

Guessing life expectancy for disasters

Mattias Lindgren
Mattias.lindgren (at) gapminder.org
Gapminder
2014 May

This documentation concerns the indicator Life Expectancy at Birth. It documents the estimates done for version 7 of the indicator, but this information is likely to be relevant for subsequent versions of the indicator. More information is available at:
<http://www.gapminder.org/data/documentation/gd004/>

Gapminder World, at www.gapminder.org, displays life expectancy (henceforth LEX) for all countries since 1800. Many of the observations are, obviously, based on rough guesstimates for the most likely long-term level. This means that the data for LEX is missing for most major disasters (such as wars, famines, epidemics, genocides and natural catastrophes). If data is missing the Gapminder World tool will automatically interpolate between the closest available years. This gives the false impression that mortality was unaffected during such occasions.

The purpose of the data is to display the broad global patterns of development, and not as a tool for statistical analysis. Hence, we would argue, that it would be better to try to guesstimate the LEX during these disasters, even a rough guesstimate would be better than implying that nothing happened at all. In most cases we could never hope to find a proper estimate for the “real” LEX, but we could at least give some impression on the magnitude of the disaster. Here we will describe the general principles we use to guesstimate the impact of such disasters from the limited data available. The data itself will be available at <http://www.gapminder.org/data/documentation/gd004/>

We will first look at how crude death rates relates to life expectancy (section 1), which leads to some simple “models”, described in section 2 to 5. We go into details for a number of “disaster types” in section 6 to 9 (wars, the Spanish flu, famines, epidemics and mortality due to general underdevelopment).

1. The relationship between crude death rates and life expectancy

In the typical case we have figures for the total deaths attributed to a certain disaster, the population at the onset of the disaster, and the duration of the disaster. From that we can get a “crude excess death rate per year”, i.e. excess deaths / (population at onset * duration in years). We will only attempt to do these guesstimate where we have some kind of figure for the “baseline LEX”, which will often also be a guesstimate. We usually also have some qualitative information on the type of disaster, e.g. how severe the situation was in a particular year.

Our main data input will be either a crude death rate (i.e. number of dead per year per 1000 people, henceforth CDR), or excess crude death rate (i.e. the crude death rate during the disaster in excess of the normal death rate). How could we transform this into life expectancy? The relationship between LEX and crude death rates depends on many factors, such as the age distribution of the population and the age distribution of the dead. However, a log-linear relationship seems to represent a fair approximation, as can be seen in figure 1 below. Hence, our starting point will be to assume the following relationship (the values for alfa and beta will be discussed in due time):

$$LEX = \alpha * \exp(-\beta * CDR)$$

We have no real theoretical basis to assume this, but it fits the actual data relatively well. We also made some crude simulations with a model life table, which also fits this model relatively well. A log-linear relationship also has the advantage that LEX never takes on negative values.

Figure 1 below display the life expectancy (on a log-scale) in 1953, plotted against the CDR. Each bubble is a country (the size of the bubbles is proportionate to the population of that

country). The color of the bubbles displays the median age in the countries (as shown by the legend). We choose 1953 because the spread of LEX was the widest for this year.

