An Assessment of the Young Lives Sampling Approach in Andhra Pradesh, India

Neha Kumra

Young Lives
An International Study of Childhood Poverty

Contents

Executive summary

Young Lives is a longitudinal research project investigating the changing nature of childhood poverty. The study is tracking the development of 12,000 children in Ethiopia, Peru, India (Andhra Pradesh) and Vietnam through qualitative and quantitative research over a 15-year period. Since 2002, the study has been following two cohorts in each study country. The younger cohort consists of 2,000 children per study country aged between 6 and 18 months in 2002. The older cohort consists of 1,000 children per country aged between 7.5 and 8.5 in 2002. The key objectives of Young Lives are: (i) to improve the understanding of causes and consequences of childhood poverty, (ii) to inform the development and implementation of future policies and practices that will reduce childhood poverty.

The sampling methodology adopted by the Young Lives team in Andhra Pradesh is known as a sentinel site surveillance system. It consisted of a multi-stage, purposive and random sampling to select the two cohorts of children. This methodology randomised households within a study site while the sites themselves were chosen on the basis of predetermined criteria, informed by the objectives of the study. To ensure the sustainability of the study, and for resurveying purposes, a number of well-defined sites was chosen. The sites were selected from three different agro-climatic areas and, in accordance with the project aims, had a pro-poor bias with districts and sites being ranked according to a number of development indicators.

This paper assesses the sampling methodology by comparing the Young Lives sample with a larger, nationally representative sample. In doing this, the Andhra Pradesh team sought to:

- analyse how the Young Lives children and households compare with other children in Ethiopia in terms of their living standards and other characteristics;

- examine whether this may affect inferences between the data;

- establish to what extent the Young Lives sample is a relatively poorer or richer sub-population in Andhra Pradesh;

- determine whether different levels of living standards are represented within the dataset.

We used the Demographic and Health Survey 1998/99 (DHS) as comparison sample and applied two different methodologies to assess the Young Lives sample. We first compared wealth index scores for the Young Lives households with those for DHS households. This provided a graphical illustration of the relative wealth of the Young Lives sample relative to the population of Andhra Pradesh. We went on to use standard t-tests to test for statistical significance of the differences in several living standard indicators between Young Lives and the DHS samples. In order to ensure comparability of the different samples we imposed constraints on the comparison samples to accommodate the fact that the Young Lives sample only includes households with at least one child aged between 6 and 18 months.

Given the Young Lives sampling procedure, we expected to find significant differences between Young Lives and the DHS sample. First, the Young Lives aim to document child poverty meant that sentinel sites were sampled over-proportionally from poorer areas. Second, sentinel sites were selected to ensure a balanced representation of the state's regional diversity and the rural and urban divide, so both rural and urban and regional weights within the Young Lives sample are not necessarily consistent with the true state-level population weights. Third, differences in the collection year between the two samples and improvements in the overall living standards in Andhra Pradesh between 1998-99 and 2002 imply that Young Lives households may appear richer than they actually are.

We found that households in the Young Lives sample were slightly wealthier than households in the DHS sample. A similar picture emerged when we use t-tests to compare the means for a range of living standard indicators between the Young Lives and the DHS samples. Young Lives households had better access to public services and owned more assets. However, households were less likely to own their house, the children's caregivers were less educated, and Young Lives mothers were less likely to breastfeed, receive an antennal visit or to have been vaccinated from tetanus.

We conclude that the Young Lives sample includes a wide range of living standards akin to the variability found in the Andhra Pradesh population as a whole. On average, we find that Young Lives households are slightly better-off than the average Andhra Pradesh household, even if in individual indicators they appear to be worse-off. We find that these differences might be partly explained by differences in the survey year between the DHS and the Young Lives samples and the fact that poverty rates in Andhra Pradesh decreased substantially between 1999 and 2005.

Despite these biases, it is shown that the Young Lives sample covers the diversity of children Andhra Pradesh in a wide variety of attributes and experiences. Therefore, while not suited for simple monitoring of child outcome indicators, the Young Lives sample will be an appropriate and valuable instrument for analysing causal relations, modelling child welfare, and its longitudinal dynamics in Andhra Pradesh.

1. Introduction

Young Lives is a longitudinal research project investigating the changing nature of childhood poverty. The study is tracking the development of 12,000 children in Ethiopia, Peru, India (Andhra Pradesh) and Vietnam through qualitative and quantitative research over a 15-year period. Since 2002, the study is following two cohorts in each study country. The younger cohort or 1-year-old cohort consists of 2,000 children per study country aged between 6 and 18 months in 2002. The older cohort or 8-year-old cohort consists of 1,000 children per country aged between 7.5 and 8.5 years in 2002. The key objectives of Young Lives are: (1) to improve the understanding of causes and consequences of childhood poverty, (2) to inform the development and implementation of future policies and practices that will reduce childhood poverty.

To fit the main objectives of the project, Young Lives employed a sentinel site sampling method, which is a multistage sampling approach and uses both purposive and probability sampling methods. While households within the sites were selected by a method equivalent to random sampling, the site selection process was not random and over sampling of poor sites took place.

