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

Natalia S. Gavrilova, Leonid A. Gavrilov

Presented at The Living to 100 and Beyond Symposium
Sponsored by the Society of Actuaries

Orlando, Fla.

January 12-14, 2005

Copyright 2005 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.

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.

Introduction

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. Thorough and comprehensive studies of survival at advanced ages requires search for new data sources in addition to careful re-evaluation of already known ones.

Our previous search for additional data resources (see Gavrilov, Gavrilova, 1998; Gavrilova, Gavrilov, 1999) revealed an enormous amount of new family lifespan data that could be made readily available for subsequent full-scale studies. Millions of genealogical records are already computerized and could be used for the study of familial clustering of human longevity (after strict data validation). Most of these genealogies are a product of family reconstitution, carried out both by professional genealogists and by family members tracing their ancestry back to the founder who brought their surname to America or even to their European family roots. The compilers of genealogies aided this time-consuming task by using many different sources: genealogical libraries, The Church of Jesus Christ Latter-day Saints (Mormon) family history centers, genealogical search engines available on the Internet, computer CDs with census, marriage, land and probate records, and many other resources for genealogical research.

Computerized genealogies provide the most complete information on the lifespan of centenarians' relatives, compared to other data sources (death certificates, census data, Medicare database). Census records provide information on birth years of parents and siblings, but no information on death dates is available. The Medicare database allows identification of spouses (see Iwashyna et al., 1998), but no information on parents and other relatives is available. Social Security Administration NUMIDENT file contains information on the names of parent/child pairs for Medicare beneficiaries (65 years and older). In the latter case, however, one cannot obtain information on distant ancestors (e.g., grandparents) as well as other relatives (e.g., first cousins), so there is no opportunity for reconstruction of pedigrees.

In this paper we describe our experience in identification, collection, verification and analysis of data taken from computerized genealogies for long-lived individuals. The process of data quality evaluation and centenarians' age verification is described in detail because it appears to be the first attempt in systematic assessment of quality for this new and potentially promising data source on family factors of longevity. We also test several hypotheses on the effects of early-life conditions and family factors affecting lifespan. In addition to that, we use data extracted from the Social Security Death Master File for detailed estimates of mortality rates at advanced ages.

I. Computerized Genealogies as a Potentially Useful Data Resource

Survey of the Existing Computerized Genealogies

At the first stage of the project, we made a survey of the relevant data resources and identified computerized family histories for over 75 million deceased individuals using online data resources (Ancestry.com and Genealogy.com) identified in our previous studies (Gavrilov, Gavrilova, 1998; Gavrilova, Gavrilov, 1999). Centenarian family histories were drawn from computerized family trees using the following selection criteria: (a) persons should have birth and death date information and have lifespan 100 year and over, (b) persons should be born in the United States after 1875 and (c) persons should have pedigree information for at least 3 generations of ancestry (both on paternal and maternal side) as well as information on birth date and death date of parents.

The decision to exclude foreign-born centenarians from our study was conditioned by the difficulties of their age verification. The main obstacle here is that it may be difficult to find many foreign-born persons in the available early U.S. censuses (1900 and 1910) used for birth date verification, because many of these persons could immigrate later. Also, in the case of foreign-born persons, the U.S. census data are useless in providing information about early-life conditions because foreign-born children spent a part of their childhood abroad in unknown living conditions. Thus, for the purpose of this particular project focused on the role of early-life conditions, the foreign-born persons are less informative. In addition to that, it is particularly difficult to verify the quality of the genealogical data for foreign-born centenarians. Thus, by excluding the genealogies for foreign-born centenarians, we excluded the most questionable part of the data, which are particularly difficult to cross-validate through early U.S. censuses. It should also be noted that foreign-born children comprised a small proportion (3 percent) of all children below age 10 enumerated in the 1900 census (we obtained this estimate from the IPUMS 1 percent random sample of the 1900 U.S. census population; for more details on IPUMS project see Ruggles et al., 2004).

Using online genealogical data resources, we identified over 2,000 genealogies, which contained detailed information about long-lived persons as well as detailed information about their parents and grandparents. The obtained genealogies were recorded in the so-called GEDCOM data format, which is used for genealogical data exchange (Gavrilova, Gavrilov, 1999) and is described below.

What is the GEDCOM File?

Although each particular genealogical software has its own data format, the genealogical data are shared among other genealogists through the GEDCOM format. "GEDCOM" stands for the "GEnealogical Data COMmunication" standard proposed by the Family History Department of The Church of Jesus Christ of Latter-day Saints (LDS Church) and adopted by many developers and users of genealogical software (Family History Department, 1996). The purpose of GEDCOM is to simplify the exchange of computerized historical and genealogical information. GEDCOM files are created in ASCII (text) format with special tags at the beginning of each line related to specific family information (variables). The most common variables contain personal information (name, birth date and place, death date and place) and family information (links to spouses and children and links to parents and sibs). In many cases, GEDCOM files contain more detailed information (occupation, education, residence, title, religion, cause of death, burial place and special notes). Data on living individuals are eliminated in the majority of computerized genealogies (to protect their privacy) except for their names and family links.

Information contained in GEDCOM files cannot be immediately used in statistical analyses because it needs to be converted to the relational database and cleaned. Thus, after collecting data in the form of GEDCOM files, they were converted to the relational database (MySQL) for their further verification and analysis.

Database on the U.S. Centenarians

The collected GEDCOM files were screened for long-lived individuals and converted to the MySQL database using specially developed program scripts. As a result, we obtained information for 2,004 long-lived individuals in the form of relational database. Out of these 2,004 records for long-lived individuals, we selected 991 records for centenarians born in the United States after 1875 and having detailed information about relatives (including information on parental names and their lifespan and grandparent names). As any new data resource, this data set has an uncertain quality, which requires additional efforts for data verification and quality control using several independent data sources. Our primary concerns were about the possibility of incorrect

dates reported in genealogies. Previous studies found that age misreporting and age exaggeration in particular are more common among long-lived individuals (Hill et al., 2000; Rosenwaike, Stone, 2003; Shrestha, Preston, 1995). For this reason the focus of our study was on the age verification for long-lived individuals rather than for other members of genealogy (which could be done later if time permits).

Verification of Centenarian Birth and Death Dates

Data Consistency Checks

To verify the centenarian's birth date, we first compared the person's birth date with birth dates for the person's parents, as well as with birth and marriage dates for the person's spouses (data consistency test). Our data consistency checks revealed a surprisingly small number of obvious data inconsistencies. In one case, an alleged centenarian had parents with incorrect birth dates (born later than the person himself). This case was dropped from the database. In another case, the centenarian's father was rather old (62 years) when the centenarian was born. This is not an impossible situation, so this case was left for further validation. All other records did not reveal obvious inconsistencies in event dates, so that 990 records were left for further verification.

Validation of Death Dates for the U.S. Centenarians Using the Social Security Administration Death Master File

In this project we followed the approach of age verification and data linkage developed by the group of demographers at the University of Pennsylvania (Rosenwaike, Logue, 1983; Preston et al., 1996; Rosenwaike et al., 1998; Hill et al., 2000; Rosenwaike, Stone, 2003).

Verification of death dates is an important step in quality control because it eliminates cases with potential mistakes and misprints in death dates reported for alleged centenarians. Verification of death dates was accomplished through a linkage of genealogical data to the Social Security Administration Death Master File. This is a publicly available data source that allows a search for individuals using various search criteria: birth date, death date, first and last names, Social Security number and place of last residence. This resource covers deaths that occurred in the period 1937-2003 (see Faig, 2001, for more details). Many researchers suggest that the quality of SSA/Medicare data for older persons is superior to vital statistics records because of strict evidentiary requirements in application for Medicare, while age reporting in death certificates is made by proxy informant (Kestenbaum, 1992; Kestenbaum, Ferguson, 2001; Rosenwaike et al., 1998; Rosenwaike, Stone, 2003). We also based the death date

verification on linkage to the SSA Death Master File, which is publicly available at the Rootsweb Web site (Faig, 2001).

The overwhelming majority of genealogical records when linked to the DMF had revealed an identical birth and death year as well as birth and death month in both databases (687 out of 764 cases, or 89.9 percent). These matched records were additionally verified using information about first and last names (or last names of spouses for women), places of death (in genealogy) and places of last residence (in DMF). When months of birth or death and years of birth or death did not match, then potential matches were established using information about place of death (in genealogical file) and place of last residence (in DMF file). Thus, in addition to 687 (out of 990) persons having exactly the same birth and death dates in both databases, it was possible to add some records with birth or death dates not identical in genealogies and DMF. In most cases these differences were related only to disagreement in *exact month of birth or death*, and in 731 cases (96 percent), the death year was the same both in genealogy and DMF. One problem for successful linkage to the DMF in our case was surname change by women after marriage. We resolved this problem by using surnames of spouses, which are available in genealogical database, so that linkage success was approximately the same in both sexes. The number of successful links strongly depends on centenarian's year of birth; persons born before 1890 were less likely to be found in the DMF (see Table 1 and Table 2 below). This result is consistent with previous reports that quality and coverage of DMF database was lower for persons born before 1890 (Faig, 2001).

TABLE 1
Data on Centenarians Born in 1875-1900
Linkage Success Rate With the Social Security Death Master File (DMF)

Sex	Found in DMF	Total Number of Persons	Percent Found
M	207	275	75%
F	557	715	78%
Total	764	990	77%

TABLE 2
Data on Centenarians Born in 1890-1900
Linkage Success Rate With the Social Security Death Master File (DMF)

Sex	Found in DMF	Total Number of Persons	Percent Found
M	130	160	81%
F	418	511	82%
Total	548	671	82%

Thus, the proportion of successful links is 75 percent for males and 78 percent for females in total. For centenarians born after 1889, the percentage of successful links is higher at 82 percent. Among 764 persons found in the DMF, their centenarian status was confirmed in 744 cases. 731 centenarians had the same calendar year of *death* both in genealogy records and in DMF. 714 centenarians had *both the birth and death years* identical in genealogy records and DMF. Centenarian status could not be confirmed only for 20 alleged centenarians from computerized genealogies (2.6 percent). The detailed breakdown of records found in DMF is presented in Tables 3 and 4.

TABLE 3
Results of Centenarian Death Dates Verification Using Social Security Administration Death Master File (DMF)

	All Centenarians Born in 1875-1900	Centenarians Born in 1890-1900
Total found in DMF	764	548
Centenarian status confirmed	744	532
Death year is exactly the same in genealogy and DMF	731	524

TABLE 4
Comparison of Death Year Reporting in Genealogy and the Social Security Death Master File (DMF)

Age at Death Reported in Genealogy	Number of Cases	Difference Between Death Year Reported in Genealogy and DMF
100	1	-1
	219	0
	1	1
	6	10
	1	20
101	4	-1
	243	0
	4	1
	2	10
	3	20
102	2	-1
	156	0
	1	2
	1	20
103	74	0
	1	1
	1	22
104	23	0
105	9	0
106	12	0
	2	20
107	5	0
	1	17
109	1	0
110	1	30
114	1	30

Note that the overall linkage success rate to the SSA DMF was moderate at 75-78 percent (Table 1). Also note (Table 4) that in 10 cases (1.3 percent) the difference between death year in genealogy and DMF was expressed in round numbers (e.g., 10, 20 or 30 years), which seems to be caused by misprints in genealogies. Thus all cases of exceptional longevity in genealogies should be verified using the SSA data. The lack of match with DMF could occur for a number of reasons: a misprint in genealogy, missing Social Security record (particularly if person did not use Medicare benefits), difficulty to match person with a common name when the dates are not identical, etc. In addition to that, DMF covers 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). Further work with non-matched cases using additional data sources (e.g., obituaries, references to death certificates, National Death Index) could probably improve the linkage success rate. It should be noted that the linkage success rate to DMF was substantially higher for persons born after 1889 at 82 percent. The 534 records for persons with confirmed centenarian status born after 1889 and matched to the DMF were used further in verification of centenarian birth dates through linkage to early censuses.

Verification of Centenarian B*irth Dates* Using the Early U.S. Censuses

In order to verify centenarian birth date, the data for centenarians were checked against the early U.S. census records collected when the centenarian was a child or young adult. For validation purposes, the early U.S. censuses (1900, 1910, and 1920) are particularly important, because they provide information on future centenarians during their childhood and early adulthood years when age exaggeration is less common compared to claims of exceptional longevity made at old age. The preference was given to the 1900 census because it is more complete and detailed (in regard to age verification) compared to the 1910 and 1920 censuses. Specifically, the 1900 U.S. census provided year and month of birth, not just an age at enumeration date.

