

Giulia Roli

An adaptive procedure for estimating  
and comparing the old-age mortality  
in a long historical perspective:  
Emilia-Romagna, 1871-2001.

Quaderni di Dipartimento

Serie Ricerche 2008, n.8

ISSN 1973-9346



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

## Abstract

When the goal of a study is to assess and compare the evolution of old mortality, the researchers often face with problems of defects in data, such as loss of accuracy and numerical inconsistency of information. In the present work, an adaptive procedure for the issue of survival estimations is proposed as a combination of several methods: spline interpolation to split grouped data into age-specific quantities; use of mortality models for estimating the risk of death at advanced ages; methods for choosing the fitting age range based on the analysis of *life-table aging rates*. The procedure is applied to Emilia-Romagna data in order to construct complete period life tables corresponding to the occurrence of each population census from 1871 to 2001, and, finally, to compare the measures of old-age mortality by time and sex.

*Keywords:* old mortality; complete life tables; mortality models; spline interpolation method.

1.	Introduction.....	4
2.	Notation and main definitions.....	5
3.	Description of original data.....	6
4.	Steps for computing probabilities of dying and period life tables .....	8
5.	Splitting 5-year age grouped data by using spline interpolation.....	12
6.	Models of old-age mortality.....	19
7.	Fitting the Kannisto model.....	22
7.1.	Choosing the fitting age range.....	22
7.2.	Parameter estimation .....	27
7.3.	Replacing observed with fitted death rates.....	28
7.4.	Comparisons with some other mortality models .....	28
8.	Period life tables.....	30
9.	Discussion .....	32
	References.....	44

## 1. Introduction

The development of human longevity is certainly a relevant and widespread topic that has drawn the attention of researchers and demographers in the last years. The growing interest in the matter is mainly due to the great increase of octogenarians and centenarians occurred during the last decades, as until recently they were unusual or even rare.

The main aim of this work is the study of old mortality in a long historical perspective, in order to assess the path and the evolution of great longevity from the past up to now. Our attention focuses on elderly people of a specific Italian region, namely Emilia-Romagna, in the period from 1871 to 2001. Actually, calendar years corresponding to population censuses within such period are considered in order to draw a picture of mortality every about 10 years. Comparisons among year-specific old mortality are made by referring to corresponding probabilities of dying at each age, whose estimates become more difficult as age increases. Once properly estimated, such quantities are arranged into period life tables for women and men separately and together.

The problem of survival estimations in old ages is mainly due to a couple of reasons. Firstly, the accuracy of age information decreases as age increases and we often face with cases of age heaping and age overstatements. Secondly, at very high ages the number of deaths becomes numerically inconsistent in order to estimate a risk of death representing the true value in the population. If it is the case, strong fluctuations in the estimates would be observed depending on the variations between the numerator and the denominator in the specific death rates. Analogously, this internal inconsistency is revealed by implausible developments of probabilities of dying.

Especially in historical periods when life expectancy is low and the number of survivors decreases consistently after 65 or 70 years of age, the estimation problems described above become stronger, so that a more careful assessment is needed.

In the present work, the Kannisto model of mortality (1994) is used in order to estimate probability of dying at highest ages and then supplement the final construction of life tables.

Another question concerns availability of data. Especially for earliest periods and sometimes just for elderly people, official statistics on deaths and population are published only into age grouped data, usually 5-year age groups. As a consequence, a method splitting such data into single year quantities is needed. Here the cubic spline interpolation method is used to split both death and population counts for all ages below (and sometimes also over) the last open age interval given by the official statistics.

## 2. Notation and main definitions

Quantities firstly considered in this work are data on mortality, namely death counts and population numbers by sex and single year of age for each census year, and annual counts of live births by sex. They are here denoted by  ${}^tD_x$ ,  ${}^tP_x$ , and  ${}^tB$ , respectively, where  $x$  represents the age and  $t$  the calendar year. Their original availability from official statistics is described in detail in the next section.

There are three main quantities actively involved in the old mortality estimation process:

- *central death rate* at age  $x$  and year  $t$ , denoted by  ${}^tm_x$ ;
- *force of mortality* (or instantaneous death rate or hazard rate) at age  $x$  and year  $t$ , denoted by  ${}^t\mu_x$ ;
- *probability of dying* within 12 month of reaching age  $x$  in year  $t$ , denoted by  ${}^tq_x$ .

These quantities are very closely related to each other and generally the choice among them is uniquely a matter of convenience depending on the nature of the problem and on the available data. Fortunately there are simple approximations which easily enable to estimate any one of them from any other. Computational details on these approximations will be introduced during the development of this document.

In the present investigation, we are firstly interested in probabilities of dying,  ${}^tq_x$ , and corresponding estimation to be arranged in period life tables and then compared among census years. In other words, estimation of age-specific probabilities of dying is our final objective.

On the other hand, central death rates,  ${}^tm_x$ , are the first quantities which can be computed from data on deaths and individuals. So they represent our first step toward probabilities of dying.

Finally, forces of mortality,  ${}^t\mu_x$ , are involved in the fitted mortality model, namely the Kannisto model, which, as most of the theoretical models, is expressed as regards to the force of mortality.

### 3. Description of original data

The sources of the data we consider are official statistics published by the Italian institutional agency for statistical data collection, formerly named DIRSTAT and then ISTAT, and by the regional authorized agency of Emilia-Romagna.

As we introduced above, the data consist at first of numbers of deaths by sex and age for calendar years centred around the date of population censuses from 1871 to 2001. The availability of detailed death counts changes during the period. From 1930 on, all deaths are classified by single year of age; for previous years only deaths on 5-year age groups are provided. Centenarian deaths are tabulated in detail only from 1950 on. In 1901 they are nested into the last open age interval 90+. Otherwise, the deaths at age 100 and over are given as a lump sum.

Furthermore, census data on present population by sex and age are considered. The age classification of population as published by official agencies changes during the census years. In 1871 and 1931 all individuals are classified by single year of age up to age 99. From 1881 to 1921, only population by 5-year age groups is available up to the last open interval 90+ in 1901 and 100+, otherwise. In 1936 and 1951, a part of elderly people are collapsed into 5-year age groups (from age 75 and over in 1936 and from age 80 in 1951). Finally, from 1961 on, individuals are classified by single year of age (up to age 109 in the last two censuses and up to age 99, otherwise). In all the cases, except for 1901, 1991 and 2001 censuses, centenarians are given as a lump sum.

For ages from 0 to 4 both death and population counts are always provided by single year of age.

**Table 1** summarizes the structure of available data on number of deaths and individuals by sex as regards age classification for each census year occurred from 1871 to 2001.

<b>Census Year</b>	<b>Deaths</b>	<b>Population</b>
1871	5-year age groups from age 5 up to 100+	single year of age up to 100+
1881	5-year age groups from age 5 up to 100+	5-year age groups from age 10 up to 100+
1901	5-year age groups from age 5 up to 90+	5-year age groups from age 5 up to 90+
1911	5-year age groups from age 5 up to 100+	5-year age groups from age 15 up to 100+
1921	5-year age groups from age 5 up to 100+	5-year age groups from age 20 up to 100+
1931	single year of age up to 100+	single year of age up to 100+
1936	single year of age up to 100+	5-year age groups from age 75 up to 100+
1951	single year of age up to 103	5-year age groups from age 80 up to 100+
1961	single year of age up to 107	single year of age up to 100+
1971	single year of age up to 109	single year of age up to 100+
1981	single year of age up to 108	single year of age up to 100+
1991	single year of age up to 111	single year of age up to 109
2001	single year of age up to 114	single year of age up to 109

***Table 1: Number of deaths and present population by available age classification in censuses from 1871 to 2001.***

In order to estimate infant mortality into a complete life table, annual counts of live births by sex are also collected for the census calendar year ( $t$ ), the next one ( $t+1$ ) and the two previous years ( $t-2, t-1$ ).

#### 4. Steps for computing probabilities of dying and period life tables

There are 7 steps involved in computing probabilities of dying and period life tables for each census year. A part of them are taken up and discussed in detail in the later sections. If not specified, all subsequent calculations are made by sex separately.

1. Distributing death and population counts of unknown age. First of all, observations (both deaths and population counts) where age is unknown are distributed proportionately across the age range. The problem of unknown ages more often occurs for past years where the quality of data is generally lower. On the other hand, in the last period death and population numbers where age is unknown become rare.
2. Centring deaths around census date. Deaths counts are generally collected for more calendar years and then summarized by the arithmetic mean in order to centre such information around the census date. For example, 1961 census occurred in October, so the death numbers for 1960, 1961 and 1962 years are gathered and then summarized. The resulting age-specific death count is denoted by  ${}^{\bar{t}}D_x$ , where  $\bar{t}$  indicates we centred the numbers of deaths around the date when census at year  $t$  occurred.
3. Splitting 5-year age grouped data on deaths and population. Death and population counts are often available only in 5x1 configuration, so they are split into finer age categories, namely single year of age. The spline method is here used to split these observations for all ages below the open age interval. The main advantage of this method lies in the fact that it requires only that the data include counts for the first year of life and for the first five years of life. Other than these two restrictions, it does not matter whether the data are strictly in five-year



age groups (after age five) or in some other configuration. Moreover, there can be an open age interval above 90, 100, or some other age. Details of the computational methods are given in the later section.

4. Computing central death rates. For each age  $x$ , death rate consists of death counts divided by the number of person years. In our case, for one-year age group  $x$  and a single census year  $t$  (i.e., a 1x1 period death rate), we have the following formula:

$${}^t m_x = \frac{{}^t D_x}{{}^t P_x} \quad (1)$$

5. Estimating central death rates at high ages. The problem of old mortality estimation and its generating reasons have been already introduced at the beginning of this document. This question is here specified in terms of death rate estimation. At older ages where the randomness of mortality process is most noticeable, death rates are usually smoothed by fitting a mathematical function in order to obtain an improved representation of the underlying mortality profile. By this way, one takes into account for both numerical inconsistency of deaths and individuals at risk and loss of data quality which could occur at higher ages especially in past years<sup>1</sup>. Thus, we begin by smoothing observed death rates at older ages by fitting a logistic function based on ages 70 and above, for males and females separately. Indeed, the mortality model described by Kannisto (Kannisto *et al.*, 1994) as a special case of logistic function is applied. The main features of the Kannisto model and comparisons with other mortality functions are well described in the next sections. Once the parameters characterizing the logistic function have been estimated, fitted death rates can be computed. Then, they replace observed death rates for all ages at or above  $x^*$ , where  $x^*$  is defined as the lowest age where there are fewer than 100 deaths, but is constrained to  $80 \leq x^* \leq 95$ .