The aim of this report is to assess the chosen sampling methodology for Andhra Pradesh (India) and to analyse the comparability of Young Lives first round data with data from the Demographic and Health Survey 1998/99 (DHS) (IIPS 2000). The DHS is a household survey that provides data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition. The DHS 1998/99 provides regionally representative data for Andhra Pradesh. The sample consists of 4,032 women and 3,872 households.

By comparing data from Young Lives and the DHS, the study attempts to answer two questions: (i) what are we looking at in the Young Lives data set? (ii) to what extent is the Young Lives sample reflective of the population of 1-year-old and 8-year-old children in Andhra Pradesh?

We realise that the Young Lives and DHS data sets were collected in different times. The decline in poverty head count ratio between 1999 and 2005 in Andhra Pradesh could have created bias which could create the impression that the Young Lives households are wealthier than they really are (Dev and Ravi 2007). However, approximately 75 per cent of households in the DHS and Young Lives samples are located in rural areas where poverty decline was much slower and lower compared to urban areas.

The outline of the report is as follows. Section 2 presents the sampling approach that was adopted by Young Lives. Section 3 presents variables used for the comparison and discusses the results of the statistical comparison tests. Section 4 concludes.

2. Young Lives sampling approach in Andhra Pradesh

The sampling strategy followed by Young Lives in Andhra Pradesh was semi-purposive. The districts and the 20 sentinel sites (mandals) from within the chosen districts were selected following a set of criteria. Then the selection of 100 households with a 1-year-old child and 50 households with an 8-year-old child per sentinel site was random.

Andhra Pradesh is divided into 23 administrative districts that are subdivided into a number of mandals. Generally, there are between 20 and 40 villages in a mandal. In total, there are 1,125 mandals and 27,000 villages in Andhra Pradesh. Villages are normally composed of a main village site with a small number (two to five) of associated hamlets.

Andhra Pradesh state can be categorised into three distinct agro-climatic regions: Coastal Andhra, Rayalseema and Telangana (Young Lives 2007). The sampling scheme adopted for Young Lives was designed to identify interregional variations with the following priorities:

- a uniform distribution of sample districts across the three regions to ensure full representation

- the selection of one poor and one non-poor district from each region, with district poverty classification based on development ranking

- when selecting poor districts and mandals, consideration was given to issues which might impact upon childhood poverty, including the presence or non-presence of the Andhra Pradesh District Poverty Initiative Programme (APDPIP).

Hyderabad district is urban and metropolitan and therefore different selection criteria were applied.

Selection of districts

The first step of selecting the sentinel sites was to select the districts. The selection attempted to (i) ensure that there was a uniform distribution of sample districts across the three geographical regions and (ii) that a poor and a non-poor district was selected from each region.

In order to classify districts from the three regions according to their poor/non-poor status, districts were ranked according to their relative level of development, based on three categories of indicators: economic, human development and infrastructure development indicators. Economic indicators included per cent of gross irrigated land, per capita income, and per cent of urban population. Human development indicators included proportion of the population who belongs to Scheduled Castes/Scheduled Tribes (SC/ST), female literacy rate, infant mortality rate, per cent of children aged 5 to 14 years who do not attend school. Infrastructure development indicators included total road length per 100km square, number of banks per 10,000 people, number of hospital beds per 10,000 people (Galab et al. 2003). A relative development index was constructed using a ranking method. Sectoral ranks (economic, human development and infrastructure) were aggregated using the following weights: economic 30, human development 40, and infrastructure 30. The overall scores, given by the weighted sum of the sector ranks, were ranked to give the final ranking of the districts. Based on these ranks a representative group of poor and non-poor districts was selected (Young Lives 2007)

From Coastal Andhra region, three poor districts were initially selected: Srikakulam, Prakasam and Vizayanagaram, with the APDPIP being implemented in two of the three districts. The APDPIP baseline survey (Dev et al. 2002) identified migration as an important issue in Srikakulam. Therefore, Srikakulam was chosen because the longitudinal aspect of Young Lives will enable exploration of whether a key policy intervention such as APDPIP, and a significant demographic phenomenon such as migration, impact upon child well-being. West Godavari was selected as representative of the non-poor districts of Coastal Andhra.

Anantapur was selected as the poor district of Rayalaseema region. It has a pattern of low rainfall (553mm per annum) and is a desert-prone area. Anantapur is the only district in Rayalaseema region where APDPIP is being implemented. In addition, the UNDP–South Asia Poverty Alleviation Programme (UNDP-SAPAP) was launched in Anantapur in 1995. Two non-poor districts were initially selected in Rayalaseema, Chittor and Cuddapah. Chittor has a higher level of urbanisation and development than the other non-poor districts in the region and therefore finally Cuddapah was chosen as being more representative of the non-poor, non APDPIP districts.