The 1900 U.S. census provides the following information for household and its members: state, county, and township of residence; street and house number (where available); relationship to head-of-household; gender and ethnicity; month and year of birth and age at last birthday; marital status and, if married, length of marriage; for married women, number of children born and number living; birthplace of person and birthplaces of mother and father; for aliens or naturalized citizens, year of immigration and citizenship status; occupation of each person 10+ years and number of months not employed; information about school attendance and literacy; and information about home ownership or farm residence. An important advantage of the 1900 census is the

availability of information about year and month of birth providing an additional source for birth date verification.

In our study, the linkage of centenarian records to the early census data is facilitated by online availability of the entire indexed U.S. 1900, 1910 and 1920 censuses—a service provided by the Genealogy.com and Ancestry.com companies. This service allows researchers to conduct an online search by head-of-household across nearly 75 million individuals. Microfilm images of the 1900, 1910 and 1920 United States Federal Censuses also are available online for subscribers. In our project we conducted a linkage of 534 centenarian records (for centenarians found in the Social Security Death Master File with confirmed centenarian status and born after 1889) to the early U.S. censuses. If individuals were not found in the 1900 census, then attempts were made to locate them in the 1910 and 1920 censuses. The overall matching success rate was 91 percent, which is significantly higher than in other studies on linkage to early censuses: 39-56 percent (Rosenwaike, Logue, 1983; Guest, 1987; Rosenwaike et al., 1998), 69 percent (Hill et al., 2000) and 54 percent overall and 69 percent for whites (Rosenwaike, Stone, 2003). Table 5 shows the results of record linkage to the early U.S. censuses.

TABLE 5
Number and Percentage of Genealogical Records That Were Successfully Linked to Early U.S. Census Records Among Records Confirmed Through Linking to the Social Security Administration Death Master File (DMF)

	Males		**Females**		**Both Sexes**	
U.S. Census	**Number Linked to Early Census Record**	**Percentage Linked to Early Census Record**	**Number Linked to Early Census Record**	**Percentage Linked to Early Census Record**	**Number Linked to Early Census Record**	**Percentage Linked to Early Census Record**
1900	90	78%	292	79%	382	79%
1910	24	21%	76	20%	100	20%
1920	1	1%	2	1%	3	1%
Total	**115**	**100%**	**370**	**100%**	**485**	**100%**

The reasons for the relatively high success rate of linkage to the early censuses in our study can be explained by availability of detailed supplemental information in genealogical records. The most important piece of information for successful search in census records was information on *places of birth* for siblings born close to the census date. Thus, if family moved to another state after the birth of alleged centenarian,

his/her family could be easily traced using information about birthplaces of other siblings. This is an important advantage compared to the traditional studies of record linkage to the early U.S. censuses based on information taken from the Social Security SS-5 forms (Rosenwaike et al., 1998; Hill et al., 2000; Rosenwaike, Stone, 2003).

We had no need to apply the scoring system of match rating suggested in previous studies (Hill et al., 2000; Rosenwaike, Stone, 2003), because the availability of supplemental information in genealogy made the judgment about match or non-match perfectly clear. If names and years of birth for parents and siblings are in a good agreement in both genealogy and census, the match is considered as very confident. On the other hand, if names of parents are the same in census and genealogy but siblings have different names, it is quite clear that the match is not acceptable. In some rare cases of small families with one or two children, additional information about places of birth for parents and children was used to resolve the problem. Unlike previous studies of linkage to early censuses, we did not encounter problems with persons having common first and last names because detailed information about place of birth for potential centenarian and his/her siblings (state, county, township) helped to identify the correct match among many potential matches. The detailed information about names, ages and places of birth for parents and siblings available in genealogies helped us to avoid ambiguous matches, which should be common in linkage studies based only on the information about parental names and places of birth and residence (Rosenwaike et al., 1998). The main difficulty we encountered in our search was related to rare and unusual first names, which were spelled in a variety of ways in census indexes. As a result of record-matching to early censuses, we downloaded and studied 601 image files with the U.S. census pages (some persons had images for more than one U.S. census).

The agreement between years of birth recorded in computerized genealogies and years of birth reported by the 1900 census as well as age reported by the 1910 census was surprisingly good; there was 92 percent of complete agreement in birth year between genealogy records and census records. In one case only, the centenarian's year of birth was three years less than in genealogy, i.e., centenarian was in fact *older* than was reported in genealogy. In 4.5 percent cases, the birth year of centenarian in the U.S. census was one year less than the birth year indicated in genealogy and in 3.5 percent cases centenarian was one year younger than reported in genealogy. Disagreements between birth years reported in census and genealogies were more notable for parents (about 15 percent of all cases) than for children, but in the majority of cases the differences did not exceed one year.

As a result of this record linkage study, we could verify birth dates for 485 centenarians born after 1889. The steps of age verification for this group of centenarians are presented in Table 6.

TABLE 6
Summary of Results of Genealogy Records Linkage First to the Social Security Administration Death Master File (DMF) and Then to the Early U.S. Censuses

Steps of Data Verification	Number of Records for Centenarians Born After 1889		
	Males	**Females**	**Both Sexes**
Initial Number of Records	160 (100%)	511 (100%)	671 (100%)
Found in the DMF	130 (81%)	418 (82%)	548 (80%)
Found in the Early Censuses	115 (72%)	370 (72%)	485 (72%)

Thus, we obtained 485 records for centenarians with verified birth dates, confirmed centenarian status and detailed genealogies. We did not find many cases of significant age exaggeration among centenarians with known genealogies and verified death dates. In other words, the birth year is recorded more accurately in genealogies than the death year. The 25 cases of one-year discrepancy with census records are more likely caused by inaccurate birth date reporting during census enumeration rather than inaccuracy of genealogical records. Most genealogical records provide detailed date of birth (day, month and year) taken from birth certificates or family bible records while census records are based on verbal reports during enumeration.

Description of the Verified Data Sample of Centenarians

As a result of this validation study, a cohort of 485 centenarians born in the United States in 1890-1900 was identified. A general overview of data collection, verification and linkage used for identification of these 485 cases is presented in Figure 1.

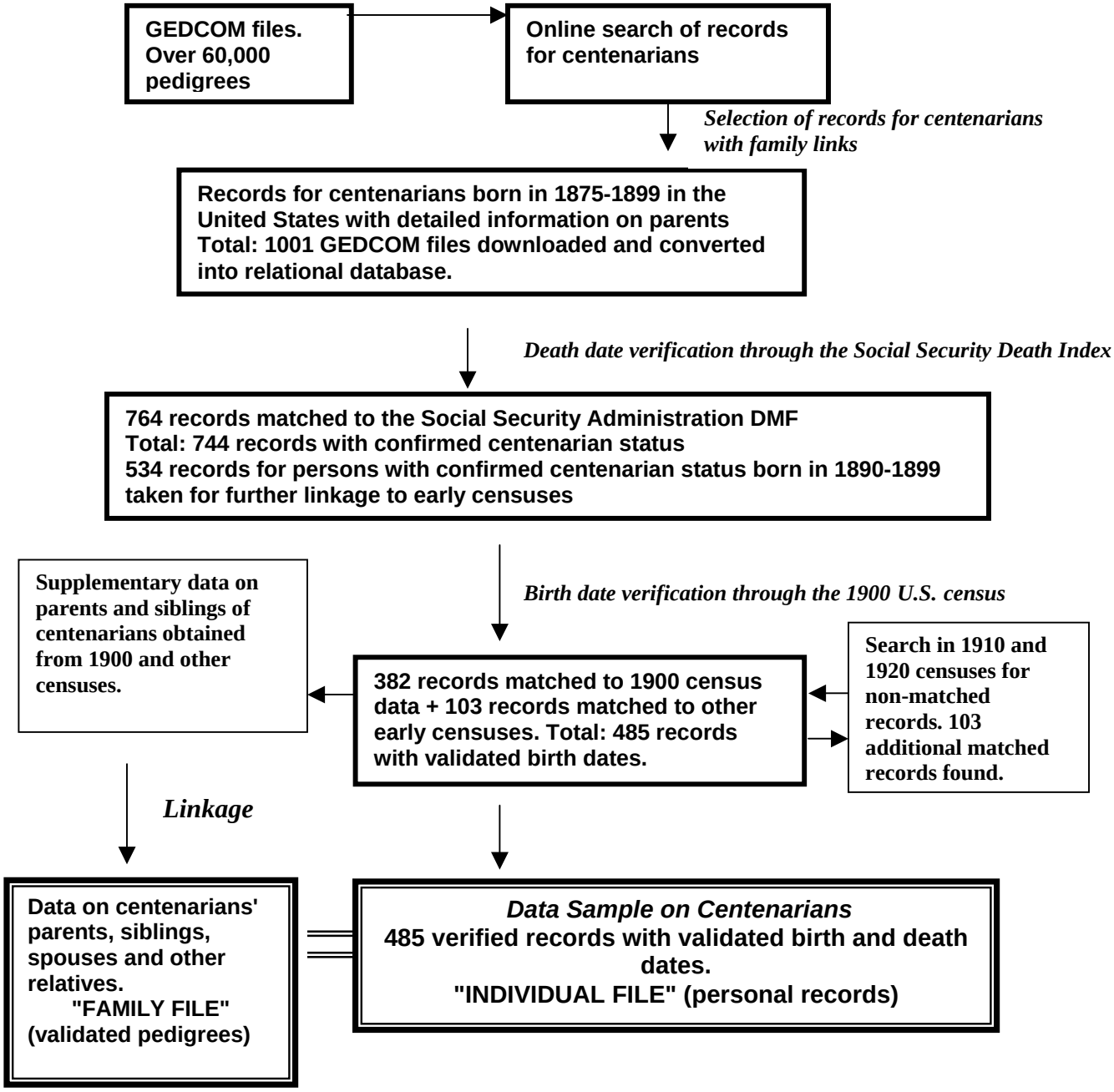

Figure 1. General Overview of Data Collection and Data Processing Protocol.

Beginning in the upper left, we searched online genealogical database (Ancestry.com). Then, records for centenarian individuals born in 1875-99 in the United States with detailed information on both parents and grandparents were downloaded for further verification and analysis.

All centenarians had verified dates of birth and death and known information for parents, siblings, spouses and other relatives. Table 7 below shows the age and sex breakdown of centenarians with verified ages. This developed database on long-lived persons combines information on family characteristics with data on the early-life conditions taken from the 1900-1910 U.S. censuses. This database was used to test a number of hypotheses on the factors affecting exceptional longevity (to be seen later).

TABLE 7
Distribution of Centenarians Confirmed Through the Social Security Administration Death Master File and Early U.S. Censuses by Age and Sex

Age at Death Reported in Genealogy	Males	Females
100	40	102
101	34	134
102	28	83
103	9	37
104	2	7
105	1	0
106	1	6
107	0	1
Total	115	370

As a result of this verification study, we may conclude that the quality of available computerized genealogies is good enough to conduct scientific studies when these genealogies satisfy the criteria of detailed birth and death dates reporting. Thus, such data may serve as a valuable source of information on determinants of exceptional longevity.

During the centenarian birth date verification process, we also tested a suggestion that deceased elder siblings of the same name might be incorrectly cited as centenarians in genealogies. Cases of "identity thefts" are well known in centenarian studies. For example, Pierre Joubert, who appeared in the Guinness Book as a 113-year-old man, in reality died at 65 years, whereas his namesake—his son—died 48 years later (see Jeune, Vaupel, 1999). Such scenario, however, is highly unlikely when detailed genealogies are available, and it was a Canadian genealogist and demographer, Hubert Charbonneau, who demystified the Pierre Joubert case using genealogical methods. In complete and detailed genealogies, this scenario of "identity theft" looks highly unlikely. Almost all genealogies with families having deceased children (88 percent) reported all children, including those who died in infancy. Only in two out of 198 such

families was the younger child named after his or her elder sibling (and this younger sibling was not a centenarian in both cases). Thus, appearance of centenarian with false identity in genealogy should involve a combination of three relatively rare events: naming child after deceased elder sibling, non-reporting of deceased child in genealogy and survival of sibling to advanced age (even younger sibling should become at least octogenarian or nonagenarian). Thus, it seems that "identity theft" by centenarians is not a likely phenomenon in detailed and complete genealogies.

Methodological Approaches to the Studies of Exceptional Longevity Using Genealogical Data

Computerized genealogies contain important information about family and life-course events, which otherwise is difficult to collect: lifespan of parents and other relatives, number and sex of siblings, birth order, ages of parents when person was born, age at marriage, number of spouses and lifespan of spouses and other non-blood relatives, number and sex of children and timing of their birth, place of birth and information about residence during the life-course (derived using places of birth for siblings and children).

