Anthropometrics

Introduction

Malnutrition remains a widespread problem in developing countries, in particular among the poorest and most vulnerable segments of the population. Malnutrition is typically caused by a combination of inadequate food intake and infection which impairs the body’s ability to absorb or assimilate food. It is an important cause of low birth weight, brain damage and other birth defects, and contributes to developmental (physical and cognitive) retardation, increased risk of infection and death, and other problems in infants and children.

One approach to studying nutrition is to assess nutritional status on the basis of anthropometric indicators. These are based on physical body measurements such as height or weight (related the age and sex of the individual), and have the benefit of being both inexpensive and non-intrusive to collect. From an anthropometric perspective, nutritional status can be seen as the output of a health production function, where nutrient intake is one input, but where other individual, household, and community variables also feature.

Anthropometric indicators are useful both at an individual and population level. At an individual level, anthropometric indicators can be used to assess compromised health or nutrition well being. This information can be valuable for screening children for interventions and for assessing the response to interventions. At the population level, anthropometry can be used to assess the nutrition status within a country, region, community, or socioeconomic group, and to study both the determinants and consequences of malnutrition. This form of monitoring is valuable both for the design and targeting of health and nutrition interventions.

This note focuses on the construction, interpretation, and use of anthropometric indicators, with an emphasis on the infants and children. The first section provides an overview of anthropometric indicators. The second section discusses practical and conceptual issues in constructing anthropometric indicators from physical measurements, and illustrates these issues with examples based on household data from Mozambique. Finally, the third section highlights some key issues and approaches to analyzing anthropometric data.

On overview of anthropometric indicators

Survey data often contain measures of weight and height, in particular for children. These measures have limited value as indicators of malnutrition in their own right. In part, this is because weight and height depend on both age and gender. Moreover, physical characteristics are affected by many intervening factors other than nutrient intake, in particular genetic variation. However, even in the presence of such natural variation, it is possible to use physical measurements to assess the adequacy of diet and growth, in particular in infants and children. This is done by comparing indicators with the distribution of the same indicator for a “healthy” reference group, and identifying “extreme” or “abnormal” departures from this distribution. For example, the three of the most commonly used anthropometric indicators for infants and children—weight-for-height, height-for-age, and weight-for-age—can be constructed by comparing indicators based on weight, height, age, and gender with reference data for “healthy” children [1, 2].

Weight-for-height (W/H)

W/H measures body weight relative to height, and has the advantage of not requiring age data. Weight-for-height is normally used as an indicator of current nutritional status, and can be useful for screening children.

---

1 In what follows, we do not distinguish between indices and indicators. In principle, however, there are important conceptual differences. An index is simply a combination of different measurements. In contrast, an indicator relates to the use or application of an index to measure (or indicate) a specific phenomenon or outcome. For example, anthropometric indices are typically used as indicators for nutritional status. However, the extent to which the anthropometric index is an appropriate indicator depends on the nature of the relationship between nutrition and the index in question (formally, the sensitivity and specificity of the indicator). For a more detailed discussion, see WHO [2].
at risk and for measuring short-term changes in nutritional status. Low W/H relative to the a child of the same sex and age in a reference population is referred to as “thinness”. Extreme cases of low W/H are commonly referred to as “wasting”. Wasting may be the consequence of starvation or severe disease (in particular diarrhea), but it can also be due to chronic conditions. It is important to note that a lack of evidence of a wasting in a population does not imply the absence of current nutritional problems such as low height-for-age.

**Height-for-age (H/A)**

H/A reflects cumulative linear growth. H/A deficits indicate past or chronic inadequacies nutrition and/or chronic or frequent illness, but cannot measure short-term changes in malnutrition. Low H/A relative to a child of the same sex and age in the reference population are referred to as “shortness”. Extreme cases of low H/A, where shortness is interpreted as pathological, is referred to as “stunting”. H/A is primarily used as a population indicator rather than for individual growth monitoring.

**Weight-for-age (W/A)**

W/A reflects body mass relative to age. W/A is, in effect, a composite measure of height-for-age and weight-for-height, making interpretation difficult. Low W/A relative to a child of the same sex and age in the reference population is referred to as “lightness”, while the term “underweight” is commonly used to refer to severe or pathological deficits in W/A. W/A is commonly used for monitoring growth and to assess changes in the magnitude of malnutrition over time. However, W/A confounds the effects of short- and long-term health and nutrition problems.

