

Section 14: Methods for Analysis

May 2017

As a longitudinal and policy-directed study comprising several distinct components, Young Lives entails many layers of analysis which rely on a variety of tools and approaches. Young Lives researchers use many different methods of analysis to discover and interpret patterns in the data collected in the field.

Units of analysis vary. Analysis of individual children allows researchers to build up detailed, contextualised life histories comprised of several types of data collected at different times, while analysis of groups of children, defined according to factors such as wealth, age, gender or location, is used to evaluate the impact of social differences on child outcomes. Although data cannot be directly compared between countries because of sampling differences, some analysis takes place at this level, highlighting how change over time is affecting the children in different countries, and comparing and contrasting the processes which are having most impact on their lives.

The structure of Young Lives, with its successive, linked cycles of survey and qualitative data collection, presents opportunities for cross-sectional and longitudinal analyses, and analysis often proceeds in both dimensions. Cross-sectional data are examined at a particular point in time, for example when a round of data is collected, and panel data are examined longitudinally across time points. Some variables are only suitable for one type of analysis.

The longitudinal character of Young Lives is a defining feature of the study. Longitudinal analysis focuses on the temporal dimensions of childhood poverty – change, process, continuity, transitions and turning points – and has many challenges. It is time-consuming and complex, involving very large quantities of data. There are particular difficulties in establishing causal relationships between different variables. Methods and software for longitudinal analysis are relatively undeveloped, especially for complex statistical models. Researchers also face the challenge of conducting longitudinal analysis of both quantitative and qualitative data in a way that produces policy relevant findings.

The complex design of Young Lives lends particular significance to defining the key terms and concepts used in the study, as these frame analysis, setting the boundaries of what can and cannot be concluded from the data. The most important are:

- **Poverty**, which is defined as multi-dimensional, complex and dynamic, subject to contextual specificity and multiple, interacting causes (Boyden and Dornan 2011). It is seen as comprising diverse material deficiencies, but also susceptibility to risk and constraint on choice.
- **Childhood**, which is seen as a distinct and universal life phase. Mirroring the holistic definition of poverty, children's development and well-being are seen as encompassing the material, physical, cognitive, psychological and socio-cultural domains. Children are viewed not as passive recipients of experience or victims of circumstance, but as social actors who have some agency in shaping their own lives.
- **Households**, which are understood as groups of people who live together, usually pool their income and eat at least one meal together when they are at home. Young Lives also recognises, however, that households are not necessarily cohesive units and that household members may act autonomously of – or in conflict with – each other.
- **Policy**, which is seen as non-linear and dynamic, and includes a full cycle of activities from conceptualisation to planning and implementation. Young Lives considers three broad types of policies: those that are specifically child-focused; general poverty reduction, development and social policies that have an impact on children; and those that shape the effectiveness of institutions that are responsible for child-related issues.

Statistical analysis of survey data

Establishing the reliability and validity of data is a keystone of good quantitative research. An early priority for statistical analysis of quantitative Young Lives data was therefore to carry out preliminary analyses of certain aspects of the whole dataset. These included:

- Analyses of the reliability and validity of the psychometric tests administered to Young Lives children. The reliability analyses examined the accuracy and stability of test scores using methods derived from Classical Test Theory and Item Response Theory. The validity analyses relied on estimating correlations between test scores and other variables such as child age and level of education, and checking these against empirical evidence reported in the wider literature on psychometric tests (Cueto et al. 2009).
- Attrition probit tests, which were used to analyse estimated models for anthropometric scores at age 1 and school enrolment at age 8 in order to enumerate any biases due to sample attrition. Further [statistical tests also examined whether attrition biases](#) might lead to biased inferences (Outes-Leon and Dercon 2008).
- Tests to compare [the Young Lives sample](#) with larger, nationally representative samples in order to be

transparent about intended and unintended sample biases. Methods included graphic comparisons of poverty rates, t-tests for the statistical significance of differences in living standards, and post-stratification, a technique used in survey analysis to incorporate population distribution of key characteristics into survey estimates (Escobal and Flores 2008).

