MPIDR-NES Training Programme

Moscow, New Economic School,14th January - 1st February 2013

Population and Health

Лекция 2. Агрегированные меры смертности и функции таблицы смертности

Lecture 2. Aggregate measures of mortality and life table functions

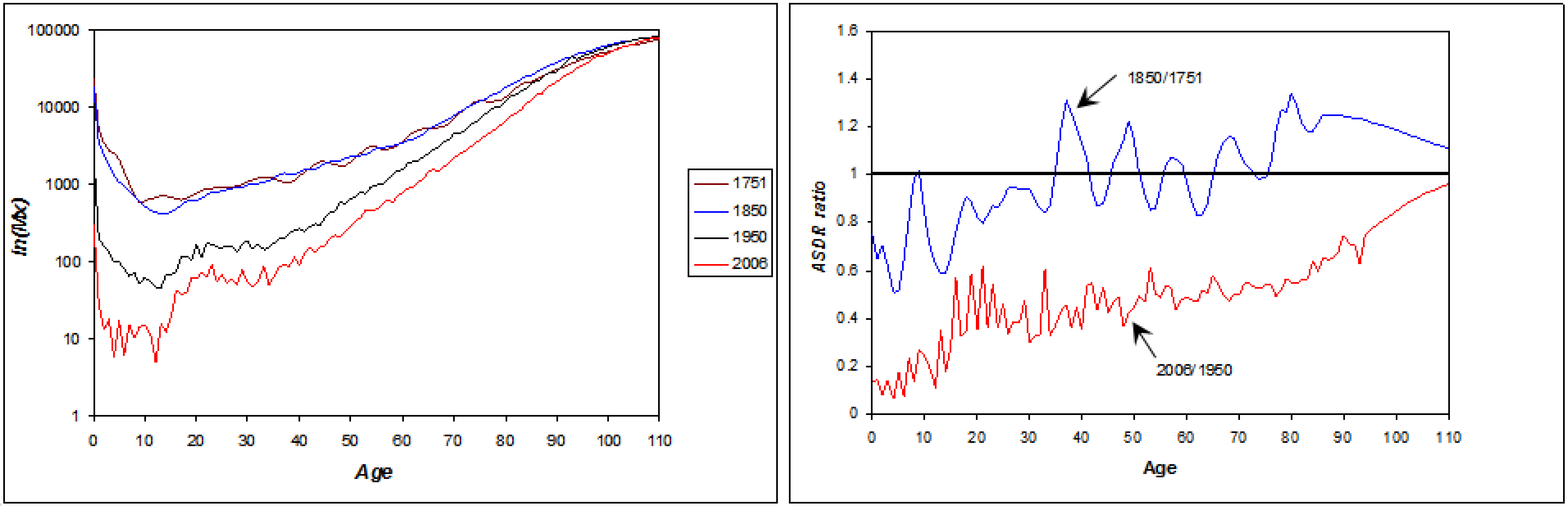
MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG

РЭШ Российская экономическая школа

A world of age-specific death rates

Analysis of age- and cause-specific death rates provides valuable information about mortality changes across time and space.



Evolution of all-cause mortality in Sweden 1751-2006: a very slow progress in 1751-1850 and a rapid progress in 1950-2006.

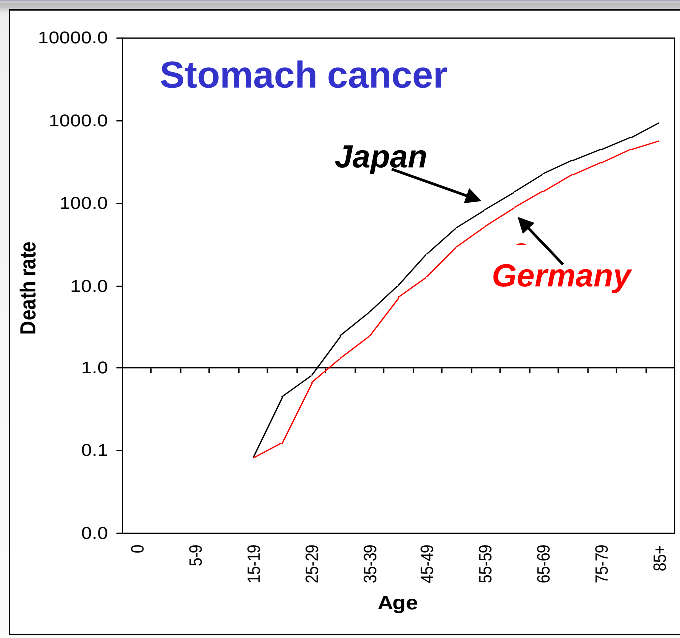
Data source: Human Mortality Database (http://www.mortality.org)