---

<sup>1</sup> A great improvement in the data quality is certainly represented by the collection of individual data by 'year of birth', rather than simply the age, introduced at regional level only in the first years after the second world war.

6. Converting into probabilities of dying. After obtaining smoothed age- and census-specific death rates for males and females, they are converted into probabilities of dying,  ${}^tq_x$ , by a standard method. Let  $a_x$  be the average number of years lived within the age interval  $[x, x + 1)$  for people dying at that age. We assume that  $a_x = 1/2$  for all single-year ages except age 0. Then we compute  ${}^tq_x$  from  ${}^tm_x$  and  $a_x$  according to the formula,

$${}^tq_x = \frac{{}^tm_x}{1 + (1 - a_x) \cdot {}^tm_x} = \frac{2 \cdot {}^tm_x}{2 + {}^tm_x} \quad (2)$$

for  $x=1, 2, \dots, 120$ .

Several methods for converting central death rates into probabilities of dying are available from demographic literature (e.g. the exponential approximation exploited in the seminal work of Kannisto *et al.* (1994), where  ${}^tq_x = 1 - \exp(-{}^tm_x)$ ), but they generally produce the same results<sup>2</sup>.

For infants, we adopt a different method for computing probability of death. At first the infant death rate is obtained as the ratio of deaths at age 0 and a weighted average of live births in the last two years. Indeed, we calculate this quantity for each census year  $t$  and the previous and next one,  $t-1$  and  $t+1$ , respectively, that is

$${}^j m_0 = \frac{{}^j D_0}{b^{t,j} B + b^{t+1,j-1} B} \quad (3)$$

for  $j=t-1, t, t+1$ . The weights,  $b'$  and  $b''$ , represent the proportion of deaths at age 0 in the study year and the previous one, respectively. Indeed, they change over the historical period depending on the infant mortality level (Livi Bacci, 1999). **Table 2** resumes the values of  $b'$  and  $b''$  employed for each census year.

---

<sup>2</sup> Another matter is when age grouped death rates must be converted into probabilities. Indeed, when the values of death rates are high the available methods usually give different results and the exponential approximation as firstly introduced by Reed and Merrell is shown to be preferable (Livi Bacci, 1999; Pressat, 1969).

Census Year	$b'$	$b''$
1871	0.6	0.4
1881	0.6	0.4
1901	0.61	0.39
1911	0.667	0.333
1921	0.667	0.333
1931	0.75	0.25
1936	0.75	0.25
1951	0.8	0.2
1961	0.85	0.15
1971	0.85	0.15
1981	0.95	0.05
1991	0.98	0.02
2001	0.98	0.02

**Table 2: Proportions of deaths at age 0 in the study year ( $b'$ ) and in the previous one ( $b''$ )**

Finally, a general estimate of infant probability of dying for each census year  $t$  is obtained by averaging the three death rates at age 0, that is

$${}^tq_0 = \frac{1}{3}({}^{t-1}m_0 + {}^t m_0 + {}^{t+1}m_0). \quad (4)$$

7. Arranging period life tables. The series of probabilities of dying by sex and age for each census year  $t$  are used to compute the other sex- and

age- specific quantities involved in life tables, namely the number of survivors and deaths,  $l_x$  and  $d_x$ , respectively, the probability of surviving,  $p_x$ , the person-years lived by the life-table population,  $L_x$ , the person-years remaining,  $T_x$ , and the life expectancy,  $e_x$ . The period life tables with sexes combined are then obtained as a weighted average of male and female probabilities of dying at each age  $x$ . Finest details are given in the specific section.

## 5. Splitting 5-year age grouped data by using spline interpolation

As we introduced above,  $n \times 1$  death and population counts are split into a  $1 \times 1$  format using cubic splines fitted to the cumulative distribution of data within each census year. In our case these aggregated data are always configured into 5-year age groups ( $n=5$ ).

Let  $Y_x$  the cumulative number of death or individuals within census year  $t$  up to age  $x$ , that is

$$Y_x = \sum_{u=0}^{x-1} {}^t D_u \quad \text{or} \quad Y_x = \sum_{u=0}^{x-1} {}^t P_u . \quad (5)$$

A cubic spline to  $Y_x$  can be expressed in the following form (McNeil *et al.*, 1977; Wilmoth *et al.*, 2007)

$$Y_x = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \sum_{i=1}^n \beta_i (x - k_i)^3 I(x \geq k_i) \quad (6)$$

where  $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \dots, \beta_n$  are the  $n+4$  parameters that must be estimated;  $I(\cdot)$  is an indicator function which equals one if the logical statement within brackets is true and zero if it is false;  $k_1, k_2, \dots, k_n$  are commonly named *knots* and they are those values of age  $x$  for which  $Y_x$  is known from the data. Actually, this interpolation method requires that  $k_1 = 1$  and  $k_2 = 5$ , that is deaths and population counts for the first year of life and for the first five years of life must be known from the data. Finally,  $k_n$  equals the lower limit of the last open age interval and there are no restrictions about it. The lower and upper boundaries are denoted by  $a$  and  $b$ , respectively, and we always have  $a = 0$  and  $b = \omega$ , where  $\omega$  is set arbitrarily to the maximum age of  $k_n + 5$ . Thus, we have  $n+2$  equations

from the  $n+2$  known values of  $Y_x$ , but  $n + 4$  unknown parameters that must be estimated.

In order to compute the coefficients involved into the spline function, two constraints specifying as many additional equations are needed. Typical solutions involve constraining the slope of the function at the boundaries. In other words, two equations based on the first derivative of  $Y_x$  at (or close to) the boundaries are usually considered. For a cubic spline the first derivative of  $Y_x$  is

$$Y'_x = \alpha_1 + \alpha_2 2x + \alpha_3 3x^2 + \sum_{i=1}^n \beta_i 3(x - k_i)^2 I(x \geq k_i). \quad (7)$$

Then, from the definitions of  $Y_x$  and first derivative it follows that

$$Y'_x \approx \bar{D}_x \text{ or, analogously, } Y'_x \approx P'_x. \quad (8)$$

Before to go on, a brief digression on the relationship between constraints and degrees of the spline function is needed. The number of constraints actually influences the choice of the degree. This is the main reason why in practice the constraints are firstly specified as desirable conditions for having a particular trend of the spline function. Then, the degree of the spline is directly drawn out. For example, in the seminal work on spline interpolation of demographic data, McNeil *et al.* showed that in order to find a smooth curve of an age-specific fertility schedule for which the only data available were average fertility rates by standard five-year age groups, four restrictions are needed. In fact, it is desirable that the cumulated fertility rates (i.e., the spline function) have zero derivative at the boundaries and also that for the age-specific fertility schedule to have zero derivative at its extremes, so that it reaches zero smoothly at ages 15 and 50. As a consequence, a spline of degree at least 5 is required. In general, the choice of the degree of the spline is arbitrary and the smallest value consistent with the number of restrictions is usually chosen. Then, if this value produces results which are unacceptable (e.g., negative values), a higher number of degrees is involved.

In the present case, at the upper boundary, the slope is constrained to be zero for both death and population numbers. Indeed, this choice is consistent with the usual tapering of the distribution of deaths and individuals at the oldest ages: there are no survivors and, as a consequence, no deaths at the maximum age  $\omega$ . On the other hand, the same constraint at the lower boundary would not be appropriate: deaths are highly

concentrated at age 0 and annual live births are (fortunately!) different from zero. Actually, another constraint is needed in order to fit at least a cubic spline. As a consequence, since for ages from 0 to 4 we always have both death and population counts by single year of age, we constrain the slope of the function at age 1 to equal deaths or, analogously, individuals at age 1<sup>3</sup>. Therefore, our constraints yield the following additional equations

$$Y'_\omega = 0 \quad (9)$$

and

$$Y'_1 = {}^i D_1 \text{ or } Y'_1 = {}^t P_1. \quad (10)$$

These two restrictions are shown to provide good estimates and, especially, to avoid the drawback of having negative death or population counts along the age  $x$ -axis. Therefore, we think there is no need to increase the degree of the spline function.

Then, fitting the cubic spline function consists of solving a system of  $n+4$  linear equations, which can be written in matrix form as follows

$$A\theta = y \quad (11)$$

where

$$y = \begin{bmatrix} Y_0 = 0 \\ Y_{k_1} = Y_1 \\ Y_{k_2} = Y_5 \\ \vdots \\ Y_{k_n} = Y_n \\ Y_\omega \\ Y'_1 = {}^i D_1 \\ Y'_\omega = 0 \end{bmatrix} \text{ or } y = \begin{bmatrix} Y_0 = 0 \\ Y_{k_1} = Y_1 \\ Y_{k_2} = Y_5 \\ \vdots \\ Y_{k_n} = Y_n \\ Y_\omega \\ Y'_1 = {}^t P_1 \\ Y'_\omega = 0 \end{bmatrix}, \quad (12)$$

---

<sup>3</sup> An exception is made for 1871, when population counts at age 1 were proved to be wrong. In this case, the constraint is made by using individual counts at age 2.

$$\theta = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \quad (13)$$

and

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & k_1 & k_1^2 & k_1^3 & 0 & 0 & \dots & 0 & 0 \\ 1 & k_2 & k_2^2 & k_2^3 & (k_2 - k_1)^3 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & k_n & k_n^2 & k_n^3 & (k_n - k_1)^3 & (k_n - k_2)^3 & \dots & (k_n - k_{n-1})^3 & 0 \\ 1 & \omega & \omega^2 & \omega^3 & (\omega - k_1)^3 & (\omega - k_2)^3 & \dots & (\omega - k_{n-1})^3 & (\omega - k_n)^3 \\ 0 & 1 & 2 & 3 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 2\omega & 3\omega & 3(\omega - k_1)^2 & 3(\omega - k_2)^2 & \dots & 3(\omega - k_{n-1})^2 & 3(\omega - k_n)^2 \end{bmatrix} \quad (14)$$

The parameters in the spline are thus given by

$$\theta = A^{-1}y. \quad (15)$$

In other words, the problem of fitting the spline reduces to the problem of inverting matrix  $A$ .

