YOUNG LIVES TECHNICAL NOTE **NO. 23**
April 2012

Surveying Siblings: The case for tracking brothers and sisters of Young Lives Index Children in Round 4 and Beyond

Catherine Porter
Alan Sanchez
Manisha Nair

Young Lives
An International Study of Childhood Poverty

Contents

Executive summary

This note sets out the arguments for extending the sibling cross-section collected in Young Lives Round 3 (2009) to a panel in Rounds 4 and 5. It is a non-technical synthesis of two background notes from an economic and health perspective[1], with some additional information and arguments. We point out three main gains in terms of content from collecting the panel of sibling data, as well as noting that extending the amount of information collected would considerably add value. Finally, without making calculations we assert that collecting sibling data would be a relatively low-cost addition to the existing survey.

The Young Lives quantitative survey has been designed and implemented as a panel survey of index children, surveyed thrice since 2002, and there are currently two more rounds of fieldwork planned. In Round 3 (2009), information was collected on the next sibling down in three of the study countries, in the domains of anthropometrics, cognitive development, education and time use. We now point out the principal methodological innovations of sibling data:

1. The sibling is our best comparison of what would have happened to the index child at the same age, under different circumstances.

- The younger sibling in the current round can be compared to the index child in the previous round to answer a variety of questions around how household or community circumstances affect child outcomes *at the same age*.
- For example, using Young Lives Peru data, the IDB project analysed the food price crisis and its impact on nutrition, which subsequently impacted cognitive development. Young Lives index children aged 5 in Round 2 were compared with their younger siblings aged 5 in Round 3, who suffered food price rises as infants. (Outes, Porter and Sanchez, 2011)
- The list of questions that can be analysed in this way using the Round 3 sibling data has not yet been exhausted. Similar studies can be done in Round 4 and 5, as the children age.
- If major events occur across the whole country or most of the study communities (macroeconomic events, droughts, food price crisis, introduction of new social programmes), the sibling comparison is especially useful – as the whole index cohort will experience the same event (i.e. we have no control group).
- The scope of questions that can be answered can also be expanded if Young Lives questionnaires in Rounds 4 and 5 are expanded to include other dimensions of child well-being – for example psychosocial competencies, health-seeking behaviour, or alcohol consumption.

2. Collecting sibling data in Round 4 and beyond means that there will be a *panel of siblings*, not just a *panel of index* children.

- The range of questions that can be addressed can now include the influence of past events and circumstances on current outcomes. We can estimate a 'difference in

1 This summary produced by Catherine Porter, Department of Economics, University of Oxford. Background papers included 'Siblings and the Young Lives Project' Alan Sanchez, and 'Sibling studies in health' Manisha Nair.

differences' of the siblings, which addresses several econometric concerns outlined by Sanchez (2011) – these are not addressed by a panel of individual children.

- For example if any sibling migrates, it will be possible to better estimate the impact of this on outcomes than with a single-child panel.
- This can be between two consecutive rounds, or in the longer term – for example whether early child circumstances impact on adult outcomes (depending how long the survey continues – this line of research may be even more fruitful when there are pairs of sibling adults to compare).

3. Collecting sibling data in Round 4 and beyond allows research to explore issues *within* the household that cannot currently be addressed.

- For example, we can observe whether differences in siblings' outcomes (nutrition, health, cognitive ability for example) narrow or widen as the children age. Plus, we can see whether households try to narrow these differences between their children, or reinforce them; and if they try to narrow, whether household resources constrain them from doing so.
- The sibling combinations will include brother-brother, sister-sister, older brother-younger sister and older sister-younger brother. Comparing the dynamics of outcomes and their differences across these different groups will allow research to unpack gender dimensions of household dynamics, across the range of outcomes for which we have data.

4. Extending the amount of data collected – by breadth and length – about the sibling may be worth exploring for future rounds.

- Since early childhood circumstances often matter for future outcomes (at a minimum, we may wish to control for them) Young Lives should discuss the merits of collecting historical data for the sibling. For example, whether they attended pre-school, immunisations, early child diseases, length of breastfeeding, birth weight.
- As noted above, Round 3 included anthropometrics and (in some countries) cognitive development. Expanding to other dimensions would increase the range of questions that Young Lives data can address. As this is one of the principal merits of the Young Lives data vis-à-vis other existing studies, it is worth extending to the siblings.

