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

Natalia S. Gavrilova, Leonid A. Gavrilov

Abstract

[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 crosschecking them with other Internet resources, including the Social Security Administration (SSA) Death Master File (DMF) and the early U.S. 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 records and then to the records of the U.S. censuses for years 1900, 1910 and 1920. Data crosschecking with the Social Security DMF revealed only a small proportion (2.6 percent) of death date misreporting in selected genealogies and/or DMF itself. We also found that inaccuracies in birth date reporting as detected through linkage to the U.S. censuses are relatively rare (8 percent) and usually small (only one-year difference 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 U.S. 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 link between exceptional longevity and a person's birth order. We found that the first-born daughters are three times more likely to survive to age 100, compared to later-born daughters of higher birth orders (7+). First-born sons are twice more likely to become centenarians compared to sons having birth order between four and six. There is also a profound sex difference in the effects of birth order on human longevity. For sons, this dependence has an unusual U-shaped form, with highest longevity chances for both the first-born and the last-born (9+) sons.

We also compared the data set of households where centenarians were raised (obtained through linkage of genealogies to early U.S. censuses) with control households drawn from the Integrated Public Use Microdata Series (IPUMS) for the 1900 U.S. census. This comparison suggests that the farm background (farm ownership by parents in particular) and the Western region of child residence in the United States may be predictive for survival to age 100.

Data from the Social Security Administration Death Master File 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 it is considered by some researchers to be superior in data quality in comparison to the official U.S. vital statistics, particularly for old age groups and recent time periods. Social Security DMF data allowed us to reconstruct cohort life tables describing survival patterns after age 80 years for those birth cohorts that are almost extinct now (born in 1891 and earlier). 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 obvious signs of expected mortality deceleration. After age 105, the mortality estimates become less reliable because of significant statistical noise. We also found that life expectancy at age 80 depends on month of persons' birth: individuals born in January live longer lives than persons born in other months and in April-June in particular. This periodicity repeats in every studied birth cohort starting from birth year1885 to 1899. However, by age 100 this dependence of survival on month of birth fades out, indicating that centenarians indeed represent a selected population.