

Documentation for

# Life Expectancy at birth (years)

for countries and territories

Gapminder Documentation constitutes work in stepwise progress.  
*We welcome all sorts of comments, corrections and suggestions through e-mail to the author.*

---

**Gapminder Documentation 004**  
Version 7  
Uploaded: 2014 April  
Published by: The Gapminder Foundation,  
Sweden, Stockholm, 2014

Author: Klara Johansson & Mattias Lindgren  
E-mail: [mattias.lindgren \(at\) Gapminder.org](mailto:mattias.lindgren@gapminder.org)  
Gapminder is a non-profit foundation  
promoting sustainable global development  
by increased use and understanding of statistics.  
[www.gapminder.org](http://www.gapminder.org)



## 1. Introduction

This is the documentation for the Gapminder compilation of global estimates for average life expectancy at birth.

The Gapminder dataset for life expectancy at birth stretches over 200 years and contains

- a) estimates of average life expectancy from different sources (see section 3)
- b) where no estimates are available, a constructed, simple model for showing approximate levels and changes in historical life expectancy by country (see section 4)

For some countries, there are no reliable estimates before 1950, while for others, there is relatively reliable data all the way back into the 19th century. This means that the quality of data varies greatly between countries and years.

The ambition has been to include as many countries and territories as possible and the latest and most updated estimates known to us. A guiding principle has been to use the most transparent and well-documented source whenever possible, however, preference was also given to sources with as wide coverage of countries and territories as possible.

The main purpose of the data is to produce graphical presentations that display the magnitude of difference in life expectancy at birth in the world over time, for the interactive moving graph [www.gapminder.org/world](http://www.gapminder.org/world). We have not been able to make sure that every single observation is based on the best estimates available. We discourage the use of this dataset for statistical analysis and advice those who require more exact data to investigate the available data more carefully and look for additional sources, when appropriate.

## 2. Definition

Life expectancy at births is defined as the average number of years a newborn child would live if current mortality patterns were to stay the same (also written as  $e_0$ ).

The life expectancy estimates used here are period estimates and not cohort estimates. This means that the life expectancy for each year gives information about the mortality during that year and not for the birth cohort that was born that year.<sup>1</sup>

---

<sup>1</sup> Period life expectancy is the type of life expectancy that is commonly used in almost all circumstances. In order to use cohort estimates it is basically necessary to wait until the cohort has died out, which means it would be impossible to have data for the last 90-100 years. Also, to get an overview of changing global health it is generally more valuable to see the year-to-year variations in mortality than to see the differences between cohorts.

### 3. Estimates from different sources: Data sources used

A number of different sources have been used for the time series of life expectancy data used in the graphic presentation in Gapminder World ([www.gapminder.org/world](http://www.gapminder.org/world)) . This section includes short descriptions of each of the datasets used to compile the estimates of average life expectancy at birth for the Gapminder World graph.

Two big datasets were the main two sources: a) Human Mortality Database (see more info below) and b) UN Population divisions World Population Prospects. As a first priority, data from Human Mortality Database (HMD) was used where available. For countries and/or time periods where the HMD did not have data, World Population Prospect (latest revision) was used, if available. The estimates from HMD and WPP are very close. The reason for using HMD in the first hand is that 1) there is data for single years, instead of interpolations and 2) there is a good, online documentation of sources and methodology, available at [www.mortality.org](http://www.mortality.org) (see individual countries).

Both World Population Prospects and HMD are regularly updated, which also include some backward revisions. When we update we always try to use the latest available update of each datasets. However, for practical reasons it is sometimes easier to keep data from older version for specific years. Hence, you will find data from different revisions of these two sources in our dataset. The differences between the revisions is usually very small.

Where no estimates from either of the above were available, we used estimates from other sources, in the general priority order listed below.

In order to know the source(s) and metadata for a specific year or country, go to the excel spreadsheet with detailed information, available at [www.gapminder.org/data/documentation/gd004/](http://www.gapminder.org/data/documentation/gd004/) This document also tries to explain exceptions and special cases, for instance if one source was used even though another source of higher priority also had estimates for that country-year.

#### ***3.1 Human Mortality database***

The Human Mortality Database is a joint project by Department of Demography at the University of California, Berkeley, USA, and at the Max Planck Institute for Demographic Research in Rostock, Germany. The project is described in more detail at [www.mortality.org](http://www.mortality.org) .

The Human Mortality Database only contains data from countries and periods where death registration and census are virtually complete, meaning that the quality of this data is generally very good.

