New life expectancy data

Mattias Lindgren, 2014

Life expectancy data is often missing during periods of war, famines and other disasters, although mortality typically increase substantially during such periods. For this reason, we chose to include "guesstimates" for life expectancy during such events, to illustrate the effect of each disaster.

In this document we try to answer the most common questions regarding these guesstimates.

## 1. What do you mean by "guesstimates"?

It means that we use the scraps of information available for a specific disaster and try to estimate how much the life expectancy dropped during this period. The alternative would be to assume that mortality was unchanged during the disaster. This would give a more erroneous impression than a guesstimate, even a rough guesstimate.

2. Is the methodology for the guesstimate documented?

Yes, in the documentation "Guessing life expectancy for disasters" it is available at http://www.gapminder.org/data/documentation/gd004/

3. How do guesstimates change the data?

To understand how guesstimates change the data, see this example: below we show data for Serbia, both the data without guesstimates (red line) and the data with the guesstimates (blue line). In the old data neither the Spanish flu, nor the Second World War was visible. The old graph gave the impression that mortality was not affected by these two major disasters. With the new data this is corrected.
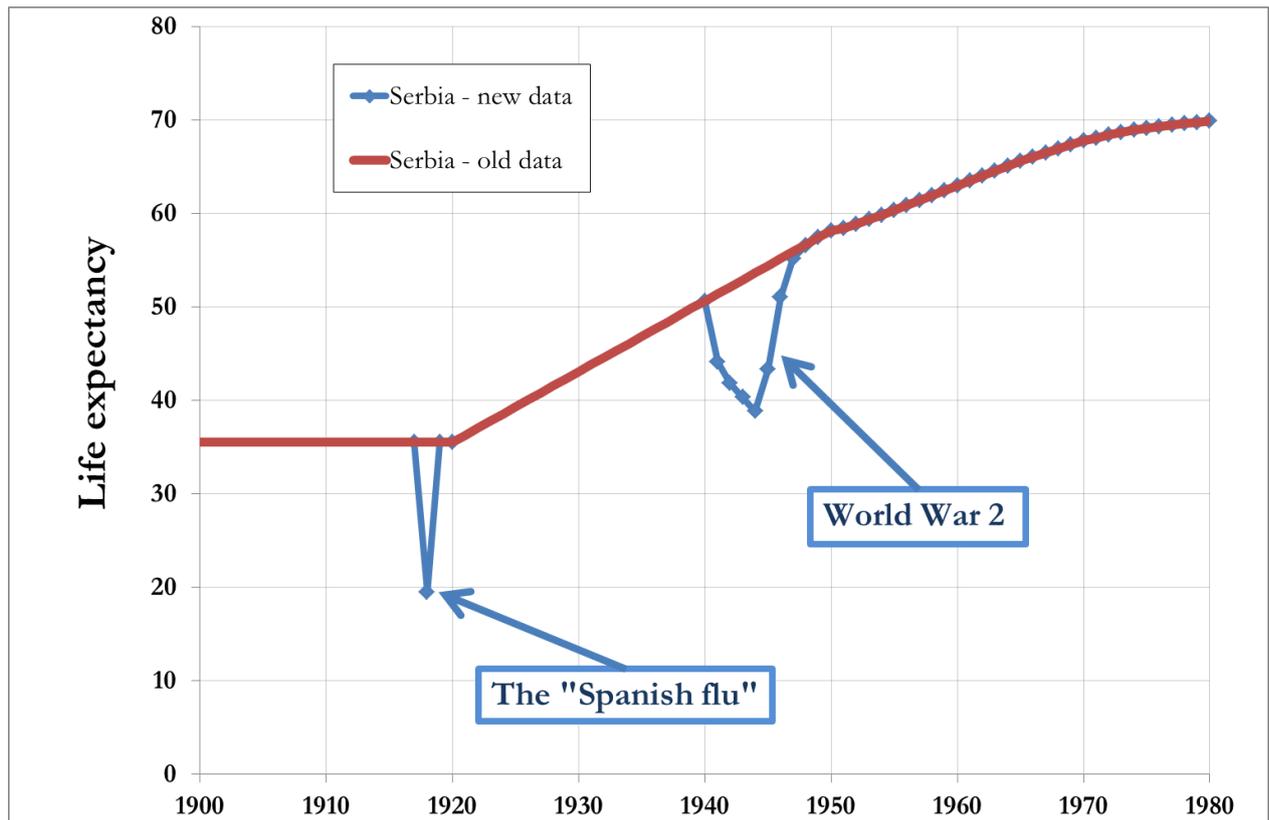
4. Why weren't the disasters reflected in the original data?

