McKinsey Global Institute

Technical Appendix

Pixels of Progress: A granular look at human development around the world

By Chris Bradley, Marc Canal, Sven Smit and Jonathan Woetzel

December 2022

Introduction

This technical appendix provides a detailed explanation of the complete data sources, methodologies, and sensitivities and constraints of the estimates and analyses in this report. It is organized by variable—GDP, population, and life expectancy. We begin with an explanation of information that cuts across all of them.

Definition of microregion, boundary limits, and levels of granularity

Countries are divided into subnational administrative units at several levels, and the levels depend on each country's territorial organization. Administrative units vary across countries, which may break themselves down into states, provinces, counties, districts, and municipalities, among other types of units. We define all such units with the generic term "microregion." In other words, a microregion in this report is defined as an administrative unit smaller than a country. It follows that countries can have several microregional levels. For example, the United States has states and counties.

It is important to note that our microregions are administrative units; they do not share the same area size, amount of population, or any other metric. In fact, they vary substantially in size, population, and GDP. For example, the province of Guangdong, China, has almost 125 million inhabitants, while Ningxia, another Chinese province, has about seven million inhabitants. However, both Chinese provinces are at the same administrative level.

We have used this system of administrative units for two primary reasons. First, development variables relate to policy, and policy is determined at different levels of administration. For the same reasons that we typically compare, for example, outcomes for European countries instead of units of the same size, it makes sense to compare subnational administrative units. Second, most information sources collect data by administrative unit, and so we can access a significantly larger amount of data at that level.

We have three levels—level 1, level 2, and level 3—of administrative units, defined as follows:

- For EU countries and the United Kingdom, the EU Nomenclature of Territorial Units for Statistics (NUTS)
- For the United States, the US Census Bureau TIGER/Line shapefiles for counties
- For all other countries, the GADM database of Global Administrative Areas version 4.0

Depending on their size and political organization, not all countries have all three levels of administrative units. While NUTS, for instance, breaks EU countries into three levels, NUTS 1, NUTS 2, and NUTS 3, our research considered the United States on only two levels, states and counties. Additionally, we cannot estimate certain levels of granularity due to data limitations, so we eliminated some levels. For example, while we can estimate GDP for more than 2,300 Indian microregions using nighttime satellite imagery, we lacked adequate data to estimate life expectancy in India at a level of granularity more detailed than its 33 states.

Additionally, "level" can mean different things in different countries. Consider Spain and Brazil. Spain is divided at level 1 into seven microregions—south, northwest, northeast, east, center, Comunidad de Madrid, and Canarias—according to NUTS 1. Level 2 in Spain is based on NUTS 2, or what are known as "autonomous communities," which in turn are home to NUTS 3 "provinces." In Brazil, states are level 1 and municipalities level 2, based on GADM. Brazil has 5,495 Level 2 units, while Spain has only 19.

These breakdowns have the advantage of being globally accepted and used, and they often stem directly from subnational divisions that countries themselves provide to international organizations like the OECD or statistical agencies like Eurostat. However, this also means they have limitations in comparability across countries and microregions. To resolve these, we were careful to aggregate our microregions at levels that are comparable when making direct comparisons and to indicate the level at which our analyses were conducted.

Finally, for some analyses, we aggregated countries into larger units, or subcontinents. We defined ten subcontinents: Advanced Asia, China, Eastern Europe, Emerging Asia, India, Latin America and Caribbean, Middle East and North Africa (MENA), North America, Sub-Saharan Africa, and Western Europe. Exhibit A1 lists the subcontinents and the countries allocated to them in our research.

Exhibit A1

List of subcontinents and countries

SUBCONTINENT	COUNTRIES
Advanced Asia	Australia, Japan, New Zealand, Singapore, South Korea
China	China, Hong Kong SAR, China, Macao SAR, China
Eastern Europe	Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Georgia, Hungary, Kosovo, Latvia, Lithuania, Moldova, Montenegro, North Macedonia, Poland, Romania, Russia, Serbia, Slovakia, Slovenia, Ukraine
Emerging Asia	Afghanistan, Bangladesh, Bhutan, Brunei, Cambodia, Fiji, Indonesia, Kazakhstan, Kiribati, Kyrgyzstan, Laos, Malaysia, Maldives, Micronesia, Mongolia, Myanmar, Nepal, Pakistan, Palau, Papua New Guinea, Philippines, Samoa, Solomon Islands, Sri Lanka, Tajikistan, Thailand, Timor-Leste, Tonga, Turkmenistan, Uzbekistan, Vanuatu, Vietnam
India	India
Latin America and Caribbean	Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, Turks and Caicos Islands, Uruguay, Venezuela
MENA	Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Syria, Tunisia, Türkiye, United Arab Emirates, Yemen
North America	Canada, United States
Sub-Saharan Africa	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Côte d'Ivoire, Democratic Republic of the Congo, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Republic of Congo, Rwanda, São Tomé and Príncipe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe
Western Europe	Andorra, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Liechtenstein, Luxembourg, Malta, Netherlands, Norway, Portugal, San Marino, Spain, Sweden, Switzerland, United Kingdom

Data sets and variables

We have built three variables worldwide: GDP, population, and life expectancy. We combined these variables to create other variables such as GDP per capita (GDP divided by population). Our GDP data were largely derived from nighttime satellite imagery; population data came from the global WorldPop data set, and we estimated life expectancy using a combination of sources. The way each variable was built is explained in the following sections.

Not all variables are available at the same level of granularity. With nighttime satellite imagery and the WorldPop database, we can estimate GDP and population for about 52,000 microregions worldwide, but for life expectancy, we can produce data for only 43,000 microregions. Hence, our research used two data sets, one containing GDP and population data at maximum granularity and one containing GDP, population, and life expectancy data. When using the second data set, we adopted a minimum common denominator approach: when life expectancy data were not as granular as GDP or population data, we aggregated these two variables into less granular levels to match life expectancy.

When performing analyses, we used one data set or the other, depending on the purpose of each analysis. For example, if we compared Indian and Indonesian microregional GDP, we used the first data set because both countries can be divided into many microregions—more than 500 in Indonesia at level 2 and over 2,000 in India at level 3. However, when comparing life expectancy in the same two countries, since India has data for life expectancy only at level 1, or the level of its 33 states, we aggregated Indonesian life expectancy to level 1 as well for comparability purposes.

Country and time series coverage

We covered most countries and territories in the world. We excluded countries, territories, and regions for which World Bank country-level data are unavailable due to conflict or other circumstances. In particular, our research covers 178 countries that have data for GDP, population, and life expectancy. This includes all of the world's countries and territories except Andorra, the British Virgin Islands, Cuba, Eritrea, Greenland, Kosovo, Liechtenstein, the Marshall Islands, Nauru, North Korea, Palau, Puerto Rico, San Marino, Seychelles, South Sudan, Syria, Taiwan, the Turks & Caicos Islands, Tuvalu, Venezuela, Western Sahara, and Yemen.

We constructed complete time series from the beginning of 2000 through 2019. We did not include 2020 or 2021 given COVID-19 distortions and the lack of precise data at the time this data set was created. Nevertheless, most of our analyses were intended to describe and draw insights on long-term human development. Given that 2020 and 2021 were such exceptional years, it would be ill-advised to use them as an end point of our time series.