The data we used from the Human Mortality Database is the Period Life expectancy at birth, 1-year. It can be downloaded in bulk for all countries by going to “Zipped data files” [http://www.mortality.org/cgi-bin/hmd/hmd\\_download.php](http://www.mortality.org/cgi-bin/hmd/hmd_download.php) or country by country by going to each country’s heading.

The same data can also be found in the Life tables for both sexes 1x1 (1 year of age – 1 calendar year time intervals) for each country, under the heading  $e(X)$  which means the expected remaining years to

live at exact age  $X$ . Thus, data under the column  $e(X)$  and the row Age:0 means life expectancy at birth. A detailed description of methodology is also available at [www.mortality.org](http://www.mortality.org), as well as the raw data.

### ***3.2 World Population Prospects:***

The United Nations Population Division regularly publishes World Population Prospects (WPP), which includes both estimates and projections for all UN member countries on basic demographic and health indicators from 1950 onwards.

PLEASE NOTE: The uncertainty of estimates is in many cases considerable, in particular as concerns the earliest decades (1950s and 60s) and countries which lack a comprehensive mortality registration system (mostly low-income and middle-income countries).

Many countries lack complete civil registration, e.g. registration of all deaths, age at death and births. Estimates of population size and age composition are mostly based on census data, and the reliability depends both on the quality and the frequency of censuses. Infant and child mortality is often based on surveys within a sample of the population, sometimes in combination with census data. Mortality patterns in the whole population (necessary to calculate the average life expectancy) are often estimated through the use of so-called model life tables: Estimates of for instance infant or child mortality will be used as input in a model of assumed age-specific mortality. (Simple explanation: if we have information about how many persons die before their first birthday, we can guess how many persons die when they are for instance 13, 27, 56 or 78 years old.) When there is information about excess adult mortality, for instance due to HIV or internal conflict, this is taken into the analysis.

The reliability of the end estimate is thus dependent both on the quality of the input data, and the accuracy of the model used for each country.

To learn more about the sources and calculations made for each country, look at Sources in the WPP database: <http://esa.un.org/wpp/sources/country.aspx> - and click the country of your interest.

### ***3.3 Publications and files by James C Riley***

The historian James Riley has done an extensive review of data and literature on life expectancy. Going through more than 700 different sources in different languages, he also assessed which estimates are the most reliable. These results are described, country by country, in:

<http://www.lifetable.de/RileyBib.htm>

We have had the privilege of getting access to Professor Riley's spreadsheets containing these estimates. These spreadsheets can be accessed online from:

<http://www.gapminder.org/data/documentation/gd004/>

For most countries, one column indicates which estimates are "preferred", i.e. have been estimated to be most reliable. In some cases where there were only estimates which were not preferred, we have used these estimates in the absence of better estimates. In some few cases, the preferred estimates in

the files were probably biased upwards (as noted by Riley 2005b), and were excluded (Australia, Uruguay). For publications by Riley, see the reference list at the end of this document.

### ***3.4 The Human Life-Table Database***

The Human Life-Table Database, HLD, is a database of life-tables from different sources. It is a companion project of the Human Mortality Database, and a joint project of the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany, the Department of Demography at the University of California at Berkeley, USA and the Institut national d'études démographiques (INED) in Paris, France. The MPIDR is responsible for maintaining the database.

Often, data is only available separately for males and females. Estimates for both sexes combined have been calculated by Gapminder, assuming a sex ratio of 0.512 males for 0.488 females.

Most of the HLD life tables are life tables for national populations, officially published by national statistical offices and added to the database. All data is available for free. The quality of data can be assumed to be lower than in the Human Mortality Database.

Human Lifetable Database presents 6 types of life tables. When available, life tables type 3 ("Recalculated complete life tables") are used for this dataset, because they are a modern recalculation of the historical life tables and thus provide the best basis for comparisons between countries, since the same methods are used for all countries and time period (which is not always the case for the historical life tables). When type 3 is not available, life tables type 1 ("Complete life tables") are used. These are more similar to the original published life tables but can, unfortunately, be less comparable across countries. See methodology note at [www.lifetable.de](http://www.lifetable.de)

### ***3.5 Data from national statistical agencies***

Some estimates have also been obtained through the statistical agencies of national governments. Other sources has also been used for several countries, as noted in the excel file.

