

REGIONAL AWARD FOR INNOVATION IN STATISTICS

NIÑOS DEL MILENIO - PERU

YOUNG LIVES – AN INTERNATIONAL STUDY OF
CHILDHOOD POVERTY

TECHNICAL BRIEF



EXECUTIVE SUMMARY

Young Lives is an innovative long-term international data gathering and research project investigating the changing nature of childhood poverty. By making publicly available the information gathered, the project seeks to improve understanding of the causes and consequences of childhood poverty and to examine how policies affect children's well-being inform the development and implementation of future policies and practices that will reduce childhood poverty. Since 2002, the project is tracking the development of 2,860 children in Peru through quantitative and qualitative data collection methods and through research over a 15-year period. The project collects information at the child, household and community level covering a range of issues that determine and affect the welfare of children. Young Lives is tracking a representative sample of Peruvian children from two cohorts: a younger cohort who were aged between 6 months and 17 months in 2002 (born 2000) and an older cohort of children aged between 7.5 years old and 8.5 years old in 2002 (born 1994). The project is largely funded by DFID and is part of a multi-country project that includes gathering comparable information in Ethiopia, India (Andhra Pradesh) and Vietnam.

The depth and extent of the Young Lives database is unique. No longitudinal research of this size, scope and complexity has ever been undertaken in the developing world. The project not only collects process and gives access to quantitative data associated with child poverty, but also collects qualitative information that allows a very rich and in-depth analysis of children's lives and how they are affected by poverty and government policies. The high quality of the data is ensured through a meticulous process of piloting, translation and back translation (in several languages), double data-entry, checking and cleaning.

Been a longitudinal project, tracking activities, identifying the network of the families involved in the study and maintaining contact with children and their families is critical to assure the integrity of the cohort. Our attrition rates are very low in absolute terms (0.74% per year), indicating that our approach is an effective one. Indeed, compared with other longitudinal project, we can contend that the project has succeeded in minimizing attrition, with rates that are probably the lowest in the world.

Despite the numerous challenges we face to ensure the accuracy and validity of a very large – and growing – dataset, we feel that we have achieved a number of innovations:

- Sampling: the sampling frame allows the linkage the results of the project to national representative samples like ENAHO and DHS Surveys, so as to improve the effectiveness of the project when engaging in policy debates. Young Lives data comes from a multi-stage, cluster stratified random sample.
- Back translation: Two levels of back translation were developed. First, back translation from English to Spanish and then back to English assure that core questions are comparable across countries that are also part of the study. In addition back translation was also used to ensure that questions are consistent between Spanish and local languages (in particular Quechua).
- The introduction of a detailed account of family and friend networks for each family to minimize attrition rates throughout the 15-year period.
- Field work training: Through intensive role-playing, a reasonable good level of standardization was achieved. A paper entitled Selection and Training of

Supervisors for Fieldwork: Experiences from the Young Lives Study in Peru documents the process. We know of no other publication that specifically addresses the selection and training of leaders or supervisors for community field research or similar situations involving autonomous groups working as a team adhering to rigorous quality control while confronting varied and challenging circumstances.

- Strict adherence to ethical guidelines approved by both local and international ethical committees assuring that the project achieves the highest ethical standards.
- The inclusion of information about field workers and supervisors (education, gender, quality of their work as assessed by their supervisors, etc.) for ex-post evaluation of potential biases introduced during the data gathering phase.
- The construction of a data entry system that includes double entry data, automatic levels of consistency checking aimed at minimizing data entry errors (see below).
- The design and construction of a software system that administrates the scanned forms collected, giving researchers access to the actual survey forms without jeopardizing the anonymity of respondents

I. GENERAL INFORMATION

Basic information about the candidate

○ **Name of the statistical activity/program**

Niños del Milenio (PERU) / Young Lives: An International Study of Childhood Poverty

○ **Institution responsible for its production**

Grupo de Análisis para el Desarrollo, and
Instituto de Investigación Nutricional

In partnership with: University of Oxford

○ **Name and position of the person responsible for application:**

Javier A. Escobal. Principal Investigator – Young Lives - Peru

E-mail address: jescobal@grade.org.pe

Complete physical address:

GRADE
Av. del Ejército 1870
Lima 27
PERU

Phone number(s): (51-1) 264-1780

Fax: (51-1) 264-1882

○ **Websites**

YL - PERU: <http://www.ninosdelmilenio.org>

INTERNATIONAL: <http://www.younglives.org.uk>

1. Context / Background.

In the last few years of the 20th century, the focus of donors on reducing and eradicating world poverty strengthened. A range of bilateral donors, as well as the UN system, the World Bank and the IMF, have all adopted a set of international development targets or goals, mostly to be met by the year 2015. Many of the goals relate to children, including reductions in infant and child mortality, and ensuring universal primary education. At the start of 2001, a number of leading actors, including Gordon Brown, the UK's Chancellor of the Exchequer, James

Wolfensohn, President of the World Bank, and Horst Köhler, Managing Director of the IMF, came together with NGOs and representatives of world faiths in London to reaffirm the international development goals, with a special focus on international action against child poverty.

However, while international policy interest in improving child welfare in the developing world has been growing, it is increasingly recognized that the data and evidence base for policy making in this area is fairly limited. Data on children is often fragmented, and rarely based on long term studies, which are essential for an understanding of how early life events affect life experience. Whilst there are child development studies in the UK, USA and other developed countries, no such studies exist in the developing world.

In 2000, the UK government's Department for International Development commissioned the Young Lives project, with the express aim of finding out what happens to children born in poor countries, using a longitudinal survey approach. The aim of the Young Lives project is to produce good quality long-term panel data about the changing nature of the lives of children living in poverty. A database management system for housing the survey data was generated, to make sure that the data was available not only to project participants but to a broader community.

Young Lives was launched as an international longitudinal study on childhood poverty and policy that is being conducted in Peru as well as in three other countries: Ethiopia, India (Andhra Pradesh), Peru and Vietnam. YL uses a multi-method approach to trace the lives of children made up of a 'younger cohort' and an 'older cohort', over a 15 year period. In each of the study countries, 2,000 children who were around the age of one in 2002, an older group of 800 children around the age of eight in 2002 and their primary caregivers are being surveyed every three to four years. Qualitative sub-studies complement the data gathered through the survey-approach to study in greater depth some of the key YL themes including children's wellbeing and children's access to such key services as education and health at transitional periods in their lives.

It is important to stress that the depth and extent of the Young Lives database is unique. No longitudinal research of this size, scope and complexity has ever been undertaken in the developing world. The project was design as a way of achieving the following three goals: a) provide good quality quantitative and qualitative information about the lives of children living in poverty; b) better understand the intergenerational transmission of poverty; and c) trace linkages between key policy changes and children's wellbeing; and, promote effective and sustainable pro-poor and pro-child policies

The countries selected for the study reflect a range of different problems and issues faced in the developing world: high debt burden, macroeconomic and structural adjustment, liberalization, high levels of inequality, transition from planned to market economy, emergence from conflict, and vulnerability to natural shocks such as drought and flood. In the case of Peru the fact that the country can be considered a middle income country with high rates of growth but reduced capacity to translate this growth into better and more equitable opportunities for the poor, make it a particularly interest case study, as other countries in the world are following a similar path.

To run such a complex data gathering initiative in Peru and to assure that quality and integrity of the data, it was deemed appropriate to establish a partnership between the Grupo de Análisis para el Desarrollo (GRADE) and the Instituto de Investigación Nutricional (IIN). Both research

centers have more than 25 years of experience working on socioeconomic development and on health and nutrition topics.¹ The complementarity of both expertises has been critical to assure a smooth operation of the project. GRADE-IIN Partnership reports to the Department of International Development – Queen Elizabeth House, University of Oxford, which coordinates and manages the international project.