Gross domestic product

Here we provide the sources of our official statistics and explain in detail how we used nighttime satellite imagery to estimate subnational gross domestic product, or GDP. Microregional GDP estimates were based on a combination of official statistics when available and nighttime satellite imagery, adjusted using data sets built by Matti Kummu, Maija Taka, and Joseph H. A. Guillaume when official data were unavailable.[1]

Data sources for GDP

Exhibit A2 provides the GDP sources used for every country-year pair.

Exhibit A2

Data sources for GDP analysis

COUNTRY	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
All countries, where there are no official data	2000–13	Low-resolution, nighttime satellite images	DMSP OLS: Global Radiance-Calibrated Nighttime Lights Version 4, Defense Meteorological Program Operational Linescan System, US National Oceanic and Atmospheric Administration	Resolution of approximately 1km x 1km
All countries, where there are no official data	2014–20	High-resolution, nighttime satellite images	VIIRS Nighttime Day/Night Band Composites Version 2; Elvidge et al., "Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019," *Remote Sensing*, March 2021, volume 13, number 5	Resolution of approximately 500m x 500m
All countries	2000–20	GDP as measured in PPP constant prices, international dollars, 2017	GDP per capita, PPP (current international dollars), World Bank, 2021	
Armenia	2017–19	GDP at level 1	Statistical Committee of the Republic of Armenia, Armenia MDGs Indicators, Statistical Committee of the Republic of Armenia, 2021	Luminosity data used to estimate GDP data for 2000–16

Exhibit A2 continues

[1] Matti Kummu, Maija Taka, and Joseph H. A. Guillaume, "Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015," *Scientific Data*, February 2018, volume 5.

Exhibit A2 continued

COUNTRY	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
Argentina	2004, 2017–19	GDP at level 1	Producto interno bruto por jurisdicción, National Institute of Statistics and Census, Argentina, 2019	Luminosity data used to estimate GDP data for 2000–03 and 2005–16
Australia	2000–20	GDP at level 1	Australian National Accounts: State Accounts, Australian Bureau of Statistics, 2021	
Belarus	2012–17	GDP at level 1	National Accounts: Republic of Belarus, National Statistics Committee of the Republic of Belarus, 2019	Luminosity data used to estimate GDP data for 2000–11 and 2018–19
Bolivia	2000–20	GDP at level 1	Producto Interno Bruto según departemento, National Statistics Institute, Bolivia, 1988–2020	
Brazil	2002–17	GDP at level 1	OECD Regional Database, OECD, 2021	Luminosity data used to estimate GDP data for 2000–01 and 2018–19
Canada	2000–18	GDP at level 1	OECD Regional Database, OECD, 2021	Luminosity data used to estimate GDP data for 2019
Chile	2008–20	GDP at level 1	Annual GDP by region with chained prices to the previous year, Central Bank of Chile, 2020	Luminosity data used to estimate GDP data for 2000–07
China	2000–18	GDP at level 1	Regional Database, OECD, 2021	Luminosity data used to estimate GDP data for 2019
Colombia	2005–18	GDP at level 1	Regional Database, OECD, 2021	Luminosity data used to estimate GDP data for 2000–04 and 2019
India	2001–13	GDP at level 1	Regional Database, OECD, 2021	Luminosity data used to estimate GDP data for 2000 and 2014–19
Indonesia	2002–20	GDP at level 1	Produk Domestik Regional Bruto Provinsi di Indonesia, Statistics Indonesia, 2015–2019, 2020	Luminosity data used to estimate GDP data for 2000–01

Exhibit A2 continues

Exhibit A2 continued

COUNTRY	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
Japan	2001–17	GDP at level 1	Regional Database, OECD, 2021	Luminosity data used to estimate GDP data for 2000 and 2018–19
Malaysia	2005–20	GDP at level 1	GDP by State, Department of Statistics, Malaysia, 2020	Luminosity data used to estimate GDP data for 2000–04
Mexico	2000–19	GDP at level 1	Sistema de Cuentas Nacionales de México, Instituto Nacional de Estadística y Geografía, 2019, inegi.org	
New Zealand	2000–19	GDP at level 1	Regional gross domestic product: Year ended March 2020, Stats NZ	
Peru	2007–20	GDP at level 1	Producto Bruto Interno por Años, según Departamentos, Instituto Nacional de Estadística e Informática, 2020	Luminosity data used to estimate GDP data for 2000–06
Russia	2000–18	GDP at level 1	Regions of Russia, socioeconomic indicators, Federal Service for State Statistics, 2021	Luminosity data used to estimate GDP data for 2019
South Korea	2000–19	GDP at level 1	Regional Database, OECD, 2021	
Thailand	2000–19	GDP at level 1	Gross Regional and Provincial Product Chain Volume Measure 2019 Edition, Office of the National Economic and Social Development Council, Thailand, 2019	
South Africa	2000–13	GDP at level 1	Regional database, OECD, 2021	Luminosity data used to estimate GDP data for 2014–19
Ukraine	2004–19	GDP at level 1	Statistical collection "Gross Regional Product," State Statistics Service of Ukraine, 2021	Luminosity data used to estimate GDP data for 2000–03

Exhibit A2 continues

Exhibit A2 continued

COUNTRY	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
Brazil	2002–18	GDP at level 2	Population Estimates database, Instituto Brasileiro de Geografia e Estatícia, 2021	Luminosity data used to estimate GDP data for 2000–01 and 2019
China	2000–20	GDP at level 2	Municipal Bureau of Statistics from each Chinese prefecture, 2021	
United States	2000–19	GDP at level 2	GDP By County, Metro, and Other Areas, Bureau of Economic Analysis, 2020	
Albania, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Türkiye, United Kingdom	2000–20	GDP at level 3	Annual Regional Database of the European Commission's Directorate General for Regional and Urban Policy (ARDECO), Eurostat, 2021	
All countries, where there are no official data	2009–11	GDP at level 3	Matti Kummu, Maija Taka, and Joseph H. A. Guillaume, "Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015," *Scientific Data*, February 2018, volume 5	Adjustment factor applied to all countries with no official data at level 3

Data methodology

Estimating GDP at different levels of granularity

We use the World Bank's purchasing power parity (PPP), 2017 international dollars metric to establish national GDP. Consequently, since all levels of granularity are benchmarked to the national figure, they are expressed in PPP and adjusted for inflation. Below the national level, when we can find official GDP data, we use it whatever the level. Exhibit A2 shows, country by country, the cases in which we used official or peer-reviewed data at subnational levels. For example, we engaged with various national statistical agencies, the OECD, and other organizations to obtain level 1 GDP data for 22 countries in addition to EU countries and the United States.

Where we did not find official data, we used nighttime satellite imagery to estimate subnational GDP. Finally, in addition to nighttime satellite imagery, we used data from the Kummu et al. data set, which is based on census data and official subnational data sources, as an adjustment factor to our satellite imagery estimates (described below). We ensured that each level was correctly benchmarked to superior levels when coming from different data sets. For instance, we ensured that the sum of a country's level 1 microregions obtained from a national data source always added up to the national GDP estimated by the World Bank. When such aggregation at one level did not perfectly equal the immediately superior one, we adjusted the inferior level proportionally.