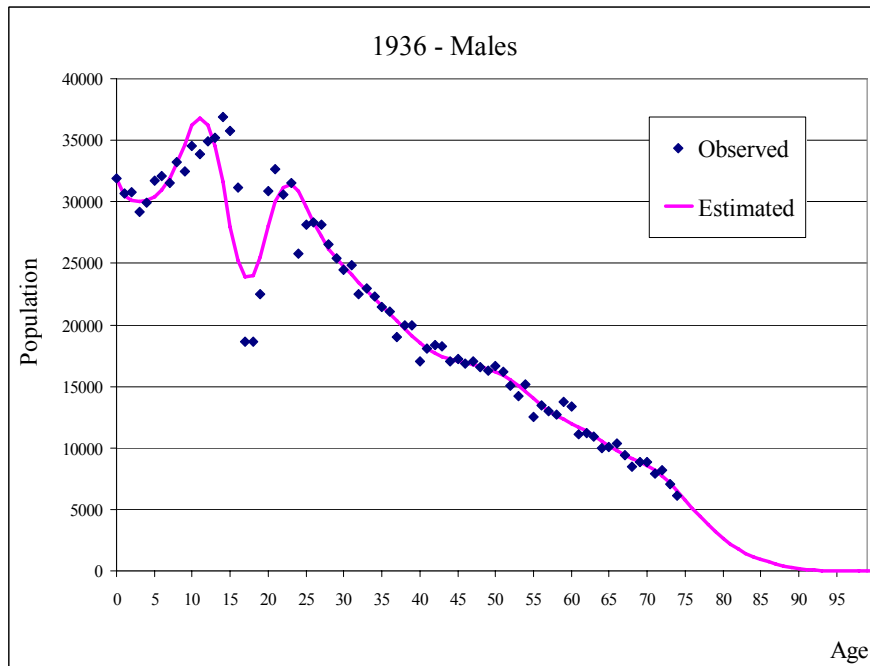
Once the vector of coefficients  $\theta$  is obtained by this method, the estimated equation is used to find the series of fitted values  $\hat{Y}_x$ . Then, deaths and individuals by single years of age are estimated by differencing successive fitted values:

$${}^i\hat{D}_x = \hat{Y}_{x+1} - \hat{Y}_x \text{ or } {}^t\hat{P}_x = \hat{Y}_{x+1} - \hat{Y}_x \quad (16)$$

for  $x=0, 1, 2, \dots, k_{n-1}$ .

Moreover, the estimated spline function is used to split death and population counts in the last open age interval, up to the age corresponding to reliable values (i.e.  ${}^i\hat{D}_x > 0$ ,  ${}^t\hat{P}_x > 0$  and  ${}^i\hat{D}_x \leq {}^t\hat{P}_x$ <sup>4</sup>).

The spline interpolation method is shown to work very well, at least on our data. As an example, **Figure 1** proves the good fitting of spline curve when both observed and estimated values are compared. This refers to male population counts in 1936, when single age quantities are available up to age 74 and then the spline interpolation method is used to split 5-year age groups up to 100+.



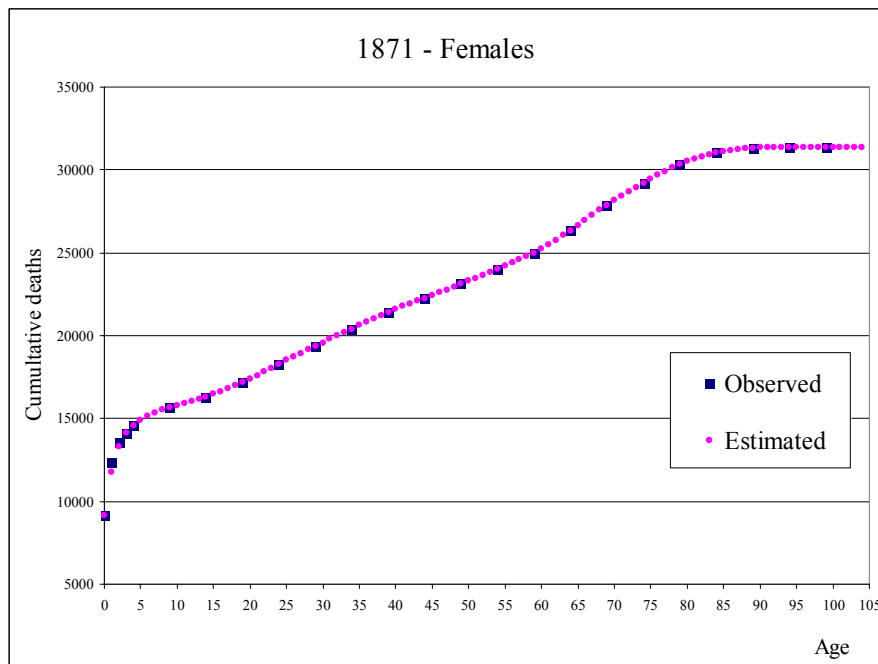
**Figure 1: Observed VS estimated values by using spline interpolation: 1936, Male population.**

---

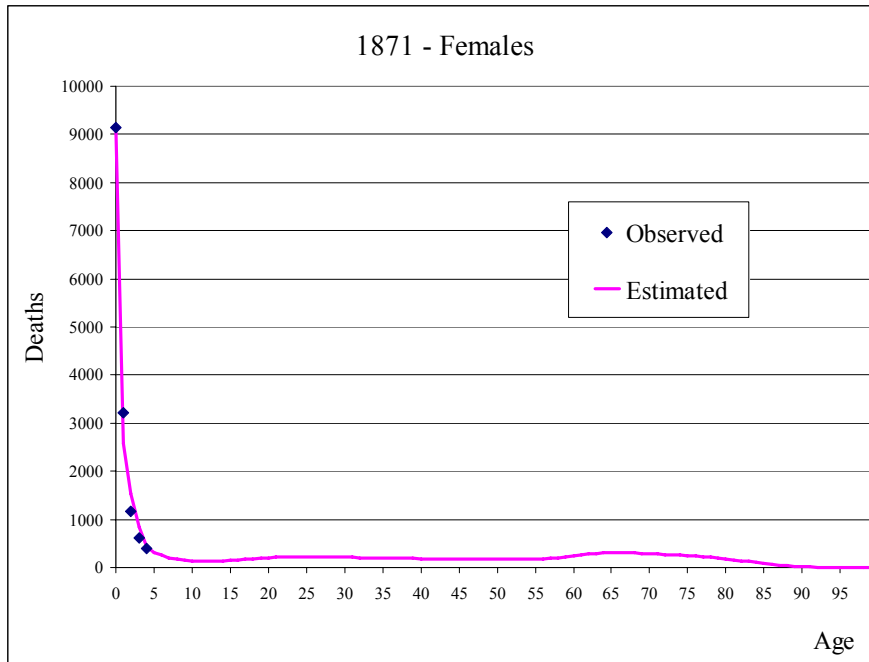
<sup>4</sup> If these constraints are not complied then the splitting procedure stops at the previous age interval.



Moreover, the spline method is shown to work well also when the observed data are aggregated into 5-year age groups even from age 5 on (*Figure 2* and *Figure 3*).



**Figure 2: Cumulative deaths distribution (observed VS estimated by cubic spline method): 1871, Females.**



**Figure 3: Spline interpolation method: 1871, Female deaths.**

Once the procedure has been applied to all aggregated data, deaths and individual counts expressed in a 1x1 format are obtained into the age intervals summarized in **Table 3**.

Census Year	Male	Female
from 1871 to 1971	0-99	0-99
1981	0-104	0-102
1991 – 2001	0-109	0-109

**Table 3: Age intervals with data expressed in a 1x1 format after the spline method.**

## 6. Models of old-age mortality

Mortality models generally attempt to find the function that summarizes the way in which the probability of dying depends on age. Literature gives them many potential and operative applications. As well as in the construction of life tables, these models are also useful in the projection of population numbers, such as the extrapolation of probabilities of dying at extreme ages, or in the validation of biological theories about the nature of the ageing process.

The earliest and still one of the most applied model is the “law of mortality” discovered by Gompertz (1825). By studying the survival curves in the life tables he found empirically that the force of mortality,  $\mu_x$ , increases constantly with age at a steady exponential rate. Thus the model can be written as

$$\mu_x = \alpha \cdot \exp(\beta x) \quad (17)$$

where  $\alpha$  represents the base level of mortality and  $\beta$  is the relative increase in the force of mortality, which is thus supposed to be constant over ages. Although this model works very well over much of the age range, it shows some problems when applied to infancy, youth and very old ages.

Several other functions of mortality have been discovered by various authors during the following years. A first attempt to improve Gompertz’s law was made by Makeham (1860). He added a constant term  $\delta$ , which includes the risk of death from all causes which do not depend on age, so that:

$$\mu_x = \alpha \cdot \exp(\beta x) + \delta . \quad (18)$$

At high ages the constant  $\delta$  is shown to be very small, so the difference between the Gompertz and Makeham laws is negligible.

In 1932 Perks found empirically that the values of  $\mu_x$  in a life table which he was examining could be fitted by a logistic function. The resulting logistic mortality model is then expressed as

$$\mu_x = \frac{\alpha \cdot \exp(\beta x)}{1 + \gamma \cdot \exp(\beta x)} . \quad (19)$$

We can note that this models includes Gompertz's law as the special case when  $\gamma=0$ .

Another mortality model was proposed by Weibull (1951). This would imply that the force of mortality is proportional to a power of age, that is

$$\mu_x = \alpha \cdot x^\beta . \quad (20)$$

Up to now, the described models of mortality are explanatory, suggesting a possible mechanism which may help to explain why the formula works. Usually the supposed mechanism is supported by biological theories on aging process. So, for example, most modern attempts to explain the Gompertz's law are linked to steady bodily deterioration, due to the accumulation of molecular and cellular damage, over the age ranges concerned.

Indeed, other models are descriptive. They give a function which simply fits the data in a particular range of ages, but no reason explains why this should be so, and hence nothing guarantees that the formula will continue to work in other circumstances.

The model proposed by Heligman & Pollard (1980), which is intermediate between the Weibull and logistic models, is a descriptive model with eight parameters and it can be formulated as follows:

$$\frac{q_x}{p_x} = A^{(x+B)C} + D \cdot \exp(-E(\ln x - \ln F)^2) + GH^x . \quad (21)$$

However, above age 50 the first two terms can be neglected and their expression reduces to

$$\frac{q_x}{p_x} = GH^x . \quad (22)$$

Another descriptive model was proposed by Coale & Kisker (1990) for the purpose of interpolating  $\mu_x$  in the range of ages from 85 to 110. Their idea was that  $\ln(\mu_x)$  can be fitted by a quadratic function of age  $x$ , that is

$$\ln(\mu_x) = a + bx + cx^2 \quad (23)$$

where  $c$  is negative.