2. Definition and Objective.

The project collects processes and makes available quantitative and qualitative data² at the individual, household and community level for a representative sample of children in Peru. In this way the project can contribute to create a detailed picture of children's experiences and well-being linked to information about their households and communities set within the national context. This is providing researchers and policy makers with a very rich and in-depth analysis of children's lives and how they are affected by poverty and government policies.

Quantitative data is gathered through comprehensive surveys that include interviews with the children themselves as soon as they are old enough to participate directly, with their parents and caregivers, and with key community members (such as teachers, village elders or elected council representatives). The quantitative data can help to answer key questions about:

- households' economic circumstances, livelihoods, assets and social capital
- coping strategies such as migration
- the relationship between parental education and other experiences and child outcomes
- the extent to which children and their parents and care-givers use services (e.g. healthcare, preschool care or education programs)
- how far the use, or lack of use, of services is associated with poverty, geographical location, ethnicity, gender, birth order and other considerations.

Qualitative information is gathered through individual interviews and group activities with selected Young Lives children, their peers and their families. The qualitative information gathered focuses on:

- children's well-being, especially related to their experiences of feeling secure and at risk, and their sense of self-worth
- children's time-use and the transitions they go through in childhood (for example, starting school or starting work)
- how the provision, delivery and quality of services such as health and education affect children's well-being.

3. Lessons learned.

Some of the uniqueness of the information gathered by the project that exemplifies the breadth, scope, and depth of initiative includes:

¹ More information on both partner's institutions can be obtained from our web sites: www.grade.org.pe and www.iin.sld.pe.

² As will be mentioned in more detail later, quantitative data is already publicly available, while qualitative data is being anonymized to be incorporated as well.

- A flexible survey scheme that maintains core data across survey waves to study changes in child wellbeing and, at the same time, introduces modules that account for the changing nature and consequences of poverty for children in different age groups as they progress from childhood to adolescence.
- An integrated community, household and child survey system to improve our understanding about the changing context in which children live.
- An introduction of in-depth qualitative longitudinal data collection and research strategy aimed at understanding some of the complex pathways through which changes in the policy context affects child welfare outcomes.
- The GPS coding of household dwellings which allows not only a better tracking of the sample through time, but also allows the possibility of connecting Young Lives data with secondary data sources, and the construction of measures of accessibility to key public services, which have also been GPS coded.
- The construction of a longitudinal picture data base that accounts for changes in children lives across survey waves. This instrument is not only a research database in its own right but has also been used to check the consistency of the data base confirming, for example, that cases registering very high or low height for their age were actually correct.
- Some of the characteristics of the data gathering methodology include the following innovations:
 - Sampling: the sampling frame allows the linkage the results of the project to national representative samples like ENAHO and DHS Surveys, so as to improve the effectiveness of the project when engaging in policy debates. Young Lives data comes from a multi-stage, cluster stratified random sample. A sampling document provides further details about the exact sampling protocol used and how census data can be used to post-stratified the sample.
 - Back translation: Two levels of back translation were developed. First, back translation from English to Spanish and then back to English assure that core questions are comparable across countries. In addition back translation was also used to ensure that questions are consistent between Spanish and local languages (in particular Quechua).
 - The introduction of a detailed account of family and friend networks for each family to minimize attrition rates throughout the 15-year period.
 - Field work training: Through intensive role-playing, a reasonable good level of standardization was achieved. A paper entitled Selection and Training of Supervisors for Fieldwork: Experiences from the Young Lives Study in Peru documents the process. We know of no other publication that specifically addresses the selection and training of leaders or supervisors for community field research or similar situations involving autonomous groups working as a team adhering to rigorous quality control while confronting varied and challenging circumstances.
 - Strict adherence to ethical guidelines approved by both local and international ethical committees assuring that the project achieves the highest ethical standards.
 - The inclusion of information about field workers and supervisors (education, gender, quality of their work as assessed by their supervisors, etc.) for ex-post evaluation of potential biases introduced during the data gathering phase.

- The construction of a data entry system that includes double entry data, automatic levels of consistency checking aimed at minimizing data entry errors (see below).
- The design and construction of a software system that administrates the scanned forms collected, giving researchers access to the actual survey forms without jeopardizing the anonymity of respondents

Although we believe we have been very successful in building and disseminating this database, there are a number of tensions that lie within the project that need dealt in order to assure long term integrity and quality of the data been produced:

The tension between respondent comfort and scientists' needs for broad in-depth data.

The time involved in administering the questionnaires and conducting interviews with children as a tension in the context of the researchers' interest in data of broad coverage and depth. Clearly this tension does exist and we acknowledge its importance. Obviously, this problem is not exclusive to Young Lives but exists in practically all research involving human subjects and is an especially important consideration in longitudinal studies. The project monitors this very closely and has developed a number of procedures to minimize respondent fatigue, which is mentioned in sections II and III.

Respondent fatigue and preserving the integrity of the data

In terms of respondent fatigue, one may contend that qualitative research qualitative may affect the integrity of the sample as it puts extra burden into a sub-sample of Young Lives children and families. We would argue that contrary to what has been suggested, qualitative interviewing is likely to reduce rather than increase this risk since it is less structured and less formal than questionnaire methodology and is dependent for success on building a strong rapport with respondents. This has already been evidenced in our first round of Qualitative work by comments often made by respondents at the end of their interviews, when they express enthusiasm rather than a sense of fatigue. At the same time, rigorous consent procedures applied during each interaction with respondents will ensure that the views of children and adults who do not wish to be involved in the qualitative interviewing, for whatever reason, are respected. This needs to be monitored very closely.

4. Future vision.

Young Lives has been designed to have four phases of research and dissemination:

- Phase 1 - Setting the baseline (2002–2005). This included the project set-up in the UK and participating countries and the first round of data collection and analysis. A substantial number of research reports and policy papers have been published from this phase of the research.
- Phase 2 - Project consolidation (2005-2006). During this stage the quantitative data from Round 1 was archived in the public domain and the research questionnaires revised in preparation for the second round of data collection. Considerable work was done on development of the databases to facilitate progressive data storage, and fieldworker Guidelines and checklists were devised. The pilots for the first round of qualitative data collection were developed.

- Phase 3 - Taking the research forward (2006-2009). In this phase of the research, we will carry out two more rounds of quantitative data collection as well as the first two rounds of qualitative research. It is at this stage of the project that we will begin to link the different strands of the research to amplify and deepen our understanding. Communicating our findings will become increasingly important in order to fulfill our policy engagement objectives and ensure that the research findings start to influence development policy and practice.
- Phase 4 - Completing the project (2009-2015). Two more rounds of research will be carried out bringing the total to 5 sets of data which can be compared and analyzed across the full 15-year period. At this stage we expect to have a broad, holistic understanding of the nature of childhood poverty – including its material, psychosocial and political dimensions – of a depth and richness never previously available.

The Spanish brand of the Project “Niños del Milenio” uses the subtitle “Información para el Desarrollo” (“Information for Development”) in its logo. The Project envisages itself as a key contributor to filling the knowledge gap that now characterizes childhood poverty in Peru. The project has already positioned itself through different mechanisms: from general branding to network construction in both academic and policy forum.

Our aim, consistent with that of the whole project, is to fill these knowledge gaps and provide solid evidence to underpin effective poverty reduction and social policies. The research that our data is fostering tracks the changing outcomes and impact of poverty for children through childhood, into adulthood and into the next generation. The data collection and dissemination and the studies it generates is designed so the research evidence can inform policy debates, and specifically link to the formulation and implementation of pro-poor policies that will be far more effective in reducing childhood poverty in the future.

We provide good quality quantitative and qualitative information about the lives of children in different contexts; we identify ways poverty affects children’s lives and their future prospects; we seek to better understand how poverty is transferred from one generation to another and trace linkages between key policy changes and children’s well-being. By doing so we expect to build capacity for and understanding of longitudinal research and promote effective and sustainable child poverty reduction policies.