This study demonstrated that quality of computerized genealogies is good enough for conducting scientific research, if only the detailed and complete genealogies are selected. If birth dates and death dates of persons, as well as their parents, are available in genealogy, then such genealogies might be considered as a basis for further studies. We found that the quality of birth dates reporting in genealogies is particularly high. Frequency of serious misprints in death dates is higher, although even in this case it is close to 1 percent only. Internal consistency check is a good way to eliminate potential misprints in genealogies, and all cases of extreme longevity require validation.

Studies of exceptional longevity using genealogical data require choice of appropriate control group. One approach is to use population-based control group. We applied this approach in studies of early-life conditions and survival to age 100 (see next section). A more elaborated approach uses as a control group either distant blood relatives (e.g., first cousins) or non-blood relatives (e.g., in-laws). In this case we eliminate unobserved shared factors and focus our study on specific effects (like number of children born, lifespan of parents, etc.).

Data from early censuses linked to computerized genealogies add additional important information about conditions during person's childhood. In this study we compared data for centenarians with population-based control. This approach allowed us to check the effect of early place of residence on the chances of survival to advanced

ages. In addition to that, there might be other approaches for choice of control group. One approach is to select a family of neighbors enumerated on the same or adjacent page of early census, which has a child of the same age and to use this family as a control. Another approach is to take a control group from genealogy (e.g., sisters- or brothers-in-law) and to link these individuals to early censuses as well. In both cases we relax a problem with potential bias caused by selection of genealogies but lose an opportunity to study geographical effects of early residence (cases and controls have the same or almost the same place of residence).

The main focus of this study was on testing the quality of genealogical data, so that the examples presented in the next section are mere illustrations of potential uses of information available in computerized genealogies.

II. Survival to Age 100 and Beyond: Evidence From Centenarian Genealogies Linked to Early U.S. Censuses

Our combined genealogical and census-related database was used in the analyses of several factors, which might be potentially important for survival to advanced ages. In particular, we focused our attention on the role of early-life conditions in determining later-life survival.

The idea of fetal origins of adult degenerative diseases and early-life programming of late-life health and survival is being actively discussed in the scientific literature (Lucas, 1991; Gavrilov, Gavrilova, 1991; 2003a; Barker, 1998; Kuh, Ben-Shlomo, 1997; Lucas *et al*., 1999; Costa, Lahey, 2003). The historical improvement in early-life conditions may be responsible for the observed significant increase in human longevity through the process called "technophysio evolution" (Fogel, Costa, 1997). Additional arguments suggesting the importance of early-life conditions in later-life health outcomes are coming from the reliability theory of aging and longevity (Gavrilov, Gavrilova, 1991; 2001a; 2003a). According to this theory, biological species (including humans) are starting their lives with extremely high initial load of damage, and, therefore, they should be sensitive to early-life conditions affecting the level of initial damage (Gavrilov, Gavrilova, 1991; 2001a; 2004).

The concept of high initial damage load also predicts that early life events may affect survival in later adult life through the level of initial damage. This prediction proved to be correct for such early-life indicators as parental age at a person's conception (Gavrilov, Gavrilova, 1997, 2000, 2003b; Gavrilova et al., 2003) and the month of person's birth (Gavrilov, Gavrilova, 1999, 2003b, Gavrilova et al., 2003; Doblhammer, 1999; Doblhammer, Vaupel, 2001; Costa, Lahey, 2003). There is mounting evidence now in support of the idea of fetal origins of adult degenerative diseases (Barker, 1998; Kuh, Ben-Shlomo, 1997; Lucas, Fewtrell, Cole, 1999), and early-life programming of aging and longevity (Gavrilov, Gavrilova, 1991, 2001, 2003a; Gavrilova et al., 2003).

Studies of Within Family Effects: Possible Links Between Birth Order and Exceptional Longevity

Information about birth order of centenarians allowed us to test a hypothesis whether the centenarians are distributed randomly within a sibship (brothers and sisters in the family) or not. If centenarian's birth order is determined by chance only and is not linked to exceptional longevity, then the ratio of [*centenarian birth order/(family size + 1)*] should be equal to 0.5 on average. If centenarians are found more often among

the older or among the younger siblings, then the observed ratio, named "centenarian birth order ratio" (CBOR), should demonstrate a statistically significant deviation from the expected value of 0.5.

Our previous studies on European aristocratic families lead us to a prediction that there might be a sex-specific link between person's birth order and exceptional longevity. Specifically, women are expected to be more likely to become centenarians, if they are born earlier compared to other siblings, when their parents are relatively young. This effect should be limited to women only and not observed among men. These predictions follow from our earlier published findings that daughters conceived to older farther live shorter lives, while sons are not affected by the fact of their late conception (Gavrilov, Gavrilova, 2000; Gavrilov et al., 2003b; Gavrilova et al., 2003). Thus, if this prediction is correct, then the ratio of [*centenarian birth order/(family size + 1)*] should be significantly below 0.5 in females, but not in males. Note that testing this hypothesis is based on within-family analysis, which allows us to relax concerns over possible confounding effects of many other predictor variables that are fixed within each family (like parental lifespan, etc.).

According to our database, the centenarians are usually born in rather large families with mean number of children equal to 7.17 ± 0.17 (data on 392 families). For comparison, the mean number of children in the total sample of 29,118 families taken from the same genealogical sources and for the same historical period is 5.65 ± 0.02 children. Thus, centenarians tend to be born in larger families on average. Further studies are required to find out whether this is an important, meaningful finding or a trivial observation caused by ascertainment bias (according to probability theory, the chances that at least one child becomes a centenarian just by chance are increasing for larger families).

To study the birth order effects, we have to remove non-informative cases where family size was equal to one and cases with less reliable information on family size (few genealogies where family size was lower than reported in census).

The results of data analyses are presented in Table 8 below.

TABLE 8
Mean Centenarian Birth Order *Ratio* for Male and Female Centenarians

Gender of Centenarian	Number of Cases (Studied Families)	Mean Value of Birth Order Ratio	Standard Error	95% Confidence Interval	
Males	114	0.479	0.022	0.435	0.523
Females	370	0.446	0.012	0.423	0.470

Note that the centenarian birth order ratio for female centenarians is indeed lower (0.45 ± 0.01) than expected (0.5) and this effect is statistically significant ($P<0.01$). In other words, the birth order ratio of centenarian-women is 12 percent lower on average than it would be expected by pure chance (random uniform distribution for cases of exceptional longevity by birth order) Thus, female centenarians can be found less likely among later-born siblings conceived to relatively old parents. In contrast to females, the birth order ratio for centenarian-men is exactly equal to theoretically predicted value of 0.5, suggesting that birth order is less important for exceptional male longevity.

Similar results are obtained using another statistic named 'centenarian birth order difference': [*centenarian birth order - (family size + 1)*/2]. If centenarians are distributed randomly by birth order within a sibship (independently of their centenarian status), then this difference should be equal to zero on average. This is what we expect to find for centenarian-males, while this birth order difference should be negative for centenarian-women, if the tested hypothesis is correct. The results of data analyses are presented in Table 9.

TABLE 9
Mean Centenarian Birth Order *Difference* for Male and Female Centenarians

Gender of Centenarian	Number of Cases (Studied Families)	Mean Value of Birth Order Difference	Standard Error	95% Confidence Interval	
Males	114	-0.13	0.22	-0.55	0.30
Females	370	-0.50	0.11	-0.72	-0.28

Note that the mean value of the centenarian birth order difference for females is lower (-0.50 ± 0.11) than zero, and this difference is statistically significant (P<0.01). In other words, the birth order of centenarian-women is lower on average than it would be expected (if birth order is irrelevant for longevity), and this difference in absolute terms corresponds to the shift of 0.5 to lower birth order on average. Thus, there is a tendency for female centenarians to be born among the first half of siblings in the family. In contrast to centenarian-women, the birth order difference for centenarian-males is exactly equal to theoretically predicted zero value, suggesting that birth order is less important for exceptional male longevity.

Thus, women seem to be more likely to become centenarians, if they are born earlier compared to other siblings, when their parents are relatively young. The birth order of centenarian-women is 12 percent lower on average than it would be expected by pure chance (random uniform distribution for cases of exceptional longevity by birth order), which in absolute terms corresponds to the shift of 0.5 to lower birth order on average (statistically highly significant, P<0.01). 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 European aristocratic families that daughters conceived to older fathers live shorter lives, while sons are not affected by the fact of their late conception (Gavrilov, Gavrilova, 2000; Gavrilov et al., 2003b; Gavrilova et al., 2003).

The results presented above are based on summary statistics, which describe the overall shift in birth order ranking of centenarians relative to other siblings. It is also interesting to study this topic in more depth and to see how exactly the odds of living to 100 depend on birth order. For this purpose, we have applied a logistic regression model with binary outcome variable (becoming a centenarian or not) and with two predictor variables (birth order and family size) included in polynomial-fitting model with non-centenarian siblings born in the same time window used as a control group. We found that the best fit of the data both for males and females analyzed separately could be achieved for the following model:

Logit (Longevity odds ratio) = $a\,x + b\,x^2 + c\,z + d$,

where x is the birth order, z is a family size, and a, b, c, and d are the parameters of the polynomial regression model.

Other interaction terms between these predictor variables were found to be statistically insignificant and therefore were not included in the model. The effect of family size, parameter c, was negative both for males (-0.11 ± 0.05, p = 0.028) and females (-0.07 ± 0.02, p = 0.002), which indicates that the odds of longevity are in fact

decreasing in larger families. Further studies are required to find out whether this is a meaningful finding or a trivial consequence of ascertainment bias (the proportion of centenarians in family is bound to decrease with increasing family size, because other siblings are likely not to be centenarians).

Figure 2 presents the results of data analysis in a graphic form. It shows the dependence of odds to live to 100 as a function of person's birth order (as predicted by the fitted polynomial logistic model). The graphs are computed for a fixed family size set to an average level of seven children (which is not particularly important, because the family size influences only the vertical location of the curves rather than their shape because there is no interaction of family size with birth order).

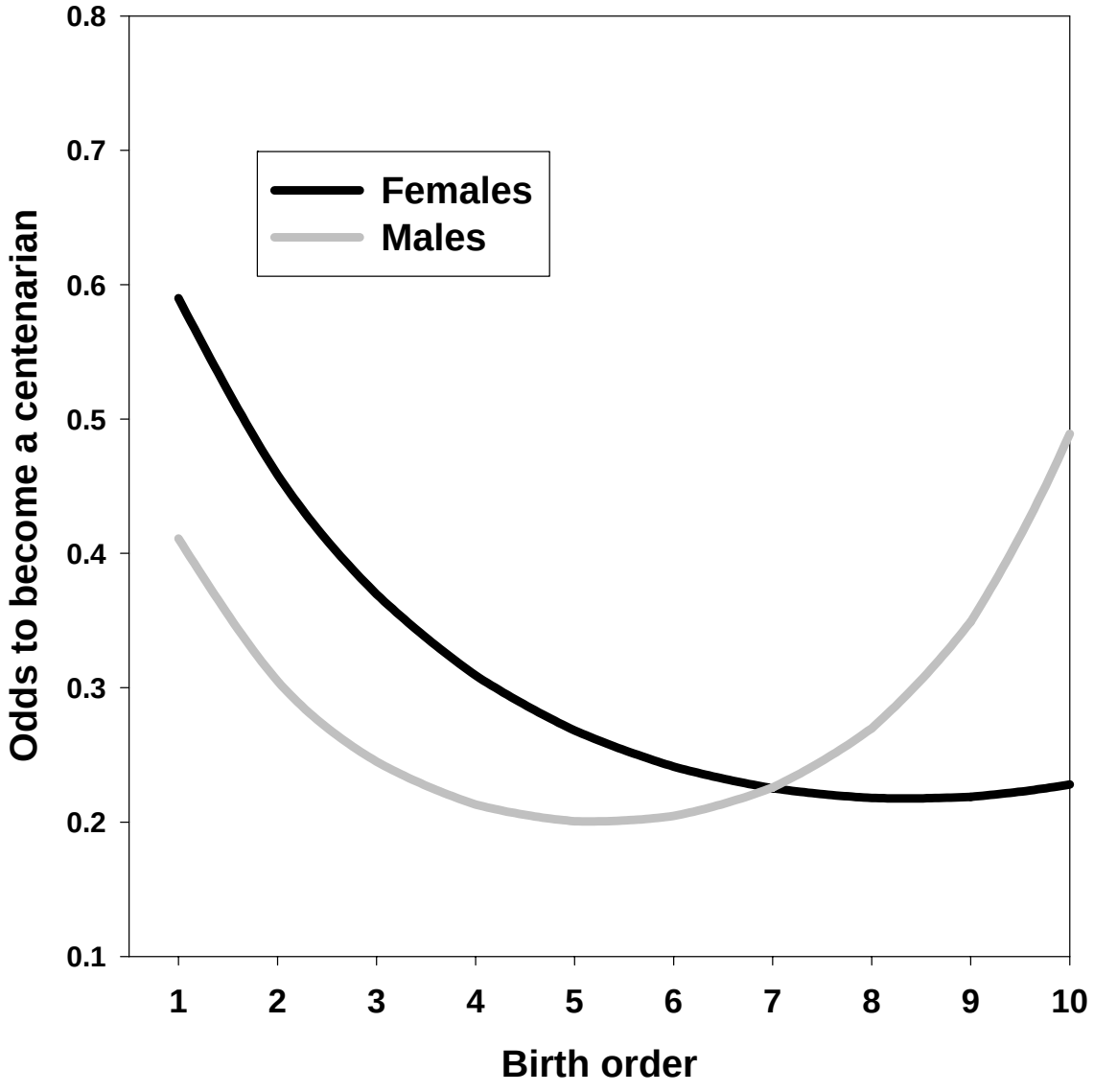

Figure 2. Dependence of Odds to Become a Centenarian on Person's Birth Order as Predicted by the Fitted Polynomial Logistic Model

Note that the odds of becoming a centenarian are decreasing with the birth order for females, which is consistent with the results of earlier data analysis based on summary measures. First-born daughters are three times more likely to survive to age 100 compared to later-born daughters of higher birth orders (7+). Note that the strongest effect of birth order is observed when it is relatively small—one to five (see Figure 2), and then the birth order effect fades out. In other words, it is good for female

longevity to be born among the first children, while for the last-born children the exact birth order is less important.