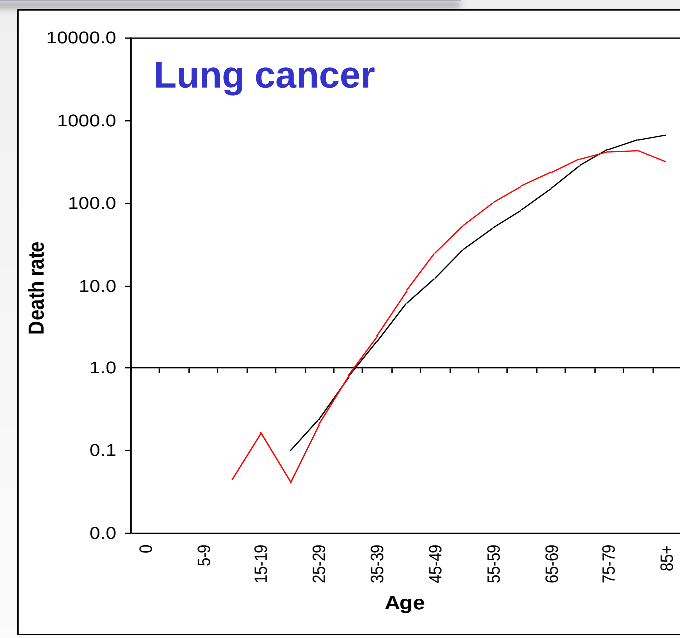
Age- and cause-specific death rates

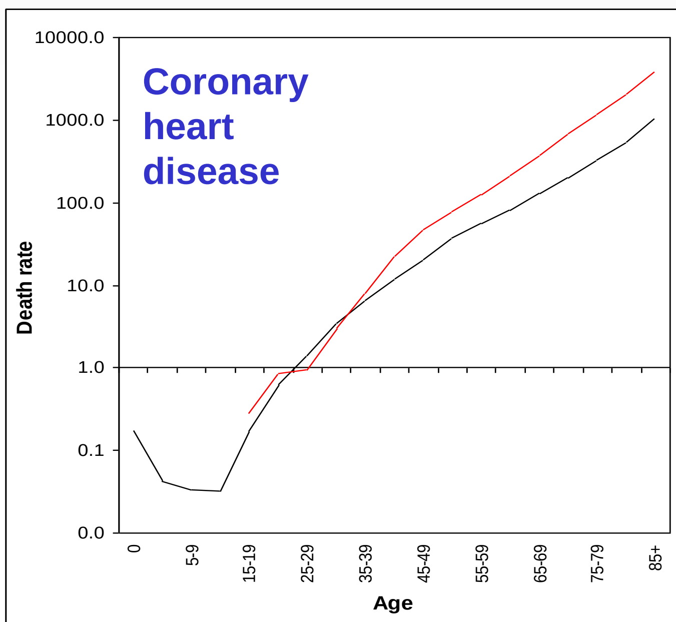
Germany vs. Japan in 2004-6.

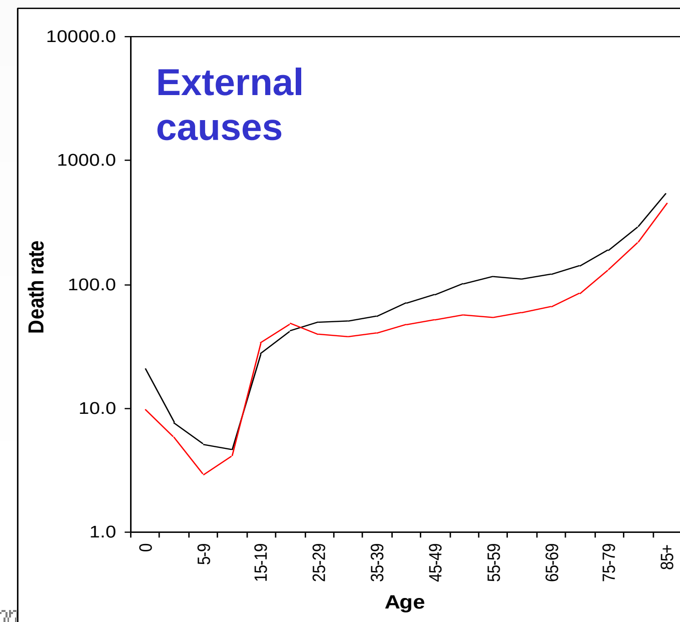
Why Japan is doing better?

Japanese disadvantage in stomach cancer and external causes is outweighed by Japanese advantage in other leading causes of death such as CHD.









Aggregation of age-specific death rates

Mortality pattern of a population is determined by array of age- and cause-specific death rates:

• Even in the abridged format, there are about 20 age groups (0, 1-4, 5-9, ..., 85+ or 95+).

• Usually at least 20 to 50 major causes of death are needed to reflect all important aspects of mortality.

This means that one should observe 400 to 1,000 real values at once for every population under study. One study usually includes several populations and several time points that further multiplies the amount of data to be analyzed.

Thus, it is desirable to find a way for a brief and aggregate description of mortality.

Crude and age-standardized death rates

Crude death rate is the simplest measure of the overall intensity of death over the entire range of ages

$$CDR = \sum_x D_x / \sum_x E_x = D \ / E$$

D_x and E_x are deaths and population exposure (mid-year population) at age x.

Inter-population comparisons can be biased by differences in age compositions of populations. Populations with higher shares of the elderly are in unfavorable position. This is so because all-cause mortality exponentially increases with age. There is a need to eliminate influence of differences among age structures.

Direct method of standardization:

$$SDR = \sum_x \theta_x^s M_x = \sum_x \theta_x^s (D_x / E_x)$$

θ_x^s are the age-specific population weights of the standard population.

Choice of the population standard

Larger difference between the age-structures of the standard and actual populations lead to greater difference between CDR and SDR.

As a particular case, age structure of one of the populations under comparison can be chosen as a standard.

WHO produces standard population age structures for geographical regions and the whole world. For example, the European and the World population standards. These standards are used in international statistical reports on mortality and also in many studies.

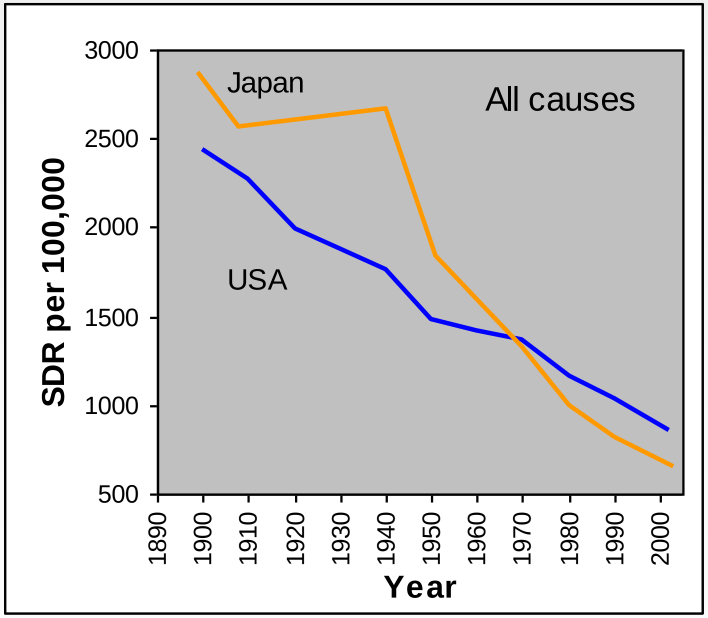
Attention!
Some countries (e.g. the US) use their own population standards.

