

KENYA CT-OVC PROGRAM DATA USE INSTRUCTIONS

OVERVIEW

This document provides information for using the Kenya CT-OVC data, a three-wave panel dataset that was created to analyze the impact of Kenya's CT-OVC cash transfer program. In addition to explaining the data structure and steps for merging files, it provides brief information about the program and the evaluation.

This dataset is released by The Transfer Project, housed at the Carolina Population Center at the University of North Carolina – Chapel Hill. The third wave of the project, conducted by the Transfer Project, was funded by the National Institutes for Mental Health (NIMH). Additional information about the project not found here or without a direct link can be found on The Transfer Project's Website: <https://transfer.cpc.unc.edu/>.

The data package contains six primary datasets (from three individual/household and three community surveys) and several auxiliary datasets that provide additional information on tracking and attrition. A sensitive section from the third wave of the survey is excluded from this package and can be obtained through special requests (see Sensitive Data section). The survey interviewed households, individuals, and community members at three time points, in 2007, 2009, and 2011.

BACKGROUND

The Program

The Kenya CT-OVC is the government's flagship social protection program, reaching over 260,000 households across the country as of the end of 2015. In response to a concern for the welfare of OVC, particularly AIDS orphans, the Government of Kenya, with technical and financial assistance from the United Nations Children's Emergency Fund (UNICEF), designed and began implementing a pilot cash-transfer program in 2004. After a successful demonstration period, the CT-OVC was formally approved by Cabinet, was integrated into the national budget, and began expanding rapidly in mid 2007 across Kenya. The objective of the program is to provide regular cash transfers to families living with OVC to encourage fostering and retention of children and to promote their human capital development. Eligible households, those who are ultra-poor and contain an OVC, received a flat monthly transfer of \$21 (U.S.) (Ksh 1500). The transfer level was increased to Ksh 2000 per household in the 2011-12 Government of Kenya budget. An OVC is defined as a household resident between 0 to 17 years old with at least one deceased parent, or who is chronically ill, or whose main caregiver is chronically ill. Beneficiary households are informed that the payment is intended for the care and protection of the resident OVC, although this is not a conditional cash grant. The website for the program is <http://labour.go.ke/ovcsecretariat.html>.

The Impact Evaluation

Prior to program expansion of the CT-OVC in 2007, UNICEF designed a social experiment to track the impact of the program on a range of household welfare indicators including child health and schooling and economic productivity. The evaluation was contracted to a private consulting firm, Oxford Policy Management (OPM), and entailed a cluster randomized longitudinal design, with a baseline household survey (and related community survey) conducted in mid 2007 and a 24 month follow-up in 2009. The ethical rationale for the design was that the program could not expand to all eligible locations at the same time, so locations whose entry would occur later in the expansion cycle could be used as control sites to measure impact. Thus within each of 7 districts that were scheduled to be included in this expansion phase four locations were identified as eligible, and 2 were randomized out of the initial expansion phase and served as control locations. Targeting of households was carried out in the intervention locations according to standard program operation guidelines. Each location forms a committee of citizens that is charged with identifying potentially eligible households based on criteria of ultra-poverty and containing at least one OVC as defined above. The list of eligible households is sent to the program's central office (located within the Ministry of Gender, Children and Social Development, the Ministry responsible for the program at the time), which then administers a detailed socioeconomic questionnaire to confirm eligibility, and to assess poverty in order to rank households. The final number of households that enter the program in each district depends on funding to that district but approximately 20 percent of the poorest households in each location are enrolled in the program. Since the program was not scheduled to be implemented during this phase in the control locations, program targeting was 'simulated' in order to identify a sample of households that were comparable to those identified as eligible in treatment locations. Households in either arm (Intervention, Control) were surveyed prior to their knowledge that they were selected into the program.

Evaluation reports produced by OPM can be found at https://transfer.cpc.unc.edu/?page_id=1254.

The Carolina Population Center obtained funding from the NIMH (1R01MH093241-01) to conduct a second follow-up survey of the evaluation sample in 2011, with a special focus on understanding the impact of the program on the successful transition of OVC into young adulthood. The 2011 survey focused on the eligible sample only, and included a special module for young people 15-25 on sexual activity, mental health and peers, administered face-to-face. The main household survey was also expanded to include more detailed information on economic activity, fertility, and time preference. Survey instruments for all three waves of data collection can be found at https://transfer.cpc.unc.edu/?page_id=1188.

