WIDER Working Paper 2023/121

An open database on inequality and polarization in length of life (1950–2021)

Vanesa Jorda,[1] Miguel Niño-Zarazúa,[2] and Mercedes Tejería[1]

October 2023

Abstract: Monitoring health is key for identifying priorities in public health planning and improving healthcare services. Life expectancy has conventionally been regarded as a valuable indicator to compare the health status of different populations. However, this measure is simply the mean of the distribution of the length of life and, as such, neglects individual disparities in health outcomes. Tracking the evolution of lifespan inequality is essential to promoting social justice and equity by identifying and addressing the root causes of these disparities. In this paper, we use life tables from the UN World Population Prospects to develop the most comprehensive database of lifespan inequality and polarization for 258 countries and areas for the period 1950–2021. To help with the visualization and the use of these estimates, we have designed a user-friendly website that enables users to compute lifespan inequality for any group of countries of their choice. These extensive series on the distribution of lifespans provide access to crucial information for researchers, policy-makers, and the general public, thus representing a fundamental step towards a better understanding of health differences within and between nations.

Key words: inequality, polarization, lifespan, life expectancy

JEL classification: I14

Acknowledgements: This study was supported by the Novo Nordisk Foundation Grant NNF19SA0060072.

---

[1] Universidad de Cantabria, Santander, Spain; [2] SOAS University of London, London, UK, and UNU-WIDER, Helsinki, Finland; corresponding author: jordav@unican.es

This study has been prepared within the project The impacts of inequality on growth, human development, and governance—@EQUAL, supported by the Novo Nordisk Foundation Grant NNF19SA0060072.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-429-8

https://doi.org/10.35188/UNU-WIDER/2023/429

Typescript prepared by Gary Smith.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland and Sweden, as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Background and summary

Life expectancy is a commonly used health indicator that measures the average number of years a person is expected to live, based on various demographic factors including age, gender, and country of residence. This indicator is seen as a valuable tool for assessing the overall health and well-being of a population (Gulis 2000). Life expectancy is used to monitor trends in health outcomes over time and to compare the health status of different populations (Fries 2002; Wang et al. 2017). For instance, countries with higher life expectancies generally have better access to healthcare and sanitary conditions.

Despite its popularity, life expectancy faces several limitations as an indicator of health. It does not take into account differences in the quality of life or health status among individuals within a population (Wolfson 1996). In this sense, a person may live a long life but experience poor health, disability, or limited mobility. Moreover, life expectancy can be influenced by factors such as socioeconomic status, lifestyle choices, and environmental factors, which can vary widely across individuals and cause differences in health outcomes. However, this life expectancy does not capture health inequality (Jordá and Niño-Zarazúa 2017; Vaupel et al. 2011).

To address this last limitation, lifespan inequality might be an appealing summary measure of differences in mortality rates across age groups. Tracking the evolution of lifespan inequality is key because it can be used to promote social justice and equity by identifying and addressing the root causes of these disparities. Measuring length of life inequality can also be used to compare health outcomes for different countries. This information helps to identify areas where countries can learn from each other and work together to improve global health outcomes. This is of particular importance in driving global action towards achieving Sustainable Development Goal 3, which aims at ensuring healthy lives at all ages.

Closely related to the concept of inequality, polarization measures capture how concentrated the distribution is into different poles (Duclos and Taptué 2015). The analysis of this phenomenon is of particular interest to analyse the distribution of lifespans. The shape of the distribution of length of life is typically characterized by two peaks due to the two underlying phenomena driving mortality patterns (see Edwards 2011; Edwards and Tuljapurkar 2005; Peltzman 2009). The combination of high infant mortality rates and high adult mortality rates in low-income countries leads to a polarized distribution of lifespans, with a high number of deaths occurring at very young ages and at older ages. The probability of dying decreases steadily until the age of 15, when the distribution becomes bell-shaped, representing adult mortality patterns. Public health officials can benefit from analysing the evolution of polarization in length of life as they can identify areas (infant or adult mortality) where health interventions are needed the most.

The main contribution of this paper resides in the development of a comprehensive set of estimates of length of life inequality and polarization for 258 countries/areas for the period 1950–2021. The full set of estimates is available at http://vanesajorda.com/dir/database_lifespan. We pay particular attention to additively decomposable inequality measures for practical purposes. Indeed, the use of this kind of measures makes it possible to compute lifespan inequality for any group of countries using only the information included in our database. We disaggregate inequality patterns by sex due to the fact that lifespan distributions have been fairly distinct for men and women. Prior research often looks at adult and infant mortality separately because the underlining factors which determine these two phenomena are aetiologically different (Edwards 2011; Edwards and Tuljapurkar 2005; Smits and Monden 2009).

Therefore, we include estimates of different inequality measures for both total population and the population aged over 15.[1]

A newly developed database of different statistics of lifespan distribution is a significant achievement in the field of demography and public health. This database has been created to measure the differences in lifespan across different populations and countries, providing valuable insights into the distribution of health and mortality between and within countries. Additionally, a user-friendly website has been designed to enable users to compute lifespan inequality for any group of countries of their choice, providing access to crucial information for researchers, policy-makers, and the general public. This database and website will help to inform policy decisions and resource allocation by highlighting the disparities in health outcomes within populations, and will aid in identifying areas where intervention is needed to address health inequalities. Therefore, these tools have the potential to contribute significantly to our understanding of global health disparities and to inform efforts to improve health equity.

2 Methods

2.1 Conceptual framework

As our database provides information about inequality and polarization in lifespans, an illustration of how to interpret these figures is fundamental. In this section we explain how these measures can be computed and how they relate to life expectancy, with a particular emphasis on the different phenomena that these two concepts measure. Life expectancy, inequality, and polarization in length of life are derived from period life tables. Abridged period life tables provide information on death rates for $K = 24$ age intervals: 0–1, 1–4, 5–9, ..., 95–99, 100+. These figures can be seen as points of the probability density function (pdf) of the distribution of length of life, which gives the probability of dying in each interval.[2] Life expectancy at birth, defined as the number of years that an individual born today might live if the current mortality patterns remained, is the mean of this distribution:

$$LE_i = \sum_{k=1}^{K} d_k * c_k$$

where d_k is the probability of dying at any age within the interval $l_k = [x_{k-1}, x_k]$ and $c_k = (x_{k-1} + x_k)/2$ is the central point of that interval.

Life expectancy is a powerful tool to track the evolution of health conditions in a given country, because this measure summarizes mortality risks at different ages into a single number. However, this measure fails to capture the differences in the length of life (or the age at death) between individuals. Even if a country shows progress on average, this does not mean that every member of the society is improving her health status. Moreover, two countries with the same life expectancy might exhibit quite different distributional patterns of length of life.

Inequality measures inform about the extent of the differences in length of life between the citizens of a country. Inequality indices can be classified into absolute or relative measures. To illustrate the

[1] The distribution of length of life typically presents two poles: one depicted by infant mortality patterns and a second one for the adult population (see Figure 1). Restricting the sample to the population aged over 15 means that the resulting distribution is bell-shaped with only one mode, so polarization and inequality will both measure the same phenomenon. Therefore, polarization measures are provided for the total population only.

[2] Since no information on the distribution of deaths within the age intervals is available, we assume that the population dies uniformly throughout each interval. Despite being a conventional assumption, this approach would underestimate inequality levels because differences between individuals that belong to the same age intervals are not considered. This potential limitation might not affect our results substantially because the data are provided for 24 intervals.

distinction between these two types of measurements, consider the following example. Suppose we want to assess inequality between two individuals in two different countries. In Country A, one person lived for 6 years while the other lived for 60 years. In Country B, one individual lived for 7 years and the other lived for 70 years. Relative inequality measures would indicate equal levels of inequality in both countries because the distribution of lifespan in Country B can be derived from Country A by increasing the age at death by 17 per cent for both individuals. As a result, the relative difference between the two individuals in these two countries remains the same, at 1/10. On the other hand, absolute measures would deem Country B as more unequal, as the absolute difference in lifespan between the two citizens is 63 years, whereas in Country A it is 54 years. Hence, the choice between absolute and relative inequality measures is not impartial and can impact not only the magnitudes but also the trends of health inequality.

There is an ongoing debate regarding the more suitable approach for evaluating health inequalities (Anand et al. 2001). Relative indicators have gained appeal when examining income variables (Niño-Zarazúa et al. 2017), whereas for variables with boundaries, such as lifespan, absolute changes are considered a better alternative for measuring health inequality (Atkinson 2013). As a result, demographers often favour absolute measures, with variance and standard deviation being the most commonly employed ones (Edwards 2011; Edwards and Tuljapurkar 2005). In this sense, prior research often relies on the variance to evaluate the levels of health inequality. For data grouped into age intervals, as presented in life tables, the variance is given by

$$\sigma^2 = \sum_{k=1}^{K} (c_k - \mu)^2 d_k$$

where μ is life expectancy at birth.

Among the relative measures, the popularity of the Gini index has spread to health variables (Edwards 2011; Smits and Monden 2009; Vladimir Shkolnikov et al. 2003). Consequently, we also compute estimates for this inequality measure, which can be calculated as follows:

$$G = \frac{1}{2\mu} \sum_{k=1}^{K} \sum_{j=1}^{K} |c_k - c_j| d_k d_j$$

The absolute version of this indicator is given by:

$$AG = \mu * G.$$

Despite the widespread acceptance of the Gini index, this measure has certain limitations. This index is sensitive to the middle of the distribution, but it does not allow for varying the weight assigned to differences in specific segments of the distribution. Inequality measures can capture diverse trends, depending on their sensitivity to different parts of the distribution. To address this concern, we calculate an alternative set of inequality measures from the generalized entropy (GE) family. This family includes a sensitivity parameter (θ) that enables us to adjust the importance attributed to disparities in lifespan among older individuals. The mean log deviation (MLD) corresponds to the GE index when the parameter is set to 0, which is more sensitive to infant mortality. The Theil's entropy measure, characterized by a parameter value equal to 1, is equally sensitive to all segments of the distribution. We also compute the GE measure with θ set to 2 to examine the evolution of lifespan inequality with greater emphasis on differences among the elderly. The GE class of inequality measures is given by,

$$GE^{(\theta)} = \frac{1}{\theta(\theta-1)} \left(\sum_{k=1}^{K} d_k \left(\frac{c_k}{\mu} \right)^{\theta} - 1 \right), \theta \neq 0, 1$$

where, for $\theta = 1$, we have the following limiting case:

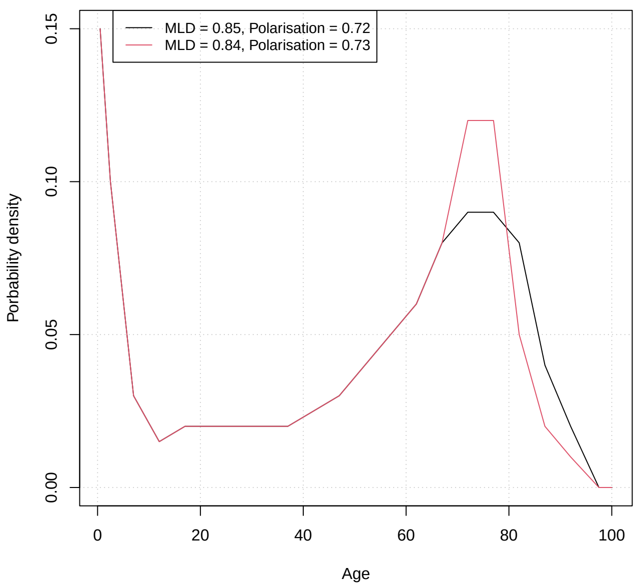
$$T = \sum_{k=1}^{K} d_k \frac{c_k}{\mu} \log\left(\frac{c_k}{\mu}\right)$$

and for $\theta = 0$, the index tends to

$$MLD = \sum_{k=1}^{K} d_k \log\left(\frac{\mu}{c_k}\right).$$

While inequality measures can provide us with valuable insights about the level and evolution of differences in length of life, some aspects of the distribution can only be analysed with polarization indices. Although closely related, inequality and polarization have distinct theoretical foundations and depict different trends over time. To illustrate the fundamental difference between these two concepts, Figure 1 presents two lifespan distributions, identical up to the age of 65. These distributions show a first spike at the lower end, which depicts infant mortality patterns. The probability of dying decreases steadily until the age of 15, when the distribution becomes bell-shaped, with a second mode at 75 years. While both distributions share similar characteristics, the red one is more concentrated around this second mode. As a result, lifespan inequality is lower under this scenario: the MLD is 0.85 for the black distribution and 0.84 for the red one.

Figure 1: Length of life distributions



Source: authors' compilation.

Polarization, instead, is higher for the red distribution. This is because the concept of polarization relies on the alienation–identification framework developed by Esteban and Ray (1994). This framework, originally developed for income variables, classifies individuals into groups (or poles). Individuals feel that they identify with other individuals in their group, while feeling alienated from those belonging to other groups. Polarization is a measure that summarizes the distance between these poles (alienation) and their representativity (identification). In the same vein as inequality, as the distance between these groups increases, the distribution becomes more spread and individuals feel more alienated from each other, also fostering polarization. By contrast, as individuals within a group become more similar to each other, particularly in terms of lifespans in the context of this study, their identification with each other grows stronger, which increases polarization. Identification is measured by the concentration of the probability mass around a pole. For lifespan distributions, we identify one pole that includes the youngest individuals and a second group of adult population. This second pole is less spread in the red distribution, so the pole is more prominent, thus leading to higher polarization levels.

The identification–alienation framework described above can be implemented using the polarization measure proposed by Duclos et al. (2004):

$$DER(\alpha) = \sum_{k=1}^{K} \sum_{j=1}^{K} |c_k - c_j| d_k^{1+\alpha} d_j \tag{1}$$

where $\alpha \in [0.25, 1]$ is the so-called 'polarization aversion' parameter, which captures the power of the identification effect . The greater the value of α, the greater the distinction between polarization and inequality. Indeed, when $\alpha = 0$, this measure is equivalent to the absolute Gini index. Although the DER index measures absolute polarization, it can be easily transformed into a relative index by multiplying Equation (1) by $LE^{\alpha-1}$, where LE is the life expectancy.

2.2 Data collection

To evaluate the level of inequality in length of life of a particular country in a given year, we need information about the death rate broken down by age and sex. Period life tables contain data on the number of deaths for every five-year age group up to 10 years for a synthetic cohort of 100,000 individuals. For a certain year, period life tables are constructed using data on the number of deaths in that particular year. Hence, mortality rates do not refer to the actual mortality patterns of a real birth cohort over its lifetime, but to the current mortality patterns of a country. Due to the large decreases in mortality rates over time—as a result of, for instance, medical advances—a person born today does not face the same probability of dying as a person born in 1900. Since we are interested in providing a snapshot of lifespan inequality trends, we use period life tables to perform the analysis as they provide an indication of the mortality situation at a particular point in time.

The data have been retrieved from World Population Prospects: The 2022 Revision, developed by the UN Population Division (UN-DESA 2022a). Among all the available sources, this database is the most appealing because of its geographical and temporal coverage. Detailed demographic estimates and projections on fertility, mortality, and migration are provided for every member state of the UN from 1950 onward. Hence, this data source provides a balanced panel of 237 countries/areas from 1950 to 2021 on an annual basis.[3] Despite the wide geographic and temporal coverage of the data, the quality and consistency of mortality series vary substantially across countries. Estimates of mortality rates are derived directly from registered deaths when civil registration data of good quality is available. Although most countries presented vital registration coverage greater than 90 per cent in recent years,[4] the period before 1990 has been plagued by the lack of empirical data. When official demographic statistics are not reported in the detail necessary for the preparation of cohort population projections, the UN Population Division undertakes its estimation by using data from different sources, including major surveys such as the Demographic and Health Surveys or the Multiple-Indicator Cluster Surveys, population censuses, and analytical reports.

[3] World Population Prospects provides country-level data for those countries and areas with at least 90,000 inhabitants in 2022.

[4] For a summary of vital statistics coverage of the last update of World Population Prospects, see the file *quality_data*, available at http://vanesajorda.com/dir/database_lifespan.

3 Data records

The dataset contains six files with information on national and regional estimates of inequality and polarization in length of life for 258 countries/areas for each year from 1950 to 2021. The datafile *APR2023_both_total* includes the estimates for the whole population, whereas *APR2023_both_15* gathers the data for the population aged over 15. Female estimates are stored in *APR2023_female_total* and *APR2023_female_15* for all women and those aged over 15, respectively. Finally, *APR2023_male_total* includes the estimates for the male population, and *APR2023_male_15* only for men aged over-15.

The file code $quality_data$ gathers information on the type of data used to build our estimates for each country and year, retrieved from past editions of the Demographic Year Book. National estimates are classified into three broad categories: 'C' is used when civil data are expected to be virtually complete, representing at least 90 per cent of deaths; 'U' is for countries where vital information is expected to cover less than 90 per cent of deaths; and 'I' indicates that the data are not gathered from civil registration systems but are considered reliable. Countries classified within this last category present demographic information derived from projections, population and housing censuses, or other estimation techniques. Finally, 'N' is used for countries with no information about the completeness of the data. In some cases, deaths are tabulated by date of registration instead of by date of occurrence, indicated by the symbol '*'. This might affect mortality estimates if the lag between the date of occurrence and the date of registration is considerable, which would lead to a substantial delay in death registrations. However, delays in the registration of deaths are less common and shorter than those in the registration of live births.

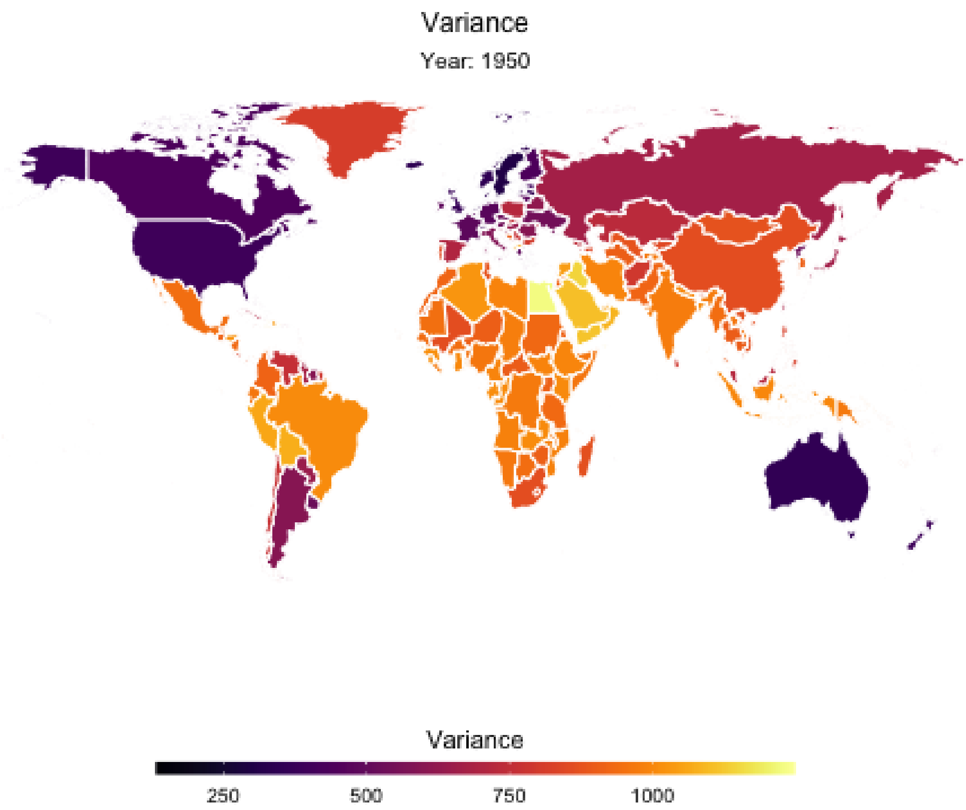
The descriptions of the variables of these data files are presented in Table 1. In addition, these new estimates on polarization and inequality in length of life can be explored using a user-friendly visualization tool available at https://vanesajorda.shinyapps.io/App-lifespans/. This tool has been designed to enable users to view, plot, and download national and regional estimates and lifespan inequalities for any group of countries. An example of the kind of graphs available at the website is presented in Figure 2. This map shows the evolution of the variance as an indicator of absolute inequality in length of life from 1950 to 2021.

Table 1: Variable description

Record	Description
country	Country name
year	Year
desa_id	Standard country or area codes for statistical use of the United Nations Statistics Division
country_year_id	Country–year identification code
ISO2	ISO country code two-digit
ISO3	ISO country code three-digit
area_type_name	The type of area: country, world
SDG_reg_name	Geographic regions defined by the United Nations Statistics Division: Australia/New Zealand, Central and Southern Asia, Eastern and South-Eastern Asia, Europe and Northern America, Latin America and the Caribbean, Northern Africa and Western Asia, Oceania (excluding Australia and New Zealand), Sub-Saharan Africa
WB_inc_name	Income groups defined by the World Bank: high-income countries, low-income countries, lower-middle-income countries, and upper-middle-income countries
WB_reg_name	Geographic regions defined by the World Bank: East Asia and the Pacific, Europe and Central Asia, Latin America and the Caribbean, Middle East and North Africa, North America, South Asia, Sub-Saharan Africa
le	Life expectancy
theil	Theil index
mld	Mean log deviation
ge2	Generalized entropy measure ($\alpha = 2$)
variance	Variance
gini	Gini index
abs_gini	Absolute Gini index
ader_25	Duclos et al. (2004) absolute polarization measure ($\alpha = 0.25$)
ader_50	Duclos et al. (2004) absolute polarization measure ($\alpha = 0.5$)
ader_75	Duclos et al. (2004) absolute polarization measure ($\alpha = 0.75$)
ader_1	Duclos et al. (2004) absolute polarization measure ($\alpha = 1$)
rder_25	Duclos et al. (2004) relative polarization measure ($\alpha = 0.25$)
rder_50	Duclos et al. (2004) relative polarization measure ($\alpha = 0.5$)
rder_75	Duclos et al. (2004) relative polarization measure ($\alpha = 0.75$)
rder_1	Duclos et al. (2004) relative polarization measure ($\alpha = 1$)
adult_pop	Adult population
total_pop	Total population
source	Coverage and type of data used to construct the estimates

Source: authors' compilation.

Figure 2: Global evolution of absolute inequality in length of life: variance

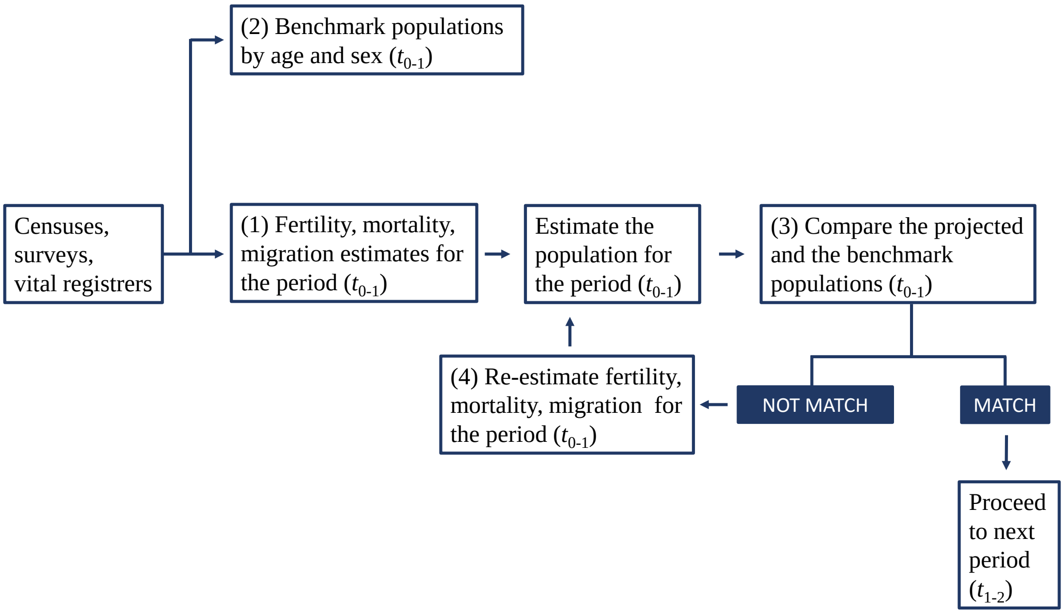


Source: authors' compilation.

4 Technical validation

The reliability of our estimates crucially depends on the quality of the data sources used for the construction of the period life tables. UN-DESA relies on both civil and survey data to construct mortality estimates. The precision of these estimates is evaluated to identify the most likely series of mortality. The core methodology for the estimation of population estimates from the 2022 Revision of World Population Prospects is the cohort-component method for projecting population (UN-DESA 2022b). This approach for producing population estimates and projections involves several steps (Figure 3).

Figure 3: Procedure implemented in the estimation of UN World Population Prospects to ensure consistency between demographic components and population projections



Source: adapted from UN-DESA (2022b).

1. The first step is to estimate the components of demographic change (fertility, mortality, and migration) using data from various sources, including censuses, surveys, vital and population registers, and analytical reports. Estimates from different sources might produce substantially different figures, which are consolidated through expert-based opinion or using automated statistical methods that generate a smooth trend.[5]

2. The second step is to estimate benchmark populations by age and sex, using population counts from censuses and other sources. These counts are adjusted as necessary, and supplemental benchmarks are considered for specific age groups.

3. In the third step, the population for period $t+1$ is computed from the population at period t, considering the births minus the deaths registered in that period and the net international migration. These figures are then compared with the benchmark population estimates from step 2 to assess their internal consistency. Another important aspect of the work at this stage is ensuring that in-

[5] See https://esa.un.org/unpd/wpp/DataSources/ for a detailed description of the sources and UN-DESA (2022b) for the different procedures applied to construct the estimates of adult and child mortality.

formation on the net number of international migrants is consistent and sums to zero at the world level.

4. The fourth step involves re-estimating the demographic components as necessary to achieve consistency between the estimated and the benchmark populations. A final consistency check is performed to ensure the geographical coverage and the demographic plausibility of these estimates are evaluated and, if necessary, adjusted to be congruent over time and across age groups.

This procedure has been deployed in each revision of the World Population Prospects. When new information becomes available, previous estimates are re-evaluated, which means that with every new revision demographic trends may be adjusted.

Although the methodological framework described above requires consistency with the estimated trends of fertility, mortality, and net international migration, our focus resides in mortality series as these data are the primary source to produce our estimates. The techniques employed to estimate mortality rates and life tables differ among countries, depending on the nature and reliability of the available data sources. These techniques can be broadly classified into two main types: *empirical* and *model-based*. The empirical method was deployed for those countries with accurate data that enabled the construction of sex- and age-specific mortality rates through vital registration or other reliable estimates across a large number of years for the period 1950–2021.

For those countries that lacked comprehensive information to use the empirical method, the model-based approach is deployed. A minimum of one parameter is required describing the mortality rate among children or overall (such as the life expectancy at birth). An extra parameter that accounts for adult mortality can be beneficial to select the most appropriate model to accurately represent the mortality age pattern for a specific country and year. To provide the necessary information to reconstruct period life tables, annual time series of complete child and adult mortality rates are estimated for each country from 1950 to 2021.

The core approach for estimating mortality is not appropriate for countries with a high prevalence of HIV and AIDS. As these diseases occur primarily among adults of reproductive age, they modify the conventional U-shaped age profile of mortality. In these countries, the method applied by World Population Prospects 2022 applies adjustments to model-based estimates because model life tables are not able to represent this atypical mortality pattern by age. A similar approach is applied to account for the excess of mortality due to conflicts, natural disasters, or epidemics.

Since the extent and the quality of the data used to construct mortality series is crucial to assess the accuracy of the estimates on lifespan inequality and polarization, our set of estimates includes information about the source of the data used by the UN World Population Prospects. This information is available in the file *quality_data* for each country/territory on an annual basis. A summary of this information is presented in Table 2, which presents the proportion of countries/territories whose mortality series are constructed from reliable estimates or civil data that can be complete, incomplete, or unknown for infant and adult mortality.

There has been a significant improvement in data coverage over time. In 1950, a substantial portion of life tables relied on data whose origin and quality were unknown (63.98 per cent). However, by 2021 this fell to 30.51 per cent, reflecting the wider availability and the improved reliability of data sources. The proportion of countries with civil data that cover more than 90 per cent of the adult and infant populations (labelled as *both complete*) also increased over the years, which reflects more comprehensive coverage of life table data.

Table 2: Data coverage of the sources used to construct period life tables

	Both complete	Both incomplete	Both unknown	Reliable estimates	Complete/ incomplete	Complete/ unknown	Incomplete/ unknown
1950	19.92	6.36	63.98	0	0.42	1.69	7.63
1970	40.25	17.37	34.32	0.42	1.69	1.69	4.24
1990	44.07	7.63	40.25	0	1.69	1.69	4.66
2010	52.97	11.86	23.31	5.08	0.42	2.54	3.81
2021	50	10.59	27.12	2.12	0.85	5.08	4.24
EAP	34.58	16.7	43.17	0.69	0.15	1.24	3.47
ECA	84.48	0.8	13.38	0.83	0.15	0.37	0
LAC	48.88	26.09	18.12	0.3	1.03	3.8	1.79
MENA	21.04	25.07	35.56	0.76	1.25	1.81	14.51
NA	94.91	0	0.46	0	2.78	1.85	0
SA	13.19	16.84	60.07	5.21	0	3.47	1.22
SSA	10.19	6.66	74.97	1.56	1.01	0.64	4.98

Note: World Bank regions: EAP, East Asia and the Pacific; ECA, Europe and Central Asia; LAC, Latin America and the Caribbean; MENA, Middle East and North Africa; NA, North America; SA, South Asia; SSA, Sub-Saharan Africa.

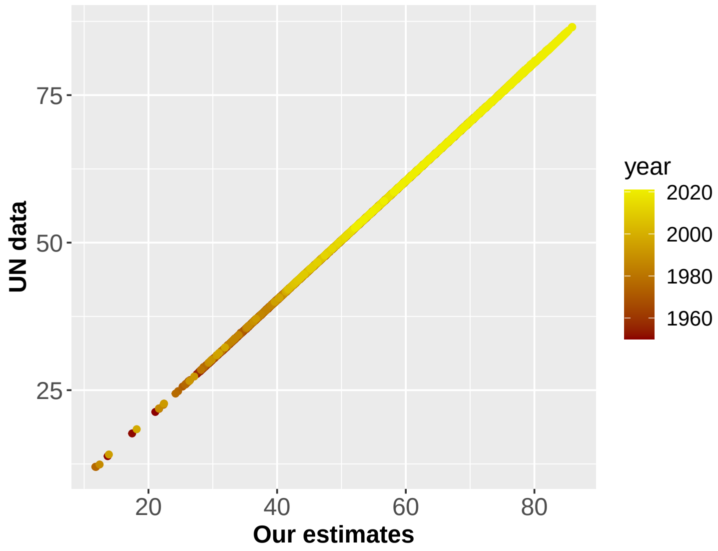
Source: authors' calculations from Demographic Year Books 1950–2021.

Despite the advances in data collection methodologies and the efforts made to enhance the quality of demographic data, Table 2 highlights regional variations in data coverage. For instance, North America and Europe and Central Asia exhibit the highest percentage of data classified as 'both complete', at 85 and 94 per cent, respectively. These figures indicate that these regions include countries with robust data collection systems. By contrast, only 10 per cent of South Asian estimates are based on complete civil data. This points towards a need for further efforts to improve data collection and quality in this region. Similarly, in the Middle East and North Africa a large proportion of life tables are constructed using incomplete or unknown data, which reveals important challenges in obtaining complete and reliable information.

Even though some data issues remain and the procedures implemented to obtain the estimates rely, in most cases, on statistical assumptions, the UN Population Division has managed to produce internally consistent and plausible estimates on fertility, mortality, and migration. Life expectancy and other life statistics provided by the World Health Organization are computed from the data provided by the UN World Population Prospects. Hence, even though we should be cautious using these series, we should also acknowledge that these figures are widely accepted and used to make cross-country comparisons, as these data satisfy the highest quality standards.

As a final exercise to validate the reliability of our estimates, we compare official data on life expectancy released by World Population Prospects 2022 with our estimates. As both estimates of life expectancy are constructed from UN life tables, we should get similar life expectancy series to those officially reported by the UN. Figure 4 plots our life expectancy estimates against UN figures, revealing a straight line, which indicates a strong relationship between them. In fact, the correlation coefficient is 0.999998, which confirms the existence of an almost perfect positive correlation between our estimates and the UN series.

Figure 4: Life expectancy estimates: UN data



Source: authors' calculations.

5 Usage notes

The database on lifespan inequality and polarization contributes to analysing health trends, understanding disease burdens, and determining priorities for public health planning. This dataset provides empirical information essential to policy-makers and public health officials to make informed decisions and to design appropriate public health actions. This information helps in identifying healthcare needs and determining priorities for public health planning. Hence, these data are helpful for allocation of resources and the development of targeted interventions to address health issues (Nsubuga et al. 2006).

Monitoring health inequality might also reveal differences in health outcomes within the population. This information is crucial for targeting interventions and policies aimed at reducing health disparities and promoting health equity. Moreover, these data are essential for the evaluation of the effectiveness and impact of health policies, interventions, and programmes. Tracking health indicators over time allows for evidence-based decision-making, and helps policy-makers to identify successful strategies and areas that require improvement (Bleich et al. 2012).

As our database covers virtually all countries, it enables international comparisons, which contributes to benchmarking and sharing of best practices. Between-country comparisons might also help to identify successful health policies and interventions implemented in countries with better health standards, allowing policy-makers to learn from other nations' experiences and improve their own healthcare systems (Blank et al. 2017; Tashobya et al. 2014).

References

Anand, S., F. Diderichsen, T. Evans, V.M. Shkolnikov, and M. Wirth (2001). 'Measuring Disparities in Health: Methods and Indicators'. In T. Evans, M. Whitehead, F. Diderichsen, A. Bhuiya, and M. Wirth (eds), *Challenging Inequities in Health: from Ethics to Action*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195137408.003.0005

Atkinson, A.B. (2013). 'Health Inequality, Health Inequity and Health Spending'. In N. Eyal, S.A. Hurst, O.F. Norheim, and D. Wilker (eds), *Inequalities in Health: Concepts, Measures, and Ethics*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199931392.003.0003

Blank, R.H., V. Burau, and E. Kuhlmann (2017). *Comparative Health Policy*. London: Bloomsbury.

Bleich, S.N., M.P. Jarlenski, C.N. Bell, and T.A. LaVeist (2012). 'Health Inequalities: Trends, Progress, and Policy'. *Annual Review of Public Health*, 33: 7–40. https://doi.org/10.1146/annurev-publhealth-031811-124658

Duclos, J.-Y., J. Esteban, and D. Ray (2004). 'Polarization: Concepts, Measurement, Estimation'. *Econometrica*, 72(6): 1737–72. https://doi.org/10.1111/j.1468-0262.2004.00552.x

Duclos, J.-Y., and A.-M. Taptué (2015). 'Polarization'. In A.B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution*, vol. 2. Amsterdam: Elsevier.

Edwards, R.D. (2011). 'Changes in World Inequality in Length of Life: 1970–2000'. *Population and Development Review*, 37(3): 499–528. https://doi.org/10.1111/j.1728-4457.2011.00432.x

Edwards, R.D., and S. Tuljapurkar (2005). 'Inequality in Life Spans and a New Perspective on Mortality Convergence Across Industrialized Countries'. *Population and Development Review*, 31(4): 645–74. https://doi.org/10.1111/j.1728-4457.2005.00092.x

Esteban, J.-M., and D. Ray (1994). 'On the Measurement of Polarization'. *Econometrica: Journal of the Econometric Society*, 62(4): 819–51. https://doi.org/10.2307/2951734

Fries, J.F. (2002). 'Aging, Natural Death, and the Compression of Morbidity'. *Bulletin of the World Health Organization*, 80(3): 245–50.

Gulis, G. (2000). 'Life Expectancy as an Indicator of Environmental Health'. *European Journal of Epidemiology*, 16: 161–65. https://doi.org/10.1023/A:1007629306606

Jordá, V., and M. Niño-Zarazúa (2017). 'Global Inequality in Length of Life: 1950–2015'. Working Paper 2017/418-6. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2017/418-6

Niño-Zarazúa, M., L. Roope, and F. Tarp (2017). 'Global Inequality: Relatively Lower, Absolutely Higher'. *Review of Income and Wealth*, 63(4): 661–84. https://doi.org/10.1111/roiw.12240

Nsubuga, P., M.E. White, S.B. Thacker, M.A. Anderson, S.B. Blount, C.V. Broome, T.M. Chiller, V. Espitia, R. Imtiaz, and D. Sosin (2006). 'Public Health Surveillance: A Tool for Targeting and Monitoring Interventions'. In D.T. Jamison, J.G. Breman, A.R. Measham, G. Alleyne, M. Claeson, D.B. Evans, P.Jha, A. Mills, and P. Musgrove (eds), *Disease Control Priorities in Developing Countries*, 2nd edition. Washington, DC: World Bank.

Peltzman, S. (2009). 'Mortality Inequality'. *The Journal of Economic Perspectives*, 23(4): 175–90. https://doi.org/10.1257/jep.23.4.175

Smits, J., and C. Monden (2009). 'Length of Life Inequality Around the Globe'. *Social Science & Medicine*, 68(6): 1114–23. https://doi.org/10.1016/j.socscimed.2008.12.034

Tashobya, C.K., V.C. da Silveira, F. Ssengooba, J. Nabyonga-Orem, J. Macq, and B. Criel (2014). 'Health Systems Performance Assessment in Low-Income Countries: Learning from International Experiences'. *Globalization and Health*, 10: 1–19. https://doi.org/10.1186/1744-8603-10-5

UN-DESA (2022a). 'World Population Prospects: The 2022 Revision'. New York: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat.

UN-DESA (2022b). 'World Population Prospects: The 2022 Revision. Methodology of the United Nations Population Estimates and Projections'. New York: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat.

Vaupel, J.W., Z. Zhang, and A.A. van Raalte (2011). 'Life Expectancy and Disparity: An International Comparison of Life Table Data'. *BMJ Open*, 1(1): e000128. https://doi.org/10.1136/bmjopen-2011-000128

Vladimir Shkolnikov, M., M. Evgeny Andreev, A. Begun, et al. (2003). 'Gini Coefficient as a Life Table Function: Computation from Discrete Data, Decomposition of Differences and Empirical Examples'. *Demographic Research*, 8: 305–58. https://doi.org/10.4054/DemRes.2003.8.11

Wang, H., A.A. Abajobir, K.H. Abate, C. Abbafati, K.M. Abbas, F. Abd-Allah, S.F. Abera, H.N. Abraha, L.J. Abu-Raddad, and N.M. Abu-Rmeileh (2017). 'Global, Regional, and National Under-5 Mortality, Adult Mortality, Age-Specific Mortality, and Life Expectancy, 1970–2016: A Systematic Analysis for the Global Burden of Disease Study 2016'. *The Lancet*, 390(10100): 1084–150. https://doi.org/10.1016/S0140-6736(17)31833-0

Wolfson, M.C. (1996). 'Health-Adjusted Life Expectancy'. *Health Reports – Statistics Canada*, 8: 41–45.