

## **Mortality Trajectories at Extreme Old Ages: A Comparative Study of Different Data Sources on U.S. Old-Age Mortality**

Natalia S. Gavrilova, Leonid A. Gavrilov  
Center on Aging, NORC at the University of Chicago, 1155 East 60th Street, Chicago, Illinois, 60637,  
USA

### **Abstract**

The growing number of persons living beyond age 80 underscores the need for accurate measurement of mortality at advanced ages. Our earlier published study challenged the common view that the exponential growth of mortality with age (Gompertz law) is followed by a period of deceleration, with slower rates of mortality increase (*North American Actuarial Journal*, 2011, 15(3): 432-447). This refutation of mortality deceleration was made using records from the U.S. Social Security Administration's Death Master File (DMF).

Taking into account the significance of this finding for actuarial theory and practice we tested these earlier observations using additional independent datasets and alternative statistical approaches. In particular, the following data sources for the U.S. mortality at advanced ages were analyzed: (1) Data from the Human Mortality Database (HMD) on age-specific death rates for 1890-1899 U.S. birth cohorts; (2) Recent extinct birth cohorts of U.S. men and women based on DMF data. (3) Mortality data for rail road retirees.

In the case of HMD data, the analyses were conducted for 1890-1899 birth cohorts in the age range 80-106 years. Mortality was fitted by the Gompertz and the logistic (Kannisto) models using weighted non-linear regression and Akaike information criterion as the goodness-of-fit measure. All analyses were conducted separately for men and women. It was found that for all studied HMD birth cohorts the Gompertz model demonstrated better fit of mortality data than the Kannisto model in the studied age interval. Similar results were obtained for U.S. men and women born in 1890-1899 and rail road retirees born in 1895-99 using the full DMF file (obtained from the National Technical Information Service, NTIS). It was also found that mortality estimates obtained from the DMF records are close to estimates obtained using the HMD cohort data.

An alternative approach for study mortality patterns at advanced ages is based on calculating age-specific rate of mortality change (or life table aging rate, LAR) after age 80. This approach was applied to age-specific death rates for Canada, France, Sweden and the United States available in HMD. It was found that for all studied 24 single-year birth cohorts LAR does not change significantly with age in the age interval 80-100 years suggesting no mortality deceleration in this age interval. Simulation study of LAR demonstrated that the apparent decline of LAR after age 80 found in earlier studies may be related to biased estimates of mortality rates measured in a wide 5-year age interval.

Taking into account that there exist several empirical estimates of hazard rate (Nelson-Aalen estimate, actuarial estimate, Sacher estimate) a simulation study was conducted in order to find out which one is the most accurate and unbiased estimate of hazard rate at advanced ages. Computer simulations demonstrated that some estimates of mortality (Nelson-Aalen estimate, actuarial estimate) as well as kernel smoothing of hazard rates may produce spurious mortality deceleration at extreme ages, while the Sacher estimate turns out to be the most accurate estimate of hazard rate. Possible reasons of finding apparent mortality deceleration in earlier studies are also discussed.

This study was supported in part by NIH grant R01 AG028620.

## 1. Introduction

The growing number of persons living beyond age 80 underscores the need for accurate measurement of mortality at advanced ages. Accurate estimates of mortality at advanced ages are essential for improving forecasts of mortality and predicting the population size of the oldest old age group.

Earlier studies suggest that the exponential growth of mortality with age (Gompertz law) is followed by a period of deceleration, with slower rates of mortality increase (Greenwood and Irwin 1939; Horiuchi and Wilmoth 1998; Thatcher 1999; Thatcher, Kannisto and Vaupel 1998). It is believed that mortality at advanced ages has a tendency to deviate from the Gompertz law (Gavrilov and Gavrilova 1991), so that the logistic model is suggested for fitting human mortality after age 80 years (Horiuchi and Wilmoth 1998; Wilmoth et al. 2007).

At the same time, making estimates of hazard rate at very old ages is difficult because of very small fraction of survivors at these ages in most countries. Data for extremely long-lived individuals are scarce and subjected to age exaggeration. In order to minimize statistical noise in estimates of mortality at advanced ages researches have to pool data for several calendar periods (Depoid 1973; Thatcher 1999). Single-year life tables for many countries have very small numbers of survivors to age 100 that makes estimates of mortality at advanced ages unreliable. On the other hand, aggregation of deaths for several calendar periods creates a heterogeneous mixture of cases from different birth cohorts. Theoretical models suggest that mortality deceleration at advanced ages may be caused by population heterogeneity even if individual risk of death follows the Gompertz law (Beard 1959; Beard 1971; Vaupel, Manton and Stallard 1979).

In addition to heterogeneity problem, there is a problem of using proper empirical estimates of hazard rate at extreme ages when mortality is high and grows with age very rapidly. This problem is sometimes overlooked by researchers who believe that mortality estimates, which work well at young adult ages (like one-year probability of death) can work equally well at very old ages.

Finally, the problem of age misreporting by older people may still be a problem affecting estimates of mortality at advanced ages (Coale and Kisker 1986; Elo et al. 1996; Hill et al. 1997; Jdanov et al. 2008; Kestenbaum 1992). In most cases the age misreporting at older ages results in mortality underestimation (Preston, Elo and Stewart 1999). Taking into account that the accuracy of age reporting is positively correlated with education it is reasonable to expect improvement in age reporting over time and less expressed mortality underestimation at older ages. Indeed it was found that mortality in older U.S. cohorts deviated more from the Gompertz model compared to mortality of younger birth cohorts (Gavrilov and Gavrilova 2011).

In this paper we analyze mortality trajectories at advanced ages using two independent sources of mortality data and two different approaches of analyzing age patterns at older age. This study is based on data from the U.S. Social Security Administration Death Master File (DMF), which allows compiling data for large single-year birth cohorts. Some already extinct birth cohorts covered by DMF could be studied by the method of extinct generations (Vincent 1951), which turns to be more accurate compared to traditional method based on death statistics and population census data. Another data source is the Human Mortality Database, which became a traditional resource for demographers and actuaries. Methods of mortality trajectory analysis include comparing alternative models of mortality using standard goodness-of-fit measure and study of age patterns for life table aging rate at advanced ages (Horiuchi and Coale 1990; Horiuchi and Wilmoth 1998).

## 2. Testing competing mortality models using two independent data sources.

The actuaries including Gompertz himself were the first who noticed the phenomenon of mortality deceleration. They also proposed a logistic formula for fitting mortality growth with age in order to account for mortality leveling-off at advanced ages (Perks, 1932; Beard, 1959, 1971). British actuary, Robert Eric Beard, introduced a model of population heterogeneity with gamma distributed individual

risk in order to explain mortality deceleration at older ages (Beard 1959). This explanation is now considered to be the most common explanation of mortality deceleration phenomenon (Horiuchi and Wilmoth 1998).

Previous studies of mortality trajectories used data of reasonably good quality for European countries and Japan, which had relatively small population sizes. As a result, few individuals in these countries survived to advanced ages, so the researchers often had to pool together data for as many as ten years to provide sufficient sample size for the study. United States has the largest population size among the advanced economies. The quality of vital statistics in the United States was not considered acceptable when the first comprehensive studies of mortality trajectories have been conducted (Horiuchi and Coale 1990; Horiuchi and Wilmoth 1998; Thatcher et al. 1998). However, the quality of U.S death data was gradually improving over time (Manton, Akushevich and Kulminski 2008) and now can be used in mortality analysis. In this study we compared two models (Gompertz model and Kannisto model) commonly used in actuarial practice by analyzing U.S. cohort mortality data taken from two independent data sources.

## 2.1. Data Sources

### *Social Security Administration Death Master File (DMF)*

Until recently DMF was a publicly available data resource, which allowed a search for deceased individuals in the United States using various search criteria: birth date, death date, first and last names, social security number, place of last residence, etc. This resource covered deaths that occurred during the period starting in 1937 (Faig 2001) and captured about 95% of deaths recorded by the National Death Index (Sesso, Paffenbarger and Lee 2000). According to other estimates, DMF covered about 90 percent of all deaths for which death certificates are issued (see Faig, 2001) and about 92-96 percent of deaths for persons older than 65 years (Hill, Rosenwaike, 2001). Unfortunately, recent developments made this data source publicly unavailable and introduced significant gaps in the number of records available in DMF.

In this study we used DMF full file obtained from the National Technical Information Service (NTIS). This is the latest available complete version of DMF where the last deaths occurred in September 2011. The advantage of this data source is that some already extinct birth cohorts covered by DMF could be studied by the method of extinct generations (Kannisto 1988, 1994; Vincent 1951). In our earlier study we conducted analyses of DMF data for both sexes pooled together. In this study mortality at advanced ages is studied separately for men and women belonging to ten recent extinct and almost extinct birth cohorts. The SSA DMF does not have information about sex of the deceased. To cope with this limitation of data sample, we conducted a procedure of sex identification using information about the 1,000 most commonly used baby first names in the 1900s provided by the Social Security Administration (<http://www.ssa.gov/OACT/babynames>). These data come from a sample of 5 percent of all Social Security cards issued to individuals who were born during the 1900s in the United States. The detailed procedure of sex assignment for DMF records was described elsewhere (Gavrilov and Gavrilova 2011). Eventually we were able to identify sex for 91-93 percent of persons in our sample. The remaining 7-9 percent of persons with unknown sex had approximately the same mean lifespan as the remaining percent of individuals with identified sex pooled together (checked with the t-test statistics), so the existence of possible sex non-identification bias in mortality seems unlikely.