5. Collecting data on siblings in Rounds 4 and 5 is likely to be cost-effective.

- The marginal cost to Young Lives of collecting sibling information in Rounds 4 and 5 is small.
- Instruments will already have been developed – e.g. the PPVT administered to the index children in Round 3 should be administered to the younger sibling in Round 4 for comparability.
- Arguably, the cost to the Young Lives households is also not too arduous: it does increase the interview burden but on a different household member.
- The main cost would incur if it is difficult to locate the sibling (i.e. if they have gone to visit a relative, moved to a different household from the index child etc).

1. Introduction

The Young Lives study collects longitudinal data at the child, household and community level in four developing countries for two different cohorts of children in each country. With three rounds of data already available, Young Lives has created a unique source of data for researchers that allows the study of different dimensions of child development and of the dynamics of child development in the context of developing countries.

An important characteristic of the study thus far has been that only one child per household is tracked over time. Methodologically, this has implications for the type of causal relationships that can be explored. The argument elaborated in this document is that while in some instances the study of causal relationships using Young Lives data is straightforward, it is important to recognize that there are situations, when we know or suspect that the occurrence of certain event is correlated with household characteristics, in which counting with additional information on siblings in multiple rounds would considerably increase the extent to which Young Lives data can be used to identify causal relationships.

For instance, consider the occurrence of events such as household economic shocks (e.g., job loss, a death of a household member) and natural disasters (e.g., frosts, droughts) on child educational and health-related outcomes. It is not unreasonable to argue that children born in households not affected by these events but that face similar living standards otherwise can be used as an adequate group of comparison of those born in the affected households: they are a good representation of how children in the affected households would do had they not been hit by these shock. It follows that by comparing both groups of children one can estimate the impact of these events on child-level outcomes. In many instances however we are interested in studying the impact of household-level or community-level events that are within the control of the family (or of the Government) on child-level outcomes. In this type of cases finding an adequate group of comparison becomes problematic. Consider two examples: (a) cash-conditional transfer programmes target poor households. Households not selected by the programme are wealthier by definition and, thus, do not make a good group of comparison of the selected ones; (b) good parents invest more in the education of their children. It follows it is very difficult to compare the cognitive results of a child that did not attend pre-school with that of a child that did attend because we suspect they were raised in very different conditions.

Given these considerations, the aim of this document is to justify why it should be of interest for Young Lives to collect longitudinal data on siblings; i.e., to create a panel of siblings. The key idea is that siblings are similar enough to be ‘comparable’ (e.g, they live in the same house, were raised by the same person, most likely attend the same school) yet their life experiences differ, in part, because of their date of birth. For instance, a cash conditional transfer programme might have a timing such that one sibling was benefited during the crucial 24 first months of life while the older sibling was not. Or, similarly, one sibling might benefit from a pre-school programme that did not exist at the time the older sibling was of pre-school age. Extreme weather events might affect one sibling when the other was not yet born. These situations can be exploited for research purposes. It opens a door not previously available in the ‘one-household, one-child’ setting.

In other words, the sibling is the best prediction available of what would had happened with the index child had her family faced different conditions. This is useful for two reasons. First, it provides a way to study the causal effect of household-level and community-level events or

time-varying characteristics when we know or presume that the affected households are not truly comparable to those not affected. Second, it provides a mean to check the validity of our assumptions in cases when we think it is safe to assume that households affected and not affected by a given event are comparable. In both cases, having data on siblings substantially increases the capacity to convince technical audiences that a causal relationship has been detected.

In Round 3 of Young Lives, conducted in 2009, information on the younger sibling of the index child was collected. Younger siblings of the Younger Cohort were weighed and measured in Ethiopia, Peru and Vietnam, and of the Older Cohort in Ethiopia (for details, see table below). Cognitive development indicators (PPVT scores) were collected for the younger sibling of the younger cohort in Peru and Ethiopia. This new dataset allows the analysis of some new topics, such as social programmes that were introduced after the Young Lives Cohort children were of an eligible age, or of economic shocks that happened similarly later than at a time when the Young Lives children were affected (e.g. the global food price crisis of 2009 is likely to have impacted younger siblings who were under 5 years more severely than the 9-year-old Young Lives Cohort children).