The picture is different for males; there is an unusual U-shaped curve for the odds of living to 100 in relation to the birth order. The chances for exceptional longevity are minimal for sons having a birth order of four to six compared to those born earlier or later. Thus the earlier studies based on summary measures, which found no birth order effect in males, seemed to overlook it, because of a complex U-shaped form of the birth order effects in males. In fact, the first-born sons are twice more likely to become centenarians compared to sons having birth order of four to six (see Figure 2). However, the last-born sons (birth order 9+) also have almost twice higher chances to survive to 100 years. It is needless to say that these preliminary findings need to be replicated with other methods and datasets. It also is obvious that these kinds of studies may have significant implications for actuarial science and practice. Interestingly, a recent study of Swedish cohorts found that children with higher birth orders have higher mortality, and this effect is particularly strong for females (Modin, 2002).

Living to 100:
Using the U.S. Census of Population Data to Study Early-Life Predictors of Longevity

The resulting dataset of 1900 and 1910 households linked to the centenarian genealogies allows us to make a comparison of these households to the general set of households enumerated in early censuses. We followed in part the methodological lines established by Preston et al., 1998, and used individual data from the 1900 U.S. census of population. The data are available as part of the Integrated Public Use Microdata Series from the University of Minnesota (Ruggles et al., 2004). The sample represents 1 percent of white households enumerated in 1900. Since the linkage to early U.S. censuses found that most of centenarians in our sample were whites (with exception of two American Indian families), we used a sample of white population from the IPUMS for comparison. At this initial stage of data analysis, we conducted a comparison of households, which raised a centenarian (linked to 1900 census) to the general sample of white households enumerated by 1900 census, which had children below age 10 (to make these households comparable to our set of centenarians who were born in 1890-1899 and hence were below age 10 in 1900).

We applied a method of multiple logistic regression (procedure 'logistic' in the Stata statistical package) in order to compare the two sets of households. Our assumption is that if early childhood conditions are important for survival to age 100, then the households of centenarians during their childhood would be different from the

general population. Tables 10 and 11 present results from multivariate logistic regression that estimates the odds for the household to be in the "centenarian" group. We conducted our analyses separately for male and female centenarians because our previous analyses demonstrated that men and women may respond differently to early-life conditions (Gavrilov, Gavrilova, 1999; 2003a).

The set of variables describing household is similar to one applied by Preston et al. (1998). We did not use the variable describing occupation of father because this variable is strongly correlated with ownership and farm status variables and because of possible problems in occupation classification. In fact, 63 percent of fathers of centenarians were farmers by occupation, almost all white-collar fathers owned their house and most low-skilled fathers were renters.

TABLE 10
Odds for Household to be in the "Centenarian" Group for Selected Characteristics in the 1900 U.S. Census. Female Centenarians

Characteristic	Odds Ratio	p-Value	95% Confidence Interval	
Census Region:				
New England and Middle Atlantic	1.00 – reference level			
Mountain West and Pacific West	3.16	0.000	1.81	5.52
Southeast and Southwest	2.05	0.002	1.30	3.23
North Central	2.42	0.000	1.58	3.70
Characteristics of Father				
Immigration Status				
Father immigrated	0.70	0.035	0.50	0.98
Father native-born	1.00 – reference level			
Literacy				
Father literate (can write)	1.29	0.352	0.76	2.19
Father illiterate	1.00 – reference level			
Survival of Siblings:				
All mother's children survived	1.02	0.917	0.75	1.37
71-99% of children survived	1.00 – reference level			
Less than 70% of children survived	0.85	0.434	0.57	1.27
Household Properties:				
Owned farm	1.00 – reference level			
Rented farm	0.63	0.007	0.45	0.88
Owned house	0.62	0.003	0.45	0.85
Rented house	0.26	0.000	0.18	0.37

TABLE 11
Odds for Household to be in the "Centenarian" Group for Selected Characteristics in the U.S. 1900 Census. Male Centenarians

Characteristic	Odds Ratio	p-Value	95% Confidence Interval	
Census Region:				
New England and Middle Atlantic	1.00 – reference level			
Mountain and Pacific West	2.68	0.041	1.04	6.90
Southeast and Southwest	1.11	0.797	0.51	2.41
North Central	1.39	0.372	0.67	2.89
Characteristics of Father:				
Immigration Status				
Father immigrated	0.40	0.019	0.19	0.86
Father native-born	1.00 – reference level			
Literacy				
Father literate (can write)	1.39	0.579	0.50	3.87
Father illiterate	1.00 – reference level			
Survival of Siblings:				
All mother's children survived	0.91	0.734	0.51	1.54
71-99% of children survived	1.00 – reference level			
Less than 70% of children survived	0.93	0.848	0.45	1.91
Household Properties:				
Owned farm	1.00 – reference level			
Rented farm	0.60	0.193	0.33	1.11
Owned house	0.28	0.001	0.13	0.58
Rented house	0.20	0.000	0.10	0.40

Data presented in Tables 10 and 11 demonstrate that region of residence and household property are the two most significant variables that affect chances to fall into group of "centenarian" households for both male and female centenarians. We would assume that these chances are related to the chances of survival to age 100. Thus, spending childhood in Mountain Pacific and West Pacific regions may highly increase chances of long life (by a factor of 3) compared to the North Eastern part of the country. Also farm (particularly owned farm) residence results in better survival to advanced ages. This result is consistent with studies of childhood conditions and survival to age 85+ (Preston et al., 1998; Hill et al., 2000). These earlier studies also based on linkage to early censuses demonstrated significant advantage in survival for children living on farm for both African Americans (Preston et al., 1998) and native-born whites (Hill et al., 2000). On the other hand, Northeast and Midwest were found to be the best regions

for survival to age 85+ (Hill et al., 2000). Both above-mentioned studies of childhood conditions and later survival found that father's illiteracy significantly declines chances of survival to age 85+. We find no such relationship for survival to age 100.

Having a father immigrant decreases chances to become a centenarian for males and females, although for females the effect is weaker. Similar negative effect of father's immigrant status was found for native-born whites, both sexes combined (Hill et al., 2000). These findings do not support a hypothesis that healthier immigrants have healthier children, thereby explaining lower old-age mortality in the United States compared to other developed countries (Manton, Vaupel, 1995). Recently Costa and Lahey (2003) came to the same conclusion that immigration status is not related to a better health. Finally, we found that deaths of siblings had no statistically significant effect on the chances to be a centenarian. Previous study found that death of siblings decreases chances to survive by age 85 years among African Americans (Preston et al., 1998), using more sophisticated methods of child mortality estimates and selection of control group. In this study we compare households where centenarians were raised, with a general population.

In general, our results support the idea that early childhood conditions might be important for survival to advanced ages (Gavrilova et al., 2003; Costa, Lahey, 2003).

Although our findings agree with previous reports (Preston et al., 1998; Hill et al., 2000), we should admit certain limitations of our study. Comparison with population samples assumes that differential survival is the only cause of differences between cases and controls. In our case computerized genealogies of good quality do not represent a random sample of population. Absence of black centenarians is one obvious bias of our sample, which can be explained by difficulties in genealogy compiling for African Americans because of paucity of historical information for blacks, lower popularity of this genealogical activity among African Americans and lower attention of African Americans to date and age recording (see Hill et al., 1995), which selected out potential African American genealogies during our initial screening. For other studied variables, the possibility of bias is not so certain. The proportion of genealogies compiled for families originated from the New England and Middle Atlantic regions by no means is lower than for families originated from the Western region. There is no reason to believe that many other household characteristics are different for families covered by genealogies and the general white population. The definite answer to this question could be obtained by comparison of computerized genealogies for "normal" (non-centenarian) individuals with population characteristics drawn from the IPUMS database, a study which we hope could be conducted in the future.

Living Beyond Age 100:
Early Life Conditions and Survival Beyond Age 100

Our database on centenarians has individuals with verified lifespans from 100 to 107 years (see Table 7). Several claims of extraordinary longevity (over 107 years) were initially found in genealogies but they were not verified using the Social Security Administration DMF and the early U.S. censuses. The data sample for centenarians contains a number of variables describing early childhood conditions taken from early U.S. censuses. Other variables for regression analysis were taken from genealogy: centenarian's month of birth, paternal and maternal lifespan, paternal and maternal ages at person's birth, family size (number of siblings) and birth year. We found that all variables describing childhood conditions in early censuses have no effect on survival beyond age 100. Although months of birth show some effects on survival after age 100, the overall model was not statistically significant. These results demonstrate that mortality at such extreme ages is less dependent on the past and appears to be more sensitive to the current conditions. Study of extinct birth cohorts for centenarians born before 1890 could shed more light on this problem (see Section III of this paper).

Comparison of Familial and Sporadic Centenarians

In this exploratory study we also tested a hypothesis that those centenarians who have family history of longevity (both parents lived over 80) are different from "sporadic" centenarians whose parents both lived less than 80. This hypothesis is based on the idea that persons having long-lived parents may be less vulnerable to adverse environmental conditions (including early-life conditions) because of favorable genetic background. As a result, offspring of long-lived individuals may survive to advanced ages despite childhood conditions being the same as in the average population. We found that both groups of centenarians do not differ from each other regarding the childhood conditions reported in early U.S. censuses as well as month of birth, birth order and family size (number of siblings). These analyses were conducted using logistic regression method with belonging to the group of familial centenarians used as a dependent variable (data not shown). Thus, farm background and other early-life conditions found in this study are equally important for both groups of centenarians.

III. Survival to Age 100 and Beyond: Evidence from the SSA Death Master File

The collection of records from the Social Security Administration Death Master File undertaken in this validation study had some interesting ramifications for mortality analyses at advanced ages.

In this study we collected information from the DMF on persons who lived 80 years and over and died before 2004. Total number of records collected is 9,014,591, including 924,222 records for persons lived 100 years and over. The information contained in this file is interesting not only for verification purposes but also for mortality estimates at advanced ages. Several birth cohorts (born in 1882-1891) may be considered as extinct or almost extinct, so it is possible to estimate mortality kinetics at very advanced ages up to 115-120 years. DMF database is unique in this regard because it represents mortality experience for one of the largest cohort of the oldest-old persons, which is readily available for survival analysis. Although the National Center for Health Statistics' National Death Index (NDI) provides superior coverage of deaths, its use is restricted and expensive, so for many researchers the DMF may be an appropriate choice (Hill, Rosenwaike, 2001).

Hazard Rate Estimation at Advanced Ages

It is now considered as an established fact that mortality at advanced ages has a tendency to deviate from the Gompertz law, so that the logistic model is often used to fit mortality (Horiuchi, Wilmoth, 1998). Actuaries including Gompertz (1825) himself first noted this phenomenon and later proposed a logistic formula for mortality growth with age in order to account for mortality fall-off at advanced ages (Perks, 1932; Beard, 1959; 1971). Greenwood and Irwin (1939) provided a detailed description of this phenomenon in humans and even made the first estimates for the asymptotic value of human mortality. According to their estimates, the mortality kinetics of long-lived individuals is close to the law of radioactive decay with half-time approximately equal to one year.

