Annex A: Data Collection Methods and Statistics

Surveys

Nutrition surveys can take a variety of formats but they accomplish the same basic task—the compilation of information about a set of nutrition issues within a population assessed from a subset of that population. Surveys vary in sample size, assessment methodologies, time frame for implementation, and cost.

Identification and description of nutritionally disadvantaged groups for advocacy, targeting, or planning purposes are the primary objectives of nutrition surveys. Survey data can predict the effect of nutrition interventions, identify indicators for monitoring and evaluation purposes, or provide a baseline for project impact evaluation.

One of the most important initial tasks is to explore thoroughly the question of need—is the survey (or the even costlier surveillance system) the only or best way to get at the desired information? If existing information is not adequate, proceed with development of the survey design. Some general guidelines for nutrition surveys are found in Box A-1.

Surveillance systems

Surveillance systems gather information about the nutrition status of a population on a regular (and usually ongoing) basis. Surveillance can be used for (1) identification of nutrition problems, (2) developing policy and program response, (3) famine or hunger crisis early warning and intervention systems, and (4) program management purposes. The disadvantage of formal surveillance systems is their cost (both initial and recurrent) as well as the tendency to generate large amounts of excess information. Before establishing a nutrition surveillance system, determine that periodic surveys cannot produce the desired information.
The health (or any other) sector alone will not solve the nutritional problems of the population.

An analysis of causes is a prerequisite to any decision-making.

A causal model is a key component of the assessment procedure.

Global does not mean total. Data collection, analysis, and action are subject to technical and resource constraints.

A nutritional assessment is not independent of the ideology of its author and its users.

The objectives of the assessment must be clearly defined at the outset.

An assessment does not only consist of collecting data and describing a situation. It is also an explanation and an identification of trends.

The maximum use of existing data is the rule: large surveys are often unnecessary.

Disaggregate the data.

A nutritional assessment is the responsibility of an inter-disciplinary team; both the underlying determinants of malnutrition and the actions to reduce the condition are intersectoral in nature.

(Adapted from Beghin, et al., 1988)
**Sentinel site surveillance**

One solution to the cost and complexity of large national nutrition surveys as well as cumbersome surveillance systems is sentinel site surveillance. Instead of selecting a nationally representative sample, a limited number of sentinel sites monitor the nutrition situation. Several advantages of this option are outlined in Box A-2. However, sentinel site surveys are also

<table>
<thead>
<tr>
<th>Box A-2: Sentinel Site Surveillance: Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better understanding of differences</td>
</tr>
<tr>
<td>More timely data</td>
</tr>
<tr>
<td>Depth rather than breadth of information</td>
</tr>
<tr>
<td>Minimize costs</td>
</tr>
<tr>
<td>Opportunity for developing a more participatory approach to nutrition status assessment</td>
</tr>
</tbody>
</table>

(Adapted from Young and Jaspars, 1995)
prone to some of the same problems noted by Beghin above. Failure to use the data is not uncommon. And, depending on the specific context in which the survey is undertaken, significant population movements over time invalidate the representativeness of the sentinel sites.

Survey design

Sampling

This statistical technique enables extrapolation of the nutrition status of a representative subgroup of a population to the total population (or that population represented by the sample such as children under 5). The sampling frame will generally include information on age, sex, geographic location, and socioeconomic status, with the objectives of the survey or surveillance system determining the specifics of each sampling frame. A detailed discussion of the methods for selecting statistically representative samples from populations of different sizes is beyond the scope of this document, but the straightforward FAO field manual (1990), Conducting small-scale nutrition surveys, is a useful guide. Box A-3 contains an overview of the main types of sampling used in survey design of quantitative studies.

“Purposive” sampling is used for qualitative data collection. For ethnographic and rapid appraisal methods, key informants who are believed to be the best sources on a topic of interest are sought out specifically, without regard for representation of the population. A richer understanding of the context of a community situation is the desired end product of these sampling techniques and data collection methods. Avoid the potential hazard of sampling by self-appointed samples.

One additional option is to abandon sampling altogether. For a small population, it may be easier to simply measure all members of the group. Community participation and ownership may increase in the absence of need for statistical expertise in the preliminary stages of data analysis.
**Box A-3: Selecting a Sample**

- *Simple random sampling* ensures equal probability that a child/individual will be selected for the sample by choosing random numbers and taking the child corresponding to the number. A disadvantage is the need for an accurate list of all members of a population (sampling frame) which is often difficult to acquire. Simple random sampling is appropriate for small surveys/small populations.

