

# Young Lives:

## A Case Study of Sample Design for Longitudinal Research

**Ian Wilson**

University of Reading

**Sharon RA Huttly**

London School of Hygiene and Tropical Medicine

**Young Lives**   
An International Study of Childhood Poverty



# Contents

<b>Acknowledgements</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>1. Introduction</b>	<b>3</b>
<b>2. Case study setting</b>	<b>4</b>
<b>3. Deciding how many children to sample</b>	<b>5</b>
<b>4. Deciding who to sample</b>	<b>6</b>
<b>5. Deciding how to sample</b>	<b>7</b>
<b>6. Deciding how to choose the sites</b>	<b>9</b>
<b>7. Semi-purposive sampling of sites</b>	<b>10</b>
<b>8. Sampling within each site</b>	<b>11</b>
<b>9. Sampling for useful analyses</b>	<b>12</b>
<b>10. Concluding remarks</b>	<b>14</b>
<b>References</b>	<b>15</b>



# Acknowledgements

The authors are grateful to the UK Department for International Development (DFID), which funded the first phase of the Young Lives project (ESCOR 7874 ESC 9800 616/532/001), and to the numerous members of the Young Lives team who contributed to the ideas contained in this paper, during numerous meetings and workshops.

# Abstract

This paper presents a case study to illustrate the range of decisions involved in designing a sampling strategy for a complex, longitudinal research study. It is based on experience from the Young Lives project and identifies the approaches used to sample children for longitudinal follow-up in four countries. The rationale for decisions made and the resulting benefits, and limitations, of the approaches adopted are discussed. Of particular importance is choice of sampling approach to yield useful analysis and particular examples are presented of how this informed the design of the Young Lives sampling strategy.

# I. Introduction

Designing the sampling strategy for a research study always involves a series of decisions and finding a balance between competing needs. A common example is the need for a sample size sufficiently large for robust statistical analysis, but limited by project resources: often this limits the scope of enquiry. Approaches to sampling as well as sample sizes are related to situational factors and project priorities.

A plethora of books and other resources exist to assist in designing sampling strategies yet they will rarely give one the entire answer to a particular scenario. Instead carefully considered decisions have to be made. However, perhaps because there are many other interesting issues at the study design stage, or perhaps since there is no "right answer" to sampling decisions, this is rarely a popular topic with students, teachers or researchers. We believe that case studies can be a useful mechanism for illustrating what is often considered to be a rather dry topic. We present here a case study of designing a sampling strategy for the Young Lives multi-country longitudinal study which because of its complexity presented the project team with some interesting questions.

## 2. Case Study Setting

Young Lives is a collaborative study investigating childhood poverty in four countries - Peru, Vietnam, Ethiopia and India (Andhra Pradesh state) - through follow-up of cohorts of children over time. The study is collecting data on a core set of child welfare indicators and their determinants in all countries, and on country-specific issues identified by researchers, government, policy-makers and other key stakeholders in each country. Funded by the UK's Department for International Development, project partners include academic institutions and Save the Children UK in each of the four countries, the UK and South Africa. Involving these different types of partner, the project aims to ensure effective linkage between the collection of primary data and its utilisation in advocacy and evidence-based policy formulation.

The project has 3 principal broad objectives:

- producing good quality long-term panel data about the changing nature of the lives of children growing up in poverty
- tracing linkages between key policy changes and child welfare
- informing and responding to the needs of policy makers, planners and other stakeholders

There is also a substantial education and media element, both in the countries where the project takes place, and in the UK.

At the heart of the study lies a cohort of children in each country who will be followed up every 3-4 years until 2015. In late 2002, the first round of data collection recruited 2000 children in each country, aged 6-17 months, to form the main cohort. In this first round, extensive data were collected to facilitate tracking of the children in future rounds.

Taking a livelihoods perspective, Young Lives sets out to trace the effects of factors acting at different levels - individual, household, community, regional and national. The project includes monitoring the social and economic policy environment and tracing the effects of macro-and meso-level policy implementation on communities and groups of individuals within them.

Further details on the rationale for the study and on its design can be found in Harpham (2002), Harpham et al (2003) and on <http://www.younglives.org.uk>



### 3. Deciding how many children to sample

The decision to enrol 2000 children in each of the four countries was made at a very early stage in the project's conception. The number has a limited statistical basis, however we justify this on the grounds of the study's objectives. Firstly the project has very broad objectives with an emphasis on producing numerous profiles, rather than a handful of simple estimates, which make it inappropriate to focus on particular indicators for formal sample size calculations. By "profiles" we mean relatively complicated summaries which take into account combinations of measurements and the structure of the sample e.g. sets of two-way tables calculated separately for interesting sub-sets of the entire sample. Secondly, experience with this type of study in the UK and other countries has shown that they tend to evolve over time in terms of their focus at different stages, responding to current interests and needs. The UK's National Child Development Study, for example, which started in 1958, has addressed a range of important health and development issues only some of which would have been predicted in the early planning stages. This too militates against a detailed statistical basis for the initial sample size.

Ultimately we had to select a number which appeared feasible to manage, was possible within the study's budget, and which we considered would yield sufficiently large numbers for statistical analysis in a general form. For example, based on the full 2000, the 95% confidence interval for a percentage will be approximately  $\pm 2\%$ . This is the accuracy which we might attribute to our estimate of the prevalence of a Yes/No outcome measure from an entire country sample. This sample size allows the detection of moderate-sized differences in such quantities between subgroups when the subgroups are at least 20% of the overall sample, say. It allows substantial cross-tabulations using the whole sample in each country and for some analyses of selected subsets with a reasonable expectation that fairly common categorisations will be represented by a sample size of a magnitude in the hundreds.

We also had to be clear about what such a sample size would prohibit - rare features such as mortality will not be captured with any meaningful degree of precision. Increasing the sample size to cover such outcomes would have had far-reaching consequences for the project design, such as limiting to one country, owing to financial limitations. More importantly, however, it was accepted that such measures were not central to the study, and could be obtained through large-scale national surveys such as the Demographic and Health Surveys, while many other important and less-researched issues could be adequately captured with a sample of 2000. Arguments are elaborated below, especially in section 9, as to why it is far too crude to treat sample size determination just in terms of claims to statistical significance: given the complex structure of our sample, the claim relates to a generalisation of a rather specialised population; given the complex objectives of the study, we would be perverting the carefully-negotiated requirements of Young Lives if low-priority "overall" prevalence estimates were the sole consideration in fixing sample sizes.

Finally, planning for the project began in the year 2000 and it was initially known as Children of the Millennium. The promotional possibilities of linking the sample size to the milestone year also contributed to the final decision, albeit not in a primary way.

## 4. Deciding who to sample

The initial aim of the project was stated as "to measure and find out about what happens to children born into poverty in the millennium". This was modified in later discussions to include the aim of capturing children moving into poverty as well as out of it. Nevertheless the focus is on the poor and results from general population sampling will not give sufficient attention to very poor children, and will not best serve to justify and bring about changes in attitudes and agendas which will favour these children.

On the other hand, there were at least four important reasons for not only sampling poor children:

- to provide more powerful arguments by comparing information on poor children with others
- to capture children moving into poverty
- to help avoid dismissal of results by potential users as being "not representative"
- to produce estimates which will more easily be able to be compared between the four countries

Thus although a sampling method designed to represent equally the whole national population of children born in the qualifying period was not considered appropriate, a method which over-samples the poor but which enables, where necessary, "national estimates" through weighted averages, is more suitable.

## 5. Deciding how to sample

A key need for the study's objectives was to obtain data at different levels - the children, their households, the community in which they resided, as well as at regional and national levels. This need thus determined that children should be selected in geographically compact sites rather than randomly selected across the country. There was, however, a much more important reason for recruiting children in "clusters" - the sites are also intended to provide suitable settings for a range of complementary thematic studies. For example, one or a few sites may be used for a qualitative study designed to achieve a deeper level of understanding of some social issues, either because they are important in that particular place, or because the sites are appropriate locales to investigate a more general concern. The quantitative panel study is seen as the foundation upon which a coherent and interesting range of linked studies can be set up.

Thus we decided on a design, in each country, comprising 20 geographic sites with 100 children sampled in each one, for the following reasons.

### **Why 20 sites?**

Returning to the project's objectives, the set of sites should serve to illuminate, and to a limited extent maybe typify, the settings in which the impact of policymaking and policy implementation can be examined. There is a vast range of ways in which the community setting can act as a mediating factor in the effect of macro-level events on the community, household and individual. Examples are the procedures followed by state or local administrations, as well as infra-structural, agro-ecological, communal and other factors. Teasing out what is germane from the mass of potentially 'explanatory' information requires concentrated commitment from someone with well-developed social understanding, local knowledge and a commitment to the project over lengthy periods of time. The above supports our decision that a comparatively small number of sites should be sampled, because they should be studied intensively.

### **Why 100 children per site?**

The study design includes thematic projects in addition to the main survey of the cohort. These might include qualitative studies such as the collection of individual narrative accounts of a displacement episode due to a dam or some such 'development' project; this would of necessity focus on one geographic site. Such studies run the risk of being dismissed as 'just' case studies, but we intend that by linking these to the survey data they can be contextualised and enriched by the existence of child, household and some community level information on each site. We believe that sites which are capable of yielding 100 compliant households are big enough to provide settings for many site level in-depth studies, but not so big that the thematic project researcher would have great difficulty tracing the index children if necessary.

Comparisons between sites will be a significant element of analysis and that objective strongly supports the supposition that sample sizes at each site should be about equal –100 index children in our case. For a quantitative, univariate measure being compared between two

groups, the t-test formula is the basis of a quite trivial demonstration that equal sample sizes in each group are more effective than unequal ones, in terms of precision. An ordinary-language version of this argument is that if one group is poorly characterised by a small, weak sample it can only provide a limited comparison with any other group, however much detail is available on the latter.

Finally, the longitudinal nature of the project means we must ensure that even after some degree of attrition there will be ongoing contact with a sample of plausible size in later phases of data collection. This militates against any smaller starting number than 100 per site.

## **What's in a name?**

It is tempting to think of these geographically compact sites in terms of the statistical procedure of cluster sampling, but this is a less than relevant paradigm. Cluster sampling theory is largely developed for one-off cross-sectional studies, and has little or nothing to say about repeated measurements through time, or about compensating for anticipated attrition through losses to follow-up whether this is due to respondents moving, or their refusing to continue as participants. Statistical considerations about cluster size are driven by the assumption that accurate estimation of one quantitative determinand is the primary focus of interest.

Our initial use of the term clusters when designing the study caused some confusion and expectations not relevant to the purpose of the study, among collaborators and stakeholders consulted. We switched to using the term sites which is in (much closer) analogy with the established terminology of sentinel site surveillance. As will be seen in a cursory web search, these are mostly used in public health applications. This procedure involves the selection of a relatively small number of settings which are then regularly studied in a consistent way. Typically (Last, 2001), this is referred to as surveillance if the observations are frequent, or monitoring if they are at longer intervals as in our case, where quite complex processes of maturation and change could take place between visits which may be some years apart.

The analogy is not perfect. In many health examples, sentinel site surveillance collects relatively simple data to provide trends in public health indicators. In contrast, with each participant in a panel study, there is a much more ambitious data collection agenda, and a greater concentration on retaining exactly the same panel members in successive waves of data collection. In particular, our concern with the livelihood trajectories of the children as they grow makes respondent substitution quite impossible.

## 6. Deciding how to choose the sites

Clusters are usually not regarded as being units worth reporting on individually, and indeed they are assumed in general statistical theory to be chosen at random. In our study, random selection of sites was not deemed appropriate for the following reasons:

- its benefit lies in situations where relatively simple variations in units of observations can be 'averaged out' by a sufficient sample size. With 20 sites, we are not close to the sample size that would be required to generate such benefits at the site level of sampling.
- on the other hand, the very terminology of 'sentinel' sites suggests, quite appropriately, that each is chosen to provide a view of a particular aspect, and that the reports on a consistent basis from each site are worthwhile and interesting in their own right. If effectively justified, that feature displaces the desideratum of randomness of selection.
- simple random sampling does not take account of information about the units sampled, but in selecting, say a commune from a Vietnamese province, a great amount and diversity of information is available and ought to be intelligently used. There is far too much such data for it to be effectively utilised just by stratifying a sample of size 20.

Reconciling our decision to over-sample the 'poor' with the particular needs for site-specific data resulted in a semi-purposive selection of sites, followed by the equivalent of random sampling of households within a site. Sampling sites rather than households on the basis of being categorised as 'poor' or not had other practical considerations. The fieldwork process within sites requires field staff whose skills must include developing good relations with community leaders, engaging potential households with the study, accurate recording of data, and establishing a robust set of tracking information so the panellists can be found when needed. It seemed inappropriate to expect that the same field staff should be required to draw complex, perhaps invidious, and even socially divisive distinctions between households as 'poor' or 'not poor'. For example, this might hamper recruitment, and it would certainly impose a severe training demand. Our decision was thus to sample with respect to 'poor'/'not poor' at the level of site selection, and within sites to sample across the board by a simple, practical, but objective, process.