Table 1: *Young Lives sibling information currently available for Round 3*

Country	Which siblings?	Dimensions of data	Number of siblings
Peru	Next sibling down, index child (YC)	Anthropometrics, Cognitive (PPVT)	
India			
Vietnam	Next sibling down, index child (YC)	Anthropometrics	
Ethiopia	Next sibling down, index child (YC and OC)	Anthropometrics, Cognitive (PPVT)	

Collecting data on siblings in Rounds 4 and 5 of the Young Lives survey allows for the tracking of both children over time and a more complex (and often more powerful) strategy for isolating the impact of external events and household responses on children's outcomes. It also allows a deeper understanding of dynamics within the household, such as whether children are favoured systematically over time, or whether gender biases diminish over time. If a child suffers bad health or an accident in early childhood, do parents compensate this misfortune compared to their siblings, or withdraw resources in order to invest in `stronger' children? These important questions can be answered when tracking more than one child per household. Section 2 outlines literature, mainly from economics and health, that has made use of sibling data. The third section discusses some technical issues that can be solved with the use of sibling pair data. The fourth and final section concludes with some recommendations for Young Lives in forthcoming rounds.

2. Literature using sibling data in economics and health

Among early applications that have exploited paired-siblings data to study the determinants of the well-being of children in the economics literature are Behrman and Taubman (1989), Rosenzweig (1986) and Rosenzweig and Wolpin (1988). More recently, Todd and Wolpin (2007) and Cunha and Heckman (2008) have used this strategy to study the process of cognitive and non-cognitive skills formation over the childhood period using data from several waves of the National Longitudinal Survey of Youth 1979, with observations for pairs of siblings at 7, 9, 11 and 13 years of age. In the context of developing countries, Alderman et al (2006) and Glewwe et al (2001) have used an equivalent strategy to study the relationship between early nutrition and educational achievement. For other studies that have exploited kinship data, see Rosenzweig and Wolpin (1995). For a historical approach to the use of this strategy, see Griliches (1979).

In the following sections we present examples that illustrate the advantages of using siblings data in a longitudinal study. Except for the first one, the examples presented have educational performance as the output of interest. However, the points made are broader and can be easily generalized to study how household-level and community-level events or time-varying characteristics affect child outcomes and how, in turn, child outcomes in a given period play a role in explaining the same person outcomes in later periods. Overall, the point made is that for research purposes, ideally for each observation of the index child over time there should be the same number of observations for at least one sibling observed at approximately a similar age in order to fully exploit the uniqueness of the Young Lives data for quantitative research. Since Young Lives did not collect siblings data in the first waves, an implication is that it might prove useful to include retrospective questions about the siblings in order to construct a panel of siblings.

2.1 Anti-poverty programmes and child outcomes

Suppose one is interested in measuring the impact of a conditional cash-transfer (CCT) programme on birth weight. This type of programme typically targets poor households (e.g., Juntos in Peru). The problem in this case is that households that do not access the programme (non-poor households) are not strictly comparable to those that do. This renders any comparison between beneficiaries and non-beneficiaries obsolete. To estimate causal effects in these conditions an alternative is to identify pairs of siblings where one was benefited by the programme during the gestational period while the other not simply by virtue of differences in date of birth. Put differently, the date of birth becomes a natural experiment. This approach was used by Amarante et al (2011) to study the impact of a poverty relief programme on birth weight in Uruguay.

2.2 Cumulative processes over the childhood period (or the study of dynamic relationships)

Another type of question that is of interest for the Young Lives project is how a child's outcome observed at a given point of time can in turn become an explanatory factor for either the same or a different outcome observed later in life. For instance, how can it be tested if a child's vocabulary knowledge at age 4-5 (measured by the standardized Peabody

test) is causally linked to writing and reading abilities at age 7-8? This case exemplifies the cumulative nature of cognitive skills formation (see Todd and Wolpin 2003 for a conceptual discussion). While the nature of the relationship of interest seems 'obvious', claiming causality is difficult for the kind of factors mentioned previously. In short, children born in 'better' households and communities are likely to perform better at school performance related tests. So, it is important to control for **community characteristics**, **household socioeconomic background**, **parental ability** and **parental preferences for education**.

In addition to this non-exhaustive list of possible household-level and community-level confounders, child-level capacities and characteristics prior to the age of 7-8 play a role in this case: according to the early childhood development literature it matters whether the child attended pre-school and whether the child was well nourished during the first three years of life. These factors matter because they could explain why a child performs better at all these tests (see Glewwe and Jacoby 1995 and Glewwe et al 2006 for conceptual discussions and evidence). Thus, a non-exhaustive list of problematic aspects that obscure a causal interpretation of this relationship includes, as before, **community characteristics**, **household socioeconomic background**, **parental ability, parental preferences for education**, as well as child **early health and nutritional status** and **pre-school attendance**.