We obtained data for persons who died before 2011 and were born in 1890-1899. Persons born in these years and alive in 2011 should survive to at least 112 years, which can be considered as very unlikely event. Thus, the 1890-1899 birth cohorts in this sample may be considered as practically extinct. Assuming that the number of living persons belonging to these birth cohorts in 2011 is close to zero, it is possible to construct a cohort life table using well-known method of extinct generations, which was suggested and explained by Vincent (1951) and developed further by Kannisto (Kannisto 1994).

DMF provides information about years and months of birth and death. However information on exact day of death is not available for all records, so we were not able to calculate lifespan in days. Nevertheless we were able to calculate individual lifespan with one-month accuracy, which is still higher than the accuracy of traditional yearly estimate of lifespan. In the first stage of our analyses we calculated an individual life span in completed months.

Using single-year birth cohort data from DMF we were able to minimize the effects of population heterogeneity on age-related mortality dynamics. We were also able to calculate more accurate monthly estimates of hazard rates, which are less prone to possible biases caused by violation of typical assumptions used in mortality estimation (Kimball 1960). However we were not able to control for possible age misreporting among the deceased. The most common type of age misreporting among very old individuals is age exaggeration (Willcox *et al.* 2008). This type of age misreporting results in underestimation of mortality rates at advanced ages (Preston *et al.* 1999).

One approach to evaluate data quality at advanced ages is to calculate female to male ratio at advanced ages. Taking into account that female mortality is always lower than male mortality (Kannisto 1994), it is reasonable to expect that the female-to-male ratio should continuously increase with age. On the other hand, old men have a tendency for age exaggeration and in populations with poor age registration there is a relative excess of men at very advanced ages (Caselli *et al.* 2006; Willcox *et al.* 2008). This approach was applied to DMF data in our earlier study and showed that data quality in DMF is acceptable up to the age 106 years (Gavrilov and Gavrilova 2011). These results are supported by independent study of age verification in DMF records, which found that invalid age claims in this source increase from 65% at age 110-111 to 98% by age 115. This finding suggests high proportion of age misreporting after age 110 and probably after age 106 years as well (Young *et al.* 2010).

In order to use mortality data with better quality we also selected DMF records for so-called “non-Southern” group. To create this group, we excluded less reliable records for those persons who applied for social security number in the Southeast states (AR, AL, GA, MS, LA, TN, FL, KY, SC, NC, VA, WV), Southwest states (AZ, NM, TX, OK), Puerto Rico and Hawaii. We also excluded all records for persons applied to SSN in New York and California states because of very high proportion of immigrants (with unknown quality of age reporting) residing in these states. We found that non-Southern group demonstrated significantly less mortality understatement (Gavrilov and Gavrilova 2011), which suggests better age reporting in this group (Preston *et al.* 1999).

#### *Human Mortality Database.*

The Human Mortality Database (HMD) was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. This is a publicly available data resource, which can be reached at <http://www.mortality.org>.

We used age-specific death rates for 1890-1899 single-year U.S. birth cohorts. These are raw mortality data not fitted at advanced ages by any parametric formula. Data for 1900 and later birth cohorts were smoothed using Kannisto formula (Jdanov, personal communication), so they were not used in this analysis. So we analyzed only 1890-1899 single-year U.S. cohort age-specific death rates separately for men and women. Mortality rates of period life tables in HMD were smoothed after age 80 by fitting a logistic function to observed death rates (Wilmoth *et al.* 2007).

For study of mortality changes using life table aging rate (LAR) measure, cohort age-specific death rates were used for the following countries: Canada, France, Sweden and the United States.

## 2.2. Statistical Methods.

Study of data quality control at advanced ages suggests that age reporting among the oldest-old in the United States is good until the age of 106 years (Gavrilov and Gavrilova 2011). It means that comparing mortality models beyond this age is not feasible because of poor quality of mortality data.

For DMF dataset, we used a subsample of deaths for persons applied to SSNs in the 'non-Southern' states (described above) and born in 1890-1899, because data for these birth cohorts have reasonably good quality. The Stata software calculates nonparametric estimates of major survival functions including the Nelson-Aalen estimator of hazard rate (force of mortality). In this study, survival times were measured in months, so the estimates of hazard rates initially had a dimension of month<sup>-1</sup>. For the purpose of comparability with other published studies, which typically use the year<sup>-1</sup> time scale, we transformed the monthly hazard rates to the more conventional units of year<sup>-1</sup>, by multiplying these estimates by a factor of 12 (one month in the denominator of hazard rate formula is equal to 1/12 year). It should be noted that hazard rate, in contrast to probability of death, can be greater than 1, and therefore its logarithm can be greater than 0 (and we indeed observed these values at extreme old ages in some cases). In this paper we focus our analyses on 1890-1899 birth cohorts, because we found that data quality for earlier cohorts is not particularly good. We tested competing models of mortality in the age interval 85-106 years using nonlinear regression method.

In the case of HMD data, weighted nonlinear regression model was applied to age-specific death rates in the age interval 80-106 years. Age-specific exposure values were used as weights (Muller, Wang and Capra 1997). Mortality data were fitted using the most frequently used models of adult mortality: the Gompertz model (Beard 1971; Gavrilov and Gavrilova 1991; Gompertz 1825) and the Kannisto model (Thatcher et al. 1998).

$$\text{Gompertz:} \quad \mu_x = a e^{bx} \quad (1)$$

$$\text{Kannisto:} \quad \mu_x = \frac{a e^{bx}}{1 + a e^{bx}} \quad (2)$$

Goodness-of-fit for Gompertz and Kannisto models was evaluated using the Akaike Information criterion, AIC (Akaike 1974):

$$AIC = -2 \ln L + 2k$$

where  $\ln L$  is the maximized log-likelihood of the model and  $k$  is the number of parameters estimated. Calculations of AIC were performed using Stata command `estat ic` after parameters of each model were estimated using non-linear regression procedure.

For both datasets data for men and women were studied separately. Calculations were done using the Stata statistical software, release 11 (StataCorp 2009).

## 2.3 Results

### a. DMF Data

Results of the hazard rate estimation for 1898 U.S. female birth cohort are presented in Figure 1. Note that mortality trajectory in semi-log scale is linear up to the age 106 years suggesting that mortality follows the Gompertz law. After age 106 years data points show very high variation probably because of declining data quality (possible age misreporting).

Next step of our study was to compare two competing models of mortality at advanced ages - the Gompertz and the Kannisto models. Study of data quality at advanced ages suggests that age reporting among the oldest-old in the United States is reasonably good until the age of 106 years (Gavrilov and Gavrilova 2011). It means that comparing mortality models beyond this age is not feasible because of poor quality of mortality data. Also, it was found that quality of age reporting is not good for older age cohorts and for persons applied for SSNs in the Southern states resulting in spurious mortality deceleration at advanced ages (Gavrilov and Gavrilova 2011). For this reason, in this paper we used a subsample of deaths for persons applied to SSNs in the 'non-Southern' states (described above) and born in 1890-1899, because these data have a reasonably good quality.

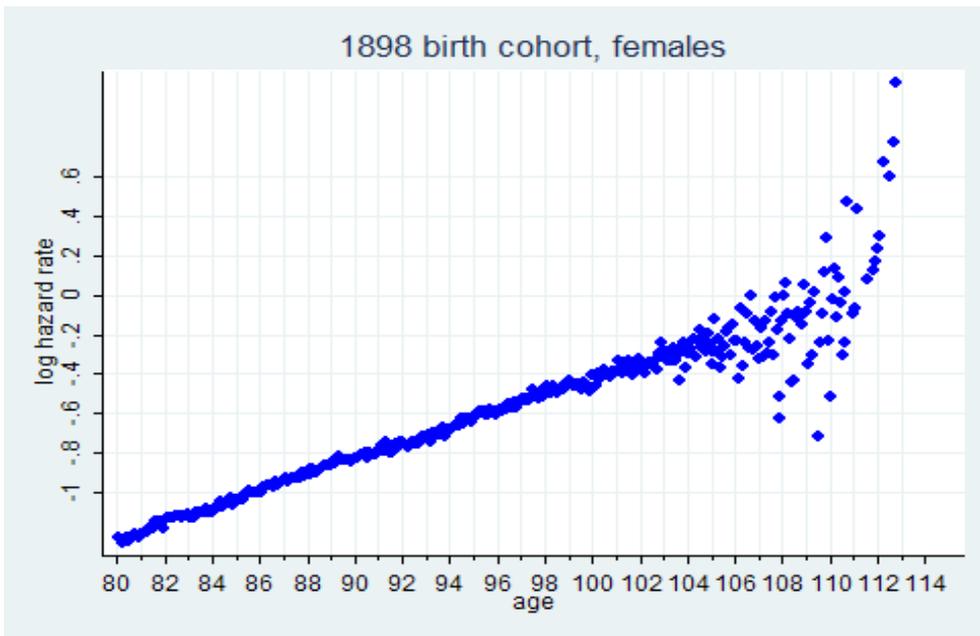


Figure 1. Monthly estimates of hazard rate at advanced ages in semi-log scale. U.S. females, 1898 birth cohort.

