

# Mortality Measurement and Modeling Beyond Age 100

Natalia S. Gavrilova  
Leonid A. Gavrilov

Presented at the Living to 100 Symposium  
Orlando, Fla.  
January 5–7, 2011

Copyright 2011 by the Society of Actuaries.

All rights reserved by the Society of Actuaries. Permission is granted to make brief excerpts for a published review. Permission is also granted to make limited numbers of copies of items in this monograph for personal, internal, classroom or other instructional use, on condition that the foregoing copyright notice is used so as to give reasonable notice of the Society's copyright. This consent for free limited copying without prior consent of the Society does not extend to making copies for general distribution, for advertising or promotional purposes, for inclusion in new collective works or for resale.

The growing number of people living beyond age 100 emphasizes the need for accurate measurement and modeling of mortality at advanced ages. However, mortality measurement at extremely old ages poses serious challenges to researchers: age reporting at extreme ages may be inaccurate; small numbers of people at advanced ages requires pooling data for people belonging to different birth cohorts, which results in data heterogeneity; and, finally, a standard approach based on annual mortality estimates may not be applicable to extremely high and rapidly changing risk of death at advanced ages.

The study uses data from the Social Security Administration Death Master File (DMF) to conduct mortality estimates for more homogeneous single-year birth cohorts with hazard rates estimated for narrow (monthly) age intervals. Mortality after age 88 was estimated for 1881-95 single-year extinct birth cohorts. Data quality control checks demonstrated that deceleration of mortality in later life is more expressed for data with lower quality. Additional quality checks using female-to-male ratio at advanced ages identified age 107 as the upper limit for acceptably good level of age reporting. Model comparison using Bayesian information criterion (BIC) showed that in the age interval 88 to 106 years and for data with reasonably good quality (U.S. Northern states), the Gompertz model shows better fitting of hazard rates than the logistic model. Hazard rate estimates obtained on the basis of DMF data agree well with estimates calculated on the basis of the 1900 actuarial cohort life table.

This study shows that mortality deceleration in humans observed at advanced ages may be an artifact caused by age exaggeration, data heterogeneity or use of improper estimates of hazard rate. The phenomenon of mortality deceleration at older ages vanishes for datasets of higher data quality.

## Introduction

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. It is now considered an established fact that mortality at advanced ages has a tendency to deviate from the Gompertz law (Gavrilov and Gavrilova 1991), so that the logistic model often is used to fit human mortality (Horiuchi and Wilmoth 1998). The estimates of mortality force at extreme ages are difficult because of small numbers of survivors to these ages in most countries. Data for extremely long-lived individuals are scarce and subjected to age exaggeration. Traditional demographic estimates of mortality based on period data encounter well-known denominator problem. More accurate estimates are obtained using the method of extinct generations (Vincent 1951). To obtain good quality estimates of mortality at advanced ages, researches are forced to pool data for several calendar periods. 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. Mortality deceleration observed in these data could be at least partially related to population heterogeneity. In addition to that, many assumptions about distribution of deaths in the age/time interval used in mortality estimation are not valid for extreme old ages when mortality is particularly high.

Thus, the estimation of hazard rates at extremely old ages poses several serious challenges to researchers:

1. The observed mortality deceleration may be at least partially an artifact of mixing different birth cohorts with different mortality (heterogeneity effect);
2. Standard assumptions of hazard rate estimates may be invalid when risk of death is extremely high at old ages; and
3. Ages of very old people may be exaggerated.

One way of obtaining estimates of mortality at extreme ages is to pool international records of people surviving to extreme ages with subsequent efforts of strict age validation (Robine and Vaupel 2001; Robine et al. 2005). This approach helps to resolve the third problem mentioned above but does not allow researchers to resolve the first two problems because of inevitable data heterogeneity when data for people belonging to different birth cohorts and countries are pooled together. In this paper, we propose an alternative approach, which allows us to resolve partially the first two problems. This approach is based on using data from the 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. Availability of month-of-birth and month-of-death information provides a unique opportunity to obtain hazard rate estimates for every month of age. Possible ways of resolving the third problem of hazard rate estimation are also elaborated.

## Hazard Rate (Mortality Force) Estimation at Advanced Ages

A conventional way to obtain estimates of mortality at advanced ages is a construction of a demographic life table with probability of death ( $q_x$ ) as one important life table function. 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. For example, if one-year probability of death follows the Gompertz law of

mortality, probability of death calculated for other age intervals does not follow this law (Le Bras 1976; Gavrilov and Gavrilova 1991). Thus, it turns out that the shape of age trajectory for  $q_x$  depends on the arbitrary choice of age interval. Also, by definition,  $q_x$  is bounded by unity, which would inevitably produce apparent mortality deceleration when death rates are particularly high.

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

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

where  $N_x$  is a number of living individuals exposed to risk of death at age  $x$ . This definition refers to a cohort closed to migration when  $dN_x$  does not involve migration. 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 famous 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 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):

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

This estimate is unbiased for slow changes in hazard rate if  $\Delta x \Delta \bar{\mu}_x \ll 1$  (Sacher 1966) and was shown to be the maximum likelihood estimate. A simplified version of Sacher estimate (for small age intervals equal to unity) is often used in biodemographic 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 intervals 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):

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

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 very old ages (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. Simulation studies showed that the bias in hazard rate estimation increases with the increase of the 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.

### **Social Security Administration Death Master File as a Source of Mortality Data for Advanced Ages**

Social Security Administration Death Master File (DMF) is a publicly available data source that allows a search for individuals using various search criteria including birth date, death date, first and last names, social security number and place of last residence. This resource covers deaths that occurred between 1937 and 2010 (Faig 2001) and captures about 95 percent of deaths recorded by the National Death Index (Sesso et al. 2000). According to other estimates, DMF covers about 90 percent of all deaths for which death certificates are issued (see Faig 2001) and about 92 to 96 percent of deaths for people older than 65 (Hill, Rosenwaike 2001).

Social Security Administration DMF was used in our study of age-related mortality dynamics after ages 85 to 90. The advantage of this data source is that some birth cohorts covered by DMF could be studied by the method of extinct generations (Vincent 1951; Kannisto 1988; Kannisto 1994). Availability of month-of-birth and month-of-death information provides a unique opportunity to obtain hazard rate estimates for every month of age, which is important given extremely high mortality after age 100. Table 1 presents information available in the Social Security Administration Death Master File, which is publicly available at [Rootsweb.com](http://Rootsweb.com).

The information from the DMF was collected for individuals who lived 88 years or more and died before 2011. The DMF database is unique because it represents mortality experience for very large birth cohorts of the oldest-old people. In this study, mortality measurements were made for cohorts, which are more homogeneous in respect to the year of birth and historical life course experiences. Despite certain limitations, this data source allows researchers to obtain detailed estimates of mortality at advanced ages. We already used this data resource for centenarians' age validation in the study of centenarian family histories (Gavrilova and Gavrilov 2007). This data resource is also useful in mortality estimates for several extinct or almost extinct birth cohorts in the United States.