Beyond these analyses of validity and reliability, the main purpose of statistical analysis of quantitative data is to interrogate the relationships between variables. In Young Lives, an important primary step towards the analysis of variables has been the definition and calculation of several composite variables from the raw data. These include anthropometric and other indices, but the most important is the Wealth Index, a composite, asset-based measure of economic well-being that is not production or location specific. The power of the Wealth Index is that it gives a measure of wealth which can be treated as a continuous variable and therefore be used for regression analyses and statistical modelling. Constructed from three other indices, the Wealth Index is used across countries and across survey rounds. It comprises the average of:

- a housing quality index, the simple average of types of flooring, roofing and walls, and the number of rooms per person.
- a consumer durables index, the scaled sum of ownership of a set of consumer durables, which includes up to three locally-specific items per country.
- a services index, the simple average of access to drinking water, electricity, toilet and fuel.

Classifying both children and households according to basic categories of difference such as wealth, location, age and gender is an important stage in generating simple statistics about the Young Lives survey data. These statistics form the foundation of country reports, produced for each round of the longitudinal survey to provide descriptive information about the whole sample and a starting point for cross-country comparisons and more in-depth analysis of particular thematic areas. The main method used to produce descriptive statistics is cross-tabulation, usually carried out using SPSS. The analysis presented in country reports has developed iteratively round by round:

- Round 1 reports provided baseline survey data in a standardised format. They contained descriptive statistics about the whole sample, and broken down by wealth and location.
- Round 2 reports presented analysis that was used to construct a profile of child poverty showing who and where poor children were, and the factors which were either keeping them in poverty or offering them a route out. They also presented preliminary analysis of the trends emerging between rounds.
- Round 3 reports were less uniform across countries, focusing more explicitly on patterns and themes that were emerging as important in each country. All, however, presented preliminary insights into trends over time, and the nature and extent of inequalities between children of different groups.

Since Round 2, country reports have also presented more complex statistics in the form of analysis based on regressions, mostly carried out using Stata software. Regressions are equations which represent the relationship between the values of two variables and allow the prediction of the most probable values of the outcome variable given the value of the observed variable. Regression analysis has the benefit of incorporating several relevant variables simultaneously, giving a better understanding of the processes at work. In Young Lives, different types of regression equation have been used to analyse different kinds of variable:

- A commonly used form of regression is Ordinary Least Squares (OLS), which is suitable for use when the outcome variable is continuous, such as a child's height or weight. For example, OLS was used to demonstrate that maternal education has a major positive effect on child nutrition in India, even when controlling for the significant effect of household resources (Galab, Reddy and Himaz 2008).
- A probit or logic model is used when the outcome variable can only be one or zero, such as whether or not a household is in poverty. For example, a probit model was used to show that parental education levels, mothers' ethnicity and household size are all statistically significant factors in explaining poverty in Vietnam (Le Thuc et al. 2008).
- An ordered probit regression is used when the observed variable has an ordinal value, for example when child well-being is measured by rating the child's place on a five-point response ladder. An ordered probit regression was used with this variable to show that children in rural Ethiopia who had a male head of household in Round 1 felt better off in Round 2, and that children in urban areas felt worse off compared to those in rural areas (Woldehanna, Mekonnen and Alemu 2008).

In some cases, numerous regression equations have been used to examine patterns emerging across all the Young Lives survey data. An analysis of the differences between boys and girls, for example, investigates gender gaps in all four countries across 13 determinants of child development including nutrition, education, subjective well-being and parental aspirations (Dercon and Singh 2011). The analysis used 156 regressions to analyse the determinants, each of which contained socio-economic and demographic controls, including the gender of the child.