The quantitative data provided by the project has already start establishing its reputation, and we expect that this will be consolidated as the second round of data goes to the public archive. We expect that qualitative information will follow these steps and will be considered by researchers and policy experts in equal foot.

We expect that quantitative and qualitative information will start playing together and reinforcing each other role. Qualitative findings will be used in refining the core quantitative instruments and developing additional modules. They can also inform the selection of variables for composite indices representing child poverty or dimensions of wellbeing, particularly where there are ordinal rankings of these components from group-based activities. The quantitative instruments have been ‘mapped’ for themes, including child health, education and child psychosocial wellbeing, which will guide future cross-cutting analyses.

Some of the areas where we envisage more work are the following:

Need to augment mechanisms ensuring publications of international standard

With the second round of survey data about to become available within the project, the next year will be a critical period for the project to produce high quality published outputs as well as broadly disseminating the data base.

The need to enhance political/policy analysis

Researching the political causes of poverty in Peru through a longitudinal project is an extremely challenging task given the importance of building long-term alliances with national and local political leaders. There is an inherent tension in use of political analysis given the YL design as a project that aims to influence policy, since to bring about policy change good practice in advocacy requires working with, rather than in opposition to, political figures.

The policy staff in Peru have been working by networking in political and policy circles and engaging in ongoing analysis of the political processes that shape policy and budgeting; they always use these contacts and analyses both in direct policy influencing and to inform their policy engagement/advocacy strategies. At the same time, as an evidence-based longitudinal project it has been crucial to hold back on making significant assertions about political and policy processes without having a second round of findings to assess impacts of policies on children. Therefore, upcoming publications will provide a greater reflection of political and policy dynamics and will build on the development and policy literatures.

Draft methodologies and conceptual frameworks for policy monitoring and analysis (PMA) and budget monitoring (CBM) have been elaborated. PMA and CBM are crucial for Young Lives for the following reasons: a) to contextualize/triangulate the evidence from the qualitative and quantitative research; b) to understand the policy, political context, and policy implementation issues of our sites; c) to inform future rounds of qualitative and quantitative research; to increase our understanding of policy and its impact on outcomes for YL children.

Capacity building and cross country and third party peer processes

In close collaboration with our UK counterparts we are involved in a range of capacity building initiatives. Capacity building is one of the four key outputs of the project in our revised project log frame and as part of the planning for 2008/09 (See annex 1).

Need to strengthen the linkages between the quantitative and qualitative components of the information gathering process

The project has generated several instances of collaboration between the qualitative and quantitative components of the project. For example the Community (Context) Instrument, which was developed and validated in 4 sentinel sites in Peru, was validated through the use of qualitative instruments (in-depth interviews and focus groups). Another is the involvement of the qualitative team in the development of the Round 2 questionnaires, who in fact took the lead on certain sections and questions, such as those concerning child outcomes. In this way, the questionnaires have been designed to provide entry points for deeper probing in the qualitative research. Likewise, the qualitative component was developed through close collaboration with both survey and policy teams.

Beyond this, we believe that the quality and usability of the data produced by the project lies in our ability to strengthen the usage of quantitative instrument to inform the qualitative research and vice versa. Integration between qualitative and quantitative components of the data gathering process can take place on four further levels: a) Conceptual: a continuously revised conceptual framework informs the development of the quantitative questionnaires, the qualitative component and the foci of the policy and budget monitoring and analysis. b) Lines of enquiry: The research is been informed by a number of grounded hypotheses; which sets the ground for synergy emerging between qualitative, quantitative and policy and budget monitoring and analysis components in the development of lines of analytic enquiry and collaborations between the teams; c) Site selection, sampling and methods: The round two questionnaires used methods drawn from qualitative and participatory research methods. Similarly, many of the group-based activities used in the qualitative research are quantifiable and can inform further analyses of the survey data. Qualitative findings will be used in refining the core quantitative instruments and developing additional modules. They can also inform the selection of variables for composite indices representing child poverty or dimensions of wellbeing, particularly where there are ordinal rankings of these components from group-based activities; d) Analysis: For logistical reasons quantitative and qualitative data collection and policy and budget monitoring will never be simultaneous. Analysis, however, will be simultaneous and collaborative, with individuals pairing up to address a particular research question or aspect of a research theme.

Future Data Management Plans for Round Three incorporating trouble shooting from Rounds One and Two

A major concern in Round Two was the length of time it was taking to not only build the databases but, once they were built and checked, to enter questionnaires into the MS Access database. As mentioned above, the Round Two questionnaires were much larger than those in Round One. On average it took 20 – 22 days to build one database and an additional five for changes and error corrections. Additionally, once these changes were made and the database was ready for entry it should have taken each data entry person one to two hours to enter a single questionnaire. You will appreciate that it is essential that the appropriate time be taken to enter in the data accurately. If the data is entered accurately this decreases the number of errors and the time it takes to check the data (both in Peru and centrally), as well as the process to put it right.

Round Three data collection is scheduled to take place beginning July 2009 in order to have data for both cohorts (both our older and younger) at eight years old. However, if the questionnaires are to be the same size as Round Two, the data manager will need to have the finalized questionnaires no later than July 2008 in order to build the databases accurately. It is hoped that the qualitative data will help decrease the size of the quantitative questionnaires by identifying specific themes and foci for moving forward.

Once the questionnaires are finalized and the databases are built in close collaboration with the central data UK manager, Peruvian data manager conducts early testing of the databases where the data manager is translating the forms into the local language.

Additionally, as the project is becoming longitudinal in Round Two, as well as the addition of the qualitative studies, it is becoming increasingly important to keep track of how the Young Lives cohorts are being surveyed in the field to ensure we do not encounter respondent fatigue and thus lose valuable data through refusals. Our strategy for combating this risk is to retain this

information in a universal ‘administrative database’ that will be reviewed by all persons planning fieldwork, and any respondent use, outside the planned quantitative rounds will need to be approved by the Peru’s Principal Investigator.

Round Two has also provided Young Lives with the unique opportunity to conduct further cleaning of the Round One data. This includes checking the demographic information contained in the household roster. For example, there are instances where the child was listed as a male in Round One and a female in Round Two. When this occurs we update the appropriate database and are working with the UK’s Economic Social Data Service (ESDS) on how we ensure these corrections are reflected in the Round One database that is already archived and being used publicly.

Rapid access to the data is vital to the researchers in order to provide results that can feed into current debates. However, data managers are reluctant to hand over any data until it is deemed fully ‘cleaned’, that is the data has gone through all the steps outlined in the deadlines. In order to meet a compromise of providing the data as soon as possible and ensuring a level of checks has been carried out, the data management team has built in a series of validation rules into the MS Access databases that will limit the types of answers a data entry person is able to enter into the database. Additionally, the double data entry comparison checks can, at the very least, ensure that the data entered is exactly that listed in the questionnaires.

II. INSTITUTIONAL CAPACITY

1. Legal framework.

This item applies partially, as the project is a private endeavor financed by international cooperation agency. However, the project does abide to an organization structure that gives consistency to the activities and acts as guaranty of the overall quality and integrity of the data provided.

Organization

The project is an innovative collaboration between the government, academic and NGO sectors. Initially funded by the UK government's Department for International Development, Young Lives is run by the Department of International Development – Queen Elizabeth House, University of Oxford, with the collaboration of the NGO Save the Children-UK. The Project has been constituted as a research centre to be run in close collaboration with partner institutions in the UK and overseas

There are research partners in each of the four countries. In the case of Peru, the project is managed by the Grupo de Analisis para el Desarrollo (GRADE) and the Instituto de Investigacion Nutricional (IIN)

In the UK, and in Peru there is also an Advisory Panel made up of experts, policy makers and others concerned with child poverty and welfare issues.