As noted, the construction of anthropometric indicators is based on comparisons with a “healthy” reference population. The international reference standard that is most commonly used (and recommended by the WHO) is based on data on the weights and heights of a statistically valid population (US National Center for Health Statistics (NCHS)) of health infants and children in the US. The validity of this reference standard stems from the empirical observation that well-nourished and healthy children will have a very similar distribution of height and weight to the US reference population, regardless of their ethnic background or where they live. In other words, although there are some differences in growth patterns across ethnic groups, the largest part of worldwide variation in anthropometric indicators can be attributed to differences in socioeconomic factors. Notwithstanding this empirical regularity, there is a long-standing debate about the appropriateness of the US reference standard for children in developing countries, in particular concerning the extent to which growth paths will depend on feeding practices. While these are important issues to address, analysts are currently recommended to use the NCHS/WHO reference data.

Anthropometric indices are constructed by comparing relevant measures with those of comparable individuals (in terms of age and sex) in the reference data. There are three ways of expressing these comparisons:

- **Z-score (standard deviation score):** the difference between the value for an individual and the median value of the reference population for the same age or height, divided by the standard deviation of the reference population.

- **Percent of median:** ratio of a measured or observed value in the individual to the median value of the reference data for the same sex and age or height.

- **Percentile:** rank position of an individual on a given reference distribution, stated in terms of what percentage of the group the individual equals or exceeds.

---

2 At the other end of the spectrum, weight-for-height can also be used to construct indicators of obesity.
3 Reference standards are available for children and adolescents up to 16 years old, but are most accurate for children up to the age of 10 years.
4 In part, this debate concerns how growth paths in infants and children can depend on feeding practices (e.g. bottle-feeding vs. breast-feeding). For further discussion of these issues, see WHO [2].
The preferred and most common way of expressing anthropometric indices is in the form of Z-scores. This approach has a number of advantages. Most importantly, Z-scores can be used to construct summary statistics (e.g. mean and standard deviation) for the population or sub-populations. This cannot be meaningfully done with percentiles. Moreover, at the extreme of the distribution, large changes in height or weight are not necessarily reflected in changes in percentile values. Percent of median also has disadvantages. In particular, percentages are not informative about where in the distribution an individual is located, and a given percentage corresponds to different Z-scores depending on the age or height of the individual.

What criteria do we use to determine whether an individual is malnourished or not? The most commonly used cut-off to define abnormal anthropometry with Z-scores is -2 standard deviations, irrespective of the indicator used. For example, a child whose height-for-age Z-score is less than -2 is considered stunted. This provides the basis for estimating prevalence of malnutrition in populations or sub-populations. The WHO has also proposed a classification scheme for population-level malnutrition.

Table: WHO classification scheme for degree of population malnutrition

<table>
<thead>
<tr>
<th>Degree of malnutrition</th>
<th>Prevalence of underweight (% of children &lt;60 months, below -2 Z-scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W/A and H/A</td>
</tr>
<tr>
<td>Low</td>
<td>&lt;10</td>
</tr>
<tr>
<td>Medium</td>
<td>10-19</td>
</tr>
<tr>
<td>High</td>
<td>20-29</td>
</tr>
<tr>
<td>Very high</td>
<td>&gt;=30</td>
</tr>
</tbody>
</table>

While weight-for-height, height-for-age, and weight-for-age are the most commonly used anthropometric indicators for infants and children, they are by no means the only ones that have been used.

Mid-upper arm circumference (MUAC)

MUAC is a measure of the diameter of the upper arm, and gauges both fat reserves and muscle mass. It is primarily used for children, but can also be applied to pregnant women to assess nutritional status. Measurement is simple and requires minimal equipment. MUAC has therefore been proposed as an alternative index of nutritional status, in particular in situations where data on height, weight, and age are difficult to collect. For children, a fixed (age-independent) cut-off point has sometimes been used to determine malnutrition. However, this risks over-diagnosing young children and under-diagnosing older children.

It should be noted that, using this criteria, approximately 2.3 percent of “healthy” children would be classified as having an abnormal deficit in any particular anthropometric indicator.

The WHO has also proposed a more general malnutrition classification, which distinguishes between mild (Z-score < -1), moderate (Z-score < -2), and severe malnutrition (Z-score < -3).

For further discussion of alternative anthropometric indicators, including indicators for adolescents and adults, see WHO [2].
children. Recent reference data, based on US children aged 6 to 60 months, have been incorporated in recent versions of some anthropometric software packages.