Figures 2a and 2b present values of AIC for Gompertz and Kannisto models for ten studied birth cohorts of men and women respectively. Note that in all cases (studied birth cohorts), the Gompertz model demonstrates better fit (lower AIC) than the Kannisto model in age interval 85-106 years.

Obtained values of AIC shown in figures 2a and 2b are negative, and this fact sometimes raises questions from researchers. Negative values of AIC are totally compatible with its definition and appear in some research papers (Lee and Russell 2004). A discussion about negative AIC values can be found on the Internet, including the ResearchGate service (ResearchGate 2013).

At this moment, we cannot make a conclusion that Gompertz model fits mortality data better than the logistic model after age 106, because of low quality of age reporting beyond this age. At the same time, the available data indicate that the Gompertz model fits mortality data well until the age 106 years. Taking into account that survival beyond age 106 years is rather rare event, it would be reasonable to suggest the use of Gompertz model of hazard rate rather than the Kannisto model for closing cohort life tables in actuarial practice. In this case, mortality modeling could be done first for hazard rate (mortality force) function and then all life table functions (including probability of death,  $q_x$ ) could be derived on the basis of modeled values of hazard rate.

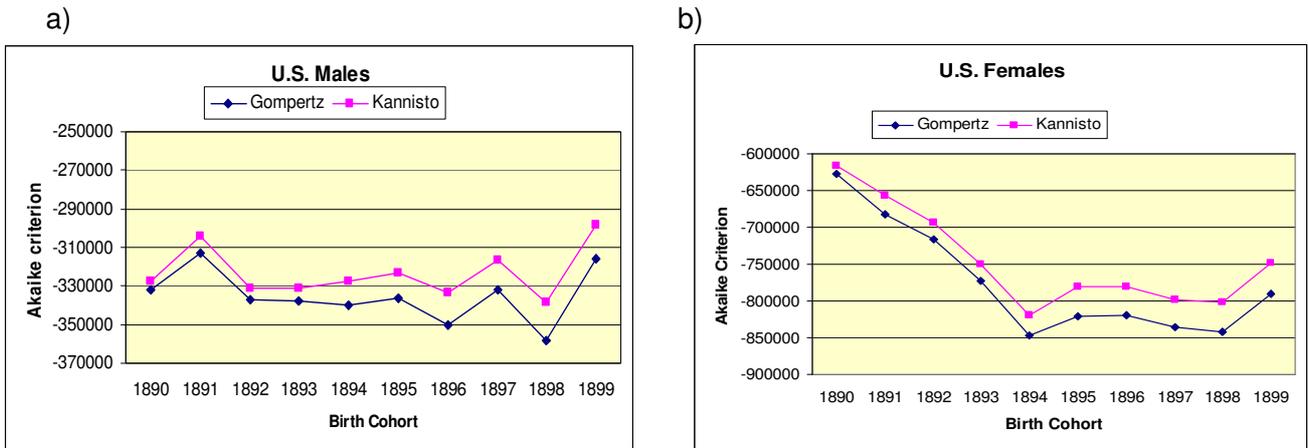


Figure 2. Akaike information criterion (AIC) for Kannisto and Gompertz models, by U.S. birth cohort (non-Southern states). a) men; b) women.

*b. Mortality of rail road workers.*

Our earlier study demonstrated that deceleration of mortality in later life is more expressed for data with lower quality (Gavrilov and Gavrilova 2011). DMF database does not contain specific information about occupation but it has information about rail road workers who receive special governmental pensions and who have specific first 3-digits (700-728) of their Social Security numbers. Our hypothesis was that rail road workers (due to specifics of their occupation) may have better age reporting compared to the rest of the same birth cohort. If this hypothesis is correct then mortality of rail road workers at advanced ages should not demonstrate mortality deceleration at advanced ages.

Figures 3a and 3b show mortality for rail road workers and for the rest of the same birth cohort (born in 1895-99) for men and women respectively. Note that mortality of rail road workers in semi-log coordinates is steeper compared to the rest of corresponding birth cohort.

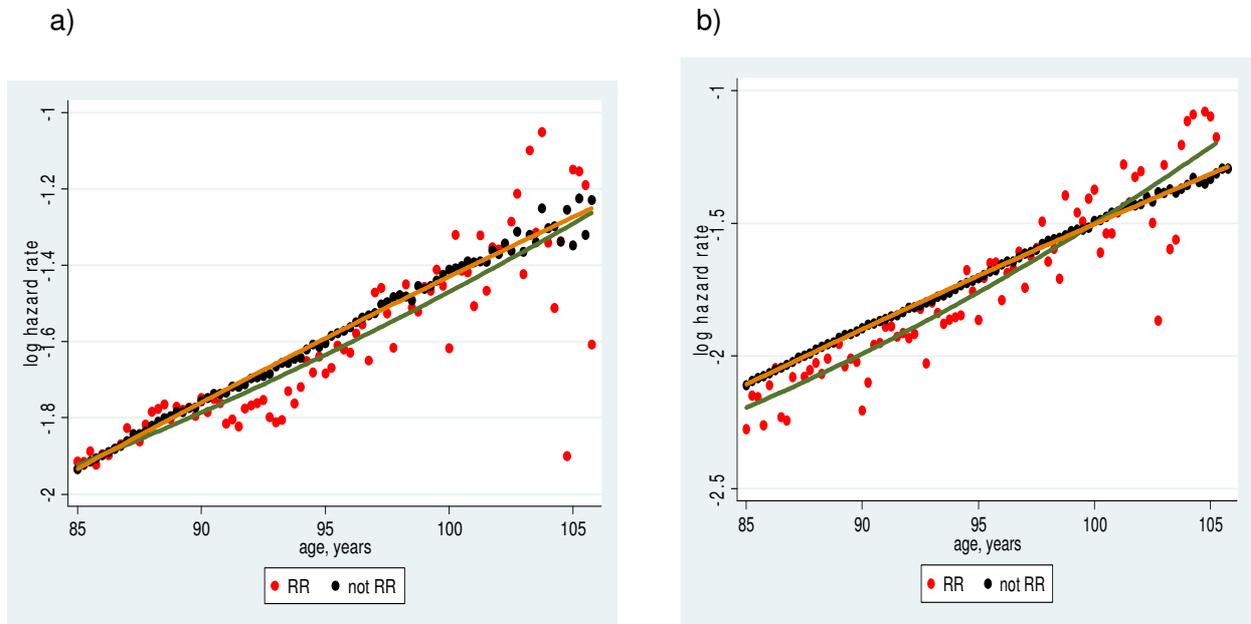


Figure 3. Mortality of railroad retirees and their non-railroad peers, 1895-99 U.S. birth cohort. a) men; b) women.

Mortality in semi-log coordinates was fitted with quadratic regression model in age interval 85-105 years. It turned out that the model has significant and positive (rather than negative) quadratic term for both male and female rail road workers suggesting an acceleration rather than deceleration of mortality. For mortality of the rest of 1895-99 cohort the quadratic term was not different from zero for both men and women suggesting that mortality at advanced ages follows the Gompertz law. We may conclude that (as predicted) mortality trajectories of rail road retirees are slightly steeper compared to the rest of cohort.

*c. Human Mortality Database*

Taking into account that complete DMF file is not readily available now, we decided to use well-known data resource (HMD), so that our results can be easily reproduced by other researchers.

Figure 4 presents age-specific death rates in semi-log scale for 1895 birth cohort of U.S. women. Note that mortality trajectory follows the Gompertz law fairly well up to very advanced ages.

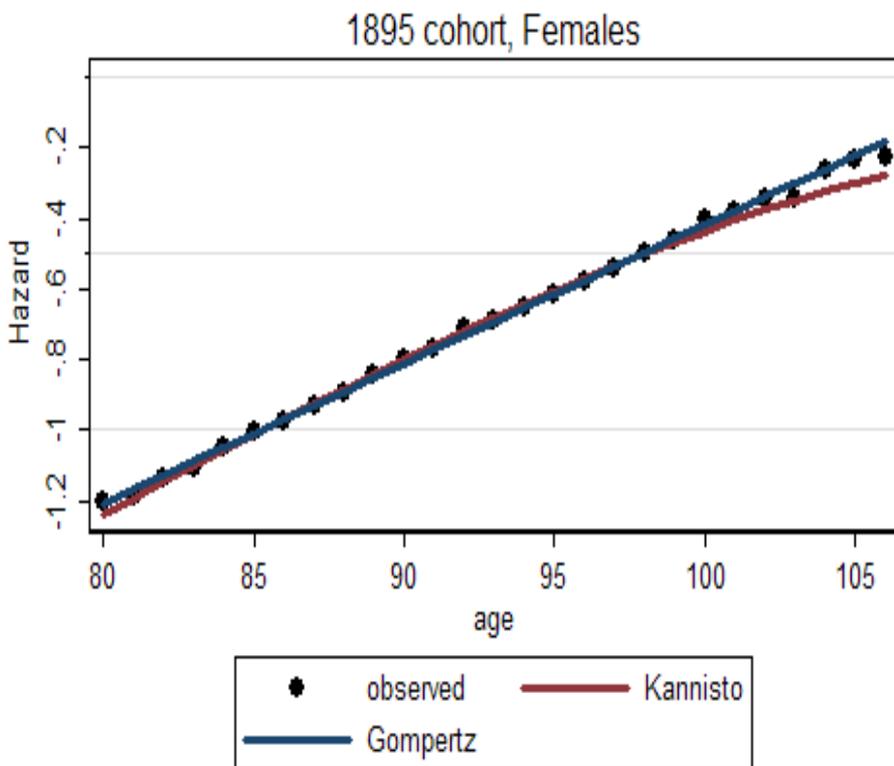


Figure 4. Age-specific hazard rates for U.S. females (1895 birth cohort) fitted by Gompertz and Kannisto models.

