

Working Paper No. 225

“Mortality Inequality”

SAM PELTZMAN

**George J. Stigler Center for the Study
of
the Economy and the State**

The University of Chicago

Mortality Inequality

Sam Peltzman

Economists and other social scientists have long been interested in measures of income inequality. Most such measures are based on individual or household income at a moment in time. However, as a measure of the distribution of welfare, the typical income inequality measure leaves out an important dimension: the length of time over which an income or consumption stream is enjoyed. (See Becker, Philipson and Soares, 2004). Clearly two individuals with the same annual income or consumption but differing in longevity do not have the same total welfare. This paper is about differences in longevity.

Specifically I seek to describe the historical evolution of longevity differences across individuals. This is motivated by the comparative neglect of the longevity component in the analysis of inequality. Think of an individual born today. He or she will have a lifetime welfare or utility related to total lifetime consumption (Y), which can be expressed

$$Y = CX \tag{1}$$

where C is some appropriately discounted measure of the individual's annual consumption and X is the number of years the individual lives. Using lower case to denote natural logs, (1) would be

$$y = c + x \tag{2}$$

The variance (V) or standard deviation (S) of the log of income across individuals in a society is a common measure of income inequality. The analogous measure of lifetime inequality as described by (2) would be

$$V(y) = V(c) + V(x) + 2r(c, x)S(c)S(x) \tag{3}$$

where $r(c, x)$ is the correlation coefficient between c and x across individuals.

The first term of this familiar formula for the sum of two random variables describes the distribution of per capita consumption. Most of the income inequality literature concentrates on something like this $V(c)$ term. Another literature on health inequality essentially focuses on some version of the correlation coefficient in the third term. The correlation between income and longevity within and between countries is positive, but most of this may reflect differences between the very poorest countries or people and the rest. (Preston, 1975; Cutler, Deaton and Lleras-Muney, 2006). This paper will focus on the mainly neglected middle term. This term measures the inequality of lifetimes, the kind of inequality that arises when one individual born today dies shortly thereafter while another lives to a ripe old age.¹

It is hard to know why the $V(x)$ term, or its square root $S(x)$, has been neglected relative to the other two terms in (3) in discussions of inequality. (Henceforth I will mainly focus on estimates of $S(.)$ rather than $V(.)$) The two kinds of inequality are not different in any fundamental way. Both are affected by or related to a similar list of background characteristics like gender, race or parental attributes. Both are affected by a production process whose endogenous correlates such as education, occupation, location and so forth are similar.

Historically, mortality inequality has contributed substantially to total inequality. To provide some context for that assertion it is helpful to start with some sense for the magnitude of the more familiar consumption inequality term, $S(c)$. For that purpose I take the standard deviation of the log of household income as a proxy for $S(c)$. Figure 1 shows some rough estimates of this proxy for the US and Sweden².

¹ The variance of log age provides a convenient description of mortality inequality in the context of the simple variance decomposition in equation (1). However, it is problematic as a stand-alone measure. The distribution of mortality inequality is skewed left, but with a significant mass in the first year of life. Hence the variance of age at death is sensitive to the infant mortality rate. This sensitivity is exaggerated by the log transformation. For this reason I show important results both inclusive and exclusive of infant mortality.

² The roughness is due to the kind of data we have available to portray long time series of the proxy. These are income distributions or even fragments thereof.

[Figure 1]

For the US, perhaps the least egalitarian of the rich countries, this proxy has ranged between .75 and .9 over most of the 20th century. For Sweden, the prototypical advanced welfare state, the much more limited data lies .15 or .2 below the US figures. Both countries are characterized by increasing income inequality over the last quarter century or so. The crudeness of the proxy needs to be emphasized. For mechanical reasons the series in Figure 1 understate the true inequality.³ For substantive reasons the data probably overstate what we want to measure – inequality of average annual consumption over a lifetime.⁴ We can, however, get a sense of magnitudes from Figure 1: think of $S(c)$ as having an upper bound around 1 in the developed world, as having no clear trend in the US but probably declining in other rich countries toward a lower bound of around .5 or more.

By comparison, as will be clear shortly, $S(x)$, the standard deviation of longevity, has taken on values even greater than 1 over much of the history available to us, and values over .5 have been common until comparatively recent times, when much smaller values emerged. Thus much of the social inequality in our history has been at least as much due to mortality inequality as to income inequality. And we shall see that the narrowing of social inequality over the last century owes more to declines in $S(x)$ rather than $S(c)$.

This paper mainly summarizes mortality data of a kind that may be less familiar to economists than demographers. Accordingly, I begin by outlining the kind of data that I will use, where they come from and what measures I will extract from them. This is followed by several sections describing various aspects of the data.

³ All households in a decile or quintile are assigned the same income. Thus within-group dispersion is suppressed.

⁴ The proxy contains transitory income elements that we might like to exclude, and it probably does not adequately capture life-cycle effects: current-income-poor older households who are dissaving.

The history of mortality inequality that I will review occurred in the context of substantial increases in average life expectancy. Accordingly, before I review the history of mortality inequality I will summarize the more familiar facts about longevity, and I will discuss the conceptual and empirical connection between the two. This will be followed by a description of the trends in mortality inequality over two long periods, one beginning around 1750 and the other beginning around 1900. The choice of these periods is driven by data availability. One remarkable feature of these data is the extraordinarily low levels to which overall mortality inequality has been driven in the developed world. Relatively little of today's social inequality is coming from the lottery of life.

I also describe the history two important sub-categories of mortality inequality: inequality by gender and place. A female born today can expect to live around 10 percent longer than her male counterpart. I document the considerable historical variability in this female premium. This is followed by a more detailed analysis of geographic inequality in the United States.

Historically, as emphasized by Becker et al (2004), where you are born matters to your expected longevity. Some of this geographic inequality is income-related, because there are also geographic differences in average incomes. The penultimate section of the paper reviews the history of mortality differences across US states and counties. Here there is more focus on the recent history (since 1970), when geographic income differences have tended to widen. I investigate whether geographic mortality differences have tended to reinforce or offset these income differences.

I will mainly, but not entirely, ignore the well-studied correlation between health and wealth in the third term in equation (1).

I. Mortality Data: Concepts and Sources

Most of the data used in this paper come from life tables. A life table is simply a function in which the y-variable is the number of survivors in a birth cohort and x is their age. The size of the cohort is set to an initial arbitrary value (usually 100,000) at age 0 (birth) and the table follows the 100,000 births to their deaths. Thus suppose 5,000 of the 100,000 die before their first birthday and another 1000 before their second. The life table would show:

age	0	1	2	...
survivors	100,000	95,000	94,000	...

The bottom row gets steadily smaller with age until it approaches zero at around 100 years. The first difference of the life table is mortality and, in percentages the mortality rate. Thus, the mortality rate at age 0 is 5 per cent; the mortality rate at age 1 is 1.05 per cent (1,000deaths /95,000 alive at the start of the year), etc. Sometimes we have only mortality rates, and we have to infer life tables. Often the mortality rates cover a range of ages – for example, we might be told only that the rate is 3.05 percent for age 0 to 2 for the population described by the above table⁵ - and we have to infer intermediate values. These details and how I handled them are discussed in the appendix.

The most commonly used kind of life table, and the one I use exclusively, is the period life table, which comes from mortality records in a particular year or group of years. This answers the question: how many survivors would there be at age k if 100,000 individuals born

⁵ Since 1 minus this rate compounded for 2 years reduces 100,000 to 94,000

today have the mortality rates that we observe today for people at various ages up to k ?⁶ This would be an unbiased estimate of actual experience only if there was no medical progress. If we can expect continued progress, then today's life table values understate the expected number of survivors at each age from a group of 100,000 born today. The reader needs to keep in mind the conservative counterfactual implicit in data derived from period life tables.

I extract two kinds of information from life tables, corresponding to the first two moments of the distribution of the age of death. One is expected longevity, and the other – the focus of the paper – is the standard deviation of (log) longevity. Expected longevity – the number most readers are familiar with – is simply a weighted average of ages at death where the weights are mortality at each age. Thus in the illustrative table above, 5000 live one year or less, 1000 live 2, and the remaining 94,000 live to ages from 3 on up to 110 or so. The total life years lived by this hypothetical group of 100,000 described by any life table is the sum of all mortality weighted ages up to, say, 110, and this sum divided by 100,000 is expected longevity. The inequality measure is just the mortality weighted standard deviation around this expected value (or, more accurately, around the expected value of the log of age.⁷)

These mean and variance estimates have been sensitive to the level and change of infant mortality and over much of the history I shall describe. Mortality rates usually trace at a U-shape when plotted against age, with the left branch declining sharply from birth through the first few years of childhood. This pattern still holds, but the level of infant mortality has declined

⁶ A cohort life table tracks a given birth cohort over time, and answers the question: of 100,000 people born in year $t-K$ how many actually survived in years $t-K+i$ (where i ranges from 0 to around 100)? This has the advantage of describing actual experience, without any need to make assumptions about technology. But it has the fatal disadvantage of cutting off much interesting history. For example, a reasonably complete history of cohort life tables would today get us only up to the birth cohorts of the first part of the 20th century, because non-trivial numbers of those born later still survive.

⁷ For this purpose I assume that mortality, including infant mortality, takes place at the end of the year. This makes the minimum value of $\log \text{age}=0$, and avoids the problem that \log of age 0 is undefined. However, in fact, most infant mortality occurs closer to the date of birth than the end of the first year

substantially. For the typical developed country of today, something like 15 per cent of all new borns died before their first birthday up to a century ago. This figure is now under 1 percent. Thus something like a sixth of the increase in expected longevity is due to reduced infant mortality. There is a much larger quantitative impact on $S(x)$ – the standard deviation of log life – because this measure is especially sensitive to the weight on the ages near zero. I deal with this sensitivity by supplementing overall measures of $S(x)$ with those excluding mortality below age 5.

The life tables I use or construct come from a variety of sources, as described more fully in the appendix. Many are from two public use data bases, one at the University of California (Berkeley) and the other at the Max Planck Institute.⁸ These were supplemented by searches of the demographic literature and national vital statistics print and electronic sources. Some of the literature tries to estimate life tables from sketchy available data, and I did not attempt to exercise quality control. Nor did I include every country for which some data are available. I used two main selection criteria: length and importance. Because the paper has a historical focus, I included any country where the data begin 1850 or before and excluded any where the data begin after 1900.⁹ For countries where the data begin between 1850 and 1900 I included large countries or less developed countries (where historical data are often especially scarce). The resulting sample of 23 countries is nevertheless heavily tilted toward today's developed economies, and especially toward Scandinavia where the data reach back into the 18th century. About half the sample has data from 1850 or before, and I begin with this half because of the insight it may offer about the role of modern medicine.

⁸ <http://www.mortality.org/> and <http://www.lifetable.de/> respectively.

⁹ For the last 50 or so years, the UN Demographic *Yearbook* has mortality rates and life table estimates for many more countries than are in my sample.

II. Mortality Inequality Since the 18th Century

Figure 2 summarizes the mean and standard deviation of longevity from birth for the 10 countries where the data go back at least to the mid 19th century. Figure 3 has the same data for the population that survives to age 5 – i.e., it excludes infant mortality and the historically perilous first few years of life. The sample is fairly homogeneous in that it is drawn completely from northern Europe plus the US.

[Figure 2] [Figure3]

Panel A in both figures repeats the familiar history of the ongoing growth of life expectancy. There is little or no progress before 1850 but continual growth thereafter. These panels also show a narrowing dispersion of life expectancy over time within the sample, though this is most clear in Figure 2. This narrowing dispersion of life expectancy across countries is the aspect of inequality emphasized by Becker, Philipson and Soares (2004). They emphasized the narrowing difference between rich and poor countries, but the figures show that this narrowing has been pervasive.

Panel B in the same figures summarize the evolution of within-country inequality, which is the focus of this paper. Inequality is measured by the standard deviation of the log of years lived, which is an approximation of the square root of the middle term in equation (3). Here, as with mean longevity, there is long-run progress following another long period of stagnation. And, again similar to the mean, the inequality measure seems to converge toward a common value as we get closer to the present. Two aspects of this history - the magnitude and timing of the decline in inequality – deserve elaboration.

The magnitude of the decline in mortality inequality is remarkable when compared to the evolution of income inequality. In the mid 18th century and for a long time thereafter (and probably before) the inequality measure hovers over 1.5 in Figure 2. Today it averages below .4. This decline of 100+ log points is at least twice the range of the similarly scaled consumption inequality proxies in Figure 1. Thus, in terms of the decomposition in equation (3), the decline of mortality inequality has been a far more important contributor to the historical decline in social inequality than any narrowing of income inequality.

The history in panel B of Figure 2 is dominated by the decline in infant mortality. However, even if we exclude infant mortality entirely, as in panel B of figure 3, the decline in mortality inequality remains remarkable. Here the long run decline is on the order of 40 log points, which might be comparable to the decline in income inequality over the same period in one of today's advanced welfare states.

While there has been a considerable reduction in mortality inequality, this process seems to have started well after the onset of increasing life expectancy. This is clearest in Figure 2. Here life expectancy begins to increase noticeably around 1850 while inequality does not begin its steep descent until the 19th century is nearly over. In Figure 3 there is some progress before 1900, but it accelerates as that date approaches. The impression that progress in inequality lagged progress in longevity is sharpened in Table 1.

[Table 1]

Here the century from 1850 to 1950 is divided into two 50 year periods, and the underlying progress in the two 50 year periods is compared. Around a third of the total progress in longevity over this century had been completed by 1900, but only a sixth to a quarter of the total decline in inequality had been completed over this initial 50 year period.

This lag between the onset of medical progress and the decline in mortality inequality is important because it shows that the connection between the two is not as straightforward as a casual glance at Figure 2 or 3 might suggest. Progress, as it has played out so far, unleashed two opposing forces on inequality. They can be illustrated in the following stylized model that captures the essence of these forces: Suppose all individuals die either at some early age (A_1) or they survive to die in old age at age A_2 . Then the mean life expectancy is

$$X = (1-p)A_1 + pA_2 \quad (4)$$

where p is the proportion who survive to old age. The variance of $\log X$ here is

$$V(x) = p(1-p)D^2 \quad (5)$$

where

$$D = (a_2 - a_1) \quad (6)$$

and the lower case again denotes logs. In this stylized model inequality depends on how many survive to old age and the difference in the ages at which death occurs. It is useful to write out how (5) changes over time

$$\frac{dV(x)}{dt} = 2p(1-p)\frac{dD}{dt} + D^2(1-2p)\frac{dp}{dt} \quad (7)$$

In principle, progress could consist simply of more people surviving to old age ($\frac{dp}{dt} > 0$). In that case, the first term in (7) would be zero and we would be left with what I will call a “Kuznets curve” for mortality. As with Kuznets’ original application to income inequality¹⁰, the connection between medical progress and inequality is non-monotonic. Inequality actually increases with progress at a decreasing rate as long as the fraction, p that survives to old age is

¹⁰ In that application growth of per capita income is driven by migration from low income rural areas to high income urban areas, and income inequality initially increases and then reverses when the fraction in urban areas passes .5.

less than half. As death in old age becomes common, the second term on the right hand side of (7) becomes negative and progress becomes the handmaiden of equality.

In practice medical progress has also affected the age at which people die in ways that offset the first branch of the Kuznets curve. Specifically, the gap between the age of those who die “early” and those who survive to old age has declined steadily. Thus the first term in (7), which captures the influence of that decline, has been negative over most of the history. Table 2, [Table 2]

which is taken from English life tables, illustrates some of the key developments in the relevant variables over the last 150 years or so. It draws the line between old and young at 80. Progress is evident throughout this period, but it mainly exempts the old. Instead, progress consists of more people surviving to old age and increased life spans for those who do not. Only in the last 50 years has there been any measurable increase in longevity of the aged. Even so, the aged today can expect to live only three years more than in the era of the dark satanic mills. The two major changes since that time are the vast increases in the numbers of old people and in the life expectancy of those who do not make it to old age.

While the cut off at 80 is somewhat arbitrary, the data in the table help to organize the historical connection between medical progress and inequality. For much of the period, the proportion of old people has been small enough to keep us on the upward sloping branch of the mortality Kuznets curve – i.e. the second term on the right hand side of (7) has been positive, though gradually diminishing. The first term has been negative over most of the history as infant mortality and other sources of premature death have been reduced, while geriatric longevity remained the same. Over the first 40 or so years after 1850 the Kuznets effect and the narrowing gap between pre and post geriatric lifetimes offset each other. Some time around the turn of the

20th century the mortality Kuznets effect became sufficiently small for the net effect of medical progress to become decisively egalitarian. This interplay between the Kuznets effect and an increase in longevity toward an apparent ceiling in old age will be important in interpreting the more recent history of mortality inequality in less developed countries.

