CHAPTER III
METHODS AND PROCEDURES

3.1 Research design

This research is based on part of the data from Young Lives study which is conducted from the year 2001 to 2015 on child well-being in Ethiopia, India, Peru and Vietnam (www.younglives.org.uk). The Young Lives study is a panel study that follows 2,000 children in each country from age 1 year until they are 15 years old. The caregiver and, when the child is old enough, both the caregiver and the child is interviewed every three to four years with a quantitative survey. The height and weight of each child is measured and community level questionnaires completed for each sentinel site at every data collection round.

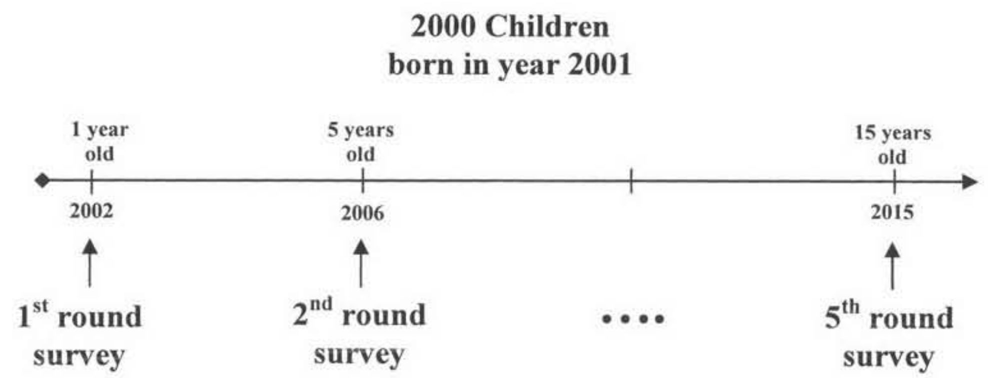


Figure 2: Young Lives study

This research uses only the ***data of Vietnamese children in the fist round survey*** conducted in the year 2002 when the children were 6-17 months. Thus, this research is similar to a cross-sectional study.

3.2 Target population

The target population of this research is all children 6 – 17 months of age in 5 provinces: Lao Cai, Hung Yen, Phu Yen, Ben Tre, and DaNang.

3.3 Research Settings

Five provinces in this research, Lao Cai, Hung Yen, Phu Yen, Ben Tre, and DaNang, show some interesting characteristics:

LaoCai: typical province for the North-East and North-West, high percentage of minority ethnic groups, mountainous, underdeveloped infrastructure.

