Search for Predictors of Exceptional Human Longevity: Using Computerized Genealogies and Internet Resources for Human Longevity Studies

Natalia S. Gavrilova, PhD
Leonid A. Gavrilov, PhD

[This paper describes the current status of the ongoing research project "Search for Predictors of Exceptional Human Longevity" supported by the Society of Actuaries, with final report scheduled for June 15, 2005].

Centenarians (people living to 100 and beyond) represent the fastest growing age group of the American population with obvious implications for actuarial science and practice. Yet, factors predicting exceptional longevity and its time trends remain to be fully understood. In this study we explored the new opportunities provided by the ongoing revolution in information technology, computer science and Internet expansion for studies of exceptional human longevity. Specifically, we explored the availability and quality of computerized online genealogies of long-lived individuals by cross-checking them with other Internet resources including the Social Security Administration Death Master File and the early US censuses. To this aim, we extracted detailed family data for 991 centenarians born in 1875-1899 in the United States from publicly available computerized genealogies of 75 million individuals identified in our previous study (Gavrilova, Gavrilov, 1999). In order to validate the age of the centenarians we linked these records to the Social Security Administration Death Master File (DMF) records and then to the records of the US censuses for years 1900, 1910, and 1920. Data cross-checking with the Social Security DMF revealed only a small proportion (1.6%) of death date misreporting in genealogies and/or DMF itself. We also found that inaccuracies in birth date reporting as detected through linkage to the US Censuses are relatively rare (8%) and small (one-year disagreement between compared data sources). The results of this cross-validation study demonstrate that computerized genealogies may serve as a useful starting point for developing a family-linked scientific database on exceptional human longevity, and that this research data could be made reliable through their cross-validation with the Social Security Administration DMF and the US censuses.

This paper also presents some preliminary studies on determinants of exceptional human longevity including familial factors and early-life conditions. Specifically, this study suggests that there may be a sex-specific link between exceptional longevity and a person's birth order. Women seem to be more likely become centenarians, if they are born earlier compared to other siblings, when their parents are relatively young. In contrast to women, the birth order of centenarian-men is no different from what would be expected by pure chance. These observations correspond well with earlier published findings obtained on other

datasets that daughters conceived to older fathers live shorter lives, while sons are not affected by the fact of their late conception. These findings are corroborated by another observation made in this study – old paternal age decreased the chances for daughters to become centenarians by one half ($p<0.01$) while the effect of paternal age was statistically insignificant for sons.

We also compared the dataset of households where centenarians were raised (obtained through linkage of genealogies to early US Censuses) with control households drawn from the Public Use Samples (IPUMPS) for the 1900 US Census. This comparison suggests that the farm background (ownership in particular) and the Western region of residence in the United States may be predictive for survival to age 100.

Data from the Social Security Administration Death Master File (DMF) allowed us to analyze mortality patterns at advanced ages, using the method of extinct generations. The DMF covers deaths that occurred in the period 1937-2003 and is considered by some researchers superior in quality to the official U.S. vital statistics. Some birth cohorts in the Social Security DMF may be considered as extinct or almost extinct. Detailed information about birth and death dates of decedents allowed us to estimate hazard rates of the oldest-old persons with resolution of single month of their age. Study of three birth cohorts (1885, 1889 and 1891) showed that mortality grows steadily with age from 85-89 to 102-105 years with almost no sign of expected mortality deceleration. After age 105 the mortality estimates become unreliable because of significant statistical noise. We also found that life expectancy at age 80 depends on month of persons' birth: individuals born in April-June live shorter lives than persons born in October-November and this periodicity repeats in every birth cohort from 1885 to 1899. However, by age 100 this dependence of survival on month of birth fades out indicating that centenarians indeed represent a selected population.