A TOOL FOR PROJECTING AGE PATTERNS BASED ON A STANDARD AGE SCHEDULE AND ASSUMPTIONS ON RELATIVE RISKS USING LINEAR SPLINES: TOPALS

Joop de Beer, Nicole van der Gaag and Frans Willekens
Netherlands Interdisciplinary Demographic Institute
The Hague

1 November 2007

1. Introduction

Using a multistate model such as MicMac[1] for making scenarios requires that assumptions need to be specified on the future values of many parameters, *viz.* transition rates distinguished by age, sex, forecast year and covariates like level of educational attainment. Particularly the distinction by age implies that many parameter values need to be specified. One solution is to specify model curves describing the age patterns and specify assumptions on the future values of the parameters of the model curve instead of values of rates for all separate ages. This paper introduces TOPALS (Tool for projecting age patterns using linear splines), a method that is capable of describing all kinds of age curves. The parameters can be interpreted easily and the fit of the curve to the data is good. The basic idea is to choose an age schedule that captures the general pattern of the demographic process and to model deviations from this curve by age-specific relative risks[2] which are modeled by a linear spline function. The values of the spline function are estimated by dividing the average values of the transition rates for successive age groups by the average values of the standard age schedule of transition rates for the corresponding age groups. The standard age pattern can be calculated by averaging an age pattern over countries (e.g. the EU15 average) or by estimating some model age schedule (e.g. the Heligman-Pollard model for mortality).

The idea of modeling deviations from a standard age schedule was developed by Brass (1974). Brass assumes a linear relationship between a double logarithmic transformation of the age pattern to be fitted and a double logarithmic transformation of a standard age schedule. The two parameters can be estimated by linear regression. One problem of this approach is that the parameters lack a clear interpretation. Zeng Yi et al. (2000) propose an alternative method for estimating the two parameters of the Brass relational model, based on the median age and interquartile range. TOPALS is more flexible than the method proposed by Zeng Yi et al., produces a better fit and is at least as simple. By using a linear spline function TOPALS is flexible in two respects. First, it can describe all kinds of age curves. Secondly, the user can choose the desired level of goodness of fit and degree of smoothness. This paper illustrates TOPALS by fitting curves of age-specific mortality and fertility rates for Italy and the Netherlands using the EU15 average as standard age schedule.

TOPALS can be used for making projections of future changes in age-specific rates by specifying assumptions about changes in the average values of the relative risks for successive age groups. Three approaches can be followed. First, assumptions can be made on the future values of the relative risks related to the standard age schedule in the base year. Secondly, the standard age schedule can be projected into the future using a random walk model with drift and a linear spline function describing changes in the age pattern over time. Assumptions on the relative risks can be specified relative to the projected standard age schedule. Thirdly, the age pattern in the base year can be used as standard age schedule. The latter approach produces projections that are similar to the Lee-Carter method. The three variants of TOPALS will be illustrated by projecting age-specific mortality rates for Italy and the Netherlands for the year 2050.

[1] "TOPALS is developed within the project MicMac: Bridging the micro-macro gap in population forecasting; see for more information: www.micmac-projections.org."

[2] Throughout the paper we will use the term 'relative risk' as a rather general term to describe ratios of rates and risks.

TOPALS allows to take into account the effects of covariates by modeling different age patterns for different categories of a covariate. For example, one may assume different age patterns of the transition rates for different levels of educational attainment. The paper shows how the linear spline function of relative risks can be used for this purpose.

2. Using TOPALS to estimate age profiles

We assume that a standard age schedule of transition rates is given. The age profile of transition rates for a given country or population category can be estimated on the basis of relative risks, i.e. on the ratio between the transition rates of that country or population category and those according to the standard age schedule. The relative risk at age x for country i (or population category i) is equal to:

$$(1) \qquad r_{i,x} = \frac{q_{i,x}}{q^*_x}$$

where q^*_x is the transition rate at age x according to the standard age schedule. All rates discussed in this paper are distinguished by sex. However, for the sake of readability we omit subscripts indicating sex. The age pattern of the relative risks can be described by a linear spline function. This is a piecewise linear curve. The ages at which the successive linear segments are connected are called 'knots'. The relative risks at each age can be estimated by the linear spline function:

$$(2) \qquad \hat{r}_{i,x} = a_i + b_{i,0}(x-m) + \sum_{j=1}^{n} b_{i,j}(x-m-k_j)D_j$$

where $D_j = 0$ if $x - m \le k_j$ and $D_j = 1$ otherwise, m is the minimum age, $x \ge m$, k_j are the knots, n is the number of knots, a_i and $b_{i,j}$ are parameters to be estimated.

The knots can be fixed a priori or they can be chosen in such a way that the fit of the spline to the data is optimal. In the latter case a non-linear estimation method is required, *e.g.* a non-linear least squares method. If the location of the knots is fixed a priori, a_i and $b_{i,j}$ can be estimated by linear regression. However, these parameter values are difficult to interpret. Therefore we suggest to estimate the linear spline in a different way. The values of $\hat{r}_{i,k_j}$ at the knots k_j are set equal to the average values of $r_{i,x}$ for successive intervals. $\hat{r}_{i,k_1}$ is set equal to the average value of the $r_{i,x}$s in the interval $1 \ldots (k_1+k_2)/2$, $\hat{r}_{i,k_2}$ is set equal to the average value of the $r_{i,x}$s in the interval $1+(k_1+k_2)/2 \ldots (k_2+k_3)/2$, etc. The values of a_i and the $b_{i,j}$s can be estimated from the values of $\hat{r}_{i,k_1}$, $\hat{r}_{i,k_2}$, $\hat{r}_{i,k_3}$, etc. by solving the following equations:

$$(3) \qquad \begin{aligned} \hat{r}_{i,m} &= a_i \\ \hat{r}_{i,k_1} &= a_i + b_{i,0}k_1 \\ \hat{r}_{i,k_2} &= a_i + b_{i,0}k_2 + b_{i,1}(k_2 - k_1) \\ \hat{r}_{i,k_3} &= a_i + b_{i,0}k_3 + b_{i,1}(k_3 - k_1) + b_{i,2}(k_3 - k_2) \\ &\textit{etc.} \end{aligned}$$

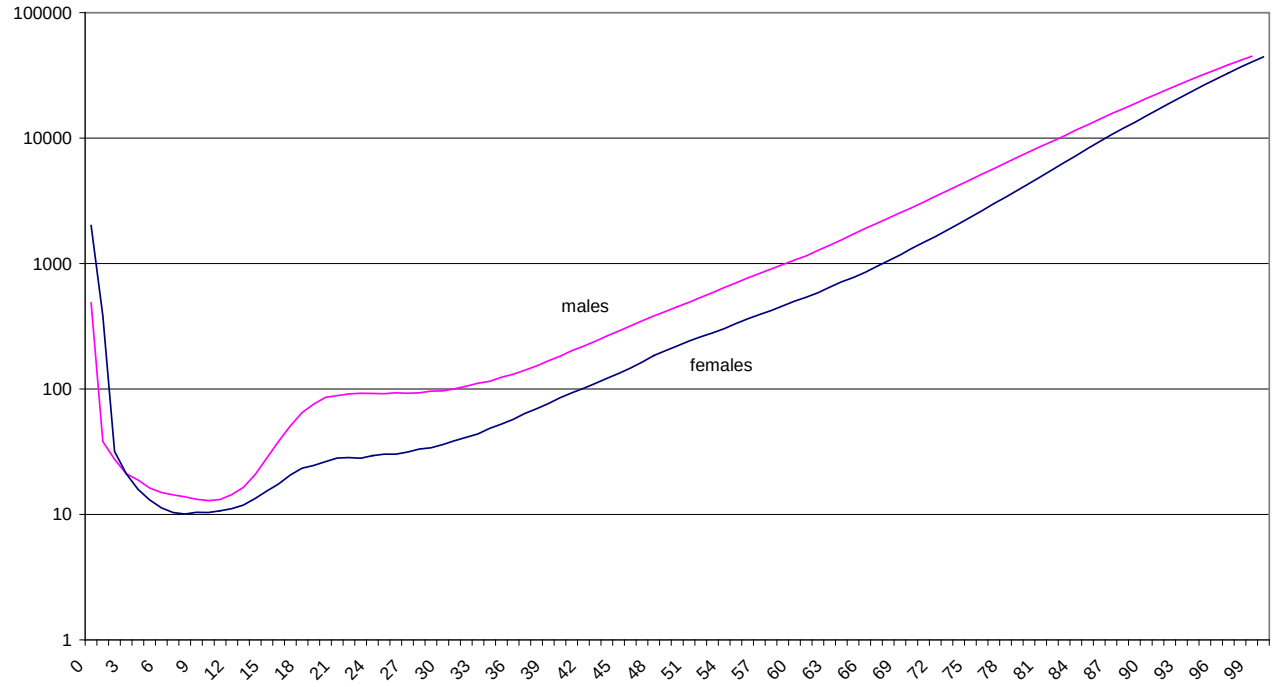
The transition rates for country i are estimated by multiplying the relative risks which are estimated by the linear spline function ($\hat{r}_{i,x}$) by the transition rates according to the model age schedule (q^*_x):

$$(4) \qquad \hat{q}_{i,x} = \hat{r}_{i,x} q^*_x .$$

3. Modeling age patterns of mortality rates

The standard age schedule that we use to model age patterns for mortality rates is the (unweighted[3]) average of the EU15 countries in 2003, for men and women separately. The age-specific mortality rates used in this paper are calculated from data on age-specific deaths and population numbers published by Eurostat. Figure 1 compares the age-specific mortality rates for men and women in 2003 (on a logarithmic scale).

Figure 1 Age-specific mortality rates, EU15 average, 2003



[3] The age-specific mortality rates based on the weighted EU15 average with population size of all countries as weights, hardly differ from the unweighted average. On average the weighted average age-specific mortality rates are 1 percent lower than the unweighted averages. The logarithms of the age-specific mortality rates differ even less: by only 0.2 percent.

Figure 2 compares the age-specific mortality rates for Italian women and men in 2000 (solid lines) with the EU15 average in 2003 (dotted lines). The overall age pattern for Italy is similar to the EU15 average, but in general the Italian mortality rates are lower than the EU15 average. Figure 3 compares the Dutch mortality rates to the EU15 average. Again, the age patterns seem rather similar, but mortality rates for Dutch men in their 20s and 30s are relatively low.