As before, exploiting differences in test performance between siblings simultaneously solves the problem of household and community unobservable characteristics that affect both siblings equally (numerated as items (a) to (d) in the previous example; they apply all the same to this example). Dealing with those child-level aspects mentioned above is not a problem conceptually, but data requirement is greater in this case, for two sets of indicators have to be observed for each sibling over time: health and nutritional indicators, Peabody scores and pre-school attendance at ages 4-5; and writing and reading test scores at age 7-8.

This type of strategy has been applied in the literature. For instance, using longitudinal data similar in nature to the Young Lives survey for the USA (the National Longitudinal Survey of Youth 1979), Todd and Wolpin (2007) used paired-siblings comparisons to measure the causal effects of cognitive skills accumulated at an early age on cognitive skills observed during adolescence. Similarly, they measured the causal effects of non-cognitive skills accumulated early in life on non-cognitive skills during adolescence. See also Cunha and Heckman (2008).

The strategy can easily be generalized to the study of other aspects of child development that are dynamic in nature; that is, situations where a child-given outcome is explained by capacities or characteristics acquired during earlier years. For instance, Alderman et al 2006 and Glewwe et al 2006 used a paired-siblings estimation as part of their strategy to identify the causal effects of early malnutrition on academic achievement during mid- and late childhood.

2.3 Sibling studies in health

Traditionally, sibling studies in health have been used for genetic and gene-environment-behavioural interaction studies (Wessel, Schork et al. 2007; Chen, Lin et al. 2009), to understand the role of genetic factors in disease causation (Wessel, Schork et al. 2007). Pairs of siblings (sib-pairs) allow for matching of genetic and environmental characteristics, for strict control of several confounding factors which may influence the pathogenesis and co-morbidities associated with diseases, including pre- and post-natal environment, household environmental exposures, sociocultural factors, thus making the genetic studies more robust (Wessel, Schork et al. 2007). However, at present the role of sibling studies

have expanded to understanding the environmental health risks pertinent to particular communities and ethnicity. The assumptions of gene and environment being independent aetiologies have been proved false by several population stratification studies which have found certain diseases and health risk factors to be more prevalent among particular communities defined based on geography or ethnicity (Chen, Lin et al. 2009). As siblings being born and brought up in similar household and community environment, sib-pair studies allow control of environmental factors to emphasize the impacts of genetic traits on diseases and health risks and thus help delineate the effects of shared environment on genes leading to differential outcomes (Chen, Lin et al. 2009).

In addition to genetic and environmental risk factors, the role of behavioural risk factors and life-stYoung Livese factors in disease causation and progression, particularly chronic / non-communicable diseases (NCDs) such as cardiovascular diseases, diabetes, obesity, cancer, etc. is well understood (Wiley and Camacho 1980; Patterson, Haines et al. 1994; Wilson 1994). Epidemiological studies use sib-pairs to understand these life-stYoung Livese and behavioural risk factors associated with NCDs. For example, 'migration studies' have compared individuals who migrated to a developed country (such as Canada, USA or UK) with their siblings still residing in their native developing or under-developed countries (such as, India, Bangladesh, etc.) (Kushi, Lew et al. 1985; Jeemon, Neogi et al. 2009). This study design allows for control of genetic and early life factors (common to siblings) to elicit evidence of the differential adult environmental and behavioural risk factors; such as dietary habits, physical activity, smoking, alcohol and other lifestyle-related factors facilitating incidence of obesity, cardiovascular diseases and diabetes (Kushi, Lew et al. 1985; Jeemon, Neogi et al. 2009). Sibling pairs control the familial, cultural and socio-economic environment to overcome the inherent problems of early life-course and ethnic comparability in such studies for better understanding of modifiable lifestyle-related risk factors in adult life (Jeemon, Neogi et al. 2009).

2.4 Behaviour – environment Interactions

As a longitudinal study following two groups of children for 15 years, Young Lives not only provides scope to study and further investigate the causal relationship between life-course events and physical and psychosocial health, but also has the advantage of a sib-pair study design. Genetic studies are not a probable area to explore with the Young Lives data and it does not have sib-pairs living across countries, but the existing sib-pair cohort can be used to study behaviour-environment interaction, to analyse the impact of non-shared environment (e.g. gender) and behavioural impacts on differential nutritional status (especially obesity), physical activity, smoking, alcohol and other lifestyle risk factors for chronic diseases. Further, if the cohort is followed until adulthood, or at least early adulthood, there is potential for examining the impact of differential lifestyle and behavioural risk factors among siblings on the initial physiological and biochemical markers of NCDs such as deranged systolic and diastolic blood pressure, impaired blood glucose, etc. (Wilson 1994).

