

# THE LESOTHO CHILD GRANTS PROGRAMME (CGP) DATA USE INSTRUCTIONS

## OVERVIEW

This document provides information for using the Lesotho CGP data, a two-wave panel dataset that was created to analyse the impact of Lesotho's cash transfer program. In addition to explaining the data structure, 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. 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 three primary datasets (individual, household and community surveys). The survey interviewed households, individuals, and community members at two time points, in 2011 and 2013.

## THE PROGRAMME

**The Lesotho Child Grants Programme (CGP)** is an unconditional social cash transfer run by the Ministry of Social Development (MSD), targeted to poor and vulnerable households. The objective of the CGP is to improve living standards of Orphans and Vulnerable Children (OVCs) so as to reduce malnutrition, improve health status, and increase school enrolment among OVCs. Households are selected through a combination of proxy means testing and community validation and registered in the National Information System for Social Assistance (NISSA).

The programme is run by the Ministry of Social Development, with financial support from the European Commission and technical support from UNICEF-Lesotho. As of December 2017, the CGP reached 26,600 households and provided benefits for approximately 65,000 children across ten districts in Lesotho. Since 2009 the nature of the CGP has transformed. From an exclusively donor-supported pilot, the CGP has developed institutional and operational systems for roll-out on a national scale. The government has taken charge of funding and is now expanding both NISSA and the CGP nationwide, with the former serving as a platform for better harmonizing social protection interventions in the country.

At the time of baseline data collection for this study in 2011, households received a flat rate of M360 quarterly. Effective April 2013, the transfer was indexed to the number of children in each household: households with 1-2 children continued to receive M360, as payments increased to M600 and M750 to households with 3-4 children and 5 or more children respectively. While beneficiaries received the total intended amount of funds over the study period, the intended payment schedule was not followed, and transfers were often made in more 'lumpy' disbursements than expected – particularly the last payment before the follow up survey. Moreover, a short-term humanitarian intervention was linked to the CGP. The Food Emergency Grant was disbursed to CGP households in 2012 and 2013, with a M400 top-up provided at least two times to each beneficiary household, and up to four times for some.

## ABOUT THE SAMPLE

This is a randomized delayed intervention study. Within 10 Community Councils selected for Phase 1 – Round 2 expansion of the programme, half of all the Electoral Divisions (EDs) were randomly assigned to be covered by the pilot, while the other half served as the comparison group for the period of the study and were only covered after the end of the two-year evaluation period. EDs were assigned to either the treatment or the delayed entry (control in public lottery events that took place in each electoral division).

The study took place in five Districts: Qacha's Nek, Maseru, Leribe, Berea and Mafeteng, covering in total 10 Community Councils and 96 EDs. The survey collected information from both a sample of CGP eligible households (48.7% of the baseline sample) and households who were not eligible for the programme (51.3% of the baseline sample). To identify CGP-eligible households in treatment EDs the CGP implemented the targeting process, selected recipients and proceeded to enrolment. In control EDs the CGP implemented the targeting process and selected recipients who should receive the transfer, but enrolment was delayed until after the follow-up data collection was completed. Two villages (or clusters) were chosen within each selected ED, in each of which 20 households (10 eligible and 10 non-eligible at baseline) were randomly selected and interviewed.

The baseline survey fieldwork took place between June and August 2011 and 3,054 households — more than 98% of the original baseline sample target — roughly equally distributed between treatment and control areas and across eligible and non-eligible households (Table 1). The follow-up survey for this panel dataset fieldwork took place at the same time of the year (between June and August 2013) to avoid seasonality bias, and reached 2,212 households. The fieldwork data collection was undertaken by Sechaba Consultants in direct liaison with Oxford Policy Management (OPM).

**Table 1: Number of households and individuals present at each round**

	Communities	Households	Individuals
Baseline	127	3,054	15,989
Follow-up	124	2,212	12,274
<b>Total</b>	N/A	5,266	28,263

### Sample Attrition

Sample attrition stems from the fact that some households that were interviewed at baseline had left their original community or were no longer available for interview at the time of the follow up survey. This problem was addressed in different ways for eligible and non-eligible households.