A set of examples further explains the process:

- Example 1 – For a country where official data exist at all levels, we used official data for all levels directly from that country.
- Example 2 – For a country with no official or peer-reviewed data at any level beyond the World Bank's national data, we used nighttime satellite imagery to estimate level 3 GDP and then added level 3 microregions to calculate level 2 microregions. We did the same from level 2 to level 1. We ensured that the sum of all level 3 microregions equaled national GDP, so nighttime luminosity became a way to geographically distribute national-level World Bank GDP data to a next level of granularity. If a country's maximum level of granularity was level 2, we followed the same process, but only up to level 2.
- Example 3 – For a country with level 1 data from an official source or a peer-reviewed paper but no further levels of granularity, we used level 1 data from the official source, benchmarked to the national World Bank GDP data. We then used nighttime satellite imagery to estimate level 3 GDP, ensuring that each level adds up to the superior one. In this case, nighttime luminosity can be thought of a way to distribute geographically level 1 official figures to next levels of granularity. If a country's maximum level of granularity was level 2, we followed the same process, but only up to level 2.

Using nighttime luminosity to estimate GDP

We relied on luminosity, based on nighttime satellite imagery, as a proxy for economic activity. The use of such data as a proxy for economic activity was introduced in 2012 in the *American Economic Review*, one of the world's leading economics journals.[2] Since then, rigorous peer-reviewed research, including by Nobel Prize winners, the World Bank, and the IMF, has scrutinized luminosity as a proxy for GDP.[3] These studies consistently find a strong relationship between the two variables. We built on this research by using luminosity to predict GDP (expressed in dollars) rather than presenting the results as a measure of radiance (nW-cm-2 .sr-1). To the best of our knowledge, it is the first time this has been done globally, for more than 40,000 microregions, and for 20 years.

We used low-resolution nighttime images with a spatial resolution of approximately one square kilometer from 2000 to 2013 and high-resolution satellite images with a spatial resolution of approximately 500 square meters from 2014 to 2020.[4]

Nighttime images were cleaned and processed to address "light blooming," which is when light from one microregion spills into a neighboring microregion, and to exclude sources of "noise" such as gas flares. We then calculated luminosity scores for every subnational microregion in the world.

National GDP was measured using the World Bank's GDP, PPP (2017 international dollars) data set, as explained above.[5] We modeled the relationship between luminosity and GDP using country-level data, for which we have access to both variables.

$$GDPct = f\left(luminosity_{ct}\right) + \omega_c + \delta_t + \sigma_{ct}$$

where national-level GDP from the World Bank (GDPct) is modeled as a nonlinear function f of the level of luminosity (luminosityct), and country and year dummies (ωc + δt), plus an error term σct.

Once this relationship was established, we estimated level 3 GDP using level 3 luminosity scores, coming up with estimates for each microregion (r):

$$\widehat{GDP_r} = \hat{f}\left(luminosity_r\right)$$

2 J. Vernon Henderson, Adam Storeygard, and David N. Weil, "Measuring economic growth from outer space," *American Economic Review*, April 2012, volume 102, number 2.

3 Xi Chen and William Nordhaus, "A test of the new VIIRS lights data set: Population and economic output in Africa," *Remote Sensing*, September 2015, volume 7, number 4; *Sustainable Cities*, "Tracking light from space: innovative ways to measure economic development," blog entry by Megha Mukim and Keith Garrett, World Bank, November 2013; Robert C. M. Beyer, Yingyao Hu, and Jiaxiong Yao, *Measuring quarterly economic growth from outer space*, International Monetary Fund, June 2022.

4 Earth Observation Group, Colorado School of Mines; DMSP OLS: Nighttime Lights Time Series Version 4, *Defense Meteorological Program Operational Linescan System (2000–2013)*; Earth Observation Group, Colorado School of Mines, VIIRS Nighttime Day/Night Band Composites Version 1 (2014–2020).

5 World Bank Data, GDP, PPP (constant 2017 international $), 2022.

Hyperparameters were tuned to minimize the root mean square error (RMSE) with official NUTS 3 estimates in Europe.

For countries where we lacked official level 2 or level 3 data, we trained a second model on the relationship between these GDP estimates and GDP estimates provided for the years 2009 to 2011 in a data set built by Kummu et al. That work provides granular GDP estimates for most microregions in the world derived from official subnational data sets and census data, as well as from a 2013 paper on regional development.[6]

This second model captured differences between our estimates and estimates from Kummu et al. We were able to correct our initial estimates using the output of the second model across years. This ensured that our estimates were brought into line with GDP estimates based on census data, as used in the Kummu et al. data set.

Estimating data gaps

As explained above, we obtained level 1 GDP data for 22 countries beyond EU countries and the United States. However, many of these 22 countries lack data for the full time series of our analysis.

We estimated missing years as follows:

1. We trained a model using the same process described above at the country level, but instead using level 1 GDP data and luminosity for years for which official or peer-reviewed GDP data exist.

2. We predicted GDP data for missing years using luminosity.

3. We benchmarked our GDP estimates to World Bank GDP data at the national level.

This produced a full time series of level 1 real GDP data from 2000 through 2019. We applied the same process in the cases where we had an incomplete level 2 or level 3 time series from official or peer-reviewed sources.

6 Nicola Gennaioli et al., “Human capital and regional development,” *Quarterly Journal of Economics*, June 2011, volume 128, issue 1.

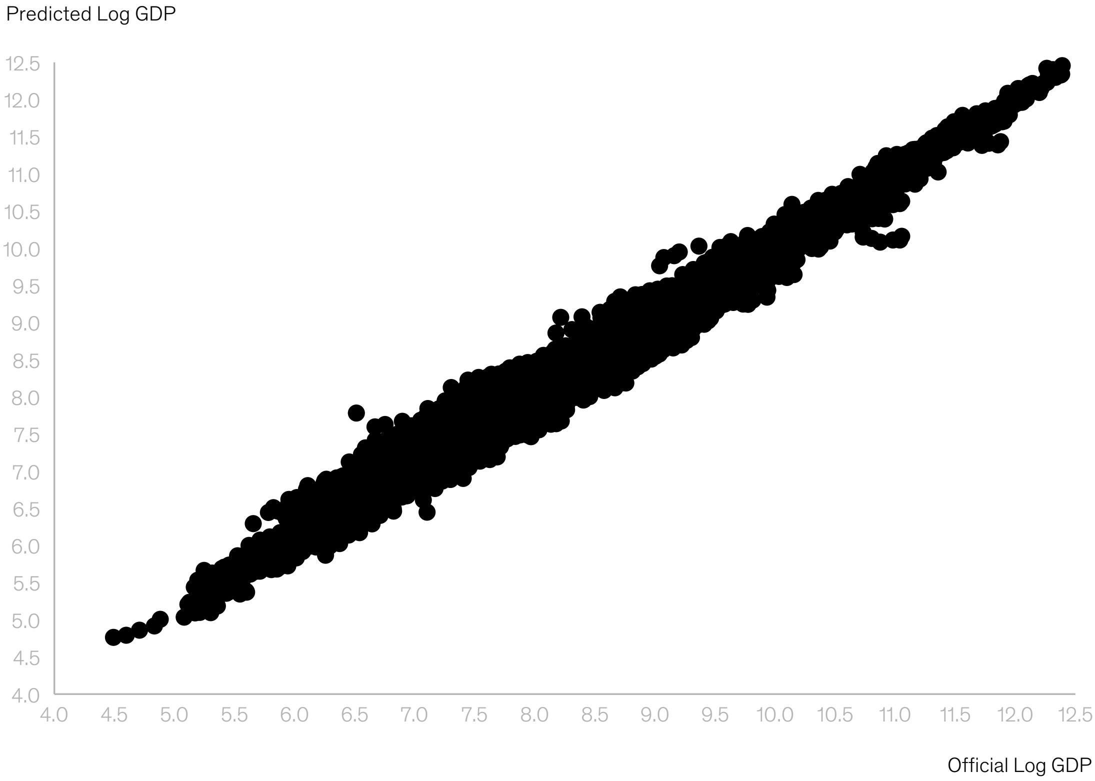
Assessing accuracy

We assess the accuracy of our models by estimating level 3 GDP from 2006 to 2020 for European microregions and comparing the output with official level 3 data from Eurostat.