From Telengana region, three poor districts were initially selected, with APDPIP being implemented in two of them: Adilabad and Mahaboobnagar. Adilabad is a largely tribal district and not representative of the region as a whole, therefore Mahaboobnagar was chosen for the survey. It should be noted that tribal communities were captured from a number of other districts throughout the regions. Three non-poor districts were initially selected in Telangana. Rangareddy is located on the outskirts of Hyderabad. Since poor slum communities were to be surveyed within the city, it was deemed unsuitable to work in a district within such close proximity of the city sites. The other two non-poor selections were Karimnagar and Nizamabad. Following close consultation with NGO agencies working in the region, Karimnagar was selected as the non-poor region in Telangana.

The districts accounted for about 28 per cent of the state's population and covered 318 mandals out of 1,119 (excluding Hyderabad).

Selection of sentinel sites/mandals

The second step was the selection of mandals within the selected districts. Young Lives defined a sentinel sites as equivalent to an administrative mandal. A mandal was considered to be urban if more than two-thirds of the population lived in urban areas. Since there are relatively few urban mandals, the district capital was invariably chosen for urban sentinel sites.

One sentinel site was chosen from the urban slums of the state capital Hyderabad. The remaining 19 sentinel sites (excluding Hyderabad) were selected from the six poor and non-poor identified districts. In order to select rural sentinel sites, mandals were classified according to their relative level of development based on a set of indicators (economic, human development and infrastructure). The development indicators differ from those used for district-level selection due to the scarcity of data at the mandal level, although mandal ranking was carried out in the same way as district ranking.

Selection of villages

The next step of sampling was the selection of villages within the mandals. Villages and their associated hamlets were defined as communities in rural areas. Each mandal was divided into four contiguous geographical areas and one village was randomly selected from each area. Care was taken to ensure that the four villages selected from each sentinel site had threshold populations sufficient to give rise to 100 households with a 1-year-old child and 50 households with an 8-year-old child. In a number of cases, additional villages had to be included, where sufficient children were not identified from the selected sample villages.

In urban areas, municipal wards were defined as communities. In Hyderabad city, three slum areas were selected in different areas of the city and included neighbourhoods with an ethnic and religious composition representative of the cultural diversity characteristic of Hyderabad.

Before data collection began in selected communities, a door-to-door listing schedule was completed in order to identify eligible children.

3. Comparison of Young Lives with the Demographic and Health Survey

Table 1 presents a list of common variables from the Young Lives data and the DHS data that are used in the comparison.

Table 1. *Comparison variables, Young Lives and DHS*

Household characteristics

Location of household
Household size
House ownership
Sewing machine ownership
Refrigerator, television ownership
Motorbike, bicycle, car/truck ownership
Fan, clock ownership
Agricultural assets ownership
Electricity supply
Landline phone ownership
Cooking fuel usage
Source of drinking water
Sanitation facility
Livestock owned in the last year
Recent death of a household member
Work in last year

Head of household and primary caregiver characteristics

Age of household head
Sex of household head
Education level of primary caregiver
Religion of primary caregiver
Ethnicity of primary caregiver
Marital status of primary caregiver

Pregnancy, delivery, and breastfeeding

Number of antenatal visits
Received antenatal care
Tetanus injections during pregnancy
Breastfeeding
Location of delivery

3.1   Household wealth and assets

The Wealth index is the primary instrument used in the Young Lives survey to measure the socioeconomic status of households. It produces values between 0 and 1, whereby a higher wealth index indicates a higher socioeconomic status. It is computed as simple average of three individual indexes that also range between 0 and 1: housing quality, consumer durables and access to services (see Appendix 2 for a detailed description).

Table 1 shows the ranking of Young Lives sentinel sites by wealth index scores. It shows that 16 out of the 20 sentinel sites selected for the Young Lives study have wealth index values below the mean wealth index of 0.5 and can be classified as poor. This finding is consistent with the pro-poor sampling approach adopted by Young Lives.

Table 2. *Ranking of Young Lives sentinel sites by wealth index*

Ranking	Sentinel site code	District	Wealth Index	Wealth Index below 0.5
1	16	Mahabubnagar	0.186	Yes
2	19	Mahabubnagar	0.204	Yes
3	04	Srikakulam	0.204	Yes
4	11	Anantapur	0.210	Yes
5	02	West Godavari	0.223	Yes
6	12	Anantapur	0.242	Yes
7	13	Anantapur	0.250	Yes
8	15	Karimnagar	0.250	Yes
9	18	Mahabubnagar	0.253	Yes
10	17	Mahabubnagar	0.264	Yes
11	09	Cuddapah	0.276	Yes
12	08	Cuddapah	0.285	Yes
13	07	Srikakulam	0.327	Yes
14	06	Srikakulam	0.330	Yes
15	05	Srikakulam	0.352	Yes
16	10	Anantapur	0.499	Yes
17	20	Hyderabad	0.566	No
18	14	Karimnagar	0.609	No
19	01	West Godavari	0.617	No
20	03	Srikakulam	0.633	No

We calculate a comparable version of the wealth index for Young Lives and the DHS survey and compare the cumulative distribution functions (cdf) graphically. This is not a standard comparison test but it provides us with first glance on the differences between the surveys.