While lifestyle and behavioural risk factor studies are used to assess and predict diseases, it is important to understand the developmental and life-course origins of these risk factors. Studies use sibling pairs to understand this complex process of behaviour and environmental interaction which leads to differential risk factors. Important examples include understanding adolescent smoking behaviour and substance use (alcohol, narcotic drugs, etc.). Studies in the last ten to 15 years, conducted across the world including the USA (Rajan, Leroux et al. 2003; Slomkowski, Rende et al. 2005), Australia (Madden, Heath et al. 1999), Sweden (Madden, Heath et al. 1999; Kendler, Thornton et al. 2000), Finland (Madden, Heath et al.

1999), Holland (Koopmans, Slutske et al. 1999; Vink, Beem et al. 2004) and Vietnam (True, Heath et al. 1997) have showed the effects of shared environment and social connectedness on smoking initiation and frequency among siblings during their adolescence. The findings were robust even after controlling for parents' and peers' smoking behaviour (Rende, Slomkowski et al. 2005). For example, Slomkowski et. al. used 1,421 sibling pairs (403 pairs of twins, 672 pairs of full-siblings, 165 pairs half-siblings and 181 pairs unrelated) from the National Longitudinal Study of Adolescent Health (Add Health) in the USA, to demonstrate the role of sibling influence on smoking initiation and frequency (Slomkowski, Rende et al. 2005). The study found that irrespective of genetic connectedness, adolescent smoking was influenced by social connectedness of siblings (Slomkowski, Rende et al. 2005). Likewise, studies on substance use and abuse have shown shared social and environmental influence to be greater predictors of initiation and dependency during adolescence and adulthood (Kosten, Ball et al. 1994; Rhee, Hewitt et al. 2003; Rende, Slomkowski et al. 2005).

Despite the existence of a large number of studies assessing familial risk factors and predictors of adolescent smoking and substance use, Avenevoli and Merikangas in their review of 87 such studies found that many of them were not methodologically sound because of their cross-sectional study design, lack of standard definitions for adolescent tobacco use and absence of reliable and valid tools for assessing tobacco use and the familial risk factors (sibling influence, socioeconomic environment, parents and peer influence) (Avenevoli and Merikangas 2003). This highlights the need for prospective longitudinal sibling studies, to further understand the sibling risk factors for smoking and substance use in general (Avenevoli and Merikangas 2003), and to identify appropriate phases in the life-course for targeted behavioural interventions (BoYoung Livese, Sanford et al. 2001). Young Lives has an advantage over the other studies (mentioned above) in being a prospective longitudinal study, following sib-pairs over a period of 15 years in multicultural and multi-ethnic settings. Thus it can be the platform for exploring the role of household and community environment, and parent's, peer's and sibling influences on initiation and dependency on substances including tobacco and alcohol.

3. Technical issue: Parental choices, child outcomes and the problem of isolating household unobservable characteristics

Suppose one is interested in testing the relationship between one specific strategy followed by parents to improve their child's performance at school and its effect on cognitive achievement measured a few years later. For illustrative purposes, consider the study of the impact that access to pre-school might have on child´s achievement on the Peabody Picture Vocabulary test at age 4-5; that is, the impact of pre-school on vocabulary knowledge. For simplicity, suppose pre-school attendance is entirely a parent's choice (i.e., is not compulsory). Pre-school enrolment is likely to be positively correlated to later cognitive achievement. Indeed, this is what one observes using Young Lives data. However, from this association it is not straightforward to claim that pre-school enrolment leads to a higher Peabody score. A non-exhaustive list of confounding factors includes:

a) **Community characteristics:** access to pre-school is limited to certain, typically wealthy, communities, and children in wealthier communities are more likely to attend better schools and, thus, to score higher in the Peabody test.

b) **Household socioeconomic background:** wealthier families find easier to afford both the monetary expenses associated to pre-school attendance as well as the monetary investments required to secure a successful performance at school.

c) **Parental ability:** some parents are better at raising well-educated children: they are more likely to send their kids to pre-school.

d) **Parental preferences for education:** parents that choose to send their children to pre-school are likely to have a higher valuation of the benefits of education and, thus, they are more likely to invest in the education of their children anyway.