In order to quantify this finding, we compared Gompertz and Kannisto models using AIC as a goodness-of-fit measure. The analysis included 10 single-year U.S. birth cohorts (1890-1899) and was conducted separately for men and women. The results of this study for men and women are presented in Figures 5a and 5b respectively. Note that for all ten cases Gompertz model demonstrates better fit than Kannisto model for both men and women in age interval 80-106 years

Due to some limitations of the DMF data (incompleteness of death registration before the 1970s), we were unable to estimate mortality before the age 85-88 years. There is also a question whether our estimates of Gompertz parameters based on DMF data (slope parameter in particular) are applicable to

a wider age interval or they are specific only for advanced ages (so-called two-stage Gompertz model). Age-specific death rates from the Human Mortality Database are available for much younger ages, so there is an opportunity to compare results obtained from these two independent data sources.

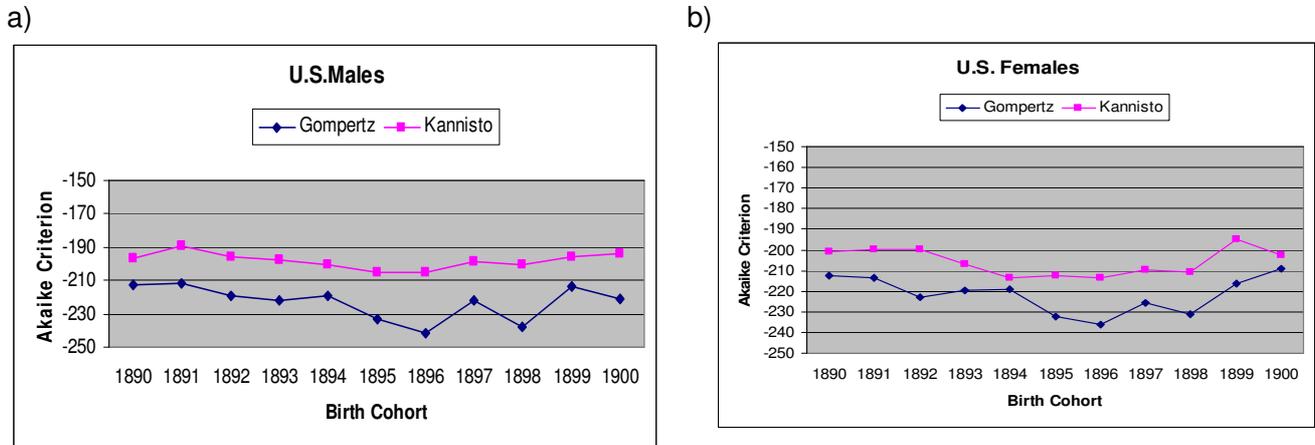


Figure 5. Akaike information criterion (AIC) for compare Kannisto and Gompertz models, by birth cohort (U.S. data from HMD). a) men; b) women.

Figure 6 shows the trajectories of age-specific hazard rates for 1898 birth cohort of females based on data from HMD and DMF over a broad interval starting at age 60 years. Note that hazard rate estimates for the DMF birth cohort are practically identical to the age-specific death rates obtained from HMD. Also note that mortality of DMF birth cohort has the same slope in semi-log coordinates as mortality of HMD birth cohort calculated for much wider age interval, which does not support the suggestion about the two-stage Gompertz model of mortality at advanced ages.

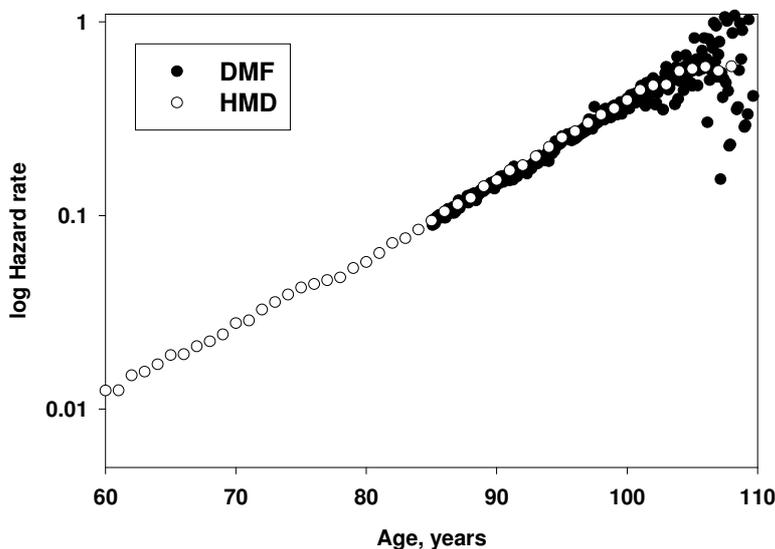


Figure 6. Comparison of DMF and HMD mortality data for 1898 birth cohort of U.S. women.

Indeed, the maximum likelihood estimator of the Gompertz slope parameter for mortality in 1898 cohort measured in the interval 88-106 years ( $0.0786 \text{ year}^{-1}$ , 95%CI: 0.0786-0.0787) does not differ from the slope parameter calculated over the age interval 40-105 years for HMD data:  $0.0785 \text{ year}^{-1}$ ,

95%CI: 0.0772-0.0797. Thus, we may conclude that the estimates of hazard rates at advanced ages based on individual mortality data practically coincide with hazard rate estimates calculated on the basis of age-specific death rates available in HMD. This finding was discussed at the international conference on mortality in Tallinn, Estonia (2012) with representative from the Human Mortality Database team (Dmitry Jdanov) who acknowledged that HMD uses Kannisto model to fit period mortality after age 80. In the case of cohort data, mortality of U.S. birth cohorts born before 1900 is not subjected to fitting. However, mortality of cohorts born in 1900 and thereafter are fitted using the Kannisto formula.

### 3. Analyzing mortality trajectories using life table aging rate

In 1990 Ansley Coale and Ellen Kisker proposed a method to calculate mortality schedules at advanced ages (Coale and Kisker 1990). This method is based on calculating a measure of mortality change that the authors called the age-specific rate of mortality change with age, or  $k_x$ . This measure is defined in the following way:

$$k_x = \ln(m_x) - \ln(m_{x-1})$$

where  $m_x$  is mortality rate at age  $x$ .

If  $k_x$  is a constant function of age then mortality follows the Gompertz law. If  $k_x$  declines with age then there is a mortality deceleration. Similar measure was also proposed by Horiuchi and Coale (Horiuchi and Coale 1990) and later this measure was called a life table aging rate (LAR) (Horiuchi and Wilmoth 1997).

This approach, based on calculating mortality change, was applied in several earlier studies of mortality trajectories at advanced ages (Horiuchi and Coale 1990; Horiuchi and Wilmoth 1998; Thatcher et al. 1998; Wilmoth 1995). These earlier studies demonstrated that values of  $k_x$  tend to decline after age 80 years suggesting mortality deceleration at advanced ages. Most of these studies used cross-sectional data. As we noted earlier, mortality deceleration may be caused by age misreporting at older ages and previous studies of mortality trajectories have been conducted almost twenty years ago. Thus, it is reasonable to repeat these earlier analyses using data on more recent extinct or almost extinct birth cohorts. In this paper we analyze mortality data for the following four countries: Canada, France, Sweden and the United States. Analyses are based on age-specific cohort death rates for the most recent extinct birth cohorts available in the Human Mortality Database (1894, 1896 and 1898 birth cohorts). Linear regression model was applied to verify if  $k_x$  is declining with age after age 80.

Figure 7 presents age pattern of  $k_x$  for Swedish female 1895-99 cohort. Note that values of  $k_x$  do not show any decline with age up to very advanced ages and after age 100 years random variation of  $k_x$  is very high. Overall, the  $k_x$  age pattern is in a good agreement with the Gompertz law and does not demonstrate any significant decline with age.

To quantify this finding, we conducted linear regression analyses for  $k_x$  as a dependent variable and age as a predictor variable in the age interval 80-100 years. If  $k_x$  does not change with age then the slope coefficient for this regression model should not be significantly different from zero. Table 2 presents regression slope coefficients and corresponding p-values for all 24 studied single-year birth cohorts.