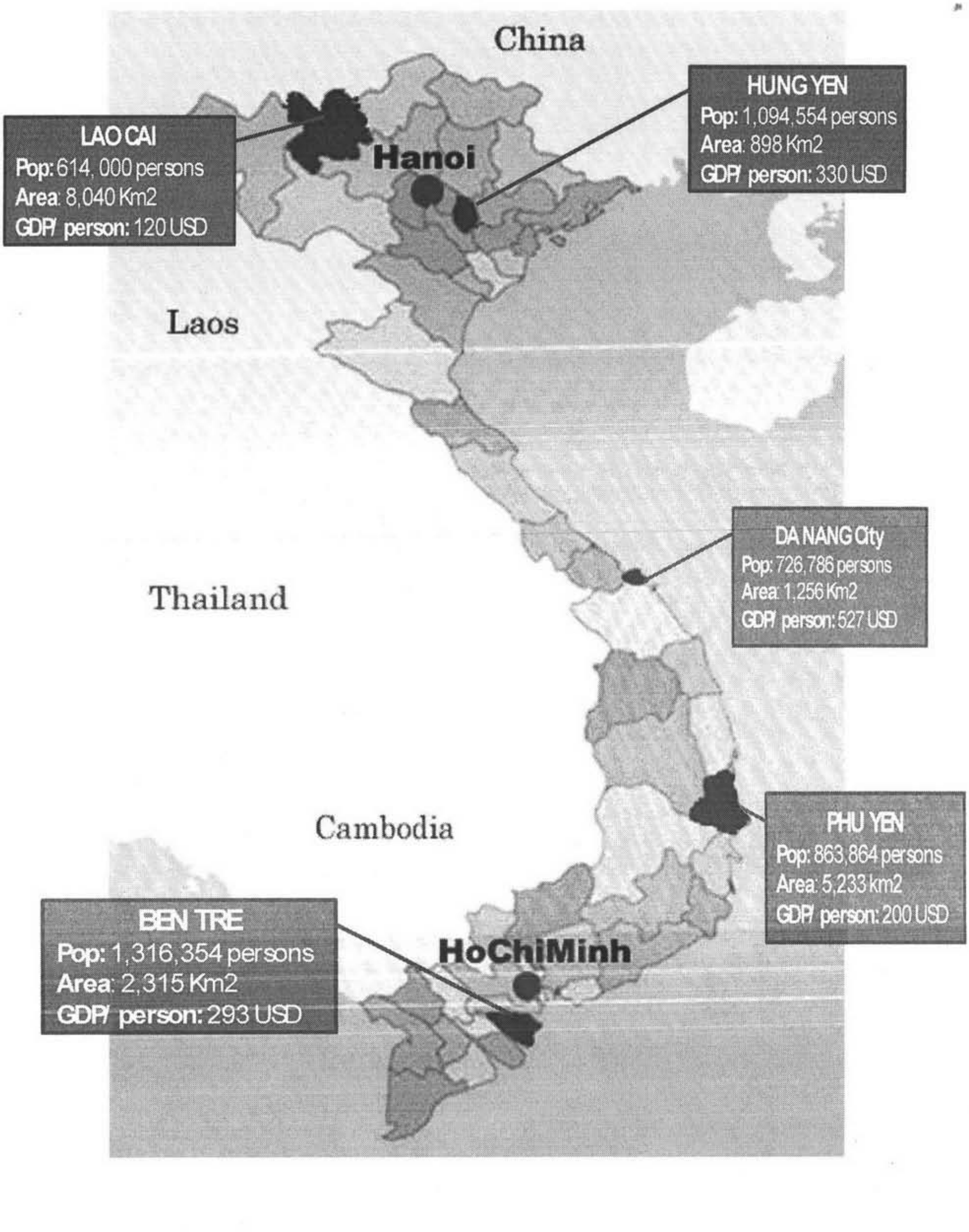
HungYen: typical rural representative of the Red River Delta region, populous; high population density, main source of income: rice farming, good infrastructure, near big cities (between Hanoi and Hai Phong), being influenced by urbanization.

PhuYen: typical province for central coast of Vietnam, suffering severe natural disasters, severely damaged by wars, mix of coastal, midland and highland, poor infrastructure, main sources of income from agriculture and seafood.

BenTre: typical province for Mekong River Delta, severely damaged by flooding, high percentage of landless families in rural area, low educational level.

DaNang: ranking in the middle of the five biggest cities in Vietnam, rapid urbanization, high level of education, receiving big investment from international sources and government, poverty rate is low.

Figure 3 – Study areas: LaoCai, HungYen, DaNang, PhuYen and BenTre



3.4 Sampling and sample size

Sampling process of Vietnam Young Lives study:

The sampling of Young Lives study is *sentinel site sampling with semi-purposive of sites* [45]. A sentinel site is defined as a commune. The sampling steps were: (1) In each of five provinces, ranking all communes in the province by poverty level: poor, average, better off and rich; (2) selecting 2 communes in the poor group, 1 in average, and 1 in the better off and rich which were most representative for each group providing four sentinel sites in each province; (3) screening and listing eligible children; (4) selecting a sample of 100 children in each sentinel site using simple random sampling [39].