Figure 2 Age-specific mortality rates, Italy 2000 and EU15 average 2003

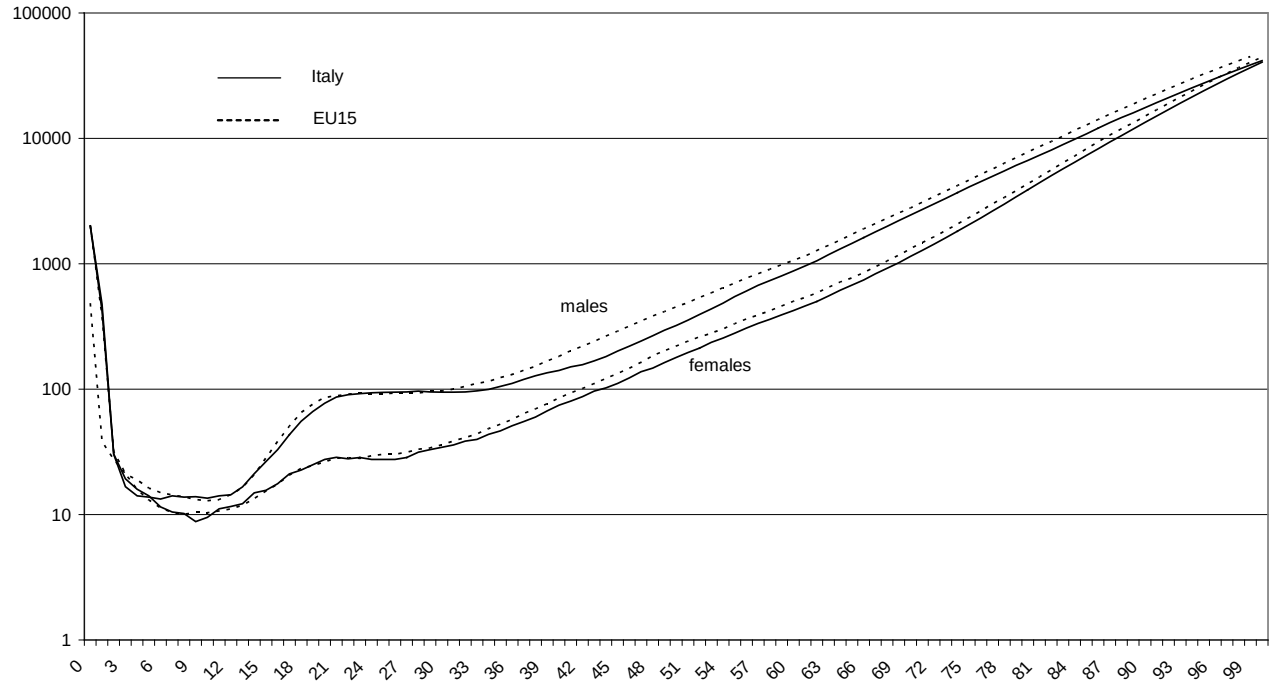
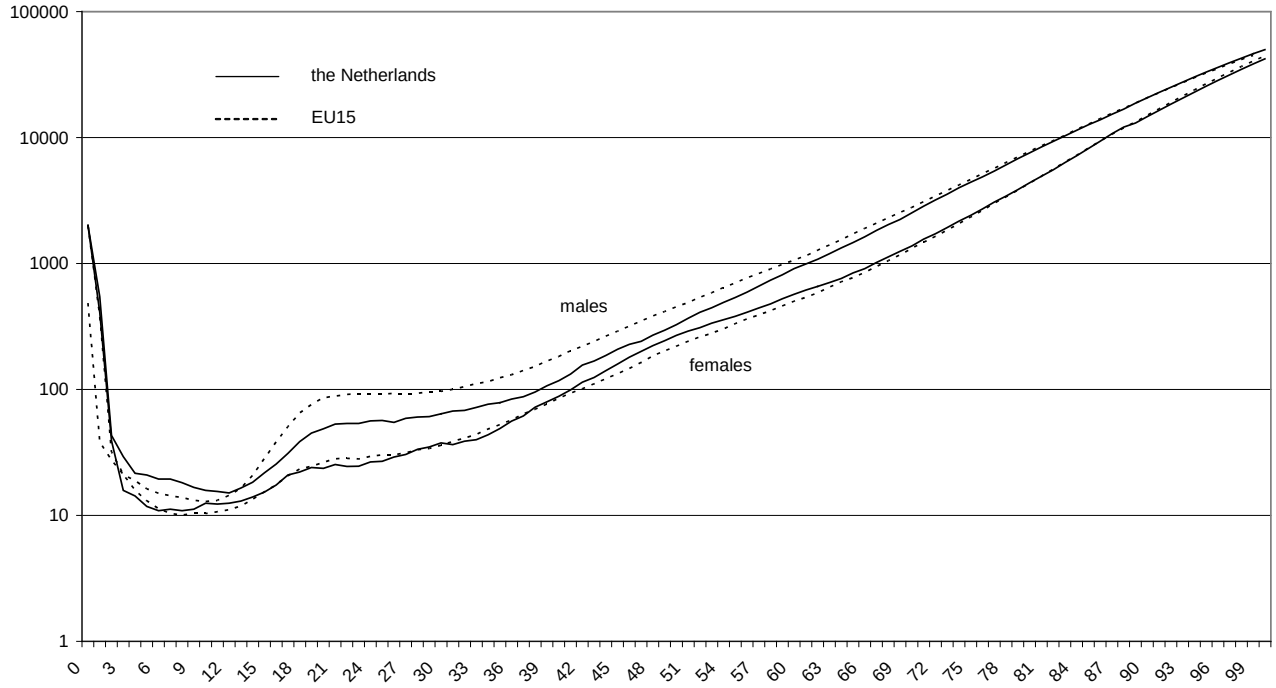


Figure 3 Age-specific mortality rates, the Netherlands and EU15 average, 2003



The solid lines in figures 4a and 4b show the ratios of the age-specific mortality rates of Italian men and women compared with the EU15 average. The figures show that particularly the mortality rates for both Italian men and women in their 40s are clearly lower than the EU15 average and the differences at young and old ages are smaller.

The dotted line is a linear spline function. As discussed in section 2 the linear spline function is estimated on the basis of average values for successive age intervals. For each 10-year age group the average mortality rates for Italian men and women and for the EU15 averages for men and women are calculated. As the mortality rates for age 0 are considerably higher than for other ages, this age group is taken as separate age category. See table 1. For each age group the ratio of the Italian average mortality rate and the EU15 average is calculated. See table 2. The linear spline is estimated by setting the value for age 5 equal to the average of the quotients for the age group 1-10, the value for age 15 to the average of the quotients for age group 11-20, etc. For age 90 the value is set equal to the average of the quotients for the highest age group. These are the knots of the linear spline function. For the other ages the values of the linear spline can be calculated on the basis of the formulas given in section 2. This corresponds with linear interpolation between ages 5 and 15, 15 and 25, etc. The figure shows that these linear splines describe the age pattern of the relative risks rather well. One benefit of this approach is that the parameters of the spline (i.e. the values in table 2) can be interpreted easily: they simply indicate to what extent age-specific rates in different age groups are lower or higher than the EU15 average.

Figure 4a Relative mortality rates, (EU15 2003 =1), Italy, men, 2000

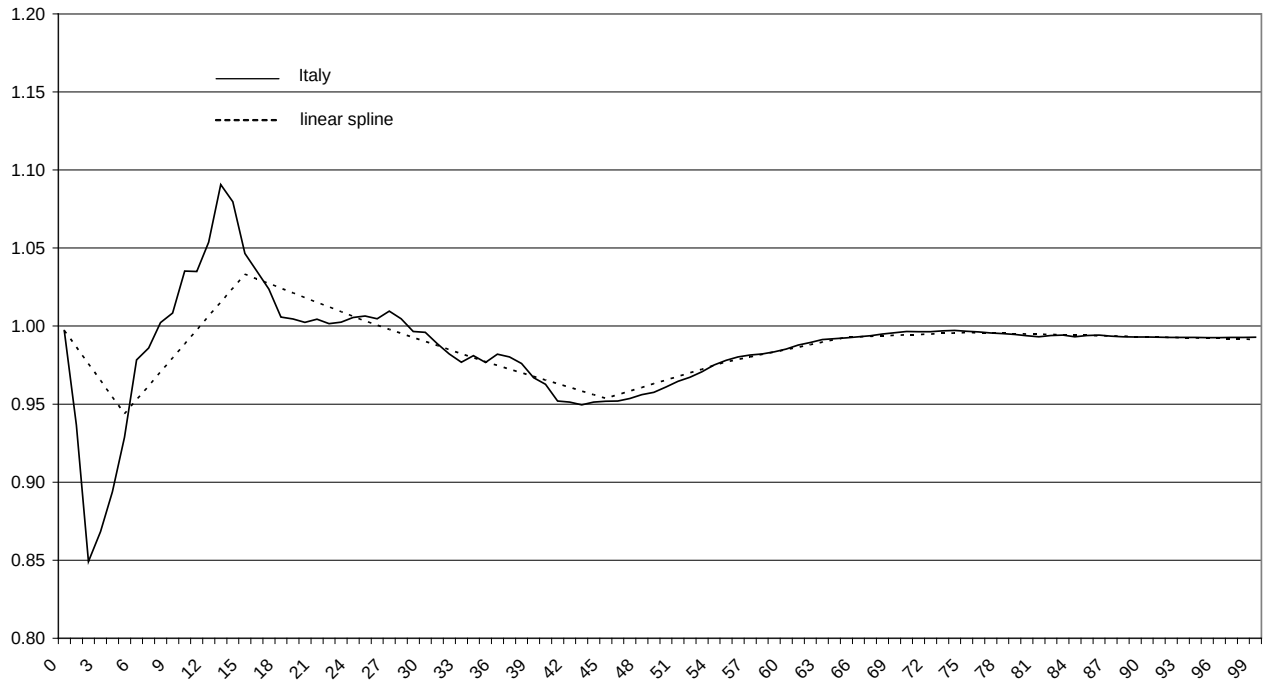


Figure 4b Relative mortality rates, (EU15 2003 =1), Italy, women, 2000

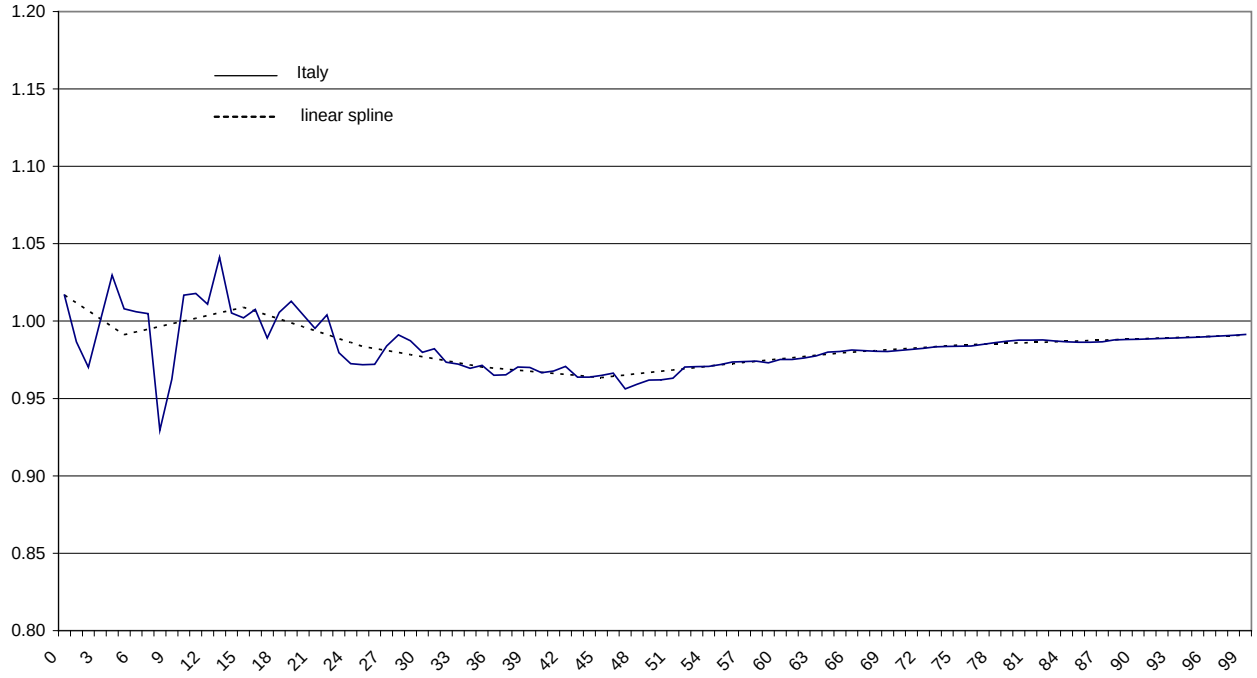


Table 1. Age-specific mortality rates (per 100,000), 2003

	EU15		Italy*		Netherlands	
age	men	women	men	women	men	women
0	486	382	477	423	546	410
1-10	19	15	16	14	22	15
11-20	41	19	44	20	31	19
21-30	93	32	94	30	59	30
31-40	133	63	118	56	92	63
41-50	313	163	241	134	245	195
51-60	750	385	651	326	659	443
61-70	1869	953	1779	829	1819	1022
71-80	5111	2996	4922	2659	5424	3057
81+	22945	19968	21361	17936	25370	19268
* Figures for 2000						

Table 2. Relative mortality rates, 2003 (EU15 average = 1)				
	Italy*		Netherlands	
age	men	women	men	women
0	0.98	1.11	1.12	1.07
1-10	0.83	0.97	1.15	1.03
11-20	1.08	1.02	0.77	0.98
21-30	1.01	0.94	0.63	0.95
31-40	0.89	0.88	0.69	1.00
41-50	0.77	0.83	0.78	1.20
51-60	0.87	0.85	0.88	1.15
61-70	0.95	0.87	0.97	1.07
71-80	0.96	0.89	1.06	1.02
81+	0.93	0.90	1.11	0.96
* Figures for 2000				

Figures 5a and 5b show the rate ratios and linear splines for Dutch men and women. The relative risks for Dutch men in their 20s and 30s are low, whereas mortality rates at older ages are higher than the EU15 average. For Dutch women the mortality rates for most ages are higher than the EU15 average.