That history is included in Figures 4 through 7. These include 13 countries where the data begin sometime in the latter half of the 19th century. This sample is considerably more heterogeneous than the previous one. Some countries in this sample have been near the top of the world income distribution all along (e.g., Australia), some rose considerably (Japan) and others remain well down (India). For comparison the heavy line in each figure is the average of the 10 country sample described earlier. To reduce clutter and gain some further insight the 13 countries are divided into two quasi-geographic groups. One group of 8 includes European countries plus Australia and New Zealand. All of these, with the possible exception of Russia, are near the top of the world income distribution today (though not necessarily throughout the depicted period). The other 5 countries are in Asia and Latin America, and of these only Japan would qualify as a rich country today.

The broad patterns in these two diverse groups are:

- Long run progress is evident in both the mean and inequality of longevity for both groups. The singular case is Russia, where progress stopped, slowed or even reversed, depending on the measure, some time in the 1950s.
- Convergence in mean life expectancy, as discussed in Becker et al (2004) is also evident, but the degree of convergence is sensitive to economic growth. The convergence is obvious in countries that were poor and became rich (Southern Europe, Japan), less so in countries that remain poor today (Brazil, India).

- By contrast, the degree of inequality has not converged in any straightforward way. For the initially poor countries, the Kuznets effect visibly retards progress until well into the 20th century, long after the reduction in inequality had begun in the developed world. For Southern Europe and Japan, for example, the inequality at birth measure is nearly the same around 1940 as it was in the developed world a century earlier. For Brazil and India, decisive progress in this measure does not begin until much later. On the whole there is more variety in inequality at birth in the middle of the period than at the beginning or end, and greater rich country-poor country differences today than 100 years ago
- Some of the convergence in the mean and divergence in inequality is driven by the development of infant mortality. If we take out infant mortality (Figures 6 and 7), the common pattern is one of considerable variety in both the levels and improvement of both measures until c.1940, and then a gradual convergence

Taken together, the long run trends summarized in Figures 1 through 7 reveal an important role for mortality inequality in overall social inequality. In some countries, like the United States, where income inequality has apparently declined little over the last century, the link between mortality equality and social equality may have been decisive. Today's poor countries have considerably more mortality inequality than rich countries, but their recent experience suggests that this gap will be steadily eliminated thereby ameliorating the social inequality in these countries.

III. Gender and Geography

Inequality, in mortality as well as in the command over resources at a moment in time, has a group dimension. The odds of success depend on, among other things, where you are born and your sex and race. For example, a girl born in a rich country today can expect to live around 5 to 10 percent longer than a boy born the same day. Here I summarize what we know about the historical evolution of this and related facts. I also try to describe the trend of geographic disparities within one country (the United States). The US data emphasize the last quarter century or so, when overall income inequality has increased, in an attempt to assess whether geographic mortality disparities have complemented or offset the income inequality trends.

A. Male-Female Mortality Differences since 1750

Females are paid less than males per hour, but they live longer. Indeed, they have been the harder gender from the beginning of our data. Panel A. Figure 8 shows the history of the mean ratio of female to male life expectancy across the long-period-sample of ten countries.

[Figure 8]

Since 1750 the female advantage in life expectancy has averaged around 7 per cent and fluctuated in a range between 2 and 10 per cent. These fluctuations trace clear cycles.¹¹ The last cycle spans the great medical advances of the last 2/3 of the 20th century. The peak around 1980 coincides roughly with the emergence of advances in treatment of heart disease, which has an especially high incidence among males in late middle age. There is a hint of a slow decline in the female advantage over the century or so preceding the discovery of antibiotics (say 1840-1940), but that sub-trend is not significant statistically. Overall, there is a remarkable stability in

¹¹ The significant first order serial correlation is around .5

this aspect of mortality inequality: the relative female advantage is the same today as it was in 1750 and the same as it has averaged since then.

This long-run stability of the average masks considerable differences across countries at any moment and over time. Panel B of Figure 8 shows the cross-country standard deviation around the mean values plotted in panel A. There is a clear (and statistically significant) decline in this standard deviation in the last quarter of the 19th century, but the standard deviation is never trivial. For example, even the lower values after 1900 (around .02 or so) suggest a range across these uniformly rich countries in any year which is roughly as wide as the historical band since 1750. Figure 9 illustrates the variety in time series behavior as well as the cross-country [Figure 9]

variety. It shows the post WWII history for the four largest countries in this sample. They all show some evidence of the inverted U-shape that characterizes the female advantage in this period. But the pattern differs greatly among the four. In every year the range across these countries is around .04, or half the historical range of the sample average that is plotted in panel A. of Figure 8. The peak of the inverted U occurs around 1970 in the US and England and more than 20 years later in Germany. The rank correlation between the countries at the beginning and end of the period is negative.¹² The interesting question is why such variety survives similar levels of development and presumably similar access to the frontier of medical technology.

The overall temporal stability of the female mortality advantage masks some significant cross-currents. In Figure 10 I break the female advantage into three components based on age: [Figure 10]

first in early childhood, then in the broad window from early childhood to late middle age and finally in the years beyond 50. Higher and rising values signify greater female advantage. The

¹² But this is not significant in this small sample.

first two panels show the male to female ratio of the probabilities of death from birth to age 5 and then from age 5 to 50. Prior to the twentieth century the overall probabilities of death over these two intervals were roughly the same – on the order of .3 or .4. Today both probabilities are well below .1. The last panel shows the female to male ratio of the extra years lived for those who survive to 50, a number which has ranged from around 15 to 30 extra years over the sample period.

Females have had the better mortality experience in all life stages since the mid 18th century.¹³ And until recently medical progress tended to widen that female advantage at every stage from birth to old age. The advantage in early childhood begins growing in the late 19th century. Then around 50 years later, there is a marked acceleration of the female advantage across the later life stages. This growth in female advantage stops in the last two or three decades, and it is decisively reversed for the population that reaches age 50.¹⁴ That decisive reversal at older ages is driving the downward part of the inverted U traced in panel A of Figure 8 by the female advantage in overall life expectancy over most of the last century.¹⁵

The larger sample of countries available over the last century (not shown) mainly confirms the patterns evident in figures 8 through 10. That is, even for countries that are or have been well below the top of the world income distribution, there is a female advantage in life

¹³ That is, all the ratios in all panels exceed 1 for all years. A few ratios dip below 1 for individual countries, but this is extremely rare. For example, every one of the country-years that are averaged in panel A exceeds 1.

¹⁴ The reversal reflects advances in treatment of heart disease, where male morbidity at 50 is considerably greater than female morbidity. However, as shown in panels A and B of figure 10, the trend toward a widening female advantage ended at all life stages at roughly the same time. This suggests that more is at work here than reduced heart mortality, which is essentially non-existent below age 40.

¹⁵ Because mortality by age 50 became very rare the dramatic growth in the female mortality advantage did not translate into similarly dramatic growth in the overall lifetime advantage. Even if a male today has twice the mortality risk as a female every year from birth to age 50 his cumulative risk is on the order of 5 per cent. Reducing that risk by half or even to zero would have little effect on overall male life expectancy. The future of gender differences in life expectancy will depend almost entirely on developments in old age mortality, such as the advances in heart disease treatment that underlie the reversal of the female advantage in recent years.

expectancy at birth that averages on the order of 5 to 10 per cent, and that flattens or reverses in the last quarter century.¹⁶

While I can safely spare the reader most details about female-male differences in lower income countries, no discussion of this source of inequality can ignore the experience of the former Soviet Union. We have already seen how singular this experience has been in the aggregate, with declining life expectancy and stagnating inequality (Figure 6). Every age and gender shows symptoms of this deterioration, but male-female differences are an especially important part of this story.

[Figure 11]

Figure 11 summarizes some salient facts. Here data for Russia¹⁷ are shown relative to corresponding data for eighteen OECD countries. Panel A. is a summary of relative life expectancy at birth. It shows improvement for both males and females over all of the first part of the 20th century with an acceleration following the Bolshevik Revolution. By the mid 1950s Russian life expectancy had come to nearly equal that of the typical advanced country. Then the improvement stopped and a relative deterioration set in that accelerates after the collapse of the Soviet Union, particularly for males.

Perhaps the most notable aspect of this widening gulf between the sexes is the experience of mature adults, some of which is summarized in panel B of Figure 11. For some Russian age-sex groups any deterioration in mortality has been relative but not absolute. For example, infant

¹⁶ India is a prominent exception. Female and male life expectancy over the twentieth century is the same on average, and the time pattern is broadly opposite to much of the rest of the world. Female relative life expectancy in India fell 6 percentage points from 1900 to 1940, then reverses around 1970, when the opposite pattern first becomes evident in the developed world.

¹⁷ Our data are for the territory of the Soviet Union from 1897-1987. So it includes the Tsar's empire up to 1917. After 1987 we have some data for some of the constituents for some years, including one year of overlap (1990) between the Russian Federation and the Commonwealth of Independent States (the immediate successor to the Soviet Union). For present purposes the differences in any relevant datum between these two entities were trivial. Accordingly I used Russian Federation data after 1990.

and early childhood mortality has continued to improve absolutely and substantially in Russia, as it has in the rest of the world. However, for adult Russian males the recent deterioration has been absolute and substantial. The sharp relative deterioration of life expectancy for mature Russian males since 1990 is clear in panel B of Figure 11. But the underlying data show more than a relative decline. Life expectancy at 50 stopped increasing for both sexes from the mid 1950s to the end of the Soviet Union. That already peculiar pattern continues for females up to the present. However, male life expectancy at 50 declined on the order of 3 to 4 years over the decade of the 1990s. A decline of this magnitude is essentially unprecedented in all the data we have since 1750.¹⁸ A 50 year old Russian male today can expect fewer years of life than his counterpart at the coronation of the last Tsar. Younger Russian males have not been spared. They are better off today than a century ago, but not by much. And their mortality experience has also deteriorated considerably in the last half century.¹⁹

The recent Russian experience on gender differences is unique and extreme. But it highlights the lack of any clear trend over time in these differences. Since overall inequality has declined, the contribution from the gender component has grown, but it remains tiny.²⁰

¹⁸ There are considerably larger declines for males during wars, which have been taken out of our data. In addition, very short period substantial declines in life expectancy sometimes occurred during famines. The unique aspect of the Russian adult male decline is that it does not appear to be temporary. It is risky to place great weight on year-to-year changes in these data, but they do show a halt of the decline in adult male and female mortality in the last few years. (The last available year at this writing is 2005.)

¹⁹ The probability that a 5 year old Russian boy will die by age 50 has doubled since 1957, compared to a roughly 20 per cent increase for a 5 year old girl. Most of that increase occurred in the 1990s. (Of course, both these probabilities have decreased substantially in the rest of the world). The current 5 to 50 mortality rate for males exceeds 1/4; it was over 1/3 in the 1890s.

²⁰ To illustrate, consider the contemporary values for the standard deviation of log life in the whole population as shown in figures 2 or 3. These are on the order of 20 to 40 log points depending on the definition. Mean gender differences on the order of 7 log points contribute less than 1 log point to these standard deviations.

B. U. S. States and Counties Over the Last 100 Years

Just as expected lifetimes and their dispersion vary across countries they vary geographically within countries. This geographic variability undoubtedly encapsulates a variety of personal and social characteristics, such as income, education and (particularly important in the U. S. context) race. Here I summarize some broad trends in geographic differences in the U.S. and also explore some of the underlying socio-economic components of the most recent trends. I use two readily available local area data for the U.S.

1. STATES: 1900-2005

State level data are available from 1900 onward for a gradually increasing number of states and for all states after 1930.²¹ The top panel of figure 12 shows the dispersion of expected lifetimes across two samples of states: the 20 states for which data are available from 1910 and [Figure 12]

then all states from 1930 on.²² The 20 state sample excludes Southern states, and that exclusion drives the considerable level difference between the two samples. The broad trend here is similar to international trends – a substantial decline in differences among states (and regions). The decline is driven by the ubiquitous near-elimination of premature mortality, so current geographic differences arise almost entirely from differences in geriatric mortality.

A notable twist is the very sharp second derivative of this decline in geographic dispersion: it is essentially over by 1950 outside the South, and the small regional differences reached by 1980 have persisted since then. One source of this persistence is a mild increase in

²¹ The states adopted death registration systems at different times.

²² The data are from estimated life tables for each available state-year. The life tables are based on mortality rates for 10 year age groups. Age specific mortality within a 10 year group is estimated from age specific shares of the 10 year group in contemporary US life tables.

the relative dispersion of non-geriatric age-specific mortality (around the declining average). An illustration of this is in the lower panel of figure 12. This shows the relative (log) geographic dispersion of the probability that a 5 year old dies by age 50. This measure seems to have turned upward around 1950 or 1960.²³

While the persistent differences across US states appear small historically, they loom larger in international context. For example, the standard deviation of current log life expectancy across the 18 OECD countries in our data (.017) is around 2/3 as large as the cross-US state counterpart (.025).²⁴

The trends in geographic longevity differences and income differences have been broadly complementary, though the timing has been different. Prior to around 1940 the cross-state standard deviation of log per capita income was on the order of .3 to .4, and then declined by about half to 1970, where it has remained since. Longevity inequality had been declining well before 1940. But the overall pattern of sharp decline that ends abruptly is similar. Even the broad magnitudes of the decline – 15 or 20 log points – seem similar.²⁵ Thus, in the sense conveyed by equation (1), the decline in longevity differences has contributed substantially to the reduction in geographic differences in social conditions.

The recent playing out of the geographic trends in both income and longevity inequality has occurred against a background of rising personal income inequality. This suggests some questions about the current connection between income and mortality inequality that we explore with county level data from this recent period.

²³ The sharp blip in 1990 outside the South is due to the cresting of the AIDS epidemic, which is concentrated geographically. AIDS mortality peaked in the 1990s and then declined sharply.

²⁴ The range in incomes is similar across the two samples – a standard deviation of around 15 log points.

²⁵ Even the historical magnitudes may be similar, though we are constrained here by lack of early mortality data from the South. To illustrate the point, consider the 20 states for which we do have pre 1930 mortality data. The standard deviation of log per capita income over these states is .16 for 1919, which is similar in magnitude to the contemporary standard deviation of log longevity for these states.

2. COUNTIES: 1970-2005

I constructed life tables for the over 3000 US counties for five year periods in the early 1970s, 1980s and 2000s.²⁶ I combined them with Census data on county socio-economic characteristics to begin addressing some questions about social inequality that have been implicit in the foregoing history. Tables 3 and 4 summarize respectively the mortality data extracted [Table3] [Table 4] from the county life tables and the socio-economic data I will focus on. I use these two sets of data to estimate regressions of the general form

$$Y = \sum b_i X_i + error \quad (8)$$

where Y is some county mortality measure, such as expected life or the inequality of life within a county and the X's are county characteristics. These descriptive regressions bring together some strands in the literature on the socio-economic conditions associated with geographic differences in mortality. For this reason I opt for a compact list of X's that includes the main broad categories of interest – income, education, race and family structure.

The period covered by the data is particularly interesting because of the widening income inequality over its last two decades. Table 4 shows that this widening inequality also had a geographic component: both the cross-county and within county measures of income inequality rise from 1980 to 2000 after declining or remaining stable in the previous decade. One obvious question, in the spirit of equations (1) - (3), is whether the associated trends in mortality tended to follow a similar pattern. The answer seems to be that they do. For example, in Table 3 the standard deviation of expected life across counties contracts from 1970 to 1980 then expands.

²⁶ See Appendix for details. 1990s data were available, but I decided not to use them to avoid distortion from the AIDS epidemic, which began in the mid 1980s and crested in the early 1990s. Data prior to 1970 are not readily available in electronic form.

Panel A in Table 5 focuses on the extremes of the geographic mortality distribution. Here counties are classified according to 1970s predicted values from a regression like (8)²⁷. Specifically the first two columns describe mortality in the 10 percent of the population living in the counties with the lowest and highest predicted life expectancy in the early 1970s. The remaining columns trace the experience of these two groups of counties over the next 30 years.²⁸ The last two lines in panel A show, perhaps unsurprisingly, that these populations differ considerably in income and education. The broad pattern in the mortality data is complementary with the income inequality trends. For the first decade, when income inequality was stable or declining, the mortality measures at the extremes tend to move together. Then they tend to move further apart, as do incomes, after 1980.²⁹ This pattern holds for mortality at all life stages. Panel B shows the same kind of co-movement between income and life expectancy more directly by defining the extremes by income rather than mortality.³⁰ Another message from the Table is the overwhelming importance of long-term progress in mortality in overcoming historical inequality. Compare the highest decile in 1972 to the lowest in 2002. They look very much alike, with the latter having a slight advantage on most measures. Thus the least healthy part of the population today has moved past the healthiest in 1972.

