CBPS HFT '20-21: Using the data

The datasets are available in Stata format in zipped folders by round. As the survey collected information at household and individual levels across different modules, the datasets have been provided by module for ease of use.

The list of datasets available are:

Dataset name	Level of observation	Topics included
cbps_tracking_r1_basic needs.dta	Household	Access to basic needs, assistance, access to health services
cbps_tracking_r1_labor.dta	Individual adult	Labor force participation, employment status, job search, earnings, work hour
cbps_tracking_r2_basic needs.dta	Household	Access to basic needs, assistance, coping, access to health services
cbps_tracking_r2_education.dta	Individual child	Education engagement, reasons for not studying, plans on future education
cbps_tracking_r2_labor.dta	Individual adult	Labor force participation, employment status, job search, earnings, work hour
cbps_tracking_r3_basic needs.dta	Household	Access to basic needs, assistance, coping, household vaccination
cbps_tracking_r3_education.dta	Individual child	Education engagement, reasons for not studying, plans on future education
cbps_tracking_r3_labor.dta	Individual adult	Labor force participation, employment status, job search, earnings, work hour

Notes on using the data:

➢ The questionnaires for each round are available along with the dataset. The modules in the questionnaire will appear in order of labor, basic needs, and education. All modules are available in the round-wise questionnaires.

➢ The stratum and settlement_type variables can be used to produce representative statistics. The sample does not allow for representative findings below these levels. The stratum variable identifies the three strata in the sample: High exposure, low exposure and camps. The settlement_type variable identifies camps versus hosts.

➢ Household level questions were asked to any adult member who was reachable. However, the labor module was only conducted if one of the two or three selected CBPS adults from the baseline were reached. This means that there may be different labor interviews with different adults across the rounds and labor data should always be merged using the uid1 and id variables, not at household level. For example:
For the basic needs household modules:
  ○ merge 1:1 uid1 using "$baseline_hh_consumption"

For the labor adult modules:

- o   merge 1:1 uid1 id using "$baseline_labor"
- o   merge 1:1 uid1 id using "$cbps_tracking_r2_labor.dta"

Note that even though the labor module datasets are unique by uid1, merging with other labor, roster or adult level datasets must always be done using uid1 and id.

➢   The education module questions were asked for children aged 3-18 years at baseline. The dataset for each round is provided in long format at child-level. These datasets are mergeable with baseline education information collected in the roster using uid1 and id variables. For example:
- o   merge 1:1 uid1 id using "$baseline_roster", keepusing(ls02_age)

➢   The two types of baseline weights – household weight and adult weight – have been readjusted for each tracking round based on the response rate. The respective tracking round weights are available on each dataset. Labor datasets contain aa_weight_r# and the basic needs datasets contain hh_weight_r#. The education datasets do not contain weights variable as individual roster member weights were not created during the baseline. It is expected that education outcomes will be analyzed at the household level using the hh_weight_r# weights.

➢   The household level datasets contain some basic baseline household characteristics which may be useful for analysis. Similarly, the adult and child level datasets contain basic baseline variables on the individual in question. The full baseline dataset for more detailed analysis can be found at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910/DVN/V614VB