1.	First and last names, and social security number (SSN) date, month and year of birth
2.	Month and year of death
3.	State of the SSN issuance
4.	Town, county, state and ZIP code of the last residence
5.	Death date verification code

## Hazard Rate Estimates at Advanced Ages Using Data From the Social Security Death Master File

In this study, we collected information from the DMF publicly available at Rootsweb.com. The total number of records collected from this resource was about 9 million with more than 900,000 records belonging to people who lived 100 years or more. Several birth cohorts (those born before 1895) may be considered extinct or almost extinct, so it is possible to apply the method of extinct generations (Vincent 1951) and estimate age-related mortality dynamics at very advanced ages—up to 115 years.

We obtained data for people who died before 2011 and were born in 1875-95. People born in these years and alive in 2010 should survive to at least 115 years, which can be considered as very unlikely event. Thus, the 1875-95 birth cohorts in this sample may be considered extinct or almost extinct. Assuming that the number of living people belonging to these birth cohorts in 2010 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). In the first stage of our analyses, we calculated an individual life span in completed months:

**Lifespan in Months = (Death Year – Birth Year) x 12 + Death Month – Birth Month**

Having this information, it is possible to estimate hazard rates at each month of age by standard methods of survival analysis. All calculations were done using the Stata statistical software, Release 11 (StataCorp 2009). This software calculates nonparametric estimates of major survival functions including the Nelson-Aalen estimator of hazard rate (force of mortality). Note that hazard rate,  $\mu(x)$ , in contrast to probability of death,  $q(x)$ , has a dimension of time frequency because of the time interval in the denominator (reciprocal time,  $\text{time}^{-1}$ ). Thus the values of hazard rates depend on the chosen units of time measurement ( $\text{day}^{-1}$ ,  $\text{month}^{-1}$  or  $\text{year}^{-1}$ ). In this study, survival times were measured in months, so the estimates of hazard rates initially had a dimension of  $\text{month}^{-1}$ . For the purpose of comparability with other published studies, which typically use the  $\text{year}^{-1}$  time scale, we transformed the monthly hazard rates to the more conventional units of  $\text{year}^{-1}$ , 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 this at extreme old ages in some rare cases as will be described later). In this paper, we focus our analyses on 1881-95 birth cohorts because data quality for cohorts born before 1881 is not particularly good.

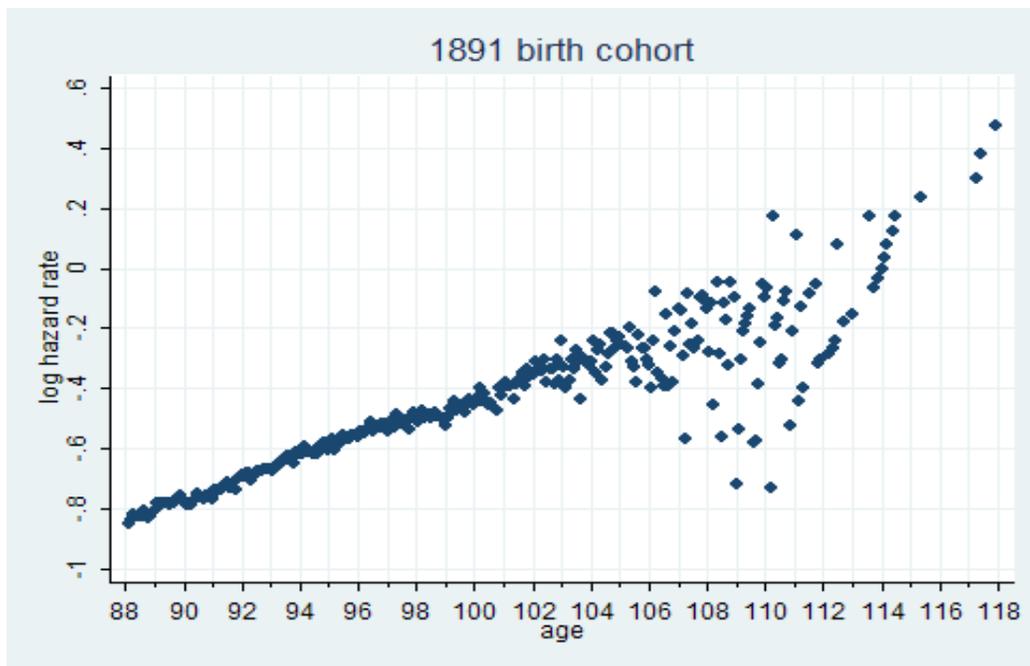
In our study, we used the Nelson-Aalen hazard rate estimates calculated by Stata statistical software (StataCorp 2009). In fact, the Nelson-Aalen method was initially proposed for cumulative hazard rate estimation (particularly for right-censored survival data) (Klein and Moesberger 1997). The way of 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 the age interval. At advanced ages when mortality is high and for relatively large age intervals, the number of people exposed to risk of death in the middle of the age interval is substantially lower than the number alive at the beginning of the age interval. This would result in downward bias in hazard rate estimates at advanced ages, which we observed in our study when the Nelson-Aalen estimates (used in Stata package) were applied to yearly age intervals (data not shown).

However, for smaller monthly age intervals, the problem described above is not so crucial and the Nelson-Aalen method still can be applied (as we did in our study).

Results of the age-specific hazard rate estimates for the 1891 birth cohort (in semi-log scale) is presented in Figure 1. Note that mortality in this large single-year birth cohort follows the Gompertz law up to very advanced ages (linear increase of logarithm of mortality force with age). Previous studies of mortality at advanced ages used aggregated data combining several birth cohorts with different mortality levels and this aggregation may lead to apparent mortality deceleration and subsequent leveling-off as suggested by the heterogeneity model of mortality proposed by British actuary Eric Beard (Beard 1959; Beard 1971).

Note that hazard rate demonstrates steady exponential growth up to the age 105. After age 105, mortality trajectories show a small tendency for deceleration with age. Estimates of hazard rates for several other single-year birth cohorts (1886-95) revealed similar age patterns for mortality at advanced ages. Using single-year birth cohort data from the SSA 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. 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. Thus, we may suggest that mortality deceleration after age 105 among DMF cohorts may be caused by poor data quality (age exaggeration) at very advanced ages. If this hypothesis is correct, then mortality deceleration at advanced ages should be less expressed for data with presumably better quality. To test this hypothesis, we conducted a data quality study for DMF birth cohorts.

