Technical Note 63

Young Lives

# A Guide to Young Lives Constructed Datasets: Rounds 1 to 7

Amanda Woodman Deza, Marta Favara
Maria de los Ángeles Molina and Alan Sánchez



www.younglives.org.uk

DECEMBER 2025

# A Guide to Young Live's Rounds 1 to 7 Constructed Datasets

Amanda Woodman Deza, Marta Favara,
Maria de los Ángeles Molina, Alan Sánchez

First published by Young Lives in December 2025



Printed on FSC-certified paper from traceable and sustainable sources.

**About Young Lives**

Young Lives is an international study of poverty and inequality, following the lives of 12,000 children in four countries (Ethiopia, India, Peru and Vietnam) since 2002. **www.younglives.org.uk**

Young Lives is funded by UK aid from the UK Government.

The views expressed are those of the author(s). They are not necessarily those of or endorsed by, Young Lives, the University of Oxford, the UK Foreign, Commonwealth & Development Office or other funders.

**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@qeh.ox.ac.uk

# Summary

The Young Lives study has conducted seven survey rounds since 2002, with the latest survey (Round 7) completed in 2023–2024. To make the data accessible and support the data users, the Young Lives study archives both the raw datasets and a constructed dataset for each of the study countries. The aim of this technical note is to provide detailed information to Young Lives data users on the information available in the constructed datasets from Round 1 to Round 7.

# The authors

Amanda Woodman Deza (Young Lives, University of Oxford) led the preparation of this technical note. Marta Favara (Young Lives, University of Oxford), Maria de los Ángeles Molina (Young Lives, University of Oxford) and Alan Sánchez (Young Lives, University of Oxford) contributed to reviewing and editing this note. All co-authors were part of the team that oversaw the Round 7 data collection, and they all contributed to the conceptualization and methodology of this note. Amanda Woodman Deza led the preparation of the Round 1 and Round 7 constructed datasets—under the supervision of the other co-authors—and she and Molina contributed to related data curation and validation activities.

# Acknowledgements

The constructed datasets and this technical note are an update of previous work by Sophie von Russdorf, Kristine Briones, Bridget Azubuike and Patricia Espinoza.

Thanks to Penny Rudling for copyediting and designing this report and for oversight of its publication. We also thank the UK's Foreign, Commonwealth & Development Office (FCDO) for funding Young Lives at Work and enabling this work.

Young Lives is a collaborative partnership between research institutes, universities and NGOs in the study countries and the University of Oxford. The constructed datasets include data from Round 7. We are grateful to the fieldworkers in Ethiopia, India and Peru for their dedication and invaluable contributions throughout the administration of Round 7. A special thanks to the Young Lives participants and their families, whose generosity in sharing their time and experiences have made this work possible. Round 7 has been core-funded by FCDO, with additional funding from Old Dart Foundation in Peru.

# Contents

# 1. Introduction

Young Lives is a longitudinal, mixed-methods, cohort study following the lives of 12,000 children and their families in Ethiopia, India (states of Andhra Pradesh and Telangana), Peru and Vietnam (Favara *et al.*, 2021), over more than 20 years. The study has conducted seven survey rounds since 2002 (five in-person and one over the phone), with the latest survey (Round 7) completed in 2023-2024 (see details in Box 1).

As per previous survey rounds, the data collected is publicly archived. To make the data accessible and support the data users, the Young Lives study archives two versions of the data: a set of raw datasets for each Round of data collection and a constructed dataset for each of the study countries, exploiting the longitudinal nature of the data. The previously released constructed datasets (Briones, 2018) included data from Round 1 to Round 5. The updated version Round 1 to Round 7 constructed dataset provides: (a) all information of the previously archived Round 1 to 5 datasets; (b) updated with data from Round 6—the "Listening to Young Lives at Work: COVID-19 Phone Survey" administered in 2020–2021 over five Calls —and Round 7; and, (c) adds additional information relevant to the age of our participants and related to labour market, family formation, obesity and overweight indicators for adults and mental health.

**Box 1: The Young Lives study**

The Young Lives sample consists of two cohorts born seven years apart in each country: a Younger Cohort (~2,000 participants) aged one when the first round of the survey was carried out in 2002 and an Older Cohort (~1,000 participants) then aged eight.

Between 2002 and 2016, five rounds of in-person surveys were completed, complemented by qualitative fieldwork. In response to the global COVID-19 pandemic, Round 6, originally scheduled for 2020, was conducted via phone (Molina *et al.*, 2025) over five phone calls between 2020 and 2021 (Molina *et al.*, 2025).

In 2023–2024, Young Lives completed its seventh survey round (Round 7) with the Younger and Older Cohort participants in Ethiopia, India and Peru (Molina *et al.*, 2025). On this occasion, data was not collected in Vietnam due to a change in governmental restrictions for the international transfer of personal data. Due to ongoing armed conflict in a few study sites in the Amhara region of Ethiopia, 205 participants living in conflict-affected zones (8.7% of the Young Lives sample in Ethiopia) were interviewed over the phone and a shorter survey questionnaire was designed for this purpose.

Twenty-one years since the Young Lives study began, an average of 81.0% of the original participants remain part of the study across all three countries, with attrition rates of 25.6% in Ethiopia, 11.5% in India and 19.8% in Peru (Molina *et al.*, 2025).

Figure 1 provides an overview of when each round was conducted, as well as the average ages of participants in each round for both cohorts. In the case of Round 6, the age of the participants at the moment of call 3 and call 5 is reported.

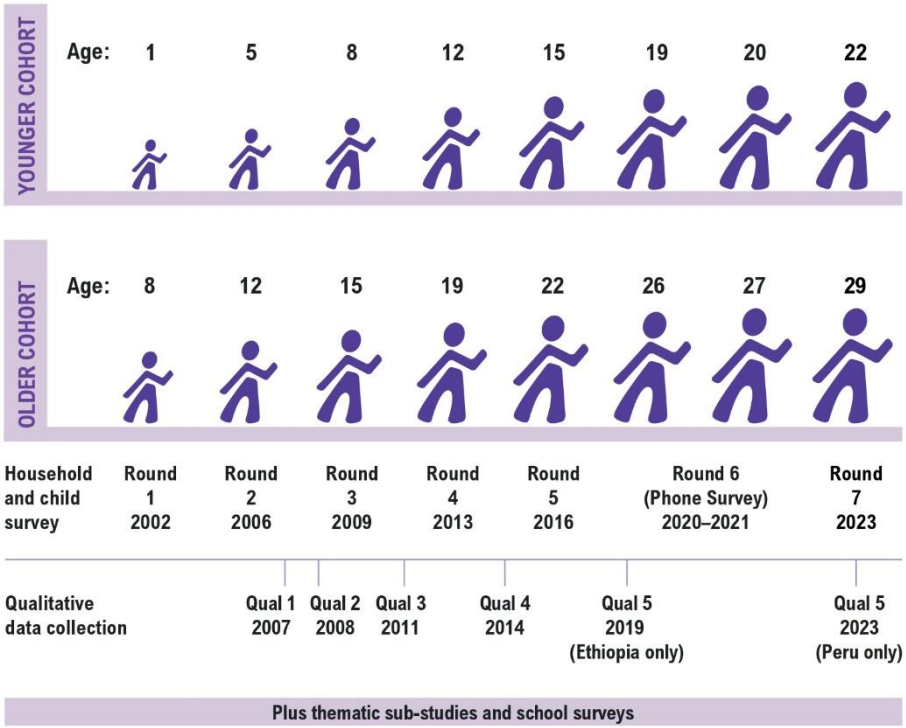


*Figure 1: Average age in years by Round and Cohort*

The aim of this technical note is to provide information to Young Lives data users on all the variables available in the constructed datasets from Round 1 to Round 7. First, it provides their definition—or refers the reader to where this information can be found—and availability in each round. Second, it describes how constructed variables were calculated. Third, it highlights the relevant differences in how questions were phrased or defined across rounds, cohorts and countries and outlines the assumptions made when questions changed over time. The note also helps users become familiar with the data collected across rounds and directs them to relevant additional information available in the archived raw datasets on the UK Data Service.

The document is organised as follows: ***Section 2*** describes the criteria used for selecting variables in the constructed dataset. ***Section 3*** presents unique identifiers and variables indicating the panel structure. ***Section 4*** summarises location-related variables. ***Section 5*** covers individual-level characteristics such as anthropometric measurements, learning and education outcomes, family formation, knowledge of sexual and reproductive health, smoking and drinking habits, immunisation, well-being, mental health and labour market outcomes. Finally, ***Section 6*** focuses on parental and household characteristics, including participants in social protection programmes, experiences of food insecurity, ownership of livestock, exposure to shocks, access to public services and basic characteristics of the mother, father, household head and caregiver.

Where applicable, three icons are added in each subsection to aid with discussion.

| | |
|---|---|
| ! | Reminders when using the variables. This includes the availability, country specific variations and changes in definition of variables across rounds. |
| ⬇ | Identifies other data relevant to the variables in the subsection that are available for download from the raw datasets. |

Provides information on whether this section was administered during the Round 6 and by phone in Ethiopia in Round 7 and pinpoints changes in the administration of specific sections or questions.

# 2. Rationale for Variable Selection in the Constructed Dataset

Since 2002, the Young Lives study has collected a broad range of child welfare indicators, in relation to physical health, growth, nutritional status, cognitive development, social and emotional well-being and life skills development (Barnett *et al.*, 2013). In Rounds 4 and 5, the survey expanded to include age-relevant topics for the cohorts, such as labour market participation, participation in economic activities, family formation and fertility and digital skills (Favara *et al.*, 2021). In Rounds 6 and 7, the study collected for the first-time data on the mental health of our participants, as well on the social impact of the pandemic or the exposure to the conflict in Ethiopia. For a detailed overview of the data collected across all survey rounds, cohorts and countries, see Mihaylova *et al.*, (2025). The key evidence emerging from Young Lives research are summarised in the four thematic Legacy Reports on Health and well-being, Education and learning, Labour Market and Family lives ((2025), Young Lives; forthcoming)

The Young Lives survey consists of three different questionnaires administered at the individual, household and community levels in the primary communities where the participants lived. In addition, a self-administered questionnaire (SAQ) has been introduced since age 15 (since Round 3 for the Older Cohort and since Round 5 for the Younger Cohort). In Round 7, a blended version of the household and individual questionnaires from previous rounds was administered to the participant (previously referred to as 'child'). Unlike earlier rounds, where their caregivers usually answered household-related questions, participants now respond to these themselves because they are now adults. Additionally, the community questionnaire was not administered in Round 7.

For the Round 1 to Round 7 constructed dataset, we select variables that were included in the constructed datasets from Rounds 1 to 5. Some of these questions were also asked in Round 6 (in at least one of the Calls) and are therefore included in this dataset. Additionally, we include information collected for the first-time during Round 6, which was also collected in Round 7 (such as anxiety and depression symptoms). Finally, we add new variables asked in previous rounds, mostly to the Older Cohort since Round 4 and to both cohorts during Round 6 and 7, such as marital status and questions related to main work activity during the past week. These additional variables are introduced to support researchers in gaining preliminary insights about Young Lives data and to reflect the relevant questions asked to the household and participants in early adulthood. The constructed datasets include information consistently collected across rounds and across countries.

There are four constructed datasets, one for each Young Lives country, available for download.[1] The datasets and Stata syntax (.do) files are available for download at the UK Data Service website.

[1] Given that data was not collected in Vietnam in Round 7, the constructed dataset in Vietnam includes information from six rounds (Round 1 to Round 6).

Throughout the technical note, Young Lives participants are referred to as index participants or participant. However, previously, we used to refer to them as 'index children' and some labels and variables names still do so. Variable names are written in italics.

# 3. Unique identifiers and panel structure

The unique identifiers and variables representing the structure of the data are reported in Table 1. In Round 1, a unique identifier (*childid*) was assigned to each Young Lives participant. This ID was retained in subsequent rounds to track participants over time. The *childid* is an alpha-numeric identifier composed of eight characters. The first two characters represent the country code (**ET** for Ethiopia, **IN** for India, **PE** for Peru, **VN** for Vietnam). The next two characters indicate the cluster ID (*clustid*)[2], followed by a single digit that identifies the cohort: ***1*** for the Younger Cohort and ***0*** for the Older Cohort. The final three digits form a sequential numeric identifier ranging from ***001*** to over ***100***. Starting from Round 4, the raw datasets no longer include the full *childid* but instead use the *childcode*, a six-digit numeric identifier derived from the last six characters of the original *childid* (i.e., excluding the country code).[3]

Our constructed dataset includes one observation per participant for each Round and Call. The Round 1 sample serves as the reference for each country, with the following sample sizes: 2,999 in Ethiopia, 3,019 in India, 2,766 in Peru and 3,000 in Vietnam. A *panel* variable identifies participants interviewed in all rounds. For Round 6, participants are considered part of the panel if they were interviewed in at least one call. Additionally, a *status* variable has been created to outline reasons why the participant was not interviewed during each survey round.[4] The variable includes the following categories; 0 – interviewed completed, 1-refused, 2- deceased, 3-migrated abroad, 4- not found (moved, not answered, not reached), 5- others.[5]

If a participant refused to participate during a specific round, all variables were set to missing, except for the basic identifiers (*childid, yc, call, round*).

