Calculating weight for the analysis of a TB patient cost survey

1. Non-stratified sample
   1. Base weight for a non-stratified sample

Weight is defined as an inverse of sampling probability.

For observations selected by non-stratified, single-stage cluster sampling method, a sampling probability of an individual observation is calculated as follows.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Selection of  a PSU by PPS |  | Selection of a fixed number (Nc) of patients |  | Conditional probability for a specific observation |  |  |
| Sampling probability |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| Base weight |  |  |  |  |  |  |  |  |

As shown above, if the PPS procedure is followed as intended, weight is not needed because the sampling probability is equal across all observations regardless of the size of PSU.

* 1. Base weight for a stratified sample

In a stratified sampling design, by definition, each stratum should represent a mutually exclusive different population group. This means sampling probability of each observation in different strata are not usually equal. By using above equation, weight values assigned to each stratum is equal to the total population size for each stratum. For example, in a survey with two strata for drug-susceptible and drug-resistant TB patients, weight values can be derived from the numbers of all notified TB cases in the country disaggregated by drug-resistant status. Using this approach, one can calculate overall estimates for all TB patients in the country, instead of estimates within the survey sample. A simple average, or weighted average within the survey sample will provide different results. Investigators need to be well aware of the differences and able to interpret estimates appropriately.

* 1. Adjustment for under- or over- enrolment

In many instances, actual numbers of enrolled patients might depart from originally assigned cluster size. In such cases, it is necessary to adjust the summary estimates by providing weight. In clusters with a smaller number of patients than the originally intended cluster size, a weight should be given so that they can represent more than their actual size. It is vise-versa for over-enrolled clusters. The weight is calculated as follows.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Number enrolled as per protocol |  | Actual number enrolled in a cluster |  |  |
| Weight |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

For example, if a cluster only enrolled 20 patients, failing to enroll 25 patients as per protocol, the weight value will be 1.25. This mean that the individual observations in this cluster will have 1.25 times of weight so that 20 patients can represent effectively 25 patients.