Sometimes it is relatively simple to isolate the effect that certain household or community characteristics might play in explaining both the chance that a child is enrolled at pre-school and the subsequent performance of the same child at school. This is provided there is enough variation in terms of the living conditions faced by households within a given community and in terms of the characteristics of the communities sampled. Different studies about the profiles of the Young Lives samples in each country suggest that is the case.[2]

For instance, in the example at hand, one could limit the comparison to communities where there is a pre-school available within certain reach. This helps to convincingly deal with aspect (a). Similarly, the researcher can identify households comparable in terms of socioeconomic background and proceed to perform the comparison between households

2 See Young Lives Technical Notes 1, 2, 3 and 4. http://www.younglives.org.uk/our-publications/technical-notes

from the same background, thus partially dealing with aspect (b).[3],[4] However, in the case at hand, parental talent to raise well-educated children (c) and parents' motivation (d) could still be the driving force behind the results and can not be controlled for. This illustrates a common problem that afflicts studies that look for linkages between household-level choices and child-child-level results using 'one child-one household' information: household characteristics that are not observed could explain the correlation observed between any pair of household-level, child-level variables.

Possible solution 1 - Quasi-experiments: To deal with this, one possible approach would be to look for some kind of quasi-experiment.[5] For instance, one might look for changes over time in pre-school enrolment policies within the communities sampled. While this is a plausible approach, the extent to which it can be implemented using data from the Young Lives index cohorts is limited by the fact that the vast majority of these children were born between 2001 and 2002: one would have to be lucky enough to observe a meaningful change in pre-school enrolment policies between 2001 and 2002 in order to implement this strategy. The availability of data on siblings would be helpful; because to compare siblings means, by necessity, to compare children born over a wider time range (e.g., children born between 2001 and 2005 rather than between 2001 and 2002), it is more likely to apply this kind of strategy with sibling data than without it.

Possible solution 2 – Comparing siblings: Even outside the quasi-experimental setting, in situations like the one stated, the availability of sibling-level information becomes particularly handy. Suppose information was available for pairs of siblings across households on pre-school enrolment and Peabody scores at age 4-5. Differences in pre-school enrolment within the household can occur because of temporary shocks that affected one of the siblings at the age of expected enrolment (Young Lives dedicates a section of its surveys to the recollection of these events); differences could also arise due to differences in the month of birth, which typically affects enrolment decisions. Provided such variation exists, it is possible to make a comparison purely between siblings. In doing so, all the characteristics that are common between them -such as, crucially, parents' unobserved traits - no longer poses a problem for the study of the relationship of interest, because one is comparing children born from and raised by the same parents. Note that this solution is not particularly data intensive. In principle, only one or two pieces of information are required for the younger siblings of the index children: Peabody scores at the age of 4-5 and the history of the younger siblings' enrolment in pre-school institutions.

For instance Garces et al (2002) use this approach to evaluate the impact of a pre-school programme for disadvantaged children in the USA (Head Start) on schooling attainment as

3 In the context of developing countries, it is considered a good practice to measure differences in socioeconomic background in terms of differences in household monthly consumption. For instance, one can distinguish between households that live below and above the poverty line and proceed to compare differences in pre-school enrolment within poor and non-poor households, respectively. Similarly, one can proceed to compare households that are similar in terms of other characteristics associated to socioeconomic background, such as access to basic services.

4 More generally, one strategy would be to use as much information as possible about community characteristics and family and child background, since this help making a more precise comparison, i.e., to study the relationship of interest across 'comparable' children (born in the same community, raised in similar households, etc. For instance, consider a selected group of children that, say, live above the poverty line, with access to basic services, with well-educated parents that reportedly help their children with homework at home and attend most of the school/teacher meetings. Suppose that, within this group, it is still observed that children that were enrolled at pre-school score higher marks than their counterparts at school. This makes the case of a pre-school-cognitive achievement nexus stronger.

5 A source of variation in pre-school enrolment that one can argue is not associated to parental characteristics.

well as on earnings and criminal behavior. See also Berlinski et al (2008), a study of the impact of pre-school attendance on primary school performance in Uruguay. Conceptually, the same idea can be used to analyze other situations where it is likely that families self-select. For instance, Ruhm (2000) exploit changes over time in parental employment status to study the impact of maternal employment during the first three years of life of the child on pre-school cognitive development. Rees and Sabia (2009) use a similar approach (differences in parents' behaviour over time) to measure the impact of breastfeeding on educational attainment. Inasmuch as there are differences in life experiences between siblings driven by differences in date of birth, siblings data offers an opportunity to measure impacts on child-level outcomes.

As the index children age, this strategy could also be used to study the effects of their choices on later outcomes. Different studies in the USAs analyze the impact of teenage childbearing on socio-economic outcomes using a sisters comparison approach (see, for instance, Geronimus and Korenman, 1992). In the same spirit, the siblings approach has been used to study the impact of alchohol consumption by young adults on socio-economic status (Kenkel. 1994).