The variables *inround* and *incalls* help users identify which participants completed specific rounds and Calls. The variable *inround* and *incall* takes the value of one if the Young Lives participant consented to be part of the study.[6] For Round 6, *inround* takes the value of one if a participant was interviewed in at least one call, whereas *incall* provides information on participation in each specific call. In each constructed dataset, all six in-person rounds (Round 1 to Round 5 as well as Round 7) and all five Calls that constitute Round 6 are included, resulting in *11 rows* per participant.[7] The *roun*d variable takes the value of 6 for all five Calls. For Ethiopia, we include an additional variable (*in_phone_r7*) to help

[2] More details about the sampling design are available for Ethiopia (Outes-Leon and Sanchez, 2008), India (Kumra, 2008), Peru (Escobal and Flores, 2008) and Vietnam (Nguyen, 2008).
[3] For example, in the *childid* **PE011001**, **PE** refers to Peru, **01** is the cluster ID, **1** denotes the Younger Cohort and **001** is the child's unique number within that cluster and cohort. In this example, **PE011001** becomes **11001**.
[4] This variable replaces the variable *deceased* in the constructed dataset from Round 1 to Round 5.
[5] Others includes country-specific reasons. For India, this category includes participants who were not interviewed because they were serving in the army or were found, but their interviews could not be recorded. For Ethiopia, one additional category was considered: 6 – unable due to conflict.
[6] There are cases, especially in Round 5, where the household was interviewed but the Young Lives participant was not. In these cases, *inround* takes the value of zero, and all variables at the child and household level are missing.
[7] As data were not collected in Vietnam in Round 7, the constructed dataset for this country contains only 10 rows.

users identify those participants who were interviewed via phone in Round 7 (see Box 1). All variables and their availability in each Round are reported in Table 1.

*Table 1 : Unique identifiers and panel structure*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **childid** | Participants ID | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **yc** | Younger Cohort = 1; Older Cohort = 0 | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **round** | Survey Round (takes the value of 6 for the five Calls that are part of the 2020-2021 phone survey (Round 6) | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **call** | Survey call | 1,2,3,4,5 | |
| **inround** | Participant was interviewed in round or at least one of the five Calls that are part of Round 6 | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **incall** | Participant was interviewed in call | 1,2,3,4,5 | |
| **panel** | Young Lives questionnaire was administered in all rounds and at least one of the Round 6 Calls. | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **status** | Whether participant was interviewed, refused to participate, died, migrated abroad, not found or other reason | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **in_phone_r7** | Participants was interviewed via phone in Round 7 (Ethiopia only) | | 7 |
| **dint** | Household date of interview | 1,2,3,4,5 | 1,2,3,4,5,6,7 |

Young Lives also administered community questionnaires from Round 1 to Round 5. The questionnaire provides background information about the social, economic and environmental situation of each community where Young Lives participants live. It covers several topics, including population, ethnicity, religion and language, economic activity and employment and infrastructure. It also provides detailed information of the health, education and child protection services that are available to community members.

# 4. Location variables

Table 2 indicates all location-related variables included in the constructed dataset. Cluster IDs are assigned to every sentinel site that Young Lives visited in Round 1, with the variable *clustid* corresponding to the cluster ID of the location the participant is living in at the time of

the call. If the participant was not living in one of the original Young Lives clusters at the time of the Round, a value of 99 was assigned.

## Household region of residence

The type of information provided by the variable *region* varies by country. In Ethiopia and India, it corresponds to the first official administrative division. In Peru and Vietnam, however, this is not the case. Specifically, in Peru, the *region* variable provides information on the three main geographical and climatic level regions, whereas in the case of Vietnam, it provides information on sub-regions. Therefore, the variable *region* includes the following:

- **Ethiopia:** Tigray, Afar, Amhara, Oromiya, Somalia, Benishangul Gumuz, SNNP, Harari, Addis Ababa City Administration, Dire Dawa, Sidama
- **India:** Andhra Pradesh, Telangana, Other, Not known
- **Peru:** Jungle, Mountain, Coast, *Other (abroad)*
- **Vietnam until Round 5:** Northern Uplands, Red River Delta, Phu Yen, Da Yang, Highlands, South Eastern, Mekong Delta, Other
- **Vietnam Round 6**: Northern Uplands, Red River Delta, Central Coast, Mekong Delta, Other

In India, until Round 6, the region variable provided information on the historical regions Coastal Andhra and Rayalaseema, both of which have been recoded as Andhra Pradesh. In Ethiopia, Sidama, which was previously part of the SNNP region, is now an independent region and has been added as a new category since Round 6. In Peru, the new category *Others* includes participants who were interviewed during the phone survey (Round 6), even though they were living abroad. In Vietnam during Round 6, the rural province Phu Yen and the city of Da Nang were coded as part of the Central Coast region, while the South Eastern and Highlands were grouped under *Others*. We did not harmonise the region categories in Round 5 with those in Round 6 in Vietnam to preserve greater granularity available in earlier rounds.

Additional geographic information can be found in the raw datasets. As part of our ethical commitments, we have guaranteed to protect the anonymity and confidentiality of our study sites and respondents so we only make geographical identifiers available at region level in Ethiopia, district level in India, province level in Peru and province level in Vietnam. We do not release any location data on a lower level.

*Table 2 : Location variables*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **clustid** | Household sentinel site ID | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **Commid (Placeid, PE only)** | Household community ID | | 1,2,3,4,5 |
| **typesite** | Household area of residence (urban/rural) | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **region** | Household region of residence | 1,2,3,4,5 | 1,2,3,4,5,6,7 |

! Peru's community ID is named *placeid* in the constructed files and raw datasets. During Round 6 and Round 7, community ID was not collected in any of the four study countries.

# 5. Participant's characteristics

All the variables related to the participants' demographic characteristics and included in the constructed datasets are reported in Table 3.

## Background characteristics

The time-invariant participant's characteristics sex (*chsex*), ethnicity (*chethnic*) and religion (*chldrel*) are obtained from Round 1 data, while the participant's first language (*chlang*) was taken from Round 2.[8] In Round 7, the participants' sex variable of a small number of participants has been corrected to amend for past mistakes.[9] We have created a new, corrected sex variable *(chsex_correct)* while still retaining the old sex variable in the dataset (*chsex*).

Participants' age in months (*agemon*) at the time of the interview is estimated by taking their age in days (date of interview minus date of birth) and dividing this number by 365/12 (number of days per month). The final number is rounded up. Dates of birth are not publicly archived to preserve anonymity and therefore, it is not possible to estimate the following variable using the archived datasets or to replicate that section of the do-files.

*Table 3: Participant's background variables*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **Chsex** | Participant's sex | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **chsex_correct** | Participant's sex (corrected version chsex) | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **Chlang** | Participant's first language | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **Chethnic** | Participant's ethnic group | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **Chldrel** | Participant's religion | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **Agemon** | Participant's age - in months | 1,2,3,4,5 | 1,2,3,4,5,6,7 |
| **Ageyea** | Participant's age - in years | 1,2,3,4,5 | 1,2,3,4,5,6,7 |

## Family formation and fertility

Since Round 4, we have asked about the marital status and fertility history of the Older Cohort participants, as they were 19 years old. Similarly, since Call 4 (Round 6), we have asked Younger Cohort participants about their marital status and fertility history *(marital_status, marrcohab*).

In addition to *marrcohab*[10], which indicates whether a participant has ever been married or cohabited in their life, we now include detailed information on participants' marital status in each Round. The categories are as follows: 0 - single (never married), 1 - married, 2 - cohabiting, 3 - divorced, 4 -

[8] Participant's first language is not available in Round 1.
[9] In India, five participants previously recorded as male have now been corrected recorded as females (four in the Younger Cohort and one in the Older Cohort).
[10] To construct this variable, we combine information from marital status and other question related to whether participants have ever been married or cohabitated

widowed, 5 - separated. In Peru, however, "single" does not mean "never married", but rather whether the participant is single at the time of the interview.

The age at which a participant first married or cohabited was asked in different survey rounds. Therefore, the variable *marrcohab_age* includes only the information from the first-time participants reported being married or cohabiting.

Similarly, *birth* takes the value of 1 if a participant has ever had a child, while *birth_age*—the participant's age at the birth of their first child—is taken from the first round in which they reported having a child. *Birth_age* is estimated using the date of birth of the firstborn child and the Young Lives participant. Since the date of birth is not publicly archived, this variable cannot be replicated.

*Pregnant* was also asked to the Older Cohort participants since Round 4 and to the Younger Cohort since Round 6. The variable is only available for women in the constructed dataset.

Table 4 summarises information related to family formation and fertility in the constructed dataset.

*Table 4: Family formation and fertility*

| Variables names | Definition | Calls | Rounds |
|---|---|---|---|
| **Marital_status** | Participant's marital status | 4,5 | 4(OC),5(OC),6,7 |
| **marrcohab** | Participant ever been married | | 4(OC),5(OC),6,7 |
| **Marrcohab_age** | Age of participant at first marriage or cohabitation | | 4 (OC),5(OC),7 |
| **Birth** | Participants has ever had a child. | 4,5 | 4 (OC),5(OC),7 |
| **birth_age** | Age of participant when first child was born | | 4 (OC),5(OC),7 |
| **Pregnant** | Whether participant is currently pregnant? | 4,5 | 4 (OC),5(OC),6,7 |

Note: Information in Call 5 was only asked to those participants who were not interviewed in Call 4.

In Round 4 and Round 5, we did not ask information on cohabitation in India because couples are not routinely found living together unless married. Moreover, in India only participants who have ever been married answer the fertility history section. In Ethiopia, Peru and Vietnam we asked the fertility history section to all participants. *Marrcohab_age* has been constructed slightly different across rounds. In Round 4 and Round 5, we asked the month and year of first marriage and estimate the age using the date of birth of participants, whereas in Round 7 we directly asked about the age at first marriage.

More questions on marital and living arrangement, characteristics of partner and their family are available for the Older Cohort since Round 4 and for the Younger Cohort in Round 7. A number of characteristics about Young Lives participants' children, such as method of delivery, vaccinations and education are available in the raw datasets for Rounds 4 and 5 for the Older Cohort. Antenatal care and delivery methods were also asked in call 4 to both cohorts. Information on birth weight is available for both cohorts in Round 7 and Call 4 and in Rounds 4 and 5 only for the Older Cohort.

The Round 6 did not include the date of birth for the children of the Young Lives participant or the age at first marriage or cohabitation.

The section on Fertility history was not administered to participants interviewed over the phone in Ethiopia in Round 7.

## Anthropometrics information

Participant's weight (*chweight*), height (*chheight*) and body mass index (*bmi*) are available in the constructed files for both cohorts in all in-person rounds.[11] Using these three indicators, z-scores for weight-for-height, height-for-age and BMI-for-age were estimated using World Health Organization (WHO) reference tables and software.[12] WHO-defined flags for each indicator were also computed to identify outliers or observations that are considered to be biologically implausible values. The z-scores indicators were estimated using the participant's age in days using the date of interview and date of birth. Since the latter information is not publicly archived, the results provided cannot be reproduced exactly. Age of participant in months, however, provides very close estimators. WHO calculates age in completed months based on the assumption that one year equals 365.25 days, which results in one month being equivalent to 30.4375 days. In contrast, we estimate age in months by dividing 365 by 12.[13]

As z-scores for BMI are defined only up to age 19 (228 months), for adults (older than 228 month) we have estimated the body mass index (bmi) using the standard formula (weight/(height/100)²). [14] Overweight (*overweight_adult)*, underweight (*underweight_adult*) and obesity (*obese_adult*) indicators for adults have also been included.[15] It is important to keep in mind the differences in BMI indicators and how they were estimated, particularly when comparing children with adults.

All anthropometric information included in the constructed dataset, as well as the definitions of the malnutrition indicators and their respective categories, are reported in Table 5.

