GDL Area Database

Sub-national development indicators for

research and policy-making

(version 4.2.1)

Jeroen Smits

Global Data Lab (GDL)

Institute for Man**a**gement Research

Radboud University

P.O. Box 9108, 6500 HK Nijmegen, The Netherlands

www.globaldatalab.org

Abstract

The GDL Area Database contains development indicators at the national and sub-national level for low and middle income countries (LMICs). These indicators were created by the Global Data Lab team and are made available to the global community to increase our knowledge and understanding of the situation and developments in these countries. The indicators are created by aggregating data from household survey datasets. This paper provides information on the indicators themselves, the method that was used to create them, and the survey datasets from which they were aggregated. With each new version of the database, this paper will be updated, hence the version number should be included when referring to it.

Contact information
Global Data Lab, Institute for management Research, Radboud University
P.O. Box. 9108, 6500 HK Nijmegen, The Netherlands
Phone: +31 (0)24-3612319/5890
E-mail: info@globaldatalab.org

GDL Working Paper Series: www.globaldatalab.org/publications

1. Introduction

For many years, the World Development Indicators program of the World Bank has made hundreds of indicators available for almost all countries of the world in a broad range of fields (http://data.worldbank.org/products/wdi). There are also many other databases that provide indicators for specific fields, like education, work, human development, health, corruption, etc. These indicators provide a unique source of information on the situation and development of mankind in relation to its environment. They have been used extensively for scientific and applied research and for policy development at regional and global level.

Unfortunately, most of these databases provide only indicators at the national level. This is a great disadvantage, as countries are no homogenous entities, but show substantial internal variation in almost all aspects relevant for researchers and policy makers. There are huge differences between urban and rural areas and between different regions within countries and many of the problems and developments that require policy attention are concentrated within specific areas or groups. National indicators are very useful for providing insights into global processes and developments, but are of little use for addressing such within-country issues.

For wealthier countries, like the US and the EU countries, in the last decades subnational indicators have become available. However, for most low and middle income countries (LMICs) -- where administrative systems are less equipped and capable to generate reliable and representative information -- such indicators are still lacking. This is regrettable, as the need for subnational development indicators is becoming increasingly urgent. The use of disaggregated indicators is for instance considered to be one of the key elements of the Sustainable Development Goals monitoring program (UN, 2016).

To improve the statistical infrastructure in LMICs will take time and does not produce the indicators for earlier years required for tracing developments over time. Therefore it is important to look for other ways of generating subnational indicators. An interesting alternative is to produce such indicators on the basis of available household survey data.

In the last decades, over 600 large representative household surveys have been held in low and middle income countries. These surveys contain a wealth of socio-economic, health and demographic information at the individual and household level that can be used to create indicators for sub-national areas (provinces, states, governates, etc.) within countries.

The Global Data Lab (GDL) has brought together and harmonized many of these datasets into a Big Data infrastructure called the Database Developing World (DDW), which currently contains information on over 30 million persons in 115+ LMICs. This information is used by the GDL staff for scientific research and for developing knowledge instruments, specialized databases and indicators for measuring and analyzing the status and progress of societies.

One of the databases created by the Global Data Lab is the GDL Area Database, which contains a broad set of indicators for over 1000 sub-national regions in more than 100 LMICs. These

indicators are made freely available through the Global Data Lab website to the global community.

The indicators in the GDL Area Database are created by aggregation from the household surveys datasets in the DDW. Aggregation means taking the average of the values of a characteristic of individuals or households in each area.

The procedure is straightforward; the educational level of an area is created by taking the mean years of education of individuals in a certain age group, the area's vaccination coverage by the percentage of children who received a specific vaccination, etc.

In this way many indicators are created in a broad range of fields (e.g. education, fertility, wealth and poverty, asset ownership, population shares, demographic window, gender, child mortality, vaccinations, housing characteristics, public services, children's stunting and wasting). The database is regularly updated and extended with additional indicators, countries and years.

Aggregation from survey data is a flexible technique that can be used to create sub-national indicators for any topic that is covered in household surveys. And the available surveys include data on many more topics than the ones mentioned here.

With the infrastructure developed at GDL, we also can aggregate to other levels, like wealth quartiles or educational categories, as well as create combinations of indicators for specific issues, like stunting levels by birth order or contraceptive use by age group.

In the following sections of this paper, detailed information is provided on the indicators themselves, on the methods that are used to generate them and on the datasets from which they are generated. In Appendix A the Terms of Use and Disclaimer information is presented.

2. Data

The indicators available in the GDL Area Database were created by aggregating data from household survey datasets. These household survey datasets are all part of the Database Developing World (DDW, www.globaldatalab.org/ddw), a huge Big Data infrastructure for LMICs held by the Global Data Lab in which over 400 of household survey datasets are connected and harmonized.

The DDW currently contains information on over 35 million persons in 130+ countries from all regions of the developing world. It combines survey datasets from major data sources like Demographic and Health Surveys(DHS), UNICEF Multiple Indicator Household Surveys (MICS), World Health Surveys (WHS), Living Standard Measurement Surveys (LSMS), Pan Arab Family Project Surveys (PAPFAM), and Barometer Surveys and is continuously updated.

The datasets included in the DDW were obtained from the producing organizations for research by GDL-connected researchers. The aggregated indicators in the GDL Area Database are outcomes of this research. Making them available through our website is one of our ways of publishing research results. Below an overview is given of the survey datasets that were used for the current version of the database.