**Figure 1**  
**Data From the Social Security Administration Death Master File**



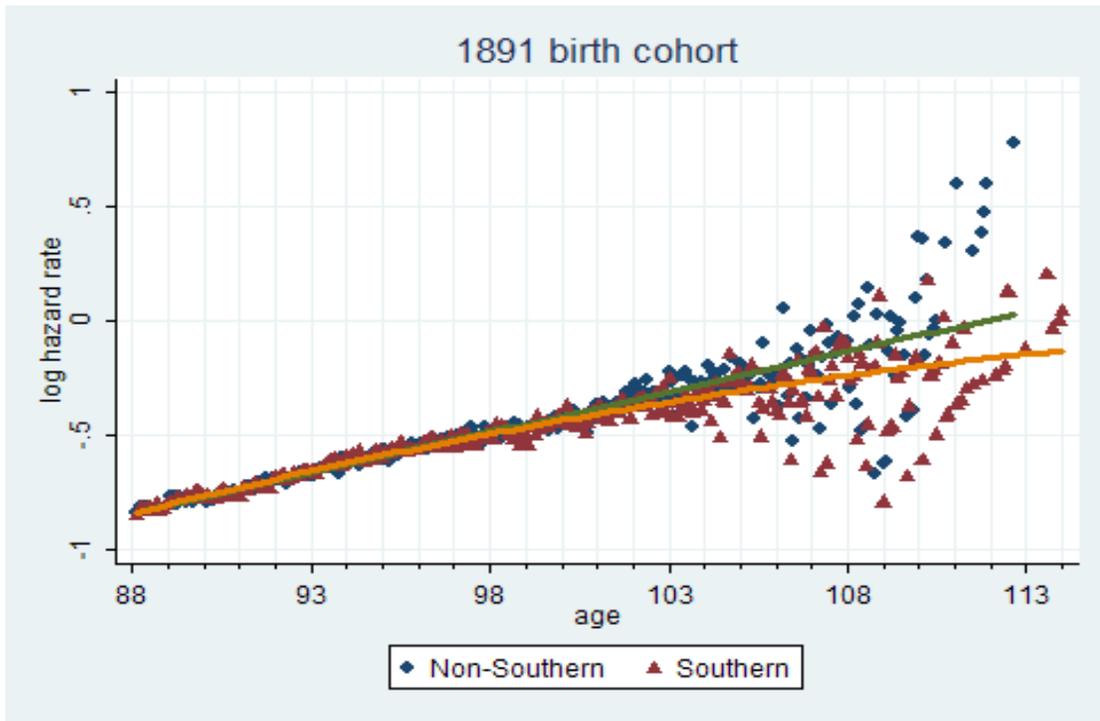
## Measures to Improve Data Quality

Mortality deceleration and even decline of mortality is observed usually for data of low quality. On the other hand, improvement of data quality results in straighter mortality trajectory in semi-log scale (Kestenbaum and Ferguson 2001). Thus, we may suggest that mortality estimates at advanced ages should be lower in populations with less accurate age reporting compared to populations with more accurate age reporting. To test this hypothesis, we compared mortality trajectories at advanced ages for populations with presumably different quality of age reporting.

Study of age validation among individuals age 110 or more (Rosenwaike and Stone 2003) demonstrated that age reporting among supercentenarians in the Social Security Administration database is rather accurate with the exception of people born in the Southern states. To compare populations with presumably different data quality, we split DMF records into two groups. In the first group, we included records for those people who applied for a social security number in the Southeast states (Arkansas, Alabama, Georgia, Mississippi, Louisiana, Tennessee, Florida, Kentucky, South Carolina, North Carolina, Virginia and West Virginia), Southwest states (Arizona, New Mexico, Texas and Oklahoma), Puerto Rico and Hawaii. We also added to this group all records for people who applied for a SSN in New York and California because of very high proportion of immigrants (with unknown quality of age reporting) residing in these states. In the second group with presumably better data quality, we included people applying for a social security number in all other states as well as retired railroad workers.

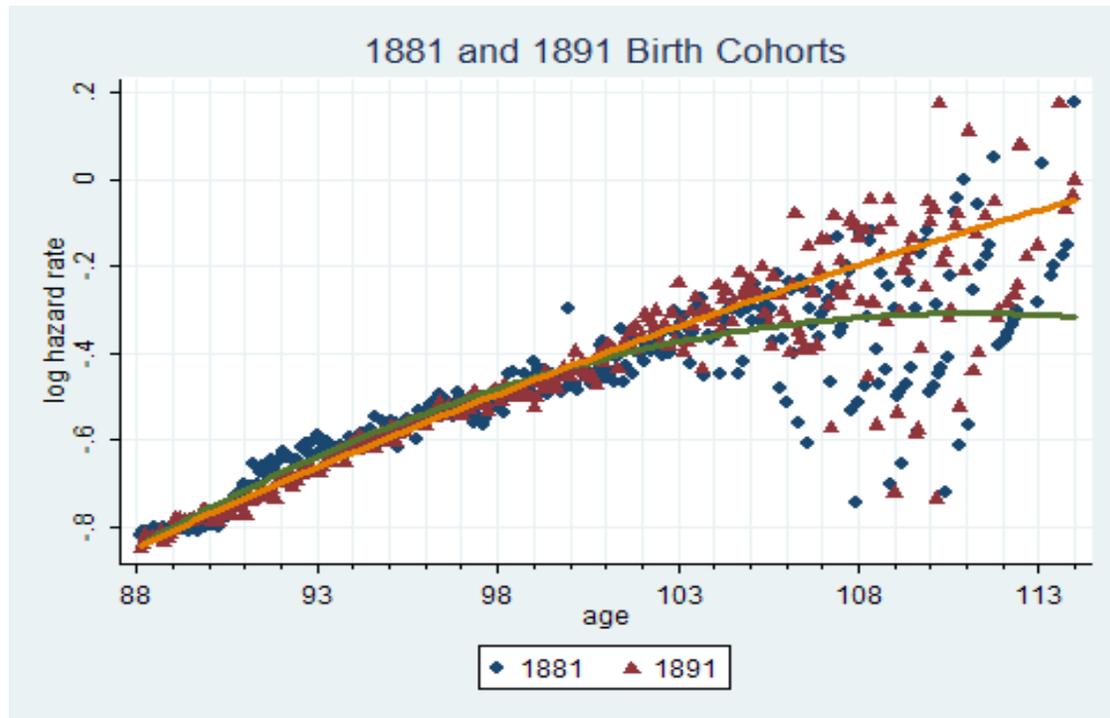
Estimates of age-specific mortality for these two groups demonstrated straighter trajectory of mortality in semi-log scale for population of non-Southern states with presumably better quality of age reporting (verified using quadratic fit of mortality data). Figure 2 shows age-trajectories of mortality for people born in 1891 who applied for a SSN either in Southern or non-Southern states. Note that the line corresponding to quadratic fit of the logarithm of mortality with age is straighter for the non-Southern group compared to the line corresponding to mortality of the Southern group.