The same phenomenon of "almost non-aging" survival dynamics at extreme old ages is detected in many other biological species (Sacher, 1966; Economos, 1979; 1980; 1983; 1985; Curtsinger et al., 1992; Carey et al., 1992; Vaupel et al., 1998). In some insects, mortality plateau can occupy a sizable part of their life (Carey et al., 1992). The existence of mortality plateaus is well established for a number of lower organisms, mostly insects. In the case of mammals, data are much more controversial. Although Lindop (1961) and Sacher (1966) reported short-term periods of mortality deceleration in mice at advanced ages and even used Perks formula in their analyses, Austad (2001)

recently argued that rodents do not demonstrate mortality deceleration even in the case of large samples. Study of baboons found no mortality deceleration at advanced ages (Bronikowski et al., 2002). In the case of humans, this problem is not yet resolved completely. Data for extremely long-lived individuals are scarce and subjected to age exaggeration. In order to obtain good-quality estimates of mortality at advanced ages, researches are forced to pool data for the several calendar periods (like in the Kannisto-Thatcher database). Mortality deceleration observed in these data might be a result of data heterogeneity as was demonstrated by heterogeneity models (Beard, 1959; 1971). Thus, we need more research efforts to obtain reliable estimates of mortality at advanced ages.

The estimates of mortality force at extreme ages are difficult because of small numbers of survivors to these ages in most countries. Traditional demographic methods of mortality analysis based on period life tables suffer from the well-known denominator problem. More accurate estimates of mortality at advanced ages can be obtained using the method of extinct generations (Vincent, 1951; Depoid, 1973). In the Kannisto-Thatcher database (Thatcher, 1999) mortality is estimated by the method of extinct generations and data are aggregated for several calendar periods in order to accumulate enough cases of survivors to older ages. This aggregation, however, creates a heterogeneous mixture of cases from different birth cohorts.

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 Δx for which it is calculated, which 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 interval does not follow this law (see Gavrilov and Gavrilova, 1991, and Le Bras, 1976). Thus, it turns out that the shape of age-dependence for q_x depends on the arbitrary choice of age interval. Also, by definition q_x is bounded by unity, which makes difficult the studies of mortality at advanced ages.

It seems that a more useful indicator for mortality studies at advanced age is instantaneous mortality rate or hazard rate, μ_x, which is defined as follows:

$$\mu_x = -\frac{dN_x}{N_x dx},$$

where N_x is the number of living individuals at age x.

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 (time^{-1}). It should also be noted that the famous law of mortality, the Gompertz law, was proposed for fitting the hazard rate 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; 1966):

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

This estimate is unbiased for slow changes in hazard rate if $\Delta x \Delta \mu_x << 1$ (Sacher, 1966).

A simplified version of Sacher estimate (for small age intervals equal to unity) often is used in the studies of mortality: $\mu_x = -\ln(1-q_x)$. This estimate is based on the assumption that hazard rate is constant over age interval.

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. Violation of these assumptions may lead to biased estimates of hazard rates calculated on annual basis. Fortunately, the narrowing the age interval from one-year to one-month period for estimation of hazard rates might help to relax this problem.

In this study we obtained monthly hazard rate estimates for single-year birth cohorts using data taken from the Social Security Administration Death Master File, which collects deaths for persons who received SSA benefits and covers over 90 percent of deaths occurred in the United States (Faig, 2001) and 93 percent to 96 percent of deaths of individuals aged 65 or older (Hill, Rosenwaike, 2001). 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 (see above). This data resource is also useful in mortality estimates for several extinct or almost extinct birth cohorts in the United States.

The last deaths in the DMF available at the RootsWeb Web site occurred in January 2004. We obtained data for persons died before 2004, because only two

individuals born in 1885-1891 (birth cohorts that we studied) died in 2004. Thus, 1885-1891 birth cohorts in this sample may be considered as extinct or almost extinct. Assuming that the number of living persons belonging to these birth cohorts in 2004 is close to zero, it is possible to construct a cohort survivorship curve. In the first stage of our analyses we calculated individual lifespan in completed months:

Lifespan in months = (death year – birth year) x 12 + death month – birth month

Then it is possible to estimate hazard rate for each month of age using standard methods of survival analysis (using Nelson-Aalen estimator of hazard rate). All calculations were done using Stata statistical package (procedures stset and sts). This program provides estimates of hazard rate per month's period. In order to obtain more common annual rates, we multiplied these estimates by 12. We estimated hazard rates for three single-year birth cohorts: 1885, 1889 and 1891.

Results of the hazard rate estimates for three birth cohorts (1885, 1889 and 1891) are presented in Figures 3-6.

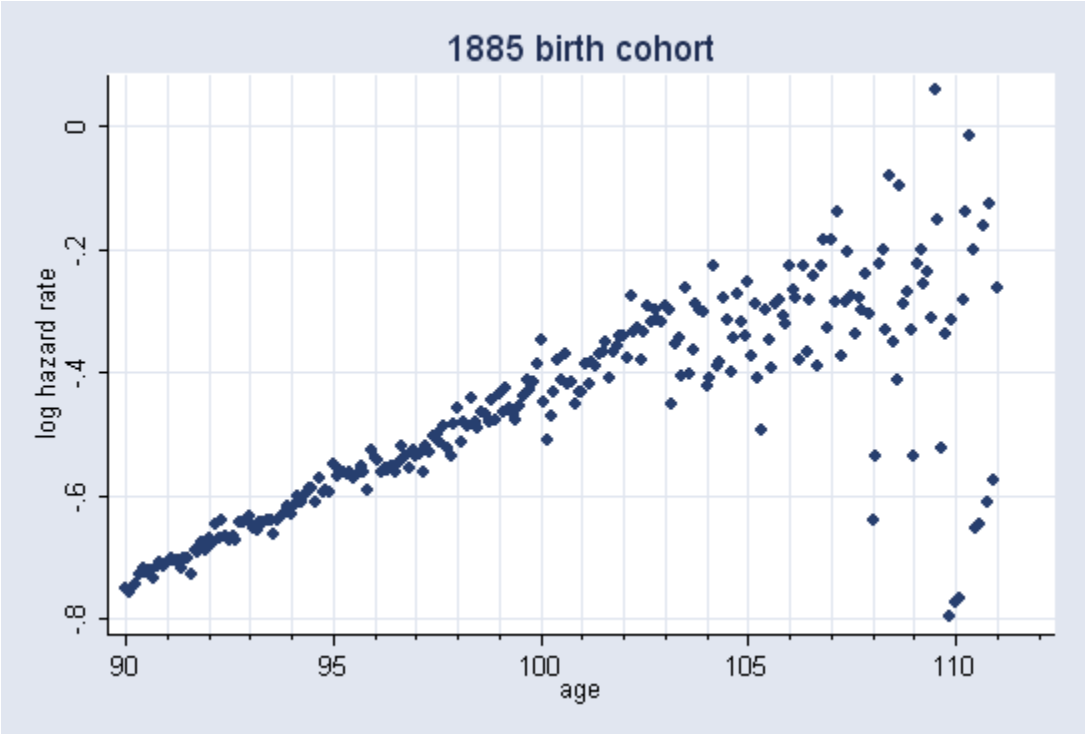

Figure 3. Hazard Rate (per year) for 1885 Birth Cohort
Data From the Social Security Administration Death Master File

Recent study of age validation among supercentenarians (Rosenwaike, Stone, 2003) showed that age reporting among supercentenarians in SSA database is rather accurate, with exception of persons born in the Southern states. In order to improve the quality of our dataset when estimating mortality rates, we excluded records for those persons who applied for Social Security number in the Southeast (AR, AL, GA, MS, LA, TN, FL, KY, SC, NC, VA, WV) and Southwest (AZ, NM, TX, OK) regions, Puerto Rico and Hawaii. This step of data cleaning however, did not change significantly the overall trajectory of mortality at advanced ages, but decreased the number of too-low mortality estimates and increased the number of higher-mortality estimates after age 105 years (see Figures 3-4).

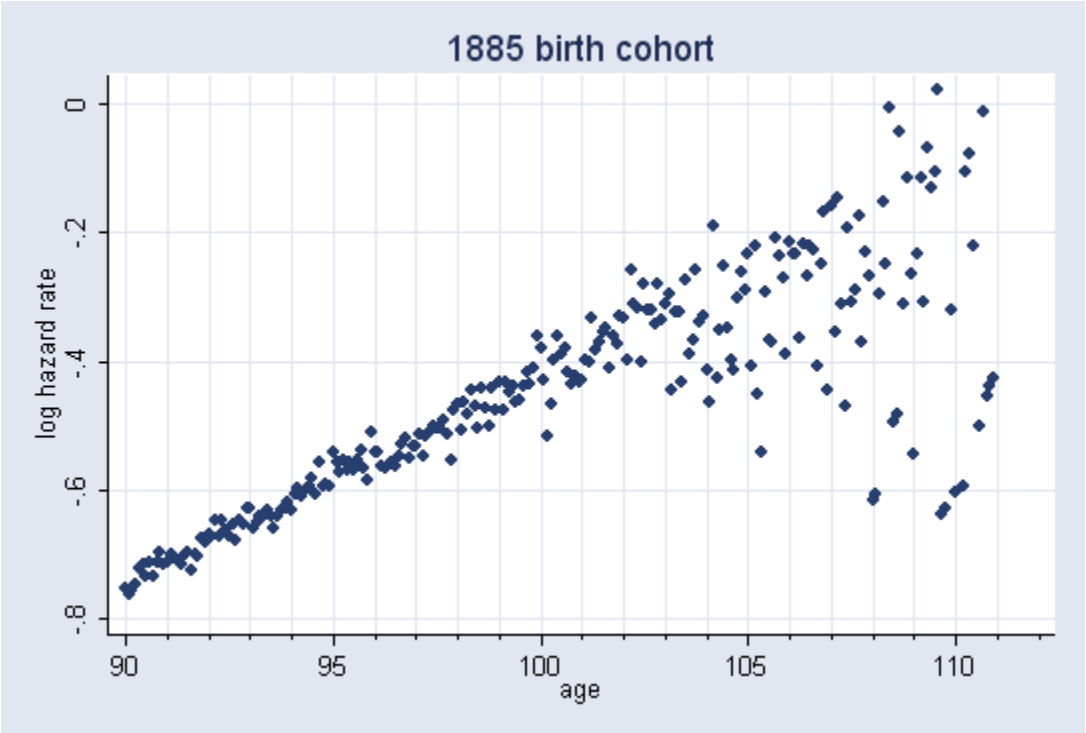

Figure 4. Hazard Rate (per year) for 1885 Birth Cohort
Data From the Social Security Administration Death Master File
Less Reliable Data for Southern States, Puerto Rico and Hawaii are Excluded

Note that from ages 85-89 up to ages 102-105 years, mortality grows steadily without obvious deceleration. Only after age 105 years mortality tends to decelerate, although high statistical noise makes mortality estimates beyond age 105 years less reliable. Also, for cohorts born after 1890, mortality over age 110 years is affected by data truncation. These figures demonstrate that for single-year birth cohort mortality agrees well with the Gompertz law up to very advanced ages. Previous studies of

mortality at advanced ages used aggregated data combining several birth cohorts with different mortality, and this aggregation apparently resulted in early mortality deceleration and subsequent leveling-off as it was demonstrated by heterogeneity model (Beard, 1971). Mortality deceleration and even decline of mortality often are observed for data with low quality. On the other hand, improvement of data quality results in straighter mortality trajectory in semi-log scale (Kestenbaum, Ferguson, 2001). In our study, more recent 1891 birth cohort demonstrates straighter trajectory and lower statistical noise after age 105 than older 1885 one (see Figures 4 and 6). 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. Testing this hypothesis now is hampered by the problem of data truncation for non-extinct birth cohorts.