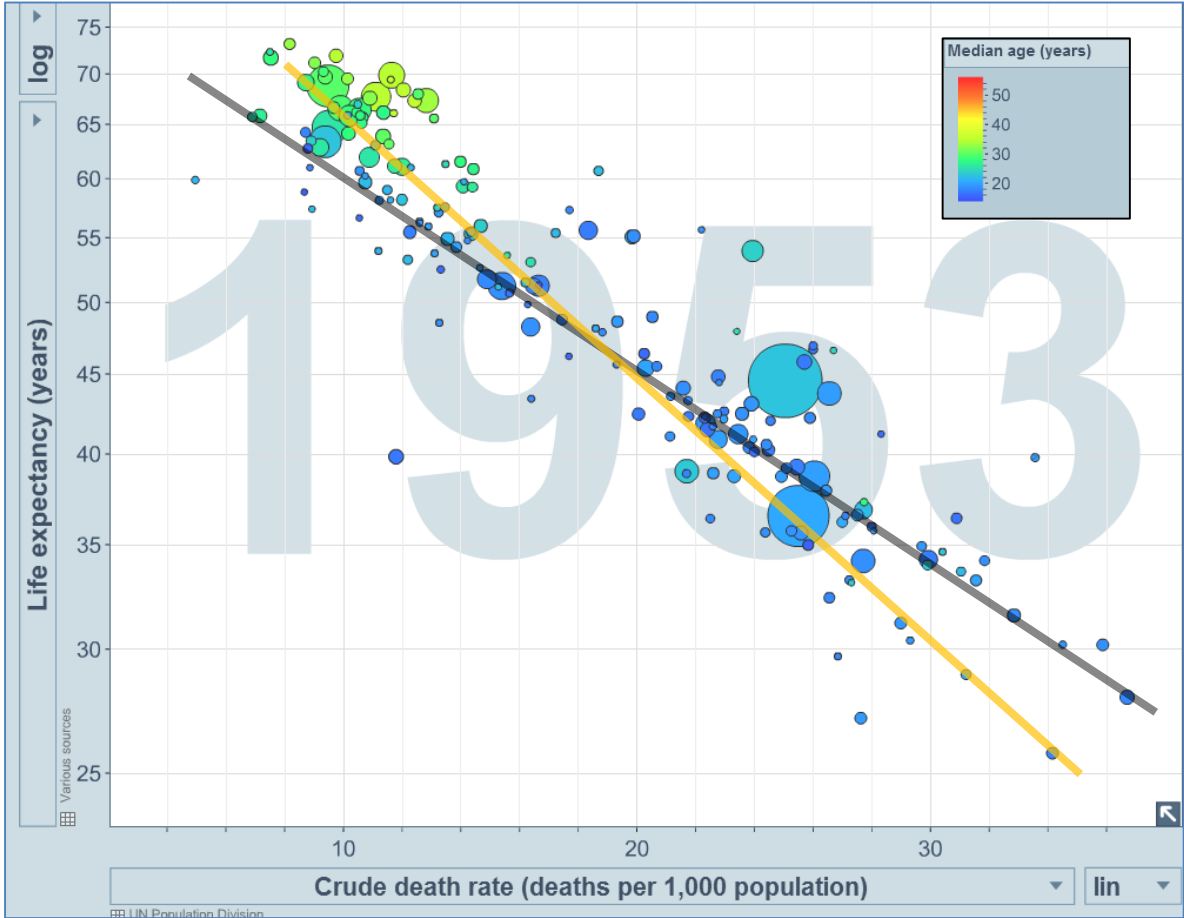


Figure 1. Source: Life expectancy from Gapminder (before the adjustments discussed here), and the CDRs are from the World Bank.

An “eye-ball regression” on this figure implies that the relationship is, roughly, like the yellow line, or:

$$LEX = 85 * \exp(-0,0384 * CDR)$$

If we try to compare only countries with a “young” population, i.e. only the blue bubbles, we get a relationship that looks more like the grey line, or roughly as:

$$LEX = 75 * \exp(-0,0285 * CDR)$$

The relationship holds even if we look at other years, or add more extreme observations for which we have data (e.g. life expectancy for Sweden in the 18th century), or 19th century Berar (India). However, it is also apparent that the predicted values often are wrong by at least five years, and sometimes wrong by more than 10 years. Hence, it would be too crude to just use the model above, even for our purposes. But how well do the *movements* in CDR fit the movements in LEX?

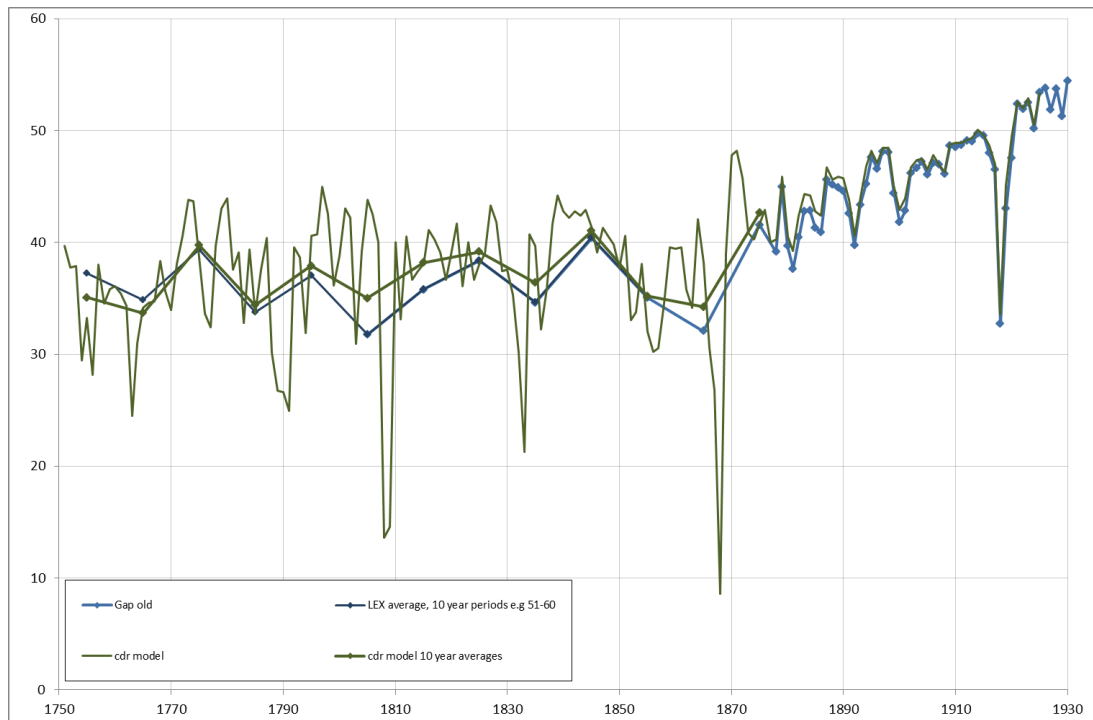


Figure 2. Finland, Life expectancy. Blue lines are actual life expectancy (sources are mainly Human Mortality Database and Turpeinen (1979). Before 1875 only 10-year averages are available. The green lines are based on the CDR-model, with the coefficients adjusted so they match the LEX as much as possible. The thicker green line is the 10-year averages, to allow comparisons with the actual LEX.