**Figure 2**  
**Mortality Data With Presumably Different Quality**  
**Hazard Rate (Per Year) for the 1891 Birth Cohort in a Log Scale;**  
**Comparison of Less Reliable (Southern Group) and More Reliable**  
**(Non-Southern Group) Data; Solid Lines Show Fit With Quadratic Function;**  
**Data From the Social Security Administration Death Master File**



Similar phenomenon is observed when mortality data for an earlier birth cohort are compared to mortality data for a later birth cohort. Records for later-born people are supposed to be of better quality due to improvement of age reporting over time. Taking into account that age reporting is improving, we may suggest that mortality for later birth cohorts follows the Gompertz law more closely compared to the mortality of earlier birth cohorts with presumably lower data quality. This suggestion is confirmed by comparing 1881 and 1891 birth cohorts (Figure 3). The more recent 1891 birth cohort demonstrates straighter trajectory and lower statistical noise after age 105 than the older 1881 set (see Figure 3). Thus, we may expect that cohorts born after 1891 would demonstrate even better fit by the Gompertz model than the older ones because of improved quality of age reporting.

**Figure 4**  
**Mortality Data With Presumably Different Quality; Comparison of**  
**Less Reliable (1881 Birth Cohort) and More Reliable (1891 Birth Cohort) Data;**  
**Solid Lines Show Fit With Quadratic Function; Data From the Social Security**  
**Administration Death Master File**



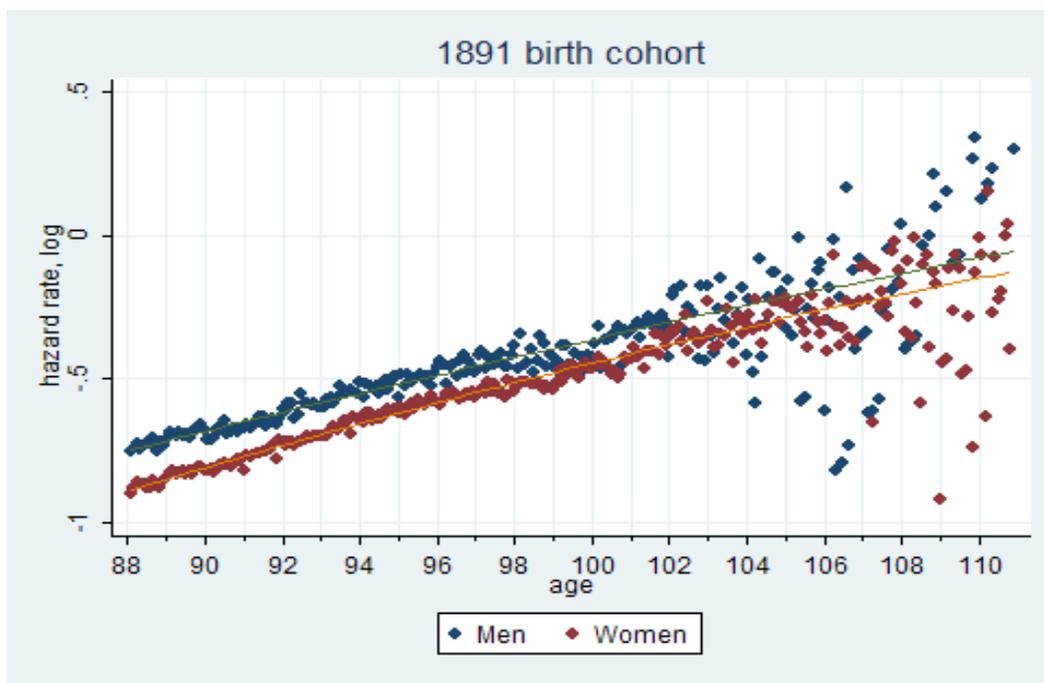
Based on the results presented above, we may conclude that the better the quality of mortality data at advanced ages, the closer mortality trajectories are to the Gompertz function. However, we need to admit that the data at very old ages are noisy and have quality problems even for the non-Southern group and the most recent birth cohorts. Thus, our next step in data quality control is to identify the age interval with reasonably good data quality and then to compare competing mortality models in this age interval.

One approach to test the data quality at advanced ages is to calculate the female-to-male ratio at advanced ages. Taking into account that female mortality is always lower than male mortality at advanced ages, it is reasonable to expect that the female-to-male ratio should continuously increase with age. On the other hand, 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).

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 born during the 1900s in the United States. From the lists of male and female names, we removed names consisting of initials and potentially unisex names (such as Jessie or Lonnie). It is interesting to note that the

SSA male list contains some obviously female names (Mary, Elizabeth) and the same problem was observed for the female list, which indicates that the SSA data apparently contain many sex misidentifications. These female names were removed from the male list and the same procedure was done for the female list. Additional male or female names found in the DMF were also added to our sex identification name lists. Using the final lists of male and female first names, we applied the sex identification procedure to the sample of 1879-95 birth cohorts. Eventually, we were able to identify sex for 91 to 93 percent of records for people in our sample. The remaining 7 to 9 percent of people 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 is unlikely. This data sample of 774,926 individuals with known sex (out of 846,982 individuals) was used for identifying the age interval of reasonably good data quality using information on the female-to-male ratio at advanced ages.

**Figure 5**  
**Age-Specific Hazard Rates (Log Scale) for Men and Women Born in 1891;**  
**Solid Lines Show Fit With Quadratic Function; Data From the Social Security**  
**Administration Death Master File**



The result of hazard rate estimation for men and women born in 1891 is presented in Figure 5. Note that male mortality continues to exceed female mortality up to very advanced ages and this mortality differential narrows very slowly with age. At age 110, the number of remaining males (nine) and females (61) is too small for accurate estimates of hazard rate after this age.

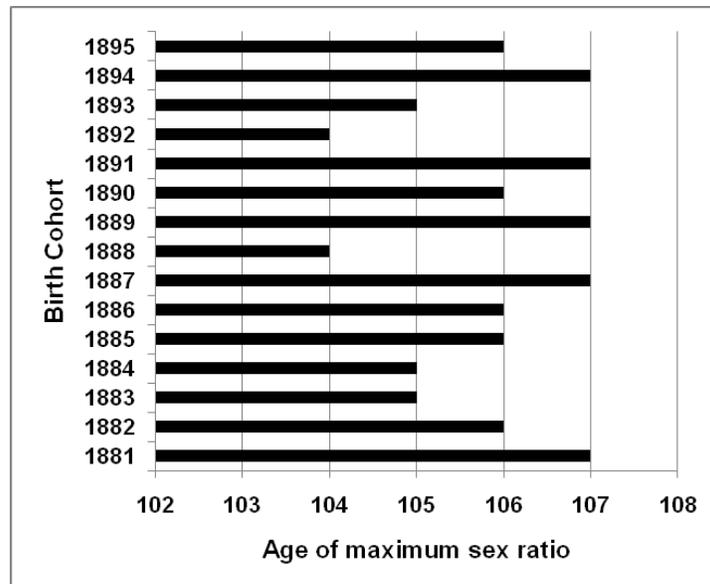
We calculated female-to-male ratio after age 95 for 1881-95 U.S. birth cohorts from the SSA DMF. Figure 6 demonstrates the age-specific dependency of this ratio for the combined sample of 1887-92 birth cohorts (these single-year birth cohorts have similar levels of mortality).