Demographic and Health Surveys (DHS)

DHS are large representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition. Standard DHS Surveys have large sample sizes (usually between 5,000 and 30,000 households). Each survey consists of a household interview, in which basic information is collected on all household members, and separate women's and men's surveys. In the women's survey, all usual resident women aged 15 to 49 are invited for an oral interview. The men's surveys are generally smaller and show more variation across countries. In the interviews, information is obtained on socioeconomic, demographic, and (reproductive) health related issues. Since the 1980s over 300 DHS surveys have been held in all regions of the developing world. For most countries, several surveys have been held in different years, allowing comparisons over time. Producer of the DHS datasets is ICF International and major sponsor of the program is the U.S. Agency for International Development (USAID). Further information is available at http://www.dhsprogram.com

In the GDL Area database currently data from the following DHS surveys are used: Afghanistan (2015); Albania (2009,2017), Angola (2011,2016), Armenia (2000,2010,2016), Azerbaijan (2006), Burundi (2010, 2017), Benin (1996,2001,2006,2011), Burkina Faso (1998,2003,2010), Bangladesh (1996,2004,2007,2011,2014), Bolivia (1998,2003,2008), Brazil (1996), Central African Republic (1994), Cote d'Ivoire (1994, 1999,2005,2011), Cameroon (1998,2004,2011), Chad (1997,2004), Colombia (1995,2000,2005,2010,2015) , Comoros (1996,2012), Congo Democratic Republic (2007,2013), Congo Brazzaville (2005,2011), Dominican Republic

(1996,2002,2007,2013), Egypt (1995,2000,2005,2008,2014), Ethiopia (2000,2005,2011,2016), Eritrea (1995, 2002), Gabon (2000,2012), Ghana (1998,2003,2008,2014,2017), Guinea (2005,2012), Gambia (2013), Guatemala (1995,1999,2015), Guyana (2009), Honduras (2005,2011), Haiti (1994, 2000,2005,2012,2017), Indonesia (1997,2003,2007,2012), India (1992,1999,2006,2016), Jordan (2002, 2007,2012), Kazakhstan (1995,1999), Kenya (1993,1998,2003,2008,2014), Kyrgyzstan (1997,2012), Cambodia (2000,2005,2010,2014), Liberia (2007,2013), Lesotho (2004,2010), Malawi (1992,2000,2004,2010,2016), Morocco (1992,2003(jointly with PAPFAM)), Moldova (2005), Madagascar (1992,1997,2004,2009), Maldives (2009), Mali (1995,2001,2006,2013), Mozambique (1997,2003,2011), Myanmar (2016), (Namibia (1992,2000,2006,2013), Nepal (1996,2001,2006,2011,2016), Niger (1998,2006,2012), Nigeria (1999,2003,2008,2013), Nicaragua (1998,2001), Pakistan (1991,2007,2012), Peru (1996,2000,2004,2006,2008,2009,2010,2011,2012), Philippines (1998,2003,2008,2013,2017), Rwanda (1992, 2000,2005,2010), Sao Tome & Principe (2009), Senegal (1992,1997,2005,2011,2012,2014,2015,2016,2017), Sierra Leone (2008,2013), South Africa (1998,2016), Swaziland (2006), Tanzania (1992,1996,1999,2004, 2010,2015), Togo (1998,2014), Tajikistan (2012,2017), Timor Leste (2009,2016), Turkey (1993,1998,2003,2008,2013), Uganda (1995,2001,2006,2011), Ukraine (2007), Uzbekistan (1996), Vietnam (1997,2002), Yemen (1991,2013), Zambia (1996,2002,2007,2014), Zimbabwe (1994,1999,2006,2011,2015).

Data added for Albania, Congo Brazzaville, Lao, Sierra Leone, South Africa, Tajikistan, Thailand.

UNICEF Multiple Indicator Cluster Surveys (MICS)

UNICEF supports countries to collect data on the situation of children and women through the Multiple Indicator Cluster Survey (MICS) program. The MICS was originally developed in response to the World Summit for Children to measure progress towards an internationally agreed set of mid-decade goals. Since the initiation of the program, several rounds of surveys have been carried out and close to 300 MICS have been carried out in more than 100 countries, generating data on key indicators on the well-being of children and women, and helping shape policies for the improvement of their lives. MICS has been a major source of data on the Millennium Development Goals (MDG) indicators and will continue to be a major data source during the 2030 Sustainable Development Agenda to measure Sustainable Development Goals (SDGs) indicators. Further information is available at http://mics.unicef.org

In the GDL Area database currently data from the following MICS surveys are used: Afghanistan (2010), Algeria (2011), Angola (2000), Argentina (2011), Azerbaijan (2000), Barbados (2012), Belarus (2005,2012), Bosnia Herzegovina (2006,2011), Botswana (2000), Burundi (2005), Belize (2006), Bhutan (2010), Central African Republic (2006,2010), Chad (2010), Congo Brazzaville