Figure 5a Relative mortality rates (EU 15=1), the Netherlands, men, 2003

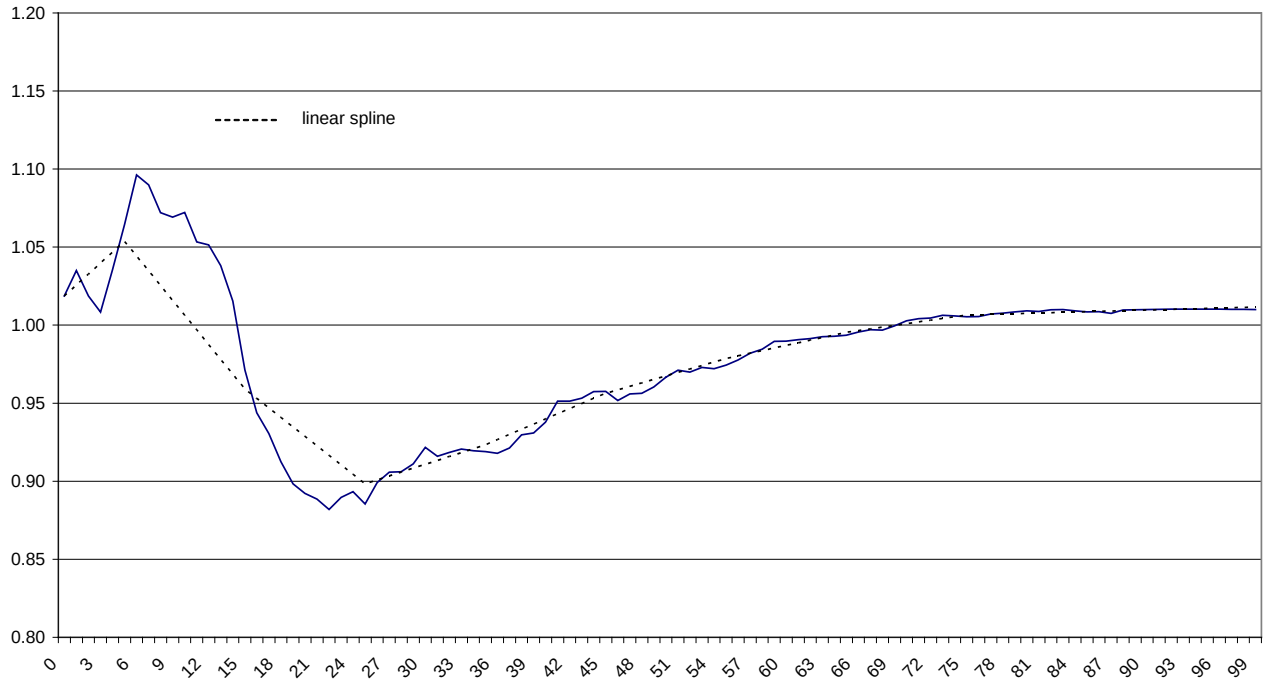
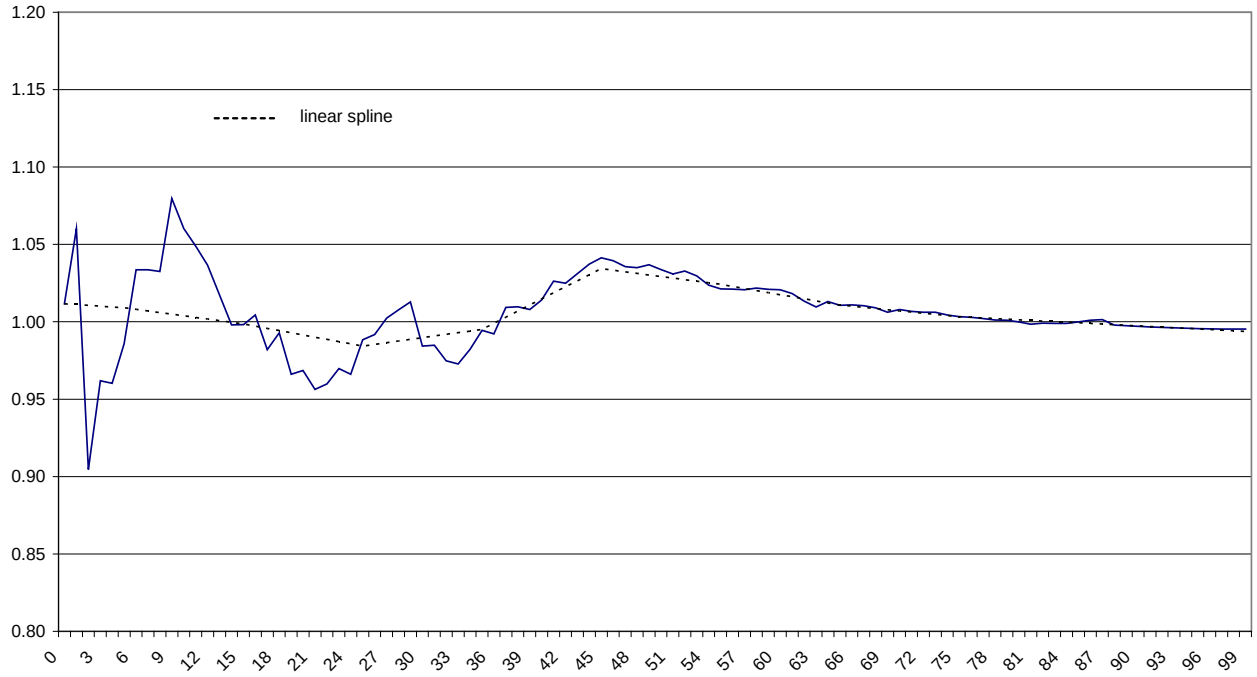


Figure 5b Relative mortality rates (EU 15=1), the Netherlands, women, 2003



Multiplying the values of the rate ratios estimated by the linear splines by the EU15 age-specific mortality rates provides the fitted curves of age-specific mortality rates. Figure 6 shows that the fit for both Italian men and women is good. Figure 7 shows the fit for Dutch men and women which is satisfactory as well. Thus based on the same standard age schedule, *viz.* the EU15 average, age-specific mortality rates for both Italy and the Netherlands can be modeled rather accurately.

Figure 6 Age-specific mortality rates, the Netherlands, 2003

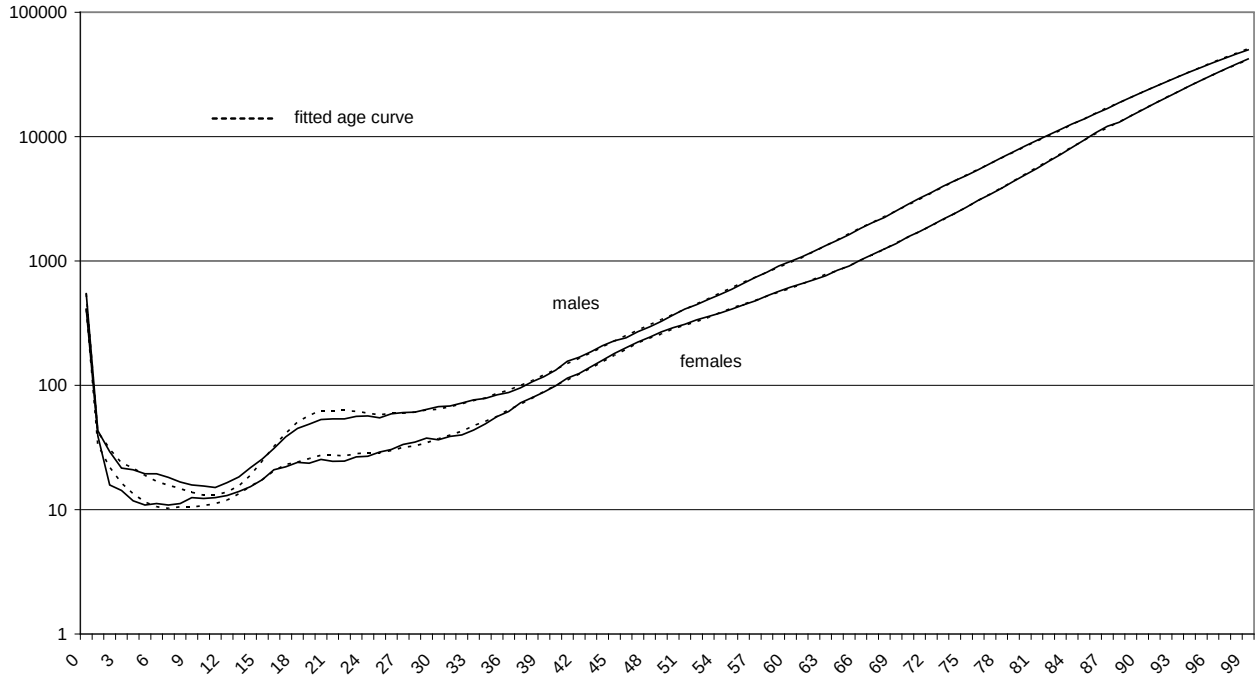
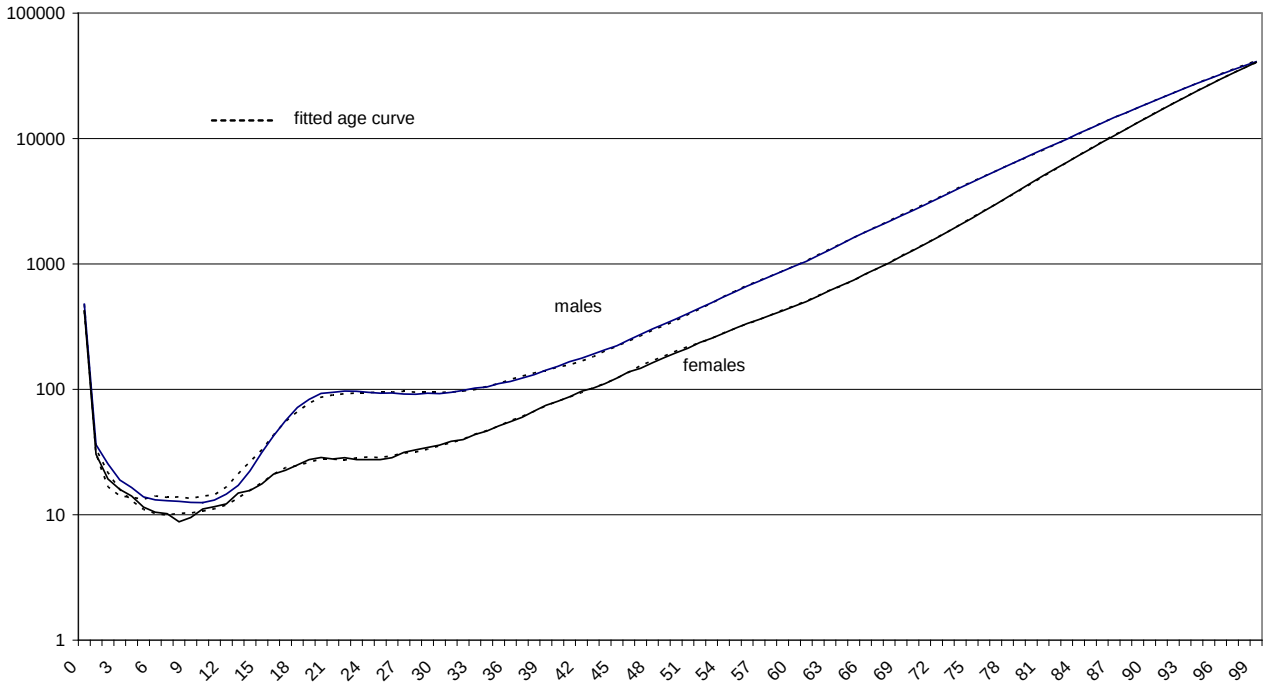


Figure 7 Age-specific mortality rates, Italy, 2000



4. Modeling age patterns of fertility rates

In order to illustrate the flexibility of the method we fit age patterns of fertility rates which have quite a different shape than mortality age curves. Figure 8 shows age-specific fertility rates for Italy and the Netherlands as well as the (unweighted) EU15 average. The overall pattern of the curves is similar, but there are clear differences as well. At most ages fertility rates in the Netherlands are higher than in Italy. Moreover, the age curve of the Netherlands is more peaked than the EU15 average. Furthermore the peak of the curve in both the Netherlands and Italy is at a slightly higher age than the EU15 average.