The model has an almost one-to-one relationship to GDP (see Exhibit A3). The mean absolute percentage error for EU data at the NUTS 3 level is 4 percent.

Exhibit A3

Relationship between predicted NUTS 3 GDP and official NUTS 3 GDP data from Eurostat.



Source: McKinsey Global Institute analysis

Sensitivities and constraints

The data and methodologies used are subject to limitations and constraints, and results may be sensitive to certain assumptions. Here are some we identified:

- **Combining low- and high-resolution imagery.** We combined low- and high-resolution satellite images to overcome constraints in data availability. Thus, biases in satellite images may not be consistent across the full time series. However, we captured variations in satellites using year dummies in our model specification.
- **Discrepancies between national sources and World Bank data.** The GDP of some countries may not equal World Bank national GDP estimates. Certain countries have so-called extra-regio territories where economic activity cannot be linked directly to a specific microregion, such as national air space, embassies, and offshore natural resource locations. These extra-regio territories are excluded from our analyses.
- **Formal versus informal economy.** Luminosity-based approaches to estimating economic activity cannot differentiate between the informal and formal economy. Our data therefore may capture informal or illicit trade that is typically excluded from official statistics. This is not necessarily a limitation, but it can be a source of discrepancy with official sources.
- **Issues with the Kummu et al. data set.** As explained, our estimates were augmented to align with GDP estimates from Kummu et al., and any significant issues in that data set may be reflected in our GDP estimates. However, that paper has been peer-reviewed and is based on underlying data that also has been peer-reviewed. As expected, the quality of underlying data varies by country. Accepting that there will always be inaccuracies, our analysis uses only official GDP estimates or peer-reviewed data sets in order to reduce them.
- **Very small microregions.** Given the small size of some level 3 microregions such as Isla in Argentina, luminosity data could not be obtained from low-resolution satellite imagery from 2000 to 2013. We therefore used high-resolution data from 2014 to 2020, where available, in order to impute luminosity for 2000 to 2013. This was done using an autoregressive integrated moving average (ARIMA). Microregions for which 2014 to 2020 imagery was also unavailable are excluded. Additionally, smaller geographical units experience more volatility in GDP estimates for two reasons. First, economic activity is distributed over less surface area in these smaller units. For example, a small level 3 area where a major infrastructure project is under way will experience a significant increase in economic growth during the years in which that project is ongoing. That same project would have a much smaller impact on a level 2 and level 1 area because it would contribute relatively less to overall economic activity in the larger area. Second, luminosity estimates may be less reliable for smaller geographical units. This is because the average pixel size is approximately one square kilometer for the low-resolution images we used for 2000 to 2013 and approximately 500 square meters for the high-resolution images we used for 2014 to 2020.
- **Gas flares.** Some small microregions were excluded from our research due to the presence of gas flares, which obscure all pixels of night light data within the microregion.
- **Location in which economic activity is generated versus where people live.** When estimating GDP per capita, there can be slight mismatches between where economic activity is generated and where people live. An extreme example of this phenomenon is mines. A mine in a specific level 3 microregion produces intense light, capturing economic activity (GDP) generated there. But people rarely live next to a mine; they may live in another level 3 microregion nearby. In this case, the microregion with the mine will appear to have high GDP per capita—very high GDP and very low population—while the microregion next to it will have almost no GDP and a large population. This is a common problem in microregional analysis that does not have a large impact on average but needs to be taken into account when, for example, comparing two specific microregions.

Population

Population data are more easily available in formats that are incorporated into our data set. We get population data directly from external sources for all granularity levels. We do not estimate any data points; we simply match existing data to our microregional boundaries.

Data sources

Exhibit A4 provides the population sources used for every country.

Exhibit A4

Primary data sources for life expectancy analysis

COUNTRIES	YEARS AVAILABLE	DESCRIPTION	SOURCE
Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, North Macedonia, Norway, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom	2000–19	Population	Eurostat, Population by NUTS 3 region, 2021
United States	2000–19	Population	County Population Totals, US Census Bureau, 2021
All remaining countries	2000–19	Population	Unconstrained, top-down 1km population counts (UN Adjusted), Version 2, WorldPop, 2020

Data methodology

As can be seen in Exhibit A4, we use subnational data from Eurostat and the US Census Bureau for population data for EU countries and the United States.

For all other countries, we use WorldPop's one-kilometer-resolution population estimates.[7] WorldPop provides a range of population estimation methodologies. Our analysis incorporates its unconstrained, top-down methodology, which WorldPop has adjusted to align with national-level population estimates from the Population Division of the Department of Economic and Social Affairs at the United Nations.

Sensitivities and constraints

WorldPop notes two limitations of this data set:

1. It may misallocate population to uninhabited areas.
2. It may underestimate urban population in some areas.

In addition to these caveats, our population estimates may also be biased by the administrative boundaries used in our analysis. To extract population estimates from WorldPop data, we overlay polygonal administrative boundaries on WorldPop's grid-level estimates. We then count the population within the boundaries of each administrative unit. When the boundaries do not perfectly map onto official administrative boundaries, this leads to slight discrepancies in population estimates.

[7] WorldPop Hub, Population Counts, 2022.

Life expectancy

Here we provide the sources and methodology for life expectancy and explain in detail how we use child mortality data to estimate life expectancy. Our microregional life expectancy estimates are based on a combination of official life expectancy statistics when available, official mortality rates by age group (life tables approach), and microregional under-5 child mortality rates.

Data sources

Exhibit A5 provides the life expectancy sources we use for every country-year pair.

Exhibit A5

Primary data sources for life expectancy analysis

COUNTRIES	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
All countries, where there are no official data	2000–19	National life expectancy estimates	World population prospects, UNDP, 2019	
Australia, Russia	2000–19	Life expectancy at level 1 estimates	Regional well-being database: life expectancy, OECD, 2021	
Azerbaijan, Fiji, Georgia, Kazakhstan, Moldova	2000–19	Life expectancy at level 1 estimates	Smits and Permanyer, "The subnational human development data- base," *Scientific Data*, 2019, volume 6, number 1	
Belarus	2000–18	Life expectancy at level 1 estimates	Accounts of the Republic of Belarus, National Statistical Committee of the Republic of Belarus, 2019	Linear interpolation used to impute life expectancy for 2019
Bolivia	2012–19	Life expectancy at level 1 estimates	Regional well-being database: life expectancy, OECD, 2021	Linear interpolation used to impute life expectancy for 2000–11 and 2019
Brazil	2000–19	Life expectancy at level 1 estimates	Complete Life Tables, Brazilian Institute of Geography and Statistics, 2019	
Canada	2000–18	Life expectancy at level 1 estimates	Life expectancy and other elements of the complete life table, three-year estimates, Statistics Canada, 2022	Linear interpolation used to impute life expectancy for 2019
Chile	2000–16	Life expectancy at level 1 estimates	Regional well-being database: life expectancy, OECD, 2021	Linear interpolation used to impute life expectancy for 2017–19

Exhibit A5 continues

Exhibit A5 continued

COUNTRIES	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
China	2000–19	Life expectancy at level 1 estimates	Data sourced from provincial government statistical agencies	
Colombia	2000, 2005, 2010, 2015, 2020	Life expectancy at level 1 estimates	Censal conciliation 1985–2005 and forecast 2005–2020, National Administrative Department of Statistics, Colombia, 2021	Linear interpolation used to impute life expectancy for 2001–04, 2006–09, 2011–14, and 2016–19
India	2001, 2006, 2011, 2016, 2020	Life expectancy at level 1 estimates	Socioeconomical statistical information about health in India, Indiastat, 2021	Linear interpolation used to impute life expectancy for 2000, 2002–05, 2007–10, 2012–15, and 2017–19
Indonesia	2000–19	Life expectancy at level 1 estimates	Life expectancy by province and gender, Statistics Indonesia, 2020	
Japan	2000, 2005, 2010, 2015	Life expectancy at level 1 estimates	Regional well-being database: life expectancy, OECD, 2021	Linear interpolation used to impute life expectancy for 2001–04, 2006–09, 2011–14, and 2016–19
Malaysia	2000–19	Life expectancy at level 1 estimates	Abridged Life Tables, Department of Statistics Malaysia, 2020	
Mexico	2000–19	Life expectancy at level 1 estimates	Indicadores demográficos de la República Mexicana, Consejo Nacional de Población (CONAPO), 2021	
New Zealand	2001, 2006, 2013	Life expectancy at level 1 estimates	Regional well-being database: life expectancy, OECD, 2021	Linear interpolation used to impute life expectancy for 2000, 2002–05, 2007–12, and 2014–19
Peru	2003, 2008, 2013, 2018	Life expectancy at level 1 estimates	Regional well-being database: life expectancy, OECD, 2021	Linear interpolation used to impute life expectancy for 2000–02, 2004–07, 2009–12, 2014–17, and 2019
South Korea	2005, 2008, 2010–17	Life expectancy at level 1 estimates	Regional well-being database: life expectancy, OECD, 2021	Linear interpolation used to impute life expectancy for 2000–04, 2006–07, and 2017–19

Exhibit A5 continues

Exhibit A5 continued

COUNTRIES	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
Ukraine	2000–12	Life expectancy at level 1 estimates	Tables of birth, death, and average life expectancy, State Statistics Service of Ukraine, 2019	
United States	2000–19	Life expectancy at level 1 estimates	US Mortality Database, University of California, Berkeley, 2021	
Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, North Macedonia, Norway, Portugal, Slovenia, Slovakia, Spain, Sweden, Switzerland, United Kingdom	2000–19	Life expectancy at level 2 estimates	Life Expectancy at Birth by NUTS 2 Region, Eurostat, 2021	
Australia	2010–19	Life expectancy at level 2 estimates	Life tables statistical area level 4, Australian Bureau of Statistics, 2021	Linear interpolation used to impute data for 2000–10
Brazil	2000 and 2010	Life expectancy at level 2 estimates	Programa das Nações Unidas para o Desenvolvimento; Instituto de Pesquisa Econômica Aplicada; Fundação João Pinheiro, Brazil; Atlas de desenvolvimento humano, United Nations, 2013	For level 2 microregions missing from the UN's data set, life expectancy values spatially imputed from average of neighboring level 2 microregions; linear interpolation used to impute data for 2001–08 and 2011–19
Chile	2014	Life expectancy at level 2 estimates	Departamento de Estadísticas e Información de Salud, Anuario de Estadísticas Vitales, Ministry of Health, Chile, 2014	Level 3 life expectancy data for 2014 calculated using life tables approach

Exhibit A5 continues

Exhibit A5 continued

COUNTRIES	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
China	2000–19	Life expectancy at level 2 estimates	Numerous sources compiled by McKinsey Global Institute, for example: Improvement in people's health and increase in average life expectancy, National Bureau of Statistics, Shanxi Provincial People's Government, 2008; Hu Guangyu and Xie Xueqin, "Prediction and analysis of life expectancy per capita during the 'Twelfth Five-year Plan' period in Beijing," *Chinese Journal of Health Policy*, 2012, Volume 5, Number 4; 2009 Statistical Bulletin of Wuhan National Economic and Social Development, Wuhan Municipal Bureau of Statistics, Wuhan Government, 2014	For missing years, life expectancy was imputed by linear interpolation
Croatia, Netherlands	2001–19	Life expectancy at level 2 estimates	Life Expectancy at Birth by NUTS 2 Region, Eurostat, 2021	Linear interpolation used to impute data for 2000
Germany	2002–19	Life expectancy at level 2 estimates	Life Expectancy at Birth by NUTS 2 Region, Eurostat, 2021	Linear interpolation used to impute data for 2000–01
Japan	2010, 2015	Life expectancy at level 2 estimates	Life expectancy by cities, towns, and villages, Ministry of Health, Labor, and Welfare, Japan, 2016	Life expectancy for all other years imputed by linear interpolation

Exhibit A5 continues

Exhibit A5 continued

COUNTRIES	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
Mexico	2000, 2005, 2010, 2015, 2019	Life expectancy at level 2 estimates	Paredes and Silva, “Estimación de la esperanza de vida a nivel municipal y por marginación sociodemográfica: una aplicación del método de Swanson para el caso de México, 2010,” *Estudios Demográficos y Urbanos*, 2017, volume 32, number 1. Oficina de Investigación en Desarrollo Humano (OIDH), 2015; municipal Human Development Index (HDI), 1990–2015 (database), UNDP, Mexico	Life expectancy for all other years imputed by holding constant the difference between level 2 life expectancy and level 1 life expectancy in 2010 for all years
Montenegro	2005–19	Life expectancy at level 2 estimates	Life Expectancy at Birth by NUTS 2 Region, Eurostat, 2021	Linear interpolation used to impute data for 2000–04
Myanmar	2014	Life expectancy at level 2 estimates	Life expectancy at birth by state/region, and district, 2014 census, Myanmar	Life expectancy for all other years imputed by holding constant the difference between level 2 life expectancy and level 1 life expectancy in 2014 for all years
New Zealand	2006, 2013, 2018	Life expectancy at level 2 estimates	Stats NZ Tatauranga Aotearoa, 2021	Linear interpolation used to impute data for 2000–05, 2007–12, 2014–17, and 2019
Poland	2000, 2002–19	Life expectancy at level 2 estimates	Life Expectancy at Birth by NUTS 2 Region, Eurostat, 2021	Linear interpolation used to impute data for 2001
Romania	2004–19	Life expectancy at level 2 estimates	Life Expectancy at Birth by NUTS 2 Region, Eurostat, 2021	Linear interpolation used to impute data for 2000–03
Serbia	2017–19	Life expectancy at level 2 estimates	Life Expectancy at Birth by NUTS 2 Region, Eurostat, 2021	Linear interpolation used to impute data for 2000–16