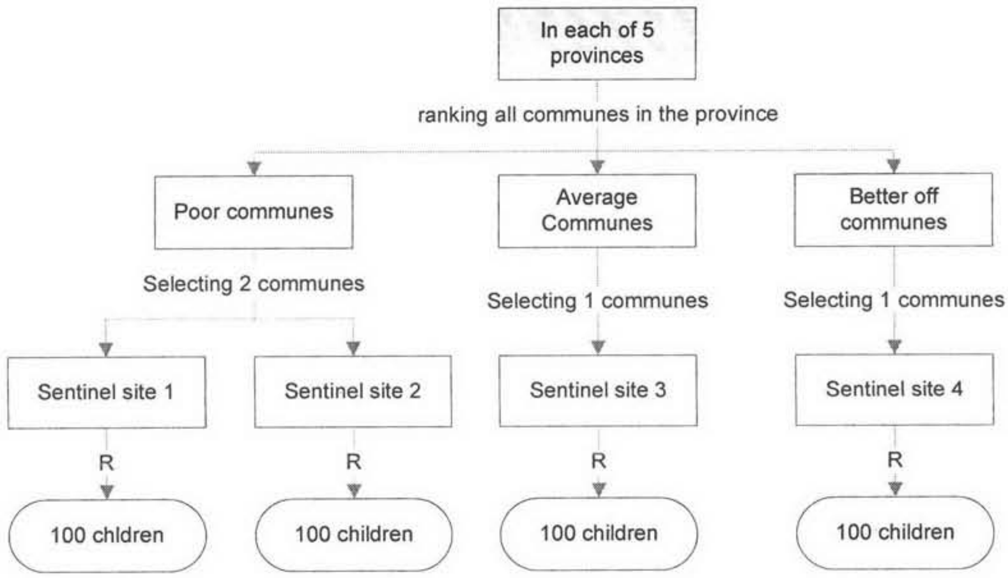


Figure 4 : Sampling process of Young Lives study

Sample size:

The sample size of Young Lives study is ***2000 children (400 children per province)***.

Inclusion and exclusion criteria

For this research, all of the children of Young Lives study will be involved. However, there were 6 children excluded from all analyses because data collectors could not measure their anthropometrics. For the analyses of prevalence of child malnutrition, there are no exclusion criteria. At first, in the thesis proposal, I planned to exclude all children who have genetic, congenital, or chronic diseases related to short stature when analyzing association between child nutritional status and other variables. Nevertheless, when examining the database I found that the information on the diseases of surveyed children was not reliable. The data collection method for the Young Lives study was to interview caregivers using a structured questionnaire. The interviewers were not doctors. There was only one question which collected information on diseases. Therefore, the information gathered was not reliable enough to diagnose diseases for any children. In the end, I decided to not exclude any of the children from the analysis.

3.5 Research Variables

Independent variables:

- **Child:** sex, age, birth order, birth interval, chronic and acute diseases

- **Caregiver:**

  + Primary information: age, ethnic, occupation

  + Education: years of education

  + Mental health: depression (Self-reporting questionnaire 20 items (SRQ20) recommended by WHO, validated in Vietnam)

  + Social support: Active member of community groups, Receiving any spiritual/physical support

- **Child care:**

  + Frequency meet the child of the biological mother and father

  + Sending child to childcare facility since born or not

  + Duration of breast-feeding:

- **Household level:**

  + Wealth index: built around the five elements of household capital (natural, physical, human, financial, social)

  + Number of economic shocks

  + Source of drinking water, toilet facility

  + Highest education in the household: years of education

- Community level:

Physical and social environment:

+ Urban/rural

+ Distance to the regional/district capital

+ Total land area, population

+ Activities used land most

+ Ecological classification for commune

+ Natural disasters occurred in the last 4 years

+ Largest ethnic groups in the commune

+ Major religious groups in the commune

Infrastructure and access:

+ Services available in commune: public telephone, piped water, electricity power.

+ Market available in the commune

+ Loudspeaker system available at hamlet level

+ Type of toilet used commonly in commune

+ Main source of drinking water

+ How rubbish is treated

+ Distance from the commune center to the nearest road accessible by car

+ Health and development programmemes in the commune

The economy:

+ 3 main economic activities undertaken by members of commune

+ Average income per capital/month in the commune

+ The main sources from which people in commune borrow money

Health care:

+ Distance to the nearest public hospital

+ Health programmes are being carried out by the commune health station

+ Number of traditional, western medicine clinics, and medicine shops in the commune

Outcome variables:

In this research, anthropometric methods were used to assess the nutrition status of children. Height and weight measurements of the children, taking age and sex into consideration, are converted into Z-scores based on the National Center for Health Statistics (NCHS) reference population to build three anthropometric indices, height-for-age z-score (HAZ), weight-for-age z-score (WAZ), and weight-for-height z-score (WHZ) [46]. The Z-score cutoff point recommended by WHO to classify low anthropometric levels is 2 SD units below the reference median for each of the three indices [47].