Procedures for eligible households: A tracking protocol was established that directed the team to find households that have moved outside their original community and seek an interview if 1) their new location was known to the field team and 2) the households had relocated to: a) a district capital in one of the regions of the study or the capital city Maseru; or, b) a location within 30 minutes or 10 km travel from the village where the household was originally sampled. No replacements were allowed for this group, so households that could not be located/reached or were not available for interview for other reasons were dropped from the study.

Procedures for non-eligible households: Replacements were available and used when households had relocated outside their original community, following the same criteria that were adopted for the baseline survey.

It is also important to note that in the follow-up survey the sample size of non-eligible households was reduced to roughly half of that of the baseline due to budgetary restrictions, leading to an overall target follow-up sample of about 2,300 households (1,484 eligible households and 803 non-eligible).

**Table 2: Household panel and sample attrition**

	Eligible for CGP	Not eligible for CGP	Total
Baseline households	1,486	1,568	3,054
Sampled at follow-up*	1,484	803	2,289
Surveyed at follow up <sup>†</sup>	1,353	797	2,150

\*Note: Due to budgetary restrictions the sample of non-eligible households was reduced by approximately half at follow-up.

†Note: Completed at least one survey at follow-up

The above table shows that 2,150 of the 3,054 households interviewed at baseline were interviewed in the follow-up study. While the overall sample attrition was 6%, it was significantly higher (9%) for CGP eligible households, for which replacements were not available. For non-eligible households, there was a 17% replacement rate for follow-up interviews. The analysis further suggests that there were some systematic differences in the response rate to the follow-up survey between treatment and control groups. The attrition rate among eligible households was much higher in the control group (12%) than the treatment group (8%), mainly because a higher proportion of households in the control group moved outside the cluster to a location where tracking was not viable.<sup>1</sup> The sampling weights have been adjusted for selective non-response by calculating the probability of households being retained in the sample on the basis of key household characteristics at baseline.

### Program Eligibility

The survey for the impact evaluation collected information on four groups (Table 3):

- A – CGP-eligible households in the programme areas (actual beneficiaries)
- B – CGP-eligible households in control areas (would be beneficiaries or delay enrolment group)
- C – Households in programme areas that are not eligible for inclusion in CGP programme (spillover group)
- D – Households in control areas that are not CGP-eligible

The comparison of trends over time in the programme recipients (group A) and CGP-eligible households in control areas (group B) provided the basis for the analysis of the direct impact of the CGP on beneficiary households and children. The study also collected a sample of households that were not eligible for the programme, both in treatment communities (Group C) and control communities (Group D). These additional observations were used to conduct an analysis of targeting effectiveness at baseline, and to study local spill-over effects at follow-up.

**Table 3: Structure of the sample**

Beneficiary Status	Area		Total
	Treatment	Control	
Eligible for CGP	747 [A]	739 [B]	1,486
Non-eligible for CGP	779 [C]	789 [D]	1,568
<b>Total</b>	<b>1,526</b>	<b>1,528</b>	<b>3,054</b>

<sup>1</sup> For details, see table 79 and 80 in the follow-up OPM report, available on the Transfer Project website or here: [http://www.fao.org/fileadmin/user\\_upload/p2p/Documents/draft\\_CGP\\_Follow\\_Up\\_v11\\_out.pdf](http://www.fao.org/fileadmin/user_upload/p2p/Documents/draft_CGP_Follow_Up_v11_out.pdf)

## Sampling Weights

**Important note: this sample is not meant to be externally representative.** Sampling weights have been generated and used to produce estimates that relate to all households living in the electoral divisions covered by the evaluation. Even though the EDs were selected randomly, the ED sampling probabilities are not reflected in the household sampling weights and therefore the estimates do not apply to any households that are located outside the evaluation EDs. As such, the EDs selected for the evaluation represent only the ‘study population’ and no inferences should be drawn about the wider population of Lesotho.

