

Section 1: Glossary of Terms Used in STEPS

Introduction This section provides an alphabetical list of all the terms used in a STEPS surveillance with definitions that are appropriate for STEPS.

Term	Definition
Age-standardisation	A process of statistically adjusting rates or prevalence or mean values from two or more populations with different age structures in order to facilitate comparisons or understand differences between the populations.
Archive	A depository containing records or documents.
Average	See Mean
Bias	Distortion of a population estimate away from the true value. Bias can arise for many reasons such as measurement error or non-response.
Cluster	A (usually geographical defined) group of individuals.
Cluster sampling	A sampling method where the target population is divided into clusters/groups and a subset of each cluster is selected instead of the entire cluster. Cluster sampling often uses enumeration areas for the primary cluster
Confidence interval (CI)	A range of values around the sample estimate in which the true population value is likely to fall. For example, a 95% confidence interval indicates that for 95 out of 100 surveys, the population mean would fall into this range of values around the sample mean.
Cross-sectional design	A study design based on observations at a single point in time. STEPS surveys will be cross-sectional unless they are especially being extended to follow the sample over time.
Database	A large amount of information stored in a file that is easily searched by a computer. STEPS mostly uses Microsoft Access.
Dataset	An electronic file consisting of a table in which each row contains data for one individual and each column represents one variable.
Demographic characteristics	The characteristics of a population, for example, age, sex, ethnicity and place of residence.
Distribution	The complete summary of the frequencies of the values or categories of a measurement made on a group of persons. The distribution tells either how many or what proportion of the group was found to have each value (or each range of values) out of all the possible values that the quantitative measure can have.
Enumeration Area	A small to medium sized geographic area that has been defined in a census.
EpiData	A freely available software package designed to facilitate data entry of survey data. Functions include immediate checking of ranges and legal values and ability to export data to a range of analysis packages.
Epi Info	A freely available statistical software package providing basic statistical functions and capable of handling complex sample designs.
Estimate	A calculated guess of the true value of a population characteristic deriving from data obtained from a sample of the population.
Household composition	The age and sex of all the residents in the household who are within the age range of the survey.

Term	Definition
Instrument	This refers to the STEPS Instrument which includes a questionnaire (Step 1), physical measurements (Step 2), and biochemical measurements (Step 3).
Inter-quartile range	The difference between the upper and lower quartiles (25 th and 75 th percentiles) in a set of values. They separate the lowest 25% and highest 75% of values, respectively, in the set of measurements
Mean	The arithmetic mean is the average of a set of values, that is, the sum of all the values divided by number of values. Because of its simplicity and its statistical properties, it is used more than any of the other measures of central tendency (e.g. median).
Measurement device	A tool used for measurement purposes, for example a blood pressure monitor.
Median	The median is a measure of central tendency that is often used for non-normally distributed variables. It is the simplest division of a set of sorted measurements into two halves - the lower and the upper half.
MET	Metabolic equivalent (MET) is the ratio of a person's working metabolic rate relative to the resting metabolic rate. One MET is defined as the energy cost of sitting quietly, and is equivalent to a caloric consumption of 1 kcal/kg/hour.
Moderate intensity physical activity	Refers to activities which take moderate physical effort and that make you breathe somewhat harder than normal. Examples include cleaning, vacuuming, polishing, gardening, cycling at a regular pace or horse-riding. Moderate intensity activities require an energy expenditure of approximately 3-6 METs.
Multi-stage sampling	Multi-stage indicates that sampling is done in several steps. First larger sampling units are selected then smaller sampling units are selected within the selected larger units.
Non-probability	Methods of sampling a population in which the probability of selection of each every individual is not known, and therefore from which reliable population estimates are not calculable. A non-probability sample is not desirable for STEPS.
Non-response	In a sample survey, the failure, for any reason, to obtain information from a designated participant.
Non-response bias	Also known as coverage bias, the error introduced by non-response.
Outlier	An observation differing so widely from the rest of the data as to lead one to suspect that a gross error may have been committed or suggesting that this value comes from a different population.
Participant	An individual who responds to the STEPS Instrument.
Pilot test	A small trial run or "dress rehearsal" of an entire process, e.g. data collection or data entry, completed before the process officially begins.
Post-stratification	A means of making sample estimates more representative of the target population after data have been collected. For STEPS surveys, it is recommended to do a post-stratification for age and sex so that differences in the age-sex distribution between the sample and the target population can be accounted for.