Above we compare our model for Finland (with the alfas and betas adjusted to fit the actual LEX) with actual LEX. After 1875 we have LEX for each year, and we see that the model captures the yearly movements fairly well, but we also see that the level trend to drift away over time. Other countries also show similarity in the short term movement, but for some countries the drift is even larger. Hence, to use CDR movements alone over any longer periods would be too crude even for our purposes, but we do think we can use these movements to illustrate short term movements, as long as we have some LEX benchmarks not too distant in time.

Collecting data during disasters is of course quite challenging, especially in low income settings where the statistical infrastructure is underdeveloped even in normal times. However, a number of data collection tools have come into use during the last couple of decades that allows us to make some assessment of the excess mortality during crisis (the Geneva declaration).

In some disasters a substantial number of people die abroad. This includes soldiers on foreign fronts, but also persons forcefully removed to concentration camps or similar. We are not sure what the “correct” principle is, but our principle has been to count their deaths towards the home population.

With these initial notes we can move on to describe our simplified models in more detail. We begin with our “main model”, followed with some alternative models.

2. Our main method: excess CDRs and baseline LEX

So what we will do in most cases is to use excess crude rates to guesstimate short term deviation from a baseline life expectancy. In the typical case we proceed as follows:

- We collect data on excess deaths, often from Wikipedia (crosschecked with other sources when we felt it was necessary)
- We divide this (times 1000) with the population (from Gapminder) the year before the onset of the disaster. This gives us an excess CDR.
- We divide this with the duration in years of the incident. This gives us the excess CDR per year.
- We identify a baseline life expectancy, typically the existing Gapminder figure for the year before the incident (we have not been entirely consistent here, sometimes we use a life expectancy for the same year, but this choice should have a quite modest effect)
- We estimate a model average life expectancy for the whole period of the incident, by using the formula below:

$$LEX_{model} = LEX_{baseline} * exp(-\beta * CDR_{excess\ per\ year})$$

- The choice of the beta is based on qualitative information on the characteristics of the disaster. Exactly how we choose the betas will be explained further on, but we typically choose 2%, 2.5% or 3%. Sometimes we also adjust the model so that it fits existing life expectancy estimates (if such exists for other years than the baseline).
- For multiple-year disasters we use qualitative information (or, if available, quantitative information) to guesstimate which years were worse and which were better. We rely heavily on Wikipedia and a number of historical atlases, but in some cases we utilize more specialized sources. We make sure that the average life expectancy for the disaster will match the average implied by our model.

3. Alternative method: Actual CDRs together with benchmark LEXs

Sometimes we have the actual CDR for several years (excess CDR is the actual CDR minus “baseline CDR”). We use those CDRs if we also have a number of benchmark LEXs. We calculate a model LEX from the following:

$$LEX_{model} = \alpha * exp(-\beta * CDR)$$

We choose the alfa and beta so that the model LEX match the available LEXs as well as possible (we typically do that with simple visual inspection). The match will never be perfect for all the available LEX figures. However, if we believe the available LEXs is of good quality we multiply the model-LEX with a year-specific adjustment factor, so that the match becomes perfect. For each benchmark year we find the adjustment factor that makes the match perfect for that year. For the years in-between we interpolate the adjustment factors. We have only done this when the variation in the adjustment factor is relatively small.

4. Another alternative method: Age specific mortality, model life tables, together with benchmark LEXs

Sometimes we have age-specific mortality. Most often this is infant mortality. From this we calculate a model life expectancy by fitting a model life table to this data. The model life table we use is based on “UN general, female”, as extended by the World Population prospects 2010.¹ The resulting life expectancy tends to “drift” even more than the CDR model, i.e. the level might fit the actual life expectancy well for some years, but deviate more and more over time. Again, we only use this for short term movements, so we use an adjustment factor to link the model with benchmark LEX.

5. The choice of the “beta” in the CDR model

In our model the “beta” described the relationship between CDR and log life expectancy. The relationship between crude death rates and LEX obviously depends on many factors, such as the age distribution of the population and the age distribution of the dead. We will not be able to take the age distribution of the population into account. However, we will make some assumptions on the age profile of mortality for a number of “disaster types”. For example, direct military deaths mostly affect young adult men, whereas the effect of “normal underdevelopment”, as reflected in the long-term trends in life expectancy, take a heavier toll amongst young children. Hence, it would be reasonable to use a lower beta for wars than the betas we found in the cross-country comparison in figure 1 (since the cross-country comparison is likely to mostly reflect the impact of under-development).

