ANNEX C Definition of variables and methodology

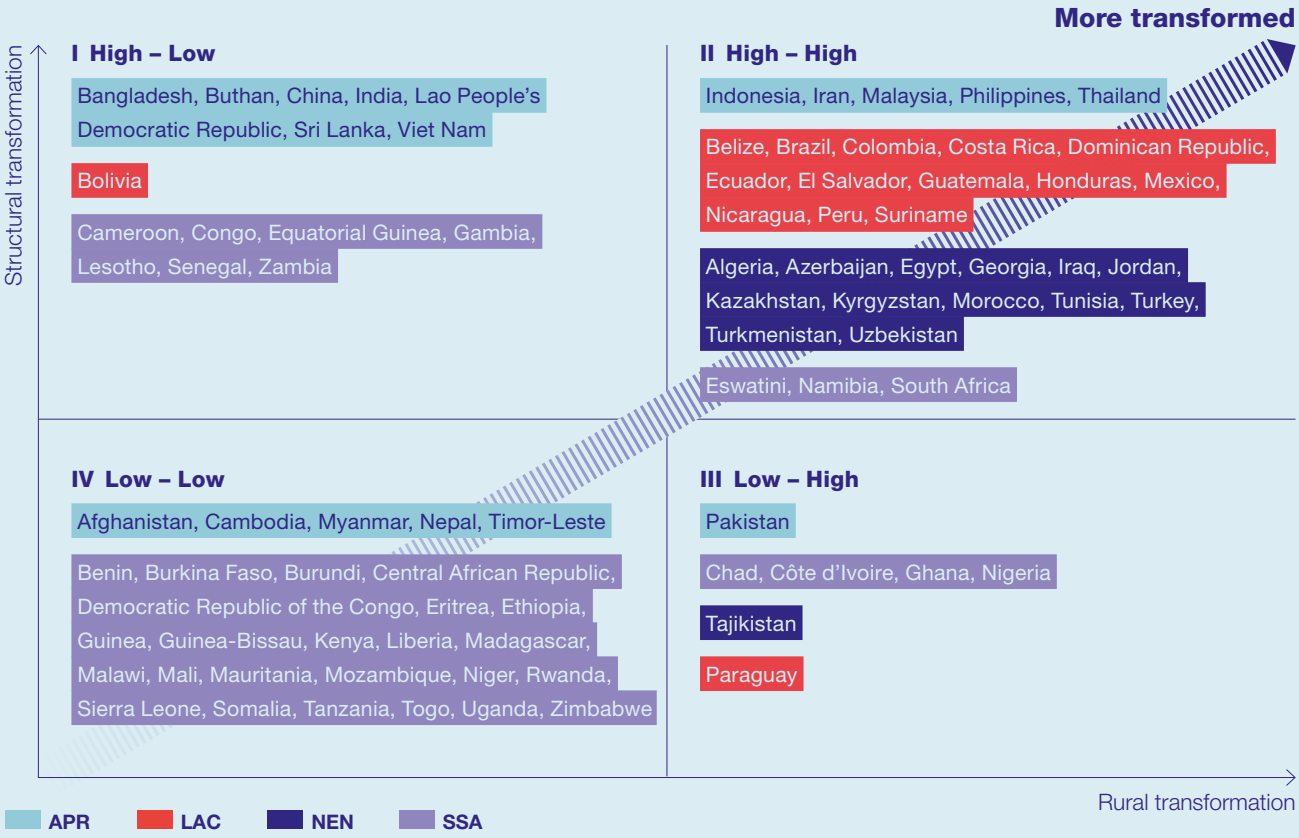
The *Rural Development Report 2019* makes use of macro- and micro-level data to analyse the level of transformation of the countries where young people live (*country transformation typology*), determine the nature of their opportunity matrix (*rural opportunity space*) and examine the characteristics of the households to which they belong (*household transformation categories*). These three typologies, taken together, provide information about the opportunities open to rural youth and the challenges that they face. This information can then be used as inputs for a systematic approach to the design of policies and programmes for fostering youth-centred rural transformation.

