November 11, 11:00pm

Paper accepted for presentation at
The 2001 Annual Meeting of the Social Science History Association
November 15-18, 2001
Chicago

Network: *Family Demography*
Session: *Early-life conditions and mortality in later life*

Early-life Predictors of Human Longevity

Natalia S. Gavrilova[*], Leonid A. Gavrilov[*],

Galina N. Evdokushkina[**], Victoria G. Semyonova[**]

(*) Center on Aging, NORC/University of Chicago, 1155 East 60th Street, Chicago, IL 60637

(**) Central Research Institute of Public Health and Informatics, Moscow, Russia.

Address for correspondence:
Dr. Natalia S. Gavrilova, Center on Aging
NORC/University of Chicago
1155 East 60th Street, Chicago, IL 60637
Fax: (773) 256-6313; Phone: (773) 256-6359
E-mail: nsgavril@midway.uchicago.edu

EARLY-LIFE PREDICTORS OF HUMAN LONGEVITY

Natalia S. Gavrilova, Leonid A. Gavrilov,
Galina N. Evdokushkina, Victoria G. Semyonova.

Center on Aging, NORC/University of Chicago, 1155 East 60th Street, Chicago, IL 60637
lagavril@midway.uchicago.edu.

Abstract

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. This idea is also important for understanding the historical changes in human lifespan through the mechanism of technophysio evolution as suggested by Robert Fogel and Dora Costa. Can this new fascinating concept also be useful to understand (at least partially) the observed sex disparities in adult health and longevity? Are the long-lasting effects of early-life conditions identical for both sexes, or, on the contrary, are they sex-specific? These questions stimulated us to conduct the present pilot exploratory study on the sex specificity of late-life health outcomes for early-life effects.

In this study we explored the effects of early-life conditions on adult lifespan of 13,000 persons using methodology of historical follow-up study of extinct birth cohorts (members of European aristocratic families born in 1800-1880). Using the method of multivariate regression with nominal predictor variables for individual lifespan as outcome variable, we found that sex differences in adult life span are indeed modulated by early-life events and conditions. Specifically, we found that such variables as (1) month of birth, (2) father's age at person's conception, (3) birth order (first-born status) have statistically significant effects on adult lifespan (life expectancy at age 30) in females, but not in males. Female lifespan has bimodal distribution according to the month of birth (M-shaped curve). Women born in May or December live longer lives compared to those born in February or August, while male lifespan is less affected by the season of birth in our historical dataset. Daughters born to younger fathers (below 30 years) or old fathers (above 45 years) live significantly shorter lives, while sons are less affected by paternal age at conception. First-born daughters live longer lives, while first-born sons are not affected by their first-born status. The findings of this pilot exploratory study provide important scientific justification for a subsequent large-scale research project on related topic.

Introduction

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; Barker, 1992; 1998; Kuh & Ben-Shlomo, 1997; Leon *et al*., 1998; Lucas *et al*., 1999; Blackwell *et al*., 2001). 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; Fogel, 1997; 1999). Additional arguments suggesting the importance of early-life conditions in later-life health outcomes are coming from the reliability theory of aging and longevity (Gavrilova, Gavrilova, 1991; 2001a). 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 (Gavrilova, Gavrilova, 1991; 2001a). All these ideas require further testing, more studies, and more data.

There are two major goals in this pilot exploratory study:

(1) To find out whether early-life conditions may have significant effects on adult lifespan. We also tried to determine whether our dataset on European aristocratic families could be useful to explore the role of early-life conditions, and to be used in future more detailed studies (e.g., perhaps to become a research component of the Program Project on related topic).

(2) To determine whether early-life conditions may have significant effect on sex disparities in adult health and longevity. These sex disparities are well documented (Van Poppel, 2000), but they still have to be explained and fully understood. For example, the following research question could be posed: are the long-lasting effects of early-life conditions identical for both sexes, or, on the contrary, they are sex-specific? This question stimulated us to conduct the present pilot exploratory study on the sex specificity of the effects of early-life conditions on adult lifespan.

The study of sex differences has also important methodological implications for research work on early-life effects. This is because in many cases the available datasets are limited in their size and/or the studied outcomes are rare events, thus creating a temptation to pool the data for both sexes together in order to increase the statistical power (Blackwell *et al*., 2001). It is important, therefore, to find out whether gender differences in response to early-life conditions are indeed similar (so that the data could be pooled together with simple adjustment for sex just by one indicator variable), or they are fundamentally and qualitatively different (so that each sex should be studied separately).