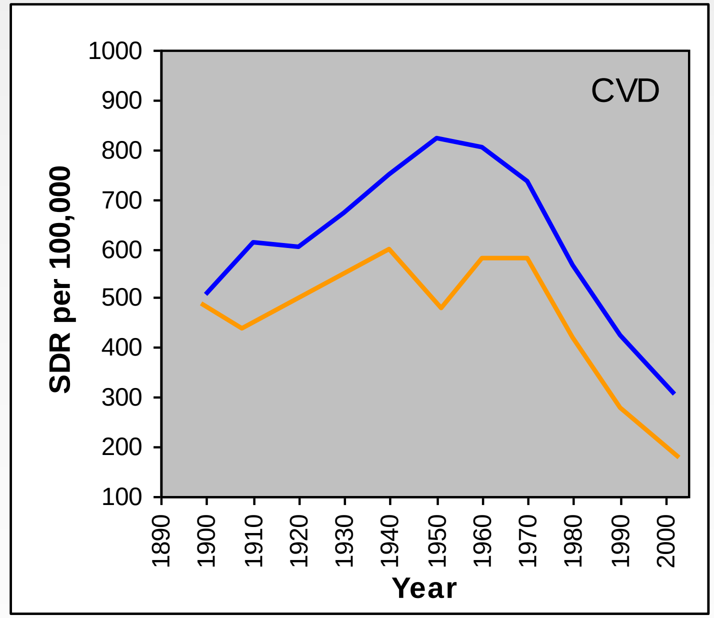
Result of the comparison sometimes depends on choice of the population standard.

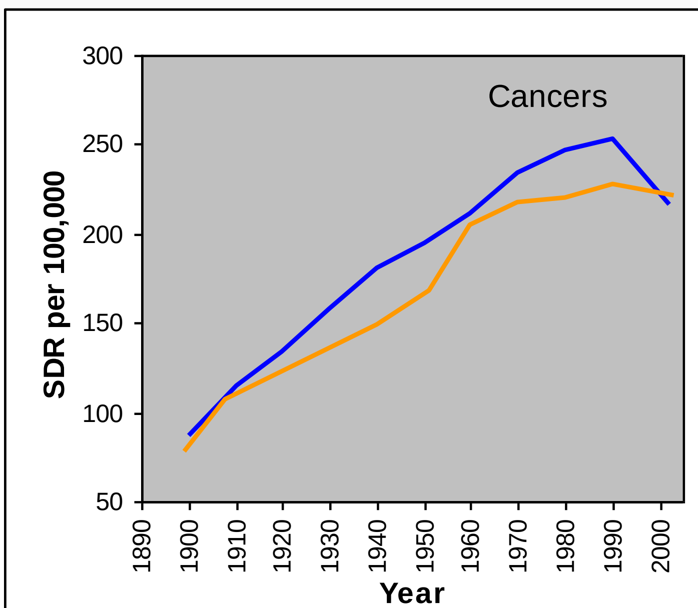
Age group (years)	European standard population
0	1600
1–4	6400
5–9	7000
10–14	7000
15–19	7000
20–24	7000
25–29	7000
30–34	7000
35–39	7000
40–44	7000
45–49	7000
50–54	7000
55–59	6000
60–64	5000
65–69	4000
70–74	3000
75–79	2000
80–84	1000
85 +	1000
Total	**100000**

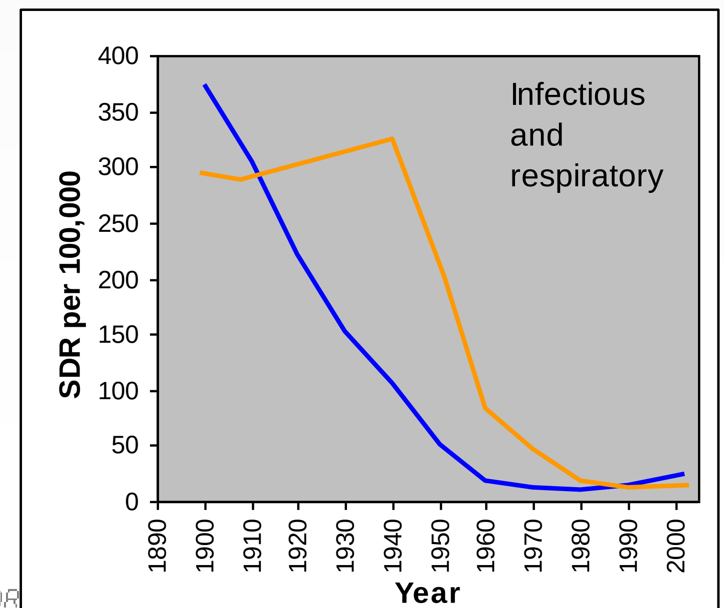
Long-term trends in cause-specific SDRs: USA vs. Japan

Data sources: Preston, Keyfitz., and Schoen, 1972; WHO Mortality Database









Indirect standardization

Direct standardization implies calculation of a death rate expected in a given population under condition that its population age structure corresponds to the standard. ***Indirect standardization* is** based on the comparison of the observed CDR with the one that would have been observed in a given population if its age-specific death rates correspond to a standard.

Advantage: no need to know age-specific death rates in populations to be compared. It is enough to know total death numbers or CDRs $\Rightarrow$ SMRs are popular in population geography.

