The 2007 Annual Meeting of the Population Association of America

Physical Markers at Young Age and Survival to 100
A study of a new historical data resource (the US WWI draft cards)

Natalia Gavrilova, Leonid A. Gavrilov
Center on Aging, NORC and The University of Chicago

Abstract

This study explored whether people living to 100 and beyond were any different from their peers at their middle age (30 years) in terms of their physical characteristics (height and 'build'). An individual's height at young adult age seems to be a good indicator of person's early nutritional and infectious disease history at least in historical data.

The study took advantage of a new unique data resource – the US WWI Civilian Draft Registration Cards collected in 1917-1918, which covers 98% (24 millions) of the entire US male population at draft ages of 18-45 years.

The US WWI draft cards were linked to exceptional longevity records obtained from the Social Security Administration database. The linked data were analyzed using a conditional multiple regression model for matched case-control studies to find out what combinations of the middle-life predictor variables are conductive for exceptional longevity in the United States.

Introduction

Centenarians (persons living to age 100 and over) represent a population, which could be useful in identifying factors leading to long life and avoidance of fatal diseases. Even if some middle-life factors have a moderate protective effect on risk of death, persons with this trait/condition would be accumulated among long-lived individuals. Thus, study of centenarians may be a sensitive way to find genetic, familial, environmental, and life-course factors associated with lower mortality and better survival.

Incorporation of physical characteristics into demographic analysis of mortality widens a scope of explanatory variables in biodemographic research on health outcomes (Crimmins and Seeman 2000). This study investigates the effects of the physical traits – height and 'build' at young age - on the chances of survival to age 100.

Height is an important indicator of childhood nutritional status and exposure to early infections. An individual's height at young adult age seems to be a good indicator of person's nutritional and infectious disease history at least in historical data (Elo and Preston 1992; Alter 2004; Alter, Neven et al. 2004). Most studies, starting with Waaler's pioneer work, found a negative relationship between body height and mortality later in life (Waaler 1984; Elo and Preston 1992). A study of Union Army veterans found that the relationship between height and subsequent mortality was negative (Costa 1993; Fogel and Costa 1997; Costa and Lahey 2005), findings similar to a study of modern Norwegian males (Costa 1993; Costa 1993; Fogel and Costa 1997; Costa and Lahey 2005). Infectious diseases (and diarrhoeal diseases in particular) can result in growth retardation leading to shorter adult height. For example, conscripts from high-mortality districts of antebellum New York were shorter than those from healthier districts (Haines, Craig et al. 2003).

In addition to nutrition and disease exposure, height is also dependent on genetics. Heritability of height is one of the highest among human quantitative traits. However, genetic influence may be suppressed by environmental factors, such as poor nutrition or early infections in the past (Lauderdale and Rathouz 1999).

It is not clear what the body size of centenarians was during their adult ages. Historical studies suggest that centenarians may be taller than average due to better nutrition and avoidance of diseases early in life. The proposed study tested this hypothesis.

Our previous study explored new opportunities provided by the ongoing revolution in information technology, computer science, and Internet expansion for studies of exceptional human longevity. Specifically, it explored the availability and quality of computerized online family histories (genealogies) of long-lived individuals by cross-checking them with other Internet resources, including the Social Security Administration (SSA) Death Master File (DMF) and early U.S. censuses.

In summation, this earlier exploratory study developed a new methodology of using online genealogical, historical, and demographic data resources; demonstrated feasibility of large-scale studies on predictors of human longevity; and yielded preliminary findings on several hypotheses on the determinants of survival to advanced ages (Gavrilova and Gavrilov 2005). These preliminary studies also allowed us to learn about the existence of the US WWI Civilian Draft Registration Cards, their recent availability online, and their rich content in terms of predictor variables, which made this study possible.

Rationale

Adult body height is affected by both environmental (early-life nutrition and exposure to infections) and genetic factors. It was found that familial resemblance in height was suppressed in the past possibly because of early environmental effects (Lauderdale and Rathouz 1999). It was suggested that population of the United States at the end of the 19th century had relatively good nutritional status but very high burden of infections (Preston and Haines 1991). Thus, we may hypothesize that low height of males born in the end of the 19th century may be related to the infectious diseases during childhood. If the hypothesis of childhood infections as a possible cause of late-life chronic diseases is correct (Finch and Crimmins 2004), we may expect that centenarians at young adult ages would be taller on the average than their peers who did not survive to advanced ages. Alternatively, if biological hypothesis about adverse effects of rapid catch-up growth on longevity is correct (Rollo 2002), we may expect the opposite result. These hypotheses were tested in this study using data from the WWI civilian draft registration cards (see below).