In this study we addressed these scientific problems (fundamental and methodological) by studying the effects of early-life conditions on adult lifespan of men and women separately, using the methodology of historical prospective study of extinct birth cohorts. We found significant sex differences in adult lifespan responses to early-life conditions that justify the need for further full-scale research project on related topic.

Data and Methods

Main Data Source. In this study we collected, computerized, cross-checked and analyzed the detailed genealogical records on lifespan of about 13,000 adult persons (6,635 men and 6,488 women survived by age 30) and their parents, using particularly reliable and complete data on European royal and noble families for extinct birth cohorts (born 1800-1880). The main advantage of these data is their high accuracy, reliability and completeness (to be discussed later). Another advantage of this kind of data is the relative homogeneity of this Caucasian population regarding social class and educational background. Since this privileged social group lived in favorable conditions for many centuries, one could expect less influence of adverse social factors (poverty, for example) on life span and hence lower bias caused by these factors. This kind of data allows us to minimize the social heterogeneity of the population under study. Thus, although the sample analyzed in this study does not represent the whole human population (as laboratory animals do not represent species in the wild), it is one of the best possible samples to test biodemographic hypotheses since the effects of population heterogeneity are minimized with regard to social status.

The database on European royal and noble families (a family-linked database) was developed as a result of 5 years of our continued efforts that proved to be both labor-intensive and time-consuming because of extensive data cross-checking and data quality control. The earlier intermediate versions of this database were already used in our previous studies (Gavrilov, Gavrilova, 1997a; 1997b; 1999a; 2000a; 2001b; Gavrilov et al., 1997; Gavrilova, Gavrilov, 2001; Gavrilova et al., 1998). To develop this database we have chosen one of the best professional sources of genealogical data available - the famous German edition of the "Genealogisches Handbuch des Adels" (Genealogical Yearbook of Nobility). This edition is known world wide as the "Gotha Almanac" - "Old Gotha" published in Gotha in 1763-1944, and "New Gotha" published in Marburg since 1951 (see Gavrilova, Gavrilov, 1999, for more details). Data from the Gotha Almanach were often used in early biodemographic studies of fertility (see Hollingsworth, 1969, pp. 199-224, for references) and proved to be useful now in the studies of human longevity (Gavrilov, Gavrilova, 1997a; 1997b; 1999a; 2000a; 2001b; Gavrilov et al., 1997; Gavrilova, Gavrilov, 2001; Gavrilova et al., 1998).

Each volume of the New Gotha Almanach contains about 2,000 genealogical records dating back to the 14-16th centuries (to the founder of a particular noble genus). More than 100 volumes of this edition are already published, so more than 200,000 genealogical records with well-documented genealogical data are available from this data source. The high quality of information published in this edition is ensured by the fact that the primary information is drawn from the German Noble Archive (Deutsches Adelsarchiv). The Director of the German Noble Archive (Archivdirektor) is also the Editor of the New Gotha Almanach. Our own experience based on cross-checking the data, has demonstrated that the number of mistakes (mostly misprints) is very low in the "New Gotha Almanac" (less than 1 per 1000 records), so this source of data is very accurate compared to other published genealogies.

The information on noble families in the New Gotha Almanac is recorded in a regular manner. The description of each particular noble genus starts with information on 2-3

generations of founders of male sex only. Then three to four the most recent generations are described in more detail, including information on individuals (e.g., first and last names; event data: birth, death, marriage dates and places; descriptive data: noble degrees, occupation if available, information on death circumstances if available), information on parents (e.g., first and last names; event data: birth and death dates and places), information on spouses (e.g., first and last names; birth and death dates and places; first and last names of parents) and information on children (detailed as for each individual).

The process of data computerization is not yet completed – instead this is an ongoing project because of tremendous amounts of published data available for further computerizing. The present study represents, therefore, our intermediate findings.

Supplementary Data Sources. Some other supplementary sources of data were used in the development of database. These data sources included:

(1) computerized database on European royalty named "Royal92" and distributed on the Internet by Brian C. Tompsett at University of Hull, UK;

(2) computerized database on British Peerage distributed on CD by S&N Genealogy Supplies;

(3) relevant computerized data for European aristocratic families available in World Family Tree Archive CDs (Gavrilova, Gavrilov, 1999);

(4) over 100 genealogical publications on Russian nobility listed elsewhere (Gavrilov et al., 1996).

These data were used as a supplement to the main data source since their quality was not as high compared to the Gotha Almanac. Although data on European royalty were recorded in computerized data sources ("Royal92", British Peerage CD, see above) with sufficient completeness, data on lower rank nobility (landed gentry) were less complete and accurate. The same was true for the data on Russian nobility. All supplementary data were matched with the Gotha Almanac data, in order to cross-check the overