometry (weight, height, MUAC, edema), child morbidity, and food consumption patterns. In addition, data at the community level (population movements, community health, market prices for key commodities) are collected from key informants. In Somalia, sentinel site surveillance data were collected from 132 sentinel sites in Central and Southern Somalia. Data collection rounds took place every 3 months, and 50 children were randomly selected at each round at each site. The sites were purposively selected according to predefined criteria similar to those used in Darfur and were designed to represent livelihood zones within each affected region.
Data collected included anthropometric features of children and women of childbearing age (15 to 49 years), child morbidity, and household food consumption. The system was discontinued in mid-2007 because it failed to detect a deterioration of the nutrition situation in specific crisis-affected regions, although this deterioration was ascertained by nutrition surveys.

**Data from mass screenings**

In this design, exhaustive mass screenings of children of target age are conducted periodically, usually every 1 to 6 months. This is usually done to identify children eligible for admission to feeding programs or to monitor growth through Road-to-Health or similar cards. In some settings, these screenings may be combined with delivery of other health and nutrition interventions, such as vitamin A supplementation, deworming programs, or distribution of micronutrient powders for home fortification. Measurements are done either in the households by door-to-door visits or at the health facility or nutrition center. Because of the large workload, normally data are collected on only one anthropometric indicator (usually weight-for-age or MUAC), and no additional data on nonanthropometric indicators are collected.

The quality of measurements in such screenings is often questioned, highlighting problems of large workload, substandard measuring skills, substandard equipment, and poor documentation of the results, as well as community pressures to enroll more children in the feeding programs. Achieving high coverage may be problematic in some settings, raising questions about potential selection bias. The directionality of such selection bias may differ depending on which population groups are better covered by screening. Better coverage among the neediest, most affected population groups would result in overestimation of the problem, whereas better coverage in the least affected population groups would underestimate the need. In both scenarios, there is a risk that programmatic response can be misguided by biased results, and knowledge of the local situation is critical to understanding which of the scenarios described above is more likely to occur.

The fact that data collection is often combined with delivery of other health and nutrition interventions can be listed as a major advantage of this approach. Because all children (rather than a sample) are measured, there is no need to calculate confidence intervals, and data can be disaggregated to small population units (individual settlements or even neighborhoods within settlements).

Examples of this approach include the Enhanced Outreach Strategy (EOS)/Targeted Supplementary Feeding (TSF) for Child Survival program implemented by the Ethiopian Federal Ministry of Health in collaboration with UNICEF [19, 20] and periodic nutrition screenings conducted in many refugee settings by the United Nations High Commissioner for Refugees (UNHCR) and its implementing partners. EOS/TSF is implemented in selected districts of Ethiopia where TSF programs are available. It is conducted as campaigns every 6 months by mobile teams going from village to village and includes vitamin A distribution and deworming for children under 5 years of age, as well as nutritional screening of all children under 5 years and pregnant and lactating women using MUAC and checking for edema. Children and women identified as acutely malnourished receive two supplementary rations of oil and corn–soya blend. In Bhutanese refugee camps in Nepal, all children aged 6 to 59 months are screened monthly using weight-for-age criteria at the nutrition centers. Children who fall under the screening cutoff have their weight-for-height measured to determine eligibility for admission into the feeding programs. These monthly anthropometric screenings are combined with blanket distribution of micronutrient powders for home fortification to children aged 6 to 59 months. The prevalence of low weight-for-age obtained during these screenings is reported through the Health Information System, a surveillance system tracking many health and nutrition indicators in refugee settings [21].

**Data on feeding program admissions**

This type of nutrition surveillance uses data from feeding programs targeting moderately and/or severely malnourished children. Data are reported as the number of new cases of moderate and/or severe acute malnutrition per given period of time. Incidence rates can be constructed if the denominator (the number of children of targeted age living in the catchment area of the program) is known. Usually, no additional indicators (nutritional or nonnutritional) are reported.

This approach has several limitations. First, serious problems with interpretation of these data can arise if access to, or use of, the feeding center changes over time. For example, a drop in the number of admissions may be due to improved nutritional status, a shortage of available commodities in the center, or decreased access due to insecurity or flooding. Second, the population denominator is often unknown or changing over time, which makes it problematic to produce reliable incidence rates. Third, the quality and completeness of reporting from feeding centers are often substandard. Lastly, it may be very problematic to interpret trends when admission criteria change (for example, change from National Center for Health Statistics [NCHS] Growth Reference % median to WHO Growth Standards z-scores or to MUAC, change in MUAC cutoffs from 110 to 115 mm, etc.). Different feeding centers may change admission criteria at different times, and such changes may not be reported to the surveillance
system in a timely manner.

