Young Lives Survey Design and Sampling (Round 5)
Peru

This fact sheet describes the survey methods and sample design in Peru and attrition rates throughout the fifteen years of the Young Lives study (since Round 1 in 2002 to Round 5 in 2016). It also provides information on the sites where Young Lives children live in order to contextualize the findings of the accompanying fact sheets, which present preliminary results of the fifth survey round in the areas of: *Education and Learning, Growth and Nutrition*, *Youth Transitions: Skills, Work and Family Formation*, and *Poverty and Intergenerational Change.*

About Young Lives

Young Lives is designed as a cohort study following the lives of 12,000 children in four low and middle-income countries – Ethiopia, India (in the states of Andhra Pradesh and Telangana), Peru and Vietnam – over 15 years. The sample in each country consists of two cohorts of children: a Younger Cohort of 2,000 children who were aged one when the first round of the survey was carried out in 2002, and an Older Cohort of 1,000 children (700 in Peru) then aged eight.

Young Lives has collected a wealth of information through a large-scale household survey of all the children and their primary caregivers, bolstered by in-depth interviews, group work and case studies with a sub-sample of the children, their caregivers, teachers and community representatives. This not only tracks the material and social circumstances of the Young Lives sample, but also captures their perspectives on life and their aspirations for the future, set against the environmental and social realities of their communities.

The fact that our work spans 15 years in the lives of these children – covering all ages from early infancy into young adulthood when some have become parents themselves – means that we are able to examine how the lives of children, living in different circumstances and in diverse contexts, change over time. The five rounds of survey data, supplemented by four rounds of nested qualitative case studies, affords Young Lives a unique cross-country longitudinal dataset exploring the causes and consequences of poverty in childhood.

Figure 1. *Young Lives longitudinal and cohort study*

Figure 2. *Young Lives study sites in Peru*
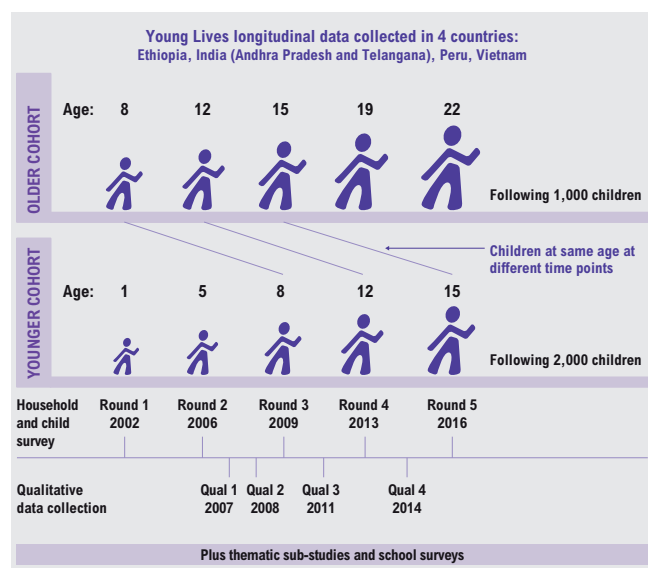
Young Lives in Peru

Five rounds of quantitative surveys of children, households and communities have been conducted by Young Lives (Niños del Milenio) in Peru. The first round was carried out between August and December 2002 when the children were aged around 1 year (Younger Cohort) and 8 years (Older Cohort) of age. The following surveys were carried out at the same time of year in 2006 (Round 2), 2009 (Round 3), 2013 (Round 4) and, most recently, in 2016 (Round 5) when the children were aged approximately 15 years (Younger Cohort) and 22 years (Older Cohort).

Sample design

The children were selected from 20 sentinel sites that were defined specifically in each country. The concept of a sentinel site comes from health surveillance studies and is a form of purposive sampling where the site (or 'cluster' in sampling language) is deemed to represent a certain type of population, and is expected to show typical trends affecting those people or areas.