Macro-level analysis

At the macro level, data from 85 countries in Asia and the Pacific (APR), sub-Saharan Africa (SSA), the Near East, North Africa, Central Asia and Europe (NEN), and Latin America and the Caribbean (LAC) were used to analyse the rural and structural transformation processes and economic and institutional indicators of the countries where rural youth live. The sample includes all low- and middle-income countries[58] except small island nations, resource-dependent nations and countries for which information was not available.

TABLE 1 Country transformation typology

Country transformation typology



Notes: APR: Asia and the Pacific; LAC: Latin America and the Caribbean; NEN: Near East, North Africa, Europe and Central Asia; SSA: sub-Saharan Africa. Countries are classified as having attained a relatively high degree of rural transformation if their value added per worker exceeds the sample median (US$1,592) and as having attained a relatively high degree of structural transformation if the share of non-agricultural value added exceeds the sample mean (80%). The sample consists of 85 low- and middle-income countries as defined by the World Bank (2018).
Source: Authors.

58 Low- and middle-income countries are defined on the basis of the World Bank classification for 2018.

Country transformation typology

This report uses two variables[59] to define levels of structural transformation (ST) and rural transformation (RT):

- ST: Non-agricultural value added (percentage of GDP)
- RT: Agricultural value added per worker (constant 2010 US$)

The country typologies were defined on the basis of a combination of the level (high/low) of these measures relative to the global median (for RT) and global average[60] (for ST) using the latest value available (data for 2016 in 90 per cent of the cases).[61]

The variables listed to the right are also used in discussing the country-level challenges and opportunities for rural youth in all 85 low- and middle-income countries included in the analysis:

Variable	Source	Year
Rural and urban poverty headcount ratio	*Rural Development Report 2016*, World Bank	Measured at 2011 $ in PPP
Income per capita in constant 2010 PPP US$	World Development Indicators (WDI), World Bank	
Government Effectiveness percentile rank	Worldwide Governance Indicators (WGI), World Bank	
Countries in a conflict situation	Brueck et al., 2018	
Countries with fragile situations	Harmonized List of Fragile Situations, World Bank	
Youth population share	*World Population Prospects: The 2017 Revision*, United Nations	1950-2050
Rural youth population share	*World Urbanization Prospects: The 2017 Revision*, United Nations	1985-2015

Micro-level analysis

At the micro level, this report uses socio-economic household surveys from national statistical offices, combined with geographic variables explained in detail in annex B. With the exception of the Indonesia Family Life Survey,[62] all the surveys are nationally representative and cover both urban and rural areas. Household and individual-level data are collated for 13 countries in the three main regions of SSA, LAC and APR (see **TABLE 2**).

59 Source: World Development Indicators, World Bank.

60 Resource-rich countries were excluded in order to avoid artificially increasing the value of the global mean and median. A country is defined as being resource-rich if the share of rents from natural resources other than forest resources amounts to more than 12 per cent of total GDP. Resource-rich countries were then classified as being at a high or low level of ST based on the non-resource, non-agricultural share of their non-resource GDP.

61 The countries for which the latest available data were for a different year are: Tajikistan (2015), Belize (2015 for ST) and Tunisia (2015 for RT).

62 The Indonesia Family Life Survey is based on a sample of households representing about 83 per cent of the Indonesian population.