WWI Civilian Draft Registration Cards as Data Resource on Person's Height, 'Build' and Other Physical Characteristics. In 1917 and 1918, approximately 24 million men born between 1873 and 1900 completed draft registration cards. President Wilson proposed the American draft and characterized it as necessary to make "shirkers" play their part in the war. This argument won over key swing votes in Congress.

Men already on active duty in the military were excluded from draft registration. Registration of eligible men has been determined to be close to 100%, which means that about 98% of adult men under age 46 living in the U.S. in 1917-18 completed registration cards. Instructions for filling in each question on the card were posted for all to read at each registration site, and the local newspapers sometimes printed copies of sample cards in the days prior to registration. In the vast majority of cases, volunteer staff at the local office filled in the information on the card, and the registrant then signed his name. Men who registered were given certificates to prove

they had registered. More detailed description of this data source is available in a seven-volume set, which is a part of ongoing study by Raymond H. Banks (Banks 2000). Table 1 describes variables available in the draft registration cards.

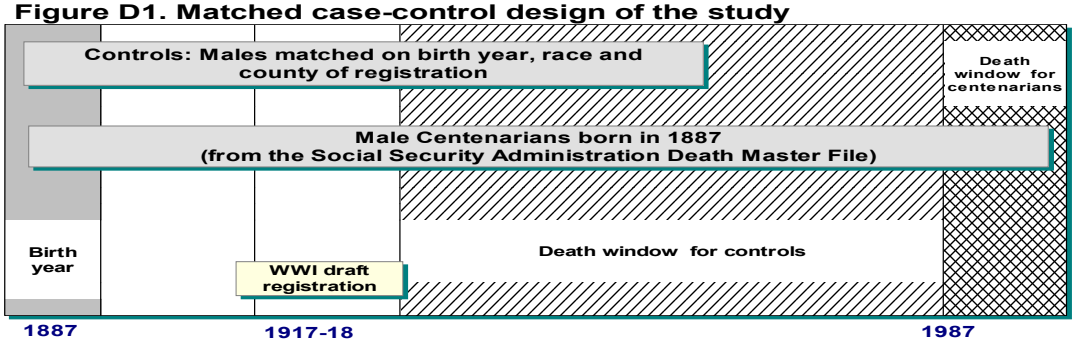
Table 1. Variables available from the WWI draft registration cards.

Variable group	Variable description
Core demographic variables	age, date/place of birth, race, citizenship
Geographical variables	permanent home address
Working characteristics	occupation, employer's name
Family characteristics	Marital status, information about dependents
Physical characteristics	height (3 categories: tall, medium, or short), build (3 categories: slender, medium, or stout), eye color, hair color, baldness, disability

The WWI civilian draft registration cards are available now online through the paid service provided by the Ancestry.com. Thus the linkage process was facilitated by availability of online indexes and actual digitized images of draft registration cards. The linkage process was facilitated by the fact that the exact birth date (day, month, year) is provided both in the WWI draft cards and the Social Security Administration Death Master File allowing us to obtain unambiguous matches.

Matched Case-Control Study Design. The study applied a *matched case-control design* where shorter-lived males matched with centenarian males by birth year, race and county of draft registration were used as controls (see Figure D1). This approach allowed us to eliminate effects of birth cohort, race and place of draft registration on survival. Using controls from the same geographical area (county) allowed us to mitigate a possible geographically-related subjectivity in height and build estimation.

Figure D1. Matched case-control design of the study



The development of the study sample was conducted in three stages:

- In the first stage, records of 120 males born in 1887 and survived to age 100 were randomly selected from the Social Security Administration Death Master File (DMF). Males born in 1887 reached age 30 in 1917, so their height stopped to grow by the time of draft registration. Taking into account that DMF covers 93-96 percent of deaths of persons aged 65+ (Hill and Rosenwaike 2001), it was possible to apply a simple random sampling design for male centenarian data. The DMF database contains about 2,500 death records of male centenarians born in 1887.

- In the second stage, the selected records were linked to the WWI civilian draft registration cards available online (a service provided by the Ancestry.com).
- In the third stage, each centenarian record has received three matched control records randomly selected from the civilian draft registration records of persons of the same birth year, race and county of registration. The 1:3 matched design is known to be an optimal way of gaining efficiency (e.g., precision of estimates) at minimal cost (number of controls) (Woodward 2005).