In a chronological order the last model was proposed by Kannisto *et al.* (1994). It is an explanatory model and it is proved to be one of the best and simultaneously simplest functions for the approximation of probability of dying at high ages. Kannisto noted that a easier case of the logistic model worked well at least on modern data. He did not propose a general law, but

simply observed an empirical finding. The basic idea is again that the relative increase in the force of mortality decreases at old ages (generally from 75-80 years of age on), instead of being constant as in Gompertz formula. Kannisto and his scholars started from the most general logistic model which has four independent parameters and can be expressed as

$$\mu_x = \frac{\kappa \cdot \alpha \cdot \exp(\beta x)}{1 + \alpha \cdot \exp(\beta x)} + \delta. \quad (24)$$

They firstly showed that this general model is not very useful in practice since a simpler version with only three parameter is more robust and that the fitted parameter  $\kappa$  tends to be near 1 (Thatcher *et al.*, 1998; Thatcher, 1999). As a consequence the three-parameter model follows

$$\mu_x = \frac{\alpha \cdot \exp(\beta x)}{1 + \alpha \cdot \exp(\beta x)} + \delta. \quad (25)$$

Such model was fitted by Thatcher (1999) to some data from the mediaeval period to the present and at all adult ages (from age 30 on), not just high ages. He proved the model fits data rather well and the estimated parameters can be interpreted revealing a long-term development of adult mortality.

Secondly, two important properties have been pointed out. At ages up to 70 or so, that is when  $\alpha \cdot \exp(\beta x) / [1 + \alpha \cdot \exp(\beta x)]$  is small, the three-parameter model (25) is the law (18) proposed by Makeham. The second property, already introduced above, is that at high ages in historical periods and at almost all ages today the parameter  $\delta$  is small compared to  $\alpha \cdot \exp(\beta x) / [1 + \alpha \cdot \exp(\beta x)]$ . Indeed, Thatcher (1999) observed a wide fall during the long-historical period in the values of  $\delta$ , becoming slight in the recent years. This pattern on  $\delta$  reflects the gradual elimination of infection diseases, which were major cause of child and adult deaths, but negligible at high ages. As a consequence, the relevant formula for the old mortality model proposed by Kannisto *et al.* (1994) can be simply expressed as

$$\mu_x = \frac{\alpha \cdot \exp(\beta x)}{1 + \alpha \cdot \exp(\beta x)}. \quad (26)$$

As we introduced above, a recent study concluded that Kannisto formula fits the mortality pattern at old ages usually better than other mortality models (Thatcher *et al.*, 1998). This is the main reason why we use it to fit our data on mortality to estimate death rates at high ages.

Moreover, if the Kannisto model is shown to work well in estimation of probabilities of dying at high ages on modern data, here we attempt to apply it even on less recent years.

## 7. Fitting the Kannisto model

### 7.1. Choosing the fitting age range

The Kannisto mortality model is not fitted to the data at the whole of the age range. Indeed, such model involves old ages, namely when the relative increase in the force of mortality starts to decrease. In modern data, it happens generally from 75-80 years of age, as for less recent periods it would be investigated. Therefore, we firstly analyze the trend of relative increases in the death rates, or analogously in the force of mortality, within each census year  $t$ . We do that to avoid the risk of misspecification of the age interval for the model fitting, which can generate biased estimates in death rates at highest ages. Actually, in our case this choice is reduced to the lower age limit, since the upper one is given by the highest age corresponding to the last death rate available from the data expressed in a 1x1 format. Besides, it is preferable to choose a single fitting age range. By this way, the same age reference for the whole of the period from 1871 to 2001 ensures reasonable comparisons in the mortality trend at old ages.

The relative mortality increase by advancing age is usually measured by a rate parameter,  $k(x)$ , defined by Horiuchi and Cole as the relative derivative of  $\mu_x$  (Horiuchi and Coale, 1990; Kannisto, 1996; Horiuchi and Wilmoth, 1998):

$$k_x = \frac{d\mu_x / dx}{\mu_x} = \frac{d \ln(\mu_x)}{dx}. \quad (27)$$

This measure, called *life-table aging rate*, has proved to be a powerful tool for detecting variations in mortality, as considerable accelerations and decelerations in the mortality pattern can escape a visual inspection such as simply plotting  $\mu_x$  in logarithmic scale. An age-related increase in  $k(x)$  implies an acceleration in the age pattern of mortality, conversely, a decrease corresponds to a deceleration (Horiuchi and Wilmoth, 1998).

For the Gompertz law of mortality we find that  $k(x) = \beta$ , according to the idea of constant increase of mortality at all ages introduced above. Conversely, for the Kannisto model we have  $k(x) = \beta \cdot (1 - \mu_x)$ , that is the relative increase equals  $\beta$  at lower ages when the force of mortality  $\mu_x$  is small, then falls tending to zero as age increases (Kannisto, 1996).

In practice, measuring accurately  $k(x)$  is not easy, as there are fluctuations in  $\mu_x$  especially at high ages. A typical solution lies in taking 5-year of age groups and then estimating aging rate as follows:

$$\hat{k}(x) = \frac{\ln(m_{x,x+5}) - \ln(m_{x-5,x})}{5} \quad (28)$$

where  $m_{x,x+5}$  and  $m_{x-5,x}$  are successive 5-year of age death rates (Horiuchi and Coale, 1990; Horiuchi and Wilmoth, 1998). Then, the corresponding standard error can be approximated by

$$\hat{\sigma}_{\hat{k}(x)} = \frac{1}{5} \sqrt{\frac{1}{D_{x,x+5}} + \frac{1}{D_{x-5,x}}} \quad (29)$$

where  $D_{x,x+5}$  and  $D_{x-5,x}$  are the number of deaths in the age intervals from  $x$  to  $x+5$  and from  $x-5$  to  $x$ , respectively (Wilmoth, 1995; Horiuchi and Wilmoth, 1998).

In **Table 4** the estimates of  $k(x)$  derived from the 5-year age groups of  $m_x$  through formula (28) and corresponding confidence intervals<sup>5</sup> are given for data here considered at each census year  $t$  and at ages from 60 to 95.

---

<sup>5</sup> The confidence intervals' width corresponds to two times the standard errors estimated through formula (29) to  $\hat{k}(x)$ .

	<b>1871</b>		<b>1881</b>		<b>1901</b>		<b>1911</b>	
<b>Males</b>								
k(60)	0.045	(0.030 - 0.061)	0.061	(0.046 - 0.075)	0.098	(0.083 - 0.114)	0.090	(0.074 - 0.106)
k(65)	0.098	(0.084 - 0.113)	0.113	(0.099 - 0.127)	0.102	(0.089 - 0.116)	0.093	(0.079 - 0.106)
k(70)	0.080	(0.066 - 0.094)	0.075	(0.062 - 0.088)	0.108	(0.096 - 0.120)	0.112	(0.099 - 0.124)
k(75)	0.099	(0.084 - 0.114)	0.079	(0.065 - 0.093)	0.090	(0.078 - 0.103)	0.105	(0.093 - 0.117)
k(80)	0.048	(0.030 - 0.067)	0.053	(0.034 - 0.072)	0.083	(0.068 - 0.097)	0.082	(0.068 - 0.096)
k(85)	0.048	(0.020 - 0.076)	0.011	(-0.019 - 0.042)	0.080	(0.056 - 0.104)	0.074	(0.053 - 0.095)
k(90)	0.043	(-0.011 - 0.097)	0.071	(0.009 - 0.133)	0.051	(0.006 - 0.096)	0.020	(-0.025 - 0.064)
k(95)	0.073	(-0.041 - 0.187)	-0.024	(-0.169 - 0.121)	-		0.042	(-0.106 - 0.189)
<b>Females</b>								
k(60)	0.065	(0.048 - 0.082)	0.062	(0.046 - 0.078)	0.100	(0.083 - 0.116)	0.100	(0.082 - 0.118)
k(65)	0.108	(0.093 - 0.123)	0.125	(0.110 - 0.140)	0.103	(0.089 - 0.117)	0.107	(0.092 - 0.121)
k(70)	0.072	(0.057 - 0.087)	0.069	(0.055 - 0.083)	0.109	(0.096 - 0.122)	0.106	(0.093 - 0.120)
k(75)	0.101	(0.084 - 0.117)	0.092	(0.077 - 0.107)	0.090	(0.077 - 0.103)	0.097	(0.084 - 0.110)
k(80)	0.035	(0.016 - 0.054)	0.049	(0.030 - 0.069)	0.075	(0.060 - 0.091)	0.095	(0.080 - 0.110)
k(85)	0.056	(0.028 - 0.085)	0.052	(0.021 - 0.083)	0.065	(0.041 - 0.089)	0.057	(0.037 - 0.078)
k(90)	-0.025	(-0.084 - 0.034)	-0.008	(-0.066 - 0.050)	0.028	(-0.015 - 0.071)	0.067	(0.026 - 0.108)
k(95)	0.070	(-0.050 - 0.190)	-0.072	(-0.223 - 0.078)	-		-0.013	(-0.122 - 0.097)

*Table 4: Estimates of  $k(x)$  (CI) in the period 1871-2001.*



	<b>1921</b>		<b>1931</b>		<b>1936</b>		<b>1951</b>	
<b>Males</b>								
k(60)	0.097	(0.082 - 0.113)	0.084	(0.068 - 0.099)	0.086	(0.070 - 0.101)	0.076	(0.060 - 0.091)
k(65)	0.094	(0.080 - 0.107)	0.095	(0.082 - 0.108)	0.093	(0.080 - 0.107)	0.093	(0.079 - 0.106)
k(70)	0.106	(0.094 - 0.118)	0.091	(0.079 - 0.103)	0.096	(0.084 - 0.108)	0.111	(0.099 - 0.123)
k(75)	0.095	(0.084 - 0.107)	0.102	(0.090 - 0.114)	0.101	(0.090 - 0.112)	0.101	(0.090 - 0.113)
k(80)	0.102	(0.089 - 0.116)	0.089	(0.076 - 0.103)	0.097	(0.084 - 0.109)	0.103	(0.091 - 0.115)
k(85)	0.064	(0.044 - 0.084)	0.078	(0.059 - 0.097)	0.073	(0.056 - 0.091)	0.088	(0.074 - 0.103)
k(90)	0.053	(0.014 - 0.092)	0.053	(0.013 - 0.093)	0.070	(0.036 - 0.103)	0.063	(0.037 - 0.090)
k(95)	0.009	(-0.100 - 0.118)	0.077	(-0.045 - 0.198)	0.080	(-0.014 - 0.175)	-0.010	(-0.093 - 0.072)
<b>Females</b>								
k(60)	0.097	(0.079 - 0.114)	0.083	(0.065 - 0.101)	0.090	(0.072 - 0.108)	0.084	(0.066 - 0.103)
k(65)	0.118	(0.103 - 0.133)	0.103	(0.088 - 0.119)	0.099	(0.084 - 0.115)	0.123	(0.107 - 0.138)
k(70)	0.106	(0.092 - 0.119)	0.113	(0.100 - 0.126)	0.109	(0.096 - 0.122)	0.119	(0.105 - 0.132)
k(75)	0.103	(0.091 - 0.116)	0.106	(0.093 - 0.118)	0.101	(0.090 - 0.113)	0.117	(0.105 - 0.129)
k(80)	0.088	(0.074 - 0.101)	0.090	(0.077 - 0.103)	0.097	(0.084 - 0.110)	0.099	(0.087 - 0.110)
k(85)	0.062	(0.043 - 0.082)	0.089	(0.072 - 0.107)	0.086	(0.069 - 0.102)	0.087	(0.074 - 0.101)
k(90)	0.049	(0.013 - 0.085)	0.066	(0.031 - 0.100)	0.077	(0.048 - 0.105)	0.082	(0.060 - 0.104)
k(95)	0.011	(-0.067 - 0.089)	0.057	(-0.028 - 0.141)	0.107	(0.037 - 0.178)	0.028	(-0.024 - 0.080)