*Table 5: Anthropometric information*

| Variable name | Definition | Categories | Calls | Rounds |
|---|---|---|---|---|
| **chweight** | Participant weight (kg) | | - | 1,2,3,4,5,7 |
| **chheight** | Participant height (cm) | | - | 1,2,3,4,5,7 |
| **bmi** | calculated bmi=weight / squared(height) | | - | 5(OC), 7 |
| **zwfa** | weight-for-age z-score | | - | 1,2 (YC),3 (YC) |
| **zhfa** | height-for-age z-score | | - | 1,2,3,4,5 (YC) |
| **zbfa** | BMI-for-age z-score | | - | 1(OC),2,3,4,5 (YC) |
| **zwfl** | weight-for-length/height z-score | | - | 1(YC),2 (YC) |
| **fwfa** | flag = 1 if (zwfa < -6 \| zwfa >5) | | - | 1,2(YC),3 (YC) |
| **fhfa** | flag = 1 if (zhfa < -6 \| zhfa >6) | | - | 1,2,3,4,5(YC) |
| **fbfa** | flag = 1 if (zbfa < -5 \| zbfa >5) | | - | 1,2,3,4,5(YC) |

[11] In the previous constructed dataset from Round 1 to 5, height and weight were not included for the Older Cohort in Peru. This information is now available in the current version of the constructed dataset.
[12] WHO reference tables and software are available at https://www.who.int/tools/child-growth-standards/software
[13] More information on how to estimate an "approximate date" can be found at the following link. When estimating the indicators using Stata, age can be provided in days, months, or years. If you are using the WHO AnthroPlus software, age in days is highly recommended for greater accuracy, although age in months can be used to approximate age in days.
[14] We estimate the BMI using the indicators appropriate for the age. In Round 4 for the Older Cohort there are some participants younger than 228 for whom we use BMI z-score, otherwise we use the standard formula. Up from Round 5 for the Older Cohort and Round 7 for the Younger Cohort, we estimated the BMI using the standard formula as all participants are older than 228 months.
[15] As the Older Cohort was 29 years old in Round 7, we did not measure their height. Therefore, to estimate BMI in Round 7, height measured in Round 5 is used.

| | | | | |
|---|---|---|---|---|
| **fwfl** | flag = 1 if (zwfl < -5 \| zwfl >5) | | - | 1(YC), 2(YC) |
| **Stunting_child** | Less than –2SD of height-for-age z-score | 1= Moderately stunted | - | 1,2,3,4,5(YC) |
| | Less than –3SD of height-for-age z-score | 2= Severely stunted | | |
| **Thinnes_child** | Less than –2SD of bmi-for-age z-score | 1= Moderately thin | - | 1,2,3,4,5(YC) |
| | Less than –3SD of bmi-for-age z-score | 2= Severely thin | | |
| **Underweight_child** | Less than –2SD of weight-for-age z-score | 1= Moderately underweight | - | 1,2(YC), 3(YC) |
| | Less than –3SD of weight-for-age z-score | 2= Severely underweight | | |
| **Underweight_adult** | BMI <18.5 | | - | 4-5 (OC), 7 |
| **overweight _adult** | BMI>25 & BMI <30 | | - | 4-5(OC), 7 |
| **Obese_adult** | BMI>30 | | - | 4-5 (OC), 7 |



Since the WHO reference tables apply only to children of certain ages, the weight-for-age z-score, defined up to age 10, is calculated for the Older Cohort only in Round 1 and for the Younger Cohort only up to Round 3. The height-for-age and BMI-for-age z-scores, defined up to age 19, are calculated for the Older Cohort until Round 4 and for the Younger Cohort until Round 5. As height was not measured for the Older Cohort in Round 7, we used height from Round 5 to estimate BMI and related indicators. The weight-for-height indicator is only defined until age 5.

In Round 7, we also measured waist circumference to assess the risk for metabolic diseases and obesity. This variable is available for download in the raw datasets.

This section was not administered in the Round 6 or to participants interviewed by phone in Ethiopia in Round 7 as measurements were only taken for in person rounds.

## Birth and immunisations

Information on the participant's conditions at birth is included in the constructed datasets. We report data from Rounds 1 and 2 on the participant's birth weight (*bwght*), whether this was documented (*bwdoc*), the number of antenatal visits (*numante*), whether the mother received at least two tetanus injections (*tetanus*) and whether the delivery was attended by skilled health personnel (*delivery*).

Additionally, the datasets include information on whether the participant received vaccinations against tuberculosis (*bcg*), measles (*measles*), polio (*polio*), diphtheria, tetanus and pertussis (*dpt*) and Haemophilus influenzae type b (*hib*). Table 6 provides an overview of all variables related to birth conditions and immunisations for the index participants.

*Table 6: Birth and immunisations*

| Variable name | Definition | Rounds |
|---|---|---|
| **bwght** | Participant's birth weight (grams) | 1 |
| **bwdoc** | The Participant's birth weight was from documentation | 1 |

| | | |
|---|---|---|
| **numante** | Number of antenatal visits of mother during pregnancy with Young Lives participant | 1 |
| **tetanus** | Mother received at least two injections for tetanus during pregnancy with Young Lives participant | 1 |
| **delivery** | Mother was attended by skilled health personnel (doctor, nurse, or midwife) during delivery of Young Lives participant | 1 |
| **bcg** | Participant has received BCG vaccination | 1,2 |
| **measles** | Participant has received vaccination against measles | 1,2 |
| **polio** | Participant has received vaccination against polio | 2 |
| **dpt** | Participant has received vaccination against DPT | 2 |
| **hib** | Participant has received vaccination against HIB | 2 |



In Vietnam, immunisations were asked only to the Younger Cohort and not to the Older Cohort.

In Round 2, the participants' caregiver was also asked if he/she believe that their children should be vaccinated and the reasons why. This information is available in the household raw dataset.

## Physical health and disabilities

The participant's history of illness and injury is asked across Rounds 1 to 5. There is information on whether or not the participant has had a serious illness (*chillness*), a serious injury since the previous round (*chinjury*), a long-term health problem (*chhprob*) and whether the participant has a serious illness and the caregiver thought their child might die (*chmightdie*).

Information about the participant's health relative to their peers (*chhrel*) was also obtained from both cohorts and across rounds. However, the availability of information varies by country. Moreover, the questions were asked to different respondents, either the caregivers or index participants, across rounds and countries. These differences are outlined in Table 7.

Young Lives has administered a series of questions regarding participants' disabilities and self-reported health. In Rounds 4, 5 and 7, participants were asked whether they have a permanent disability that affects their work capacity. This information is included in the constructed files with the variable name *chdisability*. The severity of the disability, ranging from 0 to 5, was also recorded and is included in the files with the variable name *chdisscale*. The scale is coded as follows:

0 "Able to work the same as others of this age"

1 "Capable of most types of full-time work but some difficulty with physical work"

2 "Able to work full-time but only work requiring no physical activity"

3 "Can only do light work on a part-time basis"

4 "Cannot work but able to care for themselves (e.g. dress themselves, etc.)"

5 "Cannot work and needs help with daily activities such as dressing, washing, etc."

6 "Other"

Similarly, the participant's general health (*chhealth*), scaled from 1 (very poor) to 5 (very good), is available in Round 3 and subsequent rounds. However, it is worth noticing that while this question was asked directly to the Older Cohort participants from Round 3 to Round 7, it was asked to the Younger Cohort caregiver in Round 3 and Round 4 and directly to the Younger Cohort participants in Round 5 and Round 7.

All information about general health and disability included in the constructed data from Round 1 to Round 7 are summarised in Table 8.

*Table 7: Availability and respondent information of participant's health compared to peers' questions*

| | | | | | | |
|---|---|---|---|---|---|---|
| ● | Participant | | | | | |
| ● | Caregiver | | | | | |
| | | | | | | |
| | | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 |
| Ethiopia | Younger Cohort | ● | ● | ● | | ● |
| | Older Cohort | ● | ● | ● | | |
| India | Younger Cohort | ● | ● | | | ● |
| | Older Cohort | ● | ● | | | |
| Peru | Younger Cohort | ● | ● | ● | ● | ● |
| | Older Cohort | ● | ● | ● | ● | ● |
| Vietnam | Younger Cohort | ● | ● | | | ● |
| | Older Cohort | ● | ● | | | |

*Table 8: General health and disability*

| Variable name | Definition | | Calls | Rounds |
|---|---|---|---|---|
| **Chillness** | The participant has had a serious illness since the last round | | | 2,4,5 |
| **Chinjury** | The participant has had a serious injury since last round | | | 2,3,4,5 |
| **Chhprob** | Participant has a long-term health problem | | | 1,2,5 |
| **Chmightdie** | Participant has had a serious injury/illness since last round when the caregiver though child might die | | | 1,2 |
| **chdisability** | Participant has a permanent disability | | - | 4,5,7 |
| **chdisscale** | Permanent disability scale | | - | 4,5,7 |

| | | | | |
|---|---|---|---|---|
| **chhrel** | Participant's health compared to peers | | 5 (PE) | 1,2,3,4, 5 (YC), 6 (YC-PE) |
| **chhealth** | Participant's health in general | | 5 | 3,4,5,6 (YC-PE), 7 |

! *Chhrel* in Round 1 and Round 2, only included as answers option; same, better, worse whereas from Round 3 onwards the following answer options were included: same, better, much better, worse, much worse. The variable was standardised to align with the categories used in Round 1 and Round 2.

This section was not administered in the Round 6 or to participants interviewed by phone in Round 7 in Ethiopia, except for *chhrel* and *chhealth* in Peru.

## Mental health and subjective well-being

Self-reported symptoms of anxiety and depression were measured for the first time in Round 6 (Calls 2,3 and 5), using the Generalised Anxiety Disorder-7 (GAD-7) scale and the Patient Health Questionnaire-8 (PHQ-8), respectively.[16] The tools measure the frequency of seven anxiety symptoms and eight depression symptoms over the past 14 days. Participants report their symptoms using a four-item Likert scale ranging from:

0 - “not at all”
1 - “less than half days”
2 - “more than half the days”
3 – “nearly every day” '.

The total score ranges from 0 to 21 for GAD-7 and 0 to 24 for PHQ-8. Using the raw score the GAD-7 categorical variable was created and takes value between 1 and 3s: 1 for mild anxiety (a score between 5 and 9), 2 for moderate anxiety (scores between 10 and 14) and 3 for severe anxiety symptoms (scores above 15) (Spitzer *et al.*, 2006). Similarly, the PHQ-8 score categorical variable includes four levels: 1 for mild depressive symptoms (scores between 5 and 9), 2 for moderate depressive symptoms (scores between 10 to 14), 3 for moderately severe depressive symptoms (scores between 15 to 19) and 4 for severe depressive symptoms (scores above 19) (Kroenke *et al.*, 2009).

In Round 7, a shorter version of the depression scale, including only two out of the eight items of PHQ-8 items (PHQ-2), was administered to participants interviewed over the phone in Ethiopia (Löwe, Kroenke and Gräfe, 2005). To enable cross-round comparisons, *phq2_score* has been estimated for all participants across all rounds. For PHQ-2, scores greater than or equal to 3 corresponds to at least mild depression symptoms. The constructed files include the raw scores using the 8-items and 2-items for depression and 7-items for anxiety as well as the categorical variables, from which binary indicators can be derived.

Since Round 2 for the Older Cohort and Round 3 for the Younger Cohort, participants have been asked to assess their personal well-being using Cantril's Ladder (Cantril, 1965). Respondents are asked to imagine a ladder with nine steps, where the bottom step represents the worst possible life and the top step represents the best possible life. Participants are then asked to identify the step they currently feel they stand on.

[16] For each item in GAD-7 and PHQ-8, we asked whether the symptom had been observed (yes/no) and if that was the case, we asked about the frequency.

Trust-related questions were introduced in Call 5 during Round 6 and were repeated in Round 7. Participants were asked whether "most people can be trusted," with the variable trust" recorded as either (1) "Most people can be trusted" or (2) "You can't be too careful/You cannot trust people at all."[17]

*Table 9* provides a complete list of available mental health and wellbeing variables included in the constructed dataset, along with the survey round(s) where each variable is available.

*Table 9: Mental health and subjective well-being*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **cladder** | Participant's subjective well-being (nine-step ladder) | 2,5 | 2(OC),3,4,5,6,7 |
| **gad7_score** | GAD score | 2,3,5 | 6,7 |
| **gad7_cat** | GAD categories:<br>1- mild anxiety<br>2- moderate anxiety<br>3- severe anxiety | 2,3,5 | 6,7 |
| **phq2_score** | PHQ-2 score | 2,3,5 | 6,7 |
| **phq8_score** | PHQ-8 score | 2,3,5 | 6,7 |
| **phq8_cat** | Depression categories:<br>1- Mild symptoms<br>2- Moderate symptoms<br>3- Moderately sever symptoms<br>4- Severe symptoms | 2,3,5 | 6,7 |
| **trust** | Believes most people can be trusted | 5 | 6,7 |



For the anxiety and depression scale, if any of the items were not answered, the whole score and categorical variable is set to missing.

In Round 7, Young Lives introduced a self-reported measure of stress using the Perceived Stress Scale (PSS-10). For the first time, the prevalence of psychotic experiences using the Prodromal Questionnaire PQ-16 was also assessed for participants in Peru and Ethiopia (limited to participants who were interviewed in person). Additionally, data on symptoms consistent with Post-Traumatic Stress Disorder (PTSD) using the ITQ in Ethiopia was gathered, aiming to capture the mental health impact of exposure to conflict. Items from the PQ-16, PSS-10, ITQ, GAD-7 and PHQ-8 are available for download in the raw datasets.