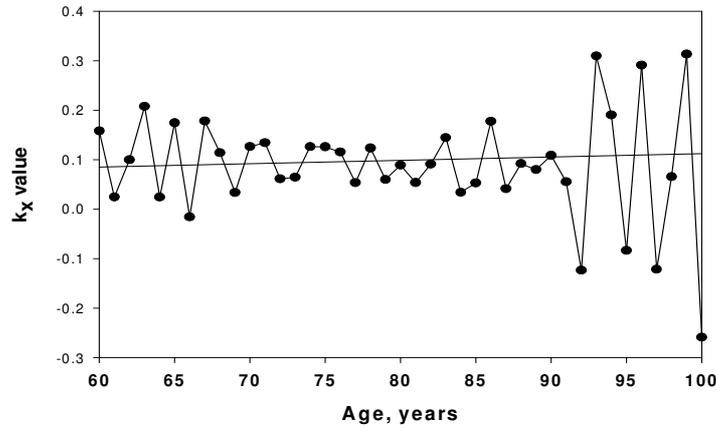


Figure 7. Age-specific rate of mortality change with age,  $k_x$ , Swedish males, 1896 birth cohort.

As follows from Table 1, the slope coefficients for all studied birth cohorts are not significantly different from zero. Thus, the shape of mortality curve after age 80, as measured by the  $k_x$  function, appears to be consistent with the Gompertz model.

Table 1. Slope parameter estimates and corresponding p-values for linear regression model of  $k_x$  as a function of age\*, by country, sex and birth cohort.

Country	Sex	Birth cohort					
		1894		1896		1898	
		slope	p-value	slope	p-value	slope	p-value
Canada	F	-0.00023	0.914	0.00004	0.984	0.00066	0.583
	M	0.00112	0.778	0.00235	0.499	0.00109	0.678
France	F	-0.00070	0.681	-0.00179	0.169	-0.00165	0.181
	M	0.00035	0.907	-0.00048	0.808	0.00207	0.369
Sweden	F	0.00060	0.879	-0.00357	0.240	-0.00044	0.857
	M	0.00191	0.742	-0.00253	0.635	0.00165	0.792
USA	F	0.00016	0.884	0.00009	0.918	0.000006	0.994
	M	0.00006	0.965	0.00007	0.946	0.00048	0.610

\* All regressions were run in the age interval 80-100 years.

These results do not agree with earlier studies that showed linear decline of  $k_x$  after age 80 years (Horiuchi and Wilmoth 1998; Thatcher et al. 1998; Wilmoth 1995). There may be several reasons why these studies found mortality deceleration in advanced ages. In most cases these studies analyzed cross-sectional data combined into 5-year or 10-year time intervals, which may be prone to mortality deceleration due to secular decline of mortality. Some studies analyzed cohort data aggregated into 5-year or 10-year birth cohorts. Such aggregation may result in mortality deceleration if mortality in aggregated single-year birth cohorts is significantly different from each other. To test this hypothesis, we calculated  $k_x$  values using age-specific death rates for 5-year birth cohorts available in HMD. Table 2 shows slope parameters and corresponding p-values for linear regression model of  $k_x$  as a function of

age in the age interval 80-100 years. Note that in the case of aggregated birth cohorts there are some cohorts where  $k_x$  is declining with age.

Table 2. Slope parameter estimates and corresponding p-values for linear regression model, by country, sex and 5-year birth cohort.

Country	Sex	Birth cohort							
		1880-84		1885-89		1890-94		1895-99	
		slope	p-value	slope	p-value	slope	p-value	slope	p-value
Canada	F	-0.00150	0.145	-0.00069	0.372	0.00015	0.851	-0.00002	0.983
	M	-0.00247	0.135	-0.00065	0.642	0.00094	0.306	0.00022	0.850
France	F	-0.00167	0.074	-0.00273	<b>0.047</b>	-0.00191	<b>0.005</b>	-0.00165	<b>0.002</b>
	M	-0.00072	0.818	-0.00082	0.515	-0.00049	0.661	-0.00047	0.412
Sweden	F	-0.00043	0.759	-0.00036	0.749	-0.00122	0.185	-0.00210	0.122
	M	0.00141	0.663	-0.00234	0.309	-0.00127	0.330	-0.00089	0.696
USA	F	-0.00131	0.113	-0.00030	0.654	-0.00027	0.685	0.00004	0.915
	M	-0.00187	<b>0.008</b>	-0.00050	0.417	-0.00039	0.399	0.00002	0.972

Still only 4 out of 32 cohorts show negative slope coefficients suggesting decline in  $k_x$  with age and mortality deceleration. This example demonstrates that even in countries with smaller populations compared to the United States mortality deceleration at advanced ages is rather an exception than a rule. Thus, we may conclude that the analysis of age-specific rates of mortality change for four countries suggests that in most cases mortality deceleration at advanced ages is not supported by existing data. These results are mixed, because for some populations (French women) mortality deceleration does exist for all studied birth cohorts while in other countries (Canada) we do not observe mortality deceleration at all.

More important factor resulting in spurious decline of  $k_x$  is related to use of inappropriate measures of hazard rate at advanced ages. In the previous studies of old-age mortality the values of  $k_x$  were calculated not for one-year but for 5-year age intervals:

$$k_x = \frac{\ln(m_x) - \ln(m_{x-5})}{5}$$

where  $m_x$  represents 5-year mortality rate.

It should be noted that 5-year age interval is very wide for analyzing mortality (and hazard rate) at advanced ages when mortality is particularly high. In this case the assumption about uniform distribution of deaths over the age interval (used for this estimate of hazard rate) does not work. As a result, hazard rate estimates become biased downward resulting in decline of 5-year  $k_x$  values with age. As an example of these effects, we present in Figure 8 the result of numerical simulation using survival data, which follow the Gompertz model with typical parameters (see Appendix A for more detail). In this example we calculate mortality rates for 5-year age interval and then calculate age-specific mortality change function ( $k_x$ ) using formula provided in (Wilmoth 1995). Theoretically we should expect to obtain constant value of  $k_x$  because our simulated data follow the Gompertz law. Instead we get declining pattern of  $k_x$  with age (see Figure 8), which is similar to trajectories reported in the previous publications (Horiuchi and Wilmoth 1998; Wilmoth 1995).

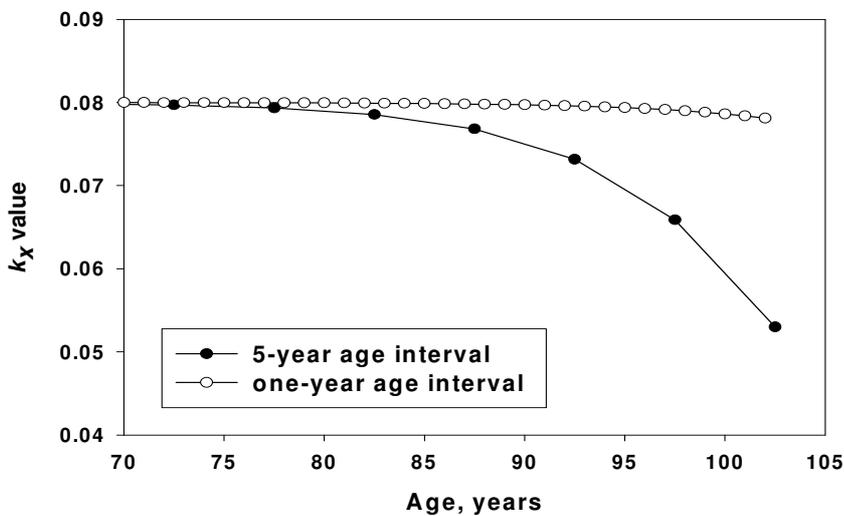


Figure 8. Age-specific rate of mortality change with age,  $k_x$ , by age interval for mortality calculation. Simulated data assuming that hazard rate follows the Gompertz law.

Thus, there is a possibility that the declining pattern of  $k_x$  with age found in earlier studies may be partially related to spurious mortality deceleration caused by using too wide age intervals for hazard rate estimates. It should be noted that in the first publication on this topic the authors used one-year smoothed estimates of mortality rates and found similar declining age pattern for  $k_x$  (Horiuchi and Coale 1990). There are two possible explanations for this phenomenon. One possibility is that the quality of age reporting at advanced ages for the studied populations was not sufficiently high for these time periods (1960s and 1970s). Another possibility is that the authors used cross-sectional data, which are more prone to show mortality deceleration.

We may conclude that the analysis of age patterns of LAR for more recent single-year birth cohorts shows no evidence of mortality deceleration in the age interval 80-100 years.