*Table 4: (Continued)*

	<b>1961</b>	<b>1971</b>	<b>1981</b>	<b>1991</b>	<b>2001</b>
<b>Males</b>					
k(60)	0.086 (0.072 - 0.100)	0.095 (0.082 - 0.108)	0.085 (0.072 - 0.099)	0.104 (0.089 - 0.119)	0.087 (0.070 - 0.104)
k(65)	0.085 (0.073 - 0.097)	0.087 (0.077 - 0.098)	0.084 (0.073 - 0.096)	0.091 (0.079 - 0.103)	0.102 (0.088 - 0.116)
k(70)	0.088 (0.077 - 0.100)	0.095 (0.084 - 0.105)	0.096 (0.087 - 0.106)	0.078 (0.067 - 0.088)	0.102 (0.090 - 0.113)
k(75)	0.089 (0.079 - 0.100)	0.096 (0.086 - 0.106)	0.099 (0.090 - 0.108)	0.102 (0.092 - 0.112)	0.105 (0.096 - 0.115)
k(80)	0.108 (0.097 - 0.119)	0.079 (0.069 - 0.090)	0.092 (0.082 - 0.101)	0.094 (0.085 - 0.102)	0.091 (0.082 - 0.100)
k(85)	0.074 (0.060 - 0.088)	0.090 (0.077 - 0.102)	0.085 (0.073 - 0.097)	0.095 (0.085 - 0.106)	0.115 (0.106 - 0.125)
k(90)	0.076 (0.052 - 0.099)	0.069 (0.049 - 0.089)	0.080 (0.063 - 0.098)	0.101 (0.085 - 0.116)	0.084 (0.073 - 0.096)
k(95)	-0.014 (-0.068 - 0.040)	-0.038 (-0.084 - 0.008)	0.067 (0.029 - 0.104)	0.072 (0.042 - 0.101)	0.074 (0.052 - 0.095)
<b>Females</b>					
k(60)	0.095 (0.076 - 0.114)	0.086 (0.069 - 0.104)	0.100 (0.081 - 0.119)	0.089 (0.068 - 0.109)	0.062 (0.038 - 0.085)
k(65)	0.116 (0.101 - 0.132)	0.097 (0.082 - 0.112)	0.088 (0.071 - 0.104)	0.090 (0.073 - 0.106)	0.093 (0.074 - 0.113)
k(70)	0.115 (0.102 - 0.128)	0.123 (0.111 - 0.136)	0.118 (0.105 - 0.130)	0.104 (0.090 - 0.118)	0.112 (0.096 - 0.127)
k(75)	0.115 (0.104 - 0.126)	0.129 (0.118 - 0.139)	0.131 (0.121 - 0.142)	0.126 (0.114 - 0.137)	0.124 (0.113 - 0.136)
k(80)	0.114 (0.103 - 0.125)	0.119 (0.109 - 0.128)	0.125 (0.116 - 0.135)	0.124 (0.115 - 0.133)	0.119 (0.109 - 0.129)
k(85)	0.085 (0.072 - 0.098)	0.097 (0.086 - 0.107)	0.111 (0.101 - 0.120)	0.130 (0.121 - 0.138)	0.143 (0.135 - 0.152)
k(90)	0.080 (0.061 - 0.100)	0.086 (0.071 - 0.102)	0.100 (0.088 - 0.112)	0.117 (0.107 - 0.127)	0.107 (0.099 - 0.115)
k(95)	0.049 (0.010 - 0.088)	0.037 (0.005 - 0.069)	0.064 (0.040 - 0.087)	0.081 (0.064 - 0.099)	0.102 (0.090 - 0.114)

*Table 4: (Continued)*

Several issues can be discussed by observing the results summarized in **Table 4**. Indeed, they will be remarked in detail in the last section. For now, we aim only to detect the lower age limit of the fitting range, namely the age at which the estimate of  $k(x)$  falls off. Under this perspective, we can note as the estimate of  $k(x)$  starts to decrease at ages varying from 70 to 95. This critical age seems to gradually increase over the period here considered: in the post-war era it always goes over or at least equals 85 years of age, as in the less recent years it touches lower values, up to the minimum of 70 years of age in 1881. As a consequence, we choose to fit the mortality model to the range from 70 years of age to the maximum age available from data (or from the splitting procedure) depending on year for the whole period here considered.

## 7.2. Parameter estimation

As we showed above, the mortality model for old ages proposed by Kannisto involves the force of mortality  $\mu_x$ , which is proved to be quite close to the central death rate  $m_x$ . Indeed, an approximate relationship among these quantities can be expressed as follows (Pollard, 1973):

$$\mu_{x+0.5} \cong m_x. \quad (30)$$

This relationship is exploited in order to smooth observed (and if it is the case split) death rates  ${}^t m_x$  by single year of age within the census year  $t$ . We do that by fitting the Kannisto model

$$\mu_{x+0.5} = \frac{\alpha \cdot \exp[\beta \cdot (x + 0.5)]}{1 + \alpha \cdot \exp[\beta \cdot (x + 0.5)]} \quad (31)$$

with an asymptote equal to one and where we require the parameters  $\alpha$  and  $\beta$  to be positive or at least equal to zero.

Parameter estimates  $\hat{\alpha}$  and  $\hat{\beta}$  are derived by using the maximum likelihood method. More specifically, we assume that  ${}^i D_x$  is distributed like a *Poisson* with parameter  $\lambda = {}^t P_x \cdot \mu_{x+0.5}$ . Thus, the log-likelihood function is derived as follows:

$$\begin{aligned} \log L(\alpha, \beta) = l(\alpha, \beta) &= \sum_x \left[ {}^i D_x \cdot \log(\mu_{x+0.5}) - {}^t P_x \cdot \mu_{x+0.5} + {}^i D_x \cdot \log({}^t P_x) - \log({}^i D_x!) \right] \\ &\propto \sum_x \left[ {}^i D_x \cdot \log(\mu_{x+0.5}) - {}^t P_x \cdot \mu_{x+0.5} \right]. \end{aligned} \quad (32)$$

Substituting the Kannisto function of  $\mu_{x+0.5}$  into the log-likelihood equation and then maximizing respect to the parameters  $\alpha$  and  $\beta$  we eventually derive the parameter estimates  $\hat{\alpha}$  and  $\hat{\beta}$ .

### 7.3. Replacing observed with fitted death rates

Smoothed death rates  ${}^t\hat{m}_x$  are easily obtained by substituting the estimates  $\hat{\alpha}$  and  $\hat{\beta}$  into the Kannisto equation and by exploiting the approximate relationship among force of mortality and death rates. Fitted death rates replace observed death rates for all ages at or above the greater of 80 or above the lowest age where there are fewer than 100 deaths, and for all ages at or above age 95 regardless of the number of deaths. The values corresponding to 100 deaths is fixed to take into account for consistency of the data and corresponding estimate of death risk. This problem typically concern male deaths which are fewer than female deaths at older ages, especially for recent years when differences between sexes become larger.

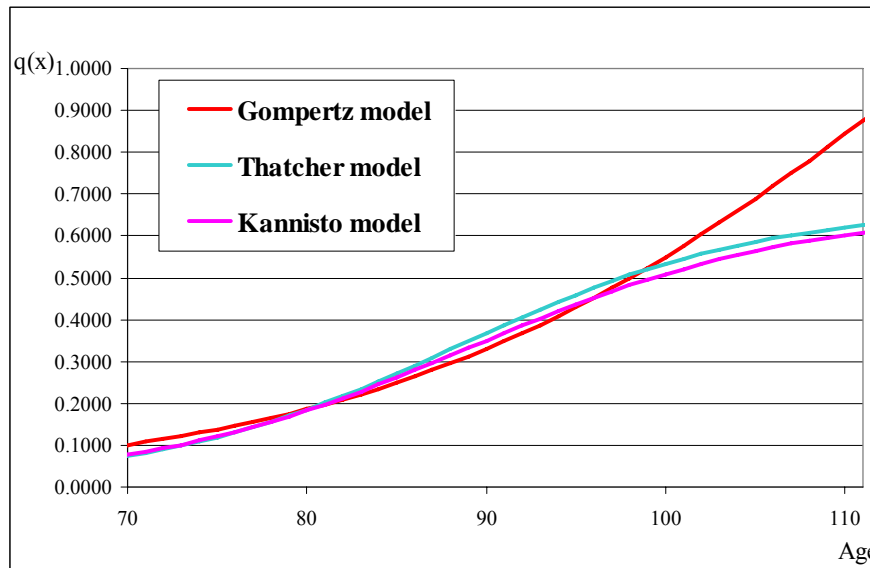
Furthermore, the fitted Kannisto model and corresponding smoothed death rates are also used in order to produce extrapolations up to age 110 for every census year within the period 1871-2001.

### 7.4. Comparisons with some other mortality models

Before going on, it is noteworthy to see how the Kannisto model compares with the Gompertz's law of mortality (17) and the three-parameter model (25) proposed by Thatcher (Thatcher, 1999). We will do that by showing the results obtained only for two years: 1991 and 1901. The first one represents modern data and the second one belongs to the historical era.

Parameter estimates are derived again by using the maximum likelihood method and by assuming  ${}^iD_x$  are distributed like a *Poisson* with parameter  $\lambda = {}^tP_x \cdot \mu_{x+0.5}$ . The fitting age range considered for the Gompertz formula is the same as employed for the Kannisto model (i.e., from age 70 on). Conversely, the Thatcher model is fitted to the data from age 30 on, according to the seminal work (Thatcher, 1999).