A shorter version of scale of symptoms of stress (PSS-4) and depression (PHQ-2) was administered to participants interviewed by phone in Round 7, whereas the same scale as for the in-person interviews was used for anxiety and PTSD.

**Smoking and drinking habits and reproductive health knowledge**

Young Lives has collected data on smoking and drinking habits since Round 3 for the Older Cohort and in Round 5 for the Younger Cohort, as listed in *Table 10*. Round 7 includes information about the frequency of the participants' drinking and smoking habits in Peru only.

[17] A trust index including different items has been administered to the caregiver in Round 2 and Round 3 and to the index participant in Round 2 and 3 for the Older Cohort and Round 3 for the Younger Cohort.

Questions related to participants' knowledge about sexual and reproductive health were asked as part of a self-administered questionnaire (SAQ). This information is available in Round 3 for the Older Cohort in Peru, in Round 4 for the Older Cohort in Peru, India and Ethiopia, in Round 5 for the Younger Cohort in Peru, India and Ethiopia and in Round 7 for both cohorts in Peru, India and Ethiopia. The variable *chrephealth1* in the constructed files, which ranges from 0 to 5, shows the participant's knowledge of reproductive health based on their correct response (true/false/Don't know) to five statements:[18]

1. A woman/girl cannot get pregnant the first time she has sex.
2. If a girl washes herself after sex, she will not get pregnant.
3. Using a condom can prevent getting a disease through sex.
4. A person who looks very healthy cannot pass on a disease through sex.
5. A person can get HIV or Aids by having sex.

The variables *chrephealth2* and *chrephealth3* are variables equal to 1 if the participants know that using a condom can prevent disease through sex and that a healthy-looking person can pass on disease through sex, respectively.

Participants were also asked where they would go when they wanted to get a condom (*chrephealth4*). While the response options for this question differ across countries and rounds, the options are grouped into four in the constructed files (1- shop or street vendor, 2- family planning services or health facilities, 3 – other, 4- participant does not know what a condom is or where to get a condom).

*Table 10: Smoking and drinking and reproductive health knowledge*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **chsmoke** | Participant's frequency of smoking | - | 3(OC)-PE, 4(OC), 5, 7(PE) |
| **chalcohol** | Participant consumes alcohol every day or at least once a week | - | 3(OC)-PE, 4 (OC), 5,7(PE) |
| **chrephealth1** | Participant's knowledge of reproductive health. | - | 3(OC)-PE, 4 (OC), 5 (YC),7 |
| **chrephealth2** | Participant knows condom can prevent disease through sex | - | 3(OC)-PE, 4 (OC), 5 (YC),7 |
| **chrephealth3** | Participant knows healthy-looking person can pass on a disease sex | - | 3(OC)-PE, 4 (OC), 5 (YC),7 |
| **chrephealth4** | Participant source of condom | - | 3(OC)-PE, 4(OC), 5 |

Since participants were 15 years old (Round 3 – OC) and (Round 5 –YC), Young Lives has collected sensitive information through a self-administered questionnaire (SAQ). In particularly, in Peru, the study gathered information on risky behaviours, sexual and reproductive health, personal experiences with different types of violence. In Round 7, this method continued in Peru. In India and Ethiopia, Young Lives employed Audio-Computer Assisted Self-Interviews (ACASI) for the first time. In this method, participants privately listened to pre-recorded questions via headphones with gender-matched speakers and responded by touching colored shapes on the tablet screen (see Tanima, *et al.,* 2025 for more information). In Vietnam, the SAQ was only administered in Round 3 for the Older Cohort. All self-administered raw datasets are available for download.

In Peru, smoking and drinking habits have been included in the self-administered questionnaire (SAQ) since Round 3 for the Older Cohort and Round 5 for the Younger Cohort. In India, Ethiopia and Vietnam these questions were included in the SAQ in Round 3 and in the individual questionnaire in Round 4 and Round 5. Knowledge about reproductive health was collected in the SAQ in Rounds 3 and 4 for the Older Cohort and in Round 5 for the Younger Cohort. In Vietnam, the SAQ was only administered in Round 3 for the Older Cohort. Due to issues during the SAQ administration in Round 3 in Ethiopia, Vietnam and India, we decided not to include this information in the constructed datasets.

In Round 7, questions on whether participants have ever smoked or consumed alcohol were included in the Audio-Computer Assisted Self-Interviews (ACASI) in Ethiopia and India. However, no data on the frequency of these behaviours was collected in Ethiopia and India but only in Peru. Additionally, in Round 7 and in Peru only, the questionnaire included questions on where the participant would go if they wanted to get a condom. The response options could not be reclassified in a way that is consistent with previous rounds and thus *chrephealth4* could not be constructed.

This section was not administered to participants during Round 6, or to those interviewed by phone in Ethiopia in Round 7.

## Time use

Time-allocation data for Young Lives participants were collected from Round 2 to Round 7 in all countries, with the exception of Round 6, as reported in Table 11. In addition, the Table also provides a detail description of what was asked in each survey Round. The participant was asked to report the amount of time (in hours) spent on eight activities during a typical day. A typical day is defined as a weekday or a normal day, excluding holidays, festivals and weekend rest days, among others. Since Round 4, additional questions on commute time to work and school (in minutes) are asked to distinguish the participants' actual hours spent at work and at school.

*Table 11: Time use variables*

| Variable name | Definition | Specification | Rounds |
|---|---|---|---|
| **hsleep** | Hours/day spent sleeping | Includes napping time (if applicable) | 2,3,4,5,7 |
| **hcare** | Hours/day spent in caring for household members | Amount of time spent taking care of other household members, such as younger siblings, elderly, or members with disabilities. | 2,3,4,5,7 |
| **hchore** | Hours/day spent in household chores | Includes work or tasks done to help at home such as fetching water, firewood, cleaning, cooking, washing, shopping and so on; excludes caring for others. | 2,3,4,5,7 |
| **htask** | Hours/day spent in domestic tasks - farming, family business | amount of time doing work inside the household which generated income; this includes farming, cattle herding, shepherding, piecework or handicrafts done at home and other family businesses. | 2,3,4,5,7 |
| **hwork** | Hours/day spent in paid activity | amount of time doing paid (remunerated) work or activities outside of the household or for someone not in the household including (if applicable) travel time to and from work. | 2,3,4,5,7 |

| | | | |
|---|---|---|---|
| **hschool** | Hours/day spent at school | time spent at school including time used to get from home to school and from school to home. | 2,3,4,5,7 |
| **hstudy** | Hours/day spent studying outside school | Time spent studying at home and doing homework or attending classes or tutorials outside school class hours. | 2,3,4,5,7 |
| **hplay** | Hours/day spent in leisure activities | Time spent playing, eating, self-hygiene, seeing friends, playing sport, using the internet, religious activities | 2,3,4,5,7 |
| **commwork** | Commuting time to place of work (out and return, in minutes) | | 4,5,7 |
| **commsch** | Commuting time to school (out and return, in minutes) | | 4,5,7 |

For the Younger Cohort, time-use information for the index participant was collected from the caregiver in Round 2 and Round 3 and from the index participant themselves starting in Round 4. For the Older Cohort, this information was obtained directly from the participants in all rounds.

Additional information available includes time-use data for all children aged 5 to 17 years living in the household for Ethiopia, India and Vietnam, while for Peru it covers all children aged 4 to 17. This data is available for the Older Cohort in Rounds 2 and 3 and for the Younger Cohort in Rounds 2, 3, 4 and 5.

This section was not administered in the Round 6 or to participants interviewed over the phone in Ethiopia in Round 7. However, in Call 2, a time-use module was included, which indicates whether participants were spending less or more time than before.

## Education and learning outcomes

Education variables available in the constructed datasets include pre-school attendance (*preprim*), age at start of Grade 1 (*agegr1*), formal school enrolment information during the survey year (*enrol*, *engrade* and *entype*) and highest school grade completed (*Hghschoolgrade***)** at the time of interview as reported in Table 12. Categories for *engrade* and *entype* are country-specific.

*Hghschoolgrade* replaces the previously used variable *hghgrade* (highest grade completed), which was included in the constructed datasets from Rounds 1 to 5. The earlier *hghgrade* variable reported the most recent grade completed, which is not necessarily equivalent to the highest, when referring to a level within higher education. However, it did not indicate whether the participant had actually completed a full qualification such as a university degree, or a technical or vocational program—as it only recorded the last grade completed.

To reduce assumptions and improve consistency, we have decided to focus exclusively on school-level grades. The new *hghschoolgrade* variable draws on information from the education section of the household roster in Rounds 1 to 5, which asked about the highest grade completed.

In Round 7, although this specific question was not asked, we were able to use responses from the education history module to determine whether the reported grade had been completed.

The age at start of grade 1 is computed based on the participant's age at the typical month when the academic year starts: April in India and Peru, September in Ethiopia and Vietnam.[19]

[19] In Ethiopia, the official theoretical entrance age is 7 years old whereas in India, Vietnam and Peru it is 6 years old.

*Table 12: Education variables*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **preprim** | Participant attended pre-primary school | | 1-7 |
| **Agegr1** | Age at start of Grade 1 | | 1-7 |
| **timesch** | Travel time to school (in minutes) | | 2-5 |
| **Enrol** | Enrolled in education during round | 2,3,5 | 1,2,3,4,5,6,7 |
| **Engrade** | Grade currently enrol in | 2,3,5 | 1,2,3,4,5,6,7 |
| **Hghschoolgrade** | Highest school grade completed at the time of the interview | 2,3,5 | 2,3,4,5,7 |
| **Entype** | Type of school enrolled during survey year | 2,3,5 | 1,2,3,4,5,6,7 |

The education history section was not administered to participants in Round 6 or to those interviewed by phone. For participants interviewed in person, we use information recorded in education history to create variables on the grade and type of institution currently enrolled in. These questions were included in the current education section for participants interviewed in Calls 2,3 and 5 and in Round 7, for those interviewed over the phone in Ethiopia.

Detailed information on day care and preschool enrolment is available for the Younger Cohort in Round 2. A complete education history for both the Younger Cohort and Older Cohort has been available since Round 3 and includes schooling information dating back to 1995 for the Older Cohort. Additionally, a household education section was included from Round 2 to Round 5, providing detailed information on grade level and type of institution attended for individuals aged 4 to 17. In Calls 2, 3 and 5 during Round 6, data on reasons for leaving education, whether classes were suspended due to COVID-19, whether classes were online or in person and learning activities during the pandemic was gathered. Furthermore, in Round 7, we administered a module focused on education changes caused by the pandemic.

For Round 6 and those interviewed via phone in Round 7 we have categorised the following options as being currently enrol in education: (1) Yes, attending regularly; 2) Yes, but attending irregularly; 3) Registered for this school year but classes are suspended 4) Registered for this school year but classes are not started.

## Learning outcomes

In the constructed files there is information on the participant's reading (*levlread*) and writing skills (*levlwrit*) and whether they can read and write a sentence without difficulty (*literate*), which was asked to the Older Cohort in Round 1 and 2 (see Table 13).

*Table 13: Writing and reading ability codes*

| Variable name | Definition | Rounds |
|---|---|---|
| levlread | 1- Can't read anything<br>2- Reads letters<br>3- Reads words<br>4- Reads sentences | 1,2 |
| levlwrit | 1- No<br>2- Yes, with difficulty or errors<br>3- Yes, without difficulty or errors | 1,2 |
| literate | | 1,2 |

Additionally, Young Lives has administered a series of cognitive tests to both cohorts and their siblings across multiple rounds.[20] This data is not available in the constructed datasets. For an overview of all available information on cognitive and achievement competencies tests administered, see Figure 2 and refer to Revollo and Scott, 2022. For more information on all education and learning related outcomes available in each Round see Mihaylova *et al,.* (2025)

The available data enable the analysis of differences in learning achievement over time and across cohorts. To ensure comparability over time, cognitive scores from different rounds and cohorts have been equated using the Item Response Theory (IRT). For details on this methodology, see Cueto *et al.*, 2009; Cueto and Leon, 2012; Leon and Singh, 2017; Leon, 2020.

Three panel datasets containing the IRT scores for mathematics, reading and PPVT test administered to the participant are available for download separately. Furthermore, there are two additional datasets: one includes PPVT-IRT scores from Rounds 3 to 5 for the siblings of the index participants in Ethiopia, Vietnam and Peru and the other contains maths-IRT scores from Rounds 4 to 5 for the siblings in India.

The Peabody Picture Vocabulary Test (PPVT), a measure of receptive vocabulary consisting of different vocabulary items arranged in increasing order of difficulty, was administered to both cohorts in Rounds 2 and 3. However, in Rounds 4 and 5, the PPVT in Ethiopia, India and Vietnam underwent notable changes and was henceforth referred to as the Receptive Vocabulary test (Leon and Singh, 2017b). In Peru, the Spanish version of the PPVT-R was administered without modifications in both Rounds 4 and 5. For more information on the development and use of the PPVT in Young Lives see Cueto *et al.*, 2009; Cueto and Leon, 2012; Leon and Singh, 2017.

The mathematics test was first administered to the Older Cohort in Round 2 and has since been adapted to adjust participants' age and provide enough variation to capture a wide range of abilities. It was administered to the Younger Cohort from Rounds 3 to 5 and to the Older Cohort from Rounds 2 to 4. For details on the items administered in each round and country refer to Revollo and Scott, 2022.

The reading comprehension test was administered in Round 4 to both cohorts and in Round 5 and 7 only to the Younger Cohort. Items varied across rounds, cohorts and countries. For more information on the items administered in each round and country, see (Revollo and Scott, 2022).

[20] In addition to the above-mentioned cognitive tests, in Round 4, the Young Lives (YL) study introduced a series of computer-based tasks for participants in Ethiopia and Peru. These tasks were part of RACER (Rapid Assessment of Cognitive and Emotional Regulation), a tablet-based application featuring five brief tasks designed to assess four key cognitive skills (Behrman *et al.*, 2022). In Round 7, the first two tasks, which measured executive function (EF), were re-administered to the Younger Cohort (YC) in Ethiopia and Peru.

**1.7. Cognitive and achievement tests**

| | YOUNGER COHORT | | | | | | | OLDER COHORT | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Round 1 2002 | Round 2 2006 | Round 3 2009 | Round 4 2013 | Round 5 2016 | Round 6 2020/2021 | Round 7 2023♣ | Round 1 2002 | Round 2 2006 | Round 3 2009 | Round 4 2013 | Round 5 2016 | Round 6 2020/2021 | Round 7 2023♣ |
| | 1 year | 5 years | 8 years | 12 years | 15 years | 19/20 years | 22 years | 8 years | 12 years | 15 years | 19 years | 22 years | 26/27 years | 29 years |
| **Reading** | | | | | | | | | | | | | | |
| Self-reported english test | | | | *IN only | | | | | | | | | | |
| Reading and writting item | | | • | | | | | • | • | | | | | |
| Reading comprehension | | | | • | • | | • | | | | • | | | |
| Early Grade Reading Assessment | | | • | | | | | | | | | | | |
| Close test | | | | | | | | | | • | | | | |
| **Vocabulary** | | | | | | | | | | | | | | |
| Peabody Picture Vocabulary test (PPVT) | | • | • • | • • | • • | | | | • | • | | | | |
| Receptive vocabulary (modified PPVT) *Not PE | | | | • • | • • | | | | | • | | | | |
| **Numeracy** | | | | | | | | | | | | | | |
| Numeracy item | | | | | | | | • | | | | | | |
| Mathematics | | | • | * IN only • | * IN only • | | | | • | • | • | | | |
| **Early cognition** | | | | | | | | | | | | | | |
| Cognitive Development Assessment (CDA) | | • | | | | | | | | | | | | |
| Ravens Coloured Progressive Matrices (Raven's CPM) test | | | | | | | | • | | | | | | |

*Figure 2: Overview of cognitive test administered over time*

Given that tests differ across rounds and countries, scores for each cognitive test haven been equated with support of a 2-parameter item response theory (IRT) model to generate standard scores which are comparable across time and the four countries (see León 2020).

All reading, maths and PPVT items, as well as the percentage of correct answers, are available to download in the raw datasets. Additional information on the language of administration and the duration of the cognitive test is also available in the raw datasets. In Ethiopia, Peru and Vietnam, from Round 3 to Round 5, the siblings that were closest in age to the Younger Cohort children were tested using the PPVT and receptive vocabulary tests. In India, the siblings who were closest in age to the Younger Cohort children were also tested in maths during Round 4 and Round 5 (same test as YC). All information is available for download in the raw datasets. Moreover, the IRT scores for maths, PPVT and reading are also available for download in the UK data service.

This section was not administered in Round 6 or to participants interviewed by phone in Ethiopia in Round 7.

## Labour market outcomes

Data on work activities and earnings in the year prior to the interviews has been collected for the Older Cohort since Round 4 and for the Younger Cohort since Round 5. Additional questions on job characteristics such as economic sector, job satisfaction, working conditions and work-related training have also been administered. For more detail information about employment related question asked over time see Mihaylova *et al.*, (2025).

In Round 6 and more specifically in Calls 2, 3 and 5, we administered a comprehensive employment section focusing on the main work activity (paid or unpaid) during the last week. Additionally, we also collected information about the main work activity before the outbreak of COVID-19 and in March 2021. In Round 7, we asked both cohorts about their main work activity (paid or unpaid) during the last week and last year and gathered information on the economic sector, type of activity, employer and earnings.

Variables in the constructed dataset focus on the main work activity during the past week and include information on the type of activity (*type_act*), employer (*who_work*), economic sector (*econ_sector*), whether the participant is self-employed (*typemp*), works in the agricultural sector (*agri*) and has a written contract (*written_contract*). Table A 1, Table A 2, Table A 3 indicate the categories for each country for type of activity (type_act), economic sector (econ_sector) and employer (who_work). Table 14 provides a list of all available variables related to the labour market included in the constructed dataset.

*Type of employment*

The variable describing participants' type of employment, *typemp*, distinguishes between dependent workers and self-employed individuals. It is defined by combining information about participants' type of activity (*type_act*; see Table A.1 in the Annex) and who they work for (*who_work*; see Table A.2 in the Annex). Since *type_act* is a country-specific variable, the way *typemp* is constructed differs slightly across countries.

For Ethiopia, India and Vietnam, *typemp* is constructed as follows:

• Participants are classified as dependent workers (*typemp*=0) if they meet the following criteria:

  – They are working for a private company/enterprise, a household member, another private individual/household, the public sector/government or a rural public works programme (who_work=1,2,3,4,5,7) and they are not self-employed (*type_act*=5,6,7,12,13,15,19);
  – They are working for a private company/enterprise, a household member, another private individual/household, the public sector/government or a rural public works programme (*who_work*=1,2,3,4,5,7) and *typemp* would be missing otherwise.

• Participants are classified as self-employed (typemp=1) if they meet the following criteria:

  – They are working for a household member or on their own account (who_work=2,6) and they are self-employed (type_act = 1,2,3,4,8,9,10,11);
  – They are working on their own account (who_work=6) and typemp would be missing otherwise.

For Peru, typemp is constructed as follows:

• Participants are classified as dependent workers if they meet the following criteria:

  – They are working for a private company/enterprise, a household member, another private individual/household, the public sector/government or a rural public works programme (who_work=1,2,3,4,5,7) and they are working as a salaried farmer, labourer, rancher, in forestry, in fishing, salaried worker, casual worker, artisan, trader, household employee or other (type_act=2,3,4,5,6,8,11,12,13,14,16,23);
  – They are working for a private company/enterprise, a household member, another private individual/household, the public sector/government or a rural public works programme (who_work=1,2,3,4,5,7) and typemp would be missing otherwise.

• Participants are classified as self-employed if they meet the following criteria:

  – They are working for a household member or on their own account (who_work=2,6) and they are self-employed or working as a farmer on their own farm (type_act=9,10,21,22);

- They are working on their own account (who_work=6) and working as a labourer, rancher, in forestry, in fishing, unpaid family worker, casual worker, artisan, trader, or other (type_act=3,4,5,6,7,8,12,13,14,15,16);
- They are working on their own account (who_work=6) and typemp would be missing otherwise.

*Table 14: Labour market outcomes*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **Work_lastyear** | Worked in the last year | | 4(OC),5(OC),7 |
| **work_week** | Worked in the past 7 days | 2,3,5 | 4 (OC),5 (OC),6,7 |
| **no_work_job** | Did not work in past week but has job | 2,3,5 | 4 (OC),5 (OC),6,7 |
| **econ_sector** | Economic sector in past week | 2,3,5 | 6,7 |
| **type_act** | Type of activity in past week | 2,3,5 | 6,7 |
| **who_work** | Who worked for in past week | 2,3,5 | 6,7 |
| **Agri** | Worked in agriculture in past week | 2,3,5 | 6,7 |
| **Typemp** | Type of employment (dependent worker vs selfemployed) in past week | 2,3,5 | 6,7 |
| ***Written*_contract** | Whether written contract for the last 7 days activitiy? | 2,5 | 6,7 |

In Round 4 and Round 5, we asked to the Older Cohort about type of employment, employer, economic sector, net earnings and working hours related to the most important work (unpaid or paid) activity in the last year. However, information for the Calls is only available for the last week.

More information about the economic sector, type of activity, employer and time spent on the top 3/5/all work activities during the last year is available for both cohorts in previous rounds. For household members living in the household, information on the type of activity and time spent on the main work activity was collected from Rounds 1 to 4 for the Younger Cohort and Rounds 1 to 3 for the older one. In the Calls raw datasets, detailed information is available on employment status prior to the outbreak of the COVID-19 pandemic and in March 2021. Additional data on the average number of working hours, net earnings and working arrangements for the main activity in the last week and last year—as well as information on reservation wage, willingness to work for pay and unemployment over the past 12 months—is available in the raw Round 7 datasets for both cohorts.

A labour market section about the main activity in the last year was not administered during the Calls and to those participants interviewed by phone in Ethiopia in Round 7. Still, all questions regarding the most important work activity (unpaid or paid) during the last week, have been asked during the phone-survey in Round 6 and the phone survey in Round 7 in Ethiopia.

# 6. Household and parental characteristics

## Parent characteristics

Variables corresponding to the participant's biological mother and father are available in the constructed files. This includes their age (*momage*, *dadage*), a variable indicating whether they are alive and living in the household (momlive, dadlive) and the year of death if they have passed away (*momyrdied*, *dadyrdied*) from Rounds 1 to 7. The highest school grade completed (*momedu*, *dadedu*) was asked until Round 5. In Round 2, parents were asked if they are able to read in any language and if they are able to read in Spanish in the case of Peru (*momcantread, dadcantread*). An overview of all available variables, their definitions and for which rounds there is available information can be found in Table 15.

*Table 15: Mother, father and caregiver characteristics*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **dadid** | Father's id in roster | 1,4,5 | 1,2,3,4,5,6,7 |
| **dadage** | Father's age | 1,4,5 | 1,2,3,4,5,6,7 |
| **dadlive** | Location of father | 1,4,5 | 1,2,3,4,5,6,7 |
| **dadyrdied** | Year the father died | 1,4,5 | 1,2,3,4,5,6,7 |
| **dadcantread** | Father cannot read… any language (ET, IN, VN); Spanish (PE) | | 2 |
| **dadedu** | Father's level of education | | 1,2,3,4,5, |
| **momid** | Mother's ID in roster | 1,4,5 | 1,2,3,4,5,6,7 |
| **momage** | Mother's age | 1,4,5 | 1,2,3,4,5,6,7 |
| **momlive** | Location of mother | 1,4,5 | 1,2,3,4,5,6,7 |
| **momyrdied** | Year the mother died | 1,4,5 | 1,2,3,4,5,6,7 |
| **momcantread** | Mother cannot read… any language (ET, IN, VN); Spanish (PE) | | 2 |
| **momedu** | Mother's level of education | | 1,2,3,4,5 |

Note: Information available in call 5 only includes those participants who were not interviewed during call 4.

While the constructed files compiled information about the participant's biological parents, basic characteristics of other members of the household are available in the raw dataset in all rounds, alongside detailed information on the education of the siblings and other family members aged 5 to 17.

## Caregivers' characteristics

Basic information on the primary caregivers is also available in the constructed files from Rounds 1 to 5, including id (*careid*), age (*careage*), sex (*caresex*), level of education (*caredu*), relationship to Young Lives participants (*carerel*), relationship to the head of the household (*carehead*), whether they cannot read (*carecantread*), as well as their current subjective wellbeing (*careladder*) and their

wellbeing four years prior to the interview, i.e. since the previous survey round (*careldr4yrs*) (See Table 16).

*Table 16: Caregiver's characteristics*

| Variable name | Definition | Rounds |
|---|---|---|
| **careid** | Caregiver's id in roster | 1,2,3,4(YC),5(YC) |
| **careage** | Caregiver's age | 1,2,3,4(YC),5(YC) |
| **caresex** | Caregiver's sex | 1,2,3,4(YC),5(YC) |
| **carehead** | Caregiver's relationship to head of household | 1,2,3,4(YC),5(YC) |
| **carerel** | Caregiver's relationship to Young Lives participant | 1,2,3,4(YC),5(YC) |
| **carecantread** | Caregiver cannot read | 1,2 |
| **caredu** | Caregiver's level of education - harmonised var | 1,2,3,4(YC),5(YC) |
| **careladder** | Caregiver's ladder - subjective well-being | 2,3,4(YC),5(YC) |
| **careldr4yrs** | Caregiver's ladder (four years from now) - subjective well-being | 2,3,4(YC),5(YC) |

## Household head characteristics

The characteristics of the household's head have been gathered from Rounds 1 to 7, excluding Round 6. Information in the constructed files includes the head of household's age (*headage*), sex (*headsex*) and relationship to Young Lives participant (*headrel*) as well as the education level (*headedudu*). Similar to parental characteristics, Round 7 surveys did not include questions related to the education level of the head of household (*headed*). Table 17 summarises all available information for the head of household included in the constructed dataset from Round 1 to Round 7.