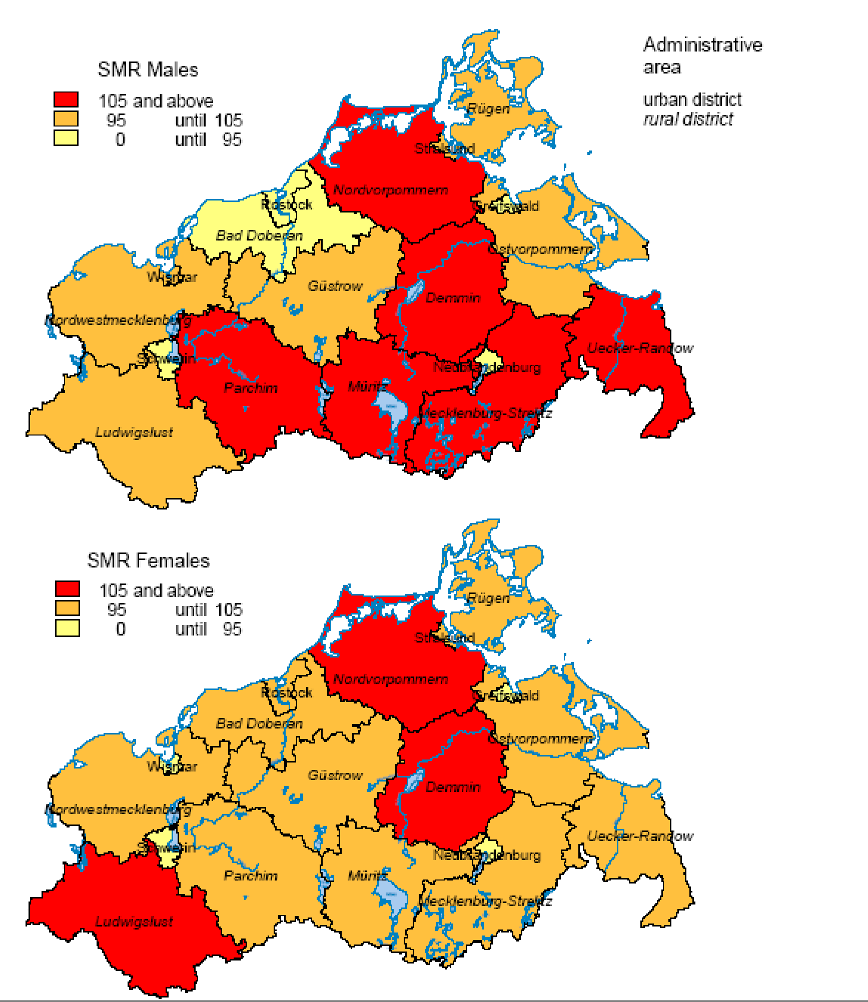
$$SMR^i = \frac{D^i}{\sum_x (M_x^s E_x^i)}$$

SMR-standardized mortality ratio, M^s_x – age-specific death of the standard population

$$ISDR^i = SMR^i \cdot CDR$$

ISDR – indirect-method SDR for area *i*

Figure 11: Standardised mortality ratios in districts of Mecklenburg-Vorpommern in 1997–2001 for males and females



Kibele, 2005.

Life table in demography and in public health

In classic demography the life table is a model of mortality as a component of the population reproduction.

Each life table produces a stationary population. This population has age structure that would be formed under regime of constant fertility and the life table mortality rates.

In epidemiology and public health, life table is a way to describe population's mortality and survival.

Life table functions in continuous notation

Let $l(0)$ be a cohort size at starting age 0. Knowing the force of mortality $\mu(x)$, the decrease of the cohort size $l(x)$ with age can be expressed as

$$l(x) = l(0)e^{-\int_0^x \mu(t)dt} \qquad (0)$$

The monotonously decreasing function *l(x)* is called survival function.

$l(0)$ is called radix of the LT. Usually it is assumed to be 1 or 100,000.

Life expectancy at age 0 is defined as the average (or expected) age at death (or length of life) in the cohort

$$e(0) = \int_0^\infty l(x)dx = \int_0^\infty e^{-\int_0^x \mu(t)dt} dx \qquad (1)$$

The survival function and the life expectancy can be also calculated for any range of ages

$$l(x \mid X) = l(x)e^{-\int_x^X \mu(t)dt} \qquad (2a) \qquad\qquad e(x \mid X) = \int_x^X l(y)dy = \int_x^X e^{-\int_x^y \mu(t)dt} dy \qquad (2b)$$

LT computation from real discrete data

To build a LT means to be able to calculate integrals (1) and (2) from the empirical data on deaths and population exposures by one- or five-year age groups. Imagine a cohort with its initial size equal to l_0. For elementary age intervals $[x_i,x_{i+1})$:

$$l_{x_{i+1}} = l_{x_i} \cdot (1 - q_i)$$

where q_i is the probability of dying between exact ages x_i and x_{i+1}, $i=0,1,2, ..., w-1$ (w – the number of the last age interval 85+, 90+, 100+ ...).

LT deaths in $[x_i,x_{i+1})$:

$$d_i = l_{x_i} - l_{x_{i+1}} = l_{x_i} q_i,\ \ i = 0,1,2,..., w-1;$$

person-years (LT stationary populations) in $[x_i,x_i+1)$:

$$L_i = l_{x_i} n_i - d_i (1 - a_i) n_i,\ \ i = 0,1,2,...w-1$$

a_i is the share of the age interval lived by those dying within the interval.

LT death rate in $[x_i,x_{i+1})$:

$$m_i = \frac{d_i}{L_i},\quad i = 0,1,2,..., w.$$

LT computation from real discrete data

There are several ways to connect the LT probabilities of death with the empirical data. For example, it is natural to postulate that the LT death rates m_i are equal to the empirical death rates M_i yields

$$m_i = M_i = \frac{d_i}{L_i} = \frac{d_i}{n_i(l_{x_i} - d_i(1-a_i))} = \frac{q_i}{n_i(1-q_i(1-a_i))}\,, \quad M_i = \frac{D_i}{E_i} \qquad (3)$$

$$q_i = \frac{n_i M_i}{1+(1-a_i)n_i M_i} \qquad (4)$$

(4) is the most widely used formula connecting the LT q_i with the empirical death rates

The amount of person-years lived after age x_i

$$T_{x_i} = \sum_i^w L_i. \qquad (5)$$

Finally, the life expectancy at age x_i

$$e_{x_i} = T_{x_i} / l_{x_i} \qquad (6)$$

Because all people will eventually die, for the last age group w we have:

$$q_w = 1,\ d_w = l_{x_w},\ L_w = T_{x_w} = l_w / M_w,\ e_{x_w} = 1/M_w.$$

LT computation from real discrete data

Finally, we need to obtain values a_i, e.g. shares of the age interval $[x_i, x_i+1)$ lived by those dying in this interval.