To assure an effective partnership at the national and international level, memorandum of understanding (MOU) between the different partners have been developed and signed. Annex 2 shows such MOU for partners working in Peru.

The Project has a Log Frame that gives coherence to all the activities carry out. (Annex 1)

The agreement between DFID (UK) and University of Oxford, as well as the sub-agreements between University of Oxford and each of the country institutions (including GRADE and IIN in Peru) explicitly states that the data generated by the project has to become public.

2. Resources.

Overall the Young Lives Project has a long term financing commitment from the UK Department for International Development (DFID). During the first two phases of the project 6.5 million pounds were allocated, of which 800,000 pounds were allocated to fund YL-Peru. Subject to evaluation it is expected that the Project will continue operating in its phase 3 until 2010, with an additional budget of 7 million pounds, of which 1.1 are allocated to YL-Peru.

Most of the budget has been allocated to the core activity: producing and maintaining the YL data base. Of course this has meant fewer funds for research. However it is expected that as

the project reaches its third round of data collection several other funding agents can come on-board. Currently pre-commitments from two donor agencies have already been obtained.

The fact that the funds are allocated with priority to Data collection and data management activities, has allowed us to produce the innovations mentioned in section I. Further appropriate funding has allowed documenting most of the processes. For example, during each phase detailed justification documents of the content of the household questionnaires, children questionnaires and context (community) questionnaires are developed. This document allow for maintaining coherence not only between the different sections of the questionnaires, but more importantly between different questionnaires across time. As an example of the type of background documents produced by the project Annex 4 presents Peru Sampling Design and Post-Sampling Adjustments Report. Annex 5 presents the justification of the content of the 7.5-8.5 year old child questionnaire for the first round. Similarly Annex 6 presents the validation of the Food Security Measurement Module introduced in Questionnaire 2 and Annex 7 presents the Community Quantitative Questionnaire Concept Note and Context Instrument Manual, which describes the rationale behind such instruments and the validation procedure for Round 2. Many other examples are available upon request.

Scanning of Questionnaires

Young Lives has recognized the importance of digitizing all the questionnaires for long term storage and safeguarding, but also as a way to use the questionnaires as a research tool. In this way it is possible to incorporate data drawn from any handwritten annotations on questionnaires into the analysis. We have already scanned all questionnaires from Round One and we are in the process of scanning questionnaires from round Two. Special software has been build to facilitate the search of these scanned forms. Forms have been scanned anonymizing those areas were reference to names or addresses appear.

Implementation of the WebDav Server

The size of the Round Two questionnaires and databases are quite large and therefore it is unfeasible to send them via email. In order to combat this (and avoid using CDs and mail services) we have implemented a web based server to transfer the files. The server is located at <https://webdav.geh.ox.ac.uk> and is a password-protected, encrypted website. This means that you are not allowed to access it unless you have a username and password. Equally, the encryption means that the data is protected when it is uploaded and downloaded. It is backed up and currently has 100GB of space. At the moment it consists of two top level folders: Quantitative and Qualitative. The central Data Manager has access to all folders. Peruvian Data Manager has access only to its specific country folder in each top level folder. This can be used as a way of 'sharing information' across the project.

Qualitative Data Management

The qualitative team is generating a mass of data in a variety of formats including field notes, audio recordings, photographs and videos and from a data management view, we realize that working with a dataset that is multilingual and contains multimedia requires careful monitoring and maintain a consistent data set sites within Peru and across counties for the overall project. This has included agreeing on an accurate and systematic method for data recording, clear and systematic labeling of data files in all formats, building a consistent file and folder structure within Peruvian country teams, ensuring our data is secure and backed up via the server,

methods for confidentiality, and that transcription and translation of audio files and other data records are similar across countries participating in the study. It is equally vital that the data collected in the qualitative phase link with other components of Young Lives (i.e. qualitative and quantitative for all current and future rounds). In order to achieve a collaborative set of guidelines encompassing these issues, the Oxford Data Manager took part in the qualitative workshop that was held in July 2007. This produced a set of guidelines that were further revised prior to data collection.

An important and exciting part of the qualitative study for the data management team is the work with the image and audiovisual materials such as children's drawings, photographs and videos. In doing so we will encounter a variety of complex issues including who to allow access and technical issues such as the most effective way to store and retrieve these images.

3. Management quality.

Annex 8 shows the Job description (terms of reference) of the main personnel working in the project.

Cohort maintenance is the responsibility of the Principal Investigator. He is involved in all decisions affecting our respondents and sites, such as who visits which sites and households when, for how long and for which activities. The issue of respondent fatigue is one aspect of cohort maintenance that they monitor very closely, since respecting respondents' schedules and avoiding becoming a burden to them is crucial to the successful outcome of the project. Any problems or issues that arise in the field are reported in detail to the PIs, who then take appropriate action in consultation with other members of the team. Our current attrition rates are very low (see Annex 22).

Annex 9 presents Data Management including detailed flow charts that show the full operation, starting from data collection to data entry, to data consistency to data backup management.

Young Lives – Peru has a full-time Data Manager who works closely with the Oxford based Data Manager. This team works in every aspect of Young Lives from Quantitative data collection to the Qualitative. Each individual working on the data management team is skilled in MS Access, which is the program used to build our databases, and SPSS which is used to archive the data.

The project has implemented a system of Quarterly Data Management Reports that are used to monitor various components of the project and ensures we are meeting our Data Management deadlines and quality targets

4. Innovations for improved cost-effectiveness.

Being a longitudinal project, cost-effectiveness is measure by the capability of having low attrition, minimum attrition bias and good quality data at a reasonable cost. We don't know of other longitudinal data gathering efforts of this scale and complexity (including quantitative and qualitative information, reciprocity activities and policy engagement) so as to compare our costs to comparable projects.

To assure cost-effectiveness the project has gone at length to build up several activities that will reduce the time of administering the questionnaires (Piloting). Each of the instruments has been extensively piloted. Further, through a carefully devised selection and training process of Supervisors for fieldwork the project has been able to standardize the data gathering process and reduce time allocated to the field operation.

Protocols for standardization have been constructed, tested and validated (see Annex 10). Such protocols cover the areas related to practice interviews, role playing, leadership skills, psychological tests that allow us not only to identify the best fieldworkers and supervisors for the job but also how they should be teamed together. As the fieldwork is quite a long process, good team relations are critical for the successful operation. The annexed document, also reports on several weakness especially regarding the importance of account for greater leadership skills.

On the other hand, we have been able to reduce the time spent to track those children that have moved (within a district) or migrated to other district, province or even outside the country (gathering information on the network of people/friends/family that are available to each family. (see questionnaires, round 1 and round 2, at http://www.ninosdelmilenio.org/ingles/inv_cuestionarios.shtml)

Annex 11 shows the Peru 2004-05 Tracking Report describing the challenges faced.

III. QUALITY OF DATA

To start with, the quality of data is as good as the conceptual framework that drives the questions that are to be answered through the data gathering exercise. The conceptual framework developed informs the development of the questionnaires, the qualitative component and the foci of the policy and budget monitoring and analysis. In line with the dynamic nature of our conceptualization and analysis and new insights brought to bear by further theorization and exploration of the literature, the conceptual framework is continuously revised, building on papers generated across the whole team on topics such as the multidimensional nature of poverty, the political economy of childhood, life-course transitions, and children's risk, resilience and wellbeing.

1. Integrity.

The high quality of the data is ensured through a meticulous process of piloting, translation and back translation (in several languages), double data-entry, checking and cleaning. These processes have been carried out by a dedicated team managed that includes not only full-time Data base Manager and Data Manager Assistant but a system that assures involvement of a central (UK) Data Manager. At the same time the Principal Investigator have worked hard to ensure that as many field supervisors who participated in Round 1 of the survey as possible were retained for Round 2, and this also ensures quality data.

As mentioned, new supervisors underwent careful training and their skills were assessed through role play and research piloting. Only those individuals with good communications skills and an ability to fill out the questionnaires effectively were hired.

The project has given priority to the training of fieldworkers as a key element to assure the quality of the data been collected. At this stage several innovations have been developed

As mentioned, protocols for standardization have been constructed, tested and validated (see Annex 10). Such protocols cover the areas related to practice interviews, role playing, leadership skills, psychological tests that allow us not only to identify the best fieldworkers and supervisors for the job but also how they should be teamed together. As the fieldwork is quite a long process, good team relations are critical for the successful operation. The annexed document, also reports on several weakness especially regarding the importance of account for greater leadership skills.

Some of the steps followed include

1. Publicise the posts through personal contacts, notices in universities and word of mouth.
2. Request candidates send via email, their background information in a predetermined format (education, work history, any relevant experiences, any additional skills, languages spoken) and answer a number of questions handwritten on a separate sheet to be faxed or hand delivered.
3. Revise replies and select a shortlist of 13 candidates who were invited to next step of the program
4. Two day intensive selection-training program
 - Icebreaking games
 - Exam of health knowledge, common sense, maths (calculating ages), open questions on management of situations;
 - Psychology tests to test teamwork, interpersonal skills, leadership aptitude, personality
 - Group work on attributes of a good supervisor
 - Test of skill of organizing logistics in the field
 - Videos of mock encounters between fieldworkers and families in the field for comment in group discussion
 - Practising filling the survey instrument
 - Practice of “Good clinical practice” techniques, how to cross out and correct errors on a form
 - Sociodramas in pairs of enrolment, with video which we showed afterwards.

During each of the activities individuals were evaluated or graded independently by different people and we had a final session with all the examiners (Investigators, psychologist, project coordinator).

As a longitudinal project with many types and levels of data it is essential that the project achieves a rigorous process for data management to ensure long-term quality and validity of the data. Many of the innovations listed below do not have a single purpose but provide benefits at various stages of the project.

The quality of Young Lives data is our highest priority. The first line of quality assurance happens at the field during data recollection. Field workers maintain close contact with the Lima-

based data management team, through their supervisors using cellular phones, to resolve critical issues. The issues solved through this mechanism have been consolidated and passed to all the field supervisors to assure consistency throughout the field team. Once the questionnaires are collected from the field we conduct 5 levels of data quality checks at various stages of the program in an effort to reduce error rates. The first is a verification of the questionnaire. The data management team has developed a checklist (in close coordination with members of the research team) that identifies areas where general mistakes could be made during data collection. A team of individuals conduct manual checking of the questionnaires against these checklists for inconsistencies and complete any missing information with a standard coding scheme for Missing or Not Applicable codes. Once each questionnaire is verified, it is sent to a data entry assistant who keys the data into a pre-designed database. The databases have been built with validation rules to limit the types of data that can be entered, thus further cleaning the data. Each questionnaire is entered twice into two separate databases. After data entry is complete, a double data entry comparison is run using Epi-Info. Epi-Info automatically generates a report outlining instances where data varies between questions in the databases. We work with the international standard of 1% as an acceptable error rate. Our error rate to date is 0.4% – well below the accepted norm.

After the completion of these checks, we send the database to the Data Manager at the University of Oxford where a series of random checks are conducted. After the data has been cleaned and validated, it is transferred into a series of SPSS files that are publicly archived through the UK's Economic Social Data Service (ESDS).

Young Lives has recognized the importance of digitizing questionnaires for long-term storage and safeguarding as well as for its usage as an additional research tool. As well as ensuring the longevity of the data, we are able to quickly compare Round 1 and Round 2 data according to sections and questions. The use of the digitized questionnaires has also proved beneficial during the verification of the questionnaires of the subsequent round.

Another added benefit of our data is the use of a picture database. A standardized protocol is used to capture pictures of the children, mother/child and their home that allow us to verify the consistency of the data as well as providing an innovative research tool on itself..

Monitoring the household roster for 15 years is a complicated task. As households can quickly change through migrations, new marriages or death we provide each fieldworker with the most recent household roster as well as at least two known contacts, with addresses, for the household they are visiting. This household roster is carried by fieldworkers on each visit so that if they are unable to locate the household they can use the contacts provided to locate the family. These documents are updated with each visit and are reflected in our administration database. This administration database is also used to keep notes about the community and is particularly useful for our parallel projects (such as PhD studies) that are using the Young Lives sample. The fieldworkers are asked to keep notes of any occurrences or challenges they have faced with the household, as well as any natural disaster or community issues. This ensures that we are not saturating our respondents and jeopardizing our retention through over-use of our sample, as well as taking an interest in our communities and ensuring we are working with them.

Attrition up the second round of data collection (five year period) has been estimated at 0.7% per year. This figure is dramatically lower than comparable longitudinal studies (with figures ranging between 2% and 6% percent on comparable studies like the Cebu longitudinal health

and nutrition survey on Philippines, the Pelotas Birth cohort in Brazil or the Birth to 20 project in South Africa). Moreover, detailed analysis of the attrition rate shows that it is concentrated in urban areas where migration to other countries is a major explaining factor. For rural areas, despite the remoteness of the areas covered, the attrition rate is below 0.5% per year.

The project follows specific ethical guidelines (Annex 13) to assure the Children wellbeing is not compromised. These guidelines are translated into Local Ethics Committee approvals (Annex 14). Finally Annex 15 shows the actual consent forms used during the fieldwork.

In terms of community reciprocity, there are several issues that merit commenting. First, we have passed through an Ethical Committee review from our partner institutions in Peru (the IIN), beyond the ethical review from Oxford. Thus we have acted in those few cases where ethical issues were at stake (for instance referring children in abandonment to public services). Second, we have proceeded in a way that we could reciprocate families for the information and time they've given us without interfering with the longitudinal nature of the study (not being an intervention study, we did not want to interfere with the developmental trajectory of children). The information we have returned to families is sometimes related to the results of the children (e.g. results of height and weight according to WHO tables). After round 1, during tracking, we provided children with a small picture of each. More recently, during qualitative work, we have agreed that each time a child is interviewed we will provide each with a small gift (i.e. not to exceed \$3.3 US dollars, for instance a coloring book). Finally, in the next months we plan to pilot test a picture exhibit of children's issues. The exhibit will rotate among YL sites, but will exclude pictures of children from the same community. We plan to include legends to each picture, which will be based on anonymous testimonies of children participating in Yong Lives, raising issues of children's well being. We are planning this exhibit to be interactive in some ways (e.g. asking participants to react to the pictures and messages).

2. Methodological soundness.

Site selection, sampling and methods: The round two questionnaires used methods drawn from qualitative and participatory research such as scenarios, ladders, and focus groups. Although the former used coded responses, they still enabled a more qualitative form of interaction and generated responses that could be explored further in the qualitative research. Similarly, many of the group-based activities used in the qualitative research are quantifiable and can inform further analyses of the survey data. Primarily qualitative case studies can also be understood in terms of (ideal) 'types' and 'kinds' of cases and used within the quantitative dataset to make probabilistic assertions about causality.

As already mentioned most of the instruments introduced in the quantitative questionnaires have been validated. For example, in the case of the food security module, we are collaborating with a regional network of scientists working on the measurement of food insecurity in South and Central America and is contributing the work they have done on validation and analysis of the food insecurity questionnaire that is part of the Peru specific questions with a view to contributing to an instrument to measure the qualitative dimension of food insecurity.

We also perceived that achieving international recognition for our data management procedures can be a way of validating our procedures and ensuring data quality and the soundness of our methodologies. Our participation in this competition thus should also be seen as part of our quality assurance process.

Further to our goals for achieving international recognition, we are planning to raise the profile of our Young Lives data and data management practices through training and dissemination events. Recognizing the success of the Centre for Longitudinal Studies (CLS) in disseminating their data and methods, the Oxford Data Manager is planning to attend training events held by the CLS. These training events will provide knowledge in order to develop a Young Lives training sessions geared towards future users of our data.

3. Exactness and reliability.

Non-response rate: Only 88 children and their care-givers had to be replaced as they did not want to participate in the survey. (53 refusals for the younger cohort and 35 refusals in the older cohort) This is equivalent to a non-response rate of 3.2 per cent. To all intents and purposes, this is a very modest rate.

The project visited 36,375 dwellings to enroll 2,751 children. This may look as a high number. However if one uses Peruvian Population Census information to calculate the expected number of dwellings one may need to visit to enroll one child of the targeted age range, we found that this number is 13 which is about the same number that can be calculated with the field supervisors report.

We have investigated the incidence and potential bias arising from attrition in the Young Lives study following the conclusion of the second round data collection process. Sample attrition is particularly concerning in the context of a longitudinal study such as Young Lives, where cohort sample sizes are modest and individuals are tracked over a relatively long period of time. The report provided as Annex 16 documents the rates of attrition of the Young Lives study (for Peru and for the other 3 countries involved in the project) following the termination of the second round data collection process. Second, it investigates the extent to which sample attrition is nonrandom. Third, it analyzes whether non-random attrition in the Young Lives sample might lead to attrition bias.

Attrition bias arises when sample attrition is non-random and the variables affecting attrition might be correlated with the outcome variable of interest. More formally, attrition bias will occur if the error term in the equation of interest is correlated with the error term in the selection or attrition equation. We observe that attrition rates are very small 3.52% between rounds for both the younger and the older cohort. This amounts to an attrition of 0.72% per year which is the lowest we know of compared to similar studies (see table 2 in Annex 16)³

Second if we use the unweighted sample YL households appear to be better-off than the national average. Nevertheless, and crucially, the YL sample of households covers a wide range of Wealth Index values akin to the variability observed in the Peruvian population as a whole. If we adjust the estimations for the sample design the distribution is statistically indistinguishable. Further, comparisons of poverty rates with 2001 ENAHO data appear to imply that the YL households are more likely to be poor but these differences are statistically insignificant. (See annex 3)

³ Attrition from Peruvian data is slightly higher than those registered in the other YL countries. This has to do with two factors. First Peruvian sample is geographically more dispersed and second migration rates are much higher in Peru than any of the other countries.

We found that in general, attriting households tend to hold fewer assets, have poorer access to services and utilities and are less educated, while at the same time children have poorer health and are less likely to attend school. However these patterns of non-random attrition are particularly not very strong when compared to Ethiopia and India, two of the other three countries study under YL.

The presence of non-random attrition does not necessarily imply attrition bias. In fact, the results from attrition Probit and BGLW tests provide only very limited evidence of attrition bias. Furthermore, the R-Square values in the attrition Probit regressions provide an indication of the proportion of attrition that can explained with a full set of child and household characteristics and the lagged endogenous variable. We find that across different models, the fitted models explain less than ten percent of the observed attrition. In other words, even if it follows some non-random patterns, attrition remains overwhelmingly a random phenomenon. (Annex 16)

In summary, our detailed analysis of attrition bias of the YL sample strongly indicates that current attrition for the Peruvian YL Sample is highly unlikely to bias research inferences.

Data Reliability

We have done extensive consistency checks for the first round of data collection and we are finalizing the same procedure after the double data entry was complete for the second round. While the double data entry program checks the accuracy of the data entry from the questionnaires, the consistency checks determines the likeliness of the data – that is if the respondent answered within a pre-determined range of answers based on local knowledge. This is a time consuming process and requires keen attention to detail, patience, as well as a detailed understanding of the questionnaires. One of our major concerns when checking the data is the household roster (as it is the base for knowing how each member of the household is related to the index child). We have hopefully eliminated the problems faced in Round One by building the household roster into the database and asking for the roster to be updated. However, we do predict we may have errors and the occasional “13 year old Grandmother”, as we did in Round One. When errors such as these occur the data manager is instructed to first refer to the questionnaires, and then contact the fieldworker who collected the data for further information. If the researcher is unable to help than it is recommended that a return visit to the household is scheduled to correct the data.

Once the data has been entered, compared using double data entry (Epi-Info Program), and the consistency checks have been run both internally and externally, we look at our ‘Other, Specify’ answers. In some coded questions the option of ‘Other, Specify’ was added with a text field to enter the data. The data manager will need to look at the full lists of ‘Other, Specify’ for each question to determine if some of the data can be coded into either the existing list and/or a new code added due to a large volume of the same specify answer. This is especially important because we do not archive the Specify answers and we need to retain as much information as possible. This is a very time consuming and detailed job and subsequent time has been built into the data management deadlines to allow for this to be completed.

Once all the databases are entered separately by the data entry personnel, they need to be appended according to site. Appending code is built to notify them of any records in each table that could not be appended. We have conduct double data entry (for 20% of the sample) and in-country consistency checks (for the whole sample) site-by-site as it is easier to run the checks and programs on a smaller scale. Once all the necessary checks are made, we will need to

merge the databases into one final database. We will need to be careful during this to ensure all records are transferred and that there are no more errors in the appending codes that might go unnoticed.

Annex 9 presents an assessment of our data entry procedures. The data entry process for 1826 forms for the younger cohort and 646 for the older cohort were completed. Double entry was carried simultaneously for a random stratified 20% sample. Stratification was done by data clerk in order to assess the reliability of the data entry process. We found an error rate below 0.5% which is considered reasonable. For all data clerks data error rates were consistently below the threshold of 1% (a higher rate would have triggered a complete re-entry of the data for that person).

After the data is cleaned and appended into one database is transferred into SPSS and the appropriate documentation is finalized. The data is then lodged with the Economic Social Data Service (ESDS) for public use.

4. Innovation in the quality of data.

Data management for a project of this nature is highly complex and our data management team plays a central role in the project. Building on current best practice we have developed a data management methodology which ensures the integrity of the information we collect, the confidentiality of the children, and aims to develop capacity for data management within Peru. Our approach starts with the construction of the questionnaire and leads through to data storage and public archiving.

The key features of the data management process include:

- input to questionnaire design developing databases for each phase of the research
- providing detailed data management guidelines and support to both the central and Peruvian research teams
- ensuring consistency checks are completed in order to provide high-quality and valid data for analysis
- developing systems for storing qualitative data and for linking this with the quantitative data
- submission of the data to the UK Data Archive

As already mentioned, there are a number of innovation that have been established to improve the quality of the data including back translation of all questionnaires; double entry; scanning of all questionnaires and allowing online access to them so as to use them as double-checking throughout the analysis phase.

At the questionnaire level, to assure that tracking has a higher success rate we have introduced a detailed account of family and friend networks for each family to minimize attrition rates throughout the 15-year period. Further, GPS coding of household dwellings has been introduced to allow better tracking of the sample.

GPS coding of household dwellings also allows the possibility of connecting Young Lives data with secondary data sources which greatly enhances the usefulness of the data. The project all

GPS codes key public services in the communities where children live (see community questionnaire).

Field work training: Through intensive role-playing, as has been already mentioned, has had an important role in assuring the quality of data.

Further a data base with key information about each of the fieldworkers and supervisors finally chosen has been produced. This data base includes age, gender, education of the fieldworkers and supervisors, previous experience with the project, internal evaluation, etc.. (See Annex 12). Obviously we expect that the characteristics of the pollster should not affect the main relationships that are to be found. Although this is typically an untested claim, we have been able to translate it into a testable hypothesis. For example, we know that the relationship between mother's education and child welfare outcome (proxied here by an indicator of child malnutrition) in YL data, is not affected by the characteristics of those that conducted the interviews and generated the anthropometric measurements.