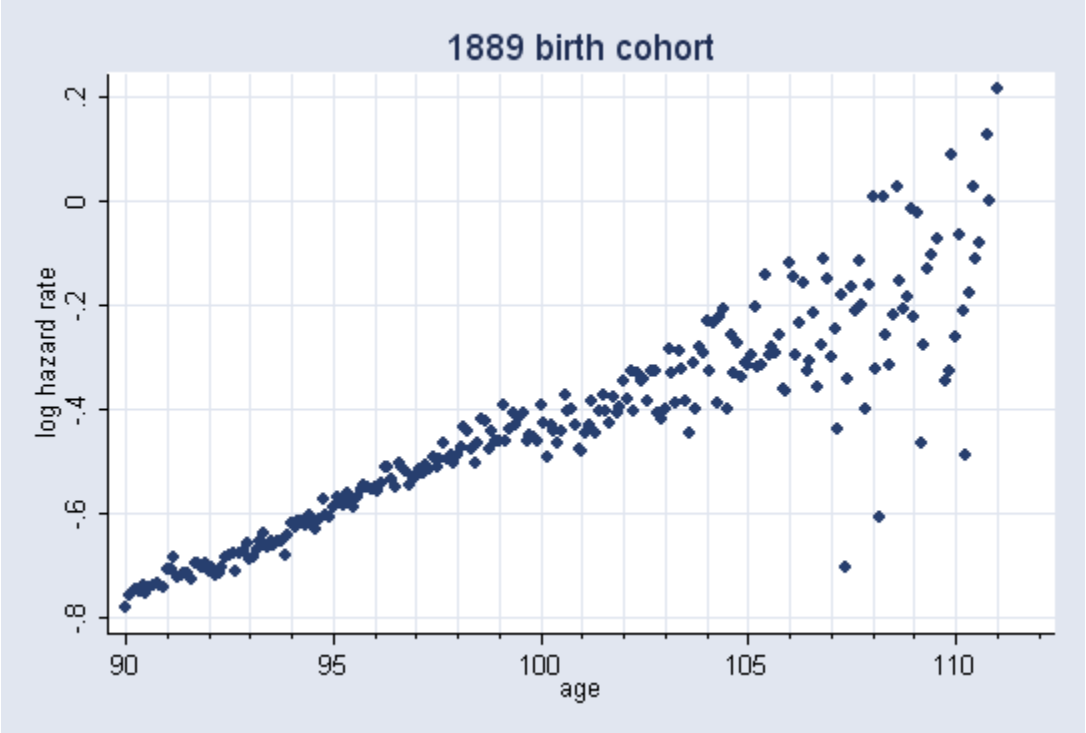

Figure 5. Hazard Rate (per year) for 1889 Birth Cohort
Less Reliable Data for Southern states, Puerto Rico and Hawaii are Excluded
Data From the Social Security Administration Death Master File

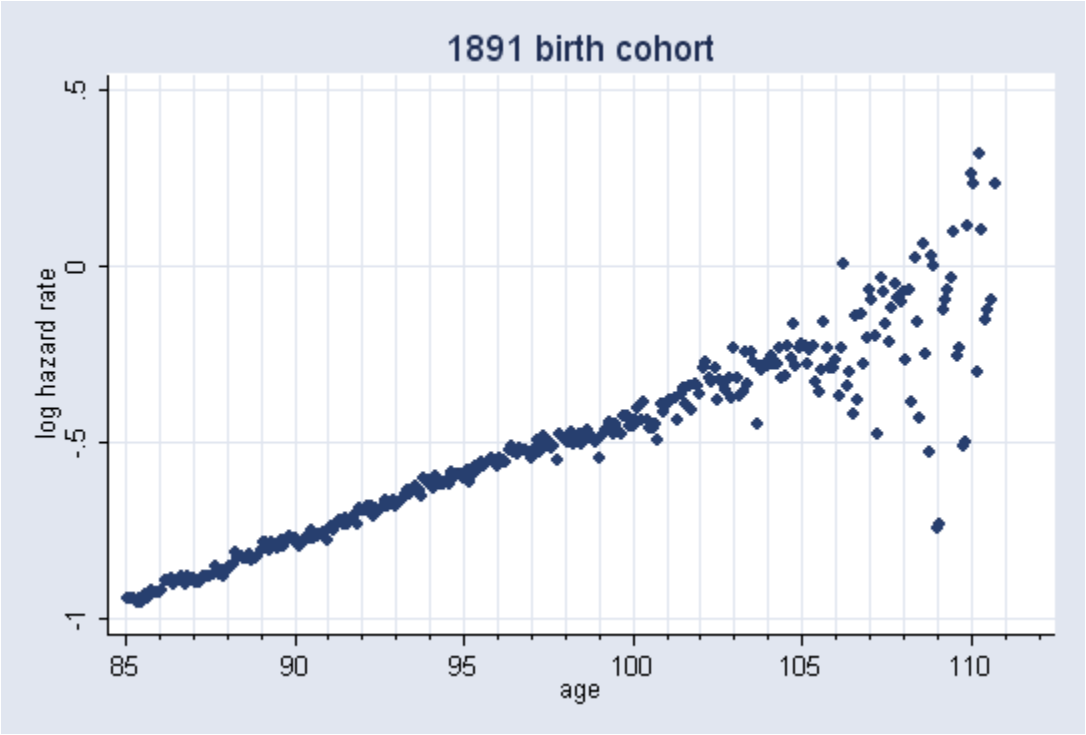

Figure 6. Hazard Rate (per year) for 1891 Birth Cohort
Less Reliable Data for Southern states, Puerto Rico and Hawaii are Excluded
Data From the Social Security Administration Death Master File

We already noted that the period of mortality deceleration in mammals is very short compared to lower organisms. It appears to be relatively short in humans, too. This observation agrees well with the prediction of reliability theory of aging, according to which more complex living systems/organisms with many vital subsystems (like mammals) may experience very short or no period of mortality plateau at advanced ages in contrast to simpler living organisms (Gavrilov, Gavrilova, 1991; 2001b; 2003a).

Month of Birth and Mortality at Advanced Ages

The Social Security Administration Death Master File data allow us to explore another interesting problem: the effects of early-life conditions and month-of-birth in particular on mortality at advanced ages. It was shown that month of birth has a significant effect on later-life mortality and lifespan (Gavrilov, Gavrilova, 1999; Doblhammer, Vaupel, 2001; Costa, Lahey, 2003). For example, Costa and Lahey (2003) used data on month of birth and mortality for the Union Army veterans at age 60-79 in 1900 and Americans of the same age in 1960-1980. They found that persons born in the second quarter had higher mortality than persons born in the fourth quarter (Costa, Lahey, 2003). Another study of month-of-birth effects on mortality in the United States

(Doblhammer, 2003) was based on the analyses of cross-sectional death certificates, which do not take into account the underlying structure of population exposed to risk. This approach could be justified only for stationary population with population structure constant over time. In real life this assumption usually is not valid, and the mean age at death (calculated from death certificates) is affected by temporal trends in population characteristics as well as temporal changes in seasonality of births and infant mortality. Because of these problems, using mean age at death as a proxy for life expectancy may lead to incorrect conclusions about better survival for low-educated or widowed persons (Doblhammer, 2003). More reliable estimates of mortality by month of birth could be obtained either by using death certificates in conjunction with population denominator data taken from censuses or by analyzing cohort mortality. In this regard, DMF containing cohort data may provide more accurate estimates of month-of-birth effects on mortality.

The Social Security Administration Death Master File contains data on month of birth for each person (in the overwhelming majority of cases), so it is possible to estimate life expectancy at age 80 years for each month of birth. In a cohort life table, mean lifespan (mean age at death in a cohort) is equivalent to the life expectancy, while this is not the case for period life tables. We use the term life expectancy here instead of mean age at death in order to avoid confusion with mean age at death calculated on the basis of cross-sectional (death certificates) data. In order to avoid possible truncation biases, we estimated life expectancy in the age range 80-110 years.

Figure 7 shows the effects of month of birth on life expectancy at age 80 for two birth cohorts: 1885 and 1891. Note that persons born in January have higher life expectancy at age 80 than persons born in April-June. Figure 8 confirms this observation for longer time period: practically all single-year birth cohorts born from 1885 to 1899 demonstrate the same monthly pattern in life expectancy. It is interesting that monthly pattern does not change for this relatively long 14-year calendar period. Thus, life expectancy at age 80 depends on month of birth: persons born in January (or December) live longer than persons born in other months and in April-June in particular. This seasonal pattern repeats in every birth cohort from 1885 to 1899.

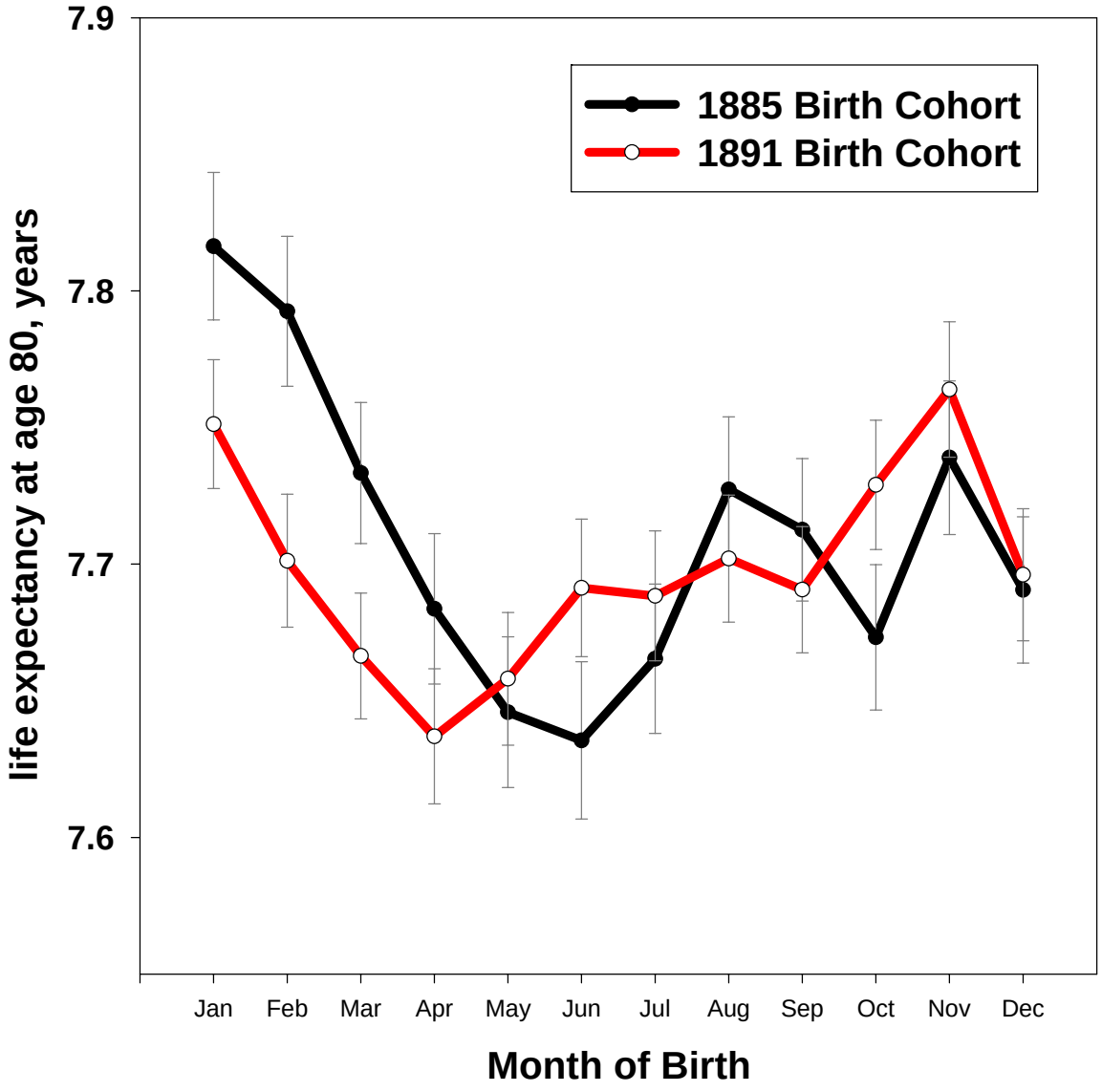

Figure 7. The Dependence of Life Expectancy at age 80 on Person's Month of Birth
Comparison of 1885 and 1891 Birth Cohorts
Data on Extinct Birth Cohorts Obtained From the
Social Security Administration Death Master File

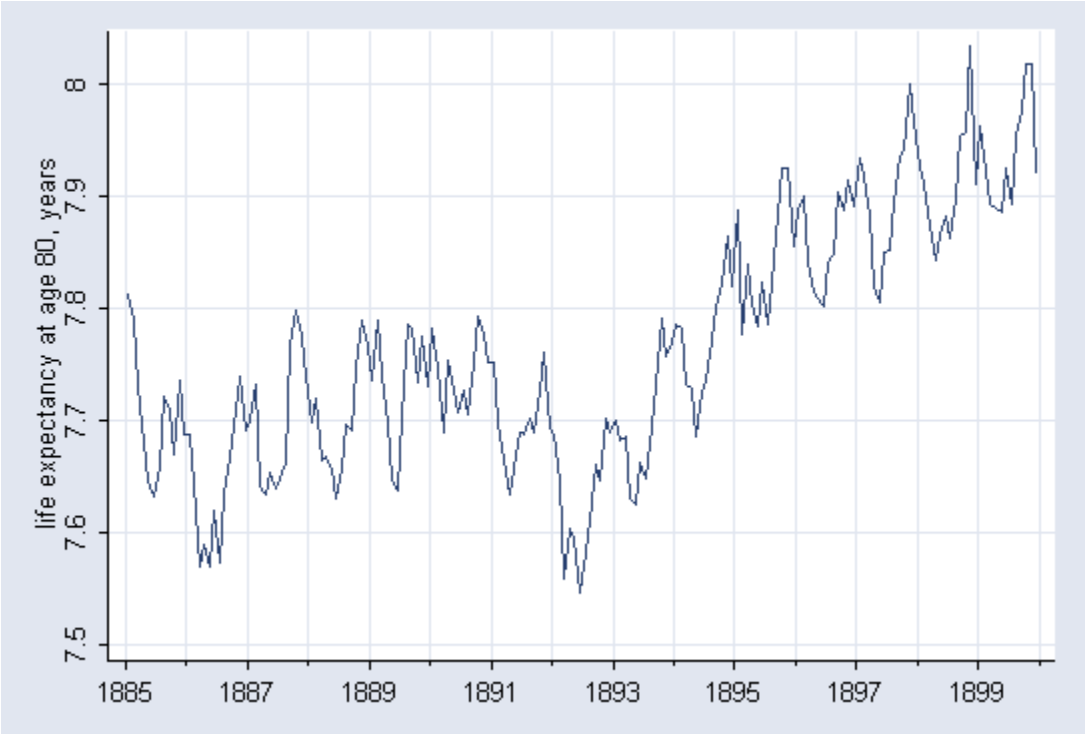

Figure 8. Periodic Seasonal Changes in Life Expectancy at age 80 for 1885-1899 Birth Cohorts Depending on Month of Birth

These seasonal patterns are consistent with earlier reports that persons born in the second quarter live shorter than persons born in the fourth quarter (Costa, Lahey, 2003). These monthly patterns also partially agree with a previous study based on aggregated death certificates, which found peak of mean age at death in September/October and trough in June/July (Doblhammer, 2003). Agreement with results obtained on the basis of cross-sectional data might indicate that effects of month of birth are indeed rather stable over time. This stability is evident at least for the 1885-1899 birth cohorts (Figure 8).