#### 4. Discussion

This study of ten single-year U.S. birth cohorts, based on two independent data sources, found that mortality deceleration at advanced ages is negligible up to the age of 106 years. Below the age of 107 years and for data of reasonably good quality the Gompertz model fits mortality better than the logistic model (no mortality deceleration). Hazard rate estimates after age 85 obtained from DMF data agree remarkably well with age-specific death rates obtained from the Human Mortality Database. Study of age-specific rate of mortality change used as a measure of mortality deceleration found no mortality deceleration in the age interval 80-100 years for single-year birth cohorts of Canada, France, Sweden and the U.S. Thus, data suggest that mortality after age 80 follows the Gompertz model not only for the United States but also for countries with smaller population (like Sweden)

It should be noted that some researchers already found no mortality deceleration at advanced ages, but did not conduct a systematic study of this phenomenon. For example, Stauffer presents mortality of German cohorts, which shows no mortality deceleration up to age 90 years (Stauffer 2002). Other researchers who found no mortality deceleration at older ages for Canadian cohorts believed that this result is associated with problems of quality in their data (Bourbeau and Desjardins 2006). On the other hand, several systematic studies of mortality at older ages conducted in the 1990s came to a conclusion that mortality does decelerate after age 80 (Horiuchi and Wilmoth 1998; Thatcher 1999; Thatcher et al. 1998; Wilmoth 1995).

There are several reasons why earlier studies, including our own research (Gavrilov 1984; Gavrilov and Gavrilova 1991), reported mortality deceleration and mortality leveling-off at advanced ages (Horiuchi and Wilmoth 1998; Kannisto 1994; Robine and Vaupel 2001; Thatcher 1999; Thatcher et al. 1998). First, mortality deceleration may be caused by age misreporting in death data for older persons (Coale and Kisker 1986; Gavrilov and Gavrilova 2011). Earlier studies, conducted more than ten years ago, used data for older birth cohorts when age reporting was not particularly accurate (Jdanov et al. 2008), even for such developed countries as the U.K (Gallop and Macdonald 2005).

Second, mortality deceleration may be a consequence of data aggregation. Most developed countries have much smaller populations compared to the United States and hence studies of mortality at advanced ages for these countries have to combine together many single-year birth cohorts thereby increasing the heterogeneity of the sample.

Finally, some researchers use inappropriate estimates of hazard rates in their studies of mortality at very high ages when hazard rate is high and changes rapidly. Many studies analyze age-specific probability of death rather than hazard rate (Gallop and Macdonald 2005; Gampe 2010; Modig, Drefahl and Ahlbom 2013; Robine and Vaupel 2001). It is not surprising that probability of death has a tendency of deceleration at advanced ages when mortality is high, taking into account that this mortality indicator has theoretical upper limit equal to one (see Appendix A). For example, a study of mortality among supercentenarians demonstrated that probability of death for this group does not increase with age (Robine and Vaupel 2001). Some authors do not distinguish between probability of death and hazard rate in their calculations (Le Bras 2005).

Mortality rates calculated for wide age intervals also produce biased estimates of hazard rates. For example, use of five-year mortality rates may produce a spurious evidence of mortality deceleration when age-specific rate of mortality change is analyzed (see previous section). Loss of individuals to follow-up in longitudinal study may also be a factor contributing to apparent mortality deceleration at advanced ages (Manton et al. 2008). Appendix A compares accuracy of various estimates of hazard rate at advanced ages and provides a degree of deviation from the correct theoretical values of hazard rate. It appears that age exaggeration, use of inappropriate estimates of hazard rates and perhaps data heterogeneity could lead to downward biases in mortality estimates at older ages reported in previous studies.

The results obtained in this study may be important for actuarial practice, particularly if mortality is analyzed for birth cohorts. These results also may be significant for projections of the size of older population. Recent mass media reports (Financial Times, September 11, 2012; Wall Street Journal, March 2, 2012) revealed that official projections significantly overstated the number of centenarians both in the United States and the UK. Inappropriate assumptions about mortality at advanced ages may be partially responsible for these projection inaccuracies.

## **5. Conclusion**

Few people survive to advanced ages and, in standard mortality tables, it is frequently necessary to compile data over an entire decade to obtain a sufficiently large sample. Our work shows that the observed deceleration in measured mortality rates could result in part from the heterogeneity of the data. The second problem we examined is frequently overlooked by demographers and actuaries: the problem of correct estimation of the instantaneous mortality rate (hazard rate). At the most advanced ages, the rates of death are so high that it is impossible to assume that the number of dying is distributed uniformly within the studied one-year age intervals. As a result, the estimates of mortality rates (or central death rates) are biased downward at advanced ages. And finally, the third problem is related to the fact that elderly people tend to round their ages up, thereby exaggerating their true age. In the United States, this may have impaired the accuracy of mortality rate estimates in the past.

## Appendix A. Hazard rate (mortality force) estimation at advanced ages: A simulation study

A conventional way to obtain estimates of mortality at advanced ages is a construction of demographic life table with probability of death ( $q_x$ ) as one of important life table functions. Although probability of death is a useful indicator for mortality studies, it may not be the most convenient one for studies of mortality at advanced ages. First, the values of  $q_x$  depend on the length of the age interval  $\Delta x$  for which it is calculated. This hampers both analyses and interpretation. Also, by definition  $q_x$  is bounded by unity, which would inevitably produce apparent mortality deceleration when death rates are particularly high.

More useful indicator of mortality at advanced age is instantaneous mortality rate (mortality force) or hazard rate,  $\mu_x$  which is defined as follows:

$$\mu_x = -\frac{dN_x}{N_x dx} = -\frac{d \ln(N_x)}{dx} \approx -\frac{\Delta \ln(N_x)}{\Delta x}$$

where  $N_x$  is a number of living individuals exposed to risk of death at age  $x$ . It follows from the definition of hazard rate that it is equal to the rate of decrease of logarithmic survival function with age. In actuarial practice, hazard rate is often called mortality force as it was done in the original paper by Benjamin Gompertz (Gompertz 1825). Hazard rate does not depend on the length of the age interval (it is measured at the instant of time  $x$ ), has no upper boundary and has a dimension of rate ( $\text{time}^{-1}$ ). It should also be noted that the famous law of mortality, the Gompertz law, was first proposed for fitting the age-specific hazard rate function rather than probability of death (Gompertz 1825).

The empirical estimates of hazard rates are often based on suggestion that age-specific mortality rate or death rate (number of deaths divided by exposure) is a good estimate of theoretical hazard rate. One of the first empirical estimates of hazard rate was proposed by George Sacher (Sacher 1956; Sacher 1966):

$$\mu_x = \frac{1}{\Delta x} [\ln(l_{x - \frac{\Delta x}{2}}) - \ln(l_{x + \frac{\Delta x}{2}})] = \frac{1}{\Delta x} \ln \left[ \frac{l_{x - \frac{\Delta x}{2}}}{l_{x + \frac{\Delta x}{2}}} \right]$$

This estimate is unbiased for slow changes in hazard rate if  $\Delta x \Delta \mu_x \ll 1$  (Sacher, 1966) and was shown to be the maximum likelihood estimate (Gehan and Siddiqui 1973). A simplified version of Sacher estimate (for small age intervals equal to unity) is often used in demographic studies of mortality:  $\mu_x = -\ln(1-q_x)$ . This estimate was initially suggested by Gehan who called it a 'Sacher' estimate (Gehan 1969; Gehan and Siddiqui 1973). It is based on the assumption that hazard rate is constant over age interval and is shifted by one half of a year to younger ages compared to the original Sacher estimate.

Another empirical estimate of hazard rate, often used in life table construction (Klein and Moesberger 1997), is the actuarial estimate, which is calculated in the following way (Kimball 1960):

$$\mu_x = \frac{2q_x}{\Delta x (2 - q_x)} = \frac{2}{\Delta x} \frac{l_{x - \Delta x} - l_x}{l_{x - \Delta x} + l_x}$$

This estimate assumes uniform distribution of deaths over the age interval and is bounded by  $2/\Delta x$ , so this is not the best estimate of hazard rate at extreme old ages when death rates are particularly high (Gavrilov and Gavrilova 1991).

At advanced ages, when death rates are very high, the assumptions about small changes in hazard rate or a constant hazard rate within the age interval become questionable. The same is true for the assumption of uniform distribution of deaths within the age interval.

We conducted a simulation study in order to compare and evaluate the accuracy of different empirical estimates of hazard rate. For this purpose values of survivors at each age were calculated assuming that age-specific hazard rate follows the Gompertz law. The theoretical equation was drawn by integrating the Gompertz formula (Gavrilov, Gavrilova and Nosov 1983):

$$\frac{N_x}{N_0} = \frac{N_{x0}}{N_0} \exp \left[ \left( -\frac{a}{b} \right) (e^{bx} - e^{bx_0}) \right]$$

where  $N_x/N_0$  is the probability of survival to age  $x$ , i.e. the number of hypothetical cohort at age  $x$  divided by its initial number  $N_0$ .  $a$  and  $b$  are the parameters of Gompertz equation (see formula 1). The simulation assumed that the Gompertz law works for the entire age interval and the initial cohort size is equal to  $10^{11}$  individuals. The Gompertz parameters are typical for the U.S. birth cohorts: slope coefficient ( $b$ ) = 0.08 year<sup>-1</sup>;  $a$  = 0.0001 year<sup>-1</sup>. The main focus of this study was on older ages beyond 90 years. Accuracy of various hazard rate estimates (Sacher, Gehan, and actuarial estimates) and probability of death is compared at ages 100 and 110 years.