(2015), (Congo Democratic Republic (2010), Costa Rica (2011), Cuba (2011), Djibouti (2006), El Salvador (2014), Equatorial Guinea (2000) Georgia (2005), Ghana (2006,2011), Gambia (2000, 2006), Guinea Bissau (2006,2014), Guyana (2006,2014), Iraq (2006, 2011), Jamaica (2005; 2011), Kazakhstan (2006,2010,2015), Kosovo (2014), Kyrgyzstan (2006,2014), Lao (2012,2017), Macedonia (2005,2011), Mongolia (2005,2010), Mauritania (2007,2011), Mexico (2015, Montenegro (2005,2013), Myanmar (2000), Palestine (2010,2014), Panama (2013), Sao Tome & Principe (2000,2014), Serbia (2005,2010,2014), Sierra Leone (2017), Sudan (2000,2010,2014), Somalia (2006,2011), South-Sudan (2010), St. Lucia (2012), Suriname (2006, 2010), Swaziland (2000,2010), Syria (2006), Togo (2006), Thailand (2006;2012,2016), Tajikistan (2000,2005), Trinidad & Tobago (2006), Turkmenistan (2006,2015), Tunisia (2011), Ukraine (2005,2012), Uruguay (2013), Uzbekistan (2006), Vietnam (2006,2010,2014), Vanuatu (2007), Yemen (2006).

Integrated Public Use Microdata Series (IPUMS) International

IPUMS-International is an effort of the Minnesota Population center of the University of Minneapolis to inventory, preserve, harmonize, and disseminate census microdata from around the world. The project has collected the world's largest archive of publicly available census samples. The data are coded and documented consistently across countries and over time to facilitate comparative research. IPUMS-International makes these data available to qualified researchers free of charge through a web dissemination system. The IPUMS project is a collaboration of the Minnesota Population Center, National Statistical Offices, and international data archives. Major funding is provided by the U.S. National Science Foundation and the Demographic and Behavioral Sciences Branch of the National Institute of Child Health and Human Development. Additional support is provided by the University of Minnesota Office of the Vice President for Research, the Minnesota Population Center, and Sun Microsystems. Further information is available at http://international.ipums.org

In the GDL Area database currently data from the following IPUMS are used: Brazil (2000,2010), Botswana (2001,2011), Chili (2002), Fiji (2007), Iran (2006,2011), Malaysia (2000), Mexico (2000,2010,2015), Paraguay (2002), South Africa (2011), South Sudan (2008), Sudan (2008), Venezuela (2001). These data were derived from the Integrated Public Use Microdata Series, International: Version 6.3 [Machine-readable database].

Afrobarometer Surveys

Afrobarometer is an African-led, non-partisan research network that conducts public attitude surveys on democracy, governance, economic conditions, and related issues across more than 30 countries in Africa and are repeated on a regular cycle. Each of the barometers is implemented independently. In each country, a national research team administers a country-wide face-to-face survey using standardized survey instruments to compile the required micro-level data under a

common research framework and research methodology. Core donors for Afrobarometer Rounds 5 and 6 include: The Mo Ibrahim Foundation, The Swedish International Development Cooperation Agency (SIDA), Department for International Development (DFID), UK and The United States Agency for International Development (USAID). Supplemental funding is provided by: The World Bank, Institute for Security Studies (South Africa), United States Institute of Peace, Transparency International and The Bill & Melinda Gates Foundation. Further information is available at http://www.afrobarometer.org

In the GDL Area Database currently data from 6 Afrobarometer surveys for 4 countries are used: Botswana (2012,2014), Cape Verde (2012,2014), Mauritius (2012,2014), Morocco (2012,2015). Because the samples are relatively small, the data for both years for these countries were combined.

AmericasBarometer (LAPOP) Surveys

The AmericasBarometer is an scientifically rigorous comparative survey that covers all of the mainland independent countries in North, Central, and South America, as well as a significant number of countries in the Caribbean. The AmereicasBarometer is part of the Latin American Public Opinion Project (LAPOP) of the Department of Political Science of Vanderbilt University. With the AmericasBarometer, LAPOP measures values, behaviors, and socio-economic conditions in the Americas using national probability samples of voting-age adults. Stratified sample designs, available on LAPOP's website, permit complex analyses of individuals nested within sub-regions. Survey participants are voting-age adults interviewed face to face in their households, except in Canada and the United States where the interviews are Web-based. Major supporters of LAPOP are the United States Agency for International Development, the Inter-American Development Bank, and Vanderbilt University. Further information is available at http://www.vanderbilt.edu/lapop/

In the GDL Area Database currently data are used from AmericasBarometer surveys for the following countries: Ecuador (2006, 2008, 2010, 2012), El Salvador (2006, 2008, 2010, 2012, 2014), Guatemala (2006, 2008, 2010, 2012, 2014), Jamaica (2006,2008,2010,2012,2014), Nicaragua (2006, 2008, 2010, 2012, 2014), Paraguay (2006,2008, 2010, 2012,2014), Uruguay (2006, 2008,2010,2012,2014), Venezuela (2007). Because the samples are relatively small, data for several years were combined.

Arab Barometer Surveys

The Arab Barometer was established in 2005 by Scholars in the Arab world and the United States. The University of Michigan and Princeton University.S. provided initial leadership, in partnership with universities and research centers in Jordan, Palestine, Morocco, Algeria and Kuwait. In 2010, the Arab Democracy Barometer formed a partnership with the Arab Reform Initiative (ARI, www.arab-reform.net) to expand the project's scope and range of activities, building off ARI's regional survey work carried out in 2006-2008.

The Arab Barometer was developed in consultation with the Global Barometer project (http://www.globalbarometer.net). Like other regional Barometers, the objectives of the Arab Barometer are to produce scientifically reliable data on the politically-relevant attitudes of ordinary citizens, to disseminate and apply survey findings in order to contribute to political reform, and to strengthen institutional capacity for public opinion research. Further information is available at http://www.arabbarometer.org/

In the GDL Area Database currently data are used from Arab Barometer surveys for the following countries: Lebanon (2013), Libya (2014), Kuwait (2013), Saudi Arabia (2011).