Figure 8 Age-specific fertility rates, Italy, Netherlands and EU15 averages, 2005

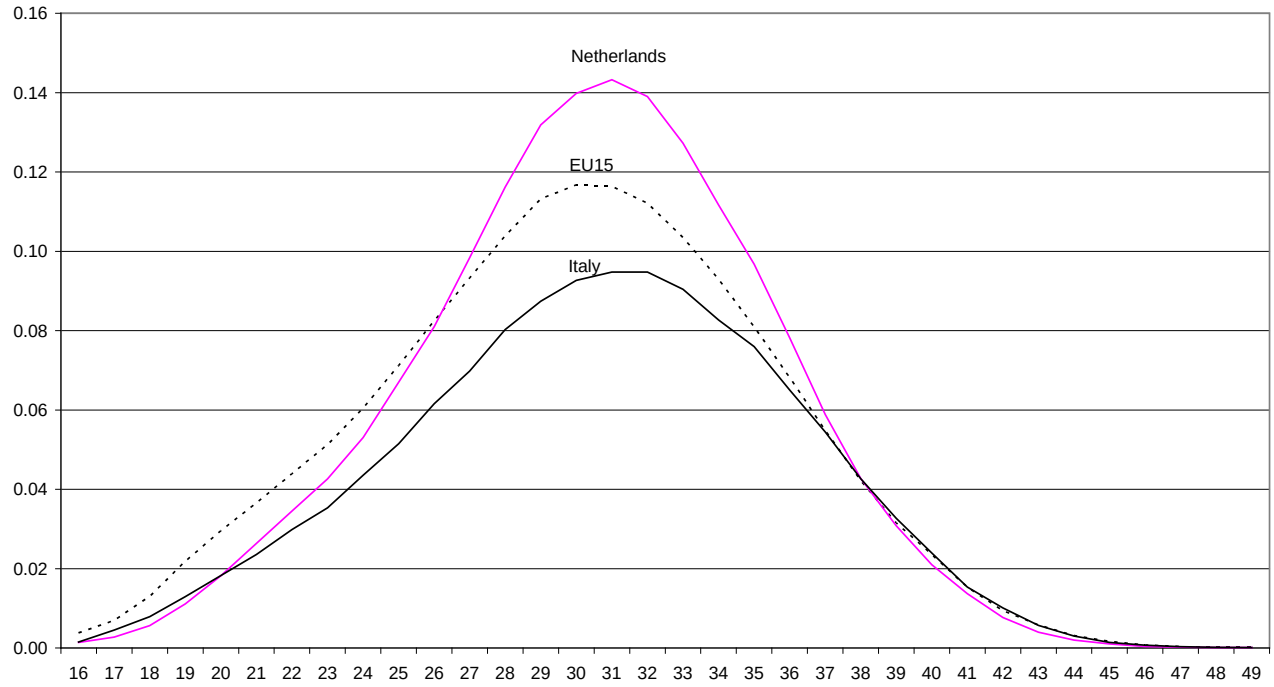
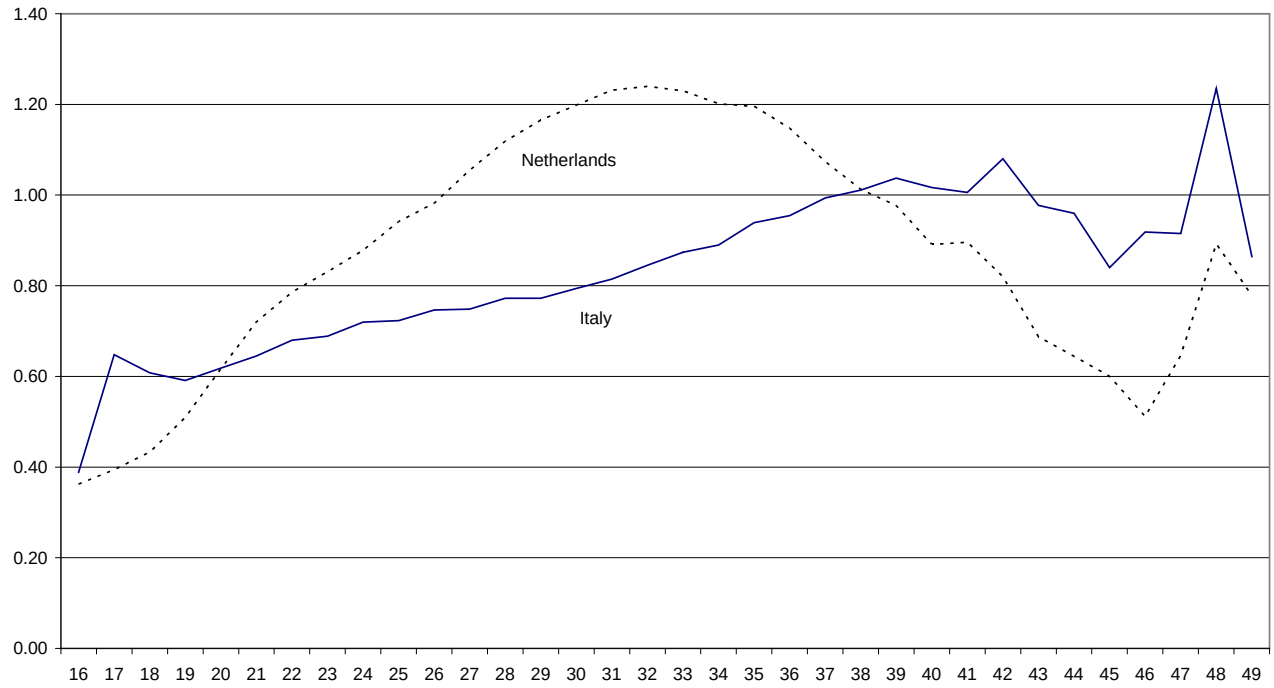


Figure 9 shows the ratios of the age-specific fertility rates of Italy and the Netherlands compared with the EU15 average. The figure shows that between ages 27 and 38 the fertility rates in the Netherlands are higher than the EU15 average, whereas for younger and older ages the opposite applies. For Italy for ages younger than 38 years fertility rates are lower than the EU15 average, whereas for the oldest ages fertility rates for Italy are on average equal to the EU15 average.

Figure 9 Relative fertility rates (EU15=1), Italy and Netherlands averages, 2005



For fitting a linear spline function to the relative risks shown in figure 9 relative fertility ratios for six age groups are calculated. See table 3. On the basis of these ratios fertility age curves for Italy and the Netherlands are fitted using the method described in section 2. Figures 10a and 10b show that the fitted curves describe the age-specific fertility rates in both the Netherlands and Italy rather accurately.

Table 3. Age-specific fertility rates, 2005

	per 1,000 women			Relative fertility rates (EU15 = 1)	
age of mother	EU15	Italy	Netherlands	Italy	Netherlands
16-20	15	9	8	0.60	0.52
21-25	53	37	45	0.70	0.85
26-30	102	78	113	0.77	1.11
31-35	101	88	124	0.87	1.22
36-40	44	44	46	0.99	1.05
41+	4	4	3	1.01	0.80

Figure 10a Age-specific fertility rates, the Netherlands, 2005

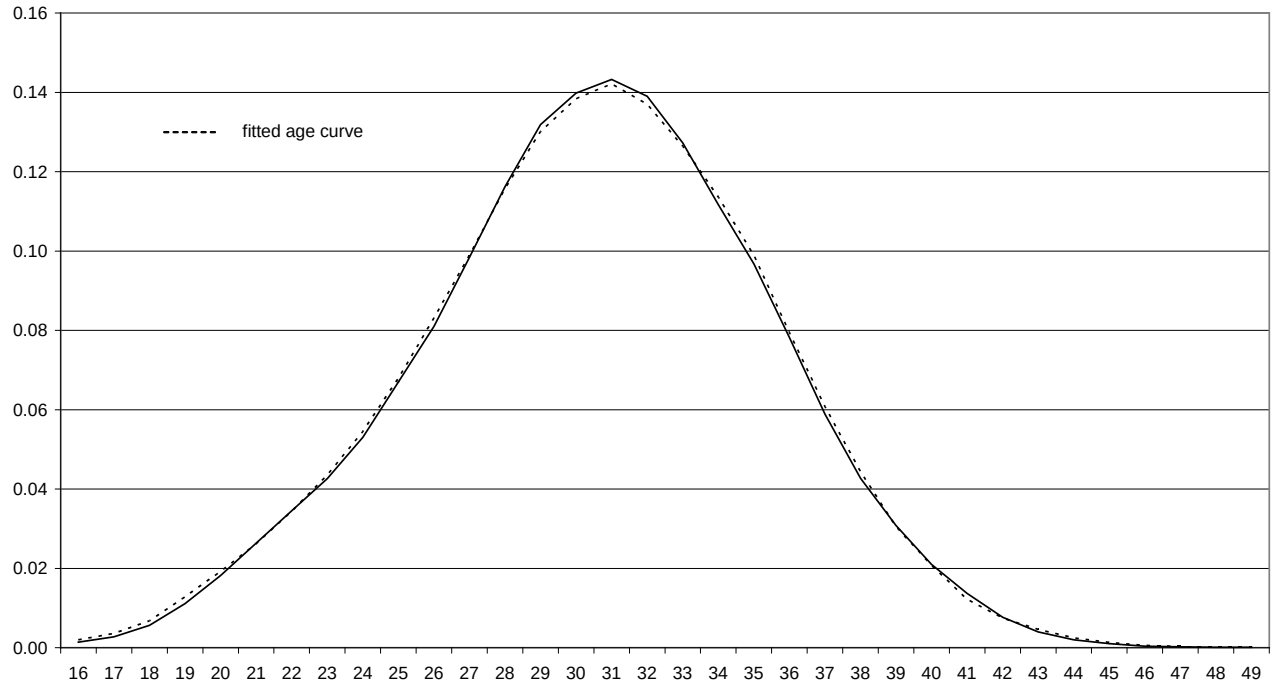
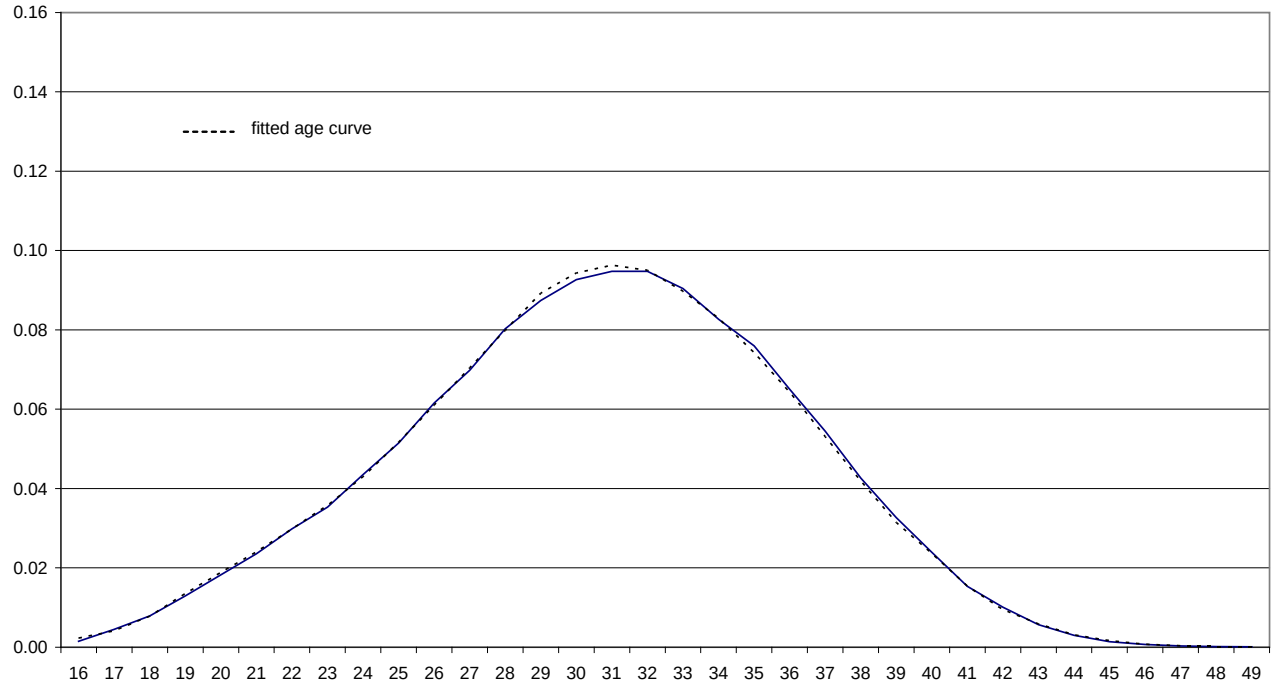


Figure 10b Age-specific fertility rates, Italy, 2005



5. Using TOPALS to project future age-specific mortality rates

Using the estimated relative risks for making projections on future values of age specific mortality rates three alternative procedures may be followed. First, one may make assumptions on future changes in the values of the relative risks for 10-year age groups and multiply these values by the age specific mortality rates according to the standard age schedule in the base year. Secondly, one may make assumptions about the future changes in the age-specific rates according to the standard age schedule and assumptions about the future values of the relative risks compared with the projected age schedule. In the second variant one may assume that the relative risks will remain constant. Alternatively one may assume that the relative risks will change. For example, one may assume that age-specific mortality rates of various countries will converge by assuming that the relative risks of each country will change in the direction of a value of 1. We will also discuss an alternative to the second variant in which we use the age profile of each individual country in the last observation year as standard age schedule rather than the EU15 average. In this variant the relative risks describe future changes in the age pattern of a country compared to the age pattern in the base year of that same country. The latter variant produces projections that are similar to the Lee-Carter method.