The descriptive regressions in Table 6 try to get behind the complementary inequality trends. For example, the positive cross-county correlation between income and life expectancy is

²⁷ Specifically, I use the predicted values from regression in the second column of Table 6.

²⁸ I classify by predicted life expectancy instead of actual to mitigate potential problems from measurement error and associated regression to the mean. However the results in the table are essentially duplicated if the deciles are defined by actual life expectancy in 1970, or any other year.

²⁹ Another way to summarize these mortality trends is to regress the change in, say, life expectancy on the initial predicted level. Over the whole period – i.e., when the change in life expectancy from 1972 to 2002 is regressed on the 1972 level – the coefficient is -.11, which implies a mild tendency for places with low life expectancy to improve more than others. This measure is the net result of a much sharper narrowing of geographic inequality from 1972 to 1982 – the corresponding regression coefficient is -.26 when the y-variable is the 1972 to 1982 change – and a widening of inequality from 1982 to 2002 (coefficient = +.14).

³⁰ Panel B. shows only life expectancy, but there is so much overlap between the counties in panels A and B that every other mortality measure for the sample in panel B follows the same pattern as in panel A

already clear from the summary statistics. One apparent implication could be that widening or narrowing of the income distribution would be reflected in a widening or narrowing of the life expectancy distribution.

Table 6, however, suggests a more complex picture. The regressions are subject to the usual caveats about causality and inferences about individual behavior. The connection between the main X-variables and life expectancy has been well-studied in a variety of contexts.³¹ My regressions include within county inequality measures for income and education. This is motivated by a robust regularity in the literature – the so-called „Preston curve“ (from Preston (1975)). This is a concave cross-country relation between income and life expectancy in which the positive slope flattens considerably beyond a modest per capita income threshold. If this concavity holds within a population it implies that wider income inequality will be associated with lower mean life expectancy. Cutler and Lleras-Muney (2006) find some equivocal evidence of a Preston curve for education³² that would carry the same implication for education inequality.

Perhaps the most important message of Table 6 is how powerful the socio-economic descriptors are in this specific geographic context. They can rationalize most of the geographic inequality in any year even without including state fixed effects. The sign pattern is essentially the same over the three decades.³³ Moreover, as judged by the summary statistics, the small local idiosyncrasies seem to be getting smaller over time. In such a context the interesting questions

³¹ On the connection between life expectancy and income see Cutler et al (2006). The benefits of marriage have been well known since at least Gove (1973).

³² They show a concave relation between life expectancy and years of education across countries, but they find a linear relationship between education and mortality across individuals in the US.

³³ One possible reason for the consistency is that these are not independent samples. As can be seen in the last line of Table 3 there is substantial positive serial correlation of life expectancy. One way to adjust for this serial correlation is to estimate a partial adjustment model. That is, think of the x-variables as describing the target or equilibrium life expectancy to which the actual value adjusts gradually. This calls for including the lagged value of life expectancy in the regression (so the model cannot be estimated for 1972) and then using the coefficient of this lagged value to estimate a speed-of-adjustment parameter and the parameters of the target life expectancy function. When I implemented this model (results not shown) the coefficients from the target function were uniformly similar to those shown in Table 6. This suggests that intercensal changes in the x's have important effects, and we are not simply replicating the same cross-section three times.

revolve around magnitudes rather than statistical significance, and this aspect is captured in the last three columns. These show beta coefficients – the number of standard deviations that life expectancy changes when the x-variable increases by one standard deviation – from the regressions with state fixed effects.³⁴

If, for expositional convenience, we think of betas of .2 or more as “substantial,” then some subset of the socio-economic betas exceed that threshold in every decade. The magnitudes tend to get larger over time as well. The individual results that stand out are those for education and family structure. The (partial) income effects are the least important quantitatively.

In terms of the direction, the only surprise is the robust and large *positive* effect of education inequality. This hints at an increasing return to education across individuals. And such an increasing return also appears across counties in all three years.³⁵ These education effects clearly merit further study and attention. The substantial and growing effects of mean education evident in Table 6, for example, suggest that conventional measures of the rate of return to education, which focus on annual earnings, may be understated. The regressions also point to education as a counterforce in the recent growth in geographic inequality of expected life. The geographic standard deviation of average education has narrowed by around .15 years since 1980 (see Table 4), and the relevant regression coefficients would translate this into a roughly comparable narrowing of the dispersion of expected life. The regressions also suggest that the important offsets to this favorable education effect are not coming from growing dispersion of

³⁴ The numbers would be essentially the same if the regressions without state fixed effects, or coefficients from the equilibrium relationship described in the preceding footnote were used to estimate the betas.

³⁵ When a quadratic term for mean education is added to any of the regressions the coefficient is significantly positive, and the resulting convexity is quantitatively meaningful. For example the derivative of expected lifetime with respect to years of education for the quadratic version of the 1972 regression (with state fixed effects) is .78 at the mean years of education, and 50 per cent larger (smaller) than this at the mean plus (minus) one standard deviation.

income but from geographic variation in the continued decline of the married couple household.³⁶

The county data also provide an opportunity to explore geographic differences in the inequality of individual lifetimes. Table 7 condenses results from regressions of within-county mortality inequality measures on the same set of descriptors as in table 6. The regressions in the two tables do not, however, describe independent phenomena. This is clear from the strong negative correlation between expected life and the inequality of life in panel C. of Table 3. Indeed, as progress moves us further away from the peak of the mortality Kuznets curve that negative correlation is becoming stronger. Accordingly, Table 7 mainly echoes the preceding results on expected life. For example, mean income seems to matter less than its variance and less than the education variables, any past advantage of large population centers has disappeared by 2002, etc. The main difference between the two tables seems to be that family structure matters less and race matters more for the inequality of life than its mean. The robust and somewhat sizeable negative relation between inequalities of lifetime and of education suggests that there is more at work here than the individual benefits of education.³⁷

The important message from these county data is probably in the declining importance of local idiosyncrasies. Overall, regional and within county mortality inequality have not changed much in the last two or three decades. However the part of these measures that cannot be explained statistically by standard observables has declined. Thus the regional and local inequalities that remain tend more and more to be a reflection of these observables.

³⁶ The standard deviation of the female head household percentage has increased by about a year since 1980 (see Table 4). In the regressions with state fixed effects that extra geographic dispersion would be sufficient to offset the effect of reduced spread of education in promoting greater geographic equality of expected life years.

³⁷ Any positive relation, linear or non-linear, between individual lifetimes and individual education would imply a positive correlation of the variance of lifetimes and the variance of education across communities

IV. Summary: The Mortality Gini

The Gini Coefficient is perhaps the most commonly used summary measure of income inequality. It is just the share of the area below a hypothetical egalitarian cumulative income distribution that is not occupied by the actual cumulative distribution. For contemporary household income distributions in developed economies this coefficient would range from the mid .20s in the Scandinavian welfare states to the .30s in Continental Europe and the .40s in the US. The difference between Scandinavia and the US probably conveys a rough sense of the limits of the historical effects of income redistribution policies – i.e., a halving of the income Gini.³⁸ If you replace “income” with “life years” in the definition a corresponding Gini emerges for inequality of lifetimes.³⁹ A motivation for this paper has been to emphasize the role of inequality in life years in overall social inequality. Accordingly, the main themes of this paper can be summarized by a history of this Mortality Gini⁴⁰ and a comparison with income Ginis.

In Sweden c.1750, where we have perhaps the best data for that time⁴¹, the Mortality Gini was .50. Over the next 150 years this figure declines slowly in what is now the developed world to something around .40. For example, England’s Mortality Gini is .378 in 1897. Thus until 1900 Mortality Ginis were hovering at or above the figures typical of income Ginis today. Over the next 50 years mortality inequality declined precipitously in all developed countries. For example, by 1952 England’s Mortality Gini had fallen to .132 – i.e., by well over half since the

³⁸ As nearly as we can tell the US Gini today is not much different from 100 years ago, when the developed world as a whole was much nearer to the peak of the Kuznets curve than it is today. So if we assume that, say, that the Swedish Gini of c. 1900 was somewhere in the .40s, didn’t decline at all until the welfare state developed its modern form and would not have otherwise declined subsequently, we get a halving of Sweden’s Gini from implementation of the welfare state.

³⁹ The necessary data for the calculation come from period life tables, which, recall, are not actual distributions of dates of death. Thus there is a hypothetical element in this Gini that is, in principle, absent from income Ginis.

⁴⁰ Specifically, the hypothetical egalitarian distribution for calculating a Mortality Gini is that all members of a birth cohort described by a particular life table die at the expected age of death for that life table. The Mortality Gini then is the share of the area below the cumulative egalitarian distribution not occupied by the actual cumulative distribution of age at death implied by the life table

⁴¹ The data are from Lutheran Church records that enable construction of life tables by single year of death.

turn of the century. In the last 50 years mortality inequality has continued to fall, but at a much more modest pace. So, England's contemporary (2002) Mortality Gini is .094, or 70 per cent of the 1952 figure. Figures 2B and 3B, with suitable translation of scales, will convey adequately the shape of this history.

This dramatic decline in mortality inequality in the 20th century has transformed a major source of social inequality into a minor one. Mortality Ginis today are only a fraction of even the lowest income Ginis. If we extend the inquiry to the less developed world, or look within developed countries, the variety of mortality inequality increases, but not by very much in historical context. Figure 13 illustrates these points. The first triplet of bars illustrates the variety of Mortality Ginis across US counties. King County in Washington, which includes the city of Seattle, is at roughly the 10th percentile of the population weighted distribution of Mortality Ginis and Duval County (Jacksonville), Florida is at the 90th percentile.⁴² Both are within 2 percentage points of the US average. The next triplet of bars shows three large less-developed countries, including the odd case of Russia where mortality inequality has not declined since the 1950s. They all have Mortality Ginis that are well above those in the developed world today, but none are anywhere near the values characterizing much of human history. As shown by the last bar, these three countries are at or below where the developed world was in the early 1930s – i.e., after the dramatic decline of mortality inequality had been well established. These countries' Mortality Ginis are also at or below the lowest contemporary income Ginis.

In much of today's world mortality inequality has become a comparatively minor potential source of further improvement in social equality. This reflects something about the

⁴² That is 10 percent of the US population lives in counties with Mortality Ginis at or below King's value and another 10 per cent live in counties at or above Duval's Gini.

nature of medical progress. All but the poorest countries find themselves well past the peak of the mortality Kuznets curve. Thus continued improvements in life expectancy translate into continued reduction of the remaining inequality. However vast reductions in infant mortality and the near elimination of scourges like tuberculosis have all but eliminated another source of improvement in mortality equality – the narrowing of the difference between the age of those who die prematurely and those who live the full measure of their years. The latter is a number like 85 years, and it hasn't changed much in the last 250 years. Over 80 percent of those born today can expect to live past 60; more than half will live past 80. But the vast majority of that half will be gone by 95. The difference between the average lifetimes of those who die early (say before 80) and those who do not is now on the order of 20 years, or half what it was when the 20th century began. Interestingly, if future progress does materially increase the full measure of our years that difference will widen and mortality inequality will reverse its historic downward course.

- *I thank Ivan Rivas, Wladimir Zanoni, Ricardo Estrada and Pablo Celhay for valuable research assistance.*

References

Preston, Samuel H “The Changing Relation Between Mortality and Level of Economic Development” *Population Studies* July, 1975 29:2, 231-48

Becker, Gary, Tomas Philipson and Rodrigo Soares, “The Quantity and Quality of Life and the Evolution of World Inequality,” Working Paper 192, George Stigler Center for the Study of the Economy and the State, University of Chicago, May 2004

Cutler, David, Angus Deaton and Adriana Lleras-Muney, “The Determinants of Mortality,” *J of Econ Perspectives*, Summer 2006, 20:3, 97-120.

Cutler, David and Adriana Lleras-Muney, “Education and Health: Evaluating Theories and Evidence,” Working Paper 12352, National Bureau of Economic Research, June 2006.

Gove, Walter R. “Sex, Marital Status and Mortality” *American Journal of Sociology*, July, 1973 79:1, 45-67

Kuznets, Simon. *Shares of Upper Income Groups in Income and Savings*. New York, 1953: National Bureau of Economic Research

US Bureau of the Census. *Historical Statistics of the US*. Washington, 1975

Sweden Central Bureau of Statistics. *Income Distribution Survey 2004* (Report HE21SM0601). Stockholm, 2006.

Table 1. Progress in Inequality and Expected Life from 1850-1950.

Distribution by Half-century

	Percentage of Total Progress Completed by c.1900	
Measured at	Inequality (SD of ln yrs)	Expected Life (Years)
Birth	15.4%	32.2%
Age 5	24.6	31.6

Note: Progress is defined as the total decline in inequality or the total increase in expected life over the 1850-1950 period. These totals are averages across the 10 countries for which data over this century are available. These are listed in Figure 1. The averages are obtained from panel regressions of the form:
 $y(it) = a + b(j) \cdot T + \text{fixed effects}$
 where
 $y(it)$ = the inequality or expected life variable for country i in year t
 T = time indicator
 $b(j)$ = a slope specific to period j , $j=1852-1902, 1902-1952$.
 Fixed effects are included for country and country-specific deviations from trend for the observations involving world war 1 and the spanish flu epidemic and world war 2. The sample period is 1852-1952.

The figures in the table show what percentage of the estimated total change from 1852-1952, as implied by the $b(j)$ coefficients, occurred in 1852-1902.

Table 2. Selected Mortality Data. England, 1842-2002

Year	Percentage Surviving Past 80		Average Age at Death of Those Dying			Ratio of Old to Young	
	of Births	if Alive at 5	Before 80		After 80	Life Expectancy	
			All	Alive at 5			
	(1)	(2)	(3)	(4)	(5)	(5)/(3)	(5)/(4)
1842	9.6%	13.0%	36.5	50.7	85.0	2.33	1.68
1902	11.9	15.1	43.2	56.3	84.6	1.96	1.50
1952	30.1	31.1	62.2	65.1	85.3	1.37	1.31
2002	54.8	55.2	66.3	67.2	87.7	1.31	1.31

Note: Data from life tables for England and Wales. Year is mid-point of 5 year interval.

Table 3. Summary Statistics. Mortality and Life Expectancy. US Counties. 1972, 1982, 2002

	1972		1982		2002	
	Mean	SD	Mean	SD	Mean	SD
A. Levels						
1. Expected Life						
at birth	70.8	2.1	74.0	1.5	76.9	1.9
at age 50	25.9	1.1	27.8	1.0	30.0	1.3
2. Inequality measures						
sd of log age (at birth)	.66	.07	.55	.05	.45	.05
sd of log age (at age 5)	.30	.03	.27	.02	.25	.02
3. Mortality below age 50						
prob death by age 5	.022	.007	.014	.003	.008	.003
prob die by 50/age 5	.090	.024	.067	.015	.057	.016
B. Changes	1972-2002		1982-2002			
1. Expected Life						
at birth	5.9	1.6	2.8	1.1		
at age 50	3.9	1.1	2.1	0.8		
2. Inequality measures						
sd of log age (at birth)	-.20	.06	-.10	.05		
sd of log age (at age 5)	-.04	.02	-.02	.02		
3. Mortality below age 50						
prob death by age 5	-.013	.006	-.005	.003		
prob die by 50/age 5	-.031	-.017	-.008	.010		
C. Selected Correlations	1972	1982	2002			
Expected life at birth and						
sd of log age (at birth)	-.75	-.79	-.89			
sd of log age (at age 5))	-.65	-.68	-.85			
expected life in 2002	.70	.82				

Note: All variables are weighted by population. (Changes are weighted by 1982 population.) All estimates are based on estimated ages at death of the cohort described by a county life table. The county life tables are based on mortality averaged over the 5 years centered on the year indicated. See appendix for details. N=3114 counties.

Table 4. Summary Statistics. Socio-Economic Data. US Counties, 1970, 1980,

Variable	1970		1980		2000	
	Mean	SD	Mean	SD	Mean	SD
Family Income						
Mean (\$000)	11.6	2.56	23.4	4.34	66.1	17.73
<i>coeff of variation¹</i>		22.0%		18.6%		26.8%
SD of log within county	.81	.07	.80	.08	.85	.10
Education						
Mean (years)	10.7	.94	11.7	.89	13.0	.74
SD within county	3.6	..39	3.5	.41	3.2	.52
Family structure						
female headed hh (%)	10.9	3.54	14.4	5.21	17.5	6.00
Race/Ethnic (%)						
black	11.2	12.24	11.7	12.77	12.3	13.19
american indian	0.4	2.31	0.7	2.73	1.0	3.27
hispanic²	4.6	8.73	10.5	17.30	12.5	15.10

Notes:

1. standard deviation as per cent of mean across counties

2. "Spanish origin" population in 1970

Means and standard deviations are weighted by number of families(income, family structure), population over 25 (education) or population (race/ethnic).