While regression equations illuminate patterns and trends in the data, they do not explain why these phenomena happen. Establishing causality statistically first requires making a hypothesis that one variable causes a change in another variable, then testing this hypothesis through fitting an explanatory model to empirical data. Such models, composed of complex sets of equations, are used to make valid predictions reflecting causality. In a study examining a multi-dimensional phenomenon like poverty, however, such predictions are subject to biases due to unmeasured variables which may be influencing or producing the measured effect. Nonetheless, some analysis of causality has been undertaken using econometric models and

innovative approaches to controlling for bias. These include one study which finds that early childhood stunting has a significant effect on cognitive development in Vietnam (Le Thuc 2009), and another that finds that shocks and adverse events have more effect on the nature and amount of work done by children than levels of household poverty in Ethiopia (Heissler and Porter 2010).

Systematic analysis of qualitative data

Qualitative data are analysed using very different methods from the positivist, statistical approaches inherent in quantitative data analysis. The analysis of longitudinal qualitative data is a multi-stage process that relies on abductive and inductive methods of reasoning. Rather than focusing solely on either cross-sectional findings or differences between two points in time, it aims to illuminate the trajectories or pathways individual children are following, and understand how different factors interact to shape them.

Most Young Lives qualitative data take the form of text, although they also include the pictures, diagrams and photographs that are the [outputs of some data collection methods](#). The majority of the text comprises transcripts of interviews, but also includes reports on group activities and texts such as diaries produced by the children themselves. Each text file is transcribed and recorded in a common, collectively agreed format which allows it to be identified by date, research method and participants.

The principal method for organising these data to facilitate systematic longitudinal analysis is the coding and categorisation of text sections according to a mutually agreed framework. Once the framework for each round is finalised, each section of text is coded according to its content using the qualitative data analysis software Atlas.ti. Once coded, the whole dataset can be searched according to different codes and intersections of code, facilitating systematic analysis. The first stage of analysis in each round tends to be descriptive, summarising the data and highlighting emerging themes, patterns and connections, while subsequent stages reflect the specific focus and research questions developed by each country team.

The principal divisions in this framework, known as ‘super-families’, follow the core themes of the longitudinal qualitative research component, which are services, transitions and changes, and child well-being. Each super-family is sub-divided into ‘families’, topics which are derived from the main category. In the case of the transitions and changes super-family, for example, the family categories are expectations, relationships, educational transitions, and other transitions.

All country teams use the same framework of super-families and families but a third level of the framework, comprising detailed sets of codes within each family, is unique to each country. In Peru, for example, the educational transitions family was further sub-divided into communications and exchange between teachers, communications and exchange between teachers and parents, school organisational arrangements, school material resources, teacher training, adaptation, attitudes, expectations, support, children’s

readiness for school, children’s feelings, importance of preschool, and difficulties (Crivello et al. 2013).

Analysis can be approached either by theme or by individual:

- A ‘horizontal’ reading of the data is undertaken through the themes and sub-themes identified in the coding framework, enabling the identification of trends, similarities and differences between and within the research sites.
- A ‘vertical’ reading of the data entails starting with selected individual children and gathering all the available information about them in order to construct a detailed, composite image of their lives, and to triangulate data collected at different times.

Making such horizontal and vertical readings was relatively straightforward with a single round of data, but has become successively more complex as subsequent rounds of data have been collected. Research teams agree strategies for summarising longitudinal data. Tables and matrices constructed to track changes in the lives of case study children over time are an important tool for condensing data, guiding researchers to extract coded material from previous rounds. Case histories are constructed by examining all interviews with one child over successive data collection rounds, dividing the data into different domains such as education, work or risk, and creating a narrative or storyline for each domain.