Figure 1 presents the wealth index cumulative distribution functions for the Young Lives and the DHS survey sites, using the household as the unit of observation. The figure shows a number of key findings. First, for a wide range of values of the wealth index, the Young Lives distribution function is below the DHS distribution. The Young Lives distribution crosses the DHS distribution close to the eightieth percentile, suggesting that fewer very rich households are included in the sample. Moreover, the Young Lives sentinel sites cover a very wide range of wealth index values.

Figure 1. *Wealth index cumulative distribution functions, Young Lives and DHS, by household*
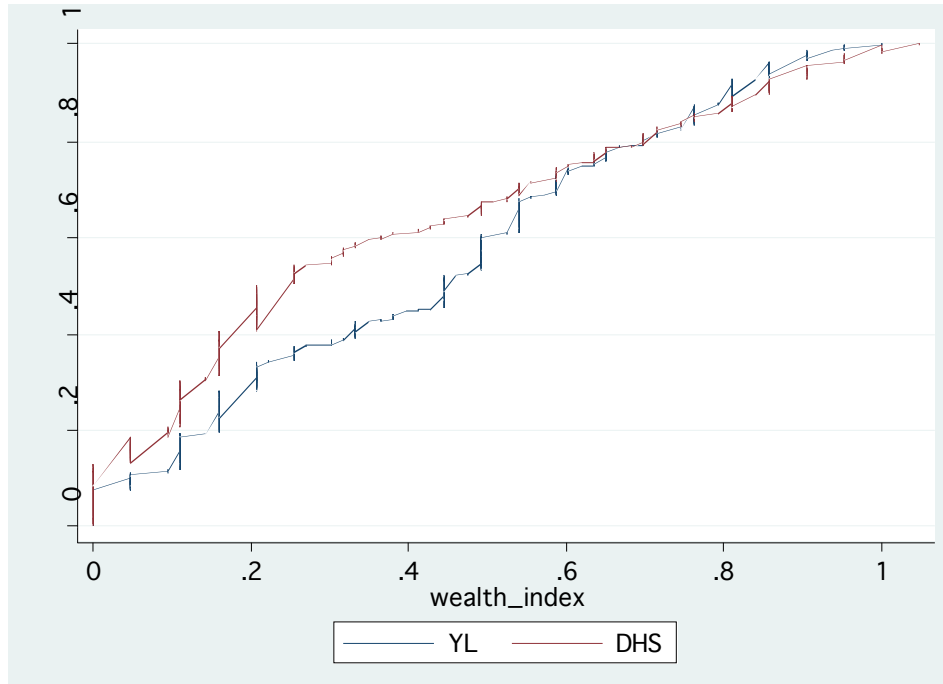
The difference in wealth index values between the Young Lives and the DHS sample should imply similar differences in the average amounts of assets possessed by households from both samples. We test this speculation with a t-test for equality of means for various household assets (see Appendix 1). For the majority of assets, the differences between DHS and Young Lives data are not statistically significant. However, we observe that significant more households in the DHS sample own their house compared to households in the Young Lives sample (89 per cent versus 82 per cent). In Appendix 4 we analyse the wealth index distribution function for Young Lives sample further.

3.2   Education, gender, ethnicity and age of household head

Figure 2 illustrates that a greater proportion of primary caregivers in the Young Lives sample completed primary school than in the DHS sample. The difference is statistically significant (see Appendix 2).

Figure 2. *Level of education attained by primary caregiver*
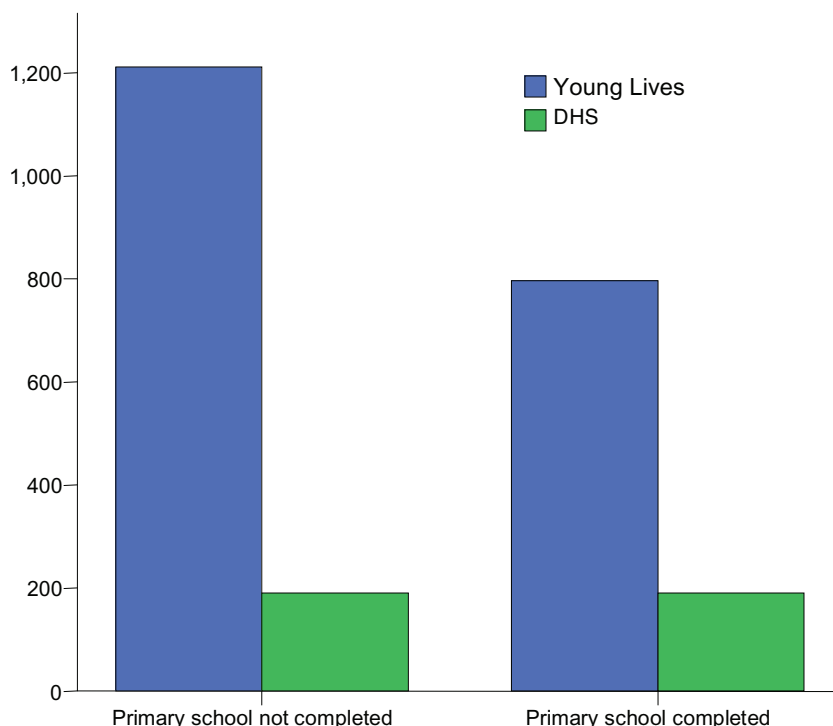
Concerning gender, we see that approximately 8 per cent of the households in the Young Lives data set are female headed while only 6 per cent of the households in the DHS data set are. Concerning ethnicity, we find a significant difference in the representation of Scheduled Tribes. While 15 per cent of the heads of household in Young Lives are from a Scheduled Tribe, only 7 per cent in the DHS are. We also find that there is a significant difference in the average age of household heads. The heads of household are much younger in the Young Lives data set than in the DHS data set. And while an average household is larger in the DHS, the absolute difference is relatively small.