After 1950 we mainly use UN data. These are "smoothed" data, which means that short-term movements have been evened out. This is reasonable since the underlying data often display a substantial error margin. But this means that even "real" short term movements are smoothed out. Sometimes they are not visible at all, or sometimes they are visible, but with a smoothened curve, which means that life expectancy starts to drop long before the disaster starts.

Before 1950 we use estimates for many countries. Originally these were just intended to illustrate the general low level of life expectancy in the past. Hence, they did not display any short term movements. Our guesstimates for disasters display at least some of the movements.

5. Have you covered all disasters now?

No, there are plenty of disasters that are still not reflected in the data. For example, Serbia was probably affected by the First World War, but that is not reflected in the graph above. We might add more estimates in the

future, so when you read this the data for Serbia might have been improved. However, the number of potential mortality crisis are quite large, so we have only been able to deal with the most well-known cases.

There is an obvious bias in that disasters are more well-known the closer in time they are, and many disasters in the past are not documented at all. For example, we have so far only included a few disasters for 19th century Africa, even though we are aware of significant number of incidents.

6. How reliable are the estimates?

They are informed guesses, so they should in no way be interpreted as an exact figure. In most cases they are mainly based on information on the number of those who died. To use that figure to estimate a life expectancy requires a range of assumptions. Furthermore, data on deaths are themselves often uncertain, and sometimes politicized. The aim is not to give an exact figure, but rather to visualize the existence of a mortality crisis, and give some sense of whether it was very large or not.

We have not been able dig too deep into the scientific debate about the sources of the figures, and we have in many cases relied on Wikipedia (although cross checked with other in-depth sources when we felt that it was required).

7. Do you have information on how the data have been calculated?

All available meta-data are displayed in the excel-file available at http://www.gapminder.org/data/documentation/gd004/. In the excel file, each row includes the data for one country-year. The text is technical, and is not spell-checked, so it requires some enthusiasm to read. However, all information should, in principle, be there to understand how we arrived at each estimate.

8. I got the feeling you got disaster X totally wrong, can I contribute?

You can send us a mail at mattias.lindgren@gapminder.org, and we will see if we can revise the estimate in our next update. Even though we did our best, there are surely correctable mistakes in the data. However, please look at our documentation first (see above) to see how we reasoned and see if we missed something. If you are aware of a better, more recent, or more in-depth estimate we would certainly be very interested to hear about it.

Some figures are controversial, and we have no resources to make our own research to settle the issues. We tried to take some kind of middle road, based on the information we had available (since assuming no impact at all would be an even worse guess). Our choices should not be considered as taking sides in any conflict.

9. Some figures are very low. Can the life expectancy really be as low as 5 years, or even lower?

It is not impossible, but we should remember that the life expectancy figure refer to a temporary situation.

Life expectancy reflects the conditions during *one* specific year. It measures the expected life for those born in a year, if the conditions during that year remain unchanged. Say 80% of the young children die during a disaster in a specific year and a high percentage of older age groups as well. Life expectancy in that year is based on the assumption that 80% of all children would continue to die year after year. Then it is easier to imagine that the expected length of life would be very low. In reality, however, the situation usually returns to normal relative quickly, so the actual average life becomes much higher than life expectancy.