Asian Barometer Surveys

The Asian Barometer (ABS) is an applied research program on public opinion on political values, democracy, and governance around the region. The regional network encompasses research teams from fourteen East Asian political systems (Japan, Mongolia, South Korea, Taiwan, Hong Kong, China, the Philippines, Myanmar, Thailand, Vietnam, Cambodia, Singapore, Malaysia, and Indonesia), and five South Asian countries (India, Pakistan, Bangladesh, Sri Lanka, and Nepal).

Together, this regional survey network covers virtually all major political systems in the region, systems that have experienced different trajectories of regime evolution and are currently at different stages of political transition. http://www.asianbarometer.org

In the GDL Area Database currently data are used from the Asian Barometer survey for Malaysia (2011).

The Pan Arab Project for Family Health (PAPFAM)

The Pan Arab Project for Family Health is a data collection programme in Arab Countries that is rather similar to the Demographic and Health Surveys and MICS programmes. PAPFAM surveys are mostly funded by the Leage of Arab States. Further information is available at http://www.papfam.org

In the GDL Area Database currently data from the Algerian (2002) PAPFAM survey and the Moroccan (2002) joint PAPFAM-DHS survey is used.

Other data sources

For China 2002, data from the Chinese Household Income Project is used. The purpose of this project was to measure and estimate the distribution of personal income and related economic factors in both rural and urban areas of the People's Republic of China. Data were collected through a series of questionnaire-based interviews conducted in rural and urban areas at the end

of 2002. Further information is available here:
http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/21741

For China 2011, data from the Chinese Household Finance Survey is used. This survey was conducted by the Research Center for China Household Finance, a non-profit institute for academic inquiry, based at Southwestern University of Finance and Economics (SWUFE). The China Household Finance Survey (CHFS) contains detailed information about household finance and assets, including housing, business assets, financial assets, and other household assets, plus information about income and expenditures and social and commercial insurance. Further information is available here: http://www.chfsdata.org

For India 2012, data from the Indian Human Development Survey (IHDS) are used. The IHDS is a nationally representative, multi-topic panel survey of 41,554 households in 1503 villages and 971 urban neighborhoods across India. The first round of interviews were completed in 2004-5. A second round of IHDS reinterviewed most of these households in 2011-12 (N=42,152). Households that could not be found back were replaced by new households in such a way that the 2011-12 sample can be used as a national representative sample of India. Given that the large majority of interviews was held in 2012, it is called the India 2012 survey here. The IHDS has been jointly organized by researchers from the University of Maryland and the National Council of Applied Economic Research (NCAER), New Delhi. Further information is available here: http://www.ihds.info; http://doi.org/10.3886/ICPSR36151.v2

For South Africa 2014, data from the South African General Household Survey (GHS) are used. The GHS is an annual household survey conducted by Stats SA since 2002. The GHS is an omnibus household based instrument aimed at determining the progress of development in the country. It measures, on a regular basis, the performance of programmes as well as the quality of service delivery in a number of key service sectors in the country. The GHS covers six broad areas, namely education, health and social development, housing, household access to services and facilities, food security, and agriculture. Further information is available at:
- http://www.statssa.gov.za/publications/P0318/P03182014.pdf
- https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/526

National Indicators

The national population figures used to compute the area population sizes are derived from the World Development Indicators of the World Bank (http://data.worldbank.org/products/wdi).

3. Methods

The indicators made available through the GDL Area Database are created by aggregation from the household surveys datasets. Aggregation means taking the average of the values of a characteristic of individuals or households in each area.

For instance, the indicator of the educational level of the population aged 20-39 in an area is the mean years of education of the respondents in this age group in the area. And the area's vaccination coverage is the percentage of children aged one who received a specific vaccination.

In all cases, the sample weight factors available in the datasets are used to get indicators that are as well as possible representative for the areas to which they apply.

Data are aggregated to the country level, the level of urban and rural areas, of poor and nonpoor households, of four wealth quartiles, and of subnational regions within the country.

Geographic subdivisions

The sub-national areas that are used for aggregation are based on the geographic information that is present in the survey datasets. The available regional coding is often, but not always, based on official administrative subdivisions used in the countries.

Generally first-level administrative units are used, but there are many deviations. Some datasets use a coding of their own, that may consist of a combination of administrative units or be completely stand alone.

Even when official classifications are used, the situation may be complicated by the fact that those classifications change over time. It often happens that new regions are created by splitting up or merging existing regions. In those cases, the subdivisions used in earlier and/or later surveys had to be adjusted to keep comparability over time. Such adjustments always imply a reduction of the number of areas.

Reductions sometimes also are made for small datasets, to increase the number of cases on the basis of which aggregation takes place. This means that the subdivision used in the GDL Area Database often contain somewhat less regions than the official ones in a given year. On average, about ten subnational regions are distinguished within the countries.

For some countries, the regional codes differed so much between years that making them comparable was not possible or would lead to much information loss. In those cases different classifications were used for different years.

Aggregation to urban and rural areas is done on the basis of the urban/rural variables available in the datasets. Depending on survey type and preferences of data collecting organizations, the definition of this variable may differ. In most countries, the definitions used by national statistical agencies were taken.