**Body mass index (BMI)**

BMI is a measure to define overweight and thinness. BMI is defined as the weight in kilos divided by the square of height in meters. In developing countries, the BMI is primarily used with age-independent cut-offs to identify chronic energy deficiencies (or obesity) in adults. Although there is some scope for using BMI for adolescents, the index varies with age for children and teens, and must therefore be interpreted in relation to BMI-for-age reference charts.

<table>
<thead>
<tr>
<th>BMI range</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;16</td>
<td>Underweight (grade 3 thinness)</td>
</tr>
<tr>
<td>16-16.99</td>
<td>Underweight (grade 2 thinness)</td>
</tr>
<tr>
<td>17-18.49</td>
<td>Underweight (grade 1 thinness)</td>
</tr>
<tr>
<td>18.5-24.99</td>
<td>Normal range</td>
</tr>
<tr>
<td>25.0-29.99</td>
<td>Overweight (pre-obese)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>Obese</td>
</tr>
</tbody>
</table>

Although overweight and obesity is a growing problem in many developing countries, prevalence is likely to be low. For example, in the figure below, a histogram of BMI for adults (over 20 years) in Vietnam illustrates that while nearly 45 percent of individuals are underweight, overweight and obesity is rare (less than 4 percent).

**Constructing and analyzing anthropometric indicators**

The easiest way to construct anthropometric indicators is to use designated anthropometric software, which contains the relevant reference data and has easy procedures for constructing the indicators of interest. This section provides a brief overview of the most popular anthropometric software packages, and a step-by-step guide to using one of these pieces of software to construct key anthropometric indicators.

**Software for anthropometric analysis**

At the simplest level, anthropometric software uses raw measurement data in combination with reference data to calculate the corresponding anthropometric indicators. Many of the available software packages also have more advanced functions, including statistical and graphical analysis. The two most popular software packages for anthropometric analysis are ANTHRO and EPI-INFO. Both can be downloaded for free.
ANTHRO is a DOS-based program which can be used to compare the growth of individual children with the growth patterns of the 1978 NCHS/CDC/WHO growth reference. It requires the sex, height, weight, and age of children to calculate normalized anthropometric z-values, percentiles and percent-of-median. It can use dBase files for batch processing and has an anthropometric calculator.8

EPI-INFO is a series of programs for Microsoft Windows 95, 98, NT, and 2000. It can be used by analysts for general data base and statistical applications (including designing questionnaires and data entry platforms). EPI-INFO contains a special purpose module called NutStat, which can be used to compare data on height, weight, age, sex, and arm circumference with international reference standards for assessment of nutritional status. NutStat can be linked to an EPI-INFO for data entry or analysis or used alone. The program calculates percentiles, number of standard deviations from the mean (z-scores), and, depending on the reference data used, percent of median. NutStat calculations are based on the following reference standards.

- The sex specific 1978 CDC/WHO normalized version of the 1977 NCHS reference curves for height-for-age, weight-for-age, and weight-for-height.
- WHO reference data for arm circumference-for-age and arm circumference-for-height.
- The sex specific 2000 CDC reference curves for head circumference-for-age, length-for-age, weight-for-length, weight-for-age, stature-for-age, weight-for-stature, and body mass index (BMI)-for-age.

EPI-INFO uses the Microsoft Access file format as a database standard, but many other file types can be analyzed, imported, or exported.9

From physical measurement to anthropometric indicators: A step-by-step guide

Many analysts are unfamiliar with anthropometric software packages, and prefer to use a more general statistical application such as STATA or SPSS to manage and analyze data. However, unless the data set contains pre-calculated Z-score values, it is necessary to use dedicated software to construct the variables of interest. Broadly speaking, this entails four steps: (i) setting up the data in STATA or SPSS; (ii) reading and processing the data in the EPI-INFO or ANTHRO; (iii) re-exporting the constructed variables to STATA or SPSS; and, (iv) basic data cleaning. Each of these steps is described in more detail below, with an illustration based on data from a living standards survey from Mozambique.10,11

1. Setting up the data in STATA or SPSS

Both EPI-INFO and ANTHRO performs batch processing to construct anthropometric indicators. However, in order to calculate the desired anthropometric indicators, the appropriate variables must first be loaded. The range of requisite variables depends on what anthropometric indicators are desired, and also to some extent on the reference data that are used in the analysis (see Table). The data do not need to contain all the listed fields. However, if a statistic is dependent on a field that is not available in the data, it cannot be calculated.