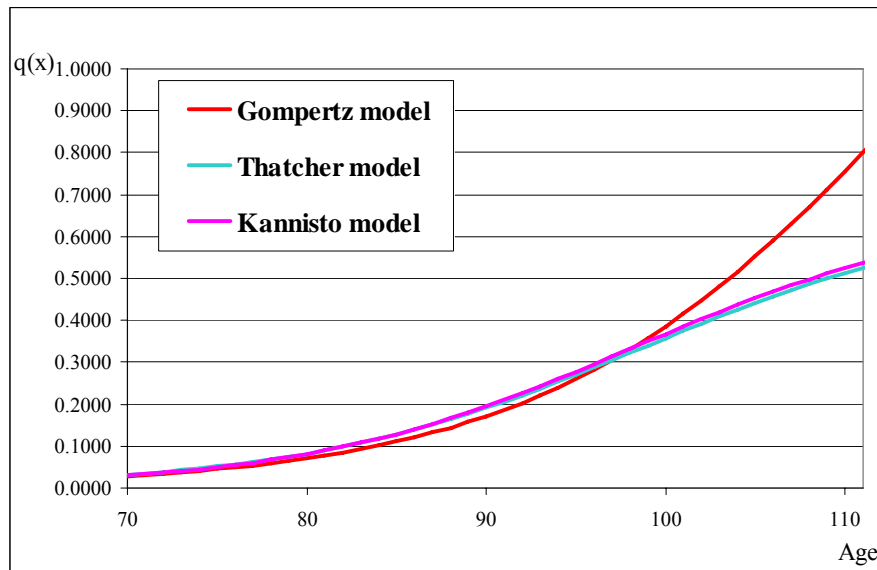
**Figures 4 and 5** show as the relationship among the Gompertz's law of mortality and the Kannisto and Tacher models is the same as reported into the reference works on mortality models (Kannisto, 1994; Thatcher, 1998 and 1999). Indeed, the plotted probabilities of dying start to significantly diverge at nearly 95 years of age independently on which particular historical data set is used to fit the models. In fact, the divergence is inevitable, since in the Gompertz model the force of mortality increases exponentially with age, as in the Kannisto and Thatcher models, it tends to a constant. The same structure is naturally reflected into the patterns of the probabilities of dying.



**Figure 4: Probabilities of dying estimated by the Gompertz, Thatcher and Kannisto models: 1901, Males.**

When the Kannisto and Thatcher equations are compared, the features firstly discussed by Thatcher (1999) can be observed even for our data. In fact, the parameter  $\gamma$  becomes negligible over time (from 0.0048 in 1901 to 0.0008 in 1991), reflecting the gradual elimination of death causes due to infectious diseases. Secondly, there are no significant differences among

the two corresponding curves at high ages (Figures 4 and 5), which proves the substantial equivalence of the two models for estimating old mortality. Therefore, this can account for our decisive choice of fitting the Kannisto model.



*Figure 5: Probabilities of dying estimated by the Gompertz, Thatcher and Kannisto models: 1991, Males.*

## 8. Period life tables

Once the series of central death rates are converted into probabilities of dying, they are employed to arrange complete period life tables, separately by sex and together. Indeed, the set of quantities involved in the life tables represent the first tools in order to compare the pattern of mortality by sex and by time.

Starting from age- and year- specific probabilities of dying,  ${}^tq_x$ , the quantities of a generic life table can be defined as following:

- the probability of surviving from age  $x$  to  $x + 1$ , denoted by  ${}^t p_x$ , is
 
$${}^t p_x = 1 - {}^t q_x \quad (33)$$

- the number of survivors at age  $x$  equals

$${}^t l_x = {}^t l_0 \cdot \prod_{i=0}^{x-1} {}^t p_i \quad (34)$$

where  ${}^t l_0$  equals to 100,000, namely the radix of the table;

- the distribution of deaths by age  $x$  is
 
$${}^t d_x = {}^t l_x \cdot {}^t q_x; \quad (35)$$

- the person-years lived by the life-table population from age  $x$  to  $x + 1$  are

$${}^t L_x = \frac{{}^t l_x + {}^t l_{x+1}}{2} \quad (36)$$

with  ${}^t L_0 = b'' \cdot {}^t l_0 + b' \cdot {}^t l_1$ <sup>6</sup>;

- the person-years remaining for individuals of age  $x$  equal

$${}^t T_x = \sum_{i=x}^{110} {}^t L_i \quad (37)$$

- remaining life expectancy at age  $x$  is

$${}^t e_x = \frac{{}^t T_x}{{}^t l_x}. \quad (38)$$

The period life tables for women and men together are built by combining the sex-specific probabilities of dying at each age  $x$  through a sort of weighted average. The method is described in detail in the following list:

---

<sup>6</sup>  $b'$  and  $b''$  are the same proportions as employed for computing  $m_0$ .

1. Male probabilities of dying, denoted by  ${}^tq_x^M$ , are employed to compute a new distribution of deaths and survivors, denoted by  ${}^td_x^{M*}$  and  ${}^tl_x^{M*}$ , respectively, based on a new radix of the table, namely  ${}^tl_0^{M*} = 106,000$ .
2. The series of these new male deaths and survivors is added to the series of female deaths and survivors in corresponding life tables at each age  $x$ , that is

$${}^td_x^{T*} = {}^td_x^{M*} + {}^td_x^F \quad (33)$$

$${}^tl_x^{T*} = {}^tl_x^{M*} + {}^tl_x^F \quad (34)$$

where superscripts T, F, and M represent total, female, and male, respectively.

3. The ratio of these quantities provides the total probability of dying at age  $x$

$${}^tq_x^T = \frac{{}^td_x^{T*}}{{}^tl_x^{T*}}. \quad (35)$$

4. Finally, to complete the life table calculation with sexes combined, the other quantities ( ${}^td_x^T$ ,  ${}^tl_x^T$ ,  ${}^tp_x^T$ ,  ${}^tL_x^T$ ,  ${}^tT_x^T$  and  ${}^te_x^T$ ) are computed through the same formulas as defined previously separately by sex.

## 9. Discussion

Once the construction of the Emilia-Romagna's period life tables separately by sex and together<sup>7</sup> has been completed for all the census years  $t$  from 1871 to 2001, they can be employed in several ways in order to make historical comparisons as regards to the old mortality development in Emilia-Romagna. Moreover, a set of additional demographic indicators can be built *ad hoc* and then compared (e.g., modal age at death). In other words, the construction of period life tables actually only represents the starting point towards a critical analysis of mortality under a long historical perspective. Indeed, this is a widespread and complex subject, that need to be developed and will be discussed in detail into a separate work. For now,

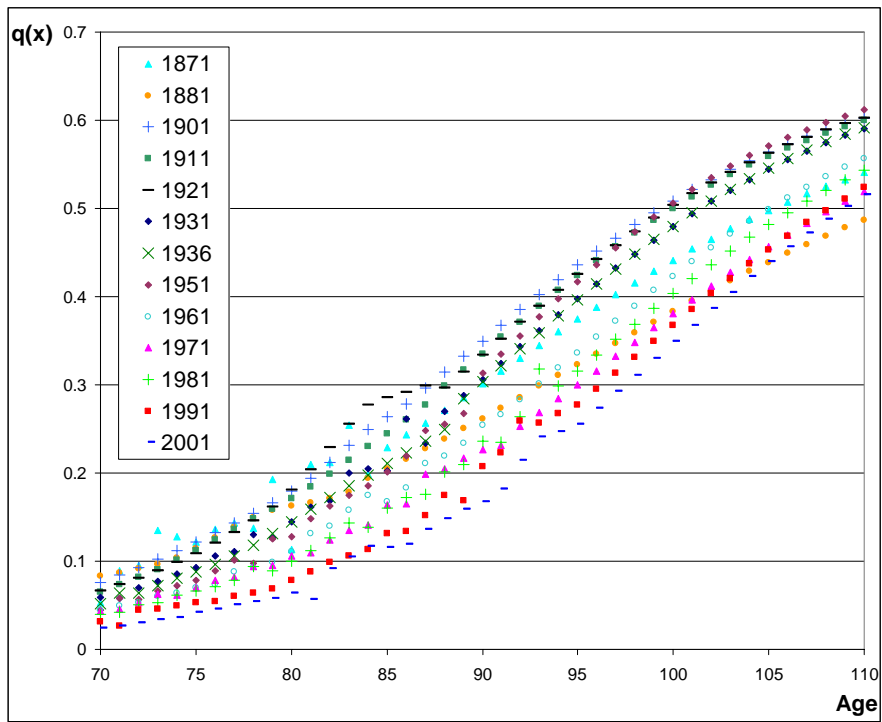
---

<sup>7</sup> The set of such period life tables is available upon request to the author.

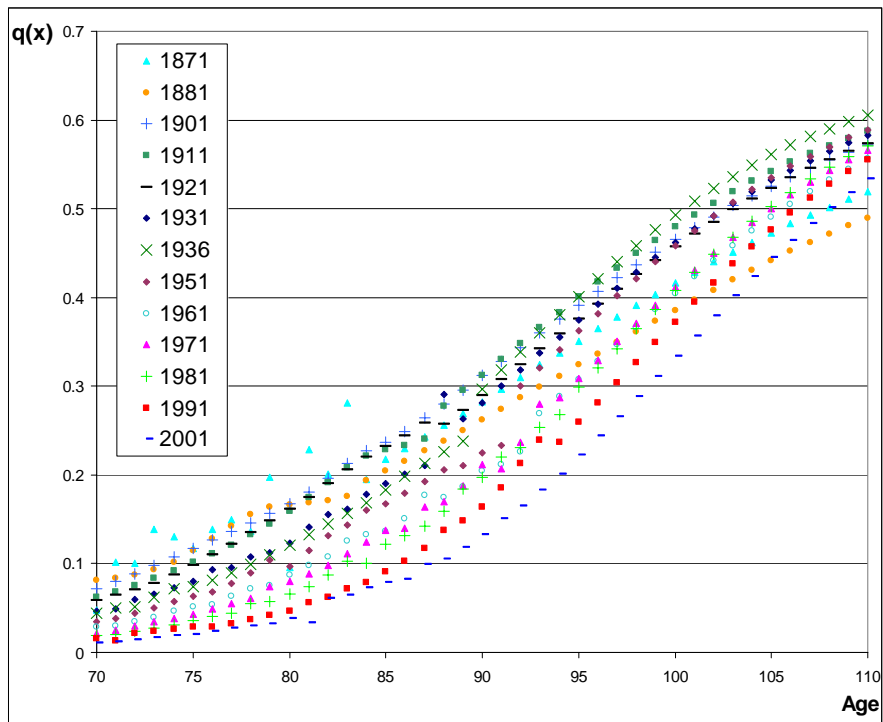


we would simply introduce some fundamental issues concerning these results based on some quantities computed during the construction. Finally, a brief comparison among the competing procedures employed by ISTAT and by Wilmoth *et al.* for the Human Mortality Database (HMD) will be made. The main reason is that ISTAT recently published the complete series of period life tables from 1974 to 2004 by sex separately for all the Italian regions. On the other hand, the HMD is certainly the fundamental worldwide source of data on old mortality and the set of life tables at country level provided by the analysis of these data represent the most influential demographic tool to document the longevity revolution of the modern era. Therefore, the methods protocol for the HMD (Wilmoth *et al.*, 2007) represents actually our main reference document for the practical construction of the life tables here presented.