### ***3.6 Other scientific literature***

In some cases, we have searched for or come across other scientific publications which include information on life expectancy. Depending on the quality and publishing year of such publications, they can have a higher priority than the above mentioned sources, even be a first hand choice. In that case, this is explained in the excel file with metadata at:

[www.gapminder.org/data/documentation/gd004/](http://www.gapminder.org/data/documentation/gd004/)

## 4. Gapminder model for life expectancy at birth

In order to show a general outline of the world's development since 1800, Gapminder has constructed a very simple, rough model to show life expectancy for a number of countries and periods where we have no estimates. This model stems from literature studies as well as discussions with demographers and historians. In practice, on the country level, the model to a large extent relies on the online bibliography and published research of Professor James C. Riley (see reference list). Any problems or faults with the model are, however, entirely the responsibility of Gapminder.

PLEASE NOTE THAT: The Gapminder Model for historical life expectancy is extremely rough, and should not be used for research or similar purposes, nor taken as a statement of fact. The model will be updated and improved continually. Please feel free to send us your comments and suggestions.

### ***4.1 Basic assumptions:***

The model is based on two basic, very simple assumptions.

- a) That all countries historically had low levels (25-40 years) of average life expectancy at birth, with periodic variations.
- b) That all countries go through a "health transition" of sustained gains in health, implying lowered crude death rates, improved child survival and increased average life expectancy. Exception: improvement in health can be halted, stagnated or reversed due to circumstances such as war, famine, pestilence etc.

CAVEAT: these basic assumptions present a very simplified description, for the purpose of constructing this model. But it is true that all countries seem to have gone through (and many are still going through) sustained periods of decreasing mortality and improving health during the past 200 years.

### ***4.1 The Gapminder Model for life expectancy consists of***

- a) Modelling a crude baseline of life expectancy prior to the "health transition"
- b) Modelling the improvements in health during the health transition

CAVEAT: A modelled baseline can never show the historical reality of health and mortality, which can be assumed to have varied greatly both on a year-to-year and over longer periods of time basis due to vulnerability to disease, war and economic circumstances etc. The "health transition" was probably seldom as smooth and straightforward as presented in this model. In addition, the periodic variations would have been big enough to confuse the picture.

NOTE: For several countries, the entire "health transition" is shown by actual estimates for that period, and not through the model. You can see this in the moving graphics for instance when a country moves around a lot rather than moving smoothly, showing there is data for each year. You can also find this information in the detailed excel file with data sources, downloadable from: [www.gapminder.org/data/documentation/gd004](http://www.gapminder.org/data/documentation/gd004)

## ***4.2 How the baseline is estimated***

The pre-transition life expectancy were estimated in a number of ways:

- a) The earliest available estimate is used as baseline in the case where there are estimates available before the estimated onset of sustained gains in average life expectancy, and the estimates are within the range of reasonable pre-transition life expectancy (which Gapminder for the purpose of this model has set to 25-40 years).
- b) When earliest estimates for a country were lower than 25, adjacent estimates were examined to assess whether the below-25 estimate can be assumed to be a single bad year or a reflection of generally high pre-transition mortality. When it seemed to reflect generally high mortality levels (=low life expectancy), baseline was set close to 25, with a spread out of countries based on their ranking in 1950. In case the earliest estimate seemed to be a single bad year, but there was data for adjacent years, a crude average of the earliest five years was used as baseline (e.g. Mexico). In case there was not sufficient data for adjacent years, Gapminder Model was used.<sup>2</sup> If the low data seemed to be biased (e.g. by being incompatible with the UN:WPP data), it was removed and Gapminder Model used instead. This was done very roughly. Each country with early estimates of 25 or under, was given a ranking, based on their relative positions in 1950 (according to WPP 2008). Countries were then given a baseline of  $25 + 0.1 \times [\text{their ranking}]$ , meaning that the country with the lowest life expectancy in 1950 (ranking 1) had the baseline 25.1 etc.
- c) When estimate(s) before assumed health transition were higher than 40, we studied the literature to see if this could be due to biased data. If the data could be assumed to be biased, it was removed and Gapminder Model used instead. If the estimate(s) seemed relatively solid, we assumed a somewhat earlier health transition than given by the literature.
- d) In the case where there are no pre-transition estimates, continent averages have been used, from Riley (2005b, table 1). For the purpose of the graphic presentation, the countries have been spread out around the continent averages, based on their relative position within their continent in 1950-55 (the first UN estimate, according to WPP 2006 revision). This was done to avoid the situation where they are placed on top of each other, or on a line (depending on the variable on the other axis). Each country on the continent was given a ranking, based on their relative positions in 1950-55. This was done whether or not the continent average was used for this country. The median country was set at the continent average, and then countries were evenly spread around the continent average, with difference of 0.1 years between each, irrespective of

---

<sup>2</sup> Riley (2005b) notes that a "...small group began health transitions with life expectancies below 25 years: Pakistan at 20.1, Vietnam at 22.5, Korea at 23.5, Kenya and Uganda at 23.9, Sierra Leone at 24.5, and Bangladesh and India at 24.9". However, Gapminder assumes that no country could sustain a life expectancy of below 25 years over ~100 years without significant effect on population size.

the magnitude of relative differences in 1950-55. Even though all countries on the continent were included in calculating the model, the estimates obtained were only used for countries lacking an estimate that could be used as a baseline. For Fiji and Solomon Islands, where the available continent average (from Riley 2005a) was lower than 25, baseline was set close to 25, with a spread out of countries based on their ranking in 1950.