It is important to note that we also have policies to retain qualified personnel at the times where no field work is being done. We have offered employment opportunities to several key fieldworkers between fieldwork periods – to be involved in data management, coding and analysis of qualitative data to encourage them to stay on with the project build their skills and enable to participate in future fieldwork phases.

Another area where quality of data can be reflected is that of validation of several instruments that have been introduced in the questionnaires. This includes, for example the validation of the food security module (see annex 6); the validation of the community questionnaire (see annex 7); as well as validation of cognitive development and achievement instruments. For the later, the Peruvian team is currently working on validation issues for the other countries that are part of Young Lives. The tests used in the second round included the Peabody Picture Vocabulary Test (PPVT) and the Cognitive Developmental Assessment (CDA) for the younger cohort; and the PPVT and the Mathematics achievement test for the older cohort.

IV. USE OF DATA

1. Relevance of the program.

As has been mentioned this projects combine not only qualitative and quantitative data gathering but also incorporates an explicit communication strategy to assure that the data is used and the research that is produce with the data informs policy. This is not just an aim but is a building block of the partnership that constitutes the core of the project. Save the Children UK an NGO specifically engaged in child policy advocacy work has been a partner of the project from the start.

Annex 17 presents Young Lives Communications Plan for Peru 2007-2008. There you can see the specific objectives and actions that are currently been carried out to ensure that the data does collected and the research outputs produce does inform policy.

The following are the goals of our communication strategy:

1. Primary research is increasingly considered by policy-makers in designing and revising social policies and programs. We expect that achieving this goal should have a relatively immediate impact on pro-poor policies.
2. Primary research is used increasingly by academics and higher-education institutions in their regular teaching. As opposed to the first objective, achieving this one is likely to have an impact on pro-poor policies in the medium to long-term, as some of the professionals that learn with primary research will become eventually the next generation of policy-makers.
3. Primary research is increasingly considered by the press in their reporting news (or is the news) and read by the public. With this objective we expect to influence a powerful sector, the press, but also create or strengthen the general public's opinion about the importance of pro-poor policies for children.
4. Topics related to infant poverty are increasingly researched by several institutions in Peru. The academic community of researchers in Peru is relatively small and so is the interest in children's topics. We expect to influence this and thus generate more information from other sources that could be used for achieving the communication goal stated above.
5. Children and community representatives are given a voice to express their views on childhood issues. This objective would be achieved mostly at YL sites.
6. Delivering relevant YL information to local authorities at YL sites. This objective has to do mostly with returning local stakeholders with information that is relevant to them as a way of giving back for the information they have provided.
7. Keeping the YL brand prestigious. We realize that in order for the main messages to be delivered by YL to be accepted there is a need for the brand to be respectable. A lot of this has been achieved by several activities in the past, some of which we expect to continue.

The kind of activities we have done on this area include

1. Public presentations
2. Closed meetings and seminars
4. Dissemination of YL publications to academics and higher education professors nationally
5. YL newsletter
6. Press notes and op ed articles
7. Workshop for representatives from the press
8. Participatory videos
9. Focus groups with children
10. Updating the website
11. Annual YL calendar
12. Budget monitoring paper

Further we are preparing a National conference on childhood research and policy that will bring together the research done by the project, by other using the data of the project plus other relevant research on child poverty. The idea is to bring together researchers and policy makers

2. Use of the program.

Annex 18 shows a list of Publications done with Young Lives – Peru data.

At the same time the data has been used for several master and PhD theses at US and UK universities and a range of working papers and publications, including articles in peer-reviewed academic journals (for example in the *Lancet*, *Social Science and Medicine*, and the *Community Development Journal*).

Dissemination

A priority for the project is to ensure that research findings reach policy makers and the general public, For Peru, the UK and as far as possible, internationally. National Coordinators will use their knowledge and experience of policy processes to maximize the visibility and clarity of messages arising from the Young Lives project. For the purpose of branding and positioning of the project we have done partnerships with the media is promoting awareness of childhood poverty as an issue (RPP radio). Although we are careful to avoid been identified as an advocacy group, we do fill that branding is essential to position the data and the research coming from the data as a key element of good policy debate and improve policy making.

The project is producing its own working paper series, with the expectation that some of these papers will then be published in peer-reviewed journals. After the first round of fieldwork and data analysis, a country report will present the analysis of the main findings, with a full international report appearing shortly afterwards. In each country, lead issue will then be pursued, with further, more in-depth analysis, leading to targeted policy briefings, workshops and other outputs. In late 2003 the project will hold an international workshop.

Young Lives First Round Country Report for Peru can be downloaded directly from www.younglives.org.uk/publications/country-reports

In the case of the FTA, I believe the policy debate that the research generated allowed public opinion and government to be aware that crop-specific compensations linked to products that were expected to be negatively affected by tariff reductions may not be the best policy option to deal with vulnerable individuals. Moving away from a crop or product centered policy towards contingency funds associated to vulnerable people is something that is happening. I have been invited several times to closed discussions within the government to discuss these issues.

In the area of child care centers the former director of the Wawa-Wasi program has acknowledged the role the project in raising awareness of its usefulness and increasing the budget allocation. The current officers have engaged with YL researchers and policy officers at other Ministries (specifically Ministry of Economics) in discussions about the future of the program. Furthermore, based on the research conducted so far, we are planning at the time a policy research on the municipalization of the program (which is scheduled for 2008).

Further Research and Policy evaluation

Although a third round of data is critical to increase the robustness of causal links, we believe that the panel data provided right now by the project can provide us with key insights on the effects of policy changes on child welfare. We may claim that there are several “natural experiments” that can arise from the data that can be used to evaluate the effectiveness of key policies. For example GRADE is working on a paper that evaluated the conditional cash transfer program that has been implemented in Peru, taking advantage that since it was phased-in our sample include children that received the transfers and children that did not but are just been incorporated into the program (that is, are definitively eligible) and did not received the program. Qualitative research done by a PhD Student (Natalia Streuli) from Oxford University will complement this research (Children’s experiences of poverty and well-being in Peru: exploring the implications of a conditional cash transfer program)

The combination of quantitative and qualitative evidence provided by the project has been extremely useful to inform public officials about the need to change the emphasis of their child care programs from “just” a day care center to a program that helps children develop their potential fully. Using Young Lives data a research study showed that the program is indeed received quite positively by participants because it provides with a safe space where children are cared for, fed and checked for health problems (also giving mothers time to work or study). However, the same study did not show any differences between treatment and comparison children in cognitive development (i.e. no differences in language, fine and gross motor skills scores), even though this is one of the aims of the program. Without the combined used of quantitative and qualitative data gathered by the project, we would not have be able to show that the loss of standardization that the municipalization may bring about could hinder the positive effects of the program

We consider this study quite unique because of it being longitudinal over 15 years, integrated (i.e. qualitative and quantitative research, communications and policy) but also because it is multi-country. In the future we expect that the later component may be stressed by emphasizing comparative research across the four countries (as the qualitative team is starting to do now), so that South-South collaborations emerge (and not only bilateral dialogues between Oxford and each country team as sometimes is the case). Further, we think that the study is becoming an interesting learning experience and source of information for researchers in developing as well as developed countries (especially from the UK). Examples of this are PhD students coming to collect additional data from our children to do their thesis or more experienced

researchers using directly the data bases (for instance a group of researchers from the U of Lancaster contacted Dr. Martin Woodhead as they were interested in analyzing the data from the Ravens from round 1). We believe that once the data base for round 2 is finalized we should make a strong campaign among researchers, university programs and international research societies with an emphasis on children's issues on the availability of the data base and the disposition of YL researchers at Oxford but also in the countries to do research.

Country interaction (north-south and south-south) adds incredible value to the project that needs to be underline. For example, just to make one point, we can expect that over the life of the project over 50 (or more) young researchers (PhD. students) will be "trained" on development issues through their interaction with south research partners...