N= 3113 to 3116 counties

Table 5. Mortality at the Top and Bottom of the Geographic Distribution. US Counties 1972, 1982, 2002

	1972			1982			2002		
Variable	Lowest Decile	Highest Decile	Difference or Ratio	Lowest Decile	Highest Decile	Difference or Ratio	Lowest Decile	Highest Decile	Difference or Ratio
<i>A. Counties with Highest and Lowest Expected Life, 1972</i>									
Expected Life	67.40	73.20	5.8	71.6	75.8	4.2	74.1	79.3	5.2
prob death by age 5	.029	.016	0.55	.018	.011	.61	.012	.006	.53
prob die by 50/age5	.128	.065	0.51	.089	.050	.56	.08	.04	.48
exp life beyond 50	24.3	27.0	2.7	26.4	28.8	2.4	28.3	31.5	3.2
SD log age at birth	.76	.58	-.18	.62	.51	-.11	.52	.40	-.12
SD log age at 5	.33	.27	-.06	.29	.25	-.04	.28	.22	-.06
Family income (\$000)	9.2	14.6	1.59	19.4	29.2	1.51	55.4	88.2	1.59
Mean yrs education	9.6	11.9	2.3	10.7	12.8	2.1	12.5	13.8	1.3
<i>B. Counties with Highest and Lowest Family Income, 1972</i>									
Expected Life	68.8	72.6	3.8	73.0	75.3	2.3	74.9	79.2	4.3

Note: In panel A. "Lowest decile" defined as the population living in counties with "predicted expected life" in 1972 at or below the 10th percentile. "Highest decile" is defined similarly for population at or above the 90th percentile. Predicted expected life comes from the regression in column (2) of Table 6 (see text for discussion). These two groups of counties, defined by their position on the predicted expected life distribution in 1972, are then followed for 1982 and 2002. The data in the table are population weighted means for these two fixed samples of counties. For panel B. counties are defined according to their position on the 1972 average family income distribution (i.e., "lowest decile" means the 10 per cent of the population residing in counties with the lowest average family income).

Table 6. Expected Life at Birth. Descriptive Regressions. US Counties. 1972, 1982, 2002

Variable	Regression Coefficients						Beta Coefficients ¹		
	1972		1982		2002		1972	1982	2002
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Education:									
Mean yrs	.61	.56	.73	.72	1.03	1.19	.25	.42	.46
SD within county	.80	.67	.58	.62	1.02	1.10	.12	.17	.30
Income									
Average	-.02	.001	-.002	-.002	.03	.01	.01	-.01	.13
SD log within county	-1.20	-2.46	-3.66	-2.70	-5.45	-3.31	-.09	-.14	-.17
Population characteristics									
female head hh (%)	-.20	-.23	-.09	-.15	-.03	-.16	-.39	-.51	-.51
black (%)	-.07	-.04	-.04	-.02	-.05	.004	-.25	-.18	.03
american indian (%)	-.10	-.08	-.06	-.04	-.04	-.01	-.09	-.12	-.02
hispanic (%)	.002	.02	.01	.02	.03	.06	.08	.24	.44
log population	.26	.19	-.03	-.02	.04	.01	.15	-.02	.01
State fixed effects	no	yes	no	yes	no	yes	yes	yes	yes
Summary statistics									
Adj R ²	.61	.69	.77	.83	.77	.86			
Standard error estimate	1.28	1.14	.73	.63	.93	.72			
N	3110	3110	3113	3113	3112	3112			

Notes:

Regressions are weighted by population. All coefficients are significant at $p < .05$ except for those in small font.

Dependent variable is expected life years at birth from a life table based on mortality over the five-year period centered on the indicated year. Independent variables are from Census data in the first year of this five year period. See Appendix for more details.

1. The change, in standard deviation units, in expected life from a one standard deviation increase in the variable. (For example, in 1972 a one standard deviation increase in mean years of education is associated with a .25 standard deviation increase in expected life.) Based on the regression with state fixed effects.

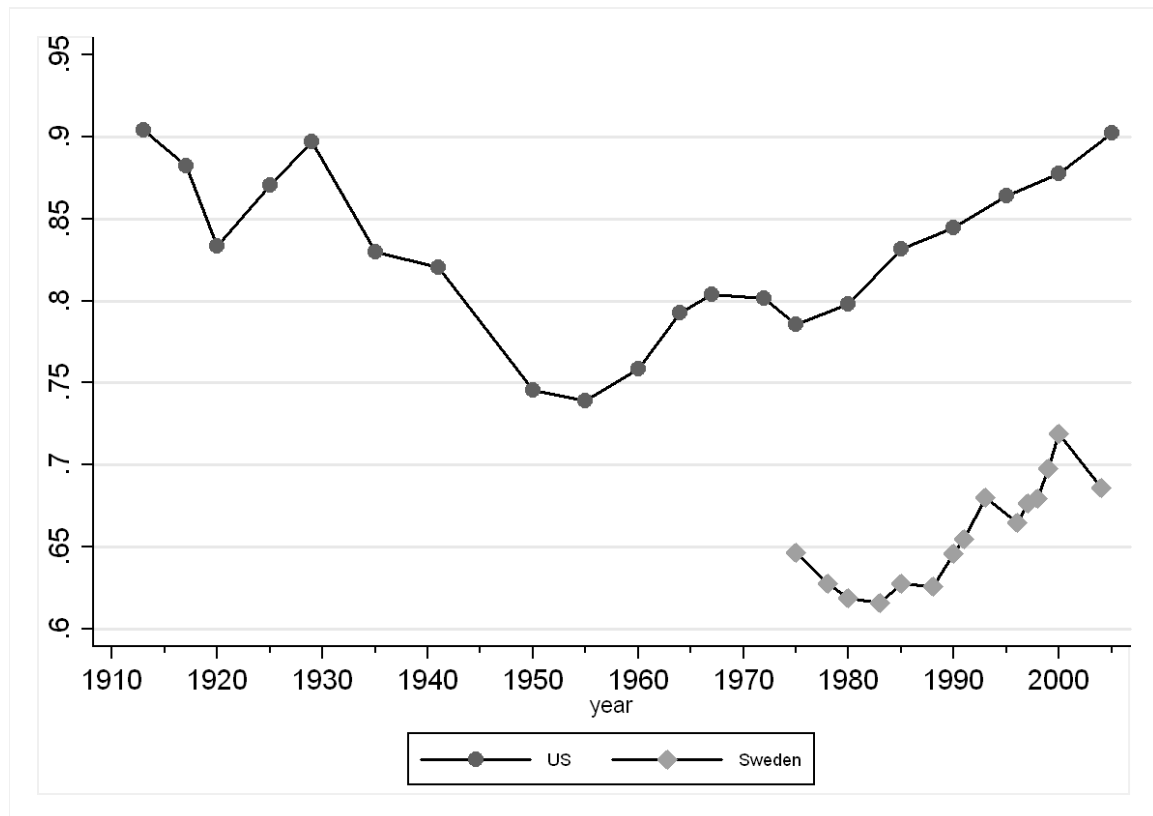
Table 7. Mortality Inequality within US Counties. Beta Values, 1972 and 2002

Independent Variable	Beta Coefficients ¹			
	SD of ln		SD of ln	
	1972	2002	1972	2002
Education:				
Mean yrs	-.18	-.18	-.63	-.51
SD within county	-.09	-.19	-.42	-.29
Income				
Average	-.07	-.33	.31	-.15
SD log within county	.18	.29	.49	.41
Population characteristics				
female head hh (%)	.30	.04	.10	.05
black (%)	.22	.34	.35	.21
american indian (%)	.04	.03	.09	.09
hispanic (%)	-.03	-.26	.06	-.29
log population	-.23	-.02	-.70	-.11
Summary statistics				
Adj R2	.65	.70	.65	.79
Standard Error Est	.043	.029	.019	.0105

Notes: Based on coefficients from regressions with state fixed effects.
Dependent variables are standard deviations of log age of death for all births and for survivors to age 5 respectively.
Regressions are weighted by population. All coefficients are significant at $p < .05$ except for those in small font.

1. The change, in standard deviation units, in mortality inequality, from a one standard deviation increase in the variable.

Figure 1. Standard Deviation of Log Household Income. US and Sweden. Estimates from Income Distributions. Various Years 1913-2005



US :

1929-2005: estimated from quintile (plus top 5 per cent) shares of household money income in Bureau of the Census (1975), p 301 and <http://www.census.gov/hhes/www/income/histinc/ie3.html>

Shares are assumed to be allocated evenly within quintiles or sub-quintiles.

1913-29: estimated from income share of top 1 per cent of income recipients in Kuznets (1953) p. 582. Kuznets estimate is available for 1913-1948. For the overlapping observations in 1929-48 the estimated standard deviation of log income is regressed on the Kuznets top 1 per cent share ($r \sim .99$). The pre 1929 estimates shown here are fitted values from this regression.

Sweden:

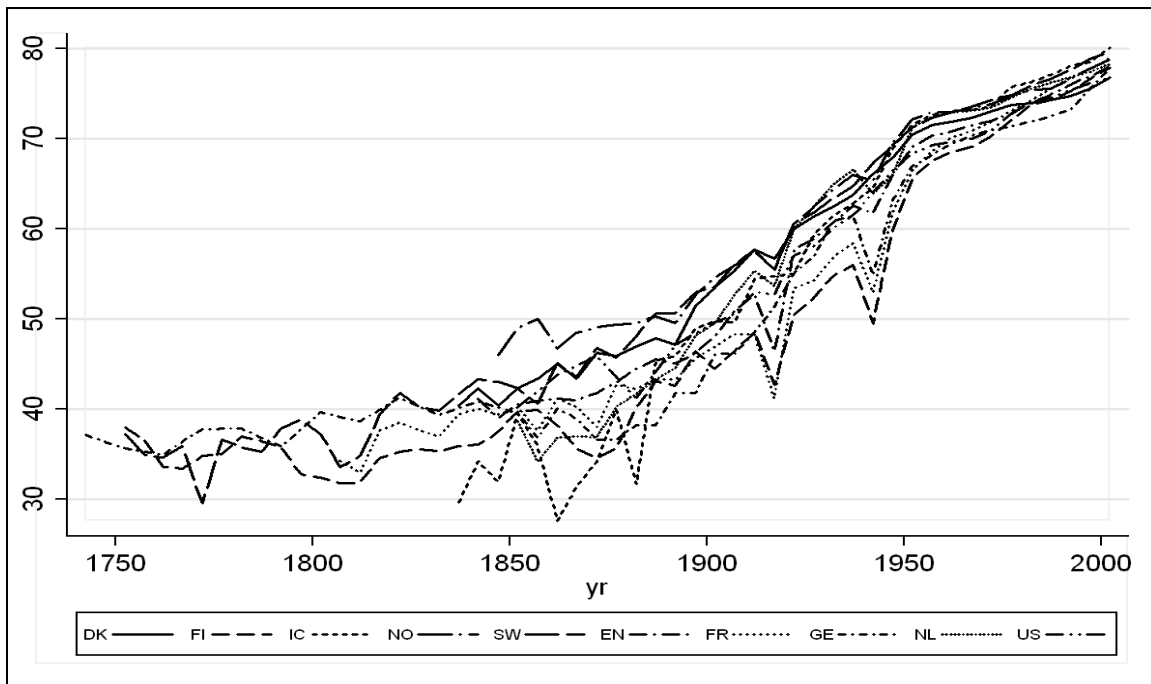
1991-2004: estimated from decile (plus top 5 per cent) mean household incomes in Sweden Central Bureau of Statistics (2006) p 14. All quintile or sub-quintile members are assigned the group average income.

1975-1991: The above source gives Gini coefficient and top 10 per cent shares of the household income distribution for 1975-2004 (p 18). These are used in regressions with the post 1991 estimates of standard deviation of log income as the dependent variable. Pre 1991 estimates here are the fitted values from this regression.

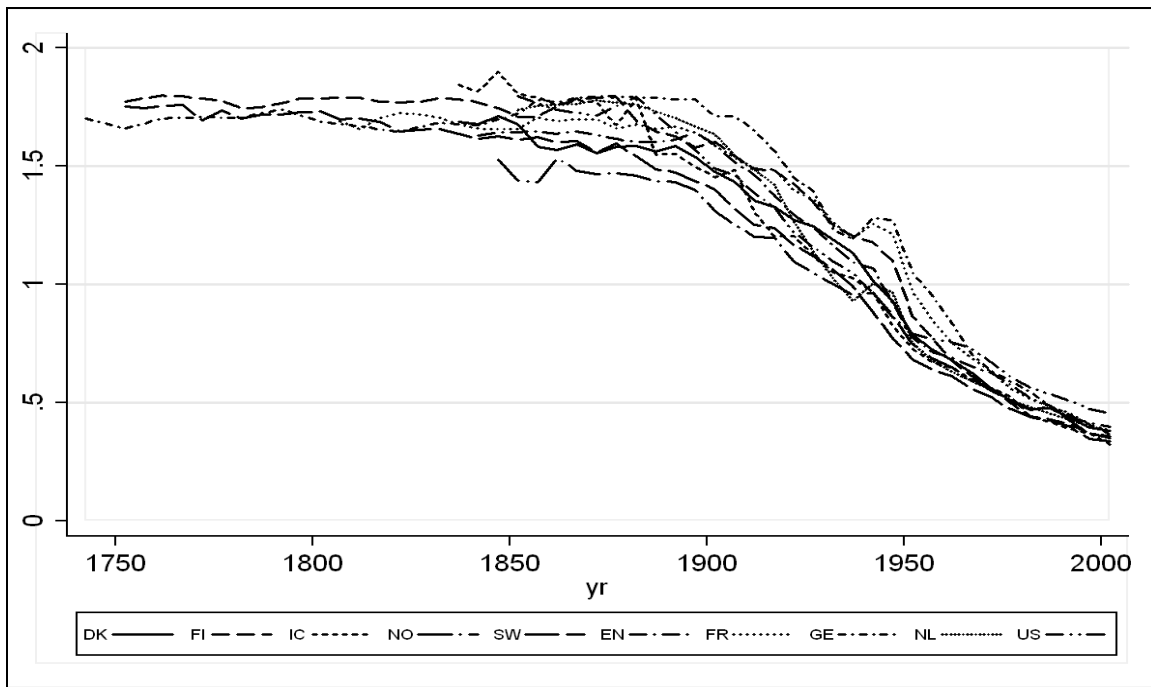
The US income concept is pre-tax but post-explicit transfers. The Swedish concept is disposable income after both taxes and transfers.