Wealth, poverty, inequality

Measuring household welfare in low income countries is difficult, as income is often unstable, unregistered, informal, or in non-monetary form. Household expenditure potentially is a better option, but is not widely available in household surveys and may suffer of measurement and comparability problems. Using an asset-based wealth index provides a useful alternative, as assets are easily observable to interviewers and information on assets is widely available in survey datasets for LMICs.

The International Wealth Index (IWI) is an asset-based wealth index that is comparable across time and place. IWI is based on household-level information about possession of durables (bike, phone, TV, refrigerator, car and a cheap and expensive utensil), housing characteristics (quality of floor and toilet and number of rooms) and access to basic services (water and electricity). IWI ranks households on a scale running from 0 to 100, with 0 for households with none of the assets, bad quality housing and no services, and 100 for households with all assets, highest quality housing and good services. Any household in the world can be placed on the IWI scale using a simple additive formula. For further information see Smits & Steendijk (2015).

The indicators for household wealth, poverty and inequality available in the GDL Area Database are based on IWI. To indicate the level of development of an area, the mean value IWI in the region is taken. To indicate the level of poverty, the percentage of households under an IWI value of 50 is used (IWI50 poverty line). To indicate the level of extreme poverty, the percentage of households under an IWI value of 35 is used (IWI35 poverty line).

To study variation between poor and nonpoor households for the indicators in the GDL Area Database, indicators are presented separately for both groups. For this distinction, the IWI50 poverty line is used. Separate indicators for poor and nonpoor households are only presented if there are at least 20 households in the smallest category.

To study inequality within countries for the indicators in the GDL Area Database, these indicators are presented separately by wealth quartiles. To allocate the households available in a dataset to a wealth quartile, country-specific wealth indices are used (like the standard wealth indices available in the Demographic and Health Surveys). These country specific indices are most precise in allocating the households within a specific survey to wealth quartiles. Because they are not meant to be comparable, they can be based on all asset information available in a specific dataset (while IWI is based on a smaller subset available in all surveys). Country-specific wealth indices are available for the Demographic and Health surveys and the UNICEF MICS surveys. For other surveys, the household's IWI value is used to determine its wealth quintile.

Education

To create indicators of the educational level of the population in an area, the available educational level variables are translated into years of education. For DHS surveys and some other surveys, already a years of education variable is present in the data.

In MICS surveys and some other surveys, both a level and a grade variable are available, whereby the grade variable indicates what the highest grade is which the respondents have completed within their highest level. For these surveys, years of education could be computed by combining these variables.

I a substantial number of surveys, only the highest completed educational level is available. This variable is then recoded into a years of education variable as follows: None=0, primary=6, lower/junior secondary=9, upper/senior secondary=12, tertiary=16.

Child mortality

The Area Database contains for most countries indicators for infant mortality and under five mortality and for many countries also neonatal, postneonatal and child mortality. Infant mortality is defined as the number of deaths of children under age 1; under 5 mortality as the number of children dying under exact age five years; neonatal mortality as the number of neonates dying under exact age one month (29 days), postneonatal mortality as the number of children dying between 29 days of age and exact age one year, and child mortality as the number of children dying between exact one and exact five year of age, in all cases per 1,000 live births in a given year.

For Demographic and Health Surveys, child mortality rates are estimated with the 'direct' estimation method on the basis of information derived from birth histories collected from women aged 15-49 in oral interviews (Rutstein & Rojas, 2006). For MICS surveys and IPUMS, the rates are estimated using the 'indirect' method on the basis of information on the number of children born and the number of those children who died derived from oral interviews with women in different age groups (United Nations, 1990, 2011). The direct method is more precise and produces all five measures. However, the required data is more expensive to collect and more error prone. If good data is lacking, the indirect method forms a reasonable alternative, particularly in stable populations (Rutstein & Rojas, 2006).

Fertility

Under the heading "Fertility", the total fertility rate (TFR) and age-specific fertility rates are presented in the Area Database. Fertility indicators are computed on the basis of the birth history information available in the DHS surveys. To compute these indicators, information is used for a period of three years before the survey, following the procedure described by Rutstein and Rojas (2006).

Age-specific fertility rates indicate the number of births per 1000 women in a specific age group per year. The total fertility rate indicates the total number of children a woman at age 50 would have given birth to, if she would have experienced the current age-specific fertility rates. To

compute the total fertility rate, the age-specific fertility rates for the age groups from 15-19 to 45-49 are summed up and multiplied by five.

GDL code

To be able to identify each possible level (row in the Area Database) a unique code was developed which we call the GDL code. This code always starts with the three-digit iso_code (in capitals) for the country, followed by one or more lower case letters and numbers indicating what kind of level the code indicates. The codes have the following meaning:

't'   The country as a whole,
'tu'   Urban areas,
'tr'   Rural areas,
'p0'   Nonpoor areas,
'p1'   Poor areas,
'w41'   Lowest wealth quartile,
'w42'   Second wealth quartile,
'w43'   Third wealth quartile,
'w44'   Highest wealth quartile
'r1'   Sub-national region, classification 1 (followed by a two digit region number)
'r2'   Sub-national region, classification 2 (followed by a two digit region number)

Quality of indicators

The fact that the indicators are created on the basis of household surveys means that they to a certain extent suffer from bias. Their quality (correctness) depends on the design, size, structure and quality of the household survey on which they are based and on the number of sub-national areas distinguished. Detailed information on survey design and quality can be obtained from the producing organizations mentioned in the data section.