Over the first years of life (age=0, 1-4), the infant mortality is rapidly falling from its high level during the first days and weeks, down to much lower levels. For the first year of life and age 1-4, there are empirical formulae by Coale and Demeny (1983).

Keyfitz, 1968:

$$a_0 = 0.07 + 1.7 M_0$$

Values of $\boldsymbol{a_i}$ for use in ages 0 and 1-4

	Males	Females
Value of $_1a_0$		
If $_1m_0 \geq 0.107$	0.330	0.350
If $_1m_0 < 0.107$	$0.045+2.684*{}_1m_0$	$0.053+2.800*{}_1m_0$
Value of $_4a_1$		
If $_1m_0 \geq 0.107$	0.338	0.340
If $_1m_0 < 0.107$	$0.413-0.704*{}_1m_0$	$0.381-0.380*{}_1m_0$

Source: Preston et al, 2001; adapted from Coale & Demeny (1983)

For a complete (one-year) LT one can set a_i=0.5 for all ages older than 0. The same can be done also for abridged LT. A more precise values can be obtained by applying more complex procedures.

Life-Table-abridged-complete.xls

A few words about notation and names

The life table cohort *lx* is also called synthetic cohort. That is to distinguish between this cohort exposed to age-specific death rate of the current year and real birth cohorts.

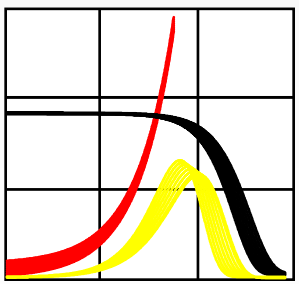
For the LT functions defined on an age interval $[x_i, x_{i+1})$, (following C.L.Chiang (1984)) we use:

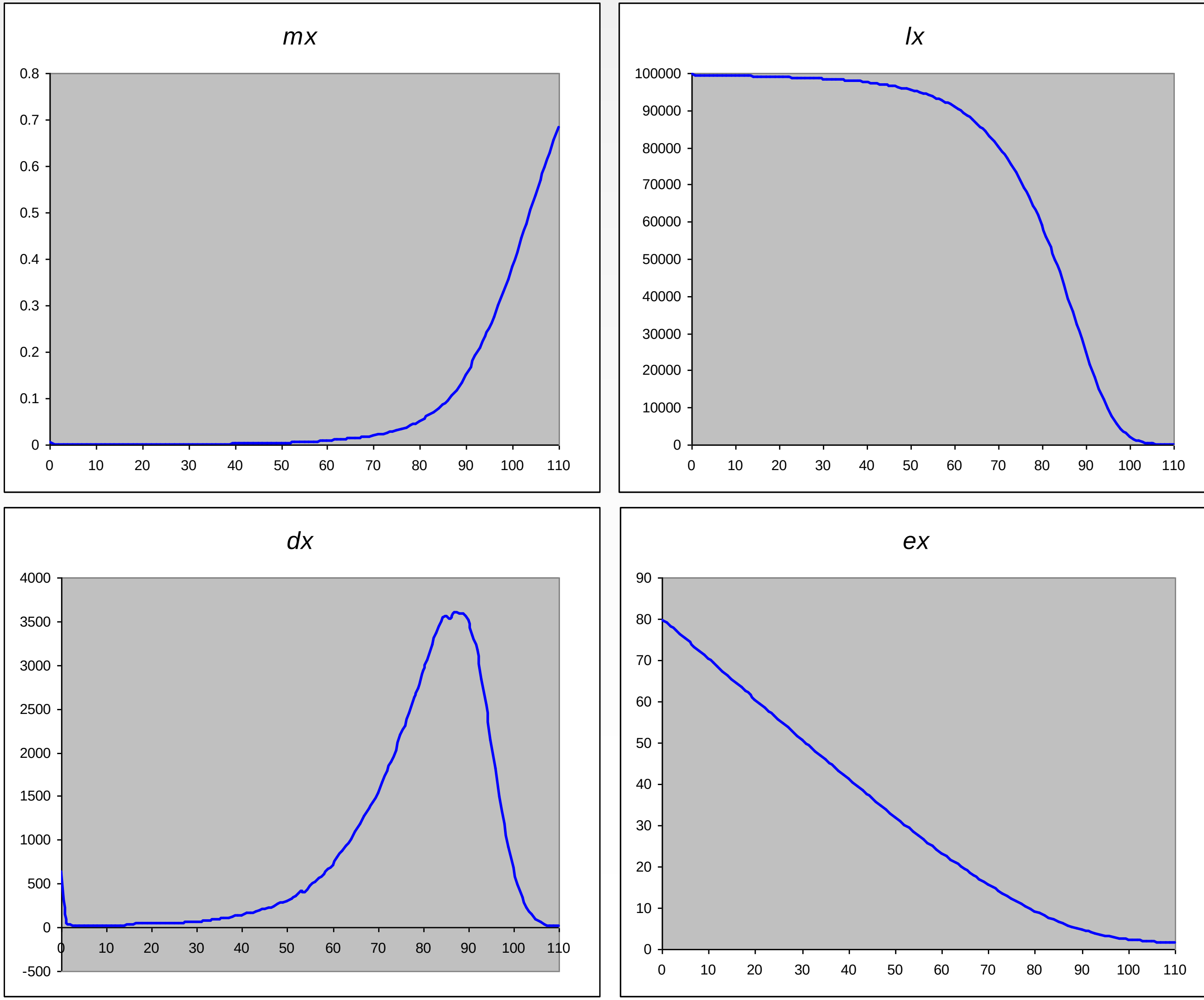
$$M_i, m_i, d_i, L_i$$

an alternative and widely used notation for an elementary age interval $[x, x+a)$ is:

$${}_aM_x,\ {}_am_x,\ {}_ad_x,\ {}_aL_x$$

Shapes of the LT curves





Some useful LT measures

A few LT measures used in public health research

e_x = life expectancy at age *x*

$p_{x|y} = l_y / l_x$ = probability of surviving from age *x* to age *y*

$d_{x|y} = l_y - l_x$ = people dying between exact ages *x* and *y*

$e_{x|y} = \frac{T_x - T_y}{l_x}$ = expected length of life between exact ages *x* and *y*

$e_{x|y} = \frac{T_x - T_y}{l_0}$ = expected length of life lived by a newborn between exact ages *x* and *y*

$\frac{l_x - l_y}{l_0}$ = probability that a newborn will die between exact ages *x* and *y*

Measures of length of life

Another formula for the **average life expectancy** at age x_i:

$$e_{x_i} = \frac{1}{l_{x_i}} \sum_{i=0}^{w} d_i \bar{x}_i, \quad \bar{x}_i = x_i + a_i n_i$$

Median age at death (X_{med}). Age to which exactly half of the LT cohort survive:

$$l_{X_{med}} = 0.5$$

Modal age at death. Age at which LT death number reaches its maximum.