The fact that the number of new admissions is part of routine feeding programs’ reporting and therefore does not require additional efforts or advanced expertise is an advantage of this approach. It is possible to disaggregate data to areas served by individual feeding centers, and data collection is continuously ongoing, rather than conducted in discrete cross-sectional rounds at specified intervals, as in the approaches described above. Data generated by this approach provide information on the incidence rather than the prevalence of malnutrition, which may be more relevant for program planning and resource allocation.

Examples of this approach include the use of feeding programs admissions data for surveillance purposes in the Darfur nutrition surveillance system [16] and reporting of feeding program admissions in refugee settings through the Health Information System [21]. In Darfur, most of the operational feeding centers report their admission data monthly to the State Ministries of Health and UNICEF. The numbers of new admissions are tallied monthly separately for supplementary (SFP) and therapeutic (TFP/OTP) programs for each of three Darfur states. Incidence rates are not calculated in the absence of reliable denominators. The average number of admissions per site is also tracked to account for the variable number of sites reporting each month [16]. In refugee settings, monthly numbers of admissions to supplementary and therapeutic programs are collected and reported through the Health Information System [21]. Since population data in refugee camps are often reliable and coverage of feeding centers is predictably high, incidence rates can often be calculated and trends interpreted with confidence.

**Data reported from health clinics**

This type of nutrition surveillance uses anthropometric data collected on an ongoing basis from health clinics. Children of target age groups are routinely measured when visiting clinics. In different settings, surveillance may include either all children visiting the clinic for any reason (sick visits, routine checkups, immunizations, etc.) or only children visiting the clinic for some specific reason (e.g., measles immunization). Anthropometric indicators reported may include weight-for-age, weight-for-height, and MUAC.

The representativeness, quality, and interpretability of these data are questioned for several reasons. First, the nutritional status of children visiting clinics is not representative of the nutritional status of children in the population, and the direction of bias is not straightforward. Children attending clinics for immunizations or routine checkups may be healthier than those who do not go to clinics for these reasons. On the other hand, children who are sick may be more likely to be malnourished [22]. Second, the age distribution of children measured at the clinics is severely skewed toward younger children (less than 1 year of age), because these children are much more likely to visit clinics for immunizations and routine checkups. Third, identifying those children who visit the clinic (and thus are measured) multiple times over a given reporting period adds additional complexity and a source of bias. Fourth, clinic staff often have substandard measuring skills, and measuring equipment may be faulty, broken, or poorly calibrated. Staff burnout from the large added workload of measuring every visiting child may contribute to poor quality of the measurements. Lastly, reporting from these facilities is often incomplete and irregular, which complicates trend analysis and other aspects of data interpretation.

Advantages of this approach, similar to admissions data from feeding centers, include the facts that data are collected continuously rather than at discrete intervals, and that data can be disaggregated to the level of individual clinics. If the primary interest of decision makers is the nutritional status of children under 1 year of age who are commonly seen for immunization visits, then routine measurement of these children in clinics could prove useful if acceptable quality of measurement and reporting and acceptable coverage rates could be achieved.

Examples of clinic-based nutrition surveillance include the collection of clinic-based anthropometric data in Somalia with the support of FSNAU [14, 15] and clinic-based nutrition surveillance operated in the West Bank and Gaza by the Palestinian National Authority Ministry of Health in collaboration with UNICEF and WHO [23]. In crisis-affected regions of Somalia, purposively selected health centers measure all children aged 6 to 59 months visiting the clinic for any reason. About three to five health centers per livelihood zone report monthly from each affected region. Weight and height are measured, and the prevalence of low weight-for-height is reported monthly. In the West Bank and Gaza, 70 clinics (44 in the West Bank and 26 in Gaza) that cover approximately 80% of children living in these regions are involved in surveillance. All children coming to these clinics for measles immunization have their height, weight, and hemoglobin level measured. Approximately 94% of all children measured are 9 to 10 months of age, with a few children being slightly younger or older (an overall range from 8 to 13 months). Indicators reported include prevalence rates of wasting, stunting, underweight, overweight (high weight-for-height), and anemia.

Some general characteristics of the five methodological approaches to nutrition surveillance described above are summarized in **Table 1**. One of the most important characteristics of the surveillance system is the frequency of data collection. In two of the discussed surveillance designs (selective feeding centers and health clinics), data collection is continuous, whereas
in the other three it is periodic, conducted at discrete time intervals. The length of the time interval between data collection rounds is critical for the interpretability and usability of surveillance data. Generally, the shorter the interval the better; however, data collection is often costly, and very frequent data collection rounds are seldom feasible. When deciding on the maximum acceptable time interval, consideration should be given to such issues as ability to detect changes in key indicators for timely programmatic action and ability to follow the seasonal trends in key indicators if such trends exist. For example, if there are only two seasons with distinctly different patterns of nutritional status during the year, then two rounds of data collection per year may be sufficient to follow trends. If, however, there are four distinct seasons (two of them “lean”), then twice-a-year data collection may produce misleading results if data collection rounds fall only in “lean” or only in “fat” seasons. In those circumstances, four rounds of data collection may have to be considered.