The indicators also suffer to some extent from random aggregation error, as they are based on samples of the total population. Given that no clear criteria exist for the number of cases to be used for aggregation of indicators, we provide for all our indicators in all our regions the number of persons/households on which they are based. This offers users the possibility to make their own choices in this respect.

Although the surveys we use are designed to be representative at the national level and in almost all cases also at the level of sub-national subdivisions, they are not always designed to be representative for the sub-national areas we distinguish. In those cases, we still present our indicators, as they often are the only available data for the region, which we consider better than no information at all.

The indicators shown at the GDL Area Database website can be downloaded from the website as a CSV or Excel file by clicking on "Download this". The complete set of the indicators can be downloaded by clicking on "Download all". You then enter a page where you find separate CSV

files with for each region the indicators and the number of households or persons on which they are based and with the labels and descriptions of the indicators.

Special requests

Organizations or persons interested in specific (combinations of) indicators can submit a request for creation of those (combinations of) indicators to the Global Data Lab.

4. Indicators

The table below provides an overview of all indicators included in the current version of the GDL Area Database, together with the acronym that is used for it, the broader category in which it is included and a description of its content.

Acronym	Category	Description
iwi	Wealth & Poverty	Mean International Wealth Index (IWI) score of area. IWI is a comparative asset-based wealth index at the household level, based on the possession of consumer durables, housing characteristics and access to basic services (for details see Smits & Steendijk, 2015). IWI runs from 0 (none of the durables, bad quality housing and no access to services) to 100 (all durables, highest quality housing, all services).
iwipov50	Wealth & Poverty	IWI-based poverty measure. Percentage of households with an IWI value under 50 in the area. Highly correlated with World Bank poverty headcount ratios.
iwipov35	Wealth & Poverty	IWI-based poverty measure. Percentage of households with an IWI value under 35 in the area. Highly correlated with World Bank poverty headcount ratios.
urban	Urbanization	Percentage of population living in urban areas in the area (according to national definition)
edyr20+	Years of education	Mean years of education of adults aged 20+ in the area
womedyr20+	Years of education	Mean years of education of women aged 20+ in the area
menedyr20+	Years of education	Mean years of education of men aged 20+ in the area
edyr2039	Years of education	Mean years of education of adults aged 20-39 in the area
womedyr2039	Years of education	Mean years of education of women aged 20-39 in the area
menedyr2039	Years of education	Mean years of education of men aged 20-39 in the area
edyr4059	Years of education	Mean years of education of adults aged 40-59 in the area
womedyr4059	Years of education	Mean years of education of women aged 40-59 in the area
menedyr4059	Years of education	Mean years of education of men aged 40-59 in the area
edyr60+	Years of education	Mean years of education of adults aged 60+ in the area
womedyr60+	Years of education	Mean years of education of women aged 60+ in the area
menedyr60+	Years of education	Mean years of education of men aged 60+ in the area
lprimary	Educational attendance	Percentage of children aged 6-8 in the area that currently attends school, or in the current school year attended school
uprimary	Educational attendance	Percentage of children aged 9-11 in the area that currently attends school, or in the current school year attended school
lsecondary	Educational attendance	Percentage of children aged 12-14 in the area that currently attends school, or in the current school year attended school
usecondary	Educational attendance	Percentage of children aged 15-17 in the area that currently attends school, or in the current school year attended school

tertiary	Educational attendance	Percentage of children aged 18-21 in the area that currently attends school, or in the current school year attended school
lprimaryg	Educational attendance	Percentage of girls aged 6-8 in the area that currently attends school, or in the current school year attended school
uprimaryg	Educational attendance	Percentage of girls aged 9-11 in the area that currently attends school, or in the current school year attended school
lsecondaryg	Educational attendance	Percentage of girls aged 12-14 in the area that currently attends school, or in the current school year attended school
usecondaryg	Educational attendance	Percentage of girls aged 15-17 in the area that currently attends school, or in the current school year attended school
tertiaryg	Educational attendance	Percentage of girls aged 18-21 in the area that currently attends school, or in the current school year attended school
lprimaryb	Educational attendance	Percentage of boys aged 6-8 in the area that currently attends school, or in the current school year attended school
uprimaryb	Educational attendance	Percentage of boys aged 9-11 in the area that currently attends school, or in the current school year attended school
lsecondaryb	Educational attendance	Percentage of boys aged 12-14 in the area that currently attends school, or in the current school year attended school
usecondaryb	Educational attendance	Percentage of boys aged 15-17 in the area that currently attends school, or in the current school year attended school
tertiary	Educational attendance	Percentage of boys aged 18-21 in the area that currently attends school, or in the current school year attended school
tv	Asset ownership	Percentage of households owning a TV in the area
fridge	Asset ownership	Percentage of households owning a fridge in the area
washmach	Asset ownership	Percentage of households owning a washing machine in the area
computer	Asset ownership	Percentage of households owning a computer in the area
cellphone	Asset ownership	Percentage of households owning a cellphone in the area
phone	Asset ownership	Percentage of households owning a phone in the area
agedifmar	Gender	Mean age difference partners (husband-wife) in the area
agemarw20	Gender	Mean age at first marriage of women aged 20-50 in the area
age1chw20	Gender	Mean age at first birth of women aged 20-50 in the area
womo50	Gender	% of women in 50+ population
patrilocal	Gender	Patrilocality Index: computed as log(% patrilocal / % matrilocal).
ppatloc	Gender	Percentage of households where couple lives with parents of husband in the area (% patrilocal)
pmatloc	Gender	Percentage of households where couple lives with parents of wife in the area (% matrilocal)
tfr	Fertility	Total Fertility Rate for the 3 years before survey
asf1014	Fertility	Age-specific fertility rate age 10-14 for the 3 years before survey
asf1519	Fertility	Age-specific fertility rate age 15-19 for the 3 years before survey
asf2024	Fertility	Age-specific fertility rate age 20-24 for the 3 years before survey
asf2529	Fertility	Age-specific fertility rate age 25-29 for the 3 years before survey