Figure 2. Expected Life and Inequality of Life at Birth. Ten Countries. 1742-2002

A. Expected Years of Life at Birth



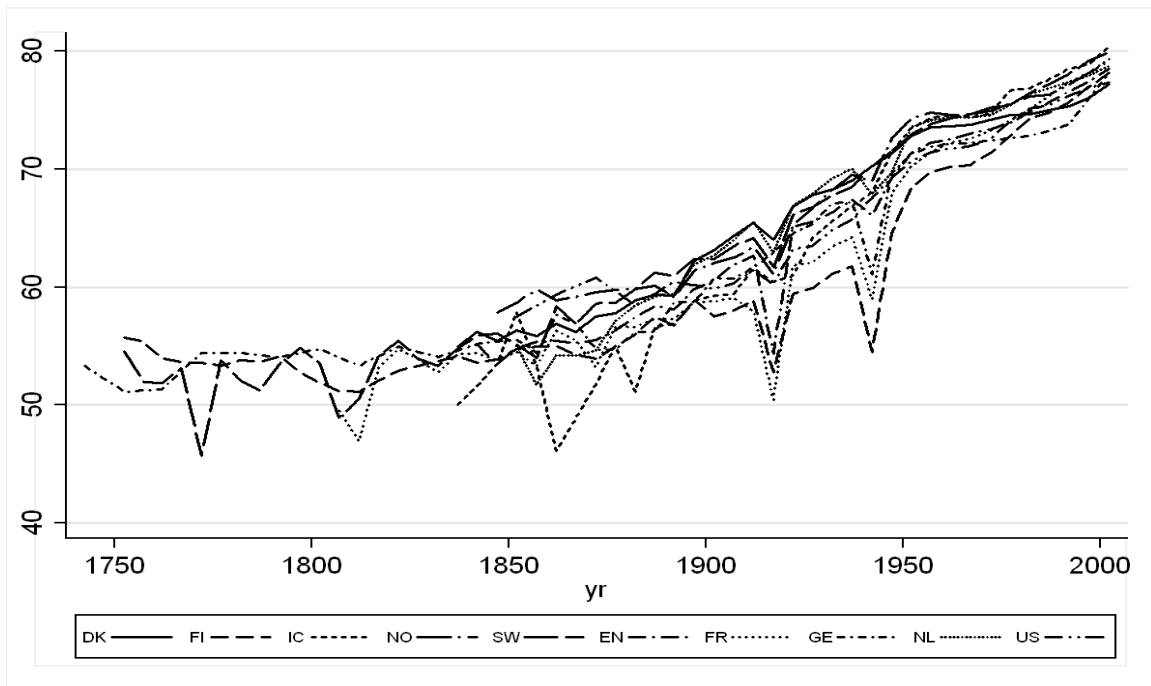
B. Standard Deviation of Log Years Lived. All Births



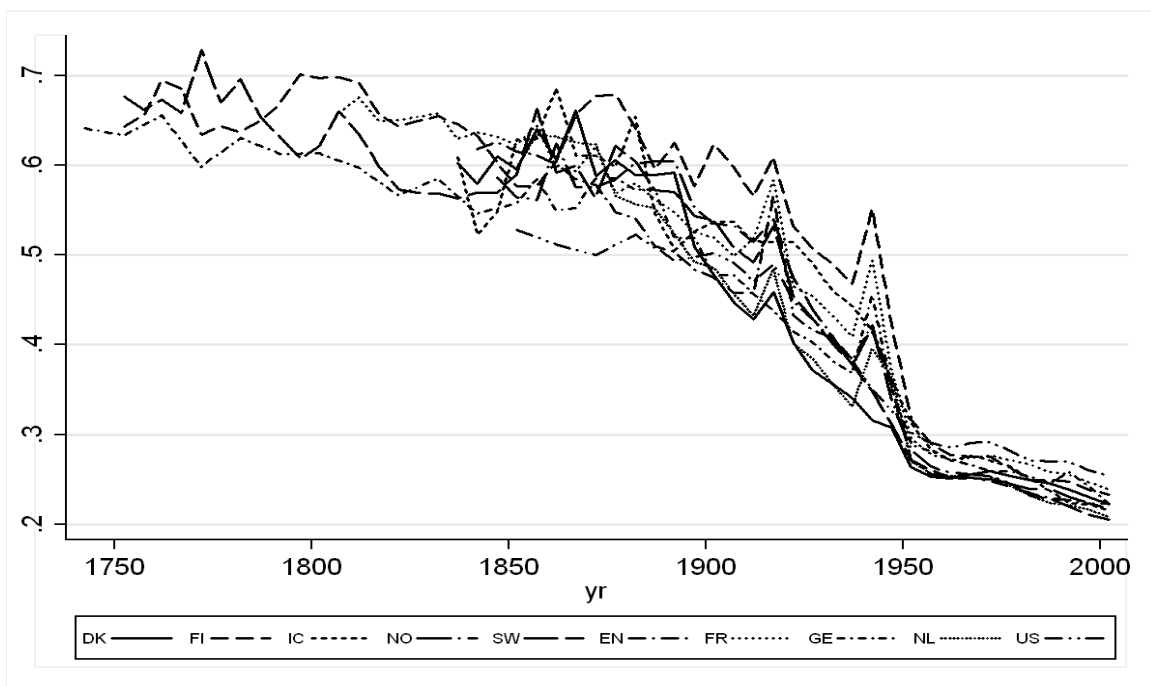
Countries[first decade with data]: Denmark[1830], Finland[1750], Iceland[1830], Norway[1840], Sweden[1750], ENgland[1840], FRance[1800], GERmany[1740], NetherLands[1850], United States[1850]

Figure 3. Expected Life and Inequality of Life at Year 5. Ten Countries. 1742-2002

A. Expected Years of Life. Survivors to Year 5



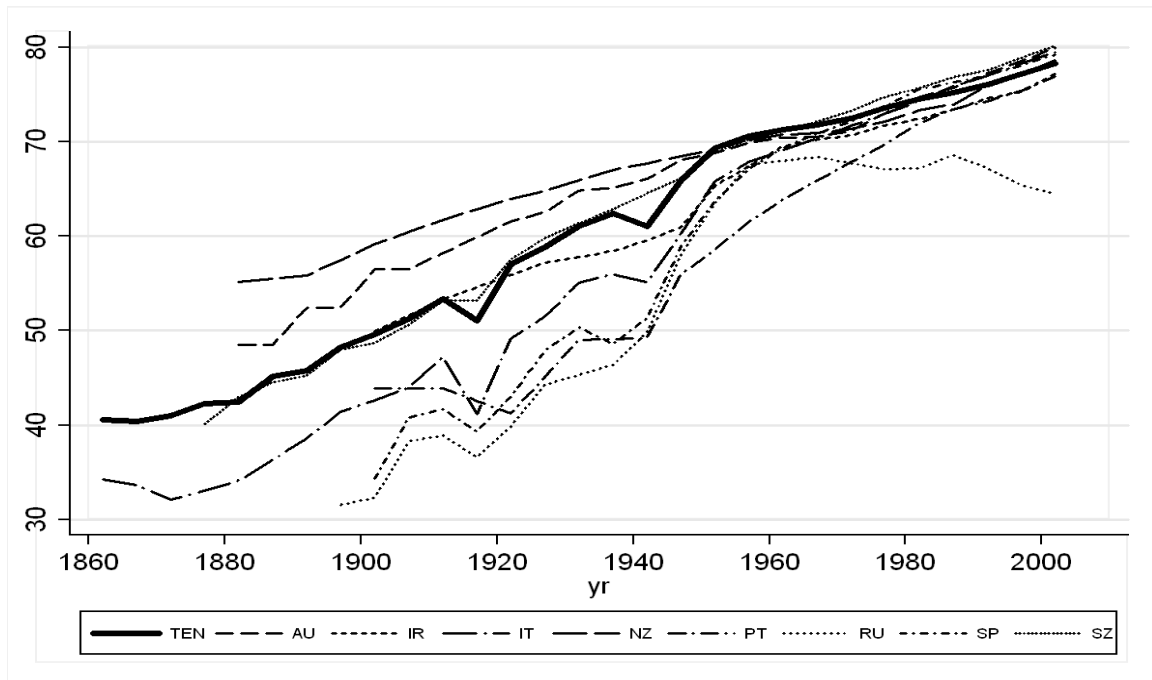
B. Standard Deviation of Log Years Lived. Survivors to Year 5



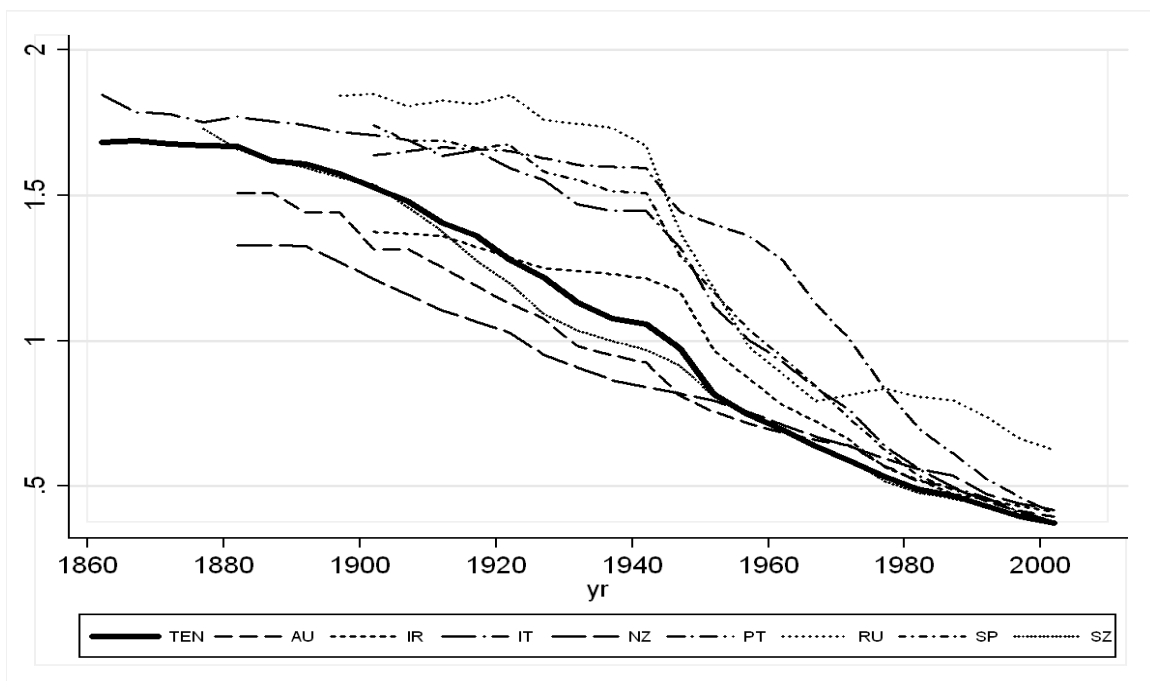
Countries: See note to Figure 2

Figure 4. Expected Life and Inequality of Life at Birth. Eight Countries. Europe and Australasia. 1860-2002

A. Expected Life at Birth



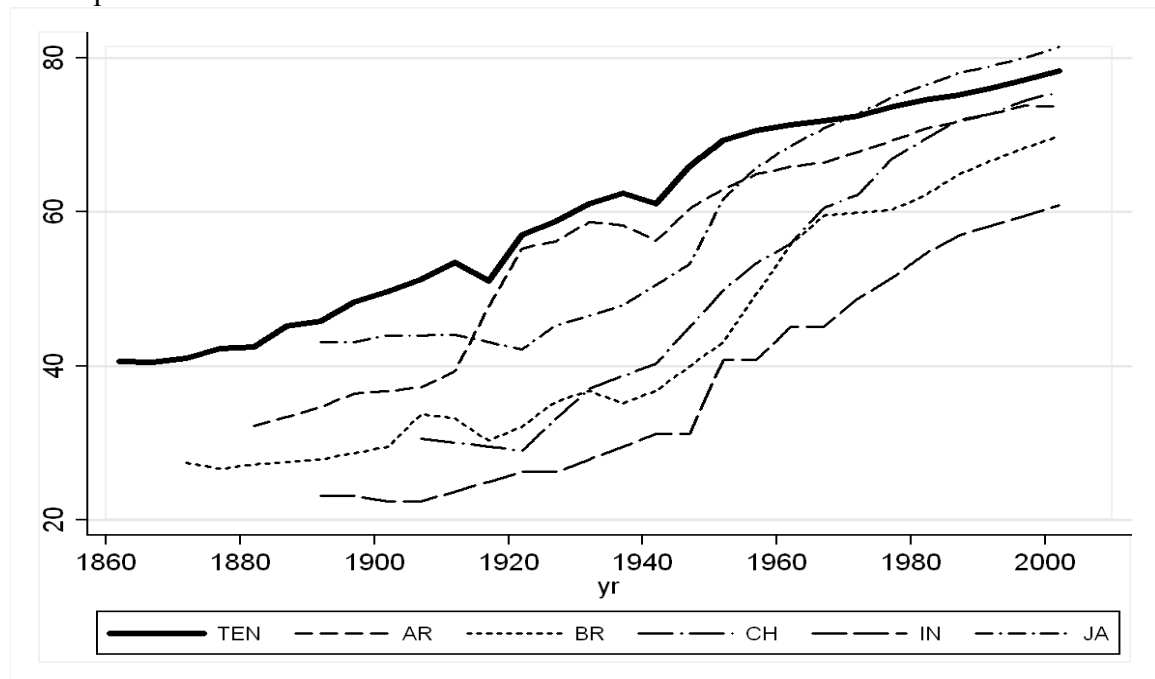
B. Standard Deviation of Log Years Lived. All Births



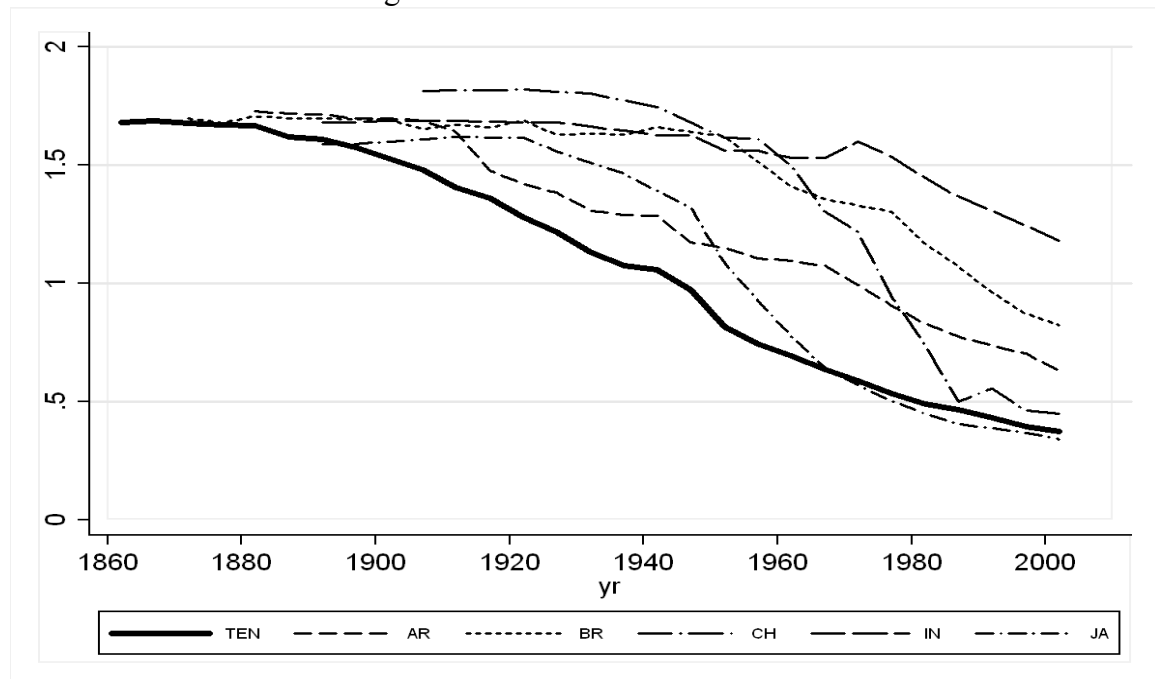
Countries: TEN countries in Figure 2 and 3 (average) AUstralia, IReland, ITaly, New Zealand, PorTugal, RUssia, SPain, SwitZerland

Figure 5. Expected Life and Inequality of Life at Birth. Five Countries. Latin America and Asia. 1860-2002

A. Expected Life at Birth



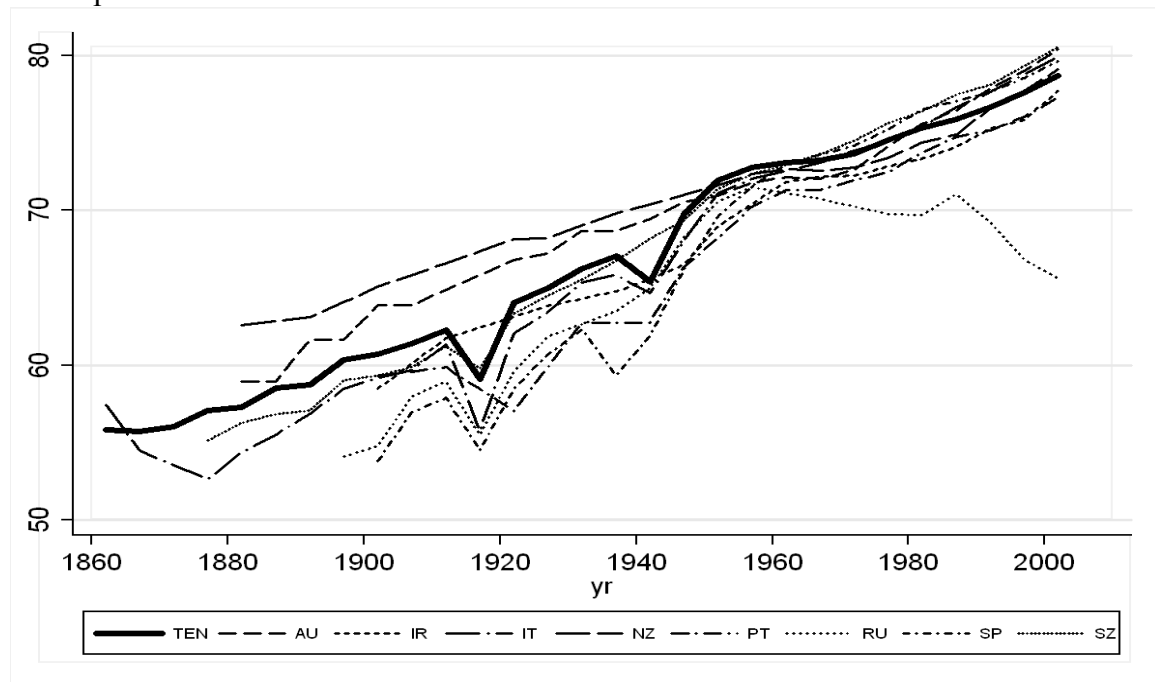
B. Standard Deviation of Log Years Lived. All Births