- *Systematic sampling* does not require a sampling frame. Individuals are chosen systematically at equal intervals (e.g., every tenth house). Also appropriate for small surveys/small populations.

- *Cluster sampling* is useful for large populations. Groups of children are selected systematically at equal intervals, usually each with the same number of individuals. This method has the advantages of reduced travel by the survey team (in comparison to random sampling), and less information (needs only community groupings and estimated size) for constructing the sampling frame.

  A disadvantage of cluster sampling is the diminished reliability relative to simple random sampling—children within clusters are often more similar in anthropometric status than children between clusters. To compensate for this bias, sample size for a cluster design is usually doubled and a correction factor applied when calculating the confidence interval (Young and Jaspars, 1995).

- *Lot Quality Assurance Sampling (LQAS)*, is a sampling method that is used to obtain reliable information on a small geographic or administrative unit using a small sample. LQAS can be used to accurately detect the extremes of performance; those which are exceeding an “upper threshold” of performance and those fail to meet a “lower threshold” of performance. It is an adaptation of a method used originally to assess the quality of industrial commodities for use in health systems. LQAS is well suited for evaluating coverage and quality of health services. It is characterized by the division of the total intervention population into smaller administrative units and the random sampling of each subdivision, yielding a pictures of the performance or impact of an intervention throughout the entire population of interest. For more information on LQAS, refer to [http://www.coregroup.org/working_groups/LQAS_Participant_Manual_L.pdf](http://www.coregroup.org/working_groups/LQAS_Participant_Manual_L.pdf).
and with the inclusion of all children (for example) in the survey process. Absentees and non-participants mean that the issue of missing data remains a potential source of bias in this scenario.

**Cross-sectional and longitudinal surveys**

Cross-sectional nutrition surveys provide a snapshot of a population's nutrition status. These may be repeated at intervals in order to capture seasonal changes. Longitudinal surveys monitor anthropometric trends in a population over time. Disadvantages of longitudinal nutrition surveys include difficulties with maintaining a sample's representativeness (e.g., children migrate, die, or drop out of growth monitoring programs) and the unique problem of children aging. As children grow older, the cumulative effects of stunting are more evident and wasting is less likely to occur, affecting the interpretation of longitudinal results.

An important result in the Millennium Summit in September 2000, are the Millennium Development Goals (MDGs). One of the MDGs (poverty and hunger) uses child underweight (weight-for-age) in children under 5 years of age as an indicator of progress toward the goal. Tabulation of trends in child underweight rates will be a focus of international data collection and analysis.

**Using anthropometry in surveys**

From Beaton et al., (1990) come a series of recommendations for use of anthropometry in survey design.

**Population level: one-time assessment**

1. Population under food crisis conditions:

   • Prevalence of wasting, measured by weight-for-length, or if this is impractical by arm circumference, provides the best anthropometric indicator for assessing current effects of food shortages.
• No universal trigger level based on wasting prevalence can be recommended. Changes in wasting prevalence may be particularly informative (or particularly labile).

• A sample based on the most vulnerable group—usually children under 5 years of age—will suffice to detect higher than expected prevalence.

• Presentation of prevalence findings in terms of appropriate targeting characteristics—e.g., age/sex, ethnic status, administrative area, socioeconomic status in relation to access to food, etc.—will assist the planner. Examination of frequency distributions (by age, if sample size is large enough) is recommended.

2. Population assessment for long-term planning

• Sampling should be on a representative basis to enable proper comparisons by region or other policy variables. In certain circumstances a sample of convenience may prove adequate (e.g., school entrants).

• The recommended measurement for long-term planning purposes is height/length-for-age. Should this indicator not be feasible to collect, weight-for-age may serve as a substitute.

• For presentation, measurements are converted to prevalence using international references (e.g., NCHS). For analysis, (e.g., differences between groups, associations with possible causal factors, etc.), use indices as continuous variables such as Z-scores (average or %< –2Z) percent of median, or percentile.