3.6 Research instruments

In the Young Lives study, as the children were less than 24 months of age at the time of the survey, *recumbent length* was measured using a standard technique [48] to the nearest 1 mm with a portable wooden measuring board (ShorrBoard, Shorr Productions, USA) with moveable head-panel. *Body weight* of lightly clad children was measured using standard techniques [48] on a Seca UNICEF solar powered electronic scale and read to the nearest 0.1 kg. If the child was unable to stand on the scale, both the mother and child were weighed together, and the weight of the mother subtracted from the combined weight by a computer function in the electronic scales.

Other variables for the children, caregivers, and households were collected by household questionnaires (Annex 1). Besides that, a community questionnaire was designed for collecting information at community level (Annex 2). Those questionnaires were designed by an academic consortium involving the University of Reading, the London School of Hygiene and Tropical Medicine, London's South Bank University, the University of Sussex, and the South African Medical Research Council, and Vietnamese Principle Investigator. The questionnaires were also commented on by the Vietnamese Policy Advisory Committee and Technical Advisory Committee members. The questionnaires were also pre-tested and adapted for the local culture.

3.7 Data collection

A training course was undertaken for both anthropometrics measurers and interviewers. Training of anthropometric measurement techniques for the measurers was carried out by an expert from the National Institute of Nutrition. Anthropometry standardization exercises were conducted twice during the training of the anthropometry survey team. These measurements were recorded in a spreadsheet and a simple method was used to assess the precision and accuracy of the trainees' measurements in comparison to the measurements recorded by the supervisor. The standardization exercises are used to identify trainees who needed more training and those anthropometric techniques that needed improvement. Questionnaire interviewers, who were staff from the General Statistics Office, went through a ten-day training course including two-day testing in fieldwork. They also had to pass an end of -course test.

In the first round survey, data was collected by 4 anthropometrics surveyors and 20 interviewers divided into 2 teams. The fieldwork was conducted from July to November 2002 under two levels of supervision and quality control. At the first level, there was a team leader and a supervisor of the data collection team who were responsible for administration and technical supervision. At the second level, quality control staff checked all completed questionnaires again before packing and sending to the central office. The first round data collection was done in a three-month period.

The respondents for the household questionnaire were the head of household and the primary caregiver. For the community questionnaire, the information was gathered by interviewing the leader of the commune or people responsible for the issue as well as from secondary data.

3.8 Data management and Analysis

Data entry and data cleaning was carried out by the Information Technology Unit of Research and Training Centre for Community Development. This process was done in Microsoft Access software with 10% double data entry.

3.8.1 Analysis strategy

The data analyses were performed in STATA version 9.0 in 3 stages below:

Stage 1 – Descriptive analysis

At the first stage, the distributions of HAZ, WAZ, and WHZ were described by descriptive statistics and graphic methods by provinces. All other variables were also be described the distributions in this step.

Stage 2 – Estimation of the prevalence of malnutrition

At the second stage, the estimated prevalence of stunted, underweight, and wasted children were calculated for each province. Since this research is designed as a complex survey (multistage sampling design), *Svy commands* in STATA [49] were applied to do these analyses that give estimates adjusted for sampling design. Options were used in *Svy commands* include *psu, strata, pweight, fpc*.

Stage 3 – Association analysis

In the analysis of association, only HAZ were involved to reflect the child nutritional status. The reason is that this research focuses on the community determinants of child nutritional status. These determinants influence child nutritional status in long-term. Height-for-age indicates the long-term nutritional status of the child. Therefore, HAZ is the most suitable.

To determine the association of potential factors and child nutritional status in this research, the *multilevel regression analysis* was applied. This is a quite complicated and newly developed technique. The rationale and basic principles are explained below.

The data structure in this research is hierarchical in which children are nested within their communes. It is also called multilevel data with level 1 is child and level 2 is commune. In such multilevel research, variables can be defined at any level of the hierarchy. Variables are measured at their natural level; for example, at commune level, I measured population, land area, and at the child level I measured sex, age. In traditional methods, it is usual to move variables by aggregation or disaggregation to one single level of interest, followed by ordinary statistical tests of association or ordinary multiple regression. Aggregation means that the variables at a lower level are moved to a higher level, for instance by computing commune mean child HAZ. Disaggregation means moving variables to a lower level, for instance by assigning to all children a variable that reflects the population of the commune they belong to.