We will use three default values for the beta in the log-linear model above:

- War 2%
- Spanish flu, most famines, measles in new environment (e.g. Fiji 1875), small pox: 2,5%
- “Normal” underdevelopment: 3%

In retrospect we feel that it would have been better to use the same default coefficient for all disasters (probably 3%), since the age profile of death differs much more *within* each type of disaster than between them. The choice of coefficient probably has much less importance than the general uncertainty with the model. Anyways, since we have used these coefficients (with some variations) we should motivate the choices.

First, we could look at the coefficients implied by the observations for which we actually have both CDR and LEX. From these we can calculate a beta in this way:

$$\beta = \frac{\ln(e_{baseline}) - \ln(e_{crisis})}{CDRplus}$$

Let us now look separately at a number of disaster “types”. We begin with wars.

¹ The table has been extended by me in the following ways: extrapolated (manually) down to LEX 10; interpolated to steps of 0.1 years of LEX; interpolated the survival chances to yearly ages, assuming constant survival chances within each age bracket.

6. The “beta” for wars

Deaths attributed to wars can be divided into military deaths and civilian deaths. Both military and civilian deaths can be divided into direct deaths and indirect deaths. Direct deaths include deaths by weapons or use of force. Indirect deaths are the deaths that have been attributed to the indirect effects of war, such as the collapse of health care, destruction of infrastructure, famines, epidemics, dislocation and so forth. Military indirect deaths are, for example, the soldiers that die from diseases spread due to the war.

The share of indirect deaths in the total deaths varies substantially. In a selection of recent conflicts the share indirect deaths range from 0% indirect deaths for Kosovo 1998-99, to 94% for Sierra Leone 1991-2002 (the Geneva declaration, calculated from table 2.3).

We can also get a sense of the share direct deaths by looking at male and female LEX separately. We do that for France below, with data from Human Mortality Database. The World War I stands out as a disaster where the direct effect seems to dominate.

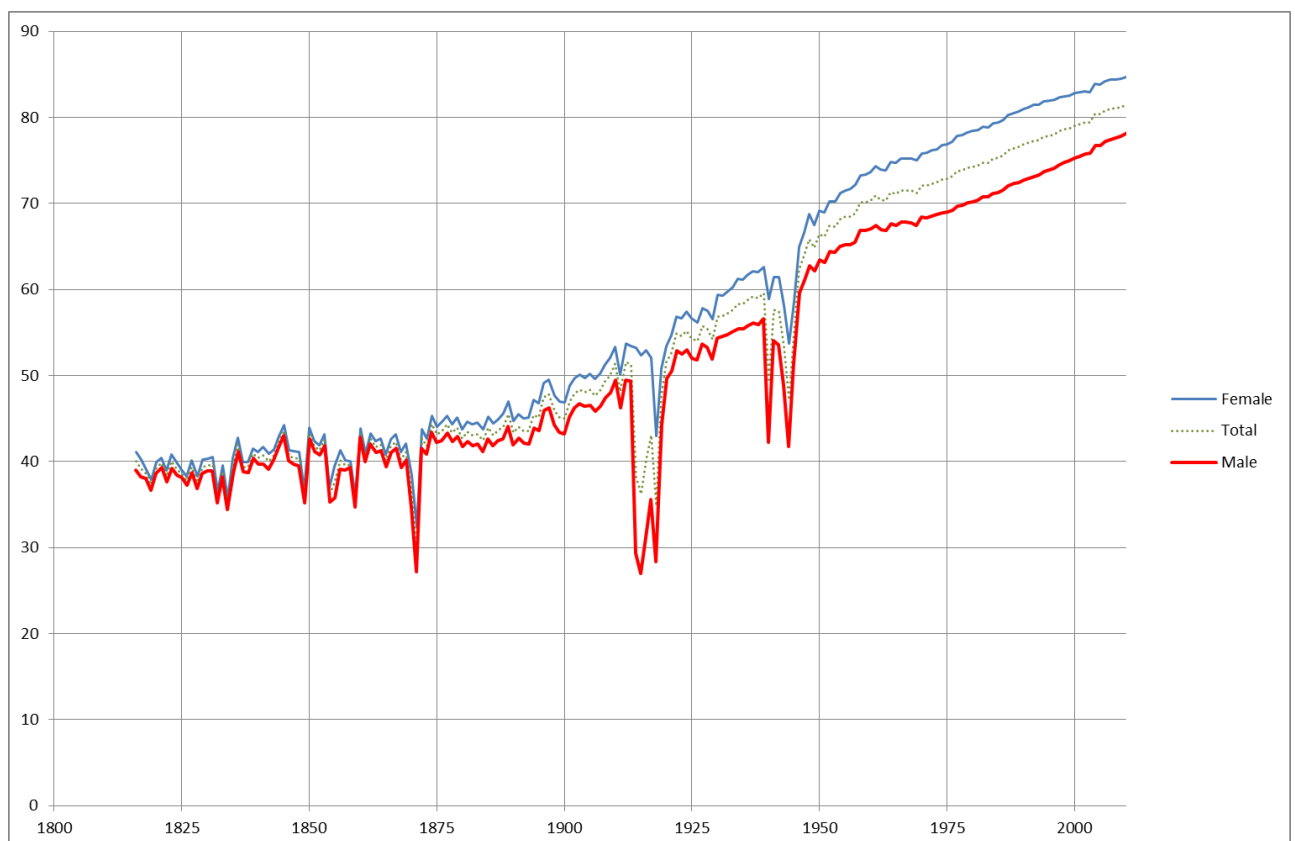


Figure 3. French Life expectancy, Human Mortality Database

Sources on mortality do vary with respect to what they include. “Official statistics often exclude military deaths that occur abroad” (Jdanov et al). This might seriously overestimate the life expectancy during a war. The Human Mortality Database has, to my understanding, made the effort to include all such deaths.