Population level: trend assessment

1. Nutritional surveillance for long-term planning:

• As linear growth is a good proxy for general development constraints, trends in length/height-for-age provide information on long-term
changes in the environment and their nutritional consequences. Weight-for-length/height or weight-for-age patterns in children will generally reflect stress resulting from inadequate food consumption and/or changes in infectious disease incidences in the short run. For adults, measures of thinness, such as body mass index, are the most relevant (except for pregnant women).

- The sampling frame—age, sex, area, and socioeconomic status—should be chosen to reflect planning needs. An important consideration is that samples should be comparable over time to allow for trend assessment.

- Presentation is in terms of trends in prevalence below the chosen cut-off point, classified by those sociodemographic and administrative variables required for planning. External or internal references are equally valid if employed consistently over time; choice of cut-off points is not critical but should be comparable over time and will need to be considered in relation to which reference data are to be used.

2. Nutritional surveillance for early warning systems:

- Anthropometry is used for assessment of effects of an impending or actual emergency, for targeting of relief supplies, and for tracking whether certain areas are showing continued deterioration when food is being distributed.

- Anthropometry may serve to identify and validate, retrospectively, agricultural and meteorological predictive indicators for future application.

- Sampling methods depend on the population of concern—usually the most vulnerable, based on historical evidence. Random sampling is encouraged, although clinic-based data often provide a convenient source. Preschool children are likely to show the first signs of stress due to food shortages, but older children and adults should not be ignored.
• Changes in weight-for-height or weight-for-age are the most useful current indicators. Reporting is in terms of trends in prevalence below the cut-off point.

3. Nutritional surveillance for program management:

• Anthropometry is relevant to program management for three reasons: a) to identify target groups, b) in monitoring progress, and c) in assessing overall program impacts.

• Indicators used must relate to the objectives of the program. The choice of indicator depends on factors such as the nature of the intervention and age of target groups. Little anthropometric response to intervention may be found in children of more than two years of age, although there may be other benefits in terms of cognition and immunity, for example.

**Assessment methods**

Nutrition status surveys and surveillance systems routinely collect data through quantitative measurement techniques to answer questions about the location, extent, and severity of malnutrition. To better understand the causes of nutritional deficiencies and the processes of change, qualitative methods, which look at why conditions of poor nutrition occur or persist, complement standard quantitative survey methods.

**Quantitative methods**

Nutrition status is determined through measurement of growth and body composition (anthropometry), analysis of the biochemical content of blood and urine, and by clinical examination of external physical signs of nutrient deficiencies. Additional information comes from quantitative assessment of food availability (food balance sheets, agriculture surveys, household expenditure surveys) and access to and intake of food (24-hr. recall, food frequency questionnaires, market surveys, household income...
and expenditure surveys). The incidence and prevalence of nutrition-related morbidity and mortality (clinic records, hospital admissions records) and availability and consumption of health, agriculture (clinic and health ministry records, agriculture ministry records), water supply, and sanitation services (water board and local public health statistics and records) all supply important details about the multiple determinants of nutrition status in a population.

Qualitative methods

Rooted in the disciplines of sociology and anthropology, qualitative methods are important companion techniques to those enumerated above. Rapid rural appraisal, participatory rural appraisal, and ethnographic methods are used more routinely in Bank poverty assessments, beneficiary assessments, and other sector studies in recent years. Qualitative research techniques are less costly than traditional quantitative study methodologies, may take less time than a large regional survey (depending upon the types of data being collected), and provide a richness of detail not possible with large sample surveys. “Instead of concentrating on representation through sampling, rapid appraisal concentrates on relevance and understanding. For this purpose, a rapid appraisal survey will address sample conditions rather than sample cases, often studying the whole population of conditions (World Bank Togo Poverty Assessment, June 1996—Internal Document).”

Methods such as direct observation, semi-structured interviews with focus groups or key informants, participatory mapping exercises (health, disease, community gardens), seasonal analysis of food availability and disease patterns, food charts, wealth ranking, market surveys, flow diagramming of the food chain and complementary foods practices in relation to seasonality and food availability, matrix ranking of food preferences, and daily activity profiles, answer questions about the behaviors of household members related to food beliefs and practices and the provision of childcare (often called nutrition caring practices). According
to Young and Jaspars (1995), “Perhaps the greatest strength of qualitative approaches . . . is their ability to reveal rapidly the underlying causes of malnutrition and show how these are interrelated and might change on a seasonal basis.”