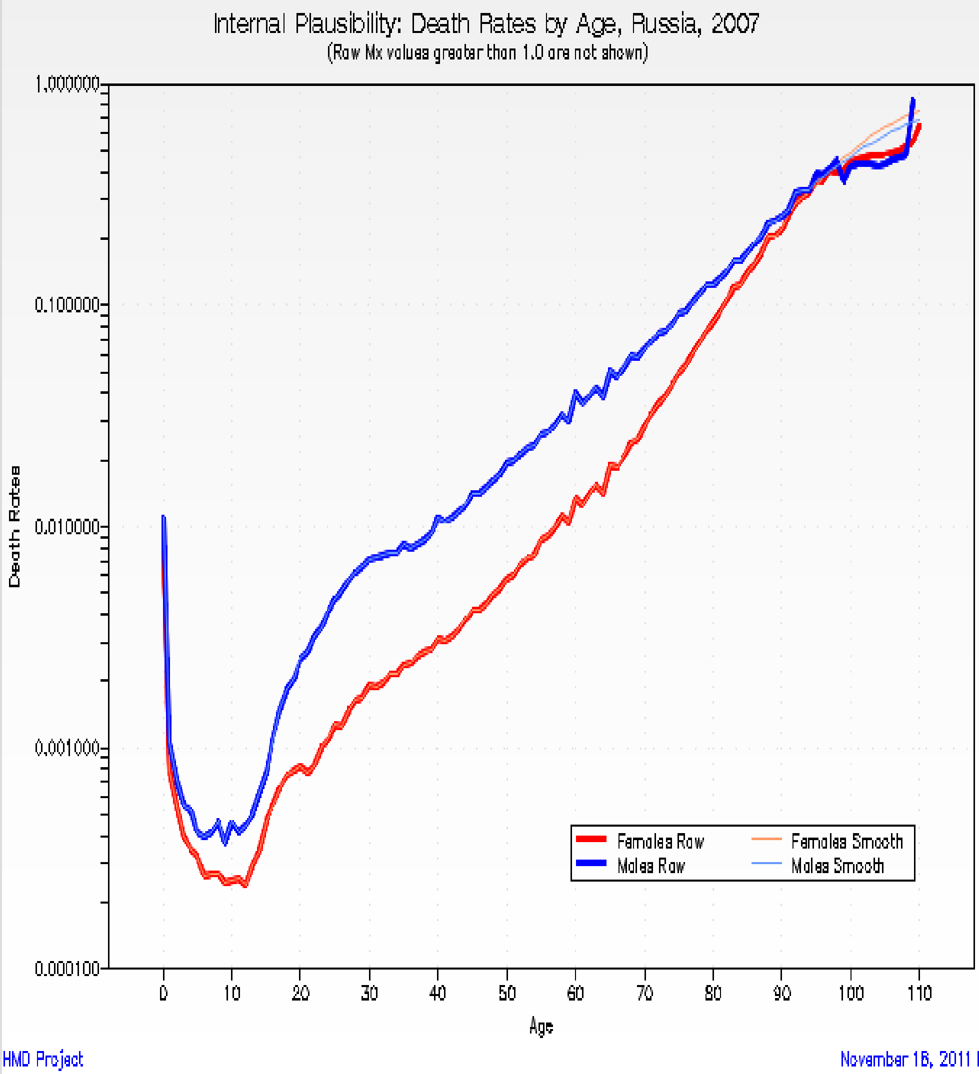
Usually in human populations: Mode >= Median >= Average