3. Accessibility of data produced by the program.

The fact that the data bases for all the countries maintain a common framework and all of them are publicly archived in the UK (<http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=5307&key=Young+Lives#doc>) allows researchers and policy makers to access the data both for monitoring child welfare outcomes, as well as to evaluate specific public policies.

The Metadata associated with the databases available can be found at:
<http://www.ninosdelmilenio.org/ingles/investigacion.shtml>

The actual questionnaire forms can be accessed online, as they have been scanned and are administrated through specially constructed software. The anonymity of the families and children is achieved by striking out sensitive information prior to scanning. The access to this data base is done only for registered researchers as the software access is password protected. (See annex 19)

Quantitative data

We are in the process of cleaning Round 2 data and aim to complete this task by March 2008. At the same time we are preparing the data dictionaries and the other supporting documentation needed for researchers to make full use of the data to submit our datasets to the UK's Economic Social Data Service (ESDS). Based on experience with Round 1, we expect the ESDS to complete their checks in around 3 months. We are keen to maximize data use and have discussed with ESDS staff ways of promoting knowledge about the project and archived data among scholars internationally.

However, we do not plan to restrict ourselves to archiving with the ESDS and have a strong commitment to facilitate use of our data by external researchers, as part of capacity building. During early 2008 we will be developing a curriculum for training of local researchers and for local dissemination of anonymised data sets using CDs.

Qualitative data

We have received the recommendation for archiving of qualitative data to enhance access for external researchers. Recognizing the challenges involved, we are working together with ESDS on best practice and they have offered to provide any assistance needed. Several steps have to

be taken to ensure proper procedures for anonymisation and informed consent are followed. Our ethics advisor is working closely with us on this issue, and we are already in discussion with lead qualitative researchers on the implications for their work. Cleaned data sets from Qualitative work will be available in early 2008, and will already have been anonymised in large measure. We will then undertake a careful review to assess additional steps needed to ensure confidentiality and/or other forms of protection for Young Lives children and their families. We anticipate that more sensitive materials might be summarized, while others made available in full (with anonymisation). We aim to share our draft proposals for archiving with the ESDS and agree plans with them.

Access for external researchers to data prior to archiving

Access to external researchers do data prior to archiving is possible following a simple procedure and signing a consent form (see annex 20)

Again, we have already taken some measures to facilitate this process. Essentially our aim is to build a large critical mass of research and data users in a short period of time, whilst also ensuring quality, good coverage of our 'key issues', and enhancing the Young Lives brand:

Graduate and Postgraduate students

Postgraduate students at Oxford and other universities (to date, Bath and the Institute of Education, London) are being strongly encouraged to work with the project on literature reviews, conceptualization, engagement with theory and analysis of our empirical data. Since 2006, at least 8 postgraduate students have worked with the project in Peru. In addition, Oxford has put out a call for proposals for graduate students to work under supervision on YL 1st and 2nd round data, with a deadline of 30 December.. Selection of the proposals will be done by senior YL staff, with relevance and academic quality of the research proposed as the key criteria. Specific research support and seminars will be organized as part of this process and a small fee paid to successful applicants. The call for proposals contains specific appeals for research that would integrate the qualitative and quantitative data, with specific support from the qualitative team of YL. A similar call for proposal at local level will be launched in Peru during the first quarter of 2008.

In preparation for an international conference planned for March 2009, a one-day seminar will be held in May 2008 to discuss the core themes of Young Lives, our key research questions, findings from the Round 2 quantitative data and initial findings from the qualitative and policy work. The aim of this meeting is threefold: a) To invite others to engage with the Young Lives data before it is publicly archived; b) To bring in analysis from other poverty- or child-focused research which will help broaden our own analysis and approach, and allow us to situate Young Lives in the wider discourse; and c) To raise the profile of Young Lives with key stakeholders and build understanding of our work at a critical time in the project's development.

Along with Young Lives staff and Advisory Panel, participants will be invited from other research organizations (academic and policy research), international NGOs and donor organizations. Papers commissioned from some of the participants will contribute to the formulation of the intellectual framework for the conference, helping to ensure that the conference agenda is informed by the latest research and can stimulate fresh thinking and new approaches to childhood poverty and policy.

Thesis work already completed using Peruvian YL data include:

Tami Aritomi (Ph. D. Dissertation Penn State University): exploring barriers to Peruvian mothers' labor force participation: a multi-disciplinary multimethod approach (2005)

Mary De Silva. (Ph. D. Dissertation Faculty of Medicine, University of London. Department of Epidemiology and Population Health London School of Hygiene and Tropical Medicine): "Context and composition? Social capital and maternal mental health in low income countries" (2005)

Other Audiences

However, in YL we are also targeting an audience of students in diverse social areas as a medium-term way of influencing policy. We expect that once the web page is renewed we will disseminate that there is a variety of relevant primary research free to download. We expect that professors at universities will use it as study material (given that there is little such type material available in Peru, with university professors relying mostly on theoretical texts written in a different language and translated into Spanish) and students using the studies and data bases in their thesis. In the long run we expect that they will develop a habit of including relevant research (YL and other) as one relevant input in designing or revising policies.

We have explored theoretical models of research use by policy makers, such as RAPID, developed by ODI. Based on these and the conceptualizations developed within YL we plan to use a model where research (both quantitative and qualitative), communication and policy engagement reinforce each other. The model includes establishing working relationships and informal alliances with key stakeholders and institutions. We will stay away from a more forceful advocacy model of procuring policy changes, as we think that advocacy is the type of model used by other types of organizations (such as children's rights NGO's); if we used such a model our whole approach would be jeopardized (indeed some times research in YL has shown results and explored diverse policy implications but not pushed particularly for any one of them, as we understand our role is to provide and discuss empirical, high quality information, but other actors, such as policy makers, need to understand and make good use of it within their scope of action). This is based on a conviction that no one single piece of research will provide definitive understanding of complex social issues, but should provide with relevant information to be considered alongside other inputs in framing policy.

4. Innovative use for the design, monitoring and evaluation of policies and projects.

Some concrete examples of how the data has been used to inform and affect the policy making decision process include:

- The evaluation of the potential impacts of the Peru-United States Free Trade Agreement on children's welfare. In particular, the project was able to show the vulnerability for children that trade liberalization may bring about even in cases where trade related growth may occur. Studying the patterns of female labor allocation and the fact that, in the absence of daily childcare facilities, girls may be taken out of school to look after their younger siblings while their mother works, the project was able to provide strong evidence in favor of complementary policies to reduce children vulnerabilities.

- The management of the childcare program Wawa-Wasi program (a public program aimed at poor children between the ages of 6 and 48 months) has used the information provided by Young Lives to convince the Administration to increase their budget allocation. The fact that Young Lives has been able to provide solid evidence of the representativeness of its sample, has allowed the project to engage in fruitful policy discussions with key officials from the Ministry of Women and Social Affairs, of both the previous and current the Administration.
- The combination of quantitative and qualitative evidence provided by the project has been extremely useful to inform public officials about the need to change the emphasis of their child care programs from "just" a day care center to a program that helps children develop their potential fully. Using Young Lives data a research study showed that the program is indeed received quite positively by participants because it provides with a safe space where children are cared for, fed and checked for health problems (also giving mothers time to work or study). However, the same study did not show any differences between treatment and comparison children in cognitive development (i.e. no differences in language, fine and gross motor skills scores), even though this is one of the aims of the program. Without the combined used of quantitative and qualitative data gathered by the project, we would not have be able to show that the loss of standardization that the municipalization may bring about could hinder the positive effects of the program.

An example of the policy notes produced for internationally audiences using Peruvian Young Lives Data can be found in annex 21 ("Educating women = healthier children?" In: id 21 insights # 56 June 2005)