On the other hand, some sources on war deaths only include the direct military battle deaths, others include all direct deaths (civilian and military), whereas other include all excess mortality

attributed to the disasters, or all mortality, whatever the cause, in excess of the baseline mortality. For example, some sources include deaths from the Spanish Flu in the World War 1 mortality, whereas others exclude it.

Hence, it is important to try to understand what is included in any mortality figure, to avoid double counting or undercounting. However, we do not need to know how much of the excess mortality can be attributed to a war; we are mainly interested in the size of the mortality. However, the causes of mortality can be interesting because they might say something about the age-distribution of deaths. Direct battle deaths would overwhelmingly affect the young adult age-group, civilian direct deaths (e.g. due to “aerial bombardments”) probably affect all age-groups more equally, whereas starvation and epidemics might take a higher share amongst the children.

The history of a country often makes it clear how the country was involved in a specific war. The involvement can take a range of different forms, with different mortality effects. We will use this information when we try to “distribute” the deaths for each year of a specific war (we often only have information on the deaths for the duration of a war, not the year-by-year deaths toll). The types of war involvement include:

- *Military forces in direct battle overseas.* Examples include the US involvement in World War 1. Direct battle deaths might be substantial, but the indirect effects are likely to be minimal (especially if the war is far away). Hence, in these cases it will not be a problem if the sources we use fail to include indirect deaths. We will generally assume that an army that are losing ground have higher mortality than an advancing army.
- *Ground-battles in the territory of the country.* In this case there might be substantial civilian deaths, but size of this varies substantially, depending on the characteristics of the war. This can depend on a number of factors, such as where the battles are fought, the length of fighting, the degree of disruption, the behavior of the combatants and the pre-conflict level of health care, infrastructure and governance. According to the GenevaDeclaration.org: “Three main factors explain the differences in proportion between direct and indirect conflict deaths: the quality of pre-existing health care systems and patterns of disease; the speed and extent of the humanitarian response; and the intensity and duration of battle.”
- *Aerial bombardment.* Civilian areas are sometimes directly affected by bombardment, even when the frontline is far away from the country. This was very common during the World War 2. I have not seen any information on the age distribution of mortality, but a guess is that all age groups are affected rather indiscriminately (although children sometimes were evacuated to the country side).
- *Guerilla warfare.* This might cause both direct and indirect deaths
- *Occupation.* The effect of occupation seems to vary substantially, very much depending to the occupational policies and the extent of resistance. The cases of Netherlands and France (in the figure below) indicate that life expectancy was somewhat depressed during the occupation, but that the main mortality peaks occurred in the years of occupation and liberation.
- *Famines and epidemics.* Sometimes some specific indirect effects of a war are noted, e.g. the winter famines in Netherlands or in Leningrad. Excess mortality figures specifically

attributed to these disasters are sometimes provided. Some of these indirect sources of mortality are probably taking a higher toll on children than the battle deaths.

- *Genocide and other “democides”*. The number of people killed in genocides could be large enough to affect the life expectancy. They are often done in connection to a war. Examples include the Holocaust and the Rwandan genocides. In most genocide all age-groups are targeted.
- *Being close to wars*. Sometimes the effects of war can spill over to neighboring countries. Trade can be disrupted, with serious effects on the population. Epidemics can spread from the war zone (the Spanish flu could be considered a case in point). Military mobilization can spread diseases geographically. One example seems to be Denmark during World War 1. The country was neutral and stayed out of battle, but nevertheless displayed a mortality peak during the war. The Wikipedia article claims that this is due to the strong impact of the trade disruption caused by the war.