Exhibit A5 continues

Exhibit A5 continued

COUNTRIES	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
United States	2000–10, 2014	Life expectancy at level 2 estimates	United States Life Expectancy and Age-specific Mortality Risk by County, Institute for Health Metrics and Evaluation, 1980–2014	Linear interpolation used to impute data for 2011–13 and 2015–19
Albania, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Türkiye, United Kingdom	2014–19	Level 3 data on population and deaths by age group	Population by age group, sex, and NUTS 3 region, and Deaths by age group, sex, and NUTS 3 region, Eurostat, 2021	Level 3 population and deaths by age group used to calculate life expectancy using life tables approach; data for 2000–13 interpolated by holding difference between NUTS 3 and NUTS 2 life expectancy estimates constant for that period
Serbia	2017–19	Level 3 data on population and deaths by age group	Population by age group, sex, and NUTS 3 region, and Deaths by age group, sex, and NUTS 3 region, Eurostat 2021	Level 3 population and deaths by age group used to calculate life expectancy using life tables approach

Exhibit A5 continues

Exhibit A5 continued

COUNTRIES	YEARS AVAILABLE	DESCRIPTION	SOURCE	NOTES
Afghanistan, Algeria, Angola, Burundi, Benin, Burkina Faso, Bangladesh, Belize, Bhutan, Bolivia, Botswana, Cambodia, Cameroon, Cape Verde, Central African Republic, Chad, Colombia, Comoros, Costa Rica, Cuba, Democratic Republic of the Congo, Djibouti, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Guatemala, Guyana, Haiti, Honduras, Indonesia, Iran, Iraq, Jamaica, Jordan, Kenya, Kyrgyzstan, Laos, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mongolia, Morocco, Mozambique, Myanmar, Namibia, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Republic of Congo, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, South Sudan, Sri Lanka, Sudan, Suriname, Swaziland, Syria, Tajikistan, Tanzania, Thailand, Timor-Leste, Togo, Trinidad and Tobago, Tunisia, Turkmenistan, Uganda, Uzbekistan, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe	2000–17	Under 5 child mortality at level 3 estimates	Low- and Middle-Income Country Neonatal, Infant, and Under-5 Mortality Geospatial Estimates 2000–17, Institute for Health Metrics, 2019	Data for 2018–19 imputed through linear interpolation

Data methodology

For life expectancy, we did not benchmark all of our granular estimates to one source, as we did with GDP. The unit of measure for life expectancy is years. We used official data directly from the source where available. For example, we use Eurostat for all EU countries across all levels of granularity. However, we benchmark estimated data to superior granularity levels of official data. For example, Colombian level 2 life expectancy, as shown in Exhibit A5, was estimated using under-five mortality data, but level 1 data was obtained from the national statistical office. In this case, we ensured that the population-weighted average of all level 2 regions added up to the country's level 1 official estimates. Finally, for countries where subnational official data were not available, we sourced country-level life expectancy from the UN Development Programme (UNDP) and benchmarked all granular estimates to it.

We used various methodologies to estimate life expectancy due to data availability constraints. The quantity and quality of available data at a subnational level vary widely. The following sections explain the methodologies used.

Estimating life expectancy at level 1

Level 1 life expectancy was either obtained directly from official sources or estimated by aggregating the next level up as the population-weighted average of level 2 life expectancy.

As Exhibit A5 shows, level 1 life expectancy in EU countries was obtained directly from Eurostat; in the United States from the United States Mortality Database housed at the University of California, Berkeley; and from the OECD for multiple OECD countries. Additionally, where national or peer-reviewed sources were available, we used them.

In instances where level 1 life expectancy data were available from both the OECD and national statistical agencies and where there were discrepancies between the two data sets, we used data from national statistical agencies.

Where level 1 official data existed and level 2 estimates did not aggregate perfectly to level 1 results, level 2 data were adjusted proportionally to match level 1. Where only level 2 estimates existed, level 1 was the direct aggregation of level 2 data, and level 1 and level 2 were adjusted proportionally around national life expectancy from UNDP.

For some countries, we had level 1 life expectancy estimates but lacked any sources at level 2 or 3, which prevented us from estimating life expectancy at those more granular levels. In those countries, the most granular unit is therefore official level 1 life expectancy data. The two large countries for which the most granular level of life expectancy data is level 1 are India and Russia.

When level 1 data did not cover the full time series of our 2000–19 analysis, we linearly interpolated the missing years in this way:

1. For each microregion at level 1, we calculated the difference between level 1 life expectancy data from official sources and national life expectancy data from UNDP.

2. For missing years, we linearly interpolated this difference.

3. We added national-level life expectancy data from UNDP for the missing years and the interpolated difference to establish a complete time series of level 1 life expectancy data.

Using level 1 life expectancy data from Global Data Labs

For five countries—Azerbaijan, Fiji, Georgia, Kazakhstan, and Moldova—we lacked subnational life expectancy data or any way to estimate them, so we instead used level 1 life expectancy data from Global Data Labs for the full time series from 2000 to 2019.

In some cases, our level 1 microregional breakdown was more granular than the level 1 boundaries from Global Data Labs. We determined which level 1 microregions fit within Global Data Labs' level 1 boundaries in order to assign life expectancy values. This sometimes resulted in the assignment of the same life expectancy value to two or more level 1 units.

We re-benchmarked this data to ensure that the population weighted average of level 1 microregions aligned with UNDP's official national-level life expectancy data.

Estimating life expectancy at levels 2 and 3 from Eurostat data

For 36 European countries, we obtained official level 2 data on life expectancy. When data were unavailable for the full time series of 2000 to 2019, we linearly interpolated life expectancy data using a similar methodology to the one described above, but at the next level of granularity:

1. For each level 2 microregion, we calculated the difference between life expectancy values at the level 2 and level 1.
2. For missing values, we linearly interpolated this difference.
3. For missing years, we used level 1 life expectancy data from Eurostat and added the interpolated difference to produce a complete time series of level 2 life expectancy data.

See Exhibit A5 for an exhaustive description of what years were interpolated and for which countries.

Additionally, for these same 36 European countries, we obtained level 3 data for mortality by age group and population by age group for 2014–19 from Eurostat. We then used a life tables approach to derive estimates of level 3 life expectancy.

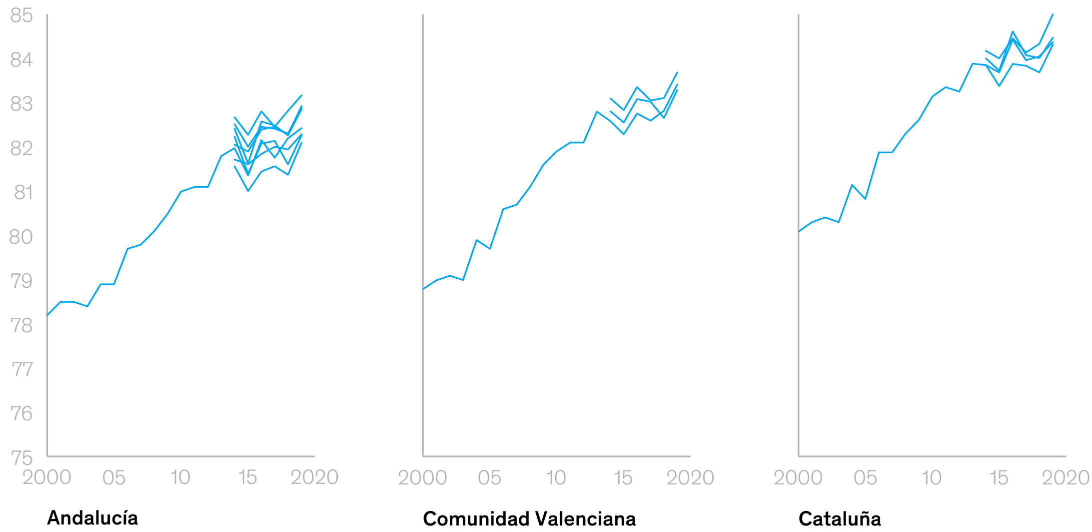
To obtain life expectancy data between 2000 and 2013, the years for which mortality and population by age group were not available, we used the three-step process described above.