- e) In some cases, the average life expectancy “baseline” can be assumed to have improved (or deteriorated) but without continued gains, for example due to the abolishment of slavery, demographic changes or similar. In a few special cases, we have chosen to illustrate this. More information about special cases can be found in the spreadsheet with detailed metadata at [www.gapminder.org/data/documentation/gd004](http://www.gapminder.org/data/documentation/gd004)

CAVEAT: these baseline estimations should not be taken as a statement regarding the life expectancy of the individual countries during historic periods, nor regarding the relative position of countries. Most likely, the actual life expectancy varied strongly year-to-year due to circumstances, and these rough baselines can be assumed to be within the reasonable range of such variations for the country in question.

### ***4.3 How the health transition is illustrated***

We used a variety of methods to get a estimate of the year when health started to improve from its pre-transition levels.

- a) For many countries Riley (2005b) suggests a period when sustained gains in health can be assumed to have begun in each country. (In some cases, such suggestions also come from other scientific literature.) The middle year of this period is used for each country as the end year of the baseline. Between the end of the baseline and the first estimate from other sources, if later, the graph will interpolate (linear).
- b) For countries with an estimated onset of health transition in 1950, but which have estimates for 1950 (World Population Prospects 2008) that are markedly higher than the continent average, the health transition is assumed to have begun from the continent average in 1945. (Example: most countries in Africa south of Sahara)
- c) For countries where there is a relatively reliable estimate higher than 40 for the suggested year or earlier, in which case a lower baseline is set, and the onset of health transition assumed to be earlier but still within the time period suggested by Riley 2005b, appendix 1. (Example: Thailand)
- d) For a number of countries, we have estimates going back to before the estimated onset of the health transition. If there are 10 years or less between the estimates, nothing further has been done to illustrate the health transition: it is illustrated by the existing observations. For countries where there is more than 10 years between the latest pre-transition observation and the estimated time



of health transition, we have used the latest pre- transition estimate to set an end year of the baseline (the year given by Riley 2005b, appendix 1) and mark the onset of the health transition.

- e) Based on Riley (2005b) three countries were assumed to have already begun sustained gains in life expectancy by 1800: Denmark, France, Sweden. For Sweden, there is civil registration data for each year back to the 1750s (though the reliability of the data is better after 1860). For France, there are estimates back to 1806, for Denmark, civil registration data go back to 1835. For France and Denmark, an interpolation has been made between the estimated level of life expectancy at the beginning of health transition (before 1800) and the earliest available estimate after 1800 – in order to have an estimate for 1800.
- f) United Kingdom (for which the earliest estimates only apply to England & Wales) is assumed to have begun sustained gains in health by the earliest years of the 19th century.

CAVEAT: this will show a very simplified version of the health transition, which was probably seldom as smooth and straightforward as suggested by the graph. The timing of the health transition is sometimes placed within a period of a few decades, where we have chosen the middle year (Riley 2005b). For more information, consult Riley (2005b) and Riley’s online bibliography (2005c).

#### ***4.4 Gapminder “guesstimates”***

In some cases, such as during World War 2, there is a lack of data although we know mortality was higher than for a “normal” year. Based on information about civilian and military deaths as well as economic development or stagnation (which is related to gains in health) Gapminder has made very rough “guesstimates” regarding life expectancy for those years. The models we used for this is described in the separate publication “Guessing Life Expectancy for disasters”, available at:

<http://www.gapminder.org/data/documentation/gd004/>

Gapminder guesstimates are a work in progress, and aims to map increasingly more of these setbacks in health.

## **5. Links to the Gapminder dataset and other links**

A detailed dataset with metadata for each observation can be accessed at:

[www.gapminder.org/data/documentation/gd004](http://www.gapminder.org/data/documentation/gd004)

The graph for which the data is used can be accessed through:

[www.gapminder.org/world](http://www.gapminder.org/world)

## 6. References

*(see the excel file with meta data)*