Disadvantages of qualitative methods include the characteristically smaller sample size with the built-in danger of non-representativeness and possible bias introduced either by the investigator or the participants in the qualitative study.

Field experience indicates that increased community participation in nutrition surveillance and intervention activities is one possible outcome when qualitative research methods are used. And from community ownership of surveys and projects may come sustainability of the intended improvements to nutrition status.

Sensitivity and specificity

Criteria for judging that an indicator is appropriate for a particular context or set of concerns include the ability of the indicator to best reflect the issue or context of concern or to predict an outcome. Other criteria are the sensitivity and specificity of the indicator; the goal is to maximize these attributes (which can be influenced through selection of cut-off points). Figure A-1 illustrates and defines several concepts useful in describing the performance of an indicator and various cut-off points for the indicator (figure and text reproduced from Tucker, et al., 1989). Sensitivity refers to the proportion of truly malnourished individuals who are correctly diagnosed by the indicator. Clearly, sensitivity decreases as the number of false negatives (c) increase. Specificity refers to the proportion of truly non-malnourished (or healthy) individuals selected by the indicator, and its precision decreases as the number of false positives (b) increases. It also declines as sensitivity increases and vice versa. Positive predictive value refers to the proportion of those diagnosed as malnourished by the indicator who are in fact suffering from malnutrition. As with specificity,
the accuracy of this indicator decreases as the number of false positives (b) increases. Its absolute value also falls as the absolute prevalence of the disease decreases.

In trying to identify the “best” indicator and cut-off point, consider the intended use of the indicator. Two major purposes are distinguished here: individual screening and estimating the prevalence of malnutrition in a population.

Choosing the most useful indicator and cut-off points for screening purposes involves trade-offs among the number of false positives, false negatives, and the cost per beneficiary. False positives (seen under low specificity) contribute to higher program costs and, if the treatment has undesirable side effects, may actually harm unintended recipients or
society at large. False negatives (seen under low sensitivity) represent truly needy individuals who are not reached by the program and, because of the inverse relationship between sensitivity and specificity, are usually associated with high specificity and a low cost per beneficiary. Seen from a policy perspective, a program that strives to meet the needs of all truly needy individuals can succeed only with high sensitivities (few false negatives). This usually implies that specificity (many false positives) will be low and, consequently, costs will be high.

Conversely, a program that seeks to exclude as many non-needy individuals as possible requires a high specificity (few false positives). Thereby, program costs are minimized but targeting costs increase. Surveillance can help maximize the benefit to a population for a given level of cost by employing indicators and cut-off points that possess intrinsically high levels of both sensitivity and specificity. However, this must be balanced against the need to use indicators and administrative systems that are relatively inexpensive, technically feasible, and appropriate to the level of training and supervision required for their use.

**Anthropometric data reporting systems**

Evaluating the anthropometric status of individuals or populations involves the use of reference growth standards against which measurements are compared. As discussed earlier, the WHO/NCHS reference is currently the recommended global standard. In order to report anthropometric data relative to the reference, there are three reporting systems commonly in use: Z-scores or standard deviation scores, percentiles, and percent of median values. The following definitions are found in *Physical Status: The Use and Interpretation of Anthropometry* (WHO—Technical Report Series No. 854, 1995).

**Z-scores (or standard deviation score)**

Reporting anthropometric indices in terms of Z-scores is the preferred reporting system because it normalizes the age-specific variances of
different indicators at different ages. An additional advantage is the ability to apply summary statistics (mean and standard deviation) to population-derived Z-scores. Z-scores are calculated by taking the difference between the value for an individual and the median value of the reference population, and dividing it by the standard deviation for the reference population.

\[ Z\text{-score} = \frac{\text{(child's measurement)} - \text{(median values of the reference population)}}{\text{standard deviation of the reference population}} \]

Z-scores for any index are distributed normally, with 2.3% of healthy children falling 2 standard deviations (−2Z) below the median, the usual cut-off used to define malnutrition. If significantly more than 2.3% of children fall below −2Z, then there is probably an undernutrition problem. If significantly more than 2.3% of children are above +2Z weight for height then overnutrition is a likely problem.