Figure 9 shows theoretical and empirically estimated values of hazard rate after age 90 using the Sacher estimate of hazard rate and one-year probability of death. Note that the Sacher estimates practically coincide with theoretical mortality trajectory. At the same time, probability of death strongly underestimates mortality after age 100.

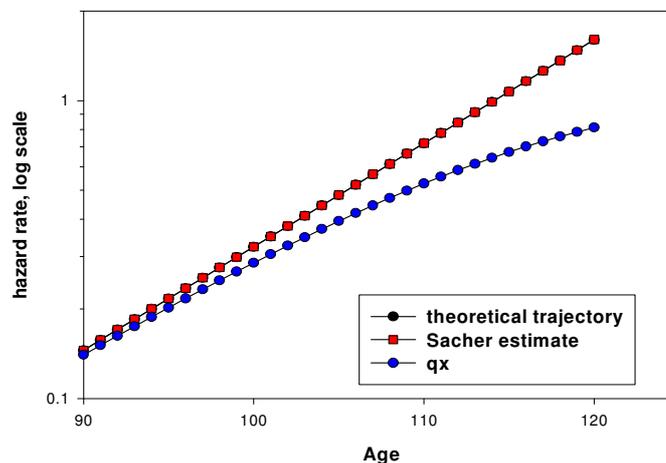


Figure 9. Comparison of Sacher estimate of hazard rate with one-year probability of death using simulated data based on the Gompertz mortality model.

Figure 10 compares theoretical values of hazard rate with empirical estimates of hazard rate using actuarial estimate of hazard rate (equivalent to the age-specific death rate or mortality rate,  $m_x$ ). The actuarial estimates underestimate mortality at later age (110 years) compared to one-year probability of death.

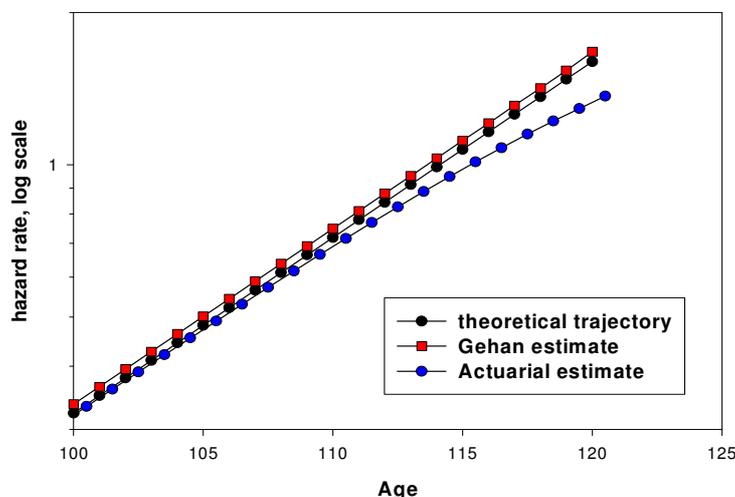


Figure 10. Comparison of actuarial estimate of hazard rate with Gehan estimate of hazard rate using simulated data based on the Gompertz mortality model.

Table 3 summarizes the results of our simulation study. It demonstrates that the Sacher estimate of hazard rate is the best one for use at advanced ages. These results underscore that the choice of proper hazard rate estimate is of paramount importance when mortality trajectory at advanced ages is analyzed. Sacher estimate turned out to be the most accurate estimate for advanced ages while one-year probability of death deviates from hazard rate function after age 85 years. Unfortunately, standard statistical packages do not use the Sacher estimate in their calculations of hazard rate.

Table 3. Comparison of different estimates of hazard rate with theoretical simulated values of hazard rate based on the Gompertz model.

Estimate of hazard rate	Hazard rate estimate at age 100 years	Hazard rate estimate at age 110 years
Probability of death	11.6% understatement	26.7% understatement
Sacher estimate	0.1% overstatement	0.1% overstatement
Actuarial estimate	1.0% understatement	4.5% understatement

Some statistical packages may produce biased estimates of hazard rates at advanced ages. This is the case for the Nelson-Aalen hazard rate estimates provided by sts command of Stata statistical software (StataCorp 2009). In fact, the Nelson-Aalen method was initially proposed for cumulative hazard rate estimation (particularly for right-censored survival data) (Klein and Moesberger 1997). In Stata, hazard rate estimation is made by taking the steps of the Nelson-Aalen cumulative hazard function (Cleves et al. 2008), so that for each observed time of death,  $x_j$  the estimated hazard contribution is:

$$\Delta \hat{H}(x_j) = \hat{H}(x_j) - \hat{H}(x_{j-1})$$

where  $\hat{H}(x)$  is an estimate of cumulative hazard function.

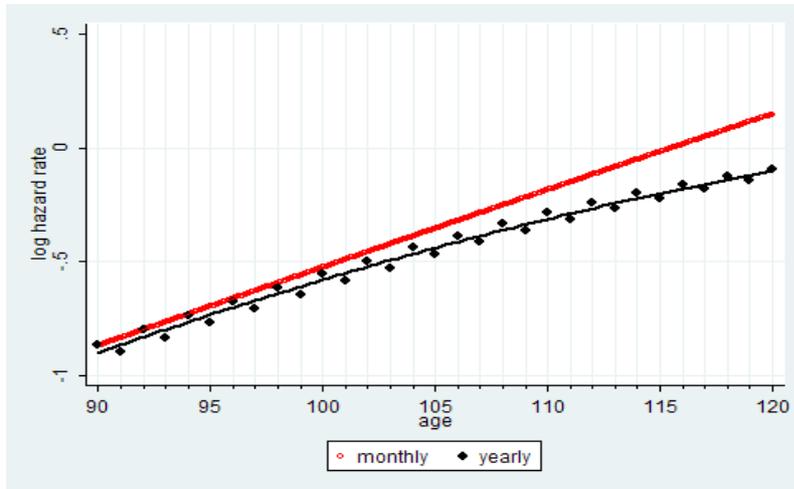


Figure 11. Comparison of monthly and yearly estimates of hazard rate using simulated mortality data based on the Gompertz model

The way of hazard rate estimation conducted in Stata is similar to calculation of life table probability of death (StataCorp 2009), i.e. the number of deaths in the studied age interval is divided by the number alive at the beginning of age interval. At advanced ages when mortality is high and for relatively wide age intervals, number of persons exposed to risk of death in the middle of age interval is substantially lower than the number alive at the beginning of age interval. This would result in downward bias in hazard rate estimates at advanced ages, which is observed when the Nelson-Aalen estimates are applied to yearly age intervals. Simulation studies showed that the bias in hazard rate estimation increases with the increase of the age interval (Kimball 1960). Narrowing the age interval for hazard rate estimation from one year to one month helps to improve the accuracy of hazard rate estimation. For smaller monthly age intervals, the problem described above is not so crucial and the Nelson-Aalen method still can be applied. Figure 11 shows Nelson-Aalen hazard rate estimates produced by sts command of Stata for yearly and monthly age intervals using our simulated Gompertz mortality data. Note that hazard rates estimated for yearly age intervals demonstrate substantial mortality deceleration while hazard rate estimates calculated for monthly age intervals follow the Gompertz model.

Mortality deceleration and even mortality decline at advanced ages may occur when hazard rates are being smoothed using kernel smooth procedure. Figure 12 shows mortality trajectory at advanced ages when Stata kernel smoothing procedure (with default settings) is applied to our simulated Gompertz mortality data.

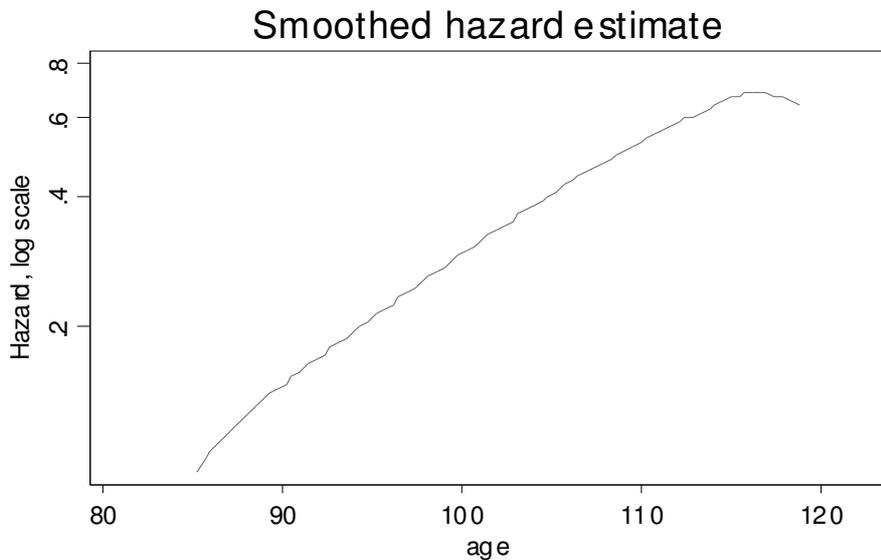


Figure 12. Simulated mortality data based on the Gompertz model after the kernel smooth procedure provided by the Stata statistical package.

Smoothing procedures assume mortality averaging over rather wide age interval (bandwidth), which leads to mortality underestimations at very advanced ages when hazard rates grow very rapidly.