The fact that such an early circumstance of human life as the month of birth may have a significant effect 80 years later on the chances of human survival is quite remarkable. It indicates that there may be critical periods early in human life particularly sensitive to seasonal variation in living conditions in the past (e.g., vitamin supply, seasonal exposure to infectious diseases, etc.).

However, by age 100 this monthly pattern in life expectancy apparently disappears, indicating that centenarians indeed represent a selected population (Figures 9-10). We already found in our previous studies that month-of-birth pattern of survival depends on age, so that the overall monthly patterns might be different in different

periods of life. For example, in the study of 1800-1880 birth cohorts of European aristocracy, we found that lifespan at age 30 is particularly low for February-born women and higher for December-born ones (Gavrilov, Gavrilova, 1999). However this monthly pattern changed when lifespan at age 50 and over was analyzed (Gavrilov, Gavrilova, 2002).

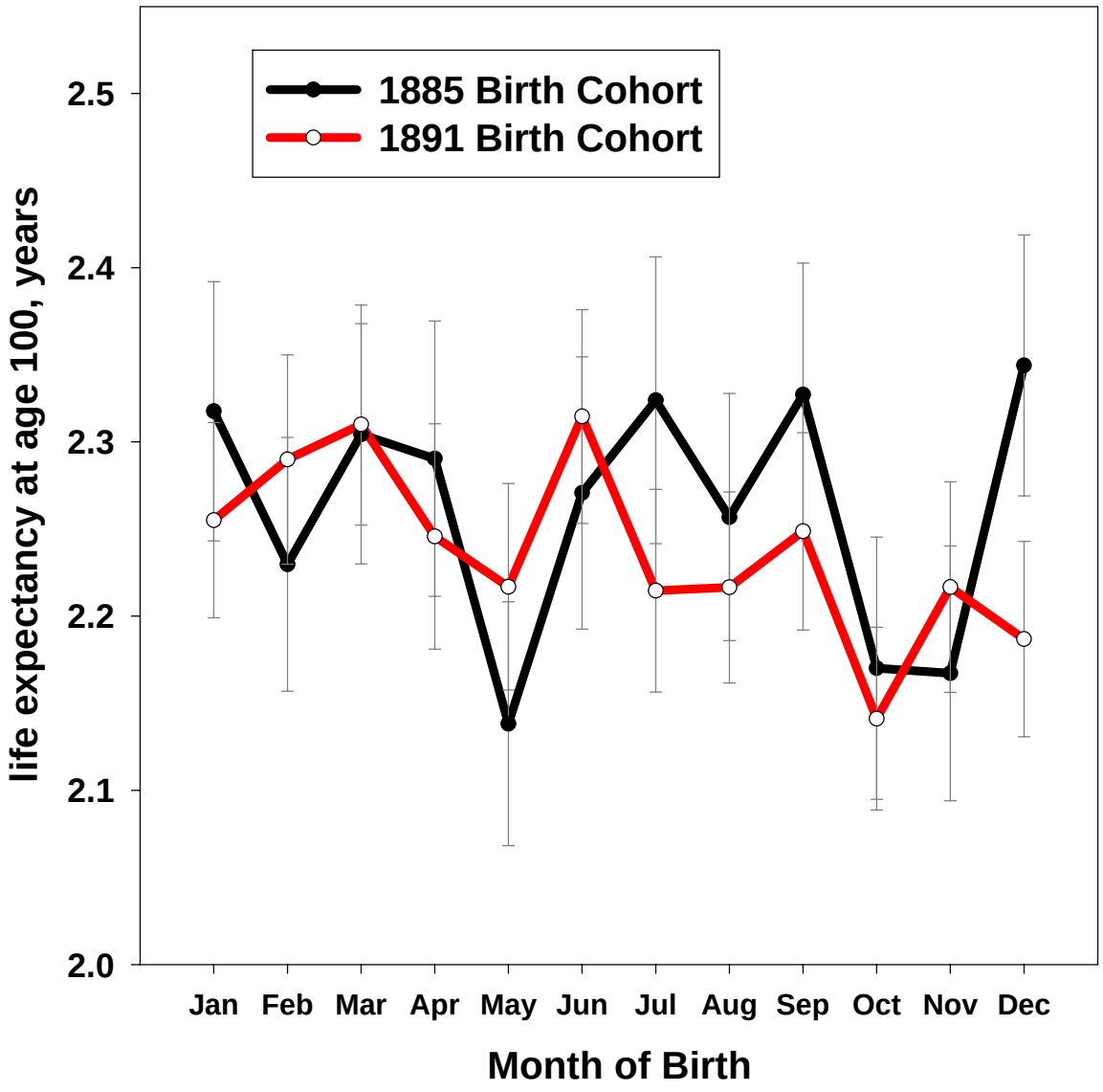

Figure 9. The Dependence of Life Expectancy at age 100 on Person's Month of Birth. Comparison of 1885 and 1891 Birth Cohorts

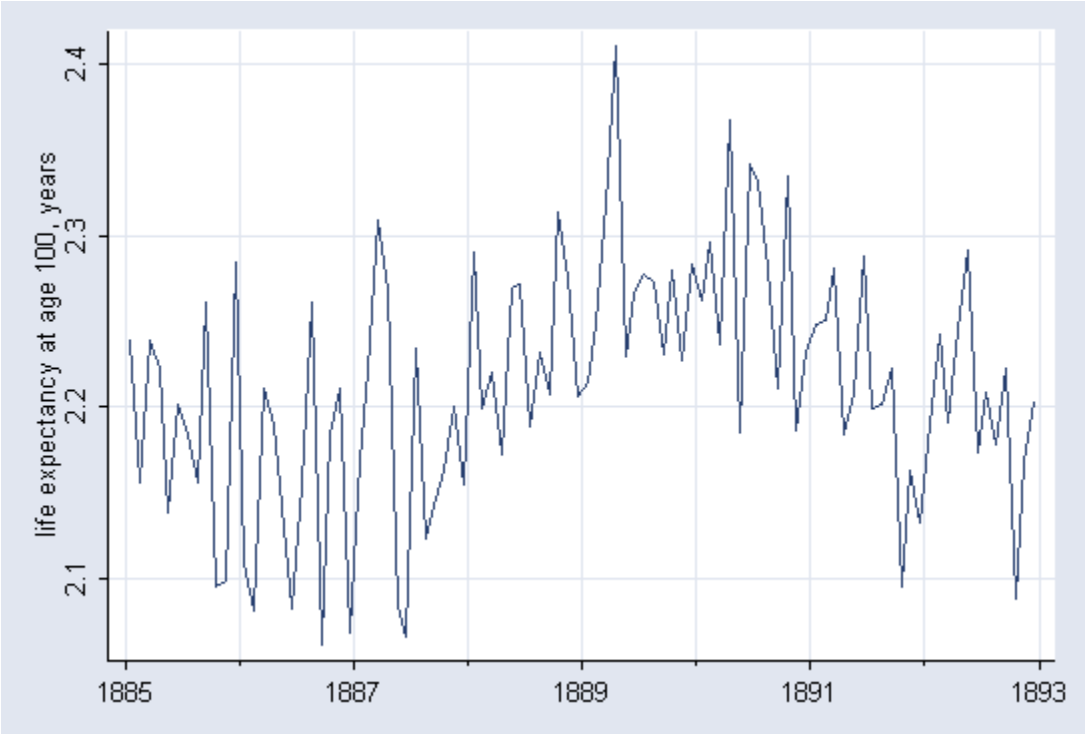

Figure 10. Life Expectancy at age 100 for 1885-1893 Birth Cohorts

The results obtained in this study are interesting but yet should be regarded with some caution. The Social Security Administration Death Master File provides no information about sex and race of decedents. Also, quality of data for earlier birth cohorts is lower than for more recent birth cohorts. Thus, we may expect that 5-10 years from now the quality of the SSA DMF data would be sufficient enough to obtain more accurate estimates of mortality at advanced ages.

Implications of This Study

This exploratory study has a number of interesting implications for actuarial science. In general, this study has demonstrated that an ongoing revolution in information technology and computer science has created new opportunities for actuarial studies on human longevity. Millions of individual records on human lifespan are now computerized and are available online (Social Security Death Master File, genealogical records, etc.). Moreover, detailed information for each member of the entire population of the United States has become available online in the form of images of the early U.S. censuses, including the most "recent" 1930 U.S. census.

This study has demonstrated how these rich information resources may be used for developing a reliable database for actuarial studies on human longevity. In this

exploratory study, we found that the best way to start the human longevity database development is to use first the family-linked data available in computerized genealogies.

We found that contrary to the common belief in poor quality of genealogical data, this information resource is highly valuable, if only we follow certain methodological guidelines uncovered in this study.

These methodological guidelines are:

1. To use only those genealogical records, which contain complete, exact and detailed dates of birth and death, place of birth, with information on parental names and their lifespan.
2. To use this genealogical data as a starting point only, subject to subsequent external validation with the Social Security Administration Death Master File, and the early U.S. censuses.

Perhaps most important, a particular procedure of data matching and cross-checking has been applied in practice, which produced a reliable dataset with several hundreds of family-linked records for individuals with exceptional longevity.

Now, when a working procedure of database development is in place, it could be applied in an industrial scale to get many thousands of family-linked records of exceptional human longevity, with obvious implications for actuarial science and practice.

Other implications of this study are related to the identified putative predictors of human longevity. It came as surprise that a geography of a birth place (or factors associated with it) within the United States seems to be such important determinants of human longevity. Our preliminary findings suggest that there may be a threefold difference in chances of survival to 100, depending on location of childhood residence. Two kinds of implications are important here. Methodological implication is that future studies should not be limited to a common practice of using geographically matched control group for comparison purposes, because this study design overlooks the importance of geographic factors. A substantive implication is that the mechanisms of this early-life location effect on human longevity need to be studied and understood, and the alternative trivial explanations (like selection bias) need to be excluded in future studies.

Acknowledgements

This study was made possible thanks to a generous support from the Society of Actuaries and a stimulating working environment at the Center on Aging, NORC/University of Chicago. We are grateful to Thomas Edwalds and Kenneth Faig for helpful comments and suggestions. We also are grateful to the members of the Project Oversight Group (POG) of the Society of Actuaries and to participants of the International Symposium "Living to 100 and Beyond: Survival at Advanced Ages" (Orlando, Florida, January 12-14, 2005) for providing useful comments and recommendations on the earlier draft of this manuscript.

References

Allison P. (1995). *Survival Analysis Using the SAS ® System: A Practical Guide*. SAS Institute.

Austad, S.N. (2001). "Concepts and Theories of Aging." In E.J. Masoro and S.N. Austad. *Handbook of the Biology of Aging*. San Diego, CA: Academic Press, 3-22.

Barker, D.J.P. (1998). "Mothers, Babies, and Disease in Later Life." (2nd ed.). London: Churchill Livingstone.