Analyzing variables from different levels at one single common level creates two different sets of problems [50]. One of them is statistical. If data are aggregated, the result is that different data values from many subunits are combined into fewer values for fewer higher level units. Information is lost, and the statistical analysis loses power. On the other hand, if data are disaggregated, the result is that a few data values from a small number of super-units are "blown up" into values for a much larger number of subunits. Ordinary statistical tests treat all these disaggregated data values as independent information from this much larger sample. The proper sample size for these variables is of course the number of higher level units. Using the higher number of disaggregated cases for the sample size leads to significance tests that reject the null-hypothesis far more often than the nominal alpha level suggests. The other set of problems encountered is conceptual. If the analyst is not very careful in the interpretation of the results, s/he may analyse data at one level, and draw conclusions at another level.

The most practical way to look at multilevel data is to investigate cross level hypotheses, or multilevel problems. Multilevel models are designed to solve this kind of problems in which all variables from different levels are analyzed simultaneously using a statistical model that includes the various dependencies. There are many multilevel models and the *multilevel regression models* are essentially a multilevel version of the familiar multiple regression model. The full *multilevel regression model* assumes that there is a hierarchical data set, with one single dependent variable that is measured at the lowest level and explanatory variables at all existing levels. This technique, *two-level regression model*, is the most suitable for the research question and data structure of this research.

The principle of constructing the *two-level regression model* is that conceptually the model can be viewed as a hierarchical system of regression equations. We can set up a separate regression equation in each separate commune to predict the dependent variable Y by the explanatory variable X as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \qquad (1)$$

In this regression equation β_{0j} is the usual intercept, β_{1j} is the usual regression coefficient, and e_{ij} is the usual residual error term. The subscript j is for the communes and the subscript *i* is for children. The difference with the usual regression model is that we assume that each commune is characterized by a different intercept coefficient β_{0j} and also a different slope coefficient β_{1j}. In other words: the intercept and slope coefficients are assumed to vary across the communes.

Across all communes, the regression coefficients have a distribution with some mean and variance. The next step in the hierarchical regression model is to predict the variation of the regression coefficients by introducing explanatory variables at the commune level, as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \qquad (2)$$

and

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \qquad (3)$$

Equation (2) states that the intercept β_{0j} can be predicted by variable Z of commune level. The equation (3) states that the relationship between outcome variable Y and the variable X depends upon the variable Z. The u-terms u_{0j} and u_{1j} in equation (2) and (3) are residual error terms at commune level, and to be independent from the residual errors e_{ij} at child level.

The model with one child level and one commune level explanatory variable can be written as one single complex regression equation by substituting equations (2) and (3) into equation (1):

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{0j} + e_{ij} \qquad (4)$$

In general, there are more than one explanatory variable at every level. Assume that we have *P* explanatory variables X at the child level, indicated by the subscript *p* (p=1..P). Likewise, we have *Q* explanatory variables Z at the commune level, indicated by the subscript *q* (q=1..Q). Then, equation (4) becomes the more general equation:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{pij} + \gamma_{0q} Z_{qj} + \gamma_{pq} Z_{qj} X_{pij} + u_{pj} X_{pij} + u_{0j} + e_{ij} \qquad (5)$$

The ordinary single level regression model would estimate only the intercept and (P+Q) regression slopes. Conversely, even with a modest number of explanatory variables at both levels, equation (5) implies a very complicated model (P regression slopes for the level 1, Q regression slopes for level 2, and PxQ regression slopes for cross level interactions). Usually, we do not want to estimate the complete model, because this is likely to get us into computational problems, and also because it is very difficult to interpret such a complex model.

Fortunately, computer programmes like HLM, VARCL, GLLAMM, and MLn allow us to specify which regression coefficients are assumed to vary and which not, and to include only a few selected cross level interactions. So that, generally we limit ourselves to parameters that have proven their worth in previous research, or are interesting in view of our theoretical problems. However, there still are so many variables in my research, so that I will use an exploratory procedure to select a model.

Analysis strategies of building multilevel regression model

This analysis was performed in STATA version 9 (using *xtmixed* command). The exploratory procedure was used to select a best model. The procedure is to start with the simplest possible model, the intercept-only model, and to include the various types of parameters step by step. At each step, we inspect the results to see which parameters are significant, and how much residual error is left at the two distinct levels.