Note that the female-to-male ratio is growing steadily with age up to ages 106-107. After this age, the female-to-male ratio starts to decrease, indicating declining quality of age reporting. Thus, the estimates of hazard rates obtained from the SSA DMF are of acceptable quality up to the age of 106. Figure 7 shows the maximum age or tipping point when the female-to-male ratio starts to decline for different birth cohorts. It demonstrates that this age varies from 104 to 107, indicating the upper limit for data of reasonably good quality in terms of age reporting.

**Figure 6**  
**Observed Female-to-Male Ratio at Advanced Ages for Combined 1887-92 Birth Cohort;**  
**If Data are of Good Quality, Then This Ratio Should Grow With Age**



**Figure 7**  
**Age of Maximum Female-to-Male Ratio, by Birth Cohort**



Thus, our study of female-to-male ratio for different birth cohorts indicates that age reporting has reasonably good quality up to the age of about 106. For this reason, we used age interval 88 to 106 years for mortality modeling.

### Modeling Mortality at Advanced Ages

The next step of our study was to compare two competing models of mortality at advanced ages—the Gompertz and the logistic models—using data of reasonably good quality. Study of data quality control at advanced ages described earlier suggests that age reporting among the oldest-old in the United States is good until the age of 106. It means that comparing mortality models beyond this age is not feasible because of poor quality of mortality data. For this reason, we used a subsample of deaths for people who applied for SSNs in the Northern states (described above) in 1886-95 because these data have relatively good quality. We applied Gompertz and logistic (Kannisto) models (Thatcher et al. 1998) to mortality modeling in the age interval 88 to 106 years using nonlinear regression method for parameter estimation. Calculations were performed using Stata statistical software, Release 11.

**Figure 8**  
**Mortality Fitting Using Gompertz and Logistic Models;**  
**1891 Birth Cohort of People Who Applied for SSNs in the Northern States;**  
**Data From the Social Security Administration Death Master File**

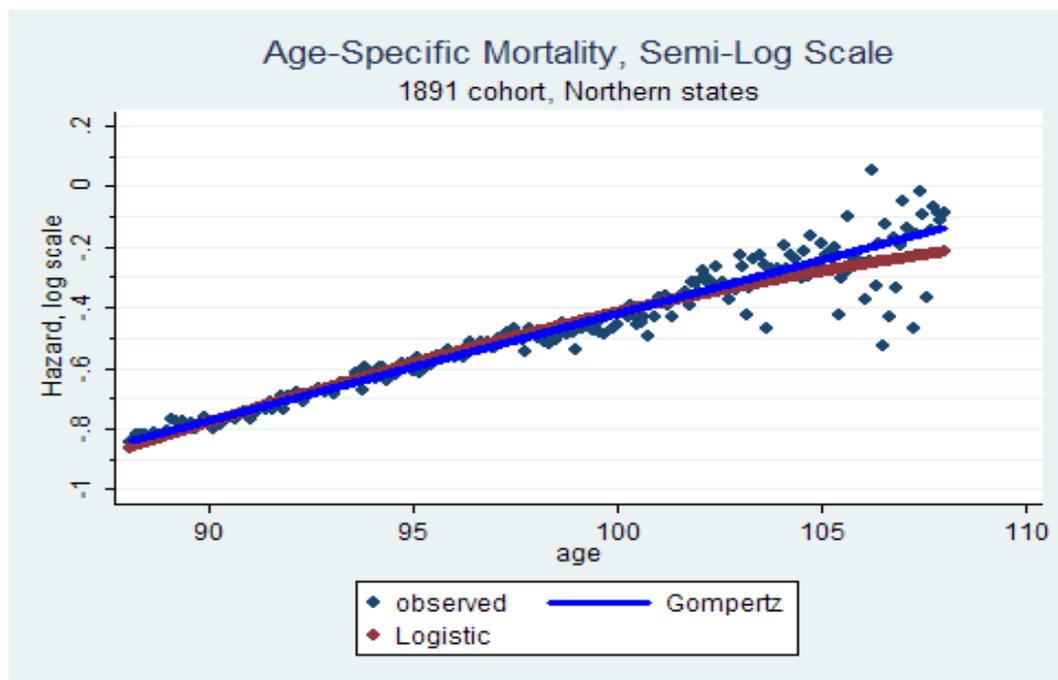


Figure 8 shows mortality fitting after age 88 for Gompertz and Kannisto models using data for Northern states of the 1891 birth cohort. Goodness-of-fit for each model was estimated using the Bayesian information criterion (BIC). Table 2 shows values of BIC for both Gompertz and logistic model for 10 studied birth cohorts. Note that in eight out of 10 cases (studied birth cohorts), the Gompertz model demonstrates better fit (lower BIC) than the logistic model for age interval 88 to 106.

**TABLE 2**  
**Comparison of Goodness-of-fit (Bayesian Information Criterion, BIC)**  
**for Gompertz and Logistic Models of Mortality in the**  
**Age Interval 88 to 106 for 10 U.S. Birth Cohorts<sup>†</sup>**

Birth Cohort Size at 88 Years	Bayesian Information Criterion (BIC)		
		Gompertz Model <sup>††</sup>	Logistic Model
1886	111,657	<b>-594776.2</b>	-588049.5
1887	114,469	<b>-625303.0</b>	-618721.4
1888	128,768	-709620.7	-712575.5
1889	131,778	-710871.1	-715356.6
1890	135,393	<b>-724731.0</b>	-722939.6
1891	143,138	<b>-767138.3</b>	-739727.6
1892	152,058	<b>-831555.3</b>	-810951.8
1893	156,189	<b>-890022.6</b>	-862135.9
1894	160,835	<b>-946219.0</b>	-905787.1
1895	165,294	<b>-921650.3</b>	-863246.6

<sup>†</sup> Data of enhanced accuracy for individuals who applied to SSNs in the Northern states (see explanation in the text).

<sup>††</sup> Cases when the Gompertz model fits data better than the logistic model are highlighted in bold.