## 7. Semi-purposive sampling of sites

A qualitative approach is needed to weigh up and balance the complex information about potential sites, but the sampling will carry little credibility with target audiences if it appears to have been ad hoc or whimsical. The first strand of our approach was to utilise in-country experts. Country partners have advisory committees whose members are people with well-established reputations, concerned with poverty matters and aware of current ideas, e.g. of poverty as a multi-faceted issue. These experts discussed key factors to cover in the selection of sites, such as geographical location and type, or particular population sub-groups. The achievement of an expert consensus in such a group proved to be stimulating and less difficult than might have been initially thought. This process also served to raise awareness of the project, and the potential of its data, in circles where results must eventually be appreciated and used. It contributed to establishing local ownership and interest in project development.

Secondly, each country developed a clear description of the protocol by which the selection of sites was made. A guiding principle was that if the process is adequately described, it should be possible for an independent researcher to replicate the same process and come out with a sample having almost all its major characteristics in common with the one chosen by the first team. A primary target was that most sites should represent 'poor areas'. There is no universally accepted definition of 'poverty' or of 'poor areas': classifications in common use in developing countries are often crudely money-metric and neglect most facets of poverty highlighted for example in livelihoods frameworks such as DFID (2002, 2003). Thus all we can say is that we selected 'poor areas' according to a particular protocol with an operational definition of 'poor areas'. This operational definition might be articulated as "the kind of area you would find if you followed the same protocol we used, with the same quality and breadth of expertise that our team brought to the task". This is discussed in Wilson (2001). Of course until resources materialise to try replicating the procedures, the plausibility of this type of argument depends entirely on its verbal description, and the standing of the experts concerned.

Countries varied in the availability of existing information which could assist in the sampling of sites. All four were able to construct some kind of poverty ranking of sites which contributed to selection and the details for each country can be found in the Preliminary National Reports accessible at <http://www.younglives.org.uk/>.

## 8. Sampling within each site

Household selection within sites operates in quite a different way from site selection. The basic sample of 2000 children and their households constitutes a large number of people who will be personally unknown to any policy-maker or central data analyst. To generalise from the data they provide, they must be sampled in an objective way, sufficiently well documented that it is unarguably clear what they do represent. This is the underlying aim of randomness of sampling.

It is recognised that at any geographic scale, e.g. within our sites, there will be a mixture of levels and types of poverty. However, we are taking a longitudinal perspective and acknowledge that manifestations of poverty may be episodic or transient, or may display either upward or downward trends. If we restricted attention to the poorest households at the start we would not be able to observe downward movements. In search of improved understanding of what drives these changes, it seems suitable to take the equivalent of a random sample within sites, rather than one driven by our pre-conceptions, which would probably be limited to rather crude categorisations of households' current poverty status.

The practicalities of fieldwork vary from place to place as there are substantial topographical and administrative differences from site to site within and between countries (e.g. there are differences, very relevant to the project's aims, between the slums of Hyderabad and areas of Andhra Pradesh largely inhabited by scheduled tribes). Choosing and walking along city streets may be the only possibility in densely populated areas, while procedures used in ecology such as line transects may suit some peri-urban or rural sites with lower density, with distance sampling methods used in low-density areas.

Partners have agreed across countries on principles for household sampling, rather than details. The exact procedures have been adapted to utilise reliable social structures and information sources in locally-appropriate ways, again with careful documentation of protocols so as to ensure:-

- a sample indistinguishable in composition from one drawn at random from qualifying households
- cost-effective field procedures for traversing each site
- reasonable control of biases, for example due to the unavailability of any respondent from a household during the listing sweep through the site, e.g. by using neighbour information

## 9. Sampling for useful analyses

There can be no serious claims that a sample is adequate except in relation to specific analysis objectives. Having described the sample selection above, we recap here on how the design selected addresses the information needs of the Young Lives study.

The sample in each country is of modest rather than large size and has not been chosen to be directly nationally representative. Of course, larger sample sizes allow greater disaggregation of samples into more tightly defined subgroups: the Young Lives study will face some restrictions, but we have satisfied ourselves that there are important general findings for which the samples will be quite adequate. There are several broad levels and types of analysis with different needs.

### Separate sites

At the site level itself we need to describe the survey results and information from any thematic projects, as well as site-level information which may include local events or trends resulting from implementation of meso- or macro-level policy decisions. Much of the analysis done on individual sites will be quite basic description, but we will certainly be interested in more sophisticated work which will require the combined or integrated results from different sources, qualitative or quantitative, at site level. This requires careful attention to linking relevant pieces of information together, for example at individual/household level.

### Comparison of sites

If one or a few sites are compared with another set of sites, there may be some numerical values where the statistical significance of differences can be assessed, but in many cases the comparisons will be more holistic, that is taking many factors into account. This requires consistency of recording practice between sites for any information which may be involved in such comparisons. Ideally it also leads to equal amounts of information or equal sample sizes in each of the entities being compared.