---

8 ANTHRO program files and supporting documentation, can be downloaded from http://www.cdc.gov/nccdphp/dnpa/anthro.htm
9 EPI-INFO programs, documentation, and teaching materials can be downloaded from http://www.cdc.gov/epiinfo/index.htm. EPI-INFO is a CDC trademark, but may be freely copied, distributed, or translated.
10 The discussion refers to EPI-INFO, but the procedures are similar for ANTHRO.
11 The 1996/97 Mozambique National Household Survey on Living Conditions (Inquérito Nacional aos Agregados Familiares Sobre as Condições de Vida) was designed and implemented by the National Statistics Institute in Mozambique, and was conducted from February 1996 to April 1997. The sample covers approximately 43,000 individuals in 8,250 households. It was selected in three stages and is geographically stratified to ensure representativeness at both at provincial level and for urban/rural areas.
Table: Variables that can be used in EPI-INFO

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description and required format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1 or M for male 2 or F for female</td>
</tr>
<tr>
<td>Age</td>
<td>Months or years</td>
</tr>
<tr>
<td>Birth date</td>
<td>Not necessary if age data are available</td>
</tr>
<tr>
<td>Date of measurement</td>
<td>Not necessary if age data are available</td>
</tr>
<tr>
<td>ID number</td>
<td>Unique individual identifier</td>
</tr>
<tr>
<td>First name</td>
<td></td>
</tr>
<tr>
<td>Last name</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>In centimeters or inches</td>
</tr>
<tr>
<td>Recumbent</td>
<td>Boolean or 1/2; whether child is lying down</td>
</tr>
<tr>
<td>Weight</td>
<td>In kilos or pounds</td>
</tr>
<tr>
<td>Edema(^{1,2})</td>
<td>Boolean or 1/2; excessive amounts of fluid</td>
</tr>
<tr>
<td>Arm circumference</td>
<td>In centimeters or inches</td>
</tr>
<tr>
<td>Head circumference</td>
<td>In centimeters or inches</td>
</tr>
</tbody>
</table>

Notes

There are a couple of important points to note in relation to the variables used in EPI-INFO. First, for many anthropometric indicators, age-specific reference data are used. When the data permits it, it is always preferable to calculate age as the difference between date of measurement and date of birth. This is almost always more accurate than age reported by survey respondents. The reference curves are based on "biologic" age rather than calendar age. Biologic age in months divides the year into 12 equal segments as opposed to calendar age in which months have from 28 to 31 days. Although this makes little difference in older children, it can have an effect on the anthropometric calculations for infants. Moreover, entering age by rounding to the nearest month and/or the most recently attained month can have a substantial effect on the anthropometric calculations, especially for infants. To calculate biologic age, anthropometric software calculates the number of days between the two dates. The age in days is divided by 365.25 and then multiplied by 12. This calculation can also easily be performed in STATA or SPSS.

Second, for younger children, height is normally measured with the child in a recumbent position. This measurement is sometimes referred to as “length”, which is contrasted with standing height measurements, referred to as “stature”.\(^{13}\) In some cases, this distinction is not important from the point of view of the analyst. For example, when using the recommended 1978 CDC/WHO reference in EPI-INFO, recumbent length is assumed from birth to age 24 months, and standing height 24 months and older. However, in EPI-INFO it is also possible to use a 2000 CDC reference standard. If this option is chosen, the user must indicate if the measurements are recumbent length or standing height for children in the age group of 24 to 36 months (for younger and older children, recumbent and standing measurements are assumed respectively).

Turning to the Mozambique illustration, the first step is to construct a data set with all the relevant variables in the appropriate format. In this case, we have data on the birth date, measurement date, sex, weight, and height of children under 5 years old. In the following code, we use the dates to construct the age in months, and adjust weight to account for the fact that a minority of children were weighed with their clothes on, before saving the relevant variables in a new data file. At this stage it is also important to ensure that all the variables are appropriately coded, and in the correct units. For example, height is sometimes entered in millimeters. If this is not converted to centimeters at this stage, the subsequent calculations of anthropometric indicators will be incorrect. Note that because we will be using the 1978 CDC/WHO reference data, it is not necessary to include variables on how the child was measured.

```stata
generate birthdate = mdy(birth_mnth, birth_day, birth_yr)
```

\(^{1,2}\) Edema refers to the presence of excessive amounts of fluid in the intracellular tissue. There is a strong association between edema and mortality. The presence of edema is therefore important for screening and surveillance purposes, and can be used to flag children as severely malnourished, independently of their wasting, stunting, or underweight status. However, although edema is included as a variable in many anthropological software for this reason, it is not generally used for evaluation purposes.