3.3 Cooking fuel, source of drinking water and health care

Nearly 74 per cent of the Young Lives households rely on wood as main fuel source compared with 73 per cents of DHS households. Furthermore, a higher proportion of households in the Young Lives sample (19 per cent) use electricity for cooking than households in the DHS sample (16.9 per cent). A standpipe was the most frequently used source of drinking water for both Young Lives (59.4 per cent) and DHS (58.6 per cent) households. On the utilisation of maternal health care services, a higher proportion of women in the Young Lives sample (97.5 per cent) received two or more tetanus injections during pregnancy than in the DHS sample (95 per cent). Moreover, while women in the Young Lives sample had an average of 4.93 antenatal visits, women in the DHS sample had 5.49. Finally, while 56 per cent of women in the DHS sample had breastfed, 97.33 per cent of the women in the Young Lives sample had done so.

4. Conclusion

In this report, we have looked at whether and how the Young Lives data set is comparable with the DHS data set for Andhra Pradesh. The plot of wealth index cumulative distributions for the Young Lives and DHS samples shows that for most values of the wealth index, the Young Lives distribution is below the DHS distribution, i.e. fewer households with low wealth index are included in the Young Lives data than in the DHS sample. This finding is partly supported when we formally test for equality of means of the distributions for a range of living standard indicators between the Young Lives and the DHS samples. We find that Young Lives households appear to have better access to services compared with an average household in Andhra Pradesh. Young Lives households also appear to possess more non-agricultural and agricultural asset. However, at the same time, Young Lives households are less likely to own their house, primary caregivers are less educated, and mothers are less likely to breastfeed or receive antenatal care. Furthermore, we note that the different survey years of the Young Lives and the DHS samples imply that our analysis might have been affected by a reduction of poverty across Andra Pradesh between 1999 and 2005. This might have resulted in an underestimation of the degree of poverty in the Young Lives sample.

References

Dev, S. Mahendra, S. Galab, M. Gopinatha Reddy, K.S. Reddy, C. Ravi, K.S. Babu and G.K. Mitra (2002) APDPIP Baseline Survey Report in Sustainable Livelihood Framework, Hyderabad: CESS

Galab S., M. Gopinath Reddy, A. Piush, A. McCoy, C. Ravi, D.S. Raju, K. Mayuri, P. Prudhivikar Reddy (2003) *Young Lives Preliminary Country Report: India*, London: Young Lives

International Institute for Population Sciences (IIPS) and ORC Macro (2000) *National Family Health Survey (NFHS-2) 1998/99, India,* Mumbai: IIPS.

Mahendra Dev S and C. Ravi (2007) 'Poverty and Inequality: All India and States 1983-2005', *Economic and Political Weekly* 42(6) 10 Feb: 509-21 h

Young Lives (2007) 'Selection of Sentinel Sites and Communities in Andhra Pradesh', http://www.younglives.org.uk/pdf/indiasampling.pdf (accessed 24 March 2008)

Appendix 1

Statistical test for comparison of means

1. T-test for equality of means

We use the Welch-Satterthwaite t-test to test for equality of means of distributions of various variables. The Welch-Satterthwaite t-test is appropriate when the assumption that two populations have equal variances seems unreasonable. It provides a t statistic that asymptotically approaches a t-distribution, allowing an approximate t-test to be calculated when the population variances are not equal. The null hypothesis is that the difference in means of the variables is not significantly different from zero.2. Methodology for data analysis

1. Two copies of the DHS data set were made. For comparability purposes, the DHS sample was narrowed down to include only households with 1-year-old or 8-year-old child. Following this approach, we create two DHS sub samples.

2. Common variables were identified across the Young Lives and DHS data sets. They were recoded and renamed in the two data sets.