That being said, the figures are based on informed guesses only, so extreme figures (as well as less extreme figures) should not be treated as exact numbers.

10. I want to know about the reasons for the sharp movement in the Life Expectancy in a specific country. Do you have any background information for the "disasters"?

Yes, the excel-file with the data include columns with explanations for what happens in a specific country and year. We also include a link to a web source, mostly Wikipedia that give an additional background. Two rows for Serbia can illustrate the content.

Year	Life expectancy	Events 1	Events 1 – link to background information
1918	19	Spanish flu.	http://en.wikipedia.org/wiki/1918_flu_pandemic
1941	44	World war 2: Germany invades Yugoslavia in April 1941 and surrenders 1945, but Serbia was surrendered already in 1944.	http://en.wikipedia.org/wiki/National_Liberation_War_(Yugoslavia)

11. Are the events mentioned in the explanation the only factor that explains the drop?

No, many factors explain the life expectancy. We mention some major events that are likely to have had an effect on life expectancy during a year. Other circumstances might have happened at the same time that contributed to a drop in life expectancy. Several factors are often interlinked, e.g. a war might contribute to a famine, and both the war and the famine trigger an epidemic.

We have noted the explanations that we found in the literature. Some sources are more inclined to focus on some specific types of causes, whereas other sources focus on others. E.g. one book might focus on wars, and another on epidemics and a third on famines. Hence, the risk is that a specific type of disasters, such as famines, might be overemphasized in some countries, whereas other types, such as cholera epidemics, are overemphasized in others.

12. Do you only explain why life expectancy drops? Why not explain why it increases?

The vast majority of explanations note reasons for a sudden drop in life expectancy. Such explanations are easier to make, and less controversial. For example, when a war occurs it is usually quite uncontroversial to assume that it had at least something to do with a concurrent mortality crisis, even if we might debate whether other factors contributed.

However, improvement in health is usually diffused over longer time-spans, and there are a whole range of possible explanations, and there is often a lack of consensus if they were indeed of any importance.

13. Do you have any information about the quality of a specific figure?

Yes. Each observation has been assigned one of four types of "data quality".

1. Yearly, data	2. Trend, data
3. Yearly, guesstimate	4. Trend, guesstimate

The observations differ with respect to whether they are based on actual *Life Expectancy data* (e.g. from conventional sources such as the UN population data) or are *guesstimated* by us.

The observations can also differ between whether the data is "smoothed" into a *trend* or based on *yearly observations*. When the data is smoothed to just display the long term *trend* (such as in the UN population data), then the lack of a drop in any given year say nothing about whether there was a crisis or not. The crisis might be there, but hidden by the smoothing.

For *yearly* data, on the other hand, the lack of a temporary drop do imply that there was no mortality crises in that year.

So this gives us four possible combinations of data-quality. The technical details for each combination is given below (the technical terms used below are explained in the main documentation files):

1. Yearly, data Based on direct estimates of LEX, from standard sources. Examples of sources include: HMD, HLD and data from national statistical offices.

2. Trend, data Based on direct estimates of LEX, but data has been smoothed, or refer to period averages. Examples include: UN population data (i.e. WPP), historical estimates for periods, or for scattered benchmark years (e.g. many of the data in the "Riley file").

3. Yearly, guesstimate Guesstimates intended to illustrate yearly movements. Examples include: our CDR model based on yearly excess mortality for wars, famines and other disasters.

4. Trend, guesstimate Guesstimates intended to just illustrate the general level for an extended period. Examples include: The "Gapminder model" for pre-transition mortality, e.g. extrapolations from earliest pre-transition estimate. All interpolations (even between good benchmark estimates). WPP projections.

Note that the quality "within" each category is likely to be substantial, so that the best "guesstimate" is probably of better quality than the worst "data".