*Table 17: Household head characteristics*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **headid** | Household head ID in roster | | 1,2,3,4,5,7 |
| **headage** | Age of household head | | 1,2,3,4,5,7 |
| **headsex** | Sex of household head | | 1,2,3,4,5,7 |
| **headedu** | Household head education - harmonised variable | | 1,2,3,4,5 |
| **headrel** | Household head's relationship to Young Lives participant | | 1,2,3,4,5,7 |

It is important to outline that household heads are not constant over the survey rounds. Reasons for this include but are not limited to: (1) change in composition of the household as members move in and out of the household; (2) movement of a Young Lives participant from one household to another; and (3) a Young Lives participant settling down, starting a new family and becoming the household head. In Round 7, due to the participants' age, it was common for them to complete the household questionnaire themselves, as well as becoming the household head.

## Household size and composition

The number of household members by sex and age groups (computed excluding the Young Lives participant) is alongside the size of the household (*hhsize*) are available in the constructed files as reported in Table 18. Note that the household size refers to the number of people currently living in the household, which may or may not include the Young Lives participant, depending on the participant's location at the time of the survey.[21]

*Table 18: Household composition*

| Variable | Definition | Calls | Rounds |
|---|---|---|---|
| **male05** | Number of males aged 0-5 in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **male612** | Number of males aged 6-12 in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **male1317** | Number of males aged 13-17 in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **male1860** | Number of males aged 18-60 in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **male61** | Number of males aged 61+ in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **female05** | Number of females aged 0-5 in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **female612** | Number of females aged 6-12 in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **female1317** | Number of females aged 13-17 in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **female1860** | Number of females aged 18-60 in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **female61** | Number of females aged 61+ in the household | 1,4,5 | 1,2,3,4,5,6,7 |
| **hhm05** | Number of hhm aged 0-5 | 1,4,5 | 1,2,3,4,5,6,7 |
| **hhm612** | Number of hhm aged 6-12 | 1,4,5 | 1,2,3,4,5,6,7 |
| **hhm1317** | Number of hhm aged 13-17 | 1,4,5 | 1,2,3,4,5,6,7 |
| **hhm1860** | Number of hhm aged 18-60 | 1,4,5 | 1,2,3,4,5,6,7 |
| **hhm61** | Number of hhm aged 61+ | 1,4,5 | 1,2,3,4,5,6,7 |
| **hhsize** | Household size | 1,4,5 | 1,2,3,4,5,6,7 |
| **Note**: Information available in call 5 only includes those participants who were not interviewed during call 4. | | | |

[21] Adding all household member groups might not always match the household size (excluding the Young Lives participant). This is because household members with missing age or sex information are not included in the specific group totals.

Other characteristics of household members are available for download in the raw dataset. These include the household members' education attainment and disability, among others. In Rounds 3 to 5 information about Younger Cohort siblings was obtained. Among the data collected are the sibling's anthropometrics, cognitive test scores, education and time use. For details about surveying these siblings see Porter et al. (2012).

Household roster was administered to participants interviewed via phone in Round 7 in Ethiopia as well as during call 1 and call 4.

## Livestock, land and house ownership variables

A section on household livelihood and assets was asked in the household questionnaire between Round 1 and Round 5. Ownership of country-specific livestock is included in the constructed dataset (see Table 19). Animals' variables are aggregated in four large groups: milk animals, draught animals, small ruminants and animals that are specific to each country. For Rounds 2 to 5, the number of livestock for specific animals is also reported.

Variable *ownlandhse* indicates whether a household member owns the land where the household's dwelling is built, while *ownhouse* indicates whether a household member owns the house. *Ownhouse* is the only question from this module that was asked in Round 7.

*Table 19: Livestock, land and house ownership variables*

| Variable | Definition | Rounds |
|---|---|---|
| **ownlandhse** | Household owns land where house is on | 1,2,3,4,5 |
| **ownhouse** | Household own the house | 2,3,4,5,7 |
| **aniany** | Household owned any livestock in the past 12 months | 1,2,3,4,5 |
| **animilk** | Number of MILK animals in the household | 1,2,3,4,5 |
| **anidrau** | Number of DRAUGHT animals owned by the household | 1,2,3,4,5 |
| **anirumi** | Number of SMALL RUMIANTS animals owned by the household | 1,2,3,4,5 |
| **anispec** | Number of OTHER animals specific to country | 2,3,4,5 |
| **anicowm** | Number of (modern) cows | 2,3,4,5 |
| **anicowt** | Number of (traditional) cows | 2,3,4,5 |
| **anicalv** | Number of calves | 2,3,4,5 |
| **anibufm** | Number of (modern) buffalos | 2,3,4,5 |
| **anibuft** | Number of (traditional) buffalos | 2,3,4,5 |
| **aniheif** | Number of heifers | 2,3,4,5 |
| **anibull** | Number of bullocks | 2,3,4,5 |
| **anihebu** | Number of he-buffalo | 2,3,4,5 |
| **anidonk** | Number of donkeys, horses, mules | 2,3,4,5 |
| **aniybul** | Number of young bulls | 2,3,4,5 |
| **anioxen** | Number of oxen | 2,3,4,5 |
| **anicaml** | Number of camels | 2,3,4,5 |
| **anishee** | Number of sheep | 2,3,4,5 |

| | | |
|---|---|---|
| **anigoat** | Number of goats | 2,3,4,5 |
| **anipigs** | Number of pigs | 2,3,4,5 |
| **anipoul** | Number of poultry/birds | 2,3,4,5 |
| **anirabb** | Number of rabbits | 2,3,4,5 |
| **anibeeh** | Number of beehives | 2,3,4,5 |
| **anillam** | Number of llamas | 2,3,4,5 (YC) |
| **aniguin** | Number of guinea pigs | 2,3,4,5 (YC) |
| **anisnai** | Number of snails | 2,3,4,5 (YC) |
| **anifish** | Number of fish ponds | 2,3,4,5 (YC) |
| **anishri** | Number of marine shrimp tanks | 2,3,4,5 (YC) |
| **anifshr** | Number of fresh water shrimp tanks | 2,3,4,5 (YC) |
| **aniothr** | Number of other livestock | 2,3,4,5 (YC) |

## Credit and food security

Beginning in Round 2 in Peru and in Round 3 in Ethiopia, India and Vietnam, households were asked to describe their food situation at home in the last 12 months (*foodsec*). Their response could be any of the following options:

1 – We always eat enough of what we want

2 – We eat enough but not always what we would like

3 – We sometimes to not eat enough

4 – We frequently do not eat enough.

In addition, the constructed dataset from Round 1 to Round 7 incorporates for the first time the Household Food Insecurity Access Scale (HFIAS) score. These items have been asked in Round 3 to both cohorts, in Rounds 4 and 5 only to the Younger Cohort and in Call 5) and Round 7 to both cohorts again. Since fewer items were included in Rounds 6 and 7, adjustments in previous Rounds were made to allow for cross-round comparison.

The food security module includes nine questions, used to create the HFIAS, which is an adaptation of the approach used to estimate the prevalence of food insecurity in the United States. However, it differs in the recall period as it asks participants about their food security situation in the past 12 months instead of the previous 4 weeks. In Call 5 and again in Round 7 only 6 out of 9 questions (Q1, Q3, Q5, Q6, Q7, Q8 of Table 20) were asked to both cohorts. Table 20 displays the list of HFIAS questions administered in each survey Round.

*Table 20: List of questions to estimate HFIAS score and categories*

| **Items** | **R3** | **R4** | **R5** | **Call 5-R6** | **R7** |
|---|---|---|---|---|---|
| Which of the following statements best describes the food situation at your home in the last twelve months? | | | | | |
| 1. In the past 12 months, did you ever worry that your household would run out of food before you get money to buy or could acquire more? | Yes | Yes | Yes | Yes | Yes |
| 2. Were you or any household member not able to eat the kinds of foods you want because of lack of money? (For example, no meat, no fish, no fruit, no sweet) | Yes | Yes | Yes | No | No |
| 3. Did you or any household member have to eat a limited variety of foods due to lack of money? (For example, only rice and one vegetable, no meat) | Yes | Yes | Yes | Yes | Yes |

| | | | | | |
|---|---|---|---|---|---|
| 4. Did you or any household member have to eat some foods that you did not want to eat because of a lack of money to obtain other types of food? (For example, wild foods, immature crops, broken rice, discarded food) | Yes | Yes | Yes | No | No |
| 5. Did you or any household member have to eat less (portion size) in a meal than you wanted because there was not enough food? | Yes | Yes | Yes | Yes | Yes |
| 6. Did you or any household member have reduced the number of meals eaten a day because there was not enough food? (For example, skip breakfast or lunch) | Yes | Yes | Yes | Yes | Yes |
| 7. Was there ever no food to eat in your household because of lack of money to get food? | Yes | Yes | Yes | Yes | Yes |
| 8. Did you or any household member go to sleep at night hungry because there was not enough food? | Yes | Yes | Yes | Yes | Yes |
| 9. Did you or any household member go a whole day and night without eating anything because there was not enough food? | Yes | Yes | Yes | No | No |

Using these 9 questions administered in Round 3, Round 4 and Round 5, we can estimate the food insecurity score following the approach used for the HFIAS score. If participants answered "yes" to any question, a follow-up question on frequency was asked. The frequency is recorded as follows:

0. “Never”
1. “Rarely”
2. “Sometimes, some months but not always”
3. “Always or nearly always for all months”

By adding up, the scale ranges from 0 to 27. Based on the raw score, the following food insecurity categories can be derived. The estimations are given in Table 21.

*HFIAS – 6 items*

In Round 6 and Round 7, items 2, 4 and 9 (see Table 20) were not administered. Moreover, if a participant answers yes to any of these questions, a follow-up variable reporting on the frequency had these categories:

0 “Never”
2. “Rarely or sometimes, some months but not always”
3. “Always or nearly always for all months”

Thereby, the scale ranges from 0 to 18. To make cross-rounds comparison, variables reporting on the frequency in Round 3, Round 4 and Round 5 were re-coded to only have three instead of four categories. More precisely, option one was re-coded to two. By doing so, we can categorise food insecurity based on the HFIAS score instruction as described in Table 21. Table 21 provides an overview of all variables related to food insecurity and credit included in the constructed dataset, in which survey Round they were asked, as well as detailed information on how the HFIAS categories were estimated using items reported in Table 20.

*Table 21: Food insecurity*

| Variable | Definition | Categories | Estimation | Calls | Rounds |
|---|---|---|---|---|---|
| **credit** | Household has obtained loan or credit… In the past 12 months (IN, PE, VN) Since the previous round (ET) | | | | 3, 4(YC), 5(YC) |
| **foodsec** | Household's food situation in the last 12 months | | | 5 | 3, 4(YC), 5(YC), 6, 7 |
| **Hfias_score** | Raw HFIAS score (Recall period: past 12 months) | | | | 3, 4(YC), 5(YC) |
| **Hfias_cat** | HFIAS categories (Recall period: past 12 months) | Food secure | [(Q1=0 or Q1=1) and Q2=0 and Q3=0 and Q4=0 and Q5=0 and Q6=0 and Q7=0 and Q8=0 and Q9=0] | | 3, 4(YC), 5(YC) |
| | | Mildly Food Insecure Access | [(Q1>=2 or Q2>=1 or Q3=1 or Q4=1) and Q5=0 and Q6=0 and Q7=0 and Q8=0 and Q9=0] | | |
| | | Moderately Food Insecure Access | [(Q3>=2 or Q4>=2 or Q5=1 or Q5=2 or Q6=1 or Q6=2) and Q7=0 and Q8=0 and Q9=0] | | |
| | | Severely Food Insecure Access | [Q5=3 or Q6=3 or Q7>=1 or Q8>=1 or Q9>=1] | | |
| **Hfias_score_6** | Raw HFIAS score (6-items) (Recall period: past 12 months) | | | 5 | 3, 4(YC), 5(YC), 6, 7 |
| **Hfias_cat_6** | HFIAS category (6 items) (Recall | Food secure | (Q1=0 and Q3 =0 and Q5=0 and | 5 | 3, 4(YC), 5(YC), 6, 7 |

| | | | | | |
|---|---|---|---|---|---|
| | period: past 12 months) | | Q6=0 and Q7=0 and Q8=0) | | |
| | | Mildly Food Insecure Access | (Q1>=2) and (Q5=0 and Q6=0 and Q7=0 and Q8=0) | | |
| | | Moderately Food Insecure Access | (Q3>=2 or Q5=2 or Q6==2) and (Q7=0 and Q8=0) | | |
| | | Severely Food Insecure Access | Q5==3 or Q6==3 or Q7!=0 or Q8!=0 | | |

The follow-up questions on frequency were coded 1/2 in Round 7 instead of 2/3 in Call 5, respectively. We have recoded the items in Round 7 to match those from Call 5 to allow for comparisons of raw scores across rounds. This does not affect the food security categorical variable values. The raw score is set to missing if at least one variable is missing. In Round 4 in Vietnam, if a participant responded 'We always eat enough' regarding the household's food situation in the past 12 months, all items used to construct the HFIAS score were skipped. In the constructed dataset, these households are categorized as food secure.