Let's start looking at the old mortality trend of the probabilities of dying separately by sex plotted in **Figures 6** and **7**. We can firstly note as for both sexes these probabilities generally tend to decrease over time. The mortality curves seem to be gradually shifted forward in the age horizontal axis from the past to now. Moreover, it is quite visible the decreasing gap from the pre- to the post-war probabilities, especially starting from 75-80 years of age. This is mainly due to the environmental improvements given by public health interventions and medical researches and discoveries. Different features can be derived by considering the pattern of nineteenth-century probabilities of dying. For both sexes, it can be seen that they assume higher values up to 80-85 years of age. Then, the slopes of the corresponding curves become the less steep over time. This can be also observed by looking at the series of the estimates of  $k(x)$  from age 80 to 95 in **Table 4**: the smallest values (namely, the lowest slopes) correspond to the 1871 and 1881 years. All the curves, if they are extrapolated at extreme ages (such as 120 or 125), converge together to a constant, according to the inner features of the fitted mortality model.

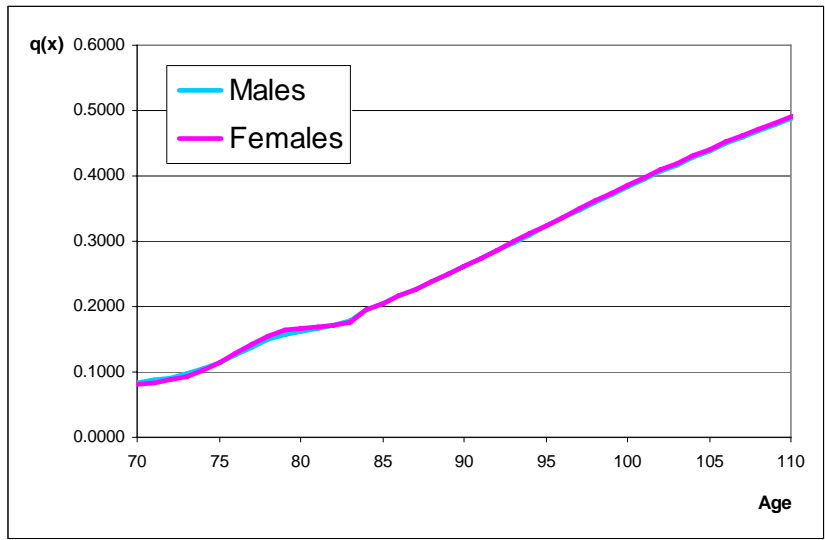


*Figure 6: Probabilities of dying in the period 1871-2001: Males.*

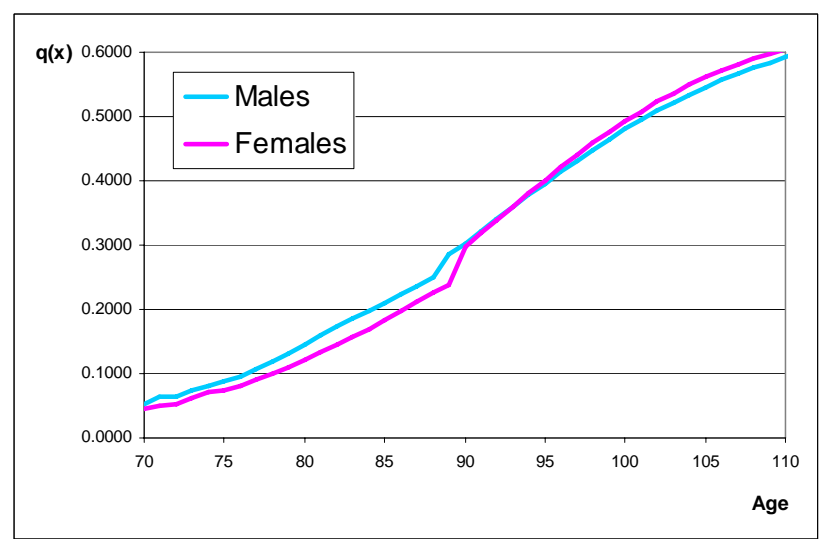


**Figure 7: Probabilities of dying in the period 1871-2001: Females.**

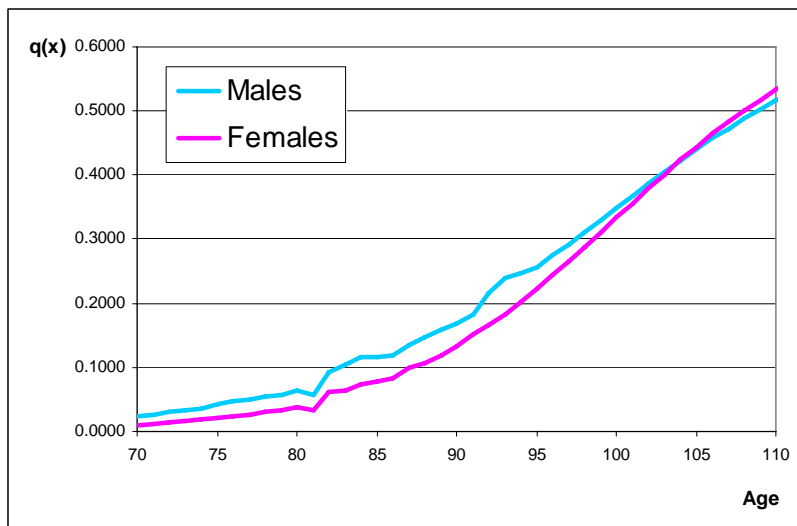
When the results are compared between sexes, for males higher values into probabilities of dying can be generally pointed out. This difference constantly increases over time, as resumed in *Figures 8, 9* and *10*.



**Figure 8: Probabilities of dying: Males VS Females, 1881.**



**Figure 9: Probabilities of dying: Males VS Females, 1936.**



**Figure 10: Probabilities of dying: Males VS Females, 2001.**

The fact that in more recent periods at highest ages the female probabilities of dying (for the last part estimated by the mortality model) exceed the male ones (*Figures 9 and 10*) is closely related to the biological theories and to the empirical findings which the Kannisto model is based on. In fact, by following the frailty theory, the  $i$ -th member of a birth cohort is subject to a particular mortality function at each age  $x$  which follows the Makeham law. Then, as a cohort becomes older, those with high frailty are more likely to die first and there is an effect which may be described as survival of the fittest (Thatcher, 1999). The result is that the average function at a certain age changes by following a decelerating curve (e.g. logistic) rather than a trend with an exponential rate such as assumed by the Makeahm or Gompertz equations (Horiuchi and Wilmoth, 1998). In the study case, despite we consider period data and not cohorts, the frailty approach can be applied by considering the overall decreasing trend of mortality over time, especially for females. As a consequence, the effect of survival of the healthier individuals and thus the mortality deceleration is postponed for women compared with men. Thus, at extreme ages and for

modern data the female probabilities can reach larger values than those for males.

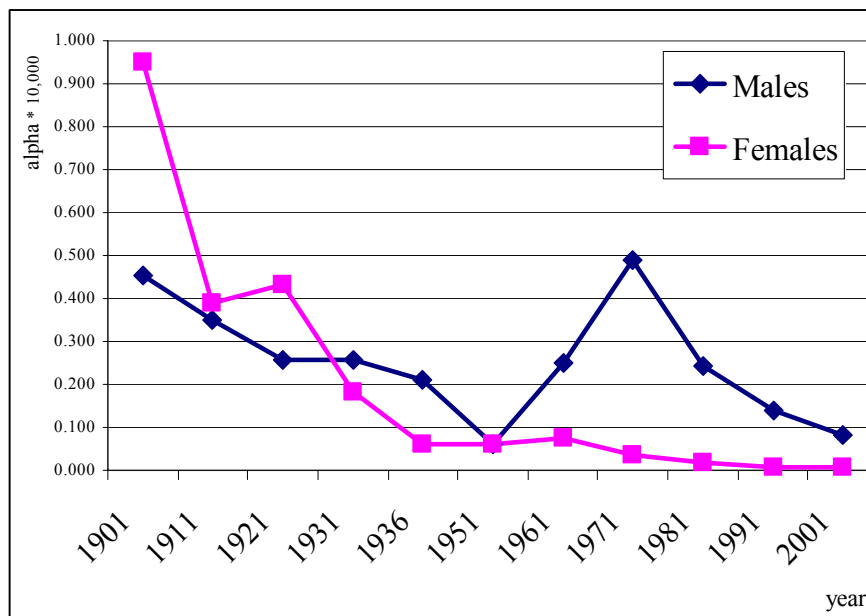
The different patterns of old mortality for males and females can be also derived by the trend of the estimates of the  $k(x)$  in the post-war era. Especially from 1961 on, the female and male values starts to diverge revealing an increasing larger steepness in the female mortality curves. It means again that the women old mortality accelerates faster than that for men and the deceleration is delayed over age.

Furthermore, by fixing the age, for males a near-complete immobility of  $k(x)$  over time can be pointed out. On the other hand, for females we can note an accelerating increase from decade to decade. This aggravation can be actually considered a consequence of their greater reduction in death probabilities for the relatively younger old as we observed before.