In detail, there are 4 models were built. In the first model, the *intercept-only model*, I analyzed a model with no explanatory variables. The second model was built with only child variables. The caregiver and household variables were added in the third model. In last, the model consisted of all significant variables in previous models and community variables. In each model, only variables having significant association with the outcome variable were kept in the model and moved to the upper model. In this analysis, significant level of p-values is less than 0.1.

3.8.2 Indices calculation

Nutritional indices: Anthropometric indices were computed using the EpiNut module of Epi Info 2002. This module uses the CDC 2000 reference and calculates Height-for-Age Z-Score, Weight-for-Age Z-Score, and Weight-for-Height Z-Score from child's height, weight, age and sex.

Wealth index: this is a composite variable that measures the poverty level of a household. The variable was constructed by Principle Components Analysis from information on household ownership of durable goods, housing characteristics, basic services, and human resource. There are some other approaches, "direct" measures, such as income, expenditure or consumption. Those are both expensive and difficult to collect, therefore, this study does not apply those approaches.

The Household Wealth Index was constructed using three steps. In the first step, 14 variables were chosen to represent household durable assets (four variables), housing characteristics (four variables), services (three variables), human resource (two variables), and poverty classification by community leaders (Table 1). In the second step, using principal component analysis, the asset index, A_i, for individual household i was defined as

$$A_i = \sum_k \left[f_k \frac{(a_{ik} - \overline{a}_k)}{s_k} \right],$$

where a_{ik} is the value of asset k for household i, $\overline{a}_k$ is the sample mean, and s_k is the sample standard deviation. The household score index, A_i, is the sum of scores of the included variables weighted by the elements of the first eigenvector.

In the third step, households were classified into living standards quintiles. Quintiles grouped by cut-off points of 20%, 40%, 60%, and 80% of the household wealth index score distribution were used stratified by provinces. In some instances, three groups: poorest (bottom 1 third), middle (next 1 third) and the rich (the top 1 third) or two groups: the poor (bottom 40%) and the non-poor /better off (upper 60%) were used.

Table 1 Household asset items used in constructing household wealth index using principal component analysis

Household asset item	Variable description
Household durable assets:	
Motorbikes	0 = No; 1 = Yes
Television set	0 = No; 1 = Yes
Telephone	0 = No; 1 = Yes
Refrigerator	0 = No; 1 = Yes
Housing characteristics	
Number of rooms per person considering household equivalence scale	room/person
House wall material	1 = Brick/concrete; 0=other
House roof material	1=concrete/tiles/slates; 0=other
House floor material	1=Cement/tile/laminated;0=Other
Services	
Main source of drinking water	1= running or bored well,0 = other
Kind of toilet facility	Water flushed latrine/semi-sealed latrine= Yes, 0 = other
Type of fuel for cooking	1=Gas/electricity; 0=other

Household human resource	
Highest educated	1=not yet going to school, 2=primary; 3=secondary; 4=high school; 5=above
Human resource = members (*able-bodied, aged 18-<60 years old, * 1.5 if completed primary, * 2 if completed secondary education, * 2.5 if completed higher education*)/household size	ratio
Household wealth status classified by commune leaders using the government's income per capita criteria	1 = Poor; 0= average or better.

3.9 Ethical considerations:

In each phase of the Young Lives study – including the selection of sentinel sites and eligible children, and conducting interviews – written consent forms for all aspects of the research were sent to participants and local government at different levels (i.e. provincial, district, and communal). Besides getting research ethics approval from London South Bank University UK, London School of Hygiene and Tropical Medicine UK, and Reading University UK, in Vietnam ethical approval for the Young Lives project was granted by VUSTA [39].

For this research, the proposal got the approval of Research Ethics Committee of Chulalongkorn University.

3.10 Identification of limitations

The Young Lives study is a well-designed study with high quality data. At the moment only the first round data is available. Therefore, my research, based on the data of Young Lives study, can only assess the associations between some variables and child nutritional status.

3.11 Generalising from the findings

The findings of this research are based on data from 5 provinces. They should not be generalised for all Vietnamese children. The results can only be generalised for children from 6 – 17 months in Lao Cai, Hung Yen, Phu Yen, Ben Tre and DaNang.