Beard, R.E. (1971). "Some Aspects of Theories of Mortality, Cause of Death Analysis, Forecasting and Stochastic Processes." In W. Brass (Ed.), *Biological Aspects of Demography* (pp. 57-68), London: Taylor and Francis.

Bronikowski, A.M., Alberts, S.C., Altmann, J., Packer, C., Carey, K.D. and Tatar, M. (2002). "The Aging Baboon: Comparative Demography in a Non-Human Primate." *Proc. Natl. Acad. Sci. U.S.A.*, 99, 9591-9595.

Carey, J.R., Liedo, P., Orozco, D. and Vaupel, J.W. (1992). "Slowing of Mortality Rates at Older Ages in Large Medfly Cohorts." *Science*, 258, 457-461.

Costa, D.L. and Lahey, J. (2003). "Becoming Oldest-Old: Evidence from Historical U.S. Data." *NBER Working Paper* No. W9933. http://ssrn.com/abstract=439614.

Curtsinger, J.W., Fukui, H., Townsend, D. and Vaupel, J.W. (1992). "Demography of Genotypes: Failure of the Limited Life-Span Paradigm in *Drosophila melanogaster*." *Science* 258, 461-463.

Depoid F. (1973). "La Mortalite des Grands Viellards." *Population*, 28: 755-92.

Doblhammer, G. (1999). "Longevity and Month of Birth: Evidence from Austria and Denmark." *Demographic Research* [Online] 1, 1-22. Available at: http://www.demographic-research.org/Volumes/Vol1/3/default.htm

Doblhammer, G. (2003). "The Late Life Legacy of Very Early Life." Rostock, *MPIDR Working Paper WP-2003-030*.

Doblhammer, G., Vaupel J.W. (2001). "Lifespan Depends on Month of Birth." *Proc. Natl. Acad. U.S.A.*, 98: 2934-2939.

Economos, A.C. (1979). "A Non-Gompertzian Paradigm for Mortality Kinetics of Metazoan Animals and Failure Kinetics of Manufactured Products." *AGE*, 2, 74-76.

Economos, A.C. (1980). "Kinetics of Metazoan Mortality." *J. Social Biol. Struct.*, 3, 317-329.

Economos, A.C. (1983). "Rate of Aging, Rate of Dying and the Mechanism of Mortality." *Arch. Gerontol. and Geriatrics*, 1, 3-27.

Elo, I.T. and Preston, S.H. (1992), "Effects of Early-Life Condition on Adult Mortality: A Review." *Population Index*, 58(2):186-222.

Elo, I.T., Preston, S.H., Rosenwaike, I., Hill, M. and Cheney, T.P. (1995). "Consistency of Age Reporting on Death Certificates and Social Security Administration Records among Elderly African-American Decedents." *Working Paper Series No. 95-03*. Population Aging Research Center, University of Pennsylvania.

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, FL.

Fogel, R.W. and Costa, D.L. (1997). "A Theory of Technophysio Evolution, with Some Implications for Forecasting Population, Health Care Costs, and Pension Costs." *Demography*, 34: 49-66.

Gavrilov, L.A. and Gavrilova, N.S. (1991). *The Biology of Life Span: A Quantitative Approach*. Harwood Academic Publisher, New York.

Gavrilov, L.A. and Gavrilova, N.S. (1998). "Inventory of Data Resources on Familial Aggregation of Human Longevity that can be Used in Secondary Analysis in Biodemography of Aging." Bethesda: National Institute on Aging. NIA Professional Service Contract #263 SDN74858. 35p.

Gavrilov, L.A. and Gavrilova, N.S., 1999. "Season of Birth and Human Longevity." *Journal of Anti-Aging Medicine*, 2: 365-366. Available at: http://longevity-science.org/Season-of-Birth.pdf

Gavrilov, L.A. and Gavrilova N.S. (2000). "Human Longevity and Parental Age at Conception." In: *Sex and Longevity: Sexuality, Gender, Reproduction, Parenthood* (J.-M.

Robine *et al*., eds), pp. 7-31. Berlin, Heidelberg: Springer-Verlag. Available at: http://longevity-science.org/Parental_Age_2000.pdf

Gavrilov, L.A. and Gavrilova, N.S. (2001a). "Biodemographic Study of Familial Determinants of Human Longevity." *Population, English Selection*, 13(1) 197-222. Available at: http://longevity-science.org/Biodemography-2001.pdf

Gavrilov, L.A. and Gavrilova, N.S. (2001b). "The Reliability Theory of Aging and Longevity." *J. Theor. Biol.* 213: 527-545. Available at: http://longevity-science.org/JTB-01.pdf

Gavrilov, L.A. and Gavrilova, N.S. (2002). "Early-Life Seasonal Programming of Adult Lifespan: Evidence from the 19th Century Birth Cohorts." Annual Meeting of the Social Science History Association, St. Louis, October 24-27, 2002. Available at: http://www.ssha.org/abstract2002/abs348.html

Gavrilov, L.A. and Gavrilova, N.S. (2003a). "The Quest for a General Theory of Aging and Longevity." *Science's* SAGE KE (Science of Aging and Knowledge Environment) for July 16, 2003; Vol. 2003, No. 28, 1-10. Available at: http://longevity-science.org/SAGE-KE-03.pdf

Gavrilov, L.A. and Gavrilova, N.S. (2003b). "Early-Life Factors Modulating Lifespan." In: Rattan, S.I.S. (Ed.). *Modulating Aging and Longevity*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 27-50. Available at: http://longevity-science.org/Early-Life-Factors-2003.pdf

Gavrilov L.A. and Gavrilova, N.S. (2004). "Early-Life Programming of Aging and Longevity: The Idea of High Initial Damage Load (the HIDL Hypothesis)." *Annals of the New York Academy of Sciences*, 1019: 496-501. Available at: http://longevity-science.org/HIDL-ANYAS-2004.pdf

Gavrilov, L.A., Gavrilova, N.S., Olshansky, S.J. and Carnes, B.A. (2002). "Genealogical Data and Biodemography of Human Longevity." *Social Biology*, 49(3-4): 160-173. Available at: http://longevity-science.org/Social-Biology-02.pdf

Gavrilova, N.S. and Gavrilov, L.A. (1999). "Data Resources for Biodemographic Studies on Familial Clustering of Human Longevity." *Demographic Research* [Online], Vol.1(4): 1-48.

Available at: http://www.demographic-research.org/volumes/vol1/s4

Gavrilova, N.S. and Gavrilov, L.A. (2001). "When does Human Longevity Start? Demarcation of the Boundaries for Human Longevity." *Journal of Anti-Aging Medicine*, 4**:** 115-124. Available at: http://longevity-science.org/JAAM-Boundaries-for-Human-Longevity.pdf

Gavrilova, N.S., Gavrilov, L.A., Evdokushkina, G.N. and Semyonova, V.G. (2003). "Early-Life predictors of Human Longevity: Analysis of the 19th Century Birth Cohorts." *Annales de Demographie Historique*, 2, 177-198. Available at: http://longevity-science.org/Early-Life-Predictors-2003.pdf

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 Irwin, J.O. (1939). "The Biostatistics of Senility." *Hum. Biol.*, 11, 1-23.

Hill, M.E. and Rosenwaike, I. (2001). "The Social Security Administration's Death Master File: The Completeness of Death Reporting at Older Ages." *Soc. Secur. Bull.*, 64: 45-51.

Hill, M.E., Preston, S.H., Elo, I.T. and Rosenwaike, I. (1995). "Age-Linked Institutions and Age Reporting among Older African Americans." Population Research Center, University of Pennsylvania, Working Paper Series No.95-05.

Hill, M.E., Preston S.H. and Rosenwaike, I. (2000). "Age Reporting among White Americans Aged 85+: Results of a Record Linkage Study." *Demography*, 37: 175-186.

Hill, M.E., Preston, S.H., Rosenwaike, I. and Dunagan, J.F. (2000). "Childhood Conditions Predicting Survival to Advanced Age among White Americans." Paper presented at the 2000 Annual Meeting of the Population Association of America, Los Angeles.

Horiuchi, S. and Wilmoth, J.R. (1998). "Deceleration in the Age Pattern of Mortality at Older Ages." *Demography*, 35: 391-412.

Iwashyna, T.J., Zhang, J.X., Lauderdale, D.S. and Christakis, N.A. (1998). "A Method for Identifying Married Couples in the Medicare Claims Data: Mortality, Morbidity, and Health Care Utilization among the Elderly." *Demography*, 35: 413-419.

Jeune, B. and Vaupel, J., eds. (1999). "Validation of Exceptional Longevity." *Odense Monographs on Population Aging,* Vol. 6. Odense: Odense Univ. Press.

Kestenbaum, B. (1992). "A Description of the Extreme Aged Population Based on Improved Medicare Enrollment Data." *Demography*, 29: 565-80.

Kestenbaum, B. and Ferguson, B.R. (2001). "Mortality of the Extreme Aged in the United States in the 1990s, Based on Improved Medicare Data." In: Proceedings of the Society of Actuaries "Living to 100 and Beyond" International Symposium, Orlando, FL.

Kuh, D. and Ben-Shlomo, B. (1997). *A Life Course Approach to Chronic Disease Epidemiology*. Oxford: Oxford University Press.

Le Bras, H. (1976). "Lois de Mortalité et Age Limité." *Population*, 31, 655-692.

Leon, D.A., Lithell, H.O., Vågerö, D., Koupilová, I., Mohsen, R., Berglund, L., Lithell, U.B. and McKeigue, P.M. (1998). "Reduced Fetal Growth Rate and Increased Risk of Death from Ischaemic Heart Disease: Cohort Study of 15000 Swedish Men and Women Born 1915-29." *Br. Med. J.* 317: 241-245.

Lindop, P.J. (1961). "Growth Rate, Lifespan and Causes of Death in SAS/4 Mice." *Gerontologia*, 5: 193-208.

Manton, K.G. and Vaupel, J.W. (1995). "Survival After Age of 50 in the United States, Sweden, France, England, and Japan." *The New England Journal of Medicine*. 333: 1232-5.

Modin, B. (2002). "Birth Order and Mortality: A Life-Long Follow-up of 14,200 Boys and Girls Born in early 20th Century Sweden." *Social Science & Medicine*, 54: 1051-1064.

Perks, W. (1932). "On Some Experiments in the Graduation of Mortality Statistics." *Journal of the Institute of Actuaries*, 63, 12-57.

Preston, S.H., Elo, I., Rosenwaike, I. and Hill, M. (1996). "African-American Mortality at Older Ages: Results of a Matching Study." *Demography*, 33: 193-209.

Preston, S.H., Hill, M. and Drevenstedt, G. (1998). "Childhood Conditions that Predict Survival to Advanced Ages among African Americans." *Soc. Sci. Med.*, 47: 1231-1246.

Rosenwaike, I. and Logue, B. (1983). "Accuracy of Death Certificate Ages for the Extreme Aged." *Demography*, 20: 569-85.

Rosenwaike, I., Hill, M., Preston, S. and Elo, I. (1998). "Linking Death Certificates to Early Census Records: The African American Matched Records Sample." *Historical Methods*, 31: 65-74.

Rosenwaike, I. and Stone, L.F. (2003). "Verification of the Ages of Supercentenarians in the United States: Results of a Matching Study." *Demography,* 40: 727-739.

Ruggles, S., Sobek, M., Alexander, T., Fitch, C.A., Goeken, R., Hall, P.K., King, M. and Ronnander, C. (2004). *Integrated Public Use Microdata Series (IPUMS): Version 3.0.* Minneapolis, MN: Minnesota Population Center. Available at: http://www.ipums.org

Sacher, G.A. (1966). "The Gompertz transformation in the Study of the Injury-Mortality Relationship: Application to Late Radiation Effects and Ageing." In P. J. Lindop and G. A. Sacher (eds.) *Radiation and Ageing* (pp. 411-441). Taylor and Francis, London.

Shrestha, L.B. and Preston, S. H. (1995). "Consistency of Census and Vital Registration Data on Older Americans: 1970-1990." *Survey Methodology*, 21: 167-177.

Thatcher, A.R. (1999). "The Long-Term Pattern of Adult Mortality and the Highest Attained Age." *J. R. Statist. Soc.* A, 162, Part 1, 5-43.

Vaupel, J.W., Carey, J.R., Christensen, K., Johnson, T., Yashin, A.I., Holm, N.V., Iachine, I.A., Kannisto, V., Khazaeli, A.A., Liedo, P., Longo, V.D., Zeng, Y., Manton, K. and Curtsinger, J.W. (1998). "Biodemographic Trajectories of Longevity." *Science*, 280, 855-860.

Vincent, P. (1951). "La mortalite des Viellards." *Population*, 6, 181-204.