At this moment, we cannot make a conclusion that the Gompertz model fits mortality data after age 106 better than the logistic model because of low quality of age reporting, which results in potential underestimation of mortality after age 106. At the same time, it is reasonable to conclude that the Gompertz model fits mortality data well until age 106. Taking into account that survival beyond age 106 is a rather rare event, it would be reasonable to suggest the use of Gompertz model for closing cohort life tables in actuarial practice.

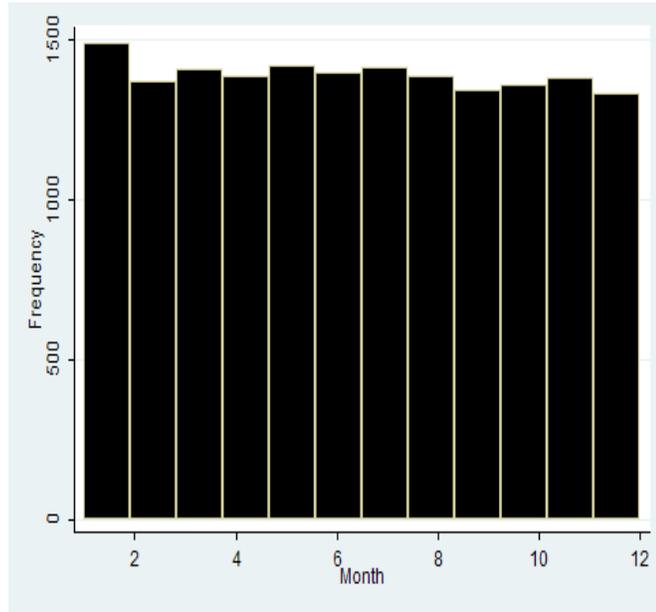
### Comparison of DMF Mortality Data with Actuarial Life Tables

Due to natural limitations of the DMF data (poor completeness of death registration before the 1970s), we were not able to estimate mortality before age 85 to 88. There is also a question of whether our estimates of Gompertz parameters (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). In this study, we compared our empirical estimates of cohort mortality at advanced ages based on DMF data with the 1900 actuarial cohort life table (Bell et al. 1992). Unfortunately, there are no reliable U.S. cohort life tables for cohorts older than 1900, so we compared our mortality data for earlier birth cohorts to later 1900 birth cohorts. It should be noted that the adult U.S. population did not experience substantial mortality changes between 1890 and 1900. According to the estimates made by historical demographers, life expectancy at age 40 in the United States in 1890 (27.61 years) was practically the same as in 1900 (27.52 years) (Haines 1998). Thus, we believe that mortality at advanced ages for the 1900 birth cohort in the actuarial study and the 1894 birth cohort we used for comparison should not be substantially different.

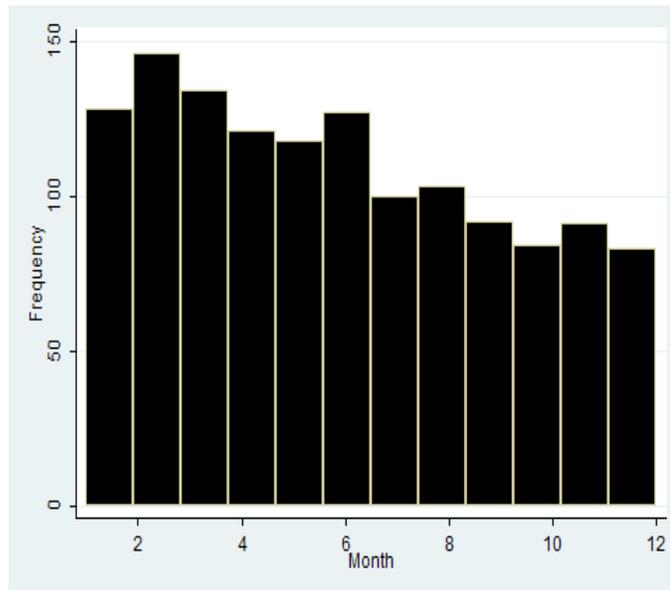
**Figure 9**  
**Deaths at Extreme Old Ages Are Not Distributed Uniformly Over One-Year**  
**Interval; 1891 Birth Cohort From the SSA DMF**

- a) **Distribution for 90-year-olds**
- b) **Distribution for 102-year-olds**

a)



b)

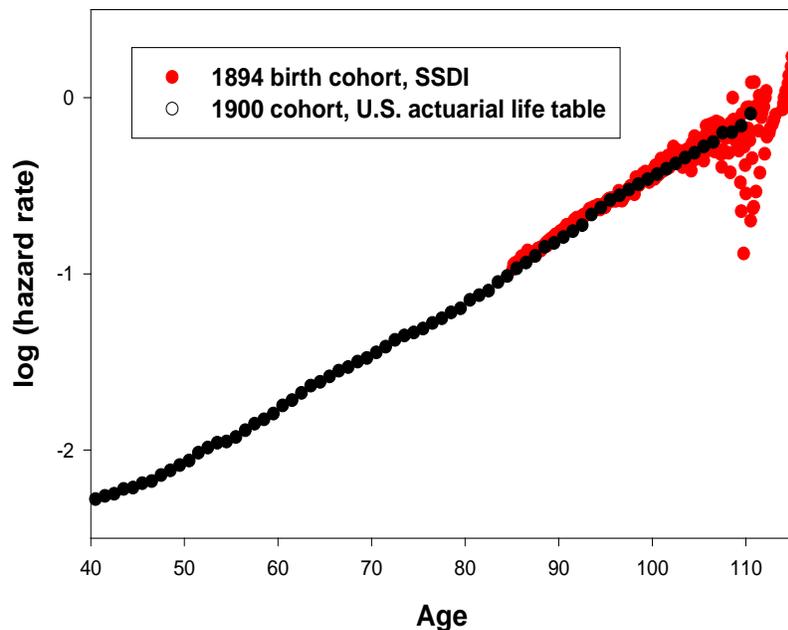


We initially compared our DMF data with probability of death estimates from the actuarial life table ( $q_x$ ) and found that life table values of probability of death show deceleration after age 100. It should be noted, however, that the annual estimates of probability of death ( $q_x$ ) and

mortality force ( $\mu_x$ ) substantially diverge at advanced ages. This happens because probability of death cannot continue its rapid growth when it approaches its mathematical upper limit equal to 1, while mortality force theoretically can increase indefinitely with age. Therefore, estimates of probability of death at advanced ages are biased downward compared to the hazard rate estimates. Hence, probability of death estimates beyond age 100 show deceleration in semi-log coordinates even if estimates of hazard rates follow the Gompertz model.