During Round 7 phone survey, a short scale of food insecurity scale as well as three questions on food situation during the first 12 months of the conflict, was administered.

The household's food and non-food consumption and expenditure are also available in the Young Lives data. From this data, consumption aggregates were computed. The dataset is also available to download. Details about how the dataset was calculated are available in Marion (2018). Moreover, in calls, we asked about food insecurity during COVID-19. In Round 7 in Ethiopia, we asked three questions about the food situation during the first 12 months of conflict.

## Public programmes

Households were also asked about their participation in several country-specific public programmes. The list of all public programmes included in the constructed dataset from Round 1 to Round 7 is available in Table 22. For more detailed information about participation in social protection programmes over time see Mihaylova, *et al., (*2025)

*Table 22: Public Program*

| Variable | Definition | Rounds |
|---|---|---|
| ***Ethiopia only:*** | | |
| **hep** | At least one member is a beneficiary of the Health Extension Programme | 4,5,7 |
| **psnp_pw** | At least one member is a beneficiary of the PSNP-Public Works Programme in the past 12 months | 3,4,5,7 |

| | | |
|---|---|---|
| **psnp_ds** | At least one member is a beneficiary of the PSNP-Direct Support Programme in the past 12 months | 3,4,5,7 |
| **eap** | At least one member is a beneficiary of the Emergency Aid Programme since the previous round | 4 (OC),5,7 |
| **othprog** | Household has received support from other security programmes in the past 12 months | 3,4,5 |
| **resettled** | At least one member has been resettled by the government since the previous round | 4,5 |
| **India only :** | | |
| **pds** | Household is accessing the Public Distribution System | 3,4,5,7 |
| **nregs** | Household has a job card under the NREGS | 3,4,5,7 |
| **nregs_work** | At least one member has worked for the NREGS in the past 12 months | 3,4,5,7 |
| **nregs_allow** | Household has received unemployment allowance under NREGS since previous round | 4,5,7 |
| **rajiv** | Respondent has Rajiv/NTR Arogyasri card | 3,4,5,7 |
| **sabla** | At least one member has accessed the SABLA programme since previous round | 5 (YC) |
| **sabla_yl** | Young Lives participant has benefitted from the nutritional component under REGSEAG/SABLA | 4 (YC), 5(YC) |
| **ikp** | At least one member has benefitted from IKP-Credit provision since previous round | 2,3,4,5,7 |
| **ikp_child** | At least one participant of the respondent has benefitted from IKP | 2,3 |
| ***Peru only:*** | | |
| **juntos** | At least one member is a current beneficiary of Juntos | 3,4,5,7 |
| **bonograt** | At least one member received transfers from Bono de Gratitud/Pension 65 programme | 4,5 |
| **sisgrat_yl** | Young Lives participant is registered in SIS gratuito | 3,4,5,7 |
| **minsa_yl** | Young Lives participant is registered in partial SIS/ SIS independiente (MINSA) | 4,5 |
| **insur_yl** | Young Lives participant has health insurance | 4,5,7 |
| **beca_yl** | Young Lives participant is a beneficiary of the Beca 18 programme | 4(OC),7 |
| **projoven_yl** | Young Lives participant has received training under the ProJoven/ Jovenes a la Obra programme | 4(OC), 5(OC),7 |
| **Vietnam only:** | | |
| **molisa06** | Household is included in the list of poor households in 2006 using the MOLISA criteria | 2 |
| **molisa09** | Household is included in the list of poor households in 2009 using the MOLISA criteria | 3 |
| **molisa10** | Household is included in the list of poor households in 2010 using the MOLISA criteria | 4 |
| **molisa11** | Household is included in the list of poor households in 2011 using the MOLISA criteria | 4 |
| **molisa12** | Household is included in the list of poor households in 2012 using the MOLISA criteria | 4 |

| | | |
|---|---|---|
| **molisa13** | Household is included in the list of poor households in 2013 using the MOLISA criteria | 5 |
| **molisa14** | Household is included in the list of poor households in 2014 using the MOLISA criteria | 5 |
| **molisa15** | Household is included in the list of poor households in 2015 using the MOLISA criteria | 5 |
| **molisa16** | Household is included in the list of poor households in 2016 using the MOLISA criteria | 5 |



In Round 7, this question was asked in the individual questionnaire instead of the household questionnaire as in previous rounds.

More details about households' participation in public programmes, such as duration of support and identification of members who participated, are available for download in the raw datasets.

During Round 7 phone survey, we asked about the Health Extension Program (HEP) and the Emergency Aid Program in a similar way as in-person interviews. However, when asking about the Productive Safety Net Program (PSNP), we did not distinguish between the Public Work Program and the Direct Support Program, unlike in the in-person interviews. Consequently, the constructed dataset includes the variables *psnp_pw* or *psnp_ds* for participants interviewed by phone.

## Household shocks

A shock module was administered since Round 1 to gather information on self-reported exposure to country-specific shocks at the household level. Note that answers are based on perceptions; that is, they do not show whether a negative event has occurred or not, but rather they show whether the respondent considers the event to have affected the welfare of the household negatively.

A long list of shocks is available in the constructed files from Rounds 1 to 7, as reported in Table 23. All shock-related variables are binary (the variable equals 1 when a shock was reported during the period between rounds, 0 otherwise). For more detailed information about household shocks over time see (Mihaylova, *et al.,* 2025).

*Table 23: Household Shocks*

| Variable | Definition | Rounds |
|---|---|---|
| **Shock A: . Has the household been the victim of any crime since previous round?** | | |
| **shcrime1** | shock-destruction/theft of tools for production | 2(All),3 ,4 ,5 (only in PE) |
| **shcrime2** | shock-theft of cash | 2(All),3,4,5 (only in PE) |
| **shcrime3** | shock-theft of crops | 1(all),2(all),3,4,5 (only in PE) |
| **shcrime4** | shock-theft of livestock | 1(all),2(all),3,4,5 (only in PE) |
| **shcrime5** | shock-theft/destruction of housing/consumer goods | 2(All),3,4,5 (only in PE) |
| **shcrime6** | shock-crime that resulted in death/disablement | 2(All),3,4,5 (only in PE) |
| **shcrime7** | shock-theft/destruction (cash, crops, livestock, housing) | 3,4,5,7 (not PE) |
| **shcrime8** | shock-victim of crime | 1 |
| **Section B Have any regulations or actions had negative impact on the household since previous round?** | | |

| | | |
|---|---|---|
| **shregul1** | shock-land redistribution | 2(All),3 ,4 (only in PE) |
| **shregul2** | shock-resettlement or forced migration | 2(All),3 ,4 (only in PE) |
| **shregul3** | shock-restrictions on migration | 2 (not in PE) |
| **shregul4** | shock-forced contributions | 2,3,4,5,7 |
| **shregul5** | shock-eviction | 2 |
| **Shregul6** | shock-invasion of property | 3,4 (PE) |
| **Section C: Have any changes to economic conditions affected the household since previous round?** | | |
| **shecon1** | shock-increase in input prices | 2,3,4,5,7 |
| **shecon2** | shock-decrease in output prices | 2,3,4,5,7 |
| **shecon3** | shock-death of livestock | 1,2,3,4,5,7 |
| **shecon4** | shock-closure place of employment | 2(All),3,4 (Only PE) |
| **shecon5** | shock-loss of job/ source of income/ family enterprise | 1,2,3,4,5,7 |
| **shecon6** | shock-industrial action | 2(All),3,4 (Only PE) |
| **shecon7** | shock-contract disputes (purchase of inputs) | 2(All),3,4 (Only PE) |
| **shecon8** | shock-contract disputes (sale of output) | 2(All),3,4 (Only PE) |
| **shecon9** | shock-disbanding credit | 2(All),3,4,5 (Only PE) |
| **shecon10** | shock-confiscation of assets | 2(All),3,4 (Only PE) |
| **shecon11** | shock-disputes with family about assets | 2(All),3,4 (Only PE) |
| **shecon12** | shock-disputes with neighbours about assets | 2,3,4,5,7 |
| **shecon13** | shock-increase in food prices | 3,4,5,7 (not PE) |
| **shecon14** | shock-decrease in food availability | 1 |
| **Section D: Have you experienced any natural disasters since previous round?** | | |
| **shenv1** | shock-drought | 2,3,4,5,7 |
| **shenv2** | shock-flooding | 2,3,4,5,7 |
| **shenv3** | shock-erosion | 2,3,4,5 |
| **shenv4** | shock-frost | 2,3,4,5,7 |
| **shenv5** | shock-pests on crops | 2,3,4,5,7 |
| **shenv6** | shock-crop failure | 1,2,3,4,5,7 |
| **shenv7** | shock-pests on storage | 2,3,4,5 |
| **shenv8** | shock-pests on livestock | 2,3,4,5 |
| **shenv9** | shock-natural disaster | 1 |

| | | |
|---|---|---|
| **shenv10** | shock-earthquake | 2,3,4,5 (PE) |
| **shenv11** | shock-forest fire | 2,3,4 (PE) |
| **shenv12** | shock-pollution caused by mining | 2,3,4 (PE) |
| **shenv13** | shock-storm | 3,4,5 (VN) |
| **Section E: Has anything happened since previous round that has affected the building you live in?** | | |
| shhouse1 | shock-fire affecting house | 2(all),3,4,5 (only PE) |
| shhouse2 | shock-house collapse | 2(all),3,4,5 (only PE) |
| shhouse3 | shock-fire or collapse of building | 2,3,4,5 |
| **Section F: Have there been any changes within the family since previous round?** | | |
| **shfam1** | shock-death of father | 2,3,4,5,7 |
| **shfam2** | shock-death of mother | 2,3,4,5,7 |
| **shfam3** | shock-death of another household member | 2,3,4,5,7 |
| **shfam4** | shock-illness of father | 2,3,4,5,7 |
| **shfam5** | shock-illness of mother | 2,3,4,5,7 |
| **shfam6** | shock-illness of another household member | 2,3,4,5,7 |
| **shfam7** | shock-divorce or separation | 2,3,4,5,7 |
| **shfam8** | shock-birth of new household member | 1,2,3,4,5 |
| **shfam9** | shock-enrolment of participant in school | 1,2,3,4,5 |
| **shfam10** | shock-imprisonment | 2 (all),3,4 (only PE) |
| **shfam11** | shock-conscription, abduction or draft | 2 |
| **shfam12** | shock-death/reduction household members | 1 |
| **shfam13** | shock-severe illness or injury | 1 |
| **shfam14** | shock-move/migration | 1 |
| **shfam15** | shock-political imprisonment | 2,3,4 (only PE) |
| **shfam16** | shock-political discrimination | 2,3,4 (only PE) |
| **shfam17** | shock-ethnic/social discrimination | 2,3,4 (only PE) |
| **shfam18** | shock-illness of non-household member | 4,5 |
| **shother** | shock-others | 1,2,3,4,5,6,7 |

! The list of shocks varies across rounds and countries. See Annexes for details.

A more detailed report of shocks by year is available in the raw data. This also includes the top three shocks that affected the household and the degree of economic loss attributable to the shocks. In Round 7 in Ethiopia, follow up questions on whether the shocks happened due to the conflict, as well as additional conflict-related shock specific questions were included.

During Round 7, a reduced number of shocks was administered. Young Lives asked about shcrime7, shecon1, shecon5 and shfam7. An additional list of conflict-related shocks was also administered by phone in Round 7.

## Wealth index

The Young Lives wealth index is the primary measure of the socio-economic status of households. It is constructed from three indices: housing quality, access to services and ownership of consumer durables, as reported in Table 24. A detailed description of how these indices were computed can be found in Briones, 2017.

During the Calls, both, flush toilet/septic tank (private) and flush toilet/septic tank (shared) have been considered safe sanitation services for participants in Ethiopia. In previous Rounds and Round 7, we did not distinguish between private and shared flush toilet.

In Peru in Calls 2 and 5, toilet with water/septic tank inside the house/lot and latrine/household manhole are considered safely managed sanitation services. Similarly, pipeline that reaches household/lot/land and well cases with manual pump inside the house are also considered as safe drinking water sources and equivalent to “piped water to the house/plot” and “well/tube with hand pump”.