\(^{13}\) For details on measurement, see WHO [2] or http://www.cdc.gov/nccdphp/dnpa/bmi/meas-height.htm.
generate measuredate = mdy(visit_mnth, visit_day, visit_year)
generate age_days = measuredate - birthdate
generate age_months = (age_days/365.25) * 12

replace weight_grams = weight_grams - weight_cloth if how_weighed==1
generate weight_kilos = weight_grams / 1000

keep id sexo age_months weight_kilos height_cm
drop if sexo==. | age_months==. | weight_kilos==. | height_cm==.
save C:\anthro_data.dta, replace

2. Reading and processing the data in EPI-INFO / NUTSTAT

*NutStat* uses data in Access data base format. There are two ways of importing external data into *NutStat*. The “Add Statistics” feature simply processes the data from an Microsoft Access data file and adds the results of calculations to the file. In contrast, the “Import Data” feature can be used to import data from an existing table into a new table that has the data structure that is required by *Nutstat*. The table can be an EPI-INFO table or a table from Microsoft Access. Both the “Add Statistics” and “Import Data” features can be accessed from the FILE menu. Under either alternative, the user is first required to choose between the two alternative reference standards (1978 CDC/WHO or 2000 CDC). The user must thereafter link variables in the imported data file with fields that EPI-INFO requires to calculate the anthropometric indicators, and, for some fields, select the unit of measurement. Finally, the user selects the statistics or indicators to calculate—e.g. Z-scores, percentiles, and percent of median for weight-for-age, height-for-age, and weight-for-height. The selection of indicators to be calculated is restricted by the variables that were imported. When the data are then processed, the new variables are either added to the original file (“Add Statistics” option), or saved in a new data file (“Import” option).

<table>
<thead>
<tr>
<th>Anthropometric indicator</th>
<th>Variable names 1978 CDC/WHO reference</th>
<th>Variable names 2000 CDC reference *</th>
</tr>
</thead>
<tbody>
<tr>
<td>H/A Z-score</td>
<td>fldWHOHAZ</td>
<td>fldCDCLAZ, fldCDCSAZ</td>
</tr>
<tr>
<td>H/A Centile</td>
<td>fldWHOHAC</td>
<td>fldCDCLAC, fldCDCSAC</td>
</tr>
<tr>
<td>H/A Percent of median</td>
<td>fldWHOHAPM</td>
<td></td>
</tr>
<tr>
<td>W/A Z-score</td>
<td>fldWHOWAZ</td>
<td>fldCDCWAZ</td>
</tr>
<tr>
<td>W/A Centile</td>
<td>fldWHOWAC</td>
<td>fldCDCWAC</td>
</tr>
<tr>
<td>W/A Percent of median</td>
<td>fldWHOWAPM</td>
<td></td>
</tr>
<tr>
<td>W/H Z-score</td>
<td>fldWHOWHZ</td>
<td>fldCDCLAZ, fldCDCSAZ</td>
</tr>
<tr>
<td>W/H Centile</td>
<td>fldWHOWHC</td>
<td>fldCDCLAC, fldCDCSAC</td>
</tr>
<tr>
<td>W/H Percent of median</td>
<td>fldWOHWHPM</td>
<td></td>
</tr>
<tr>
<td>MUAC-for-age</td>
<td>MUACAZ**</td>
<td>MUACAZ</td>
</tr>
</tbody>
</table>

* If stature and length are processed separately, two variables are created for each indicator

** Based on 2000 CDC reference

14 STATA or SPSS data can be converted into Access format using conversion software such as DMBScopy or StatTransfer. Alternatively, data exported from STATA using the outsheet command, which can then be read into Access. It is also possible to copy data from the STATA browser and paste it in Access.