Another issue is the shape of death distribution within studied age intervals. This issue is important because some estimates of hazard rate (for example, the actuarial estimate of hazard rate) are based on the assumption of approximately uniform distribution of deaths within one-year age intervals (Kimball 1960; Watson and Leadbetter 1964; Gehan and Siddiqui 1973). Indeed, this assumption is consistent with the data for 90-year-olds, presented in Figure 9a. However for 102-year-olds, the distribution of deaths is obviously not uniform, with far more deaths occurring at the beginning of the age interval (Figure 9b). To alleviate this problem, it is preferable to make mortality estimates at advanced ages for shorter (monthly) age intervals, rather than for traditional one-year age intervals if the actuarial estimate of hazard rate is applied.

**Figure 10**  
**Actuarial 1900 U.S. Cohort Life Table and DMF 1887 Birth Cohort;**  
**Source for Actuarial Life Table: Bell, F.C., Miller, M.L. Life Tables for the**  
**United States Social Security Area 1900-2100; Actuarial Study No. 116**



The formula proposed by Sacher (Sacher 1956; Sacher 1966) gives more accurate estimates of hazard rate for yearly age intervals, so we applied a simplified Sacher formula for age-specific hazard rate calculations to the 1900 actuarial cohort life table:

$$\mu_x = -\ln(1-q_x)$$

where  $q_x$  is a probability of death value from the life table.

Figure 10 shows the trajectories of age-specific hazard rates for the 1900 birth cohort from the actuarial study and the 1894 birth cohort from the SSA DMF over a broad interval starting at age 40. Note that hazard rate estimates for the 1894 birth cohort are practically identical to the hazard rate estimates calculated on the basis of the actuarial life table. Also note that mortality of the 1894 birth cohort has the same slope in semi-log coordinates as mortality of the actuarial birth cohort calculated for a much wider age interval, which does not support the suggestion about the two-stage Gompertz model of mortality at advanced ages. Indeed, the maximum likelihood estimator of the Gompertz slope parameter for mortality in the 1894 cohort measured in the interval 88 to 106 years ( $0.0786 \text{ year}^{-1}$ , 95 percent confidence interval: 0.0786-0.0787) does not differ from the slope parameter calculated over the age interval 40 to 104 years in the 1900 life table ( $0.0785 \text{ year}^{-1}$ , 95 percent confidence interval: 0.0772-0.0797). Thus, we may conclude that more accurate estimates of hazard rates at advanced ages based on individual mortality data practically coincide with hazard rate estimates calculated on the basis of the 1900 actuarial cohort life table (Bell et al. 1992). It is remarkable that estimates in the actuarial life table obtained by extrapolation of annual probabilities of death after age 95 (Bell et al. 1992) are practically identical to the observed estimates of hazard rates based on mortality experience of a very large cohort of old individuals.

## Discussion

Our study of large single-year U.S. birth cohorts revealed that mortality deceleration at advanced ages is not observed up to the age of 106. Below the age of 107 and for data of reasonably good quality, the Gompertz model fits mortality better than the logistic model (no mortality deceleration). We also found that deceleration of mortality in later life is more expressed for data with lower quality. Quality of age reporting in DMF becomes particularly poor after the age of 107. It is also interesting that the DMF mortality data agree remarkably well with mortality data reported by the 1900 actuarial cohort life table (if hazard rate values are compared).

There are several reasons the earlier studies, including our own research (Gavrilov 1984; Gavrilov and Gavrilova 1991), reported mortality deceleration and mortality leveling-off at advanced ages (Kannisto 1994; Horiuchi and Wilmoth 1998; Thatcher et al. 1998; Thatcher 1999; Robine and Vaupel 2001). First, many studies present information for age-specific probability of death rather than hazard rate (Robine and Vaupel 2001; Gallop and Macdonald 2005). 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 an upper limit equal to one. 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). The authors do not provide estimates of hazard rate for this small heterogeneous population, so, at this moment, it is difficult to make a conclusion about the real mortality trajectory in this sample of very old individuals. In the study of validated French-Canadian centenarians born in 1870-94, the authors found no growth of mortality with age for life table death rates rather than hazard rates (Beaudry-Godin et al. 2008). It is true that for young and middle ages (when mortality is relatively low), probability of death and hazard rate values are practically identical. As a result, some authors do not distinguish between probability of death and hazard rate in their calculations (Le Bras 2005).

In the studies of period mortality, deceleration at advanced ages is probably caused by mixing together data for several birth cohorts with different mortality levels (heterogeneity effect) and using cross-sectional instead of cohort data (Kannisto 1994; Horiuchi and Wilmoth 1998; Thatcher et al. 1998). In addition to that, if information about population at risk is taken from censuses, there is always a possibility of mismatch in the accuracy of age reporting between deaths and population at risk due to higher proportion of age misreporting in censuses (so-called denominator problem). It should be noted that mortality deceleration was also reported for cohort mortality data (Thatcher et al. 1998). In this case, mortality deceleration can be caused by age misreporting in death data for older people, as we already found in our study. Earlier studies, conducted more than 10 years ago, used data for older birth groups when age reporting was not particularly accurate even for such countries as the United Kingdom (Gallop and Macdonald 2005). In addition to that, 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 many single-year birth cohorts, thereby increasing the heterogeneity of the sample. Thus, age exaggeration, data heterogeneity and use of probabilities of death instead of hazard rates are the main causes for downward biases in mortality estimates at older ages found in previous studies.

In our study, we found no significant mortality deceleration at advanced age for humans while many lower organisms demonstrate very long late-age mortality plateaus (Gavrilov and Gavrilova 1991; Carey et al. 1992; Curtsinger et al. 1992; Vaupel et al. 1998). Earlier studies

showed that the period of mortality deceleration in mammalian species is very short (Lindop 1961; Sacher 1966) compared to lower organisms (Vaupel et al. 1998; Gavrilov and Gavrilova 2006). It appears to be relatively short in humans too. This observation agrees with the prediction of the reliability theory of aging that more complex living systems/organisms with many vital subsystems (such as mammals) may experience very short or no period of mortality plateau at advanced ages in contrast to more simple living organisms (Gavrilov and Gavrilova 1991; Gavrilov and Gavrilova 2001; Gavrilov and Gavrilova 2006). Thus, human mortality at advanced ages may show very short or even negligible mortality plateau despite the evidence of mortality leveling-off among lower organisms such as insects.

The results obtained in this study may be important for actuarial practice, particularly if mortality is analyzed for real birth cohorts. For cohort data, hazard rates can be extrapolated with the Gompertz formula up to 107 years of age and then probabilities of death can be reconstructed with the Sacher formula if necessary. Our study also suggests that improvement of age reporting in history should result in mortality trajectories that better follow the Gompertz function until very advanced ages. SSA DMF data confirm that the 1900 actuarial cohort life table (Bell et al. 1992) provides a good description of mortality at advanced ages, suggesting that cohort life tables obtained in this study may be useful in actuarial practice.