Another important aspect of surveillance design is its ability to collect data on additional (nonanthropometric) indicators. Of the five designs discussed above, only two (repeated surveys and community-based sentinel sites) are normally used to collect such data. General advantages and limitation of these two designs discussed above are equally applicable to collecting anthropometric and nonanthropometric indicators. Care should be taken to keep the number of additional indicators collected through surveillance to a necessary minimum, as increasing the volume of collected data may often jeopardize data quality.

### Considerations of data validity

The key question that needs to be answered before data produced by any type of surveillance can be used is whether these data are valid, i.e., whether they accurately reflect the magnitude as well as the temporal and geographic variation in malnutrition levels. In other words, whether the prevalence or incidence estimates produced by surveillance are accurate, whether the data allow one to construct valid trends, and whether the data allow one to accurately distinguish between areas with higher and lower levels of malnutrition. If the answers to these questions are negative, any further discussion about the system’s cost, sustainability, feasibility, usefulness, etc. is of little meaning.

An obvious way to answer these questions would be to compare data produced by each type of surveillance with the true values. The true value would be an exhaustive measurement of all children of targeted age in the area where surveillance is implemented. As this would be unfeasible in most situations, a “gold standard” could be used instead. Arguably, the commonly acclaimed gold standard for anthropometric indicators is data generated by good-quality, population-representative anthropometric surveys.

Several studies published in the 1980s and 1990s [24–28] examined the validity of data produced by clinic-based surveillance by comparing these results with estimates produced by nutrition surveys conducted in the same areas and in the same time periods. The results of these comparisons did not appear very encouraging. Most of the studies showed that the prevalence rates of underweight estimated by the two methods (clinic-based surveillance and population-representative surveys) were quite different. Moreover, the magnitude of the difference between clinic-based and survey estimates was not consistent between regions within countries and between countries. In two studies [27, 28] clinic-based surveillance overestimated, and in one [24] it underestimated, the prevalence of underweight compared with survey estimates, whereas in two others [25, 26] clinic-based data overestimated the prevalence of underweight in some areas and age groups and underestimated it in others. Only the study conducted in Swaziland showed relatively close correspondence between clinic-based and survey prevalence.
The time trends in the prevalence of underweight seen in clinic-based data did not correspond to the time trends seen in the surveys [24, 25]. Summarizing evidence from these studies, Pelletier and Johnson [25] concluded that “the validity of clinic-based data for nutrition surveillance for targeting purposes is highly suspect” and that “their validity in each setting should be demonstrated rather than assumed.” Unfortunately, we found no published studies examining in the same way the validity of data produced by other methodological approaches discussed in this paper (community-based sentinel sites, exhaustive screenings, or admissions to feeding programs).

With the advent of SMART survey methodology and software [29], it is much easier for field practitioners to verify the quality of anthropometric measurements and assess how representative age and sex distributions are in collected data samples. The Plausibility Check function available in SMART software can be used to assess the quality of anthropometric data produced by most of the types of surveillance discussed above, including repeated surveys, community-based sentinel sites, exhaustive screenings, and clinic-based data.

Conclusions and way forward

In spite of a long history, little evidence and consensus exists on the best methodological designs for conducting nutrition surveillance and on the validity of data produced by these designs. Multiple nutrition surveillance systems (spanning the spectrum of designs described above, and perhaps some additional designs or variations of designs not mentioned) exist in humanitarian settings, although the validity and usefulness of data produced by some of these systems are questionable at best. Therefore, additional research to assess the validity of surveillance data produced by different methodological approaches is urgently needed.

This research program could start by identifying datasets generated by different designs of surveillance and matching these data with representative surveys conducted in the same area at the same time so that these data could be compared and the validity of surveillance data ascertained. In addition, the quality of anthropometric data generated by surveys and some other surveillance designs, such as sentinel sites and exhaustive screenings, can be assessed with the use of SMART Plausibility Check techniques. These validity analyses may suggest surveillance designs that are more likely to provide valid data.

The financial costs of the different systems also need to be examined. Comparisons of system costs should ideally include considerations of the quality of data produced by systems at current cost and additional costs that would be required to improve data quality to acceptable standards. It might seem less costly to use routine data collection systems rather than create new systems. However, the former might perform poorly, and significant resource investment might be needed to reach acceptable quality.

Consequently, a series of comprehensive surveillance evaluations can be carried out in the field, using standardized rigorous methodology [30]. These evaluations should initially focus on the most promising (in terms of data validity) designs and evaluate multiple aspects of surveillance, including feasibility, acceptability, sustainability, cost, and usefulness for decision-making, among other characteristics. New improved designs could be suggested, piloted in the field, and evaluated. This work should ultimately result in better evidence and more informed technical discourse, and eventually lead to developing international consensus on best practices in conducting nutrition surveillance.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention. The authors declare that they have no competing interests.

References