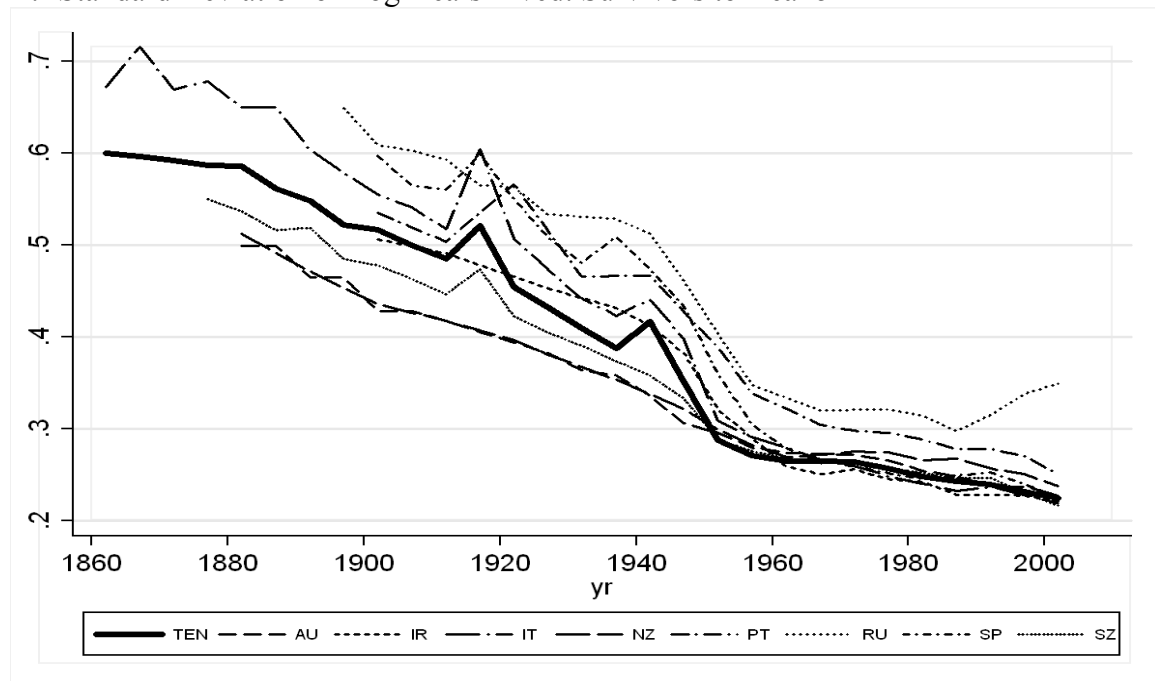
Countries: TEN countries in Figure 2 and 3 (average) ARgentina, BRazil, CHile, INdia, JApan

Figure 6. Expected Life and Inequality of Life at Year 5. Eight Countries. Europe and Australasia. 1860-2002

A. Expected Years of Life. Survivors to Year 5



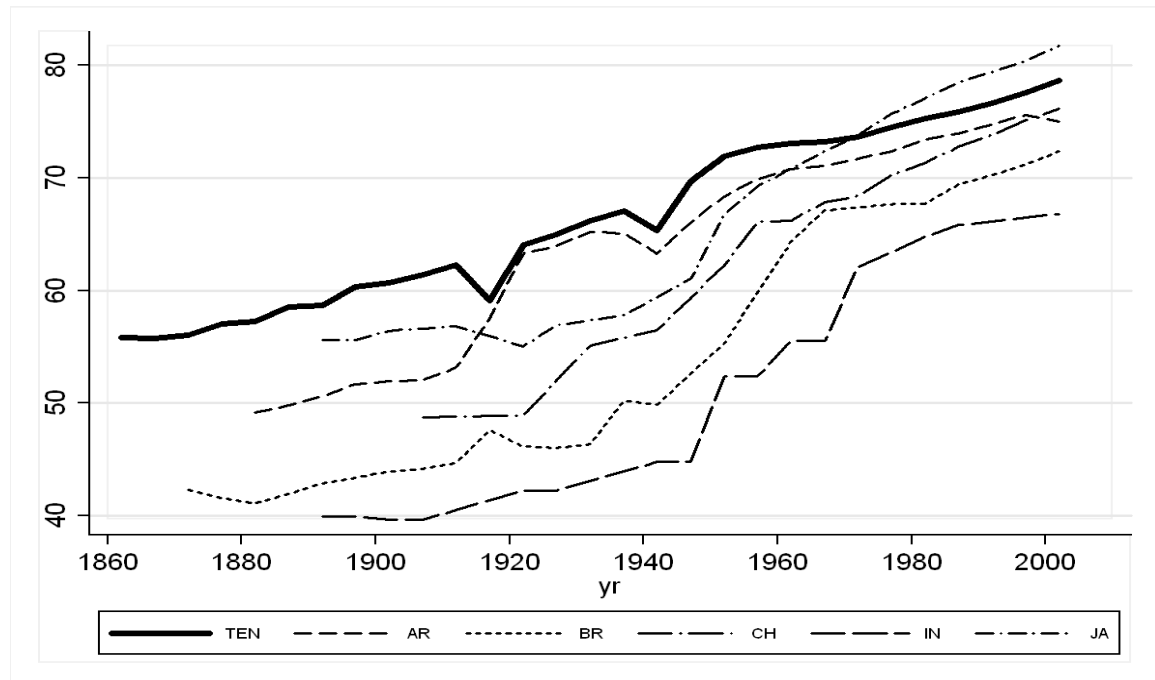
B. Standard Deviation of Log Years Lived. Survivors to Year 5



Countries: see note to Figure 4.

Figure 7. Expected Life and Inequality of Life at Year 5. Five Countries. Latin America and Asia. 1860-2002

A. Expected Years of Life. Survivors to Year 5



B. Standard Deviation of Log Years Lived. Survivors to Year 5

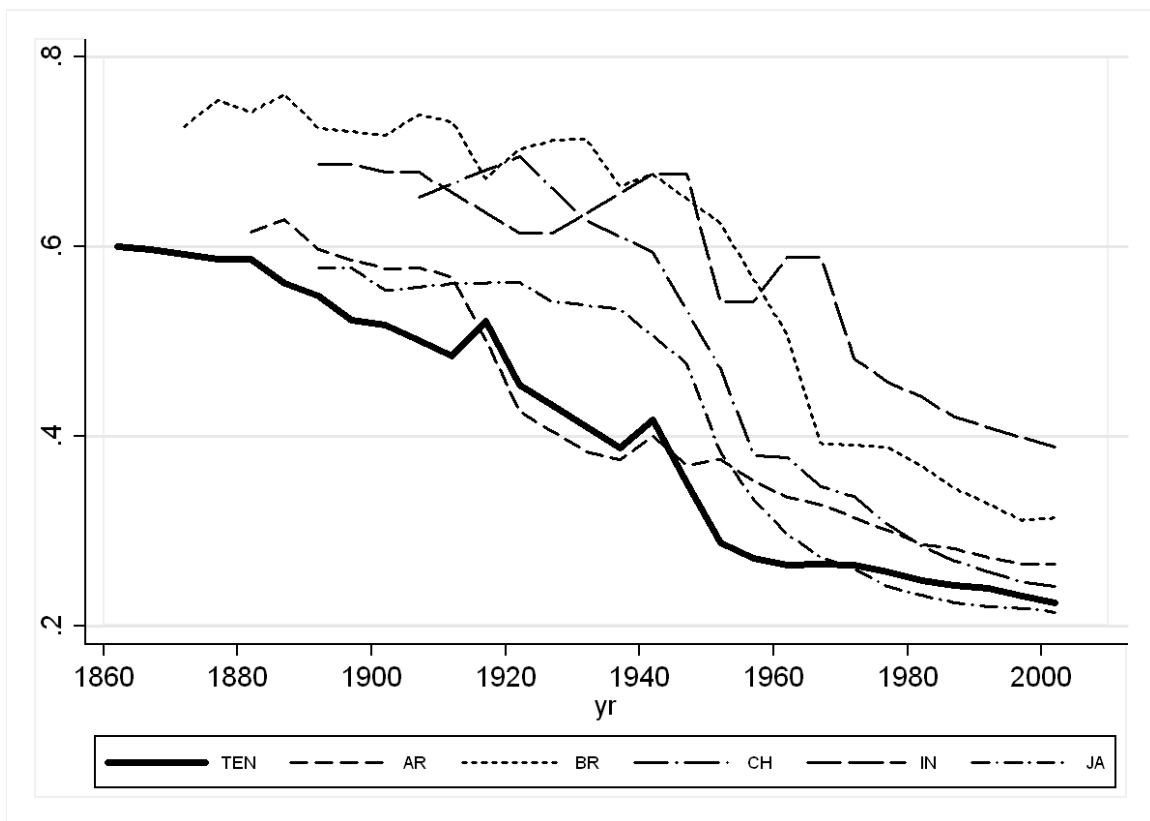
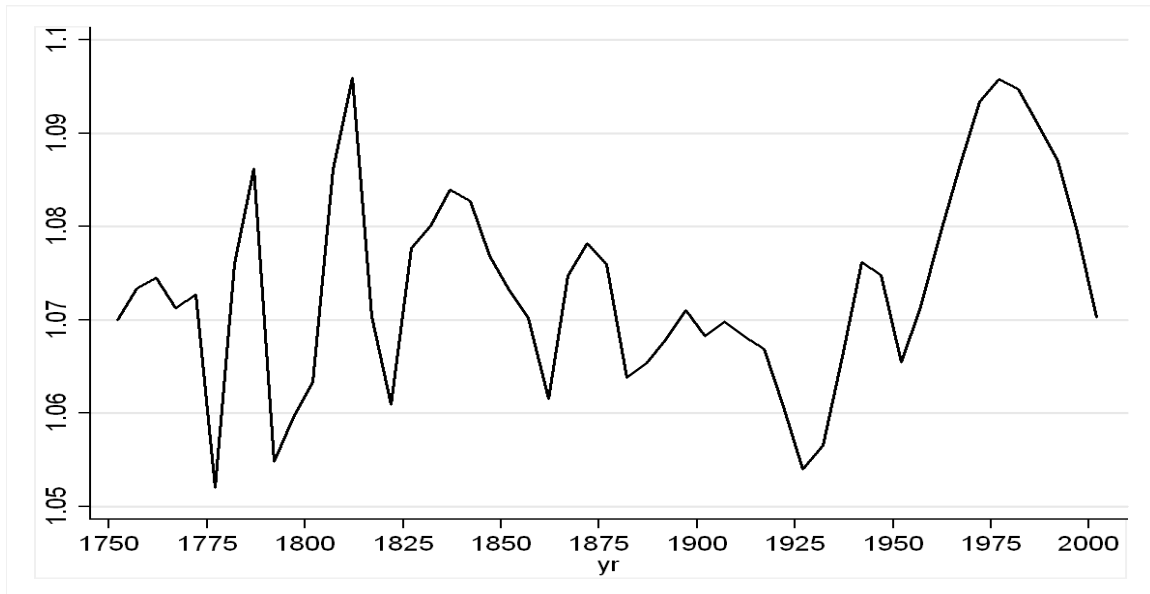


Figure 8. Female to Male Life Expectancy Ratio at Birth. Mean and Standard Deviation. Ten Countries. 1750-2002.

A. Mean Ratio of Female to Male Years of Life at Birth



B. Standard Deviation of Ratio Across Countries



Note: Data are for the 10 countries in Figure 2. Panel A shows the mean across the countries and Panel B. shows the standard deviation across the countries in the sample in any year. There are 3 countries in the sample at 1750 and 10 from 1850 onward. Country years affected by war are treated as missing observations and then interpolated (see text).

Figure 9. Female to Male Life Expectancy Ratio. Post World War II Period. England, France, Germany and US

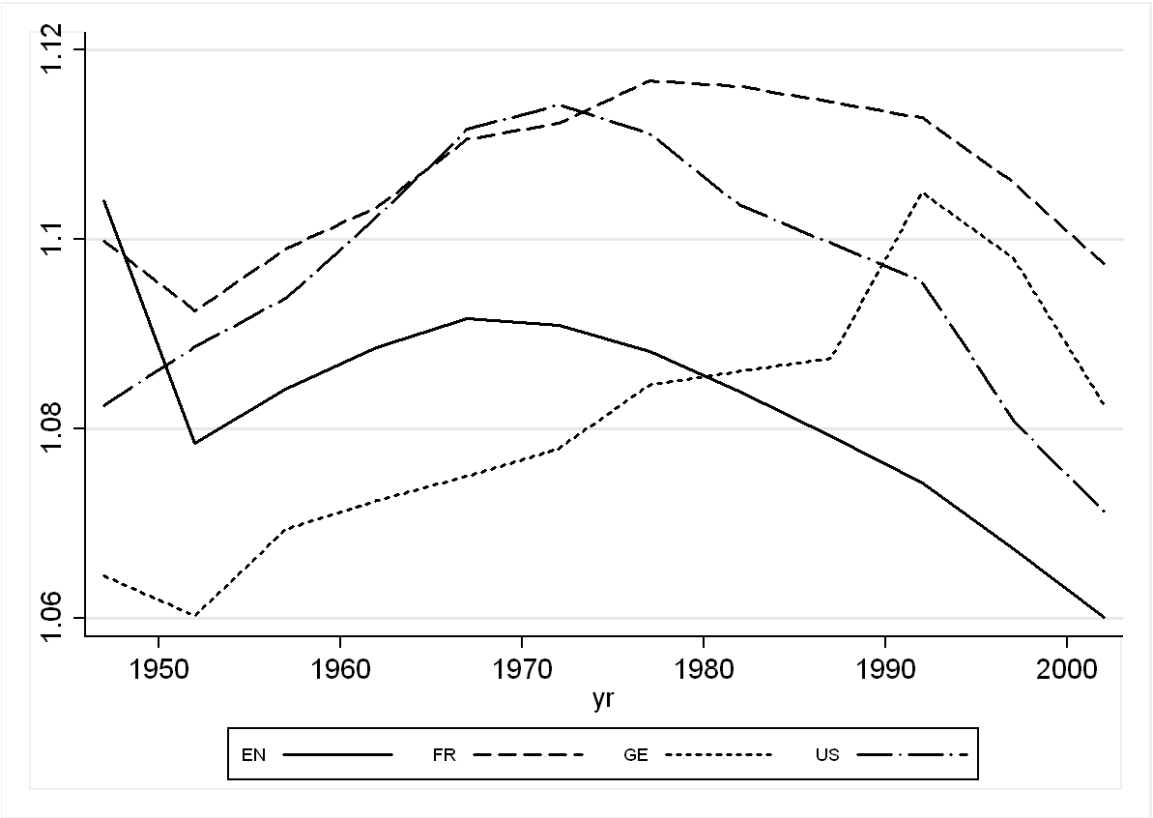


Figure 10. Gender Mortality Differences by Stage of Life. Means. Ten Countries. 1750-2002

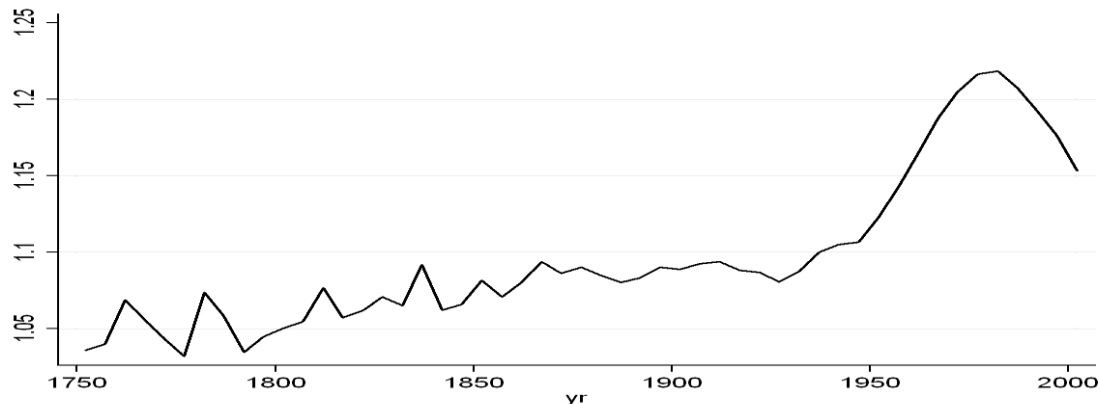
A. Male to Female Ratio. Probability of Death by Age 5



B. Male to Female Ratio. Probability of Death by Age 50 for 5 year olds



C. Female to Male Ratio. Years of Life Expected After age 50



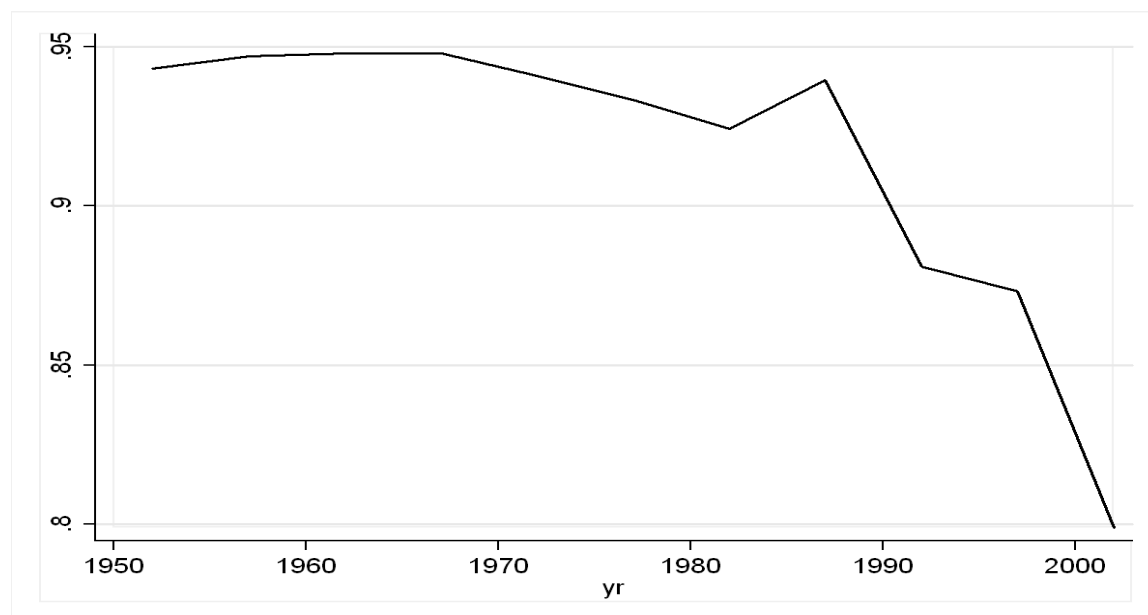
Note: See note to Figure 2 for countries.