Exhibit A6 and Exhibit A7 show three Spanish microregions with level 2 data as an example. In Figure 1, the line from 2000 to 2013 reflects the level 2 life expectancy data from Eurostat. The lines from 2014 to 2019 show the level 3 life expectancy estimates derived using the life tables approach for each level 2 microregion.

Exhibit A6

Level 2 time series with level 3 estimates of life expectancy for 2014–19, for three Spanish microregions.

Life expectancy, years

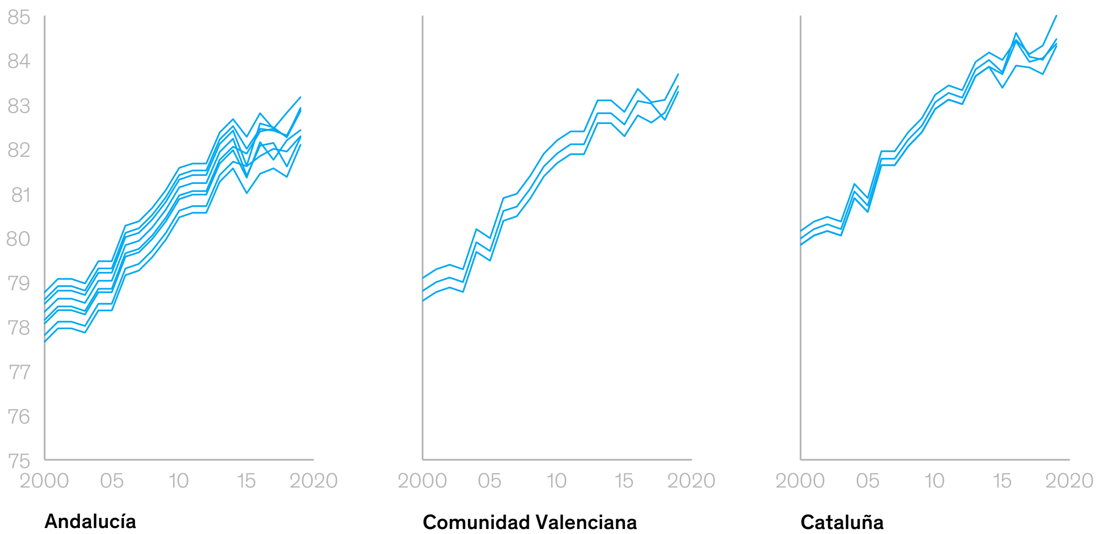


Source: McKinsey Global Institute analysis

Exhibit A7

Results of the linear interpolation methodology for the 2000–13 period.

Life expectancy, years



Source: McKinsey Global Institute analysis

Estimating level 2 and 3 life expectancy data using official data sets other than Eurostat

We obtained data for level 2 and level 3 microregions for six non-EU countries from the OECD and national statistical offices.

In cases when life expectancy data are separate for males and females, we used the average life expectancy of the two.

When life expectancy data were unavailable for the full time series 2000 to 2019, we imputed data for the missing years in this manner:

1. For each level 2 microregion, we calculated the difference in life expectancy between level 2 and level 1.

2. For missing years, we linearly interpolated this difference. If only one year of level 2 data were available, we held the difference constant. In other words, if we had only one year of data, we assumed the difference in life expectancy between level 2 and level 1 would be the same across the full time series.

3. We used level 1 level data from official sources for the missing years and added the interpolated difference to establish a complete time series of level 2 life expectancy data.

As explained above, we subsequently re-benchmarked this level 2 data to ensure that population weighted averages aligned with official level 1 life expectancy estimates when available.

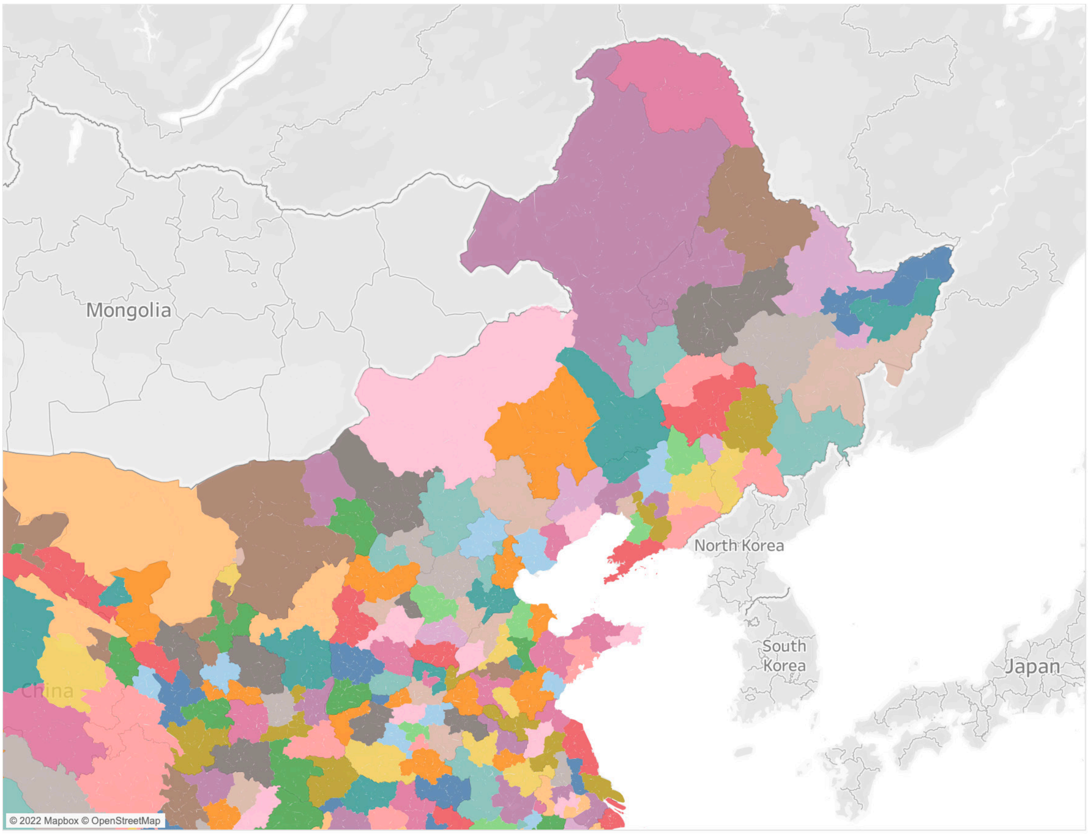
In the case of China, life expectancy data could not be obtained for all level 2 microregions.

1. Level 1 microregions such as Shanghai and Beijing have no level 2 life expectancy data. In such cases, we use only level 1 life expectancy estimates.

2. Some level 2 microregions within a level 1 microregion have available life expectancy data, while others do not. In cases when data were lacking, we collapsed level 2 microregions into one microregion, as shown in Exhibit A8 and Exhibit A9, ensuring that the population-weighted average between microregions with existing data and the consolidated ones matched the official level 1 life expectancy.