3. The two data sets were merged.

4. Statistical tests were run.

3. Recoded common variables used for the comparison

1. Cooking fuels: wood, kerosene, charcoal, gas/electricity, coal, cow dung, and other.

2. Sanitation facilities: flush toilet, pit latrine (household), pit latrine (communal), none, and other.

3. Source of drinking water: piped into dwelling, tube well in dwelling, public standpipe/common tap, unprotected well/spring/pond, and other.

4. Region: Coastal Andhra, Rayalaseema and Telangana.

Appendix 2

Comparison of Young Lives with the Demographic and Health Survey

Table 3. *Results of t-test for equality of means – households with 1-year-old child*

Variables	Young Lives	DHS	Difference in means (Young Lives – DHS)
Location (% of households, standard deviation)			
Rural	75	75	0
	(0.434)	(0.432)	
Coastal Andhra	34.81	34.48	0.0033
	(0.476)	(0.475)	
Rayalaseema	30.13	18.63*	0. 115
	(0.458)	(0.389)	
Assets (% of households, standard deviation)			
House	82	89*	-0.07
	(0.338)	(0.308)	
Sewing machine	11	10	0.01
	(0.312)	(0.297)	
Refrigerator	6	7	0.01
	(0.235)	(0.260)	
Television	30	33	-0.03
	(.457)	(.471)	
Motorbike	9	11	-0.02
	(0.293)	(0.313)	
Bicycle	30	40*	-0.1
	(0.457)	(0.491)	
Car/truck	1	1	0
	(0.097)	(0.096)	
Fan	55	55	0
	(0.498)	(0.498)	
Clock	63	64	-0.01
	(0.484)	(0.482)	
Agricultural assets (% of households, standard deviation)			
Tractor	3	1*	0.02
	(0.159)	(0.096)	
Farm equipment (pump, plough, etc.)	29	12*	0.17
	(0.453)	(0.325)	
Bullock cart	12	10*	0.02
	(0.331)	(0.297)	
Thresher	1	1	0
	(0.109)	(0.081)	

Household access to (% of households, standard deviation)			
Electricity	82	75*	0.07
	(0.383)	(0.432)	
Landline phone	6	4	0.02
	(0.237)	(0.207)	
Cooking fuel (% of households, standard deviation)			
Wood	74.4	73.6	0.008
	(0.436)	(0.441)	
Kerosene	5.52	7.04	-0.0152
	(0.228)	(0.264)	
Charcoal	0.1	0	0.001
	(0.0315)	(0)	
Electricity	19	16.9	0.021
	(0.392)	(0.375)	
Coal	0.25	0.26	-0.001
	(0.049)	(0.051)	
Cow dung	0.45	0	0.0045
	(.066)	(.44)	
Other	0.25	0.85	-0.006
	(0.049)	(0.134)	
Source of drinking water (% of households, standard deviation)			
Piped water	16.81	19.29	-0.0248
	(0.374)	(0.394)	
Tube well	7.5	1.9*	0.056
	(0.264)	(0.139)	
Stand pipe	59.4	58.6	0.008
	(0.491)	(0.492)	
Spring water	14.5	19.8*	-0.053
	(0.352)	(0.398)	
Sanitation facility (% of households, standard deviation)			
Flush toilet	18.58	19.02	-0.0044
	(0.389)	(0.392)	
Own pit latrine	7.22	5.02*	0.022
	(0.258)	(0.218)	
Community pit latrine	4.13	1.85*	0.0228
	(0.199)	(0.134)	
No toilet	69.52	74.11*	-0.0459
	(0.46)	(0.44)	
Past related (% of households, standard deviation)			
Livestock owned in the last year	44	39*	0.05
	(0.497)	(0.489)	
Recent death/reduction of members	7	15 *	-0.08
	(0.249)	(0.359)	
Whether caregiver has worked in last year	50	48	0.02
	(0.5)	(0.5)	

Household characteristics (absolute, standard deviation)			
Household size	5.42 (2.356)	6.17* (2.549)	-0.0075
Age of household head	23.79 (4.644)	41.55* (14.779)	-0.1776
% female headed households	8 (0.278)	6 (0.244)	0.02
Primary Caregiver characteristics (%, standard deviation)			
Completed primary school	40 (0.489)	50* (0.501)	-0.1
Hindu	87.47 (0.331)	85.6 (0.351)	0.0187
Muslim	7.6 (0.265)	8.8 (0.284)	-0.012
Christian	3.9 (0.194)	5.5 (0.229)	-0.016
Scheduled Tribe (ST)	14.57 (0.352)	7.23* (0.259)	0.0734
Other Backward Class (OBC)	45.9 (0.498)	45.38 (0.498)	0.0052
Married	99.25 (0.086)	98.69 (0.113)	0.0056
Divorced	0.3 (0.054)	0.7 (0.088)	-0.004
Pregnancy, delivery, and breastfeeding (absolute, standard deviation)			
Number of antenatal visits	4.93 (2.207)	5.49 * (3.732)	-0.0056
% Received antenatal care	88.11 (0.323)	5* (0.214)	0.8311
% Tetanus injection during pregnancy	97.65 (0.151)	95 (0.582)	0.0265
% Breastfeeding	97.33 (.161)	56* (.498)	0.4133
Place of delivery (%, standard deviation)			
Home	49.85 (0.5)	47.12 (0.499)	0.0273
Hospital	48.45 (0.499)	51.83 (0.50)	-0.0338
Private healthcare facility	45 (0.066)	79 (0.088)	-0.34
other	1.25 (0.11)	0.26* (0.05)	0.0099