### Summary over sites

At regional level, or for the whole state, there will be needs for summative analysis, i.e. the development of overall regional or national conclusions. Where these relate to quantitative estimates, component results can be weighted to take account of the different amounts of information from different places, different sector sizes etc. A similar but less formal process is needed for qualitative information. This requires an understanding of how big or important the components are, and ideally means samples (perhaps numbers of sites) are proportional to the component importance.

The last sentences of the two previous paragraphs propose contradictory requirements: a compromise is always necessary, having regard to which analysis purpose is most important. The design of the Young Lives project makes the information collected likely to be more powerful at the first two of these levels, i.e. for separate sites and for comparisons. The nature



of the sampling and the scale of the work involved make Young Lives quite limited as a source of national-level summative information. Young Lives is intended much more as an in-depth study of relationships between pieces of information, rather than an instrument to collect national statistical results. If the nature of the sampling was adjusted to make this latter the priority, the scale of the funding we have – and therefore of the work involved – would still make it impractical to claim that we can generate usefully accurate and fully "representative" national level data, and in the adjustment we would have lost the precious opportunity to do detailed work within sites. In relation to summative analysis, we have accepted as an inevitable consequence of its scale that this project has much less potential than, for example, the UK Millennium Cohort Study (<http://www.cls.ioe.ac.uk/Cohort/MCS/mcsmain.htm>). Nevertheless, the longitudinal element of this study suggests that in years to come there will be outputs which illustrate the life trajectories of those initially recruited as one year old index children, integrating qualitative and quantitative data. Some of the most powerful results of this type are likely to be of a qualitative character, dependent more on the integrity and intelligence of the observation and case selection than on a large mass of quantitative data.

## 10. Concluding remarks

This case study illustrates how sampling decisions have to take account of existing information and intended project analysis, as well as balancing practicalities about logistics with statistical considerations. This led to selecting a sentinel site monitoring approach in the Young Lives project.

The large ‘anonymous’ sample at household level was everywhere selected by fully-documented and objective methods closely analogous to statistical random sampling, but adapted to local settlement patterns.

The small sample of 20 sentinel sites per country acknowledged that random sampling would disregard the very extensive and diverse information available about potential sites. Instead clear, detailed, and we believe repeatable, protocols were written up to describe the structured sequence of decisions made in selecting and defining sites. The protocols and the resulting selection of sites were exposed to expert scrutiny in brain-storming sessions with relevant experts including potential critics and official users, and their endorsement was obtained for the decisions made.

Procedures for over-sampling the poor were systematised and documented in each country. They were confined to the site selection stage to avoid biases and difficulties through inappropriate, maybe invidious decisions by field staff at the within-site household selection level.

This process would generate samples reasonably representative of household wealth status within individual sites. It did not predetermine actual numbers of children in the sample whose households were ‘in poverty’, but local knowledge and pilot studies assured us these would be well-represented. In any event, existing information would in most cases only have served to predetermine ‘poverty’ with respect to narrower and weaker, largely money-metric, definitions than that favoured by Young Lives.

It is freely acknowledged that Young Lives is quite limited as a source of national-level summative information. Within the project, other objectives are prioritised over this one: primary goals are the development of in-depth understanding at site level and carefully-structured comparisons amongst sets of sites. The sampling plan is geared to the top priority objectives.

## References

DFID (2002) *Sustainable livelihoods guidance sheets* available at [http://www.livelihoods.org/info\\_guidancesheets.html](http://www.livelihoods.org/info_guidancesheets.html)

DFID (2003) *Sustainable livelihoods distance learning guide* available at [http://www.livelihoods.org/info\\_distanceLearning.html](http://www.livelihoods.org/info_distanceLearning.html)

Harpham, T. (2002) Measuring child poverty and health: a new international study. *Journal of Tropical Paediatrics*, **48**, 128-129.

Harpham, T., Huttly, S., Wilson, I.M. and de Wet, T. (2003) Linking public issues with private troubles: panel studies in developing countries. *Journal of International Development*, **15**, 353-363.

Last, J. (2001 4th edition) *A Dictionary of Epidemiology*. (Oxford: Oxford University Press).

Wilson, I.M. (2001) *Sampling and qualitative research* available at [http://www.rdg.ac.uk/ssc/develop/dfid/booklets/tp2\\_samp.pdf](http://www.rdg.ac.uk/ssc/develop/dfid/booklets/tp2_samp.pdf)