Number of households and individuals present at each round

	<i>Communities</i>	<i>Households</i>	<i>Individuals</i>
Wave 1	256	2759	15464
Wave 2	203	2255	12957
Wave 3	202	1811	10399
Total	N/A	2792	19323

Program Eligibility

The OPM evaluation sample included four groups of households: treatment and control households, and non-eligible OVC households in intervention and control localities (referred to as A, B, C and D, respectively, as described below). The latter two groups were included in the initial study in order to assess the targeting effectiveness of the program but these were not surveyed in the 2011 round.

Group A – Households with OVCs in the program areas selected for inclusion in the programme. (N = 1,540 in 2007; N = 1,556 in 2009; N = 1,280 in 2011)

- These are divided into two groups: (i) those in areas with conditions; (ii) those where there are no conditions.
- Note that conditions were not rigorously monitored during the study period though beneficiaries were told about conditions at enrolment.

Group B – Households with OVCs in control areas that were expected to have the met program criteria and should therefore (in theory) have been selected by the program if the programme had operated there. (N = 754 in 2007; N = 765 in 2009; N = 526 in 2011)

Group C – Households with OVCs in program areas that were not selected for inclusion in the program. (N=238 in 2007; N=243 in 2009)

Group D – Households with OVCs in control areas that were expected not to have met program criteria and would not (in theory) have been selected if the program had operated there. (N=227 in 2007; N=228 in 2009)

In the datasets, the variables `hhw1_househol`, `hhw2_household_type` and `hhw3_11` indicate the group to which each household belongs in the different waves (note: as mentioned above, wave 3 covers only Group A and Group B households).

DATA USE AND MERGING INSTRUCTIONS

Adjusting for sample design

In order to properly adjust for the sample design, which first sampled by location and then sampled households within the location, all standard errors should be clustered at the location (also sometimes called sub-location) variables. The location variables are as follows: `hhw1_subloca0` in wave 1, `hhw2_sublocationcode` in wave 2, and `hhw3_5i/hhw3_5ii` in wave 3 (for location and location code respectively).

Datasets included

The datasets included in this package include three individual/household datasets (from wave 1, wave 2, and wave 3) that can be merged to form a panel dataset. The package also includes three “community” datasets, which cannot be merged with each other longitudinally, but can each be joined with an individual/household dataset from the corresponding wave. Finally, there are several sections from waves 2 and 3 that deal with attrition and new household members that are provided separately from the main datasets.

Kenya CT-OVC Program Data Use Instructions

Dataset name	Level	Wave	Type	Sample	ID Variable
hh_w1	Individual	1	Main dataset	Full	<i>indivcode</i>
community_w1	Community	1	Community sections	Full	<i>commidw1</i>
hh_w2	Individual	2	Main dataset	Full	<i>indivcode</i>
hh_w2_notinterviewed	Individual	2	Separate section	Full	N/A
hh_w2_section_a2_q2	Individual	2	Separate section	Full	<i>hhcode id_code</i>
hh_w2_section_a2_q4	Individual	2	Separate section	Full	<i>hhcode id_code</i>
hh_w2_section_a2_q11	Individual	2	Separate section	Individuals new to W2	<i>indivcode</i>
community_w2	Community	2	Community sections	Full	<i>commidw2</i>
hh_w3	Individual	3	Main dataset	Full	<i>indivcode</i>
hh_w3_section_a1	Community	3	Separate section	Full + Individuals from W2 not incl. in W3	<i>indivcode</i> (for those interviewed in W3)
hh_w3_section_a2	Individual	3	Separate section	Individuals new to W3	<i>indivcode</i>
hh_w3_section_l	Individual	3	Sensitive section	Individuals aged 15-25	<i>indivcode</i>
community_w3	Community	3	Community sections	Full	<i>commidw3</i>

Merging datasets

Individual and Household Datasets

This study contains three main individual-level datasets (hh_w1, hh_w2, and hh_w3). These datasets can be merged into a three-wave panel dataset using the unique individual id, ***indivcode***. The datasets also contain other, within-wave or partial ID numbers that are marked as such. Please do not use any individual IDs other than ***indivcode*** to merge or perform other operations on the data, as they may not be reliable and are not consistent across waves.