* difference in means is significant at 5% level

Number of observations: Young Lives (n=2,011), DHS (n=383)

Table 4. *Results of t-test for equality of means– households with 8-year-old child*

Variables	Young Lives	DHS	Difference in means (Young Lives – DHS)
Location (% of households, standard deviation)			
Rural	75	75	0
	(0.433)	(0.432)	
Coastal Andhra	34.72	34.48	0.0024
	(0.476)	(0.475)	
Rayalaseema	30.46	18.63*	0.1183
	(0.46)	(0.389)	
Assets (% of households, standard deviation)			
House	85	89*	-0.04
	(0.355)	(0.308)	
Sewing machine	10	10	0
	(0.296)	(0.297)	
Refrigerator	5	7	-0.02
	(0.225)	(0.260)	
Television	33	33	0
	(0.470)	(0.471)	
Motorbike	7	11*	-0.03
	(0.263)	(0.313)	
Bicycle	30	40*	-0.1
	(0.457)	(0.491)	
Car/truck	1	1	0
	(0.083)	(0.096)	
Fan	57	55	0.02
	(0.496)	(0.498)	
Clock	67	64	0.03
	(0.472)	(0.482)	
Agricultural assets (% of households, standard deviation)			
Tractor	2	1	0.01
	(0.129)	(0.096)	
Farm equipment (pump, plough, etc.)	28	12*	0.16
	(0.451)	(0.325)	
Bullock cart	3	10*	-0.07
	(0.164)	(0.297)	
Thresher	0	1	-0.01
	(0.045)	(0.081)	
Household access to (% of households, standard deviation)			
Electricity	82	75*	0.07
	(.384)	(.432)	
Landline phone	6	4	0.02
	(.229)	(.207)	

Cooking fuel (% of households, standard deviation)			
Wood	72.52	73.6	-0.0108
	(0.44)	(0.441)	
Kerosene	5.4	8.1	-0.027
	(0.227)	(0.273)	
Charcoal	0.2	0	0.002
	(0.04)	(0)	
Electricity	21.23	16.91*	0.0432
	(0.409)	(0.375)	
Coal	0.2	0.1	0.001
	(0.044)	(0.442)	
Cow dung	72.52	73.6	-0.0108
	(.44)	(.441)	
Other	5.4	8.1	-0.027
	(0.227)	(0.273)	
Source of drinking water (% of households, standard deviation)			
Piped water	16.96	19.29	-0.0233
	(0.375)	(0.394)	
Tube well	8.33	1.98*	0.0635
	(0.276)	(0.139)	
Stand pipe	57.74	58.65	-0.0091
	(0.494)	(0.492)	
Spring water	15.77	19.82*	-0.0405
	(0.364)	(0.398)	
Sanitation facility (% of households, standard deviation)			
Flush toilet	20.44	19.02	0.0142
	(0.403)	(0.392)	
Own pit latrine	7.14	5.02	0.0212
	(0.257)	(0.218)	
Community pit latrine	3.77	1.85*	0.0192
	(0.191)	(0.134)	
No toilet	67.56	74.11*	-0.0655
	(0.468)	(0.438)	
Past related (% of households, standard deviation)			
Livestock owned in the last year	42	39	0.03
	(0.494)	(0.489)	
Recent death/reduction of members	11	15*	-0.04
	(0.311)	(0.359)	
Whether caregiver has worked in last year	71	65*	0.06
	(0.452)	(0.478)	
Household characteristics (absolute, standard deviation)			
Household size	5.54	6.17*	-0.0063
	(2.035)	(2.549)	
Age of household head	31.02	42.39*	-0.1137
	(6.39)	(12.25)	
% female headed households	8	6	0.02
	(0.269)	(0.244)	

Primary Caregiver characteristics (%, standard deviation)

Completed primary school	30	35*	-0.05
	(0.457)	(0.479)	
Hindu	87.50	85.60	0.019
	(0.33)	(0.351)	
Muslim	6.9	8.85	-0.0195
	(0.254)	(0.284)	
Christian	4.66	5.55*	-0.0089
	(0.210)	(0.229)	
Scheduled Tribe (ST)	11.01	7.23*	0.0378
	(0.313)	(0.259)	
Other Backward Class (OBC)	46.23	45.38	0.0085
	(0.498)	(0.498)	
Married	95.24	94.67	0.0057
	(0.213)	(0.224)	
Divorced	0.3	1.2	-0.009
	(0.054)	(0.109)	

* difference in means is significant at 5% level

Number of observations: Young Lives (n=1,008), DHS (n=444)

Appendix 3

Calculation of wealth index

The wealth index is intended to be the primary instrument to measure the socioeconomic status of a household. It draws on work undertaken by the World Bank and Macro International that developed a wealth index cited in the UNICEF Multiple Indicator Cluster Surveys.