Figure 11. Life Expectancy in Russia/Soviet Union Relative to 18 OECD Countries. By Gender. 1897-2002

A. Ratio of Years of Life Expected at Birth. Russia/OECD, 1897-2002



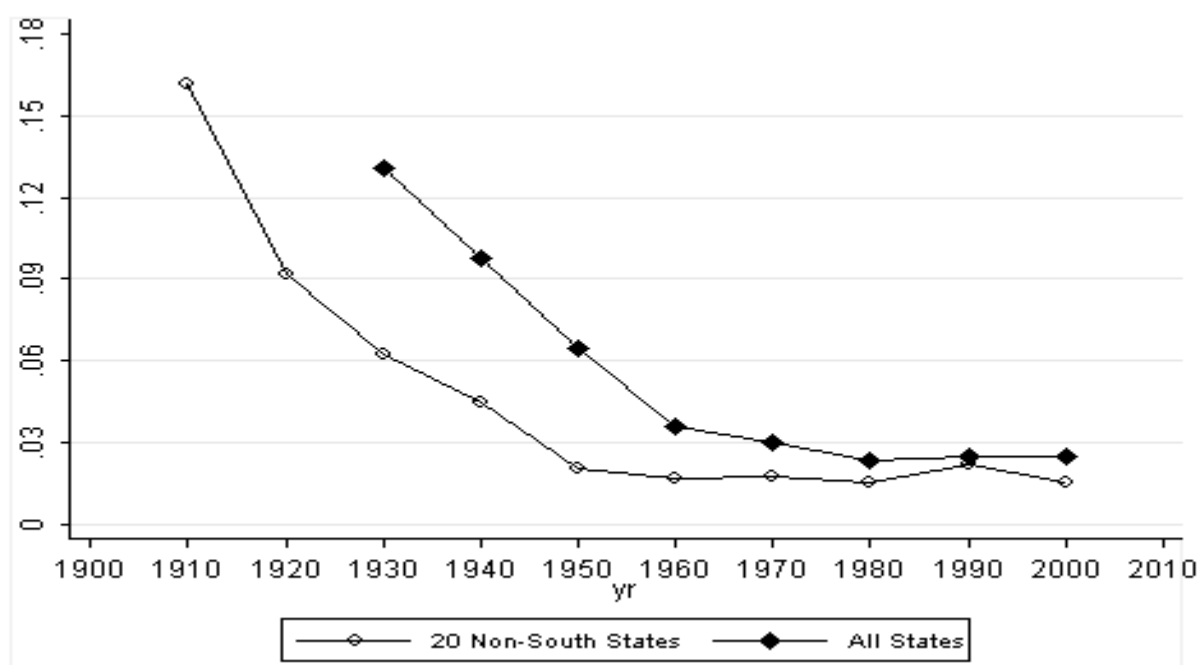
B. Male/Female Years of Life Expected at 50. Russia Ratio/OECD Ratio. Since 1950.



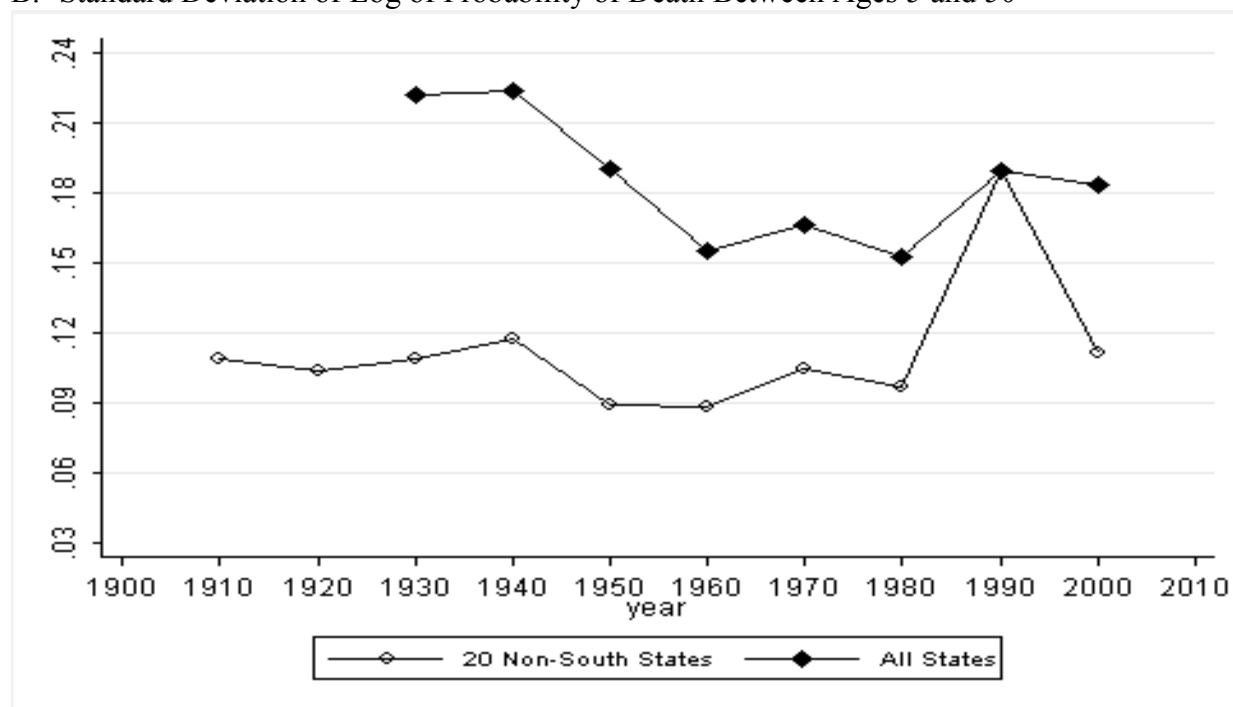
Note: All series for Russia/Soviet Union are divided by corresponding average for 18 OECD countries with data spanning the 20th century. These countries are the ten shown in the note to Figure 2 plus Australia, Ireland, Italy, Japan, New Zealand, Portugal, Spain and Switzerland.

Figure 12. Dispersion Across US States. Expected Life and Mortality. 1910-2000

A. Standard Deviation of Log of Expected Years of Life at Birth

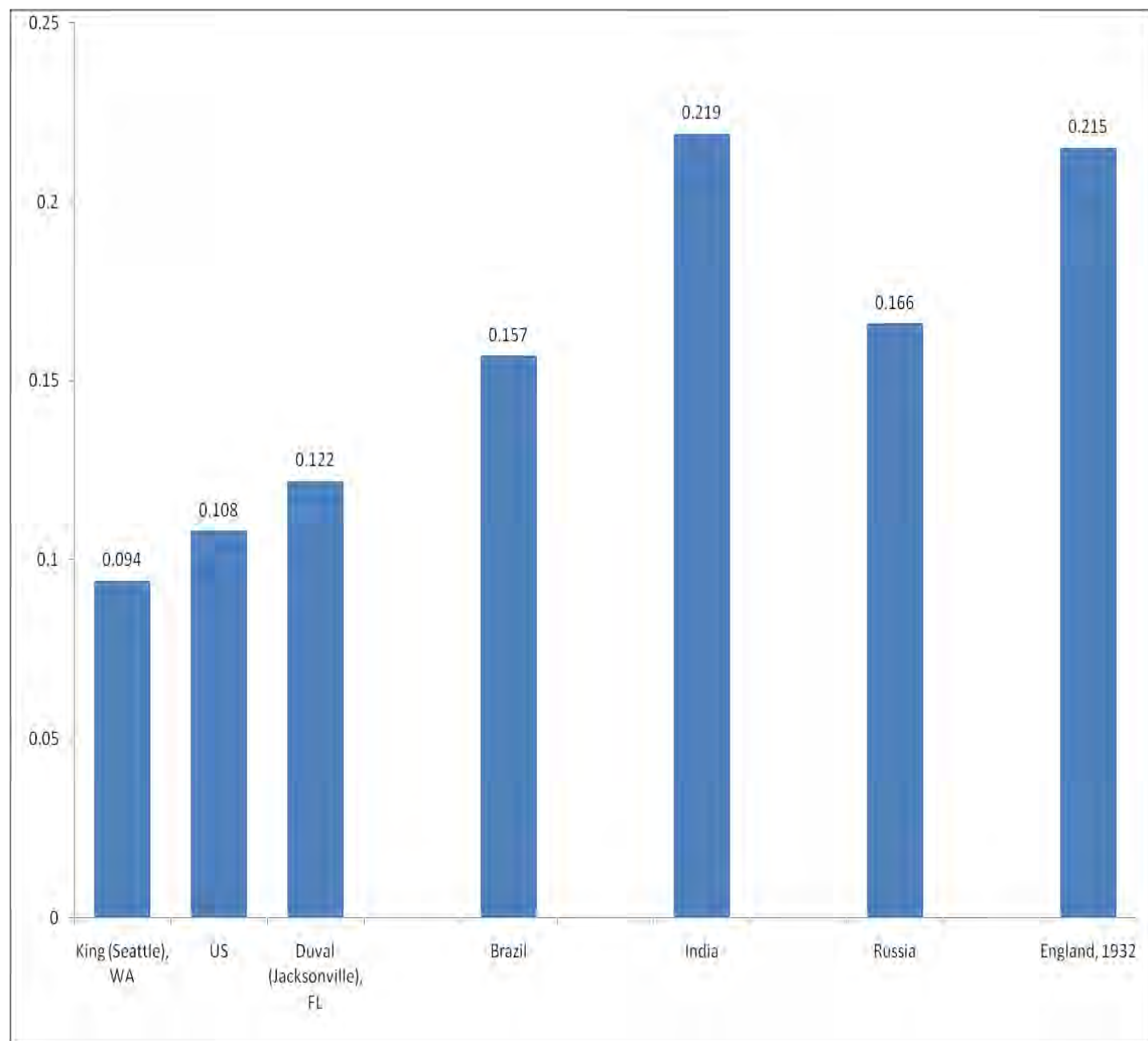


B. Standard Deviation of Log of Probability of Death Between Ages 5 and 50



Note: The 20 non-South states are: CA, CO, CT, IN, MA, MD, ME, MH, MN, MT, NH, NJ, NY, OH, PA, RI, UT, VT, WA, WI. These had death registration systems by 1910.

Figure 13. Mortality Ginis for Selected Populations



Note: Bar height is the Gini ratio for the distribution of life years in the indicated population c.2002 (except England in 1932, which is shown for comparison). The distribution is taken from period life tables. The two US counties are at approximately the 10th (King) and 90th (Duval) percentile of the population weighted distribution of mortality Ginis for all US counties.

Appendix. Data Construction and Sources

All estimates of life expectancy and its dispersion in this paper come from two kinds of related data: life tables or mortality rates. A life table lists values of the survival function

$$S(T) = 100,000 \prod_{t=0}^T s(t) \quad (\text{A1})$$

where,

$S(T)$ = number of survivors to the beginning of year T out of a cohort of 100,000 births at the start of $t=0$

$s(t)$ = probability that an individual will survive from the beginning to the end of year t

That is, the life table shows how many are expected to survive to each age out of hypothetical birth cohort of 100,000. The function is assumed to go to zero at an upper limit which is usually around 110 years. Mortality rates give the probability of dying in the next year as a function of age, and this is related to life table data because

$$s(t) = 1 - m(t) \quad (\text{A2})$$

where,

$m(t)$ = probability of dying over the next year for an individual t years old

So, the life table can be thought of as giving a stock of survivors at each age that depreciates at the mortality rate for that age. There are some conceptual and practical differences between life table and mortality measures, but for present purposes I use the two interchangeably.¹

The most commonly available life table is the period life table, which is based on contemporaneous mortality experience. I use period life tables exclusively here. The less common cohort life table follows a birth cohort through its subsequent history. The dominant reason for using period life tables is practical: a reasonably complete cohort life table of today would be tracking a cohort born in the early 20th century, so it would discard much of the available mortality information for subsequent birth cohorts.

While period life tables are often described in forward-looking terminology, they are really a snapshot of current mortality experience. Thus, the oft-cited life expectancy at birth statistic is not an estimate of how long infants born today can reasonably expect to live. It is just a summary statistic of current mortality experience. Specifically, expected life is the mortality weighted average age of death for the current population. This is an expected value for a counterfactual in which infants born today will experience the same mortality risks over their lifetimes as do the various age groups in the current population. Thus there is no allowance for medical progress, which is a historically important shortcoming of life expectancy measures if they are treated as predictions.

¹ For example, mortality rates are measured as deaths of individuals of age X in calendar year Y divided by estimated population of age X on June 30. Life tables are supposed to give the probability of dying by December 31 of Y for those alive on January 1. The simple relation in (A2) may not hold if deaths are bunched within the year. This is the case with infant mortality, where most deaths occur within a few days of birth. For this reason, mortality data are often adjusted for purposes of constructing life tables, and there may be a discrepancy between published mortality rates and the rates implied by published life tables.

A. International Comparisons

The data for the cross-country samples come from online databases that cover many countries and from print and electronic sources specific to single countries. These sources are listed in Table A1. I took available data for the country/years in my sample at face value without any attempt at „quality control.“ Accordingly, the data I used span a variety of methods, from estimates by demographers based on a few data points to summaries of all recorded deaths. My sample does not include every country I found in the literature. A necessary condition for including a country was that I could construct a reasonably complete time series for the 20th century. A sufficient condition was that I could start such a series no later than the mid 19th century forward. The ten countries which met the sufficient condition are sometimes referred to as the „long“ sample here and in the main text. (Half of the long sample is from Scandinavia, where 18th and 19th century church records were unusually detailed.) I included the other countries (the „short“ sample) based mainly on their size and how far back before 1900 I could begin their time series.

Table A2 lists the countries in each sample, the start date of their time series, data sources and details on the handling of the missing value problems described below.

Each time series used in the paper – such as the mean and dispersion of lifetimes in a specific country- is derived from a corresponding historical inventory of life tables. For purposes of comparison and aggregation, these inventories were sorted into 5 year groups beginning 1740-44 and ending 2000-04. Each group of years is identified in the graphs and other analyses by its mid-point 1742, 1747... 2002. Complete life tables use one-year age intervals. That is, in terms of equation (A1), they give $S(0)$, $S(1)$... $S(110)$. These are denoted in the

literature as $1 \times K$ life tables, where K is the number of years of mortality experience used to estimate each of the points on $S(t)$. K is often set greater than 1 to reduce the measurement error inherent in using a single year's mortality within a one-year age interval. Wherever feasible I used or estimated a 1×5 life table and assigned it to the year group closest to the mid-point of the life table.²

However, it was not always feasible to use a 1×5 life table. And there were gaps in the historical record. Accordingly there are two kinds of missing data problems, which I treated as follows:

1. Incomplete or “abbreviated” life tables, or equivalently and more commonly, mortality rates over an age interval. Most available mortality rate data cover 5 or 10 year age intervals. So they yield only a few points on $S(t)$. A typical example might give $S(1)$, $S(5)$, $S(10)$, $S(20)$... $S(80)$. When I encountered such data I converted them to complete life tables. To do this I assumed constant mortality rates within any age interval from 5 to 65. This assumption imparts some bias,³ but it is inconsequential empirically. The consequential problems arise at very young and very old ages. Historically the vast majority of early childhood mortality occurs in the first year. So there can be considerable distortion, especially of dispersion measures, from assuming constant mortality over the early childhood years. At the other end, there is a distinct acceleration of mortality risk beyond age 70, so values of $S(70+)$ are not well approximated by a constant mortality rate in old age.