Model specification. The statistical analyses were performed using a conditional multiple logistic regression model for matched case-control studies to investigate the relationship between an outcome of being a case (survival to age 100) and a set of predictor variables (Breslow and Day 1993; Hosmer and Lemeshow 2001). When each matched set consists of 3 controls (1-3 matched study), the conditional likelihood is given by:

$$L(\beta) = \prod_{i=1}^{N} \frac{\exp(u_i'\beta)}{\exp(u_i'\beta) + \exp(v_{1i}'\beta) + \exp(v_{2i}'\beta) + \exp(v_{3i}'\beta)}$$

where u_i and v_{ki} are vectors representing the predictor variables of case and 3 controls, respectively, of the *i*th matched set (Hosmer and Lemeshow 2001). We began with a matched univariate analysis for the "height" variable, using conditional logistic regression. Then other predictor variables (build, eye/hair color, disability status, citizenship status, occupation and marital status) were added to the model. Computations were done using Stata (StataCorp 2005).

Some limitations and ways of their overcoming. Although draft registration cards contain valuable information on individual physical markers, this resource is not free of limitations. The main difficulty we face here is using height and build data measured in a categorical rather than continuous scale - in three broad categories, which are less precise than measures provided in specialized health surveys like NHANES. During the WWI draft registration, local staff were asked to classify individual men as to height and weight. The three categories provided were rather vague, and occasionally staff wrote in actual weight and height instead. Some other errors in reporting physical characteristics were also mentioned (Banks 2000). Nevertheless, the data were measured by the volunteer staff in the registration office at the time when centenarians were young adults and hence are not subjected to self-report and recall bias. Also, using county-matched controls helps to avoid possible regional differences in defining "tallness" or "shortness." This study provides the first estimates of height and build for the future centenarians at their young ages and helps to test alternative hypotheses on links between the early growth and longevity.

References

Alter, G. (2004). "Height, frailty, and the standard of living: Modelling the effects of diet and disease on declining mortality and increasing height." Population Studies-a Journal of Demography **58**(3): 265-279.

Alter, G., M. Neven, et al. (2004). "Stature in transition - A micro-level study from nineteenth-century Belgium." Social Science History **28**(2): 231-247.

Banks, R. (2000). World War I Civilian Draft Registrations. [database on-line]. Provo, UT, Ancestry.com.

Breslow, N. E. and N. E. Day (1993). Statistical Methods in Cancer Research. Vol.1. The Analysis of Case-Control Studies. Oxford, Oxford University Press.

Costa, D. L. (1993). "Height, Wealth, and Disease among the Native-Born in the Rural, Antebellum North." Social Science History **17**(3): 355-383.

Costa, D. L. (1993). "Height, Weight, Wartime Stress, and Older Age Mortality - Evidence from the Union Army Records." Explorations in Economic History **30**(4): 424-449.

Costa, D. L. and J. Lahey (2005). "Becoming oldest old: evidence from historical U.S. data." Genus **61**(1): 125-161.

Crimmins, E. M. and T. Seeman (2000). Integrating biology into demographic research on health and aging (With a focus on the MacArthur Study of Successful Aging). Cells and Surveys. K. Kinsella. Washington, DC, National Academy Press**:** 9-41.

Elo, I. T. and S. H. Preston (1992). "Effects of early-life condition on adult mortality: A review." Population Index **58**(2): 186-222.

Finch, C. E. and E. M. Crimmins (2004). "Inflammatory exposure and historical changes in human life-spans." Science **305**(5691): 1736-1739.

Fogel, R. W. and D. L. Costa (1997). "A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs." Demography **34**(2): 49-66.

Gavrilova, N. S. and L. A. Gavrilov (2005). Search for predictors of exceptional human longevity: using computerized genealogies and internet resources for human longevity studies. Living to 100 and Beyond Monograph, Society of Actuaries.

Haines, M. R., L. A. Craig, et al. (2003). "The short and the dead: Nutrition, mortality, and the "antebellum puzzle" in the United States." Journal of Economic History **63**(2): 382-413.

Hill, M. E. and I. Rosenwaike (2001). "The Social Security Administration's Death Master File: The completeness of death reporting at older ages." Social Security Bulletin **64**(1): 45-51.

Hosmer, D. W. and S. Lemeshow (2001). Applied Logistic Regression. New York, Wiley & Sons.

Lauderdale, D. S. and P. J. Rathouz (1999). "Evidence of environmental suppression of familial resemblance: height among US Civil War brothers." Annals of Human Biology **26**(5): 413-426.

Preston, S. H. and M. R. Haines (1991). Fatal years. Child mortality in late nineteenth-century America. Princeton, NJ, Princeton University Press.

Rollo, C. D. (2002). "Growth negatively impacts the life span of mammals." Evolution & Development **4**(1): 55-61.

StataCorp (2005). Stata User's Guide, Stata Press.

Waaler, H. T. (1984). "Height, weight and mortality." Acta Medica Scandinavica **Suppl. No. 679**.
Woodward, M. (2005). Epidemiology. Study Design and Data Analysis. Boca Raton, FL, Chapman & Hall/CRC.