These examples suggest that even standard estimates of hazard rates provided by statistical packages should be treated with caution when mortality is studied at very advanced ages.

### Acknowledgements

Supported by the grant from the National Institute on Aging (NIA grant R01 AG028620).

### References

- Akaike, H. 1974. "A new look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* 19:716-723.
- Beard, R.E. 1959. "Note on some mathematical mortality models." Pp. 302-311 in *The lifespan of animals*, edited by E.W. Wolstenholme and M.O. O'Connor. Boston: Little, Brown.
- . 1971. "Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes." Pp. 57-68 in *Biological Aspects of Demography*, edited by W. Brass. London: Taylor and Francis.
- Bourbeau, R. and B. Desjardins. 2006. "Mortality at Extreme Ages and Data Quality: The Canadian Experience." Pp. 167-185 in *Human Longevity, Individual Life Duration, and the Growth of the Oldest-Old Population*, edited by J.M. Robine, E.M. Crimmins, S. Horiuchi, and Z. Yi. Dordrecht: Springer.
- Caselli, G., L. Pozzi, J.W. Vaupel, L. Deiana, G. Pes, C. Carru, C. Franceschi, and G. Baggio. 2006. "Family clustering in Sardinian longevity: A genealogical approach." *Experimental Gerontology* 41(8):727-736.

- Cleves, M.A., R.G. Gutierrez, W.W. Gould, and Y. Marchenko. 2008. *An Introduction to Survival Analysis Using Stata*. College Station, Texas: Stata Press.
- Coale, A.J. and E.E. Kisker. 1986. "Mortality crossovers - reality or bad data." *Population Studies-a Journal of Demography* 40(3):389-401.
- . 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):1-31.
- Depoid, F. 1973. "Mortality of old people over 85." *Population* 28(4-5):755-792.
- Elo, I.T., S.H. Preston, I. Rosenwaike, M. Hill, and T.P. Cheney. 1996. "Consistency of age reporting on death certificates and social security records among elderly African Americans." *Social Science Research* 25(3):292-307.
- Faig, K. 2001. "Reported deaths of centenarians and near-centenarians in the U.S. Social Security Administration's Death Master File." Presented at The Society of Actuaries "Living to 100 and Beyond International Symposium", Orlando, FL.
- Gallop, A. and A.S. Macdonald. 2005. "Mortality at advanced ages in the United Kingdom." in *Living to 100 and Beyond Monograph [online edition]*. . Shaumburg: Society of Actuaries.
- Gampe, J. 2010. "Human mortality beyond age 110." Pp. 219-230 in *Supercentenarians.*, edited by H. Maier, J. Gampe, B. Jeune, J.-M. Robine, and J.W. Vaupel. Heidelberg, Dordrecht, London, New York: Springer.
- Gavrilov, L.A. 1984. "Does the limit of the life-span really exist?" *Biofizika* 29(5):908-909.
- Gavrilov, L.A. and N.S. Gavrilova. 1991. *The Biology of Life Span: A Quantitative Approach*. New York: Harwood Academic Publisher.
- . 2011. "Mortality measurement at advanced ages: A study of the Social Security Administration Death Master File." *North American Actuarial Journal* 15(3):432-447.
- Gavrilov, L.A., N.S. Gavrilova, and V.N. Nosov. 1983. "Human life span stopped increasing: why?" *Gerontology* 29(3):176-180.
- Gehan, E.A. 1969. "Estimating survival functions from life table." *Journal of Chronic Diseases* 21(9-10):629-8.
- Gehan, E.A. and M.M. Siddiqui. 1973. "Simple regression methods for survival time studies." *Journal of the American Statistical Association* 68(344):848-856.
- Gompertz, B. 1825. "On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies." *Philos. Trans. Roy. Soc. London A* 115:513-585.
- Greenwood, M. and J.O. Irwin. 1939. "The biostatistics of senility." *Human Biology* 11:1-23.
- Hill, M.E., S.H. Preston, I.T. Elo, and I. Rosenwaike. 1997. "Age-linked institutions and age reporting among older African Americans." *Social Forces* 75(3):1007-1030.
- Horiuchi, S. 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(4):245-325.
- Horiuchi, S. and J.R. Wilmoth. 1997. "Age patterns of the life table aging rate for major causes of death in Japan, 1951-1990." *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* 52(1):B67-B77.
- . 1998. "Deceleration in the age pattern of mortality at older ages." *Demography* 35:391-412.
- Jdanov, D.A., D. Jasilionis, E.L. Soroko, R. Rau, and J.W. Vaupel. 2008. *Beyond the Kannisto-Thatcher Database on Old Age Mortality: An Assessment of Data Quality at Advanced Ages*. MPDIR Working Paper WP 2008-013. . Rostock, Germany: MPDIR.
- Kannisto, V. 1988. "On the survival of centenarians and the span of life." *Population Studies-a Journal of Demography* 42:389-406.
- . 1994. *Development of Oldest-Old Mortality, 1950-1990: Evidence from 28 Developed Countries*. Odense: Odense University Press.
- Kestenbaum, B. 1992. "A Description of the Extreme Aged Population Based on Improved Medicare Enrollment Data." *Demography* 29(4):565-580.

- Kimball, A.W. 1960. "Estimation of mortality intensities in animal experiments." *Biometrics* 16(4):505-521.
- Klein, J.P. and M.L. Moesberger. 1997. *Survival analysis techniques for censored and truncated data*. New York: Springer-Verlag.
- Le Bras, H. 2005. "Mortality tempo versus removal of causes of mortality: opposite views leading to different estimations of life expectancy." *Demographic Research* 13:615-640.
- Lee, D.G. and J. Russell. 2004. "Panel Data Analysis of Factors Affecting As-Built Roughness of Asphaltic Concrete Pavements." *Journal of Transportation Engineering* 130(4):479-485.
- Manton, K.G., I. Akushevich, and A. Kulminski. 2008. "Human mortality at extreme ages: Data from the NLTCS and linked Medicare records." *Mathematical Population Studies* 15(3):137-159.
- Modig, K., S. Drefahl, and A. Ahlbom. 2013. "Limitless longevity: Comment on the Contribution of rectangularization to the secular increase of life expectancy." *International Journal of Epidemiology*:doi: 10.1093/ije/dyt1035.
- Muller, H.G., J.L. Wang, and W.B. Capra. 1997. "From lifetables to hazard rates: The transformation approach." *Biometrika* 84(4):881-892.
- Preston, S.H., I.T. Elo, and Q. Stewart. 1999. "Effects of age misreporting on mortality estimates at older ages." *Population Studies-a Journal of Demography* 53(2):165-177.
- ResearchGate. 2013. *Akaike information criterion*. URL: [http://www.researchgate.net/post/Akaike\\_information\\_criterion](http://www.researchgate.net/post/Akaike_information_criterion). Accessed October 6, 2013.
- Robine, J.M. and J.W. Vaupel. 2001. "Supercentenarians: slower ageing individuals or senile elderly?" *Experimental Gerontology* 36(4-6):915-930.
- Sacher, G.A. 1956. "On the statistical nature of mortality, with especial reference to chronic radiation mortality." *Radiology* 67:250-257.
- . 1966. "The Gompertz transformation in the study of the injury-mortality relationship: Application to late radiation effects and ageing." Pp. 411-441 in *Radiation and Aging*, edited by P.J. Lindop and G.A. Sacher. London: Taylor and Francis.
- Sesso, H.D., R.S. Paffenbarger, and I.M. Lee. 2000. "Comparison of National Death Index and World Wide Web death searches." *American Journal of Epidemiology* 152(2):107-111.
- StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.
- Stauffer, D. 2002. "Simple tools for forecasts of population ageing in developed countries based on extrapolations of human mortality, fertility and migration." *Experimental Gerontology* 37(8-9):1131-1136.
- Thatcher, A.R. 1999. "The long-term pattern of adult mortality and the highest attained age." *Journal of the Royal Statistical Society Series a-Statistics in Society* 162:5-30.
- Thatcher, A.R., V. Kannisto, and J. Vaupel. 1998. *The Force of Mortality at Ages 80 to 120*. Odense: Odense University Press.
- Vaupel, J.W., K.G. Manton, and E. Stallard. 1979. "Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography* 16(3):439-454.
- Vincent, P. 1951. "La mortalite des Viellards." *Population* 6:181-204.
- Willcox, D.C., B.J. Willcox, Q. He, N.C. Wang, and M. Suzuki. 2008. "They really are that old: A validation study of centenarian prevalence in Okinawa." *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* 63(4):338-349.
- Wilmoth, J.R. 1995. "Are mortality-rates falling at extremely high ages - an investigation based on a model proposed by Coale and Kisker." *Population Studies-a Journal of Demography* 49(2):281-295.
- Wilmoth, J.R., K.F. Andreev, D.A. Jdanov, and D.A. Glej. 2007. *Methods Protocol for the Human Mortality Database. Version 5*. Rostock, Germany.
- Young, R.D., B. Desjardins, K. McLaughlin, M. Poulain, and T.T. Perls. 2010. "Typologies of extreme longevity myths." *Current gerontology and geriatrics research* 2010:423087.