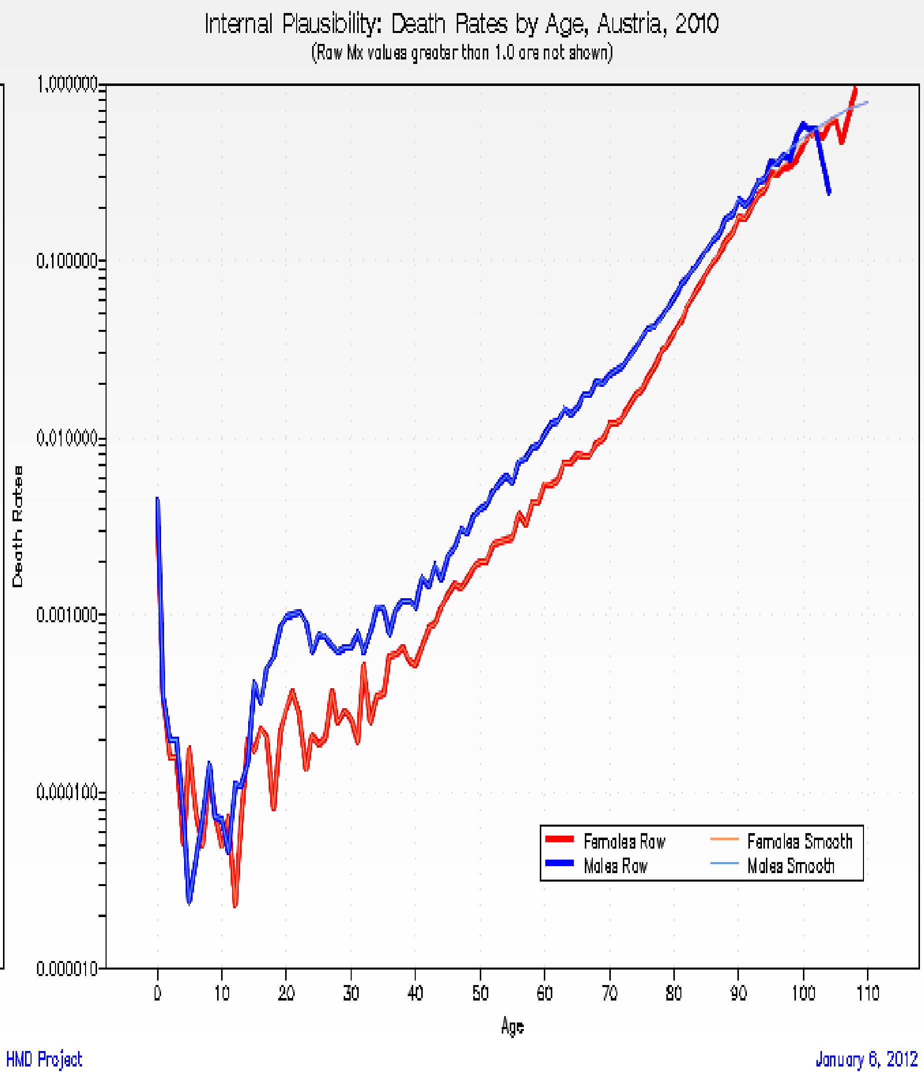
Note about the last age *w*

In the past, LT functions could be estimated quite precisely with the last age group 85+ (or even 75+) because very few people have been surviving to advanced ages. By the 1980s-90s mortality has declined very much giving a real chance to many people (especially females) to survive to ages 85+. In many industrialized countries, female life expectancy at birth exceeds 80 and male LEB exceeds 75. This makes it desirable to get data running up to the highest ages (as it is in the HMD).

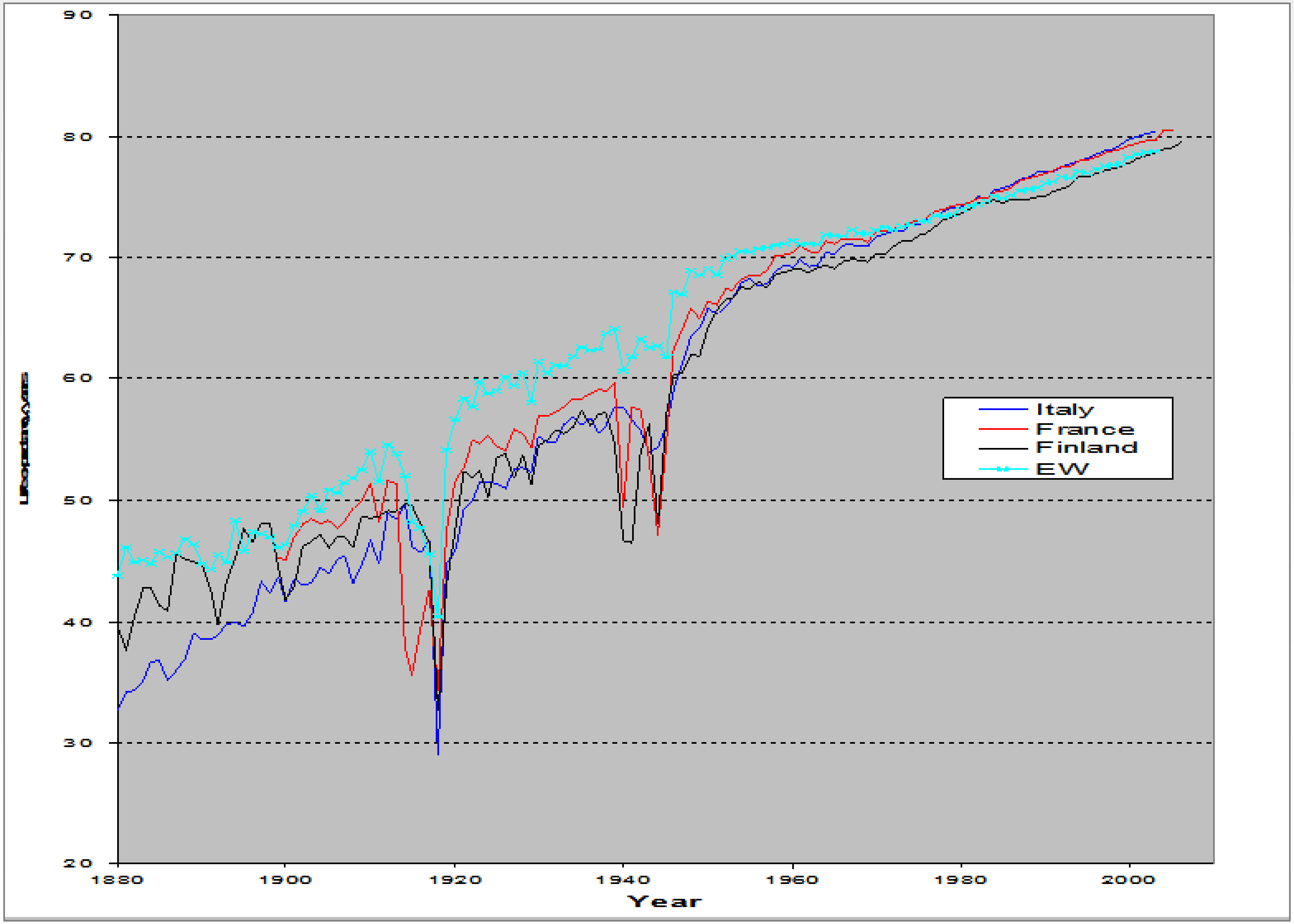
Unfortunately, cause-specific data on deaths beyond age 85 are rarely available (see the WHO Mortality Database).