5.1 Variant 1: changes in relative risks

In our first variant of using TOPALS the projections of age-specific rates are based on the same standard age schedule that is used for fitting the age profile. Projections are based on assumptions on changes in the values of the relative risks. The values of the relative risks in year $t+T$ can be projected by:

(5) $$\hat{r}_{i,x,t+T|t} = \hat{p}_{i,x,t+T|t}\hat{r}_{i,x,t},$$

where $\hat{r}_{i,x,t+T|t}$ is the projection of the value of $r_{i,x}$ in the year $t+T$ based on observations up to and including year t; $\hat{r}_{i,x,t}$ is the estimate of the relative risk at age x in year t based on the linear spline function (2); and $\hat{p}_{i,x,t+T|t}$ determines by how much $\hat{r}_{i,x}$ is assumed to change between t and $t+T$. The values of $\hat{p}_{i,x,t+T|t}$ are estimated by a linear spline function similar to (2). Thus assumptions on the values of $p_{i,x,t+T|t}$ need to be specified for the averages of age groups rather than for all

separate ages and the linear spline is used to estimate the values for each separate age. The future values of the age-specific mortality rates can be projected by:

(6) $$\hat{q}_{i,x,t+T|t} = \hat{r}_{i,x,t+T|t} q^{*}_{x,t} .$$

By way of an example we formulate assumptions on changes in $r_{i,x}$ between 2003 and 2050 by specifying values of $p_{i,x}$ which correspond closely with changes in the age patterns of mortality rates assumed in the EUROPOP baseline scenario for 2050. For Italy EUROPOP assumes a stronger decline of mortality rates for women than for men, whereas for the Netherlands a smaller decline for women than for men is assumed. Particularly for elderly Dutch women only little reduction of mortality is assumed. EUROPOP assumes a strong decline of mortality rates for Dutch men in their 50s and 60s and a much smaller reduction for men aged 70 or over. For Italy a strong decline in mortality is projected for women between ages 40 and 80 and for men between 40 and 70. For Italian men in their 20s and 30s only a small reduction is assumed, whereas for both Italian men and women a relatively small reduction is assumed for the eldest ages. Overall, EUROPOP assumes that the relatively low mortality rates in Italy will become even lower, while the relatively high mortality rates for Dutch women are assumed to decline very little. In line with this scenario we specify assumptions on the values of $p_{i,x}$, i.e. the factor by which the relative risks in the base year are multiplied in order to project the relative risks for the year 2050 (see table 4). On the basis of these assumptions, values of the relative mortality rates (compared with the EU15 average for 2003) for 2050 are calculated (see table 4). These values are used to calculate linear splines which are multiplied by the age-specific EU15 mortality rates for 2003 in order to calculate age-specific mortality rates for 2050.

Table 4. Assumptions on changes in relative mortality rates between 2003 and 2050

	Italy*		Netherlands	
age	men	women	men	women
0	0.60	0.60	0.50	0.50
1-10	0.30	0.30	0.30	0.20
11-20	0.30	0.40	0.50	0.40
21-30	0.90	0.70	0.75	0.60
31-40	0.90	0.50	0.75	0.75
41-50	0.40	0.40	0.60	0.75
51-60	0.30	0.40	0.50	0.75
61-70	0.30	0.30	0.50	0.70
71-80	0.50	0.30	0.70	0.70
81+	0.60	0.60	0.75	0.80

* changes between 2000 and 2050

Figures 11 and 12 show the age-specific mortality rates for Italy and the Netherlands in 2050 which are projected on the basis of the assumptions specified in table 4 (solid line) as well as the age-specific mortality rates according to the EUROPOP baseline scenario (dotted line). The overall patterns are similar. Due to the fact that our approach is based on a smooth standard age schedule, the projected age patterns are smoother than those according to the EUROPOP scenario, which is based on separately projecting mortality rates for individual ages.

Figure 11 Age-specific mortality rates, Italy, 2050

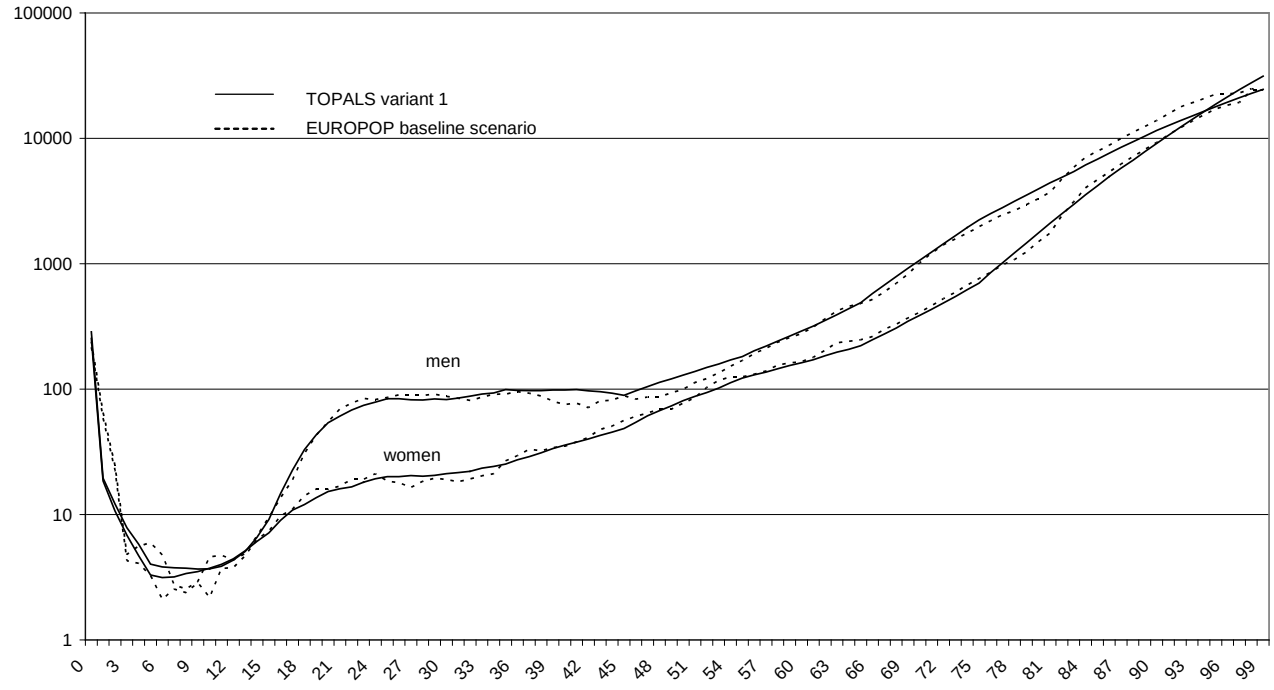


Figure 12 Age-specific mortality rates, the Netherlands, 2050

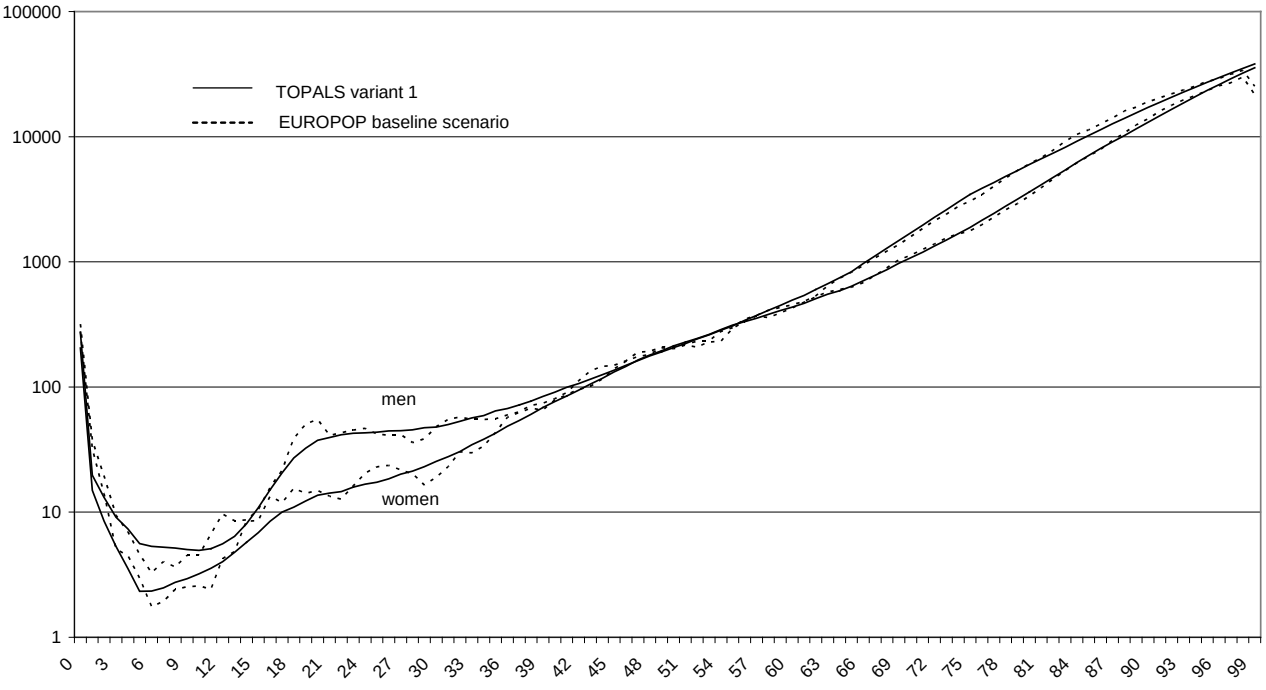


Table 5 shows the life expectancy at birth calculated from the projected age specific mortality rates. According to the EUROPOP baseline scenario life expectancy at birth in Italy will increase more strongly than in the Netherlands. As a consequence the differences between both countries will increase. Thus no convergence is assumed. This can be explained by the fact that past changes are projected into the future. During the last decades mortality rates have declined more

strongly in Italy than in the Netherlands. As a consequence a projection of these trends into the future leads to a stronger projected increase in life expectancy in Italy than in the Netherlands.

Table 5. Life expectancy at birth						
	EU15		Italy		Netherlands	
	men	women	men	women	men	women
1985	71.1	77.7	71.8	78.2	72.6	79.2
2003*	75.2	80.9	76.1	82.0	75.7	80.4
2050						
EUROPOP			83.6	88.8	80.2	83.6
TOPALS variant 1			83.6	88.8	80.2	83.6
TOPALS variant 2	83.2	86.3	83.8	87.2	83.2	86.2
TOPALS variant 3			85.0	89.5	81.2	82.7
* Italy: figures for 2000						

5.2 Variant 2: changes in standard age schedule

Our second variant of using TOPALS for making scenarios consists of two steps. First we make assumptions on future changes in the standard age schedule. Secondly, we make assumptions about the future values of the relative risks compared with the projected standard age schedule. In line with the widely applied Lee-Carter method for projecting mortality (Lee and Carter, 1992) we may assume that the standard age schedule can be projected by using a random walk model with drift. This model projects a linear change. By applying the model to the logarithms of mortality rates the model projects a constant relative change.

The random walk model with drift is

(7) $$\ln q^*_{x,t} = \ln q^*_{x,t-1} + d_x + e_{x,t}\,.$$

where $q^*_{x,t}$ is the mortality rate according to the model age schedule at age x in year t, d_x is the 'drift' and $e_{x,t}$ is the error term with $E(e_{x,t}) = 0$ and $E(e_{x,t}\, e_{x,t+i}) = 0$ for $i \neq 0$. The value of d_x can be estimated by the average change in $\ln q^*_{x,t}$ over the last n years.

According to this model the standard age schedule for year $t+T$ can be projected by

(8) $$\ln q^*_{x,t+T} = \ln q^*_{x,t} + Td_x\,.$$

since $E(e_{x,t}) = 0$.

Separately projecting $q^*_{x,t+T}$ for each x will produce a rather irregular age pattern. Therefore linear splines of the age pattern of changes d_x are estimated, similarly to estimating the splines of the relative risks according to eq. (2):

(9) $$\hat{d}_x = a + b_0(x-m) + \sum_{i=1}^{n} b_i(x-m-k_i)D_i\,,$$

where a and b_i can be estimated similarly as in eq. (3) replacing r by d.

The standard age schedule for year $t+T$ can be projected by:

(10) $$\hat{q}^*_{x,t+T|t} = \exp(\ln q^*_{x,t} + T\hat{d}_x)\,,$$

where $\hat{q}^*_{x,t+T|t}$ is the projection of q^*_x for the year $t+T$ based on observations up to and including year t.

For country i age-specific mortality rates for the year $t+T$ can be projected on the basis of assumptions on the values of the relative risks $r_{i,x,t+T}$ compared with the projected standard age schedule:

(11) $$\hat{q}_{i,x,t+T|t} = \hat{r}_{i,x,t+T|t}\hat{q}^*_{x,t+T|t}\,,$$

where $\hat{r}_{i,x,t+T|t}$ is the estimate of the value of $r_{i,x,t+T}$ based on a linear spline function similar to (2). Thus assumptions on the values of $r_{i,x,t+T}$ need to be specified for age groups rather than for all separate ages (similarly to the assumptions specified in table 4). The linear spline is used to estimate the values for each separate age.

The relative risks in the year $t+T$ may be assumed to be equal to those in the year t:

(12) $$\hat{r}_{i,x,t+T|t} = \hat{r}_{i,x,t}\,.$$

Alternatively one may assume that mortality rates of separate countries are converging (Li and Lee, 2005). This implies that the values of $r_{i,x,t+T}$ are closer to 1 than the values of $r_{i,x,t}$. This can be modeled by a partial adjustment model:

(13) $$\hat{r}_{i,x,t+j|t} = \hat{r}_{i,x,t+j-1|t} + \alpha_i (1 - \hat{r}_{i,x,t+j-1|t})$$

where $0 \le \alpha_i \le 1$ and $j = 1, 2, \ldots T$. Note that eq. (12) is a special case of eq. (13) assuming $\alpha_i = 0$.