Although the three main datasets contain individuals as the main unit of observation, most sections in the questionnaire were asked at the household level. The household-level cross-wave ID in all questionnaires is ***hhcode***. If a household-level dataset is desired, we suggest removing all individual-level sections (determined by browsing the questionnaires), and dropping duplicates.

Please note that there are some inconsistencies between waves in variables such as age and gender, which could not be reconciled. We suggest that in cases of inconsistency, baseline values for all variables should be used.

Community Dataset Merge

“Communities” in this dataset represent local units defined by this study for interviewing purposes, rather than self-defined or geographically defined pre-existing communities (which are roughly defined by sub-location codes). As such, all analysis of community variable effects on households and individuals must be interpreted very carefully.

Note that *commidw1*, *commidw2*, and *commidw3* are not the same for the same community across rounds. Some communities were assigned different community codes at follow-up and wave 3. Furthermore, in Wave 3, multiple groups were defined and interviewed in each location, resulting in non-unique community IDs (through an error in ID creation). Therefore, **the three community datasets cannot be merged into one panel dataset**, as identical numbers represent different “community” groups between rounds and within the wave 3 round. However, the sub-location codes are identical across rounds.

It is important to note that some households link to multiple community IDs. This issue arises because certain households belong to locales that had multiple “community interviews” administered to them, and thus cannot be reliably assigned to just one set of answers. Depending on the community questionnaire variable to be used, we recommend that the user take its mean, median, minimum or maximum within the communities having the same community IDs and then keep these value for these communities (alternatively, the user could randomly select any one line of data to keep from within communities with the same ID). The following provides an example:

```
use "hh_w3.dta", clear

preserve
*Example: cw3_e_price1: price of 1 kg of maize flour
use "community_w3.dta", clear
sort commidw3 //this is the community ID for wave 3
bysort commidw3: gen tmp=_n
bysort commidw3: egen maize_p=min(cw3_e_price1) //alternatively, one could take max/mean/median
drop if tmp>1 //keeps only one line of data per community so that there are no duplicates for
community IDs
keep commidw3 maize_p

tempfile tmpah
sa `tmpah'
restore

merge m:1 commidw3 using `tmpah'
drop _m
```

In this example, the minimum price is assigned to a community ID since it is presumed that individuals tend to shop where the prices are lowest.

Finally, another important note is that due to an issue with community IDs in wave 3, there are 14 communities that do not match any household, and 149 households that don't match a community that exists in the community dataset.

Separate Sections

Note that not all household or individual level sections are included in the main datasets for each wave. Both Wave 2 and Wave 3 contain tracking sections that were used to determine which individuals from previous waves are still residing in the household, which individuals moved away, and whether anyone joined the household. Since some of these individuals were

not interviewed, these sections cannot be reliably merged into the main dataset. However, since they provide useful information about reasons for attrition and for joining households, they are provided separately. Although we attempted to include a universal individual id, *indivcode*, whenever possible, some sections and observations are missing the id, or use a different variable for unique identification (listed in the dataset table above). Please do not use any IDs other than *indivcode* to merge sections within or across waves.

Status of individuals and households through waves

Not all the individuals and household are present in each wave. Each of the main files contains two variables that indicate the waves in which the respondent participated, named **WAVESTATINDIV** (calculated after wave 2) and **WAVESTATINDIV2** (calculated after wave 3). These variables can be used to identify and select respondents based on their status in waves (baseline, follow-up, and wave 3), as shown in the following table.

Respondent/Household present in:				
W1	W2	W3	Households	Individuals
+	+	+	1782	7285
+	+		440	3554
+		+	1	212
	+	+	28	1210
+			536	4413
	+		5	908
		+		1692
Other*				49

*Not interviewed in any waves but present in tracking for W3

The variables that similarly indicate wave status of households are **WAVESTATHH** (calculated after wave 2) and **WAVESTATHH2** (calculated after wave 3).

Sensitive Sections

One section is not included in the publicly released dataset package: the young person's module for all individuals aged 15-25 in Wave 3. This section includes questions about sexual behavior, and as such considered to be sensitive data. This section is available through the CPC Portal through a separate request, given appropriate justification of need and IRB approvals.

Supporting Documents

All supporting documents, such as questionnaires and codebooks, are either included in this package or can be found on the Transfer Project Website, <https://transfer.cpc.unc.edu/>, under "Instruments & Reports".