In the baseline study household sampling weights were given by:

$$w(ij) = (A_i / (m_i * a_{ij})) * N_{ijk} / n_{ijk}$$

where  $A_i$  is the total number of households in the sample frame of Cluster of Villages for ED $_i$ ,  $m_i$  is the number of Cluster of Villages sampled in ED $_i$ ,  $a_{ij}$  is the number of households in Cluster  $ij$ ,  $n_{ijk}$  is the number of households of type  $k$  interviewed in Cluster  $ij$ , and  $N_{ijk}$  is the total number of households of type  $k$  listed in Cluster  $ij$ . Sampling weight (**hh\_wt**) should be use when working with baseline data.

As noted earlier, the sub-sample of non-eligible households was cut by approximately half at the follow-up due to budgetary constraints. Besides this change, the construction of the panel was hindered by two typical challenges of sample attrition and changes in the demographic structure of the households in the sample. In order to counteract unbalanced attrition, weights (**hh\_wt\_fu**) should be used in the follow up data.

The data within this study has been corrected in the follow up survey and some households were split and/or added and assigned a new unique identifier (Table 4).

**Table 4: Dynamic codes**

	Code	N	Percent	Cumulative
<u>None</u> : no children from baseline live in the household	00	4	0.18	0.18
<u>Same household</u> : all or most of the children from baseline still live in household	01	2,051	92.72	92.90
<u>1<sup>st</sup> split</u> : the first household some children from baseline moved to	02	69	3.12	96.02
<u>2<sup>nd</sup> split</u> : the second households some children from baseline moved to	03	3	0.14	96.16
<u>New household</u> : either the whole household <u>or</u> all/most of the children from baseline moved to a different household	09	85	3.84	100.00

## MAIN CHARACTERISTICS OF THE DATASETS

This study contains a household level dataset, an individual level dataset and a community level dataset. While the latter is retrieved from a community survey, the former two both come from the household survey, in which specific sections like the roster (section 1) collect information about individual household members.

## Individual and household datasets

The household dataset includes a set of preliminary variables (such as the household ID, treatment arm, and other identifiers), followed by the variables that are included in the survey. For greater ease of viewing, we created artificial string variables (h0, h5, h6, etc) that clearly separate the various sections of the questionnaire. For instance, the variable h0 is a string equals to “-----> Section 0: Interview details” that separate identifiers from household level variables included in the cover of the questionnaire.

Among the identifiers, **year** and **wave** represent the time variables, while **idno** is the variable uniquely identifying the households at baseline only. Please check this with the following set of stata commands:

```
use lso_cgp_ie_hh13.dta, clear
keep if wave==1
isid idno
```

The purpose of the follow-up study was to track children, not households, which means that some baseline households have split in one or more households. For this reason, we have a different unique identifier at follow-up, **unihhid**. This creates some issues while analysing the data from a longitudinal perspective, because few households at follow-up may share the same baseline household identifier (**idno**).

**EXAMPLE:** Consider for instance household with idno equals to 69:

```
. list idno unihhid wave h0q7b if idno==69, noo
```

idno	unihhid	wave	h0q7b
69	6902	follow-up	1st split (first hhd some children from BL have moved to)
69	6909	baseline	.
69	6909	follow-up	new hhd (the whole hhd/all or most of the children from BL m

This household has split in two: unihhid=6902 and unihhid=6909. In the dataset we have imputed 6909 to the **unihhid** also for the baseline, based on a children prevalence criterion: if all or most of the children were in one specific follow-up household, that household **unihhid** identifier is also the baseline **unihhid** identifier. This means that unique identifiers for a panel analysis are variables **unihhid** and **wave** (or **year**):

```
use lso_cgp_ie_hh13.dta, clear
isid unihhid wave
xtset unihhid wave
```

The children prevalence criterion is based on the dynamic codes, expressed with the variable **h0q7b**, and the number of children retrieved from both waves.

**WARNING NOTE:** Data users are advised to consider carefully whether this choice is valid for their specific research questions. A different **unihhid** for the baseline can be imputed based for instance on where the baseline head of the household or the mother of the children lives, etc. In that case the baseline value for **unihhid** is no longer valid and we invite users to construct a different **unihhid** baseline value. However, it is likely this will have only minor implications, as 7.1 percent of households at follow-up report **h0q7b**=2, 3 or 9, which means we have only 7.1 percent split households (see table 4).