Note about random fluctuations and smoothing





The most popular LT output: trends in e_0

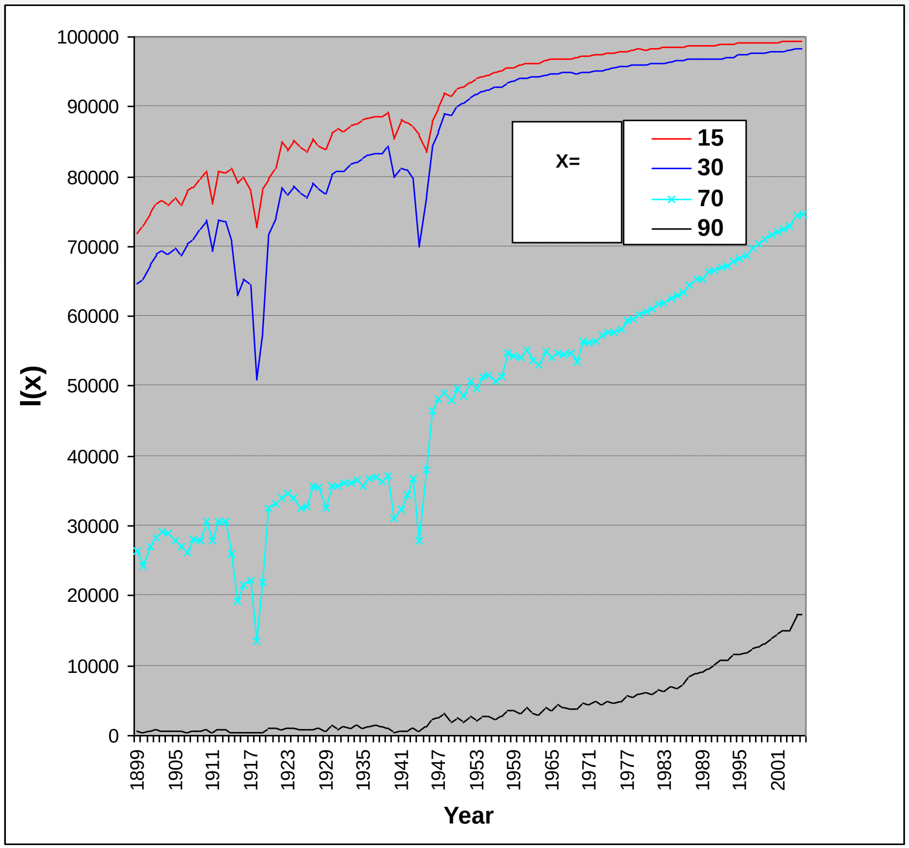


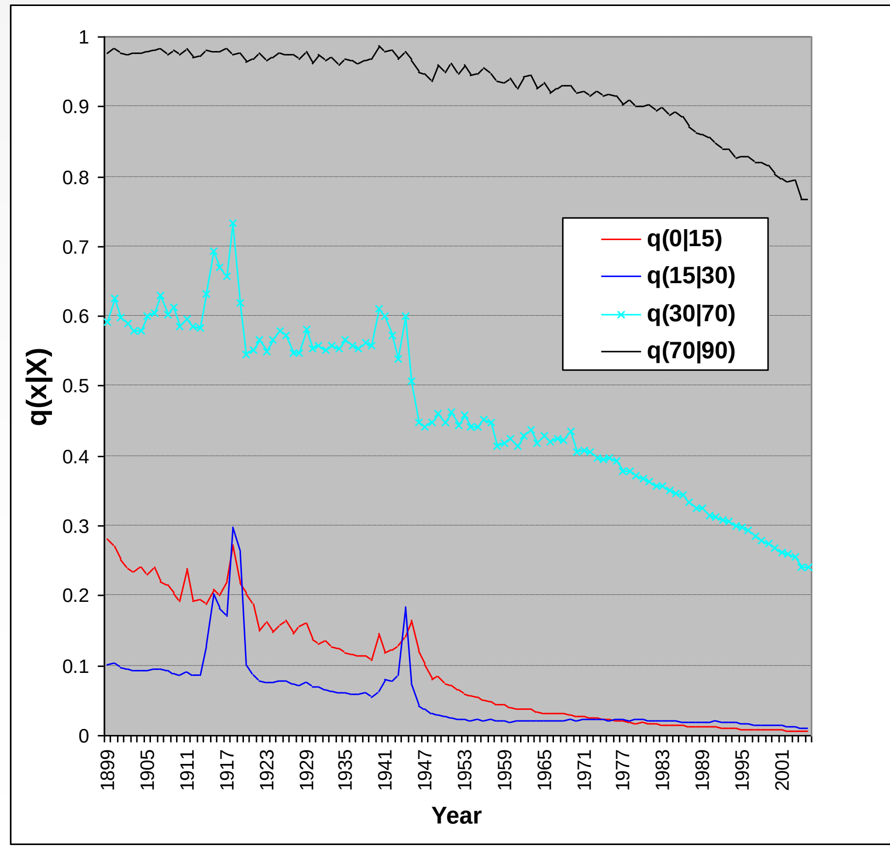
Trends in LE at birth since 1880 for Finland, France, England and Wales, and Italy.

Trends in survival and speed of mortality decline at various ages

Trends in survival to ages 15, 30, 70, and 90 (left panel) and trends in probability of dying between age 0 and 15, 15 and 30, 30 and 70, 70 and 90.

France 1899-2005





References

Chiang C. L. (1984). *The Life Table and its Applications*. Robert E Krieger Publ Co.: Malabar (FL)

Coale and Demeny (1983). Regional model life tables and stable populations. In: B. Vaughan et al. (Eds.), *Studies in Population*, Vol.8. New York: Academic Press.

Keyfitz, N. (1968). *Introduction to the mathematics of population.* Adisson-Wesley: Reading, Massachusetts.

Kibele, E. (2005). *Mortality trends and patterns in Mecklenburg-Vorpommern before and after unification: a study based on mortality, cause of death and population data*. Rostock: University of Rostock.

Preston, S.H., Heuveline, P., Guillot, M. (2001). Demography. Measuring and modeling population processes. Malden, Oxford: Blackwell Publishing.

Preston, S.H., Keyfitz, N., and Schoen, R. Causes of Death: Life Tables for National Populations. Seminar Press, New York. 1972.