Our wealth index is the simple average of three different indexes: Housing Quality Index (HQ), Consumer Durables Index (CD) and Services Index (S). Each of these is simple means of variables that are in a 0-1 range. Hence, the wealth index can take a value between 0 and 1, whereby a higher wealth index value indicates a higher socioeconomic status.

Housing quality index

- HQ1: Rooms per Person. Number of rooms divided by the number of household members. The HQ1 variable is set to take a maximum value of unity. Ratios higher than 1 are recoded accordingly.

- HQ2: Floor Quality. Has the value of 1 if the floor is made of a finished material (cement, tile or laminated material); 0 otherwise.

- HQ3: Roof Quality. Has the value of 1 if the roof is made of iron, concrete tiles or slates; 0 otherwise.

Consumer durables index

- Constructed from simple means of the following variables: ownership of (i) radio, (ii) bicycle, (iii) TV, (iv) motorbike or scooter, (v) motorised vehicle or truck, (vi) landline telephone, and (vii) a modern bed or a table.

Services index

- S1: Electricity. Has the value of 1 if the household has access to electricity; 0 otherwise.

- S2: Water. Has the value of 1 if the household's source of drinking water is piped into dwelling or yard; 0 otherwise.

- S3: Toilet. It takes the value of 1 if the household has access to its own pit latrine or flush toilet; 0 otherwise.

- S4: Cooking fuel. Has the value of 1 if the household uses electricity, gas or kerosene as cooking fuel; 0 otherwise.
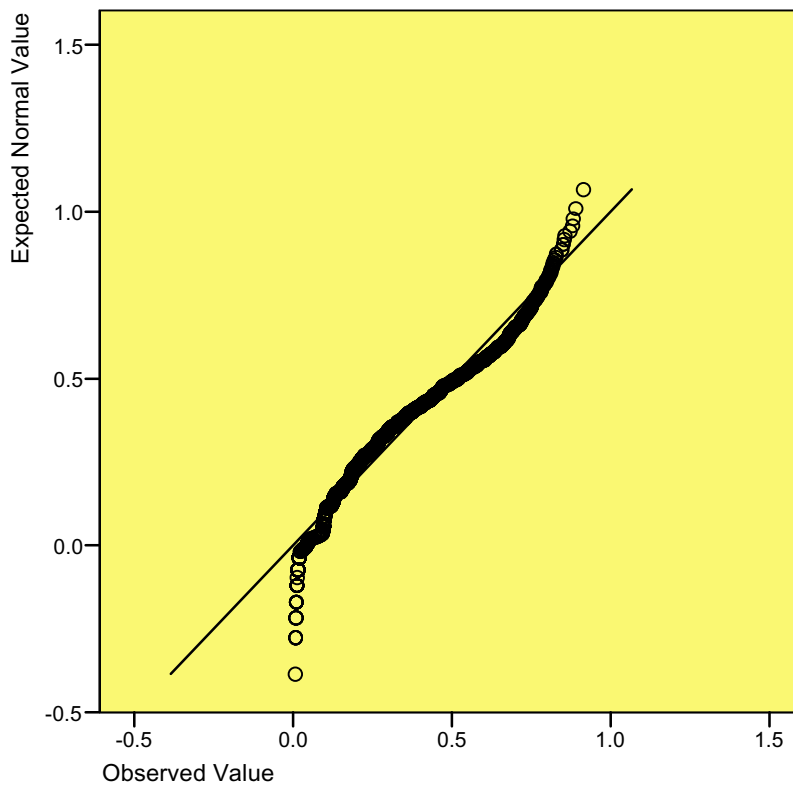
Appendix 4

Normal probability plot of the Young Lives wealth index

Figure 3: *Normal probability plot of wealth index of the Young Lives sample*

A normal probability plot or a normal Q-Q plot is the plot of the ordered data values against the associated quantiles of the normal distribution. For data from a normal distribution, the points of the plot should lie close to a straight line.

The above plot shows that the wealth index is normally distributed for the Young Lives data set for 1-year-old children. However, there are some outliers.

Normal Q-Q Plot of wealth index

Young Lives
An International Study of Childhood Poverty

3 Mansfield Road, Oxford OX1 3TB, UK
Tel: +44 (0)1865 289966
Email: younglives@younglives.org.uk

THE AUTHOR

Neha Kumra is a DPhil student in Economics at the University of Oxford. She holds a MSc degree in Economics for Development from Oxford University and a BA (Honours) degree in Economics from Delhi University. Her research interests include economics of education and health, labour economics and applied micro-econometrics.

Young Lives
An International Study of Childhood Poverty

www.younglives.org.uk

Department of International Development
University of Oxford
3 Mansfield Road, Oxford OX1 3TB, UK
Tel: +44 (0)1865 289966
Email: younglives@younglives.org.uk