For Ethiopia, we have re-estimated the wealth index, as concrete blocks were previously not considered a good-quality wall material. Although this category has been included since Round 3, fewer than 1% of participants reported using it. By Round 7, however, around 9% of participants reported having concrete block walls. The variables wi_new and hq_new were estimated by considering concrete blocks as good quality, unlike wi and hq, which do not. Across Rounds 3 and 7, 262 observations differ between wi_new and wi (1 in Round 3, 37 in Round 4, 31 in Round 5 and 193 in Round 7) .

*Table 24: Wealth index*

| Variable name | Definition | Calls | Rounds |
|---|---|---|---|
| **wi** | Wealth index | | 1,2,3,4,5,7 |
| **hq** | Housing quality index | | 1,2,3,4,5,7 |
| **sv** | Access to services index | | 1,2,3,4,5,7 |
| **cd** | Consumer durables index | | 1,2,3,4,5,7 |
| **elecq** | Access to electricity | 2,5 | 1,2,3,4,5,6,7 |
| **toiletq** | Access to sanitation | 2,5 | 1,2,3,4,5,6,7 |
| **drwaterq** | Access to safe drinking water | 2,5 | 1,2,3,4,5,6,7 |
| **cookingq** | Access to adequate fuels for cooking | | 1,2,3,4,5,7 |

For the Calls, we are unable to construct the wealth index, as no data on wall, roof and floor materials, household items needed to estimate consumer durable index, or information on cooking with adequate fuels was collected.

! Details of changes in the computation of some sub-indices, to make sure that the definitions are similar across rounds and are consistent with nationally defined standards across, can be found in Briones (2017). In Ethiopia in Round 6, flush toilet/septic tank (private) and flush toilet/septic tank (shared) are considered safe sanitation services. In Peru in Round 6, Toilet with water/septic tank inside the house and lo/latrine/household manhole are considered safely managed sanitation service. Pipeline that reaches the house/lot/land (public network) and well cased with manual pump inside the house are coded as safe drinking water sources.

# 7. References

Barnett, I., Ariana, P., Petrou, S., Penny, M., Duc, L.T., Galab, S., Woldehanna, T., Escobal, J., Plugge, E. and Boyden, J. (2013) 'Cohort profile: the Young Lives study.', *International journal of epidemiology*, 42(3). Available at: https://ora.ox.ac.uk/objects/uuid:3cee5285-30c5-49bd-ab44-4c4665b52b67 (Accessed: 14 July 2025).

Behrman, J., Briones, K., Cueto, S., Favara, M., Freund, R., Hittmeyer, A., Lopez, J., Pazos, N., Sánchez, A., Scott, D., Sheridan, M. and Woldehanna, T. (2022) *Measuring foundational Cognitive Skills in Young Lives using RACER*. Technical Note 54. Young Lives. Available at: https://www.younglives.org.uk/sites/default/files/2022-09/YL-TN54.pdf (Accessed: 8 February 2024).

Briones, K. (2017) '"How many rooms are there in your house?" Constructing the Young Lives wealth index', *Young Lives Technical Note* [Preprint]. Available at: https://ora.ox.ac.uk/objects/uuid:2858b876-8fb5-4bb6-a764-7d927f1e8ebc (Accessed: 11 October 2024).

Briones, K. (2018) *A Guide to Young Lives Rounds 1 to 5 Constructed Files*. Technical Note 48. Available at: https://www.younglives.org.uk/publications/guide-young-lives-rounds-1-5-constructed-files (Accessed: 20 February 2025).

Cantril, H. (1965) *The Pattern of Human Concerns*. Rutgers University Press.

Cueto, S. and Leon, J. (2012) *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives*. 25. Young Lives.

Cueto, S., Leon, J., Guerrero, G. and Muñoz, I. (2009) 'Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives'. Available at: https://ora.ox.ac.uk/objects/uuid:58ef6b73-aa60-40a6-929f-1eb868ecc328 (Accessed: 25 October 2024).

Escobal, J. and Flores, E. (2008) 'An Assessment of the Young Lives Sampling Approach in Peru', *Young Lives Technical Note No. 3* [Preprint]. Available at: https://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-TN3-Escobal-Sampling-Approach-In-Peru.pdf.

Favara, M., Crivello, G., Penny, M., Porter, C., Ellanka, R., Sanchez, A., Scott, D., Le, T., Tassew, W. and McKay, A. (2021) 'Cohort profile update: the Young Lives Study', *International Journal of Epidemiology*, 50(6). Available at: https://ora.ox.ac.uk/objects/uuid:e27a30db-1676-4d57-aec6-a0817bd8fbc0 (Accessed: 5 February 2025).

Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B.W., Berry, J.T. and Mokdad, A.H. (2009) 'The PHQ-8 as a measure of current depression in the general population', *Journal of Affective Disorders*, 114(1), pp. 163–173. Available at: https://doi.org/10.1016/j.jad.2008.06.026.

Kumra, N. (2008) 'An assessment of the Young Lives sampling approach in Andhra Pradesh, India'. Available at: https://ora.ox.ac.uk/objects/uuid:7c5212c5-dfd4-4a62-8bcb-071d7c123112 (Accessed: 15 November 2024).

Leon, J. (2020) 'Equating Cognitive Scores across Rounds and Cohorts for Young Lives in Ethiopia, India, Peru and Vietnam'. Available at: https://www.younglives.org.uk/sites/default/files/migrated/YL-TN51_0.pdf (Accessed: 25 September 2024).

Leon, J. and Singh, A. (2017a) 'Equating Test Scores for Receptive Vocabulary Across Rounds and Cohorts in Ethiopia, India and Vietnam'. Available at:

https://www.younglives.org.uk/sites/default/files/migrated/YL-TN40%20%283%29%20%28003%29.pdf (Accessed: 25 September 2024).

Leon, J. and Singh, A. (2017b) 'Equating Test Scores for Receptive Vocabulary Across Rounds and Cohorts in Ethiopia, India and Vietnam'. Available at: https://www.younglives.org.uk/sites/default/files/migrated/YL-TN40%20%283%29%20%28003%29.pdf (Accessed: 25 September 2024).

Löwe, B., Kroenke, K. and Gräfe, K. (2005) 'Detecting and monitoring depression with a two-item questionnaire (PHQ-2)', *Journal of Psychosomatic Research*, 58(2), pp. 163–171. Available at: https://doi.org/10.1016/j.jpsychores.2004.09.006.

Mihaylova, J., Mohsini, K., Quigua Chinchilla, J., Tanima, T. and Woodman Deza, A. (2025) 'Young Lives Round 1-7 Data Inventory', *Young Lives Technical Note 62*. Available at: https://www.younglives.org.uk/sites/default/files/2025-10/Data%20Inventory_FINAL.pdf (accessed 14October 2025).

Nguyen, N. (2008) 'An assessment of the Young Lives sampling approach in Vietnam'. Available at: https://ora.ox.ac.uk/objects/uuid:cb04de6a-91c4-4924-8982-e992f3d3214e (Accessed: 5 July 2024).

Outes-Leon, I. and Sanchez, A. (2008) 'An assessment of the Young Lives sampling approach in Ethiopia'. Available at: https://ora.ox.ac.uk/objects/uuid:1cc6b2df-3bf4-4aa9-8abd-f65d9cc1665d (Accessed: 15 November 2024).

Revollo, P.E. and Scott, D. (2022) 'Cognitive and achievement tests in the Young Lives Study'. Available at: https://ora.ox.ac.uk/objects/uuid:b58bf564-c4bb-460d-b168-9a894c74da51 (Accessed: 11 October 2024).

Spitzer, R.L., Kroenke, K., Williams, J.B.W. and Löwe, B. (2006) 'A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7', *Archives of Internal Medicine*, 166(10), pp. 1092–1097. Available at: https://doi.org/10.1001/archinte.166.10.1092

Tanima, T., Favara, M., von Russdorf, S., de los Angeles Molina, M. and Hidalgo Arestegui, A. (2025) 'Asking Difficult Questions: Design and Implementation of an Audio Computer-Assisted Self-Interview (ACASI) in the Young Lives Study', *Young Lives Technical Note No. 61*. Available at: https://www.younglives.org.uk/sites/default/files/2025-10/YL-TN61-ACASI_TN_FINAL.pdf

# 8. Annex

*Table A 1: Overview of type of activity (type_act)*

| ET, IN & VN | PE |
|---|---|
| **AGRICULTURE & ALLIED AGRICULTURE** | **AGRICULTURE & ALLIED AGRICULTURE** |
| 01=Self Employed (Food crops) | |
| 02=Self Employed (Non-food, including horticulture, sericulture and floriculture) | 02 = Wage Employment (Agriculture) |
| 03 = Self Employed (Aquaculture) | 03 = Part time agriculture labourer/Eventual labour |
| 04 = Self Employed (Livestock, bee keeping, chickens, herding) | 04 = Cattle farmer |
| 05=Wage Employment (Agriculture) | 05 = Forestry |
| 06=Annual Farm Labourer | 06 = Fishing |
| 07=Other (allied) agriculture, specify | 07 = Non-remunerated Household members (agriculture) |
| | 08 = Other (related to agriculture) (Specify) |
| | |
| **NON-AGRICULTURE** | **NON-AGRICULTURE** |
| 08=Self Employed (Manufacturing) | |
| 09 = Self Employed (Services) | 09=Self Employed (Small manufacturing business) |
| 10= Self Employed (Business) | 10 Self Employed (Services) |
| 11 = Self Employed (Other non-Agriculture) | 11 = Waged Worker |
| 12=Wage Employment (Unsalaried/ irregular; Non-agriculture) | 12 = Part time laborer |
| 13=Regular Salaried Employment | 13 = Craftsman/Artisan (independent) |
| | 14 = Independent merchant |
| 15 = Begging (IN) | 15 = Non-remunerated Household members (non-agriculture) |
| 19=Other non-agriculture, specify | 16 = Other (not related to agriculture) (Specify) |
| | **UNEMPLOYED AND OTHER NON-WAGED** |
| | 18 = Household chores/ Housewife |
| | 19= Other (non wage earner) (Specify) |
| 20=Student (IN) | 21 = Self Employed farmer (own plot)- Food crops |
| 21=Housewife (IN) | |
| 22= House maid (ET) | |
| | 22 = Self Employed farmer (own plot) - Horticulture, sericulture and floriculture |
| | 23 = Housekeeper |

*Table A 2 Overview of type of employer*

| Ethiopia, India and Peru |
|---|
| 01=Private company/enterprise or cooperative |
| 02=For a household member |
| 03=Other private individual/household (excluding own household) |
| 04=Public sector/government |
| 05=A rural public works program |
| 06=Own account/self-employed (own business or farm) |
| 07= Other, specify |

*Table A 3: Overview economic sector*

| **No.** | **Economic Sector** |
|---|---|
| 1 | Agriculture, forestry and fishing |
| 2 | Mining and quarrying |
| 3 | Manufacturing |
| 4 | Electricity, gas, steam and air conditioning supply |
| 5 | Water supply; sewerage, waste management and remediation activities |
| 6 | Construction |
| 7 | Wholesale and retail trade; repair of motor vehicles and motorcycles |
| 8 | Transportation and storage |
| 9 | Accommodation and food service activities |
| 10 | Information and communication |
| 11 | Financial and insurance activities |
| 12 | Real estate activities |
| 13 | Professional, scientific and technical activities |
| 14 | Administrative and support service activities |
| 15 | Public administration and defence; compulsory social security |
| 16 | Education |
| 17 | Human health and social work activities |
| 18 | Arts, entertainment and recreation |
| 19 | Other service activities |
| 20 | Activities of households as employers; undifferentiated goods- and services-prod |
| 21 | Activities of extraterritorial organisations and bodies |
| 77 | Not known |
| 79 | Refused to answer |
| 88 | Not applicable |

Young Lives

A Longitudinal Study of Poverty & Inequality

**About Young Lives**

Young Lives is an international study of poverty and inequality, following the lives of 12,000 children in four countries (Ethiopia, India, Peru and Vietnam). Young Lives is a collaborative research programme led by the University of Oxford in association with research and policy partners in the four study countries.

Through researching different aspects of children's lives across time, we seek to improve policies and programmes for children and young people.

---

**Young Lives Research and Policy Partners**

*Ethiopia*
- *Policy Studies Institute*
- *Pankhurst Development Research and Consulting plc*

*India (Andhra Pradesh and Telangana)*
- *Centre for Economic and Social Studies, Hyderabad (CESS)*
- *Sri Padmavati Mahila Visvavidyalam (Women's University), Tirupati (Young Lives India)*

*Peru*
- *Grupo de Análisis para el Desarollo (GRADE)*
- *Instituto de Investigación Nutricional (IIN)*

---

**Contact:**
**Young Lives**
Oxford Department of
International Development,
University of Oxford,
3 Mansfield Road,
Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Email: younglives@qeh.ox.ac.uk
Website: www.younglives.org.uk

Young Lives