TABLE 2 Microdata sources and sample sizes

Country	Survey	Source	Year	No. of households (individuals)	Geo-locations
Sub-Saharan Africa					
Ethiopia	Ethiopia – Socioeconomic Survey	Central Statistical Agency of Ethiopia (CSA)	2015/2016	4,954 (23,393)	Enumeration areas (EAs) – geocoded
Malawi	Fourth Integrated Household Survey	National Statistical Office (NSO) – Ministry of Economic Planning and Development	2016/2017	12,447 (53,885)	EAs – geocoded
Niger	National Survey on Household Living Conditions and Agriculture – Panel data	Survey and Census Division – National Institute of Statistics	2014	3,617 (22,671)	EAs – geocoded
Nigeria	General Household Survey – Panel data	National Bureau of Statistics (NBS) – Federal Government of Nigeria	2015/2016	4,291 (24,807)	EAs – geocoded
United Republic of Tanzania	National Panel Survey	National Bureau of Statistics – Ministry of Finance and Planning	2014/2015	3,352 (16,285)	EAs – geocoded
Uganda	The Uganda National Panel Survey	Uganda Bureau of Statistics – Government of Uganda	2013/2014	1,561 (9,373)	EAs – geocoded
Latin America and the Caribbean					
Mexico	Encuesta nacional de ingresos y gastos de los hogares	Instituto Nacional de Estadística y Geografía, MEX-INEGI.40.202.03-ENIGH-2016-NS	2016	69,939 (256,448)	EAs – geocoded
Nicaragua	Encuesta nacional de hogares sobre medición de nivel de vida	Instituto Nacional de Información de Desarrollo	2014	6,851 (29,381)	Municipality geocodes identified
Peru	Encuesta nacional del hogares 2016 (Annual) – Condiciones de vida y pobreza	Instituto Nacional de Estadística e Informática	2016	35,785 (134,235)	EAs – geocoded
Asia					
Bangladesh	Household Income and Expenditure Survey	Bangladesh Bureau of Statistics – Ministry of Planning	2010	12,240 (55,580)	Upazila geocodes identified
Cambodia	Cambodia Socio-Economic Survey	National Institute of Statistics – Ministry of Planning	2014	12,090 (53,968)	Village geocodes identified
Indonesia	Indonesia Family Life Survey	J. Strauss, F. Witoelar and B. Sikoki. "The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report". March 2016. WR-1143/1-NIA/NICHD.	2014	15,881 (58,312)	EA – geocoded[63]
Nepal	Nepal Living Standards Survey	Central Bureau of Statistics – National Planning Commission Secretariat, Government of Nepal	2010	5,988 (28,670)	Village geocodes identified

63 Geo-locations for the enumeration areas of IFLS-5 were obtained from the RAND Cooperation, which the authors wish to thank for its cooperation. All SSA datasets are from the World Bank's Living Standards Measurement Study – Integrated Studies on Agriculture (LSMS-ISA). The ongoing support provided by the LSMS team during the compilation work is gratefully acknowledged.

Rural opportunity space

A typology of rural opportunity spaces was created for the *Rural Development Report 2019* using two main variables: commercialization potential and agricultural production potential (Wiggins and Proctor, 2001; Ripoll et al., 2017). Data for the commercialization potential indicators were drawn from the WorldPop project (population densities). Data for agricultural potential were obtained by using the Enhanced Vegetation Index (EVI) to classify land as cropland or pastureland.

A systematic method was used to merge the population density data with the geographic locations of households obtained from survey data. In SSA, artificial boundaries were drawn around the georeferenced centroids for each enumeration area (EA) in order to capture the average EA population from the survey data based on the known densities taken from WorldPop data.

In APR, with the exception of Indonesia, and in LAC, the household data do not include georeferenced information, but GIS layers from DIVA were used to obtain centroids for municipalities/other small units with boundaries for relatively small administrative areas. The analyses presented in this report include any dataset with boundary data for an administrative unit whose average size is 1,000 square km or less.[64]

The rural-urban (i.e. rural, semi-rural, peri-urban and urban) gradients correspond to the population density quartiles for all low- and middle-income countries. The least dense quartile corresponds to rural areas and the densest quartile to urban areas. In between there are the semi-rural (second quartile) and peri-urban (third quartile) areas. Each EA has been classified along the rural-urban gradient and matched up with its level of commercialization/connectivity potential.