While following this general approach, the Peru team applied a sampling approach that differed in some respects from the other three study countries. Notably, the sampling of clusters in the other countries was semi-random/semi-purposive, while in Peru that sampling of clusters was random. The district level was used as the sample frame. The most recent poverty map of the 1,818 districts in Peru at that time (FONCODES, 2001) was used to select the 20 sentinel sites. Factors such as infant mortality, housing, schooling, road networks and access to services determined the ranking of districts.

To achieve the aim of over-sampling poor areas, the highest ranking 5% of districts (all located in Lima) were excluded, enabling a systematic selection of the remaining districts, which yielded approximately 75% of sample sites considered to be 'poor' and 25% 'non-poor'. Then, districts were divided into equal population groups. These units were ordered by a poverty index and were systematically sampled with randomisation of the starting place. Ten selection runs were made and the resulting districts were examined to cover rural, urban, peri-urban, coastal, mountain and Amazon areas and for logistical feasibility, and one of them was selected for the sampling. Since all the districts were divided into units with equal population groups, it follows that each district had a probability of being selected that was proportional to its population size.

Once the districts were chosen, a random population centre (i.e. a village or hamlet) was chosen within the district. The maps of census tracts were obtained from the National Statistics Institute (Instituto Nacional de Estadistica), and a census tract randomly selected. Within each chosen census tract, the number of *manzanas* (street blocks) was counted, and again, using random number tables, one was selected as the starting point. In each district, approximately 100 and 25-50 children were enrolled for the Younger and Older Cohorts respectively.

Finally the selected block was assigned to one fieldworker and neighbouring blocks to other fieldworkers (one each). All dwellings in each block or cluster of houses were visited to identify families with children of the right ages. On completion of one block, the next available neighbouring block was visited by the fieldworker until the required number of children was found.

The project team visited a total of 36,373 dwellings to recruit 2,751 children. Although this may seem high, we estimated (using census data) that we would need to visit 13 families to recruit one child of the right age. This is about the same ratio reported for our recruitment process.

Table 1. *Young Lives sentinel sites in Peru*

Cluster ID	Department	Anonymised name*	Short description
1	Tumbes	Tumbes	A small city on the northern coast in the Tumbes region
2	Piura	Sullana	A poor coastal rural area in the Piura region
3	Piura	Morropon	A very poor rural area in the northern Andean highlands
4	Amazonas	Chachapoyas	A very poor rural area in the northern Amazon region
5	San Martin	Rioja	A poor rural area in the San Martin region
6	San Martin	San Martín	A medium-sized city in the San Martin region
7	Cajamarca	Cajamarca	A medium-sized city in the northern Andean highlands, in the Cajamarca region
8	La Libertad	Trujillo	A shanty town on the outskirts of a medium-sized city in the La Libertad region on the northern coast
9	Ancash	Huaylas	A poor rural area in the central Andean highlands
10	Ancash	Huaraz	A medium-sized city in the central Andean highlands, in the Ancash region
11	Huanuco	Dos de Mayo	A very rural area in the centre of the Andean highlands, in the Huanuco region
12	Lima	San Juan de Lurigancho	A large urban district located in the north of the capital city, Lima
13	Lima	Ate	A large urban district located in the eastern part of the capital city, Lima
14	Lima	Villa Maria de Triunfo	A large urban district located in the south of the capital city, Lima
15	Junin	Satipo	A poor rural area in the Amazon part of the Junin region
16	Ayacucho	Huamanga	A very rural poor community in the southern-centre of the Andean highlands
17	Ayacucho	Lucanas	A poor rural area in the southern-centre of the Andean highlands
18	Apurimac	Andahuaylas	A poor rural area in the southern Andean highlands
19	Arequipa	Camaná	A small city on the southern coast in the Arequipa region
20	Puno	Juliaca	A small city on the southern coast in the Arequipa region

*Note: In order to protect the children's anonymity, we use the province name to anonymise the study sites, or the district name is used if the population is above 125,000.