The set of preliminary variables also include the baseline household weight, **hh\_wt**, and the variables identifying the treatment arm: **elig** (=0 for CGP non-eligible, =1 for CGP eligible households) and **T** (=0 for control areas, =1 for treatment areas). Information on household's eligibility and area treatment status is also provided in question **h0q5**. It is important to note that for 26 cases, households moved outside sampled clusters and during fieldwork this information has not been recorded (but was generically tagged with code=9999). For this reason there is a mismatch between the baseline and the follow-up cluster codes (**h0q3** and **h0q3b** respectively). In any case, while calculating clustered standard errors, using either of them should have minimal implications.

Some sections are reported at household level, though it could be more convenient to analyze them using a different statistical unit. For instance, section 6 reports information at plot level. In this case we could simply reshape the dataset as follows:

```
use lso_cgp_ie_hh13.dta, clear

reshape long h6q4_plot h6q4b_plot h6q5_plot h6q6_plot h6q7_plot h6q8_plot
h6q9_plot h6q10_plot h6q11_plot h6q12_plot, i(unihhid idno year) j(plot )

forvalues k=4/12 {
    rename h6q`k'_plot h6q`k'
}

rename h6q4b_plot h6q4b
keep unihhid idno year plot h6q*
```

The individual dataset includes a set of preliminary variables that identify the household member, including the **wave**, **year**, **idno** and **unihhid** fields we have described above. The personal identifier is given by the **idcode** variable. A member is uniquely identified by unihhid, idcode and wave:

```
use "$dta\lso_cgp_ie_ind13.dta"
isid unihhid idcode wave
```

### ***Creating age and gender variables***

Most inconsistencies between waves in variables such as age and gender were reconciled by Sechaba Consultants in close collaboration with OPM immediately after follow-up fieldwork. Please note that the age and gender variables have been differently named in the two survey rounds. To avoid confusion and a mismatch between datasets and questionnaire, we have left variable names as they are and have properly labelled them instead. To create unique age and gender variables, please use the code:

```
gen sex = h1q2 if wave==1
replace sex= h1q3a if wave==2
gen age = h1q3 if wave==1
replace age= h1q2a if wave==2
```

### **Community dataset**

The community dataset is uniquely identified by the **time** variable, **wave**, and the cluster variable, **caq3**:

```
isid caq3 wave
```

To simplify matching with the household dataset, we have included a variable, **h0q3**, matches the variable representing the clusters (communities) in the household dataset and can be used for merging.

**WARNING NOTE:** Data users are advised that the number of community surveys at baseline is 127, while at follow-up information on only 124 communities has been retrieved. This means that in a panel data analysis where community characteristics are included, a researcher has two choices: 1) to lose some individuals/households/plots. due to communities missing data pattern; 2) to impute community characteristics for the missing clusters at follow-up, using information from the other 124 clusters.

### Merging datasets

We match the individual level dataset with the household level dataset as follows:

```
use lso_cgp_ie_ind13.dta, clear
merge n:1 unihhid wave using lso_cgp_ie_hh13.dta, nogen
```

We match instead the household level dataset with the community level dataset as follows:

```
use lso_cgp_ie_hh13.dta , clear
merge n:1 h0q3 wave using lso_cgp_ie_cc13.dta
```

63 households from the master (the household level dataset in this example) should not link to any community dataset.

### De-identification and sensitive information

For security and privacy purposes, names, contact, and any potentially identifying information of the individuals and households have been removed, and the names of any geographic units smaller than a district have been coded. Other sensitive information that is not part of the data release include questions related to HIV/AIDS. Further, the management information system code used to track social networks in section 17 has been changed and de-identified.

---

**NOTE: More information on the evaluation can be found in the baseline and follow-up reports.**

Baseline report:

[http://www.fao.org/fileadmin/user\\_upload/p2p/Publications/Lesotho\\_baseline\\_note.pdf](http://www.fao.org/fileadmin/user_upload/p2p/Publications/Lesotho_baseline_note.pdf)

Endline report:

[http://www.fao.org/fileadmin/user\\_upload/p2p/Documents/draft\\_CGP\\_Follow\\_Up\\_v11\\_out.pdf](http://www.fao.org/fileadmin/user_upload/p2p/Documents/draft_CGP_Follow_Up_v11_out.pdf)

---