The EVI, which is a measure of the density of green vegetation, was used as a proxy for agricultural production potential. Global land use layers were used to isolate cropland and grazing land in order to exclude very densely forested areas and water bodies from the indicator. In order to avoid annual variability in the EVI caused by rainfall and temperature fluctuations and the impacts of extreme weather events, a three-year average EVI value was computed. The vegetation density was then divided into terciles in order to obtain agricultural potential gradients. The combination of the EVI terciles and the three lowest population density quartiles (rural, semi-rural and peri-urban) generates the different categories of the rural opportunity space typology. For technical details on the geospatial data that were used to develop this typology, see annex B.

Household transformation categories

The third typology used in the *Rural Development Report 2019* employs data at the household level. The household transformation categories capture the capacity of households to commercialize their agricultural production activities (akin to *rural transformation* at the household level) and their ability to diversify their sources of income towards more profitable non-farm activities (akin to *structural transformation* at the household level). The first indicator is calculated as the share of farm sales over total farm income. The second is the share of non-farm income over total income. Household income aggregates are constructed based on the Rural Income Generating Activities (RIGA) methodology, as follows:[65]

- Total income includes crop production, wage income, self-employment income from non-agricultural activities and other types of income (transfers, non-farm rent income, real estate, etc.).
- Farm income includes income from harvests, forestry, livestock and livestock products (milk, eggs, etc.).
- Agricultural income was computed by applying estimated prices for crop sales to the net (excluding crop losses) harvested quantity.
- Non-farm income includes non-agricultural wages, non-agricultural enterprise income and other kinds of income.

64 This size makes it possible to fully contain the administrative unit within a circle around the unit's centroid having a 50-km radius.

65 See http://www.fao.org/economic/riga/riga-database/en/.

Total income was decomposed into its various sources in order to measure each activity's contribution to the total value. The components of income that were considered were:

+ Own household agricultural production
+ Household enterprises
+ Wage earnings
+ Non-employment income (transfers, rental income, etc.)

Household welfare was measured using two proxies: per capita household expenditure and poverty headcount ratio at the international[66] poverty line ($1.90 a day 2011 PPP). All monetary values are expressed in per capita terms per day in PPP (constant 2011 international $). An imputation technique called winsorizing was applied to treat outliers; this involved replacing all the values above the 99th percentile of the distribution for each income component with the highest value within the 99th percentile. For the aggregate income variables, all the extreme values (above the 99th percentile and below the 1st percentile) were replaced with missing values to address outliers.

Other descriptive variables

Apart from the variables used to compute the three sets of typologies, additional data were processed in order to compile information on the demographic characteristics of the youth population, young people's level of education and their employment status. Specifically, labour force participation and the amount of time spent working in each type of employment activity were calculated for six sectoral and functional categories. The six sectoral and functional categories for which full-time equivalents (FTEs) were measured were: own farm work, on-farm agrifood system (AFS) wage labour, off-farm AFS wage labour, non-AFS wage labour, AFS enterprise work and non-AFS enterprise work.

The amount of time devoted to work in each of these categories was calculated using the concept of full-time equivalents (FTEs), which makes it possible to compare workloads across different contexts and sectors. The computation of FTEs shows how much of a household member's total labour availability (considered to be 40 hours per week) is allocated to each employment activity. A full-time work schedule is assumed to be equivalent to 12 months per year, 4.3 weeks per month and 40 hours per week. With the exception of two countries, FTEs are computed at the annual level by dividing the total number of hours worked during the year by the total labour availability (2,016 hours). In Mexico and Peru, due to data constraints, FTEs are computed on a weekly basis.[67] This indicator can range from 0 to 2; an FTE equal to 1 corresponds to full-time work, while an FTE of less than 1 signals underemployment and an FTE greater than 1 represents overemployment. This approach delivers higher estimates of workforce participation than standard labour market measures do and does not measure unemployment, since that cannot be defined for a 12-month reference period.

66 With the exception of Indonesia, where the poverty rate is based on the national poverty line due to issues with the consumption data.

67 In Mexico, the corresponding question refers to the number of hours worked per week during the last month; in Peru, it is based on the last week prior to the interview.