Analysis of mixed methods data

One of the strengths of Young Lives is that it approaches the study of childhood poverty using a variety of research methods which generate both qualitative and quantitative longitudinal data. One of the major challenges faced by researchers is how these different types of data can be analysed and used together, given their foundations in different theories of knowledge and different understandings of validity and credibility. In practice, this challenge is met in different ways at different levels of the study:

- At the level of the two main longitudinal data collection components. For example, the research framework and [design of the longitudinal qualitative research component](#) were shaped by the findings from the two survey rounds that had already been carried out. The sub-sample for qualitative case study research was derived from the full survey sample to facilitate a clear linkage between the two datasets.
- At the level of country-specific and thematic analyses. For example, in a study of healthcare user fees in Ethiopia (Barnett and Tefera 2010) researchers analysed Round 2 survey data using cross-tabulations, chi-square tests and t-tests to produce descriptive statistics about user fees. They then used this analysis to develop a question guide for household interviews about different aspects of their experiences with user fees, which was piloted and modified in one community before being applied more widely. Once the interviews were transcribed and translated, content analysis methodology was used to develop a framework of topics

and sub-topics which formed the basis for coding, cross-referencing and analysis.

- At the level of meta-analysis for policy and communication purposes, a policy paper on gender inequalities (Pells 2011) draws together survey and qualitative data from across the study to analyse the gendered differences between boys and girls in education, domestic life and subjective well-being. The paper first identifies key questions in the policy debate on gender. It goes on to present descriptive statistics as the foundation of a narrative about how gender dynamics differ when children are at different ages, and how they

accumulate over time. This narrative uses quotations from children and condensed life histories drawn from the qualitative research, to illustrate processes and practices of everyday life that lie behind and help to explain the numerical patterns. It also, however, offers some analysis of qualitative data at country level, describing the six principle obstacles children face in meeting their goals that emerge from analysis of the qualitative data collected in India and Ethiopia, and draws on multiple regression analysis of gender bias in outcome indicators (Dercon and Singh 2011). Finally, each type of data analysis contributes a composite policy message that responds to the opening question.

REFERENCES

- Barnett, I. and B. Tefera (2010) *Poor Households' Experiences and Perception of User Fees for Healthcare: A Mixed-method Study from Ethiopia*, Working Paper 59, Oxford: Young Lives.
- Boyden, J. and P. Dornan (2011) *Putting Children at the Centre of Poverty Debates*, Policy Brief 12, Oxford: Young Lives.
- G.Crivello, V. Morrow and E. Wilson (2013) *Young Lives. Longitudinal Qualitative Research: A Guide for Researchers*, Technical Note 26, Oxford: Young Lives.
- Cueto, S., J. Leon, G. Guerrero and I. Muñoz (2009) *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 2 of Young Lives*, Technical Note 15, Oxford: Young Lives.
- Dercon, S. and A. Singh (2011) *From Nutrition to Aspirations and Self-Efficacy: Gender Bias over Time among Children in Four Countries*, Working Paper 71, Oxford: Young Lives.
- Escobal, J. and E. Flores (2008) *An Assessment of the Young Lives Sampling Approach in Peru*, Technical Note 3, Oxford: Young Lives.
- Galab, S., P. Prudhvikar Reddy and R. Himaz (2008) *Young Lives. Round 2 Survey Report: Andhra Pradesh, India*, Oxford: Young Lives.
- Heissler, K. and C. Porter (2010) *Know Your Place: Ethiopian Children's Contributions to the Household Economy*, Working Paper 61, Oxford: Young Lives.
- Outes-Leon, I. and S. Dercon (2008) *Survey Attrition and Attrition Bias in Young Lives*, Technical Note 5, Oxford: Young Lives.
- Pells, K. (2011) *Poverty and Gender Inequalities: Evidence from Young Lives*, Policy Paper 3, Oxford: Young Lives.
- Le Thuc, D. (2009) *The Effect of Early Age Stunting on Cognitive Achievement among Children in Vietnam*, Working Paper 45, Oxford: Young Lives.
- Le Thuc, D., N.P. Nguyen, M.C. Tran, V.T. Nguyen and T.S. Von (2008) *Young Lives: Vietnam Round 2 Survey Report*, Oxford: Young Lives.
- Woldehanna, T., A. Mekonnen and T. Alemu (2008) *Young Lives: Ethiopia Round 2 Survey Report*, Oxford: Young Lives.