**Percentile**

Reporting data as a percentile gives the rank position of an individual on a given reference distribution in terms of what percentage of the group the individual equals or exceeds. A child whose weight falls in the 10th percentile weighs the same or more than 10% of the reference population of children of the same age. Summary statistics cannot be calculated for percentiles and at the extremes of the reference distribution, important changes in weight or height status are not clearly evident when reported as percentiles.

**Percent of the median**

This reporting system presents a child's measurement (weight or height) as a percentage of the median value of the reference population.
though percent of median is easier to calculate than either Z-scores or percentiles, it has the disadvantage of variable interpretation depending upon age and height groups and variable cut-off values for low anthropometry depending upon the index. For example, depending on the child’s age, 80% of the median weight-for-age might be above or below $-2$ Z-scores, resulting in different classification of health risk. To approximate a cut-off of $-2$ Z-scores, the cut-off for low height-for-age is 90% of the median while it is 80% of the median for low weight-for-age and low weight-for-height.

Percent of the median weight-for-age is calculated as:

\[
\text{Percent of median weight-for-age} = \frac{\text{(child's weight measurement)}}{\text{(reference median wt-for-age)}} \times 100
\]

**Alternative classification systems**

The Gomez malnutrition classification system (based on mortality in hospitalized young children in Mexico) is still used in Latin America. Based on the percentage levels of weight-for-age (W/A), it defines three degrees of malnutrition:

- First degree: $\leq 90$–76\% W/A
- Second degree: $\leq 75$–61\% W/A
- Third degree: $\leq 60$\% W/A.

While it has the advantage of being mathematically straightforward, this categorization does not take into account normal distributions of anthropometric measurements in a community, or the selection of an abnormal cut-off based on functional outcomes such as subsequent morbidity or mortality.
Presentation of Anthropometric Data

Summarizing survey results

Box A-5 presents an overview of the basic statistics used to summarize survey results.

While the *prevalence of malnutrition* is the most common way to present anthropometric survey data, the *mean* anthropometric status is preferable because a smaller sample size is needed to show significant change in the mean nutritional status compared with differences in the prevalence.

Once Z-scores for a population have been calculated, it is useful to present them as distribution curves. Illustrated in Figure A-2, the population’s Z-scores are compared to a reference population. Computer software packages (for example, ANTHRO, available from the Centers for Disease Control and WHO and EPI-INFO, from USD, Inc., Stone Mountain, Georgia) are readily available for computation of Z-scores and percentage of the median for individual children and to calculate summary statistics and plot distribution curves.

**Box A-5: Summary Statistics**

- the *prevalence of malnutrition* (the prevalence rate of children whose anthropometric status falls below the –2 Z-scores or < 3rd percentile)
- the *standard error of the prevalence and confidence interval* (a measure of variation in the sample between sampling units or clusters)
- the *mean* (the average anthropometric status—average Z-score or percentage of the median—of all children measured)
- the *standard error of the mean and confidence interval* (measure of variation in the average anthropometric status of sampling units or between each cluster)
Grouping by age

It is optimal to present anthropometric data in age (and sex) groupings for several reasons. First, patterns of growth failure vary with age (see sections on anthropometric indices for greater detail); identifying determinants of malnutrition is facilitated by separating these age groups. Secondly, as discussed in the section on international references, the irregularities in the reference curves for height at 24 months causes the prevalence of wasting (low weight-for-height) to be exaggerated for children between 12 and 24 months relative to older age groups.
WHO recommends the following age stratification:

**minimum**: 0–23, 24+ months

**acceptable**: 0–11, 12–23, 24–59, 60+ months

**best**: 0–5, 6–11, 12–17, 18–23, 24–35, 36–47, 48–59, 60–71 months, etc.

Anthropometric data are considered to be objective—this is mentioned regularly as a distinct advantage. However, as Young and Jaspars (1995) point out, to be truly objective, data must be both reliable and valid, and this may not always be the case.

**Reliability**

A reliable measure produces approximately the same response each time it is used. In the context of anthropometric surveys, standardized measurement techniques, sampling design, and training are all intended to enhance reliability of the data collected. Statistical applications such as confidence intervals describe the degree of reliability and the precision of the results. There is a direct relationship between sample size and the size of the confidence interval: as sample size decreases, the confidence interval widens.