This method is illustrated by projecting the average age-specific mortality rates for the EU15 countries. Eq. (10) is used for projecting the age-specific mortality rates of the EU15-countries for the year 2050. The values of d_x are estimated on the basis of changes in the age-specific mortality rates between 1985 and 2003. Figures 13a and 13b show the projected age-specific mortality rates for the EU15 average for men and women respectively and compare them with the 1985 and 2003 age patterns. Table 5 shows that according to these projections life expectancy at birth for the EU15 average would increase to 83.2 years for men and 86.3 years for women.

Figure 13a Age-specific mortality rates, men, EU15 average, 1985, 2003 and 2050

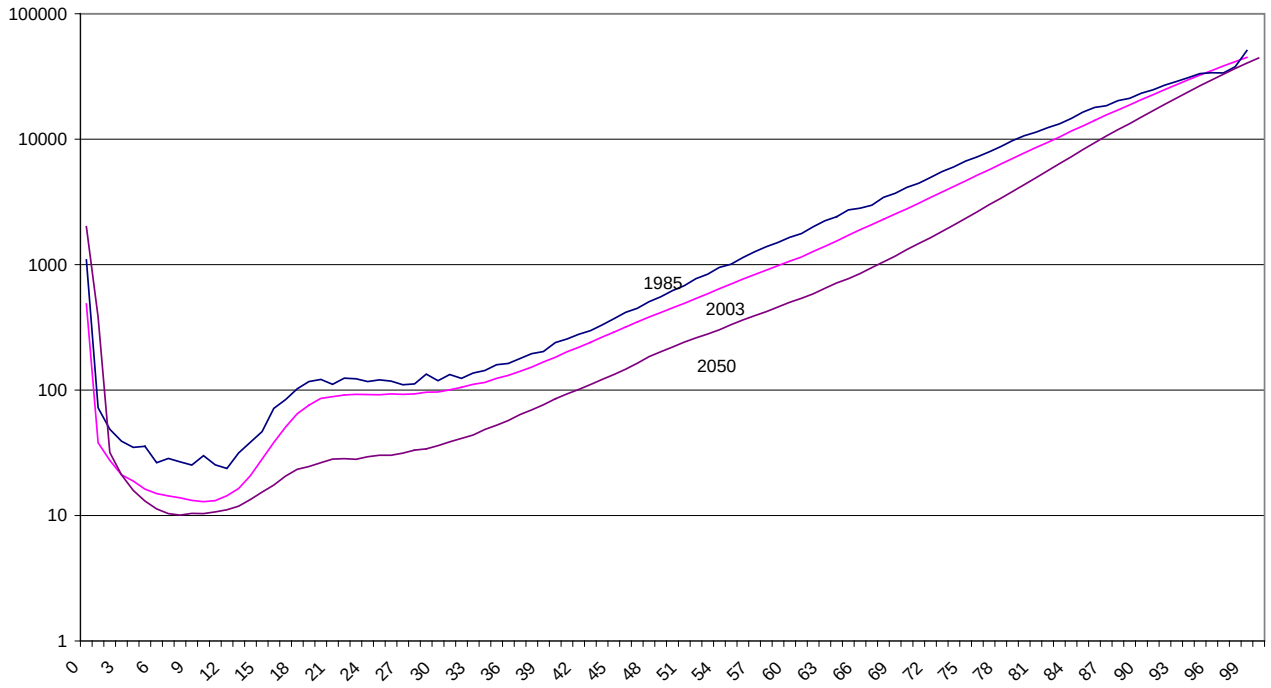
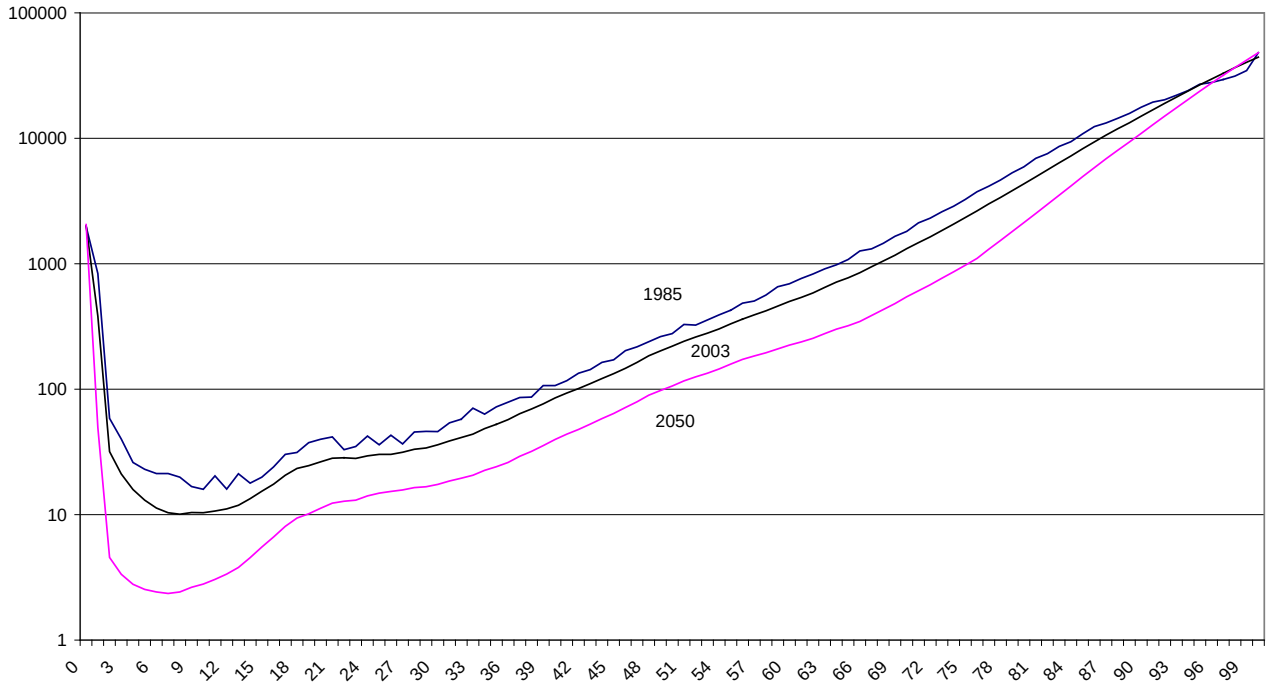


Figure 13b Age-specific mortality rates, women, EU15 average, 1985, 2003 and 2050



Based on these projections age-specific mortality rates for Italy and the Netherlands are projected assuming the relative risks in 2050 to be equal to those in 2003 (*i.e.* using eq. 12). Figures 14a and 14b show the projections for Dutch men and women respectively. The solid lines describe the projections calculated from eq. (11).

5.3 Variant 3: using age profile in base year as standard age schedule

We compare the projections of variant 2 with an alternative variant in which we do not use the EU15 average as standard age schedule but rather the age pattern of each individual country in the last observation year. We call this *Variant 3*. The projections of Variant 3 are based on equations (14) and (15) instead of equations (7) and (10):

(14) $$\ln q_{i,x,t} = \ln q_{i,x,t-1} + d_{i,x} + e_{i,x,t}$$

(15) $$\hat{q}_{i,x,t+T|t} = \exp(\ln q_{i,x,t} + T\hat{d}_{i,x}),$$

where $\hat{d}_{i,x}$ is the linear spline estimate of the drift for age *x*.

The projections according to Variant 3 are similar to the projections by the Lee-Carter method. The main difference is that Variant 3 projects a smooth age pattern, whereas the Lee-Carter method tends to produce an irregular age schedule.

The Lee-Carter model is

(16) $$\ln q_{i,x,t} = \alpha_{i,x} + \beta_{i,x}\mu_{i,t} + \varepsilon_{i,x,t},$$

where $\varepsilon_{i,x,t}$ is an error term with $E(\varepsilon_{i,x,t}) = 0$ (Lee and Carter, 1992). This method describes age-specific death rates as the sum of an age-specific component that is independent of time and another component that is the product of a time-varying parameter reflecting the general level of mortality, and an age-specific component that represents how rapidly or slowly mortality at each age varies when the general level of mortality changes (Lee, 2000). Future values of $q_{i,x,t}$ can be projected by making projections of future values of $\mu_{i,t}$, as α and β do not vary with time and the expected value of ε equals zero. Projections of $\mu_{i,t}$ can be made by a random walk model with drift:

(17) $$\mu_{i,x,t} = \mu_{i,x,t-1} + c + \eta_{i,x,t},$$

where $\eta_{i,x,t}$ is an error term with $E(\eta_{i,x,t}) = 0$ (Lee and Carter, 1992).

Thus the projection of μ for the year $t+T$ can be calculated by:

(18) $$\hat{\mu}_{i,x,t+T|t} = \hat{\mu}_{i,x,t+T-1|t} + c\,.$$

Projections of $q_{i,x}$ for the year $t+T$ can be calculated from (16):

(19) $$\ln \hat{q}_{i,x,t+T|t} = \alpha_{i,x} + \beta_{i,x}\hat{\mu}_{i,t+T|t}\,.$$

Subtracting (16) from (19) gives

(20) $$\ln \hat{q}_{i,x,t+T|t} - \ln q_{i,x,t} = \beta_{i,x}(\hat{\mu}_{i,t+T|t} - \mu_{i,t})\,.$$

From (18) it can be derived that

(21) $$\hat{\mu}_{i,x,t+T|t} = \mu_{i,x,t} + cT\,.$$

Substituting (21) in (20) yields

(22) $$\ln \hat{q}_{i,x,t+T|t} = \ln q_{i,x,t} + \beta_{i,x}cT\,.$$

Thus $q_{i,x,t+T}$ can be projected by:

(23) $$\hat{q}_{i,x,t+T|t} = \exp(\ln q_{i,x,t} + \beta_{i,x}cT)\,.$$

If it is assumed that $\beta_{i,x}c = \hat{d}_{i,x}$, the Lee-Carter projections are exactly equal to the projections by Variant 3 according to eq. (15). Usually the values of $\beta_{i,x}$ are estimated by singular value decomposition, assuming $\Sigma_x \beta_{i,x} = 1$ and $\Sigma_t \mu_{i,t} = 0$. As for each age x a separate value of β is estimated, the Lee-Carter method tends to produce a rather irregular age pattern of death rates. By using a linear spline, Variant 3 produces a smooth age curve.

Figure 14b shows that the mortality rates of Dutch women projected by Variant 2 are lower than the rates projected by Variant 3. The explanation is that during the last decades mortality rates for Dutch women have declined more slowly than the EU15 average, thus the projection of changes observed in the Netherlands (Variant 3) results in a smaller decline than the projection of changes in the EU15 average (Variant 2). Figure 14a shows that for Dutch men the projections at old ages differ between both variants since mortality rates of elderly men in the Netherlands have declined more slowly than the EU15 average.