## **Acknowledgements**

This study was supported by the National Institute on Aging (R01 AG028620 grant). We are most grateful to the participants of the 2011 International Symposium “Living to 100” in Orlando, Fla., organized by the Society of Actuaries, for their constructive criticism and useful suggestions.

## References

- Beard, R.E. 1959. "Note on Some Mathematical Mortality Models." In *The Lifespan of Animals*, E.W. Wolstenholme and M.O. O'Connor, eds., 302-11. Boston: Little, Brown.
- . 1971. "Some Aspects of Theories of Mortality, Cause of Death Analysis, Forecasting and Stochastic Processes." In *Biological Aspects of Demography*, W. Brass, ed., 57-68. London: Taylor and Francis.
- Beaudry-Godin M., R. Bourbeau, and B. Desjardins. 2008. "Data Validation and Measurement of Cohort Mortality Among Centenarians in Quebec (Canada) According to Ethnic Origin." In *Living to 100 and Beyond*, online monograph. Schaumburg: Society of Actuaries.
- Bell, F.C., A.H. Wade, and S.C. Goss. 1992. *Life Tables for the United States Social Security Area 1900-2080. Actuarial Study No. 107*. Baltimore, Md.: U.S. Department of Health and Human Services.
- Carey, J.R., P. Liedo, D. Orozco, and J.W. Vaupel. 1992. "Slowing of Mortality-Rates at Older Ages in Large Medfly Cohorts." *Science* 258: 457-61.
- 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: 727-36.
- Curtsinger, J.W., H. Fukui, D. Townsend, and J.W. Vaupel. 1992. "Demography of Genotypes: Failure of the Limited Life-span Paradigm in *Drosophila melanogaster*." *Science* 258: 461-63.
- Faig, K. 2001. "Reported Deaths of Centenarians and Near-Centenarians in the U.S. Social Security Administration's Death Master File." In *Proceedings of the Society of Actuaries "Living to 100 and Beyond" International Symposium*. Orlando, Fla.
- Gallop, A., and A.S. Macdonald. 2005. "Mortality at Advanced Ages in the United Kingdom." In *Living to 100 and Beyond*, online monograph. Schaumburg: Society of Actuaries.
- Gavrilov, L.A. 1984. "Does the Limit of the Life-span Really Exist?" *Biofizika* 29: 908-09.
- Gavrilov, L.A., and N.S. Gavrilova. 1991. *The Biology of Life Span: A Quantitative Approach*. New York: Harwood Academic Publisher.
- . 2001. "The Reliability Theory of Aging and Longevity." *Journal of Theoretical Biology* 213: 527-45.
- . 2006. "Reliability Theory of Aging and Longevity." In *Handbook of the Biology of Aging*, E.J. Masoro and S.N. Austad, eds., 3-42. San Diego: Academic Press.
- . 2007. "Search for Predictors of Exceptional Human Longevity: Using Computerized Genealogies and Internet Resources for Human Longevity Studies." *North American Actuarial Journal* 11: 49-67.
- Gehan, E.A. 1969. "Estimating Survival Functions from Life Table." *Journal of Chronic Diseases* 21: 629-44.
- Gehan, E.A., and M.M. Siddiqui. 1973. "Simple Regression Methods for Survival Time Studies." *Journal of the American Statistical Association* 68: 848-56.

- 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." *Philosophical Transactions of the Royal Society of London, Series A* 115: 513-85.
- Haines, M.R. 1998. "Estimated Life Tables for the United States, 1850-1910." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 31: 149-69.
- Hill, M.E. and I. Rosenwaik. 2001. "The Social Security Administration's Death Master File: The Completeness of Death Reporting at Older Ages." *Social Security Bulletin* 64: 45-51.
- Horiuchi, S., and J.R. Wilmoth. 1998. "Deceleration in the Age Pattern of Mortality at Older Ages." *Demography* 35: 391-412.
- 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., and B.R. Ferguson. 2001. "Mortality of the Extreme Aged in the United States in the 1990s, Based on Improved Medicare Data." In *Living to 100 and Beyond International Symposium*. Schaumburg, Ill.: The Society of Actuaries.
- Kimball, A.W. 1960. "Estimation of Mortality Intensities in Animal Experiments." *Biometrics* 16: 505-21.
- Klein, J.P., and M.L. Moesberger. 1997. *Survival Analysis Techniques for Censored and Truncated Data*. New York: Springer-Verlag.
- Le Bras, H. 1976. "Lois de Nortalité et Age Limité." *Population* 31: 655-92.
- . 2005. "Mortality Tempo Versus Removal of Causes of Mortality: Opposite Views Leading to Different Estimations of Life Expectancy." *Demographic Research* 13: 615-40.
- Lindop, P.J. 1961. "Growth Rate, Lifespan and Causes of Death in SAS/4 Mice." *Gerontologia* 5: 193-208.
- Robine, J.-M., A. Cournil, J. Gampe, and J.W. Vaupel. 2005. IDL, the International Database on Longevity. In 2005 Living to 100 and Beyond Monograph, SOA Monograph LI05-1. Society of Actuaries. Robine, J.-M., and J.W. Vaupel. 2001. "Supercentenarians: Slower Aging Individuals or Senile Elderly?" *Experimental Gerontology* 36: 915-30.
- Rosenwaik, I., and L.F. Stone. 2003. "Verification of the Ages of Supercentenarians in the United States: Results of a Matching Study." *Demography* 40: 727-39.
- Sacher, G.A. 1956. "On the Statistical Nature of Mortality, With Especial Reference to Chronic Radiation Mortality." *Radiology* 67: 250-57.
- . 1966. "The Gompertz Transformation in the Study of the Injury-Mortality Relationship: Application to Late Radiation Effects and Aging." In *Radiation and Aging*, P.J. Lindop and G.A. Sacher, eds., 411-41. 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: 107-11.
- StataCorp (2009). *Stata Statistical Software: Release 11*. College Station, Texas: StataCorp LP.

- 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.W. Vaupel. 1998. *The Force of Mortality at Ages 80 to 120*. Odense: Odense University Press.
- Vaupel, J.W., J.R. Carey, K. Christensen, T.E. Johnson, A.I. Yashin, N.V. Holm, I.A. Iachine, V. Kannisto, A.A. Khazaeli, P. Liedo, V.D. Longo, Y. Zeng, K.G. Manton, and J.W. Curtsinger. 1998. "Biodemographic Trajectories of Longevity." *Science* 280: 855-60.
- Vincent, P. 1951. "La Mortalite des Viellards." *Population* 6: 181-204.
- Watson, G.S., and M.R. Leadbetter. 1964. "Hazard Analysis. I." *Biometrika* 51: 175-84.
- 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." *The Journals of Gerontology: Series A* 63: 338-49.