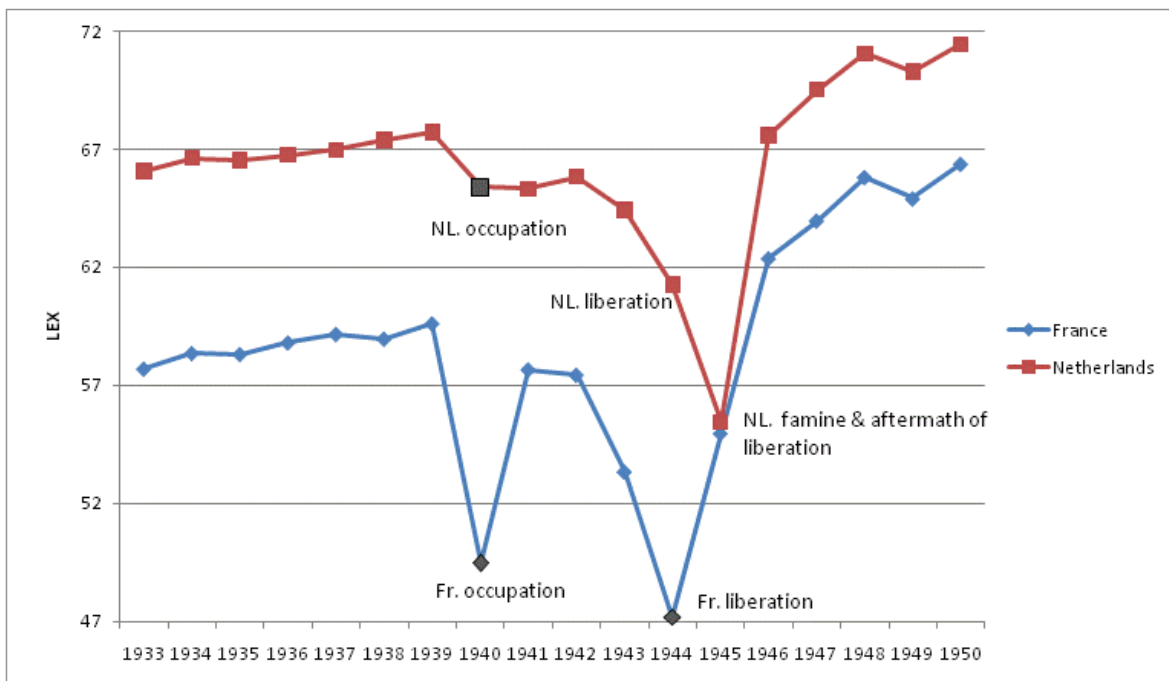


Figure 4. Life expectancy for Netherlands and France, during the World War 2. Source Human Mortality Database (accessed 2009).

Figure 4 illustrates the life expectancy during the course of a war for France and Netherlands. We marked some of the key incidents for each country.

Let us now look at wars for countries for which we have data. We have listed our available cases below, together with the calculated “beta”. The excess mortality figures are usually taken from Wikipedia (cross-checked with other sources, such as war atlases). The figures supposedly include both direct and indirect deaths, although some indirect civilian deaths might sometimes have been excluded. The Spanish flu is not included in the world war 1 deaths.

	data note	excess cdr per year	-beta (average)
<i>Sweden ww2</i>	non participant	0,02	-119,9
<i>Australia ww2</i>	overseas	1,45	-0,7
<i>Canada ww2</i>	overseas	0,67	-2,4
<i>United States ww2</i>	overseas	0,80	-2,8
<i>Denmark ww2</i>	little fighting	0,17	-3,2
<i>Spain, civil war</i>		8,53	1,1
<i>Netherlands ww2</i>	seem to be more inclusive causality figures, e.g. a major famine is included	6,90	1,1
<i>Belgium ww2</i>		2,10	1,4
<i>Finland, ww2</i>		5,24	1,6
<i>Ukraine ww2</i>		41,42	1,6
<i>UK ww2</i>	colonials included in military deaths	1,89	1,7
<i>Italy ww2</i>		2,00	1,7
<i>Norway ww2</i>		0,68	2,5
<i>Luxembourg ww2</i>	data unclear	3,39	2,6
<i>France ww2</i>		2,64	4,2
<i>UK ww1</i>		5,74	2,1
<i>France ww1</i>		9,46	2,7
<i>Italy ww1</i>		7,57	3,2
<i>average all positive</i>			2,1
<i>average all after ww1</i>			1,9

Table 1. Calculated betas for wars

The first couple of betas are in fact negative, meaning that the average life expectancy during the war increased from its pre-war level. These are all countries that had a low excess mortality, and they are either non-participants, countries that only fought “overseas”, or countries where limited fighting occurred.

The betas for the rest of the countries varies substantially, non-surprisingly, from 1.1 to 4.2. The average is around 2, which we will take as our guideline for wars. This is the lowest coefficient we use, which make sense given the supposedly low share of child death in direct mortality.

7. The “beta” for the Spanish flu

The outbreak of the Spanish Flu in 1918 is treated as a special disaster category in our data since it affected so many countries. The age profile of mortality of the Flu was quite particular since it took its heaviest toll amongst young adults and young children (Taubenberger & Morens, 2006), as illustrated in figure 5 below.