Figure 14b Age-specific mortality rates, the Netherlands, men, 2050

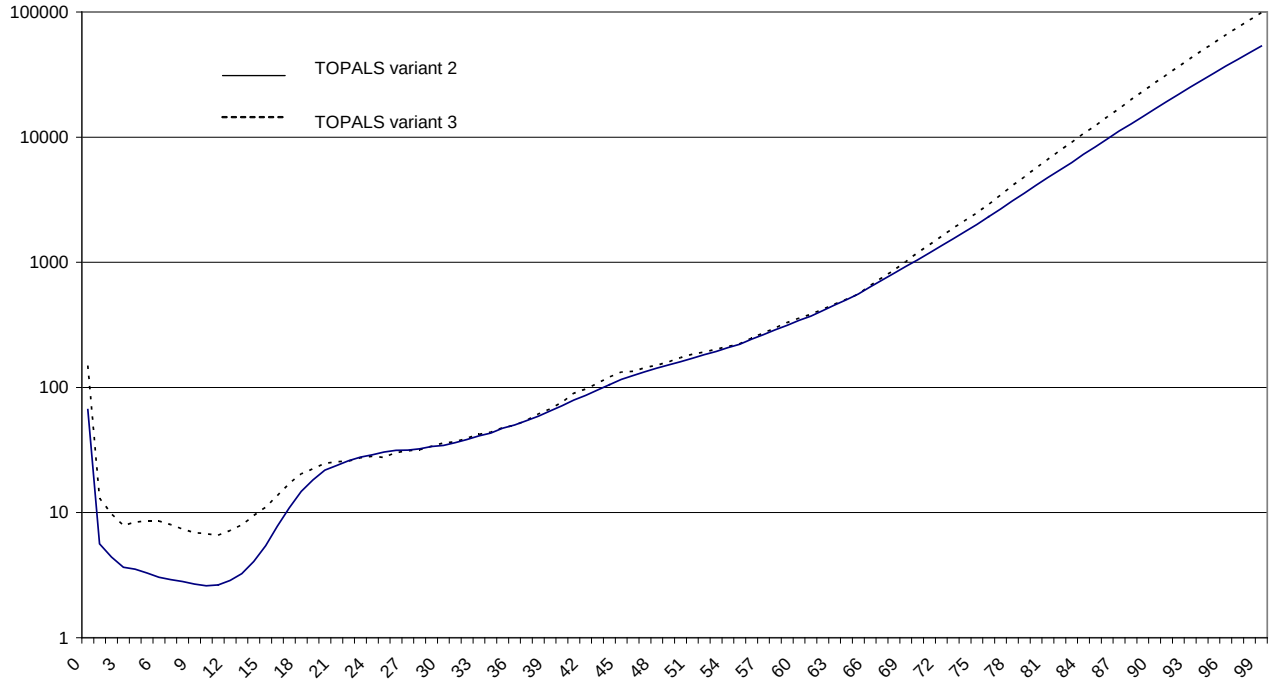


Figure 14b Age-specific mortality rates, the Netherlands, women, 2050

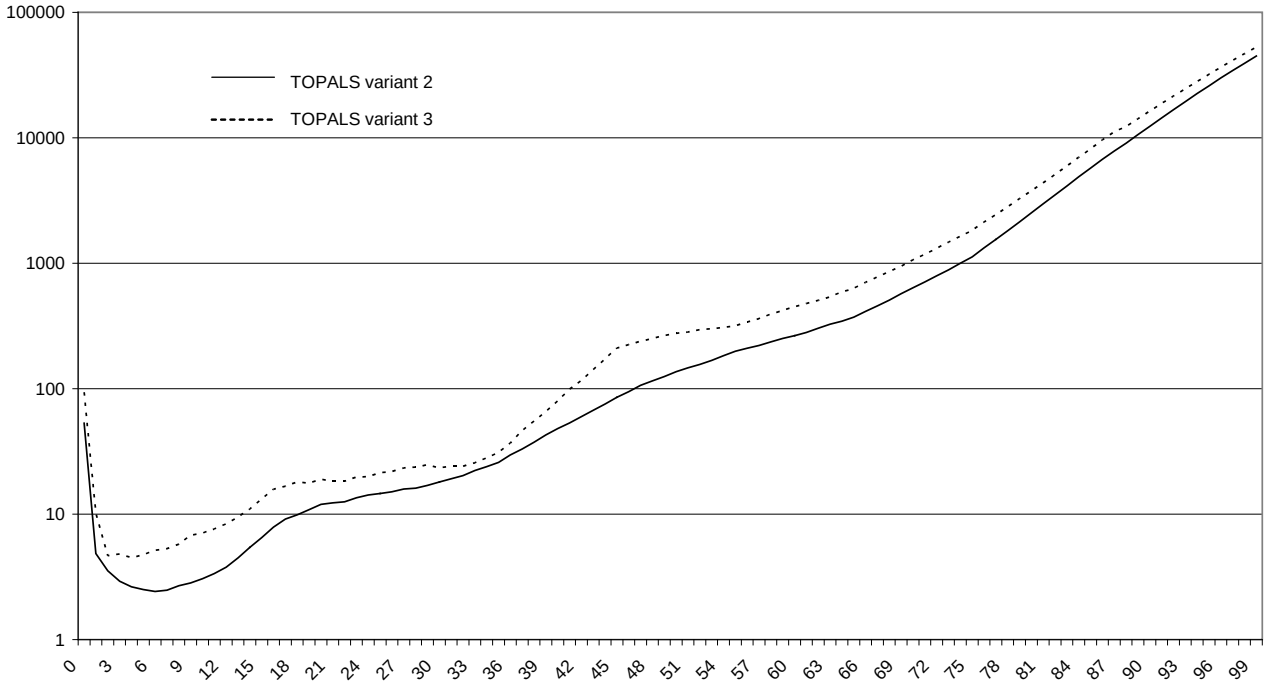


Table 5 shows that using Variant 2 life expectancy at birth for Dutch men and women in 2050 will be almost equal to the EU15 average. According to Variant 3 life expectancy would be considerably lower. Note that in using Variant 2 the relative mortality risks of the Netherlands compared to the EU15 average in 2050 are assumed to be equal to those in 2003. Thus even though no convergence of the age-specific mortality rates is assumed, there is a converging tendency of life

expectancy at birth of Dutch men and women to the EU15 average. The convergence of life expectancy can be explained by the fact that if mortality rates are lower and relative risks stay equal, the relative risks will have a smaller effect on the level of life expectancy at birth. This can be illustrated as follows. The EU15 average life expectancy of men in 2003 equals 75.2 years (see table 5). If in country *A* all age-specific mortality rates would be 20 percent lower than the EU15 average, life expectancy at birth would be 77.7 years, thus 2.5 years higher than the EU15 average. The projected life expectancy of the EU15 average in 2050 equals 83.2 years (table 5). If in country *A* all age-specific mortality rates would still be 20 percent lower, life expectancy in country *A* would equal 85.1 years, 1.9 years lower than the EU15 average. Thus the difference in life expectancy between country *A* and the EU15 average in 2050 would be smaller than in 2003, even though the relative differences of the age-specific mortality rates between country *A* and the EU15 average in 2003 and 2050 would the same. This mechanism is comparable to the fact that if age-specific mortality rates decline by a constant rate, the increase in life expectancy declines.

Figures 15a and 15b show the projections of mortality rates for Italian men and women according to Variant 2. For young ages the projections of Variant 2 are lower than if Italian mortality rates are extrapolated (i.e. Variant 3) whereas the opposite applied to older ages. Table 5 shows that following Variant 2 life expectancy at birth in Italy in 2050 will differ less from that in the Netherlands than according to the EUROPOP baseline scenario. Note that the converging tendency of life expectancy also applies to the difference between men and women. The main cause of the projected narrowing gap between life expectancy of men and women is that during the last decades mortality rates of men have declined more strongly than those of women. This trend is projected into the future. Moreover, as mortality rates will continue to decline, differences in life expectancy at birth will decline because of the same mechanism that causes convergence between countries as described above.

Figure 15a Age-specific mortality rates, Italy, men, 2050

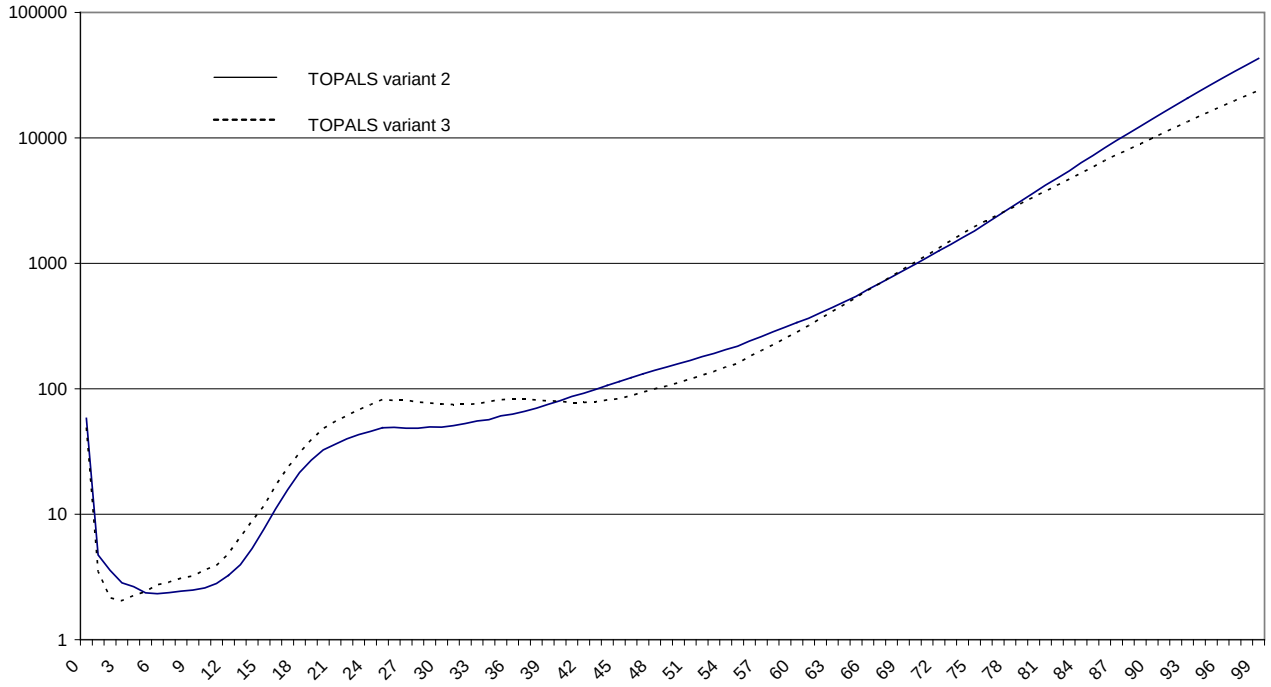


Figure 15b Age-specific mortality rates, Italy, women, 2050

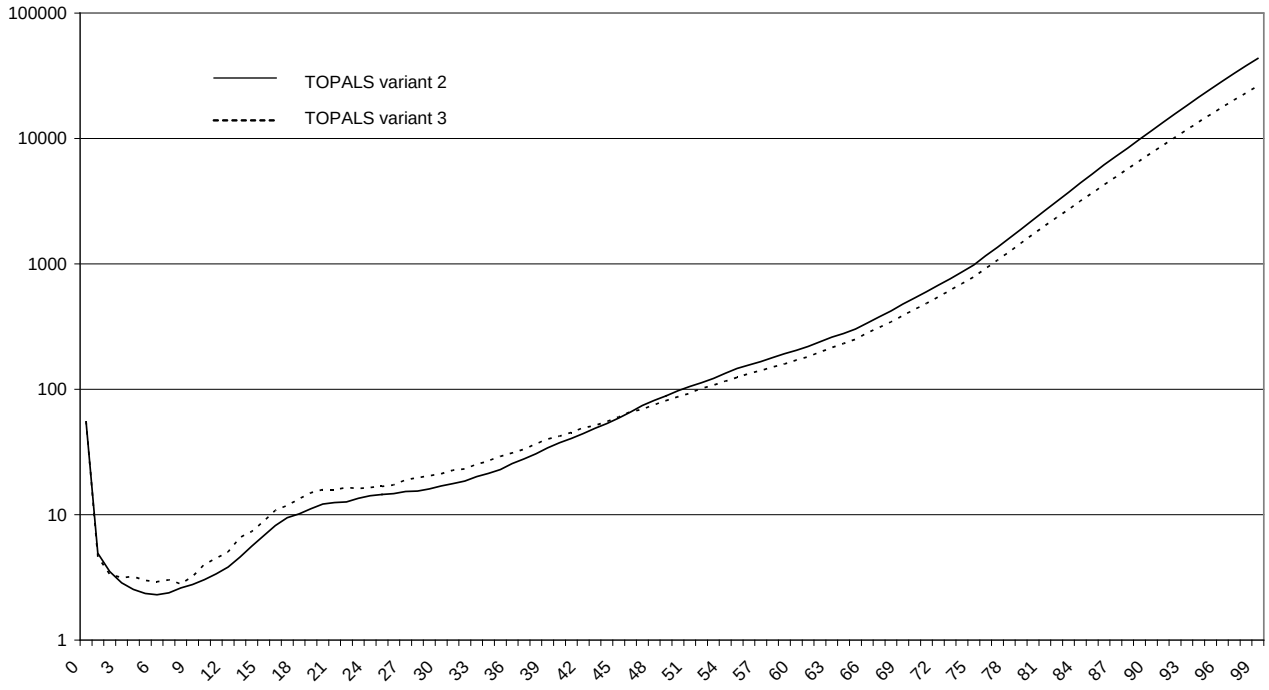


Table 6 shows the life expectancies at birth for all EU15 countries in 2050 projected by Variants 2 and 3 and compares these with the EUROPOP baseline scenario. The table shows clearly that the projections produced by Variant 2 for the separate countries are more closer to each other than those produced by Variant 3 and than the EUROPOP scenarios. The standard deviation of the life expectancies at birth in 2050 according to Variant 2 equals 0.6 for men and 0.7 for women. In 2003 the standard deviation was 1.0 for men and 1.1 for women. Whereas the projections by Variant 2 for 2050 have a smaller standard deviation than that observed in 2003, the opposite is true for Variant 3. According to Variant 2 there is a difference of 2.3 years between the countries with the lowest and highest life expectancy at birth for men in 2050. In 2003 this was 3.9 years. For women Variant 2 projects that the difference between the minimum and maximum life expectancies decreases from 3.8 years to 2.5 years. In contrast Variant 3 projects that the difference between the minimum and maximum values will increase to 6.3 years for men and 8.0 years for women.