Term	Definition
Precision	The quality of the estimate obtained from the STEPS survey. The standard error of the estimates can be taken as an indicator of the precision of the estimates with a smaller standard error indicating greater precision. See standard error.
Prevalence	The number of persons with a disease or an attribute in a given population at a designated time, e.g. % daily smoker in a country in 2008.
Primary sampling unit (PSU)	The sampling units for the first stage of sampling in a multi-stage sample design. See multi-stage sample design.
Probability	A number between 0 and 1 which represents how likely some event is to occur. A probability of 0 means an event will never occur, while a probability of 1 means the event will always occur.
Probability sample	A sample of a population (or sub-population) that has the property that each individual has an equal and known chance of being selected, and in which the chance of one item being selected does not alter or affect the selection of any other individual. Examples of probability sampling include simple random sample, cluster sampling and stratified sampling.
Probability proportional to size (PPS)	Probability proportional to size (PPS) sampling is a method for selecting a sampling unit in which the probability of selection for a given sampling unit is proportional to its size (most often the number of individuals or households within the sampling unit).
Range	The difference between the largest and the smallest in a set of values, for example in a sample in which height was measured from 135 cm to 180 cm, the range would be 45 cm.
Rank	The position of a member within a sorted set.
Rate	The occurrence of an event over a defined time amongst a defined sample or population. It may be expressed as number of events per person-years, for example 310 injury accidents per 10,000 person-years, which may be imagined as 310 of 1000 people over 10 years, or 310 of 2000 people over 5 years.
Representativeness	The extent to which a sample has the same distribution of the characteristics of interest as the target population from which it was selected.
Response proportion	The proportion or percentage of the eligible individuals sampled who did participate.
Risk Factor	Refers to any attribute, characteristic, or exposure of an individual, which increases the likelihood of developing a disease, or other unwanted condition/event.
Sample	The subset of the target population that is selected for inclusion in the survey.
Sample design	The methodology used to select the part of the population to be included in the survey. See probability sample and non-probability sample.
Sample population	The sample population is the group of individuals who have been selected from the target population (see target population) to participate in the survey.
Sample size	Sample size is the number of people selected for the sample. It should be calculated prior to conducting the survey.

Term	Definition
Sampling error	Sampling errors arise from estimating a population characteristic by looking at only one portion of the population rather than the entire population. It refers to the difference between the estimate derived from a sample survey and the 'true' value that would result if a census of the whole population were taken under the same conditions.
Sampling frame	A list of the units in the target population, for example an electoral roll, a population register, or a telephone book. For the sample to be representative of the target population, the sampling frame should include all people in the population (or sub-population) only once, will not include people who do not belong to that population, and will be up-to-date.
Sampling unit	The objects being selected for a survey. These units must cover the whole of the population and not overlap, i.e. every element in the population belongs to one, and one only, unit. In a simple random sample, the sampling units are the individuals themselves. In cluster sampling, it may be villages or other localities. In multi-stage sampling, the sampling units differ at each level of sampling.
Sampling weight	Sampling weights are weights that denote the inverse of the probability of selection.
Secondary sampling units (SSU)	The sampling units used for selection after the primary sampling units.
Serving (of fruit or vegetable)	For vegetables this refers to one cup of raw, leafy green vegetables, (spinach, salad etc.), one half cup of other vegetables, cooked or raw (tomatoes, pumpkin, beans etc.), or a half cup of vegetable juice. For fruits, this refers to one medium-sized piece of fruit (banana, apple, kiwi etc.) or a half cup of raw, cooked or canned fruit or a half cup of juice from a fruit (not artificially flavored).
Simple random sampling (SRS)	A probabilistic sampling method with only one stage of selection in which every member of the population has an equal chance.
Skew	A distribution of values that is asymmetric and therefore non-normal. Because many of the formulae for estimation are based on assumptions about normal distributions, skewness can seriously distort population estimates, and there must be a strategy for checking and coping with skewed data.
Standard deviation (SD)	A measure of dispersion, or variation. It is equal to the positive square root of the variance. It is a summary of how widely dispersed the values are around the mean.
Standard drink	The net alcohol content of a standard drink is generally 10g of ethanol depending on the country. This is the equivalent of 1 regular beer (285ml), a single measure of spirits (30 ml), a medium-sized glass of wine (120 ml), or a measure of aperitif (60 ml).
Standard error (SE)	A standard error is the standard deviation of an estimate, e.g. a mean. It can be used to calculate confidence intervals.
Strata	The plural form of stratum.

Term	Definition
Stratification	Process of dividing the sampling frame into mutually exclusive subgroups or strata. The sample is then drawn either proportionately or disproportionately from all strata.
Stratum	A partition of the population used in stratified sampling.
Systematic error	Systematic (one-sided) variation of measurements from the true values, leading to a biased estimate.
Systematic sampling	A probability sample selection method in which the sample is obtained by selecting every kth unit of the population, where k is an integer greater than 1. For example if k is 15 and the first unit is number 13, then subsequent units are 28, 43, 58 and so on. The first member of the sample must be selected randomly from within the first k units (a random start). If the target sample size is reached before all the kth members have been surveyed, recruitment must continue until all those selected have been surveyed.
Target population	The population from which the sample population is drawn. If the sample has been drawn correctly, the estimates obtained from the survey should be representative of the target population.
Variable	One item of information stored in a dataset, for example age or sex. Variables may be categorical or continuous, but should be clearly defined and consistently recorded.
Variance	A measure of the variation shown by a set of observations. The standard deviation is calculated by taking the square root of the variance.
Vigorous intensity activity	Refers to activities which take hard physical effort and which make you breathe much harder than normal. Examples include loading furniture, digging, playing football, tennis or fast swimming. Vigorous activities require an energy expenditure of greater than 6 METs.