Exhibit A8

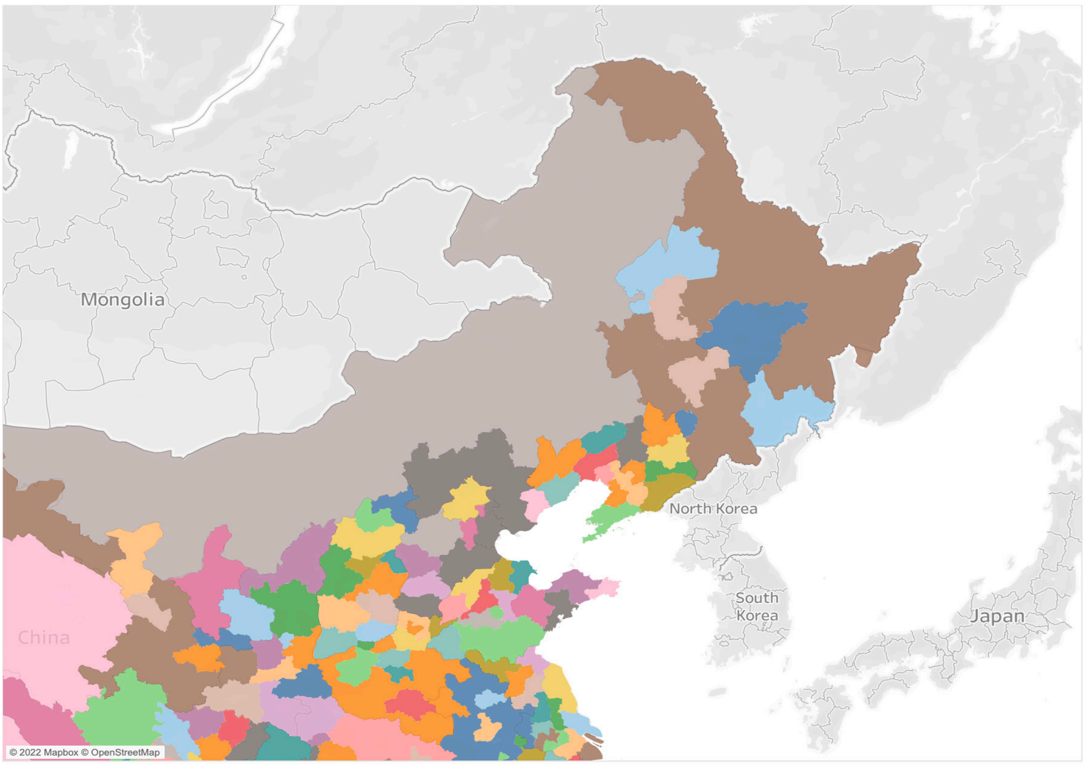
Map of Chinese microregions before collapsing microregions without level 2 life expectancy data.



The boundaries and names shown on this map do not imply official endorsement or acceptance by McKinsey & Company.

Exhibit A9

Map of Chinese microregions after collapsing microregions without level 2 life expectancy data.



The boundaries and names shown on this map do not imply official endorsement or acceptance by McKinsey & Company.

In Brazil, official estimates of life expectancy from the United Nations for 2000 and 2010 did not cover all level 2 microregions. For microregions missing values, we spatially imputed life expectancy values using the average life expectancy values of neighboring level 2 microregions. If a microregion missing level 2 data was an island and therefore had no neighboring microregions, it was assigned a null value. We subsequently obtained data for the full time series using the imputation process described above.

In Mexico, we applied a fixed effects regression model based on infant mortality rates to estimate level 2 life expectancy for 2000, 2005, 2010, 2015, and 2019, and interpolated the intervening years using the same imputation method as we used for the rest of the countries. We verified it against the Swanson method.[8]

Estimating level 2 and 3 life expectancy using child mortality rates

For 94 countries, we obtained granular geospatial data on the mortality of children under five years of age from the Institute of Health Metrics and Evaluation (IHME), as noted in Exhibit A5. This data set provided pixel-level estimates on child mortality, which we aggregated to the most granular administrative unit available for each country.

Given the strong relationship between child mortality and life expectancy, previous studies have used subnational child mortality to estimate life expectancy at birth.[9] We built on this approach by training a model on national-level life expectancy and national-level child mortality:

$$life\_exp_{cy} = child\_mortality_{cy} + child\_mortality_{cy}^2 + \Omega_c + \lambda_y + \varepsilon$$

where life_exp is national-level life expectancy data from the UNDP for country c and year y, and child_mortality is national-level child mortality data from IHME. Ωc and λy are country and year fixed effects respectively. ε is the error term.

If level 1 data were available, we used them instead of UNDP country-level data. Once the model was trained, we used level 3 data on child mortality to predict level 3 life expectancy at birth. We subsequently benchmarked the data to ensure that the population-weighted averages aligned with level 1 life expectancy values when available or with national-level life expectancy values when level 1 data was not available.

8 Israel Paredes and Eliud Silva, “Estimation of life expectancy at the municipal level and sociodemographic marginalization: Using the Swanson method for the case of Mexico, 2010,” *Estudios Demográficos y Urbanos*, January–April 2010, volume 32, number 1.

9 For more information, see Jeroen Smits and Iñaki Permanyer, “The Subnational Human Development Database,” *Scientific Data*, March 2019, volume 6, number 190038.

Sensitivities and constraints

We used a range of different methodologies and underlying data sets in creating a comprehensive subnational life expectancy data set for 2000 to 2019. The data and methodologies used have limitations and constraints, and results may be sensitive to certain assumptions:

- **Inconsistent biases across data sets and methodologies.** Mixing different methodologies means that biases in the data set may be inconsistent across countries. We have partially addressed these biases by ensuring that our data align with official life expectancy estimates either at level 1, when available, or at the national level.
- **Compression around the average.** When ensuring that life expectancy estimated values align with official data at higher levels of aggregation, the range of life expectancy values can become compressed. This means we may at times underestimate the variation in life expectancy within some level 1 microregions. A further consequence is that we sometimes overstate the effect of national or level 1 life expectancy by pulling all microregions around the average of higher levels of aggregation. For example, when two microregions in two different countries are benchmarked to their respective national averages, it is possible that we overstate the difference between the microregional true life expectancies if the two national numbers vary significantly.
- **Underlying data quality.** The quality of underlying data varies by country. To address data quality issues, our analyses used only official estimates of life expectancy or peer-reviewed data sets, but even in those cases we cannot guarantee fully consistent quality.
- **Imputation methods.** Imputation methods may add noise to the data set. We have undertaken imputation methods to build a complete time series of subnational life expectancy data from 2000 to 2019. The quality of our imputations depends on the completeness of the original time series. Microregions where we imputed one year are more accurate than microregions where we imputed several years. Moreover, many different imputation approaches could be used. We adopted the simplest — linear imputation techniques. We tested many different imputation techniques, including multiple imputation by chained equations, spline, and Kalman imputation, but the results were no more accurate than linear imputation.

Chris Bradley is a director of the McKinsey Global Institute and a senior partner of McKinsey & Company in Sydney; **Sven Smit** is director and chairman of MGI and a senior partner in Amsterdam; **Jonathan Woetzel** is a director of MGI and a senior partner, and **Marc Canal** is an MGI engagement manager in Barcelona.