15 Further anthropometric indicators can be calculated on the basis of specific WHO reference data (MUAC-for-age) or 2000 CDC reference (head circumference-for-age and BMI-for-age).
3. Exporting the data for analysis in STATA or SPSS

After the variables of interest have been constructed, the Microsoft Access database must be converted into a format that can be read by STATA or SPSS before the variables can be merged with the original data. In our example, Z-scores, percentiles, and percent of median for weight-for-age, height-for-age, and weight-for-height were constructed in the file `anthro_data.dta`, which can now be merged with the main data.

```
use "C:\anthro_results_stata.dta", clear
sort id
save, replace

use "C:\main_data.dta", clear
sort id
merge id using C:\anthro_data.dta
save, replace
```

4. Basic data cleaning

In the previous steps, we have added a set of anthropometric indicators to our original data. Before we proceed with the analysis of these indicators, there are a number of data-cleaning issues and potential sources of bias that analysts must be aware of.

The first issue to contend with concerns the problem of missing values. In most surveys, enumerators are not able to obtain all the relevant data for all sampled individuals. The most common problem concerns the age of the child, where the precise birth date may not be known by the parent, and birth records are not available. Data may also be missing for weight or height, for example because some parents do not agree for their children to be weighed, or there may have been problems with the measuring equipment.

A second problem concerns the calculated Z-scores, where errors in measurement, reporting of age, coding, or data entry sometimes result in biologically implausible values. WHO [2] recommends that, for the purpose of analysis, values outside a certain range should be treated as missing values (see Table).

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Exclusion range for Z-scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height-for-age</td>
<td>&lt;5.0 and &gt;+3.0</td>
</tr>
<tr>
<td>Weight-for-height</td>
<td>&lt;4.0 and &gt;+5.0</td>
</tr>
<tr>
<td>Weight-for-age</td>
<td>&lt;5.0 and &gt;+5.0</td>
</tr>
</tbody>
</table>

Table: Exclusion ranges for “implausible” Z-scores

*Note: If observed mean Z-score is below –1.5, WHO recommends that a flexible exclusion range is used. For details, see WHO [2].*

These problems can be explored by looking at descriptive statistics, scatterplots, and histograms. In the Mozambican data, we construct indicator variables for missing and implausible values in the variables of interest.

```
generate missing = (age_months==. | haz==. | whz==. | waz==.)
generate outrange_haz = (haz < -6 | haz > 3)
replace outrange_haz = . if haz==.
generate outrange_whz = (whz < -4 | whz > 5)
replace outrange_whz = . if whz==.
```
generate outrange_waz = (waz < -5 | waz > 5)
replace outrange_waz = . if waz==.

drop if missing==1 | outrange_haz==1 | outrange_whz==1 | outrange_waz==1

In this case, nearly 20 percent of observations have missing values for age, weight, or height. In addition, approximately 3 percent of observations have values for one or more of the anthropometric indicators that are outside the plausible range. This is quite high, and casts some doubt on the quality of the data. At this point, it is difficult to correct the problems that give rise to missing and implausible values. However, before dropping these observations, it is important to assess whether doing so is likely to result in biases in subsequent estimates. This depends both on what proportion of the sample that is being dropped, and on the extent to which the values for the variables of interest are systematically different from the rest of the sample. Of course, with missing data values for the relevant variables, we cannot provide a precise answer to this question. That said, we can look at other characteristics of the observations that we are dropping, and draw some conclusions on that basis. For example, in the Mozambique data, both average income and maternal education is significantly lower for the observations that will be dropped, and they are disproportionately from rural areas. Hence, we are dropping observations that most likely have a poorer nutritional status than the population average, and our population estimates of malnutrition (e.g. prevalence or mean Z-score) is likely to be biased to make the situation look better than it actually is.

Even after this basic data cleaning, there are many possible sources of bias that analysts need to be aware of. First, sampling bias, whereby the sample that we are analyzing is not representative of the population that we are interested in, can sometimes be important for anthropometric data. Sampling bias may be the consequence of incomplete survey coverage due to inaccurate or inappropriate sampling frame, or because the data were collected through schools or clinics which are not attended by all segments of the population. Non-sampling bias refers to systematic errors arising from the survey implementation and data management process. This includes problems of missing values, and reporting and measurement errors discussed above, but can also be due to other factors such as seasonality. In most cases, these problems cannot be corrected by the analyst. However, it is important for analysts to be aware of the and potential biases they can give rise to, and to qualify any findings accordingly.

Analyzing anthropometric data

At the most basic level, the analysis of anthropometric concerns the identification of malnourishment in a population or sub-population. However, in many cases analysts want to go beyond merely establishing prevalence to try to understand the causes of malnourishment and how can they be addressed.