Table 6. Projections of life expectancy at birth in 2050

	men				women		
	TOPALS		EUROPOP		TOPALS		EUROPOP
	Variant 2	Variant 3			Variant 2	Variant 3	
Austria	83.3	87.2	83.6		86.6	88.4	87.7
Belgium	82.7	82.0	82.3		86.3	85.9	88.3
Denmark	82.7	80.9	80.9		86.3	84.3	83.7
Finland	82.6	84.4	81.9		86.2	86.5	86.5
France	83.7	85.9	82.7		87.8	90.7	89.1
Germany	83.3	83.3	82.0		86.5	87.5	86.9
Greece	83.7	81.5	80.3		85.3	83.8	85.1
Ireland	83.0	84.6	82.4		86.2	86.2	87.0
Italy	83.8	85.0	83.6		87.2	89.5	88.8
Luxembourg	82.5	82.0	81.6		86.3	84.4	86.6
Netherlands	83.2	81.2	80.2		86.2	82.7	83.6
Portugal	82.0	81.1	80.4		85.6	86.0	86.6
Spain	83.6	82.0	81.4		87.2	88.2	87.9
Sweden	84.3	84.5	83.3		87.3	87.1	86.5
United Kingdom	83.3	83.8	82.9		86.4	86.2	86.6
Average	83.2	83.3	82.0		86.5	86.5	86.7
Minimum	82.0	80.9	80.2		85.3	82.7	83.6
Maximum	84.3	87.2	83.6		87.8	90.7	89.1
Standard deviation	0.6	1.9	1.2		0.7	2.2	1.6

6. The effects of covariates

The level of mortality rates may vary between population categories. For example, people with a high level of educational attainment tend to have considerably lower mortality rates than people with low educational levels. These differences by level of educational attainment may differ by age. For example, at middle ages these differences may be smaller than at older ages. This may be explained by a selection effect. If low-educated people with bad health have relatively high mortality risks at middle ages, relatively few low-educated people with bad health survive to older ages and consequently relatively many low-educated people at older ages have a good health.

TOPALS can simply be extended to take into account differences between population categories. The mortality rate for category y of variable Y is modeled by

(24) $$\hat{q}_{i,x,y} = \hat{c}_{i,x,y}\hat{r}_{i,x}q^*_x,$$

where $\hat{c}_{i,x,y}$ is an estimate of $c_{i,x,y}$ based on a linear spline function similar to (2) and $c_{i,x,y}$ indicates to what extent the

mortality rate for category y differs from the average level. If two covariates are included, the mortality rate for category y of variable Y and category z of variable Z is estimated by:

(25) $$\hat{q}_{i,x,y,z} = \hat{c}_{i,x,y,z}\hat{r}_{i,x}q^*_x .$$

If the effects of the two covariates on mortality are assumed to be independent, the mortality rate can be estimated by

(26) $$\hat{q}_{i,x,y,z} = \hat{c}_{i,x,y}\hat{c}_{i,x,z}\hat{r}_{i,x}q^*_x .$$

Assuming independence of the effects of variables Y and Z on mortality, assumptions on the future level of mortality of persons in category y of variable Y and category z of variable Z can be made by

(27) $$\hat{q}_{i,x,y,z,t+T|t} = \hat{c}_{i,x,y,t+T|t}\hat{c}_{i,x,z,t+T|t}\hat{q}_{i,x,t+T|t},$$

where $\hat{q}_{i,x,t+T|t}$ is projected by eq. (6) or eq. (11), and $\hat{c}_{i,x,y,t+T|t}$ and $\hat{c}_{i,x,z,t+T|t}$ indicate to what extent future mortality rates for categories y and z are assumed to differ from the future average age pattern. If the effects of both covariates are assumed to be interdependent an interaction term can be added to eq. (27). If it is assumed that the differences in the level of mortality between population categories will diminish, the future values of c are expected to move to 1. This can be modelled similarly to eq. (13). A constraint in specifying the values of $\hat{c}_{i,x,y,t+T|t}$ and $\hat{c}_{i,x,z,t+T|t}$ is that the average of $\hat{q}_{i,x,y,z,t+T|t}$ over the categories of y for each i and x as well as over the categories of z should equal $\hat{q}_{i,x,t+T|t}$. Thus:

(28)
$$\sum_y p_{i,x,y,t+T|t}\hat{c}_{i,x,y,t+T|t} = 1$$
and
$$\sum_z p_{i,x,z,t+T|t}\hat{c}_{i,x,z,t+T|t} = 1$$

where $p_{i,x,y,t+T|t}$ is the projected fraction of persons in category y of the variable Y in year $t+T$ for country i and age x and $p_{i,x,z,t+T|t}$ is the fraction of persons in category z of the variable Z.

The method is illustrated by modeling the effect of level of educational attainment on mortality rates for the Netherlands. Empirical analyses suggest that mortality differences between persons with high and low levels of educational attainment are particularly high at middle ages and decline for the oldest age groups. We assume that mortality rates of people with a low level of educational attainment at middle ages will be 50 percent higher than the average and that mortality rates of people with a medium level of educational attainment will equal the average level. See table 7. In addition we assume that 20 percent of the population aged 20 years or over has a low level of educational attainment and 50 percent has a medium level. Than eq (28) implies that the relative mortality rates of people with a high level of educational attainment at the middle ages will be 30 percent lower than the average. This can be calculated easily as follows:

(29) $$p_L c_L + p_M c_M + p_H c_H = 1,$$

where p_L is the proportion of the population with a low level of educational attainment and p_M and p_H are the proportions with medium and high levels respectively. Then if assumptions are specified on c_L and c_M, c_H can be calculated from (29) as follows:

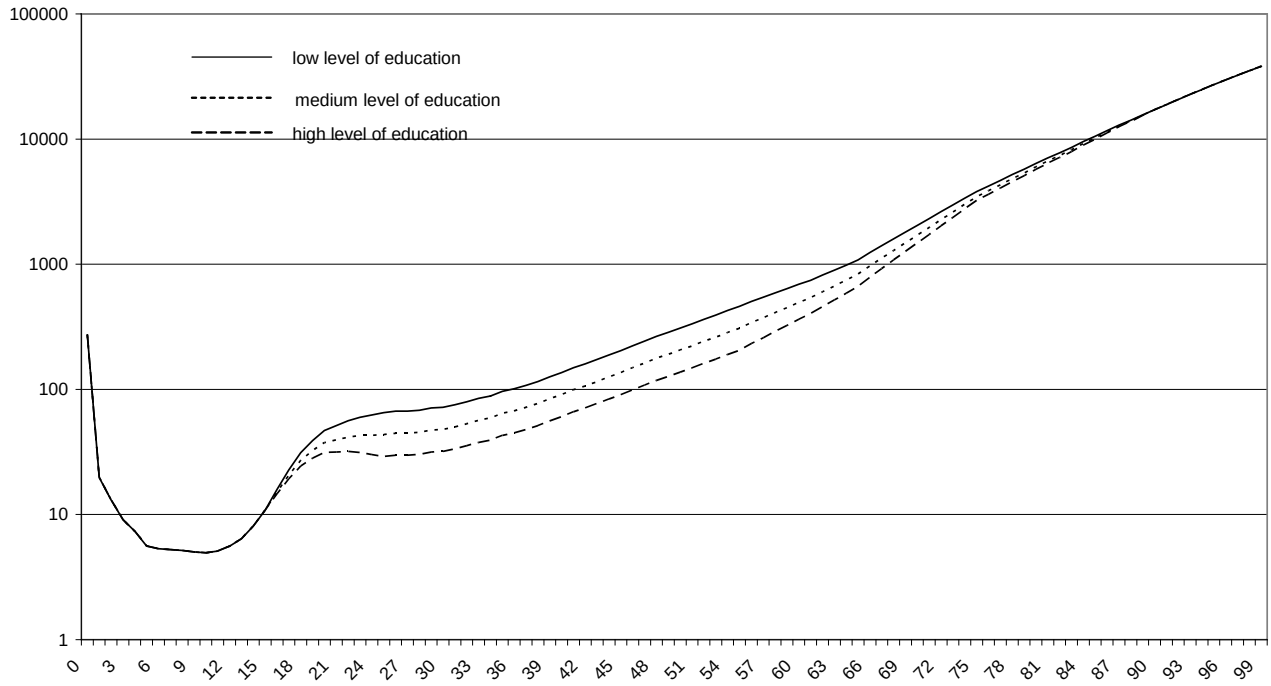
(30) $$c_H = \frac{1 - p_L c_L - p_M c_M}{p_H},$$

Instead of making assumptions on the values of c_y relative to the average level, one may make assumptions of c_y relative to one reference category. For example, if a medium level of educational attainment is used as reference category, c_M in eq. (30) equals one[4]. On the basis of these assumptions age-specified mortality rates for people with low, medium and high levels of educational attainment for Dutch men in 2050 are projected using eqs. (27) and (30). See figure 16. According to these age-specific mortality rates life expectancy at birth of a man with a low level of educational attainment would equal 78.6 years and that of a highly educated man would equal 81.4 years.

[4] In this example the value of the reference category equals the average value. This is not necessary. If the values c_y are regarded as relative risks compared with a reference category, the constraint (28) can be omitted. In that case the standard age schedule describes the age-specific rates for the reference category rather than for the average population.

Table 7. Assumptions about relative mortality rates by level of educational attainment							
	distribution of population by educational level				relative mortality rates		
age	low	medium	high		low	medium	high
0	100%	0%	0%		1.0	1.0	1.0
1-10	100%	0%	0%		1.0	1.0	1.0
11-20	50%	50%	0%		1.0	1.0	1.0
21-30	20%	50%	30%		1.5	1.0	0.7
31-40	20%	50%	30%		1.5	1.0	0.7
41-50	20%	50%	30%		1.5	1.0	0.7
51-60	20%	50%	30%		1.5	1.0	0.7
61-70	20%	50%	30%		1.3	1.0	0.8
71-80	20%	50%	30%		1.1	1.0	0.9
81+	20%	50%	30%		1.0	1.0	1.0

Figure 16 Age-specific mortality rates, by level of educational attainment, men, Netherlands, 2050



7. Conclusions

This paper introduces a simple, but flexible method for making projections of age-specific rates: TOPALS (Tool for projecting age patterns using linear splines). The method consist of two steps. First, the age-specific rates are fitted to a standard age schedule using relative risks which are modeled by a linear spline function. Second, assumptions are made on the future values of the relative risks. Using a standard age schedule has the benefit that assumptions about a limited number of parameters need to be specified rather than about all individual age-specific rates. Using a linear spline function to model relative risks has the benefit that it is flexible, as it allows to describe all kinds of variations compared to the standard age schedule, while still producing a smooth age curve.

This paper illustrates TOPALS by fitting age-specific mortality and fertility rates for Italy and the Netherlands. The (unweighted) EU15 average is used as standard age schedule. Even though the age patterns of mortality and fertility are quite different, TOPALS is capable of fitting both age curves accurately.

In using TOPALS for making projections three approaches can be followed. These are illustrated by projecting age-specific mortality rates for the year 2050 for Italy and the Netherlands. In Variant 1 assumptions about changes in the relative risks are specified. By using linear splines it is sufficient to make assumptions about average changes for age groups rather than assumptions about changes in relative risks for each age separately. This variant is illustrated by making assumptions which correspond to the EUROPOP baseline scenario. In Variant 2 the standard age schedule is projected into the future. For this purpose a random walk model with drift is used. By using linear splines projections for age groups rather than for each individual age have to be made. One benefit is that this tends to produce a smoother age pattern. Subsequently assumptions need to be made about the future values of the relative risks. They may be assumed to remain constant, but one may also assume that they will change, for example that the differences between countries will decline. It should be noted that even if the relative risks are not assumed to change and thus the (relative) differences in age-specific mortality rates between countries will not change, this method will project declining differences in life expectancy. This is caused by the non-linear relationship between life expectancy and age-specific mortality rates. If the level of mortality rates is lower, the same relative difference in mortality rates will produce a smaller difference in life expectancy. In Variant 3 the age-specific rates of each country in the last observation year are used as standard age schedule. Assumptions on future values of relative risks indicate how the age pattern will change. This variant tends to produce more differences in projections of life expectancy between countries than Variant 2. The projections by Variant 3 resemble the projections by the well-known Lee-Carter method, the main difference being that Variant 3 tends to produce a smoother age pattern than the Lee-Carter method.

TOPALS can be extended by taking into account effects of covariates. On the basis of assumptions about differences in age-specific rates by categories of a covariate, different age patterns can be modeled. One benefit of using linear splines is that this allows to take into account that the effects of covariate may vary between age categories. Nevertheless the method produces a smooth age pattern.

8. References

Brass, W. (1974) Perspectives in population prediction: Illustrated by the statistics of England and Wales. *Journal of the Royal Statistical Society*, 137, series A: 532-583.

Lee, R. (2000) The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal* 4 (1): 80-91.

Lee, R.D. and L. Carter (1992) Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association* 87: 659-671.

Li, N. and R. Lee (2005) Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. *Demography* 42 (3): 575-594.

Zeng Yi, Wang Zhenglian, Ma Zhongdong and Chen Chunjun (2000), A simple method for projecting or estimating α and β: An extension of the Brass Relational Gompertz Fertility Model. In: *Population Research and Policy Review* 19: 525-549.