² The common alternative to a 1×5 life table is a 1×1 . I assigned those to the closest matching 5 year age group. Occasionally I encountered 1×10 life tables. Here I assumed the table is valid for each 5 year sub-period and proceeded as described.

³ For example, mortality beyond age 50 or so rises palpably with age. So assuming constant mortality from, say, 50 to 60 will underestimate points like $S(55)$, or overestimate mortality rates at ages below 55.

In cases where I had to estimate the values of $S(t)$ for early childhood or old age I used a contemporary „reference country“ which had a complete life table available and an otherwise similar mortality profile. I filled out the subject country’s life table by assuming that the share of deaths at each specific age in an interval was the same as the reference country’s share.⁴

2. Gaps in time series arose when a country is missing life table or mortality data that fit within one or more of the 5 year periods between the beginning and end of the series. I filled gaps of 15 years or less by linear interpolation of the terminal $S()$ values. For longer gaps, I used a ratio-to-trend method. This entailed selecting a reference country with data available within the gap. Then I interpolated the subject and reference country’s values within the gap. The estimated value for the subject country is then set to the subject country’s interpolated value times the ratio of the reference country’s actual value to its interpolated value.

Column (3) of Table A2 indicates which countries required interpolation of either life table or time series data and lists any reference countries. In general, the less developed countries in the short sample have the less complete data, and the consequently heavier use of interpolation needs to be kept in mind when evaluating their data.

Wars often create considerable discontinuities in a country’s time series. However, I left these discontinuities in much of the data (e.g., figures 2 through 4), because their relatively small size lends some emphasis to the importance of longer run trends.⁵ However, I did remove war related effects from the analysis of gender inequality, because they often obscure everything else.

⁴ I found that the choice of reference country does not much affect the summary measures that I ultimately used. Most any contemporary life table will have a similar age distribution of mortality in the tails of the age distribution.

⁵ However, the figures obscure some of the effects of war due to interpolation around gaps in the data that occur in war time periods.

I did this by interpolating all the gender specific measures around war-affected periods. Column (4) of Table A2 gives the mid-point of any war-affected periods for each country.⁶

B. US States and Counties

a. States, 1900-2000

Data for US states come from life tables constructed from mortality rates. These are available each ten years for a growing subset of US states from 1900 to 1930, and then for all states thereafter.⁷ The mortality rates up to 1960 come from:

Linder, Forrest E. and Robert D. Grove (1943). *Vital Statistics Rates in the United States, 1900-1940*. Washington: Government Printing Office

and

Grove, Robert D. and Alice Hetzel (1968). *Vital Statistics Rates in the United States, 1940-1960*. Washington: Government Printing Office

These can also be found at

<http://www.cdc.gov/nchs/datawh/statab/unpubd/mortabs/hist290.htm>

These were converted to 1 x 1 life tables by filling in specific ages from the contemporary US life table (see previous section). I assumed that the share of deaths at a specific age within an age interval was the same for each state as the corresponding share from the US life table.

⁶ The absence of a listing in this column need not imply that the country was unaffected by war, because data for that country may be missing (and therefore already estimated by interpolation). For example, German data for 1917 is missing and therefore estimated by interpolation between 1912 and 1922. Accordingly no further adjustment for the effects of World War I is necessary for that country.

⁷ There are some state-specific life tables for the early period, but they cover fewer states and sometimes only the white population. I opted for consistency and breadth here and accordingly constructed all the state life tables over the 20th century from mortality data.

The incomplete coverage arises because the US did not adopt the internationally uniform death registration system until 1932. Prior to that, some states implemented registration systems at different times. Thus in 1900 there were 9 registration states and by 1920 there were 35. The states adopting earliest tended to be the larger and more industrialized states.

For post-1960 data we have state specific 1 x 3 life tables around each census year in

US Center for Disease Control. National Center for Health Statistics. *US Decennial Life Tables. State Life Tables* (various issues).

These can also be found at

<http://www.cdc.gov/nchs/products/pubs/pubd/lftbbls/lftbbls.htm>

b. Counties, 1970-2000

I use county mortality data for the early 1970s, 1980s and 2000s. (I leave out the early 1990s to avoid distortion from the AIDS epidemic that begins in the early part of the 1980s and crests in the mid 1990s.) I then converted these data to 1 x 5 life tables for each county/period. For each of the three decades I collected five years of mortality rates for each US county centered on 1972, 1982 and 2002 respectively. The data for the 1980s and 2000 county mortality rates are available from

<http://wonder.cdc.gov/>

For 1970 I combined mortality counts by county with census county population estimates to produce the mortality rates for that year. The number of deaths is available from the compressed mortality file for the relevant years as described at

http://www.cdc.gov/nchs/products/elec_prods/subject/mcompres.htm

This is a summary of individual death certificates which includes an indicator of the decedent's county of residence and age. County population estimates by five year age group were taken from the US census website at

<http://www.census.gov/popest/archives/pre-1980/co-asr-7079.html>

The individual death certificates for the 1970s were sorted into age group/county bins, counted and converted to rates by dividing by the population estimates.⁸

All of the county mortality rates were converted to 1 x 5 life tables by using the contemporary US life table to fill in age intervals and allocate ages beyond 85.⁹ The mortality variables are then constructed from these life tables in the same way as described previously.

The socio-economic variables used in the analysis of the county data come from 1970-2000 census data accessed through the National Historical Geographic Information System (NHGIS) at

<http://nhgis.org/>

The available income and education distributions were treated as follows:

Education. The Census distributions of completed education by the population over 25 use varying definitions of the relevant intervals. For example, after 1980 the distribution of college

⁸ We do not have county population detail for ages less than 5. To estimate the infant mortality rate (and the 1 to 4 rate) I allocated each county's under-5 population according to state shares by age and then used these estimates as the denominator of the relevant mortality rate.

⁹ Because of coding problems, Alaska was treated as one county and some areas in Virginia were dropped. For some small counties data are occasionally missing for specific years or for specific age groups. In these cases I filled gaps from the relevant average of small counties in the same state

attendees is given by degree attained rather than by years completed. I converted each distribution to a 0 to 21 year scale in a two step procedure. First I constructed a state distribution of completed years by using available information from adjacent years to allocate specific years. This is straightforward for years up through 12. For example, the 1970 distribution separates 5 to 6 years from 7, while the 1980 distribution does not. Here I allocate the 1980 5 to 7 distribution into 5 to 6 and 7 years according to the available 1970 shares. When separate years could not be estimated (as is the case here with 5 to 6 years), I allocate the shares equally to each year.

For years beyond high school, where we sometimes have years and sometimes degrees, I match degrees and years beyond 12 as follows: Associates degrees 2 years, Bachelors 4 years, Masters split evenly between 5 and 6 years, Professional 7 years and Doctorate 9 years. Then I use the 1990 shares to allocate the more aggregated 1980 and 1970 data for college attendees. These adjustments are somewhat arbitrary, but the lumpy character of education distributions renders their effects inconsequential. Typically there is a large mass at 12 years, another clump at 16, a sizeable group between the two and few in the long tail beyond 16 years.¹⁰

The state distributions are then used to allocate each county's distribution to specific years, and I extract the county mean and standard deviation of the resulting distribution.

Income. I use family income distributions, which have open-ended upper tails. To close the upper tail I fit a Pareto distribution to each county's income distribution using the upper two income classes to estimate the parameter of this distribution and then the implied mean income

¹⁰ For example in 2000, 34 per cent of the population has exactly 12 years, another 25 per cent have less than 4 years of college and 13 per cent have a 4 year degree. Of those 7 per cent who get beyond 4 years of college only 1 in 10 have a doctorate.

in the upper tail. I truncate the parameter at the 99th percentile of the sample distribution to avoid outliers, and I use the sample mean where the parameter cannot be estimated.¹¹ I then assume that all families earn either the mid-point of their distribution or the mean of the upper tail. The mean, mean of logs and its standard deviation are then extracted from this distribution.

¹¹ As when there is mass in the upper tail but none in the penultimate class.

Table A1. Sources of Data

[Brief reference used in Table A2 in brackets]

A. Online Databases Covering Multiple Countries.

[HMD] Human Mortality Database. (<http://www.mortality.org/>, <http://www.lifetable.de/>) This is a joint project of the Department of Demography, University of California, Berkeley and the Max Planck Institute for Demographic Research, Rostock, Germany. Each site has downloadable life tables for a number of countries and years. Wherever available I use the 1 x 5 (1 year age interval, 5 year average data) period life table.

[WHO] World Health Organization. WHO Statistical Information System (WHOSIS) (<http://www.who.int/whosis/en/>) Mortality rates for member countries for various years from 1979 to date and abbreviated annual life tables for member countries from 1999 to date.

B. Specific Countries

[AR1] Somoza, Jorge (1971). La Mortalidad en la Argentina Entre 1869 y 1960. Centro de Investigaciones Sociales. Instituto Tocuato Di Tella. Centro Latinoamericano de Demografia. Editorial del Instituto.

[AR2] Centro Latinoamericano y Caribeño de Demografía- División de Población (CEPAL/CELADE), Boletín Demográfico (No.67 de enero del 2001). Web site: <http://www.eclac.cl/publicaciones/Poblacion/9/LCG2119P/BD67.html>

[AU1] Australian Bureau of Statistics. Cat. no. 3105.0.65.001 Australian Historical Population Statistics. Table 50. Number of Persons at Exact Age x, Australia 1881 onwards. Available at: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3105.0.65.0012006?OpenDocument>

[BR1] Arriaga Eduardo. (1968). New life tables for Latin American Populations in the Nineteenth and Twentieth Centuries. Population Monograph series No. 3. University of California Berkeley. Department of Demography.

[BR2] Instituto Brasileiro de Geografia e Estatística (IBGE), Diretoria de Pesquisas, Departamento de População e Indicadores Sociais. Indicadores Demograficos (various issues).

[CL1] Preston, Samuel; Nathan Keyfitz and Schoen Robert. (1972). Causes of Death. Life tables for National Populations. Seminar Press. New York and London.

[CL2] United Nations. Statistical Office. Demographic Yearbook (various issues). New York

[FI1] Turpeinen, Oiva. Fertility and Mortality in Finland since 1750. *Population Studies*. Vol 33 N 1. March 1979. 101-114

[GE1] Imhof, Arthur. (1990). Lebenserwartungen in Deutschland vom 17. bis 19. Jahrhundert. (Life expectancies in Germany from the 17th to the 19th Century). Acta Humaniora. Weinheim.

[IN1] S. P Jain. (1981) Actuarial Report and Life Tables 1951-1961. Office of the Registrar General, Ministry of Home Affairs, New Delhi

[IN2] Compendium of India's Fertility and Mortality Indicators (1971-1997) Based on the Sample Registration System. Registrar General. Ministry of Home Affairs. New Delhi. 1999

[IN3] Sample Registration System, 1970-1975. Vital Statistics Division, Office of the Registrar General, Ministry of Home Affairs, New Delhi, 1983

[IN4] Sample Registration System, 1979-1980. Vital Statistics Division, Office of the Registrar General, Ministry of Home Affairs, New Delhi, 1984

[IN5] SRS Abridged Life Tables 1981-1985. Vital Statistics Division, Office of the Registrar General, Ministry of Home Affairs, New Delhi, 1990.

[IN6] SRS abridged life tables 1986-1990. Vital Statistics Division, Office of the Registrar General, Ministry of Home Affairs, New Delhi, 1994.

[IR1] Pascua, M. "Evolution of Mortality in Europe during the Twentieth Century" in World Health Statistics Report, v. 3. World Health Organization. Geneva, 1950.

[IT1] Natale, Marcello and Amedeo Bernassola. (1973) La Mortalita per Causa Nelle Regioni Italiane. Tavole per Contemporanei 1965-66 e per Generazioni 1790-1964. Facolta di Scienze Statistiche Demografiche ed Attuariali dell' Universita de Roma. Istituto di Demografia. Roma.

[JP1] Statistics Bureau and Statistical Research and Training Institute. Ministry of Internal Affairs and Communication. Historical Statistics of Japan. Chapter 2: Population and Households. Table: 2-31- b: Complete Life Table (1891--2000). Web site: <http://www.stat.go.jp/english/data/chouki/02.htm>

[US1] Haines, Michael. Estimated life tables for the United States (1850-1900). NBER Series on Historical Factors in Long Run Growth. NBER Historical paper # 59. National Bureau of Economic Research, September 1994. pp 21 (US Model life tables)

Table A2. Data Sources and Methods

Country	Start Year	Years Covered	Interpolation Conventions	War Years	Sources
	(1)	(2)	(3)	(4)	(5)
A. 10 Country ('long') sample					
Denmark	1837	all			HMD
England & Wales	1842	all		1912, 1917, 1942	HMD
Finland	1752	all	Sweden (LT)	1917, 1942	FI1 to 1877. Then HMD
France	1807	all		1807, 1812, 1912, 1917, 1942	HMD
Iceland	1837	all			HMD
Germany	1742	1742-1852 by decade. 1872-1912, 1927, 1932, 1947 on	linear, France	1942	GE1 to 1852. Then HMD
Netherlands	1852	all		1942	HMD
Norway	1847	all			HMD
Sweden	1747	all			HMD
USA	1852	1852-1892 by decade. Then 1902 on	linear		US1 to 1902. Then HMD

Table A2. Data Sources and Methods (cont.)

Country	Start Year	Years Covered	Interpolation Conventions	War Years	Sources
B. 13 country ('short') sample					
Argentina	1882	1882,1912,1917,1947 on.	Italy. Italy (LT)		AR1 to 1957. AR2 1962-97. Then WHO
Australia	1882	all		1912,1917	AU1 to 1917. Then HMD
Brazil	1872	1872,1892,1902,1922, 1942,1952,1962,1967, 1977 on.	Italy to 1902, then Spain. Linear.		BR1 to 1962. Then BR2
Chile	1907	1907,1922,1932,1942, 1952 on	Linear. Spain(LT), Portugal(LT)		CL1 to 1952. Then CL2
India	1892	1892-1907,1922,1927, 1942-1987, 1997 on	linear, India (LT)		CL2 to 1947. IN1 for 1952,1957.IN2 for 1962, 1967.IN3 for 1972. IN4 for 1977. IN5 for 1982. IN6 for1987. Then WHO
Ireland	1902	1902,1912, 1927, 1937-1952, 1957	linear, Ireland (LT)		IR1 for 1902, 1912. Then HMD
Italy	1862	all		1917,1942	IT1 to 1867, then HMD
Japan	1892	1892-1902, 1912,1922, 1927, 1937, 1947 on	linear		JP1 to 1937. Then HMD
New Zealand	1882	1882,1902,1912,1922, 1927, 1937, 1947 on	linear		CL1 to 1937. Then HMD
Portugal	1902	1902,1912,1922, 1932, 1942 on	Linear. Italy (LT), Spain (LT)		IR1 to 1947. Then HMD
Russia/Soviet Union	1897	1897,1927, 1937, 1957,1967,1977, 1982, 1987, 2002	linear, Spain		CL2 for 1897, 1927. HMD for 1937-1987. WHO for 2002. (See Note 1)
Spain	1902	all		1937,1942	HMD
Switzerland	1877	all			HMD

Table A2. Data Sources and Methods (cont.)

Start year: the mid-point of the 5 year interval with the first observation in a country's time series. (All years in the table are mid-points of 5 year intervals.)

Years Covered: years with gaps in the time series. No gaps denoted by 'all'

Interpolation Conventions: methods by which gaps in the time series and missing ages in the life tables were filled. (See appendix text for discussion). 'Linear' means that some or all gaps in the time series were filled by linear interpolation. Country name is indicated when a ratio-to-trend method is used to fill time series gaps. Country (LT) indicates which country's life tables have been used to fill missing ages in any of this country's life tables.

War Years: Years in which available data on gender differences is replaced by an interpolated value because of distortion by wars.

Sources: See Table A1 for full citations.

Notes:

1. Data are for Russian Empire (1897), Soviet Union (1927-1987), and Russia (2002). WHO data permit comparison between data for Russia and whole of former Soviet Union for 1992; they are substantially identical.