The first step in the analysis is to look at the distribution of the Z-scores and overall prevalence. This can usefully be done both graphically and with descriptive statistics. When compared with the distribution of Z-scores in the reference population, this provides a first impression of different dimensions of nutritional status in the population.

graph haz, bin(50) xlabel (-6 -4 to 6) ylabel (0 0.05 to 0.10) normal(0,1)
graph whz, bin(50) xlabel (-6 -4 to 6) ylabel (0 0.05 to 0.10) normal(0,1)
graph haz, bin(50) xlabel (-6 -4 to 6) ylabel (0 0.05 to 0.10) normal(0,1)

tabstat haz waz whz [aw=weight], stat(mean sd)
tabstat below2_haz below2_waz below2_whz [aw=weight], stat(mean)
tabstat below3_haz below3_waz below3_whz [aw=weight], stat(mean)

The results of this analysis is reported below. As can be seen from both the graphs and the table, there are deficits in both H/A and W/A, but only limited evidence of wasting.
It is also useful to look graphically at the relationship between different anthropometric indicators. In general, weight-for-height and height-for-age are not correlated, while there tends to be a positive correlation between weight-for-height and weight-for-age, and between weight-for-age and height-for-age. This pattern is confirmed in the Mozambican data.

For many purposes, anthropometric data should be presented according age and sex groups. This is because (i) patterns of growth failure vary with age, (ii) it facilitates the identification of determinants of malnutrition; (iii) as a consequence of irregularities in the reference curves, wasting tends to be exaggerated for children in the 12-24 month age group. WHO [2] recommends that at least two age disaggregations be used—under 24 months and 24 months and over—but for some purposes a more detailed disaggregation may be advisable.
It is also useful to explore how anthropometric indicators are related to different individual, household, and community characteristics. An important perspective in this regard is to assess socioeconomic inequalities in nutritional status. A simple way of looking at this issue is to look at the prevalence of malnutrition across income or wealth quintiles.

Prevalence rates of stunting, underweight, and wasting for different consumption quintiles in Mozambique are graphed below, with a disaggregation by sex for stunting. The graphs show the expected tendency for malnutrition to be worse in poorer quintiles. Although important, analysts must also avoid drawing unwarranted conclusions about socioeconomic inequalities. In particular, it is important to look beyond the simple means by quintiles to assess whether the observed differences are significant. For example, in the case of stunting, a more detailed analysis of the Mozambique data reveals that the only statistically significant difference in prevalence between consumption quintiles is between the richest quintile and the rest. Moreover, although comparing means by quintiles is a useful place to start, it is quite a crude approach to assessing socioeconomic inequalities, and analysts should consider using concentration curves and indices discussed in other technical notes.

It is important that descriptive analysis of anthropometric data be accompanied by information to assist in the interpretation of findings. This includes information such as general characteristics of the population, sample design and size, method of determining age, proportion of data missing or excluded. For some purposes it is also important to report standard errors or confidence intervals for estimates.

So far, the discussion of the analysis of anthropometric data has been concerned with descriptive statistics. Analysts may however wish to pursue more analytical questions. In this regard, it is customary to distinguish between distant and proximate determinants of malnutrition [3]. Proximal factors are inadequate dietary intake and disease. Distant factors do not influence malnutrition directly, but through their impact on proximate determinants. They include, for example, poverty, education, cultural factors (e.g. duration of breast feeding, hygiene practices), and community and environmental characteristics (e.g. availability and quality of health services, epidemiological profile). The identification of socioeconomic inequalities should therefore only be

---

16 In some context, the distinction between classifying and determining variables is used, where determining variables is used to refer to distal or proximate determinants that are amenable to be changed through interventions (e.g. feeding practices, sanitation, immunization, etc.). UNICEF has also proposed a conceptual framework on the causes of malnutrition, which distinguishes between immediate, underlying, and basic causes of malnutrition.
considered the starting point for further analysis. A socioeconomic gradient most likely confounds the influence of many complex factors. In order to understand the contribution of different factors in determining nutritional status, a multivariate modeling framework is required. This type of analysis also has to contend with difficult conceptual and empirical issues that are beyond the scope of this note.  

Useful sources of further information

WHO Global database on child growth and malnutrition: http://www.who.int/nutgrowthdb/
Practical Analysis of Nutritional Data (PANDA): http://www.tulane.edu/~panda2/

References


17 For a detailed discussion, see Behrman and Deolalikar [4] and Strauss and Thomas [5]