asf3034	Fertility	Age-specific fertility rate age 30-34 for the 3 years before survey
asf3539	Fertility	Age-specific fertility rate age 35-39 for the 3 years before survey
asf4044	Fertility	Age-specific fertility rate age 40-44 for the 3 years before survey
asf4549	Fertility	Age-specific fertility rate age 45-49 for the 3 years before survey
stunting	Child stunting & wasting	Percentage of children aged 0–59 in the area who are below minus two standard deviations from median height-for-age of the WHO Child Growth Standards
stuntingmod	Child stunting & wasting	Percentage of children aged 0–59 in the area who are between minus two and minus three standard deviations from median height-for-age of the WHO Child Growth Standards
stuntingsev	Child stunting & wasting	Percentage of children aged 0–59 in the area who are below minus three standard deviations from median height-for-age of the WHO Child Growth Standards
haz	Child stunting & wasting	Number of SD below or above the median height-for-age of the WHO Growth Standard
wasting	Child stunting & wasting	Percentage of children aged 0–59 in the area who are below minus two standard deviations from median weight-for-height of the WHO Child Growth Standards
wastingmod	Child stunting & wasting	Percentage of children aged 0–59 in the area who are between minus two and minus three standard deviations from median weight-for-height of the WHO Child Growth Standards
wastingsev	Child stunting & wasting	Percentage of children aged 0–59 in the area who are below minus three standard deviations from median weight-for-height of the WHO Child Growth Standards
whz	Child stunting & wasting	Number of SD below or above the median weight-for-height of the WHO Growth Standard
underweight	Child underweight and overweight	Percentage of children aged 0–59 in the area who are below minus two standard deviations from median weight-for-age of the World Health Organization (WHO) Child Growth Standards
underweightmod	Child underweight and overweight	Percentage of children aged 0–59 in the area who are between minus two and minus three standard deviations from median weight-for-age of the World Health Organization (WHO) Child Growth Standards
underweightsev	Child underweight and overweight	Percentage of children aged 0–59 in the area who are below minus three standard deviations from median weight-for-age of the World Health Organization (WHO) Child Growth Standards
waz	Child underweight and overweight	Number of SD below or above the median weight-for-age of the WHO Growth Standard
overweight	Child underweight and overweight	Percentage of children aged 0-59 in the area who are above two standard deviations from median weight-for-height of the WHO Child Growth Standards
obesity	Child underweight and overweight	Percentage of children aged 0-59 in the area who are above three standard deviations from median weight-for-height of the WHO Child Growth Standards

bmiz	Child underweight and overweight	Number of SD below or above the median BMI-for-age of the WHO Growth Standard
bcgage1	Vaccinations	Percentage of children aged 1 in the area who received a BCG vaccination
dtp1age1	Vaccinations	Percentage of children aged 1 in the area who received a DTP1 vaccination
dtp2age1	Vaccinations	Percentage of children aged 1 in the area who received a DTP2 vaccination
dtp3age1	Vaccinations	Percentage of children aged 1 in the area who received a DTP3 vaccination
measlage1	Vaccinations	Percentage of children aged 1 in the area who received a measles vaccination
regpopm	Population	Total area population in millions
popshare	Population	Share of national population living in the area
age09	Population	Percentage of national population aged 0-9 living in the area
age1019	Population	Percentage of national population aged 10-19 living in the area
age2029	Population	Percentage of national population aged 20-29 living in the area
age3039	Population	Percentage of national population aged 30-39 living in the area
age4049	Population	Percentage of national population aged 40-49 living in the area
age5059	Population	Percentage of national population aged 50-59 living in the area
age6069	Population	Percentage of national population aged 60-69 living in the area
age7079	Population	Percentage of national population aged 70-79 living in the area
age8089	Population	Percentage of national population aged 80-89 living in the area
age90hi	Population	Percentage of national population aged 90+ living in the area
hhsize	Population	Average household size in the area
depratio	Demographic window	Dependency ratio, dependent population compared to the working age population (15-64) in the area
youthdepr	Demographic window	Youth dependency ratio, the young (<15) compared to the working age population (15-64) in the area
olddepr	Demographic window	Old age dependency ratio, the old (>64) compared to the working age populations (15-64) in the area
dwindowphase	Demographic window	Phase of Demographic Window: 1 Traditional phase (>40% under 15), 2 Pre-window phase (30-40% under 15), 3 Early-window phase (25-30% under 15), 4 Mid-window phase (20-30% under 15), 5 Late-window phase (<20% under 15 and <15% over 64), 6 Post-window phase (>15% over 64).
popyoung	Demographic window	Percentage of population in the area under age 15
popworkage	Demographic window	Percentage of population in the area aged 15 to 65
popold	Demographic window	Percentage of population in the area over 65
nnmort	Child mortality	Number of neonates dying before 28 days of age in the area, per 1,000 live births in a given year

