

Module 2 - Definitions of survey sampling terminology¹

- **Accuracy** – how closely a measurement or observation comes to measuring the “true” value since measurements and observations are always subject to error.
- **Clusters** – Another term for the Primary Sampling Units (PSUs) selected from the sample frame to be included in the survey sample is the selected cluster. Clustering generally refers to the clustering of households within selected PSUs and/or to individuals within these households.
- **Element** – is the basic unit that represents whatever is being sampled and from which survey data are to be gathered. For health-related surveys, this is frequently the individual (e.g., what is the prevalence of anemia among pregnant women?) or the household (e.g., what proportion of households use iodized salt?).
- **Enumeration areas (EA)** – the operational geographic units for the collection of census data, also referred to as enumeration units, census tracts, or wards. These are often used to provide the sample frame and may be listed in such a way as to allow for implicit stratification.
- **Explicit Stratification** – stratification is where a geographic area is divided into mutually exclusive and exhaustive areas and a separate survey undertaken in each stratum. Explicit stratification takes place at the design stage of the survey. The main purpose of stratification is to improve the precision of survey estimates through the selection of strata that are as homogeneous as possible in terms of the micronutrient status and/or intervention of interest within each stratum, while being as heterogeneous as possible between strata.
- **Frame (sampling frame)** – A listing of Primary Sampling Units (PSUs), or of elements within a PSU, from which the sample is drawn.
- **Implicit Stratification** – a technique that can be applied prior to systematic selection of Primary Sampling Units (PSUs) from within an explicit stratum. PSUs can be sorted with respect to a variable of interest that is available from the census (or other) information being used to generate the listing. Usually the list is ordered based on geographical proximity, or by urban and rural location where these residential types occur in the same stratum, and then by geographic location within urban and rural areas. This ensures that the selected PSUs are spread across sub-groups of the population.
- **Population** – for health-related surveys, this is generally the population to which we would like to make an inference; sometimes this is referred to as the *target population*. Examples of survey populations include: households; children 12-23.9 month of age; women of childbearing age; and school-age children.
- **Precision** (reliability of the survey estimate) – how closely a repeated measurement or observation comes to duplicating the measured or observed value. The standard error is a measure of precision.
- **Primary Sampling Units (PSU)** – the first stage sampling unit in a cluster sample, usually randomly selected from a listing of PSUs using PPS sampling. The PSUs may have the same boundaries as the enumeration areas (EA); or may be some other area with clearly defined, non-overlapping geographic boundaries, for example villages in rural areas and wards or blocks in urban areas. The PSUs may also be referred to as a “cluster” since the survey elements, e.g. households, are clustered within the PSU.
- **Probability Proportional to Size (PPS)** – is the usual method used for selecting the sample of Primary Sampling Units (PSUs). The size of the PSU determines its probability of selection (with size often based on the number of households to equate to population size). Where the same number of households (or

¹ Some of these terms are based at least in part on definitions sourced from:

https://unstats.un.org/unsd/HHsurveys/pdf/Chapter_2.pdf and <http://methods.sagepub.com/Reference/encyclopedia-of-survey-research-methods>

other survey element) are randomly selected using the same methodology within each PSU, this can result in a self-weighted sample of households, i.e. all households in the stratum have the same probability of selection regardless of which PSU they are located in.

- **Sample** – the subset of elements drawn from the frame (the survey sample)
- **Sampling Units** – these are another term used for the Primary Sampling Units (PSUs).
- **Sampling weights** – complex sample survey data usually needs to be adjusted to account for unequal selection probabilities (of the Primary Sampling Unit (PSU), the household and of individuals in the household, as relevant), for nonresponse, and for known differences between the sample and the reference population. Sample weights are applied to adjust all point estimates (and to also adjust the estimate of variability around them) so that each sampled element can be considered to represent the same number of elements in the reference population.
- **Segmentation** – large Primary Sampling Units (PSUs) are sometimes partitioned into a number of reasonably sized sub-units, to make them more equal in size to other PSUs and so that survey elements within the PSU can be listed in a reasonable manner. One of these sub-units or segments is randomly selected and a listing of elements within it is determined to function as the sample frame for that PSU. The selected segment is sometimes referred to as a secondary sampling unit. This additional stage of sampling must be accounted for in the data weighting process.