Adolescent development

Similar to medical sciences, sib-pairs are powerful designs in other social sciences than economics and are used to test socialisation theory such as the psychosocial acceleration theory (Belsky, Steinberg et al. 1991) and paternal investment theory (Draper and Harpending 1982) to study the effects of life-course and life-experiences on adolescent development characteristics. Tither and Ellis compared siblings to match the genetic and environmental confounders to investigate the causal role of family disruption on early pubertal maturation of girls (Tither and Ellis 2008), which in turn is a known risk factor of adolescent pregnancy, mood disorders and other health and psychosocial risks (Ellis, McFadyen-Ketchum et al. 1999; Ellis 2004). They examined the potential causal influence of degree of familial disruption/absence of father on onset of puberty by using a differential sibling exposure design (Tither and Ellis 2008). Age of onset of menarche in biological sister pairs from disrupted families were compared with that of a control group of biological sister pairs of intact families (Tither and Ellis 2008). Nested case-control studies using sister pairs in Young Lives can help to understand the role of various familial and household factors including poverty and economic shocks on onset of menarche and other developmental characteristics of adolescent girls in this cohort. Further examples of sib-pair studies in understanding adolescent physical and psychosocial development can be drawn from the Non-shared Environment in Adolescent Development (NEAD) project. NEAD was a longitudinal family study which followed twins and siblings from adolescence to early adulthood and provided evidence of gene-environment and parent-adolescent interaction influences on developmental characteristics including coping, stability and changes during adolescence (Reiss, Plomin et al. 1994; O'Connor, Hetherington et al. 1995; Neiderhiser, Reiss et al. 2007), and parental and/or sibling influences on differential health risks and psychosocial development. Therefore, along with understanding adolescent health risk behaviours, the sib-pair cohort of Young Lives can help to delineate the complexities of household environment, poverty, economic shocks and social events on puberty and 'adolescent adjustments' (Neiderhiser, Reiss et al. 2007).

4. Implications for the Young Lives study

The discussion above from both an economic and health perspective suggests the following recommendations for the study:

- For a number of questions that are academically relevant and of policy interest (e.g., the impact of anti-poverty programmes; the impact of education and health programmes; the impact of investing in early childhood and in schooling on later outcomes, the impact of teenage childbearing and alcohol consumption on socio-economic status, etc), the collection of data for pairs of siblings would offer richer opportunities to tackle issues of causality. In a non-experimental setting, the sibling is the best prediction of how would the index child do had her life conditions had been different.
- A minimum set of variables should be chosen to be collected for both the index child and her sibling over time.
- If five observations of the index children are planned, one ideally would like to have five observations of the younger sibling at comparable ages. Given that the project did not collect data on siblings in the first two rounds, in the cases when it is possible (e.g., educational and health history of the younger sibling) it is important to consider the collection of retrospective data in order to fill the current gap.
- Collecting such data is likely to have a fairly low marginal cost, given that the survey is already underway for the Index children.

References

Alderman, Harold, Jere R. Behrman, Victor Lavy, and Rekha Menon (2001) 'Child Health and School Enrollment: A Longitudinal Analysis', *The Journal of Human Resources* 36.1: 185-205

Alderman, Harold, John Hoddinott, and Bill Kinsey (2006) 'Long-term consequences of early childhood malnutrition', *Oxford Economic Papers* 58.3: 450-74

Ashenfelter, Orley and David J. Zimmerman (1997) 'Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons, and Brothers', *The Review of Economics and Statistics* 79.1: 1-9

Amarante, Veronica, Marco Manacorda, Edward Miguel, and Andrea Vigorito (2011) *Social Assistance and Birth Outcomes: Evidence from the Uruguayan PANES*, RES Working Papers 4714, Washington DC: Inter-American Development Bank, Research Department

Avenevoli, S. and K. R. Merikangas (2003) 'Familial influences on adolescent smoking', *Addiction* 98: 1-20

Belsky, J., L. Steinberg et al. (1991) 'Childhood experience, interpersonal development, and reproductive strategy: An evolutionary theory of socialization', *Child development* 62.4: 647-70

Behrman, Jere R. and Paul Taubman (1989) 'Is Schooling "Mostly in the Genes"? Nature-Nurture Decomposition Using Data on Relatives', *Journal of Political Economy* 97.6: 1,425-46

Berlinski, Samuel, Sebastian Galiani, and Marco Manacorda (2008) 'Giving children a better start: Pre-school attendance and school-age profiles', *Journal of Public Economics* 92.5-6, 1,416-40