Large sample size does not insure reliability of results either and it is important to be aware of possible sources of bias (see Box A-6) commonly found in nutrition surveys as they can influence results as readily as random measurement errors (which can be removed during the analysis process).

**Validity**

A valid measure provides the correct answer to a question. For example, the height-for-age index does not provide a valid response to a question about wasting, but instead answers the question, “What is the rate of stunting in a population?” Problems of validity surface when anthropometric
### Box A-6: Sources of Bias in Nutrition Surveys and Surveillance

<table>
<thead>
<tr>
<th>Type of Bias</th>
<th>Cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete coverage</td>
<td>Inaccurate or out-of-date sampling frame</td>
</tr>
<tr>
<td></td>
<td>Large-scale population movements</td>
</tr>
<tr>
<td></td>
<td>Sampling subsections of the population (e.g., famine camps, feeding centers)</td>
</tr>
<tr>
<td>Clinic or school bias</td>
<td>Geographical bias towards the more accessible, affluent or urban areas</td>
</tr>
<tr>
<td></td>
<td>Fluctuating attendance</td>
</tr>
<tr>
<td></td>
<td>Selective coverage of the local population (e.g., sick, privileged)</td>
</tr>
<tr>
<td></td>
<td>Varying admission criteria</td>
</tr>
<tr>
<td></td>
<td>Self-selected</td>
</tr>
<tr>
<td>Age bias</td>
<td>Samples of varying age composition (failure to reach 0–3 yr.-old children is common; all nutritional indices vary according to the age structure of the sample)</td>
</tr>
<tr>
<td>Non-random</td>
<td>Systematic errors because of faulty weighing equipment or measurement error</td>
</tr>
<tr>
<td></td>
<td>Incorrect measuring techniques (rounding, truncating)</td>
</tr>
</tbody>
</table>

(Adapted from Young and Jaspars, 1995)

data are used to draw broad conclusions about a situation or they are used as a proxy indicator for other factors such as household food security without looking at other contributing factors.
Confounding variables

Factors external to anthropometric indicators can interact with them to mask or distort the true nutrition situation. For example, under conditions of high mortality in a refugee camp, anthropometric status might appear deceptively stable as malnourished children sicken and die and are replaced by incoming refugee children with the same poor anthropometric status. Age structure of a population can be affected by excess mortality among children under five years, making comparison with a normal population unreliable. Migration of population groups may confound survey results because the nutrition status of remaining groups is often different from those who leave.

Data quality

The quality of either existing survey data or data freshly collected and analyzed can be assessed by asking the following questions (Kostermans, 1994):

- Is the purpose for which the survey was undertaken consistent with what the information is now being used for?

- Are the data representative of the population for whom the proposed program/project is being designed? What is the source of the data—clinic, school, household and is the source properly taken into account when survey findings are interpreted? For example, if anthropometric data are collected from health clinics utilized by only 20% of the population, these data should be used for national planning with reservation. Knowledge of (1) who attends the health clinic or school, and (2) how attendance will bias the survey findings, can be used to better interpret the findings of a limited source survey.

- Is “seasonality” recognized and handled either in data collection or in interpretation? In many agriculture-based societies, food availability
and consumption are closely linked to harvests and season. An anthropometric survey undertaken during the month post-harvest may not represent the nutrition and food situation throughout the year. Knowledge of how seasonality influences food consumption and/or incidence and prevalence of diseases such as diarrhea, malaria, and acute respiratory infections in the given country can be used to interpret data that were not collected in all seasons.

- Are the data reported by sex/gender and age? Weight-for-height data may be disaggregated by height (< 85cm vs. ≥ 85cm). Is there evidence of digital preference (for instance, clumping of ages at 12-month points)?

- Are data valid and reliable? Who were the surveyors, what type of training did they receive and what level of supervision was provided? How were standardization procedures for the measurements handled? (i.e., did all surveyors take measurements and record findings in the same way and how was this monitored?)

- What type of measuring equipment was used? Who was responsible for calibrating the instruments?

- Who was responsible for overall supervision and management of the survey? Was there a detailed operation plan?

- Are the findings consistent with other countries in the region or with countries of similar economic, climatic, demographic characteristics?

- If more than one estimate is available, are they consistent?