Comparing Young Lives to other datasets

Young Lives is not intended to be a nationally representative survey such as the Demographic and Health Survey (DHS). Rather, as a longitudinal study, it is intended to show changes for individuals over time and the impact of earlier circumstances on children's later outcomes, shedding light on the differences between age, ethnicity, gender, location, wealth tercile and more. A comparison to the DHS from 2000 (the year closest to Round 1 of Young Lives in 2002), indicates that the Young Lives sample covers the diversity of children and families in Peru (Escobal and Flores 2008).

At the same time, their analysis indicates that, on average, the Young Lives sample includes households with more education, with better access to services, and more assets owned than in DHS. However, this does not take into account the fact that in the Young Lives sample for Peru, each district had a probability of being selected proportional to its population size. Once each observation is adjusted to account for this, many of the differences found between the Young Lives and DHS 2000 samples are not significant. For this reason, we report results for the Young Lives sample in Peru using this sampling frame as these results most closely resemble what is happening in the country.

Tracking and attrition

Sample attrition occurs when children who were surveyed in the first round of a survey are either not found or refuse to take part in later rounds. We made sure to track as many children as possible between survey rounds to minimise the risk of drop-out. In the case of Peru, one notorious challenge has been the high degree of geographical mobility observed among the two cohorts, exemplified by the fact that while we started visiting 27 districts in 13 regions in Round 1, by 2016 we had to visit close to 300 districts (out of 1836) and all the 24 regions to complete our interviews. To ensure cohort maintenance, we administered tracking surveys between rounds and tried to follow families everywhere within the national bounds.

Table 2. *Attrition between Round 1 and Round 5*

	Younger Cohort	Older Cohort
Initial Sample Round 1 (2002)	2,052	714
Refused	97	42
Untraceable	38	39
Living Abroad	32	19
Interviewed in Round 5 (2016)	**1,860**	**608**
Attrition (%) *	8.2	14.1

*Note: We do not include deaths within attrition rates which amount to 25 (1.2%) for the Younger Cohort, and 6 (0.8%) for the Older Cohort

Attrition rates observed in the Peru sample are relatively low compared to other longitudinal studies: 8.2% for the Younger Cohort and 14.1% for the Older Cohort since the start of the study. The larger attrition rate among the Older Cohort is likely explained by the fact that as its members became adults, many of them moved (to study, to work, or both). Many will also have started forming their own families, often making it more difficult to trace them and increasing the chances of refusals. In the case of the Younger Cohort, the large majority of children still live with their parents.

Table 3. *General characteristics of the Young Lives sample in Round 5 (2016)*

	Younger Cohort		Older Cohort	
	Number	%*	Number	%*
Gender				
Male	938	49.7	320	50.6
Female	922	50.3	288	49.4
Area of residence in 2002 (Round 1)				
Urban	1268	62.3	462	57.5
Rural	592	37.7	146	42.5
Maternal Education				
Primary incomplete or less	512	32.2	183	45.3
Complete primary or secondary	1135	58.3	372	49.7
Higher education	210	9.5	52	5.0
Mother's first language (ethnicity proxy)				
Spanish	1284	64.7	420	41.6
Indigenous language	545	35.3	177	58.4

*Percentages are weighted to adjust for the sample frame.

Note: Totals for some categories may not always add up to the total for each cohort due to missing data for some children.

REFERENCES AND FURTHER READING

Escobal. J. and E. Flores (2007). *Peru: Sampling Design and Post-Sampling Adjustment*s, unpublished document.

Escobal, J. and E. Flores (2008). *An Assessment of the Young Lives Sampling Approach in Peru*, Technical Note 3, Oxford: Young Lives.

Fenn, B, Huttley, S. and I. Wilson (2006). 'A Case Study of Sample Design for Longitudinal Research: Young Lives', *International Journal of Social Research Methodology.* 9(5). 351-365.

Outers-Leon, I. and S. Dercon (2008) *Survey Attrition and Attrition Bias in Young Lives.* Young Lives Technical Note 5, Oxford: Young Lives.

Young Lives (2017) *Young Lives Methods Guide*, available at: http://www.younglives.org.uk/ what-we-do/research-methods/methods-guide