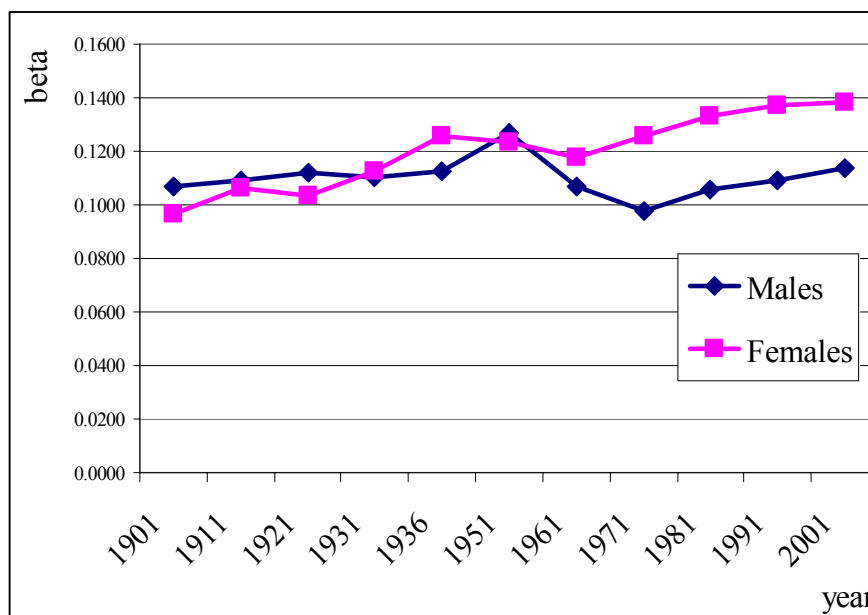
Some other interesting features of old mortality development can be derived by observing the values of the estimated parameters characterizing the fitted mortality function. In *Figures 11* and *12* the estimated values of parameters  $\hat{\alpha}$  and  $\hat{\beta}$ , respectively, over time<sup>8</sup> and by sex, separately, are plotted.

---

<sup>8</sup> We excluded from the graphical representation the values obtained for the nineteenth century for which a separate discussion would be needed.



**Figure 11: Parameter estimates of  $\alpha (x 10^4)$  from the Kannisto model: Males VS Females.**



**Figure 12: Parameter estimates of  $\beta$  from the Kannisto model: Males VS Females.**

For both sexes, the parameter  $\alpha$  has fallen over time from 0.454 to 0.081 for men and from 0.950 to 0.006 for women, bearing out the well-known decrease of the base level of mortality. Nevertheless, the male pattern reveals to be more irregular with an unusual peak in 1971 (0.488) and a particular low value in 1951 (0.061).

On the contrary, the parameter  $\beta$  has slightly risen, changing only from 0.1071 to 0.1136 for males and from 0.0964 to 0.1382 for females. As discussed above, this parameter ultimately governs the rate at which the force of mortality increases with age. As a consequence, its relative stability over such period seems to support the theory of those believing that genetic features can determine an underlying pattern of aging process.

Moreover, from the ratio of  $-\ln(\alpha)$  and  $\beta$  another important parameter, denoted by  $\phi$ , can be introduced and discussed. This quantity is shown to



correspond to the age at which the logistic curve has its point of inflection (Thatcher, 1999). By looking at **Table 5**, we can note as the values of  $\phi$  appear quite stable from 1901 to 1951 (from 93.4 to 94.6 for men and from 96 to 97.4 for women). Then, there is a great improvement in the next period up to reach age 103 in 2001. Thatcher interpreted an analogous finding by asserting that “there was a very long period when the force of mortality above 90 years of age changed very little, despite the changes at lower ages” (Thatcher, 1999, p. 21). In our case, this period likely ended in 1960s, when mortality rates actually started to fall.

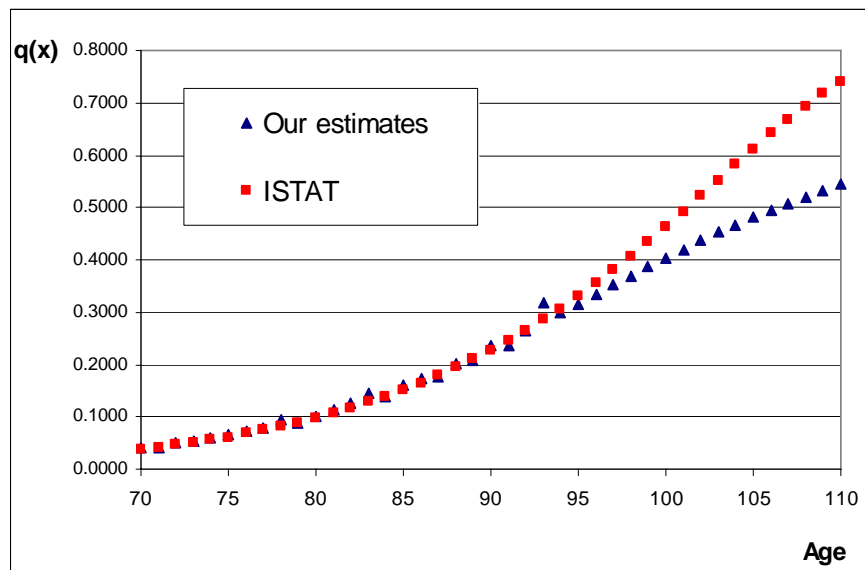
We can finally observe as the  $\phi$  seem to converge between sexes over time, suggesting the existence of an upper limit to human longevity as some biological theories assert. This finding calls certainly for further checking that will be carried out in the next and more detailed analyses through the development of specific demographic indicators.

Year	Males	Females
1901	93.4	96.0
1911	94.1	95.5
1921	94.1	96.9
1931	95.6	96.9
1936	95.7	95.5
1951	94.6	97.4
1961	99.1	100.2
1971	101.7	99.9
1981	100.3	100.1
1991	102.3	101.7
2001	103.2	103.4

**Table 5: Values of  $\phi$**

If we compare our results with those published by ISTAT we can note some significant differences. As an example, we report in **Figure 13** the estimated probabilities of dying for males in 1981 derived by our

procedure and, conversely, by ISTAT. It can be clearly noted as the two patterns start significantly to diverge from age 95 displaying constantly higher values corresponding to the ISTAT procedure. The same features are observed for all the comparable years<sup>9</sup> and for both sexes.

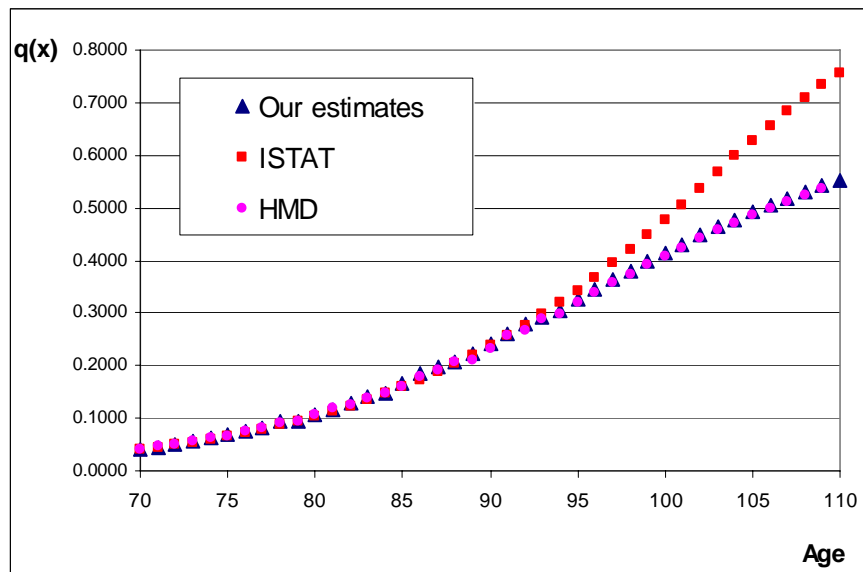


*Figure 13: A comparison with the ISTAT life tables: probability of dying, 1981, Males.*

---

<sup>9</sup> These years are: 1981, 1991 and 2001.

However, our estimated probabilities can be validated by observing wide similarities to the procedure employed for the HMD (*Figure 14*).



*Figure 14*<sup>10</sup>: A comparison with the HMD and ISTAT life tables: probability of dying, 1981, Males.

---

<sup>10</sup> The results here reported refer to the Italian case as the HMD only considers country data. Therefore, our procedure has been applied to Italian data and then compared to the HMD and ISTAT life tables.

## **References**

A.J. Coale and E.E. Kisker (1990), *Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages*, Asian and Pacific Population Forum 4:1-31.

B. Gompertz (1825), *On the nature of the function expressive of the law of human mortality etc.*, Phil. Trans. Roy. Soc. 115:513-585.

L. Heligman and J.H. Pollard (1980), *The age pattern of mortality*, Journal of the Institute of Actuaries 107:49-80.

S. Horiuchi and J.R. Wilmoth (1998), *Deceleration in the Age Pattern of Mortality at Older Ages*, Demography, Vol. 35, No. 4, (Nov., 1998), pp. 391-412.

S. Horiuchi and A.J. Coale (1990), *Age patterns of mortality for older women: an analysis using the age-specific rate of mortality change with age*, Mathematical Population Studies 2:245-267.

ISTAT (2001), *Tavole di mortalità della popolazione italiana per provincia e regione di residenza - Anno 1998*, Settore: Popolazione, Collana Informazioni – 2001.

V. Kannisto (1994), *Development of Oldest-Old Mortality, 1950-1990: Evidence from 28 Developed Countries*, Odense Monographs on Population Aging, 1. Odense University Press, Odense, Denmark.

V., Kannisto, J. Lauritsen, A.R. Thatcher, and J.W. Vaupel (1994), *Reductions in mortality at advanced ages: several decades of evidence from 27 countries*, Population and Development Review 20:793-870.

V. Kannisto (1996), *The Advancing Frontier of Survival: Life Tables for Old Age*, Odense Monographs on Population Aging, 3. Odense University Press, Odense, Denmark.

W.M. Makeham (1860), *On the law of mortality and the construction of annuity tables*, Journal of the Institute of Actuaries 8:301-310.

M. Livi Bacci (1999), *Introduzione alla demografia*, Loescher, Torino.

D.R. McNeil, T.J. Trussell, and J.C. Turner (1977), *Spline interpolation of demographic data*, Demography 14(2): 245-252.

W. Perks (1932), *On some experiments on the graduation of mortality statistics*, Journal of the Institute of Actuaries 63:12-40.

J.H. Pollard (1973), *Mathematical models for the growth of human populations*, Cambridge University Press.

R. Pressat (1969), *L'analyse démographique*, 2e édition, Presses Universitaires de France, Paris.

A.R. Thatcher, V. Kannisto, and J.W. Vaupel (1998), *The Force of Mortality at Ages 80 to 120*. Odense, Denmark: Odense University Press.

A. R. Thatcher (1999), *The Long-Term Pattern of Adult Mortality and the Highest Attained Age*, Journal of the Royal Statistical Society. Series A (Statistics in Society), Vol. 162, No. 1, (1999), pp. 5-43.

W.A. Weibull (1951), *A statistical distribution function of wide applicability*, J Appl. Mech.18:293-297.

J.R. Wilmoth, K. Andreev, D. Jdanov, and D.A. Gleijeses with the assistance of Boe C., Bubenheim M., Philipov D., Shkolnikov V., Vachon P. (2007), *Methods Protocol for the Human Mortality Database*, Revised: May 31, 2007 (Version 5), <http://www.mortality.org>