Boyle, M. H., M. Sanford et al. (2001) 'Familial influences on substance use by adolescents and young adults', *Canadian Journal of Public Health* 92.3: 206-9

Chen, Y.-H., H.-W. Lin et al. (2009) 'Two-stage analysis for gene-environment interaction utilizing both case-only and family-based analysis', *Genetic Epidemiology* 33.2: 95-104

Cunha, Flavio and James J. Heckman (2008) 'Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation', *Journal of Human Resources* 43.4: 738-82

Draper, P. and H. Harpending (1982) 'Father absence and reproductive strategy: An evolutionary perspective', *Journal of Anthropological Research* 38.3: 255-73

Ellis, B. J. (2004) 'Timing of Pubertal Maturation in Girls: An Integrated Life History Approach', *Psychological bulletin* 130.6: 920

Garces, Eliana, Duncan Thomas and Janet Currie (2002) 'Longer-Term Effects of Head Start', *The American Economic Review* 92.4: 999-1,012

Geronimus, Arline T. and Sanders Korenman (1992) 'The Socioeconomic Consequences of Teen Childbearing Reconsidered', *The Quarterly Journal of Economics* 107.4: 1,187-1,214

Glewwe, Paul and Hanan G. Jacoby (1995) 'An Economic Analysis of Delayed Primary School Enrollment in a Low Income Country: The Role of Early Childhood Nutrition', T*he Review of Economics and Statistics* 77.1: 156-69

Glewwe, Paul and Hanan G. Jacoby and Elizabeth M. King (2001) 'Early childhood nutrition and academic achievement: a longitudinal analysis', *Journal of Public Economics* 81.3: 345-68

Griliches, Zvi (1979) 'Sibling Models and Data in Economics: Beginnings of a Survey', *Journal of Political Economy* 87.5: S37-S64

Kenkel, Donald S., David C. Ribar, Philip J. Cook and Sam Peltzman (1994) 'Alcohol Consumption and Young Adults' Socioeconomic Status', Brookings Papers on Economic Activity: Microeconomics 1994, 119-75

Rosenzweig, Mark R. (1986) 'Birth Spacing and Sibling Inequality: Asymmetric Information within the Family', *International Economic Review* 27.1: 55-76

Rosenzweig, Mark R. and Kenneth I. Wolpin (1988) Heterogeneity, Intrafamily Distribution, and Child Health', *The Journal of Human Resources* 23.4: 437-61

Rosenzweig, Mark R. and Kenneth I. Wolpin (1995) 'Sisters, Siblings, and Mothers: The Effect of Teen-Age Child-bearing on Birth Outcomes in a Dynamic Family Context', *Econometrica* 63.2: 303-26

Rees, Daniel and Sabia, Joseph (2009) 'The Effect of Breast Feeding on Educational Attainment: Evidence from Sibling Data', *Journal of Human Capital* 3.1: 43-72

Ruhm, Christopher J. (2004) 'Parental Employment and Child Cognitive Development', *The Journal of Human Resources* 39.1: 155-92

Todd, Petra E. and Kenneth I. Wolpin (2007) 'The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps', *Journal of Human Capital* 1.1: 91-136

THE AUTHORS

Dr Catherine Porter is a postdoctoral research fellow in the Department of Economics and Corpus Christi College, Oxford University. She is an applied microeconomist working on poverty, health, children, and aid effectiveness in developing countries.

Alan Sanchez is an economist in the Research Division of the Central Bank of Peru.

Manisha Nair is a medical doctor and has an MSc in Global Health Science from the University of Oxford, England. At present, she is a Weidenfeld DPhil scholar at the Department of Public Health, University of Oxford, England.

ACKNOWLEDGEMENTS

We are grateful for the valuable input we received in the drafting of this paper, in particular, we would like to thank Sofya Krutikova.

About Young Lives

Young Lives is an international study of childhood poverty, following the lives of 12,000 children in 4 countries (Ethiopia, India, Peru and Vietnam) over 15 years. **www.younglives.org.uk**

Young Lives is core-funded from 2001 to 2017 by UK aid from the Department for International Development (DFID), and co-funded by the Netherlands Ministry of Foreign Affairs from 2010 to 2014. Sub-studies are funded by the Bernard van Leer Foundation and the Oak Foundation.

Young Lives
An International Study of Childhood Poverty

www.younglives.org.uk

Young Lives, Oxford Department of International Development (ODID)
University of Oxford, Queen Elizabeth House
3 Mansfield Road, Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Email: younglives@younglives.org.uk