pnnmort	Child mortality	Number of children dying between 29 days of age and one year in the area, per 1,000 live births in a given year
infmort	Child mortality	Number of deaths of children less than one year of age in the area, per 1000 live births in a given year
chmort	Child mortality	Number of children dying between 1 year of age and five year of age in the area, per 1,000 live births in a given year
u5mort	Child mortality	Number of children dying under five year of age in the area, per 1,000 live births in a given year
pipedwater	Public services	Percentage of household with piped water in the area
electr	Public services	Percentage of households with electricity in the area
flushtoilet	Quality of housing	Percentage of households with a flush toilet in the area
smallhouse	Quality of housing	Percentage of households with none or one sleeping room in the area
modsizehouse	Quality of housing	Percentage of households with two sleeping rooms in the area
largehouse	Quality of housing	Percentage of households with three or more sleeping rooms in the area
naturalfloor	Quality of housing	Percentage of households with a natural floor (earth, sand, dung etc) in the area
cookwood	Quality of housing	Percentage of households cooking on wood, straw, grass, dung etc. in the area
nprimpov50	Special requests	Percentage of children aged 6-11 who are not in school in the area living in poor households (under IWIpov50).
chnprimu50	Special requests	Number of children aged 6-11 who are not in school in the area living in poor households (under IWIpov50)
primnot	Special requests	Percentage of children aged 6-11 that currently does not attend or in the current school year did not attended school

Appendix A

Terms of use & disclaimer

The Global Data lab and the researchers connected to it (henceforth mentioned GDL) create databases and develop instruments for measuring and analyzing the status and progress of societies. The GDL Area Database is one of the databases created by GDL. The GDL Area Database is part of and can be reached through the GDL Website (www.globaldatalab.org).

The GDL Area Database contains development indicators at the national and sub-national level for low and middle income countries (LMICs). These indicators were created by the GDL team and are made available to the global community to increase our knowledge and understanding of the situation and developments in these countries

The development indicators available at the GDL Website can be freely downloaded and used by any interested person or party, provided that the Global Data Lab is mentioned as the source of the indicators and a link to the GDL Website is shown.

You may make copies of the indicators, use them for any purpose, adapt them and/or combine them with other data, and share them with others in any way you like, as long as you take full responsibility for these actions.

You may not publicly or privately represent or imply that GDL is participating in, or has sponsored, approved or endorsed the manner or purpose of your use or reproduction of the indicators.

If you encounter an error or other problem at our website, please report it to us by email, so that we can keep the website and the indicators presented through it at the best possible quality.

By using the GDL Website in any manner, you declare to agree with and be bound by these Terms of use.

Disclaimer

Although every effort was made to create indicators and other content of high quality, we cannot guarantee the correctness of the information and indicators presented at the GDL Website.

GDL is not responsible nor liable for the accuracy, usefulness or availability of any indicator or other content and disclaims all warranties of any kind related to the provision of the indicators through the GDL Website. GDL assumes no responsibility for any direct or indirect loss or damage suffered by users or third parties in connection with the use of the GDL Website.

GDL reserves the right at any time and from time to time to modify or discontinue, temporarily or permanently, any indicator or other content available at the GDL Website at our sole discretion. GDL may under any circumstances, for any or no reason whatsoever and with or without prior notice terminate the provision of indicators through the GDL Website.

GDL is not responsible for the accuracy of the content offered by websites to which the GDL Website links, nor on websites through which the GDL Website was reached.

By using the GDL Website you acknowledge that these terms constitute a non-exclusive agreement for use of the indicators. GDL may develop products or services on the basis of the indicators or other content that compete with products or services that you offer without incurring any liability.

The designations employed and the presentation of the material on the GDL Website do not imply the expression of any opinion whatsoever on the part of GDL concerning the legal status of any country, territory, city of geographic area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Lines on maps represent approximate border lines for which there may not yet be full agreement.

These Terms of use may be updated by us at our discretion and without notice to you. The new version will be posted at our website under www.globaldatalab.org/areadata/termsofuse/. Please check this page regularly for changes. Your use of the Website following the posting of any changes to this page constitutes an acceptance of the altered Terms on your part.

Last updated May 31, 2016

References

Smits, J. & R. Steendijk (2015). The International Wealth Index (IWI). Social Indicators Research, 122(1), 65-85. link.springer.com/article/10.1007%2Fs11205-014-0683-x

Rutstein, S. O., & J. G. Rojas (2006). Guide to DHS Statistics. Calverton, Maryland: ORC Macro. www.measuredhs.com/pubs/pdf/DHSG1/Guide_DHS_Statistics.pdf

United Nations (1990). Step-by-Step Guide to the Estimation of Child Mortality. United Nations publication, Sales No. E.89.XIII.9.
www.un.org/en/development/desa/population/publications/pdf/mortality/stepguide_childmort.pdf

United Nations (2011). Mortality estimates from major sample surveys: towards the design of a database for the monitoring of mortality levels and trends. Population Division Technical Paper No. 2011/2. www.un.org/en/development/desa/population/publications/pdf/technical/TP2011-2_MortEstMajorSampSurv.pdf

United Nations (2016). Progress towards the Sustainable Development Goals. Report of the Secretary General. unstats.un.org/sdgs/files/report/2016/secretary-general-sdg-report-2016--EN.pdf