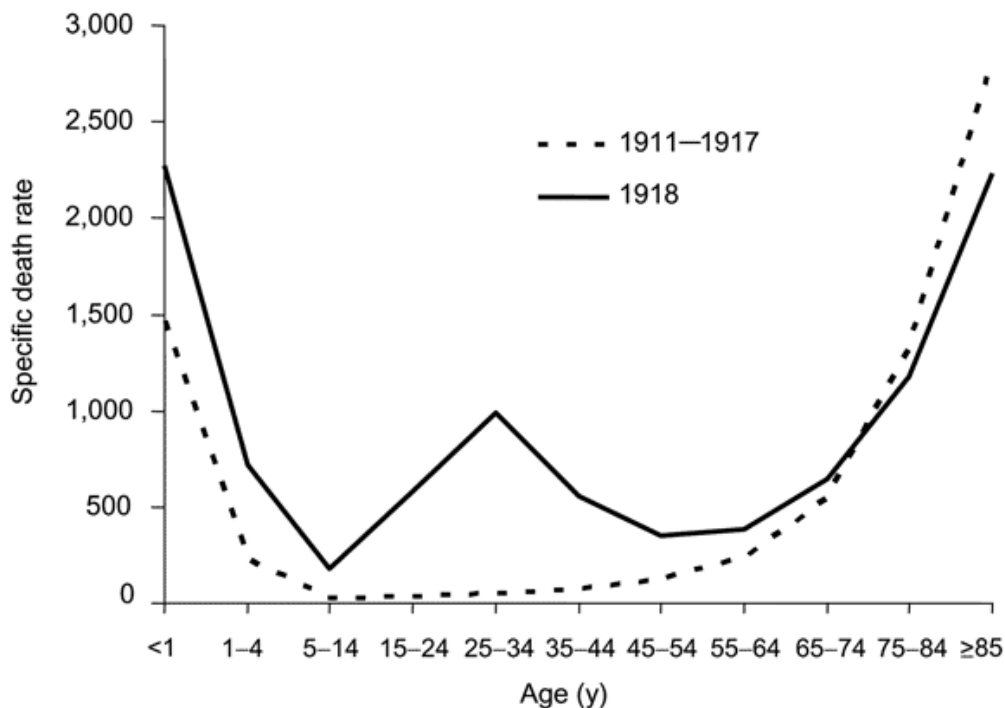


Figure 5: example of age profile of mortality in the Spanish Flu. Source: copied from Taubenberger & Morens, 2006, through Wikipedia.

This age-profile implies a higher beta than for battle deaths (due to the effects on children) but, probably, a lower beta than for normal underdevelopment, which should be even more concentrated amongst children. We calculated betas for the countries for which we had both life expectancy (generally from the Human Mortality Database) and excess CDR from the Spanish Flu (from the various sources listed later). The betas are illustrated in figure 6 below.

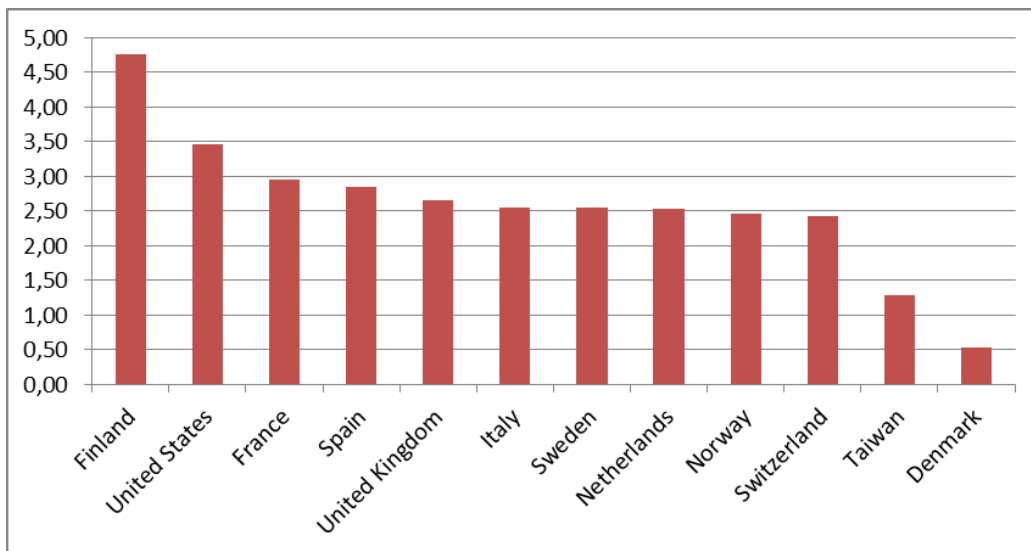


Figure 6. Calculated betas for 1918. Sources: see text

War was ongoing in several countries, which of course adds uncertainty. The biggest problem is probably Finland, since they were not directly in the war before 1918, and the Finnish civil war broke out in 1918. This might explain the high beta in Finland. We are not certain about

the data quality in Taiwan (taken from the data compilation of Riley, rather than the Human Mortality database). We have no explanation for why the impact in Denmark was so low.

If we exclude Finland from the data we get an average beta of 2.39%. If we also exclude Taiwan and Denmark we get an average of 2.71%. It also seems that 2.5% is a “typical value”. We settle for this rounded figure of 2.5%, which imply a higher beta than for wars, but a lower beta than for normal underdevelopment.

There are several studies that offer excess mortality figures for the Spanish Flu. We combined them, in the following order of preference: Ansart et al (2009) offer data for European countries, and they have the most detailed methodology. Then we use Murray et al (2006) and Johnsson & Mueller (2002). The first two sources agree quite well. After that we use a selection of ad-hoc sources, including accounts that a country totally escaped the Flu (which did happen in a few cases) and crude death rates. For the remaining countries we extrapolated from regional averages or from neighbors.

