

2018 Sierra Leone Integrated Household Survey

Survey Documentation

Guidelines for Using the Data

The Dataset Organization section below explains how different sections of the questionnaire are mapped to different datasets. All datasets contain the variables *_cluster* and *_hhno* which can be used to merge them (except **slihs2018_cluster.dta** which is cluster-level data and has just *_cluster*).

The data entry system imposed tight controls on what values could be entered, and flagged additional possible inconsistencies, resulting in fairly clean data coming in. A light approach to data cleaning was thus taken. For the most part, the only replacement of outliers was done by replacing values more than twice the 99th percentile with twice the 99th percentile. Note that this was not done when two questions (say amount paid and period for which it was paid) need to be combined to give a total value. The data from the food diaries (Section X) required more extensive cleaning, and there are still issues, particularly when giving the unit cost for items from own production or received for free. When using this data in particular, check for outliers and mistakes.

For some questions, common "other" responses were identified and codes added. These are (list possibly incomplete): b2, b24, b25, d4a, d4b, d4c, d13a, d13b, d26a, f22, o8, o14, p4

Dates are stored in Stata elapsed data format, months are stored in Stata elapsed month format.

There are some known issues with the data:

- Question b9f (class attended Sept 2018 to Dec 2018). This question should have been answered in households interviewed during the last 4 months of fieldwork. Somehow, this question was not consistently answered and entered for those 4 months, the data seems to reflect a large drop in gross enrollment which is not reflected by any other source (including following up with these households by phone later).
- Questions e25 and e27 (child height and weight). This was clearly not done correctly by Teams 1, 3, 15, 16 and 18.
- Question r25 (measurement of plot in acres). Teams 10 and 17 clearly did not do this correctly. Teams 1 and 2 (measuring just a few likely very small plots in the capital city) also likely have errors.

(note that the variable *_team* can be merged in from **slihs2018_cluster.dta** using the variable *_cluster* to merge.)