8. The “beta” for famines

The impact of a famine depends quite a lot on circumstances. Many famines take their heaviest tolls on the youngest and the oldest in the population. For example, the excess age-specific death rate in the Irish Potatoe famine was, according to estimate of Boyle & O’Grada, as in table X below.

	male %	female %
0-4 years	29	28
5-9 years	20	20
10-59 years	8	7
60+ years	31	22

Table 2: Exess mortality during the Irish famine 1845

However, the direct death of starvation in a famine is often of relative minor importance compared to the indirect effects caused by the epidemics and turbulence caused by the famine (just as direct battle deaths are often a minor part of deaths in war). Hence, the age profile of mortality during famines is likely to vary substantially, depending on what indirect effects the famine triggers.

We had relative few direct estimates of life expectancy during major famines, and there were large uncertainties with the data in each of these cases. We have tabulated the information we have in the table below. The estimates for excess mortality for the Chinese famine during the Great Leap Forward varied with more than a factor two. For the Irish famine we were uncertain about the appropriate length of the disaster. We include beta estimates during all these assumptions.

	Ukraine 1933	Great low estimate	leap, dead	Great high estimate	leap, dead	Irish famine, long duration	Irish famine, short duration
<i>lex start</i>	37,0		50,2		50,2	38,3	38,3
<i>lex average</i>	9,1		37,2		37,2	20,6	20,6
<i>dead</i>	2 600 000		20 000 000		43 000 000	985 000	985 000
<i>population start</i>	34 000 000		628 000 000		628 000 000	7 555 000	7 555 000
<i>excess cdr</i>	76		32		68	130	130
<i>Duration in years</i>	1		4		4	6	4
<i>excess cdr/year</i>	76		8		17	22	33
<i>beta</i>	1,84		3,77		1,75	2,87	1,91
<i>source LEX</i>	Vallin etal (2002)		Luo, Sheng. 1988.		Luo, Sheng. 1988.	Boyle & Ograda	Boyle & Ograda

Table 3: Calculated “betas” for three famines

As expected, the estimated betas vary substantially; between 1.75% and 3.77%. We choose, rather arbitrarily, a beta of 2.5, i.e. the same as for the Spanish Flu.

9. The “beta” for other epidemics

We have information for a large number of major epidemics, not connected to wars or famines. The age profile varies between diseases and contexts. For example, Boyle & Ograda, note that small pox often have greater impact on the young, whereas cholera often had greater impact on the old. However, we decided that we had too little information to set up rules for specific diseases. We generally choose to use a beta of 2.5 for epidemics, unless we had specific reasons to suspect a deviating age profile.

10. The “beta” for normal underdevelopment

The long term differences in life expectancy could be seen as mostly reflecting the extent of general underdevelopment of the country. Underdevelopment generally takes its heaviest toll on young children. Hence, we could assume that the beta for “general underdevelopment” is in the high end of possible beta values.

We illustrated the coefficient between (log) life expectancy and CDR in figure 1 above. We found that the general correlation implied a beta of 3.84% (the yellow line in figure 1) and a correlation between countries with roughly the same age structure imply a beta of about 2.85% (the grey line). As a compromise we use a beta of 3% for cases that are deemed to most resemble general underdevelopment.

References

- Ansart et al (2009), "Mortality burden of the 1918-1919 influenza pandemic in Europe", *Influenza and other respiratory viruses* 3 (3), 99-106. (e2717)
- Boyle and o'Grada (1986), "Fertility trends, excess mortality, and the Great Irish Famine", *Demography*, vol 23, no 4, pp. 543-562.
- Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on various dates 2009-2013, specified in the text or datafile).
- Jdanov et al (2008), "A generalized method for reconstruction of war-related mortality"
- Johnsson & Mueller (2002), "Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic", *Bull Hist Med* 76 (1) (e2001)
- Luo, Sheng. 1988. Reconstruction of Life Tables and Age Distributions for the Population of China, by Year, from 1953 to 1982. Unpublished PhD dissertation, University of Pennsylvania.
- Model life table, based on UN general, female, model life table, from World Population Prospects, interpolated and extended by the author
- Murray, Lopez, Chin, Feehand & Hill (2006), "'Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918-20 pandemic: a quantitative analysis'", *Lancet* 368: 2211-18
- Riley data, a dataset compiled by James Riley, available at <http://www.gapminder.org/data/documentation/gd004/>
- Taubenberger JK, Morens DM. 1918 influenza: the mother of all pandemics. (2006) *Emerg Infect Dis*
- The Geneva declaration (2008), "Global burden of armed violence 2008"
- Turpeinen (1979), "Fertility and mortality in Finland since 1750", *Population studie*, vol 33, no 1.
- Vallin et al (2002), "A new estimate of Ukrainian population losses during the crises of the 1930s and 1940s". *Population Studies: A Journal of Demography*, Volume 56, Issue 3
- Wikipedia – various articles
- World Bank –crude death rate available at:
<http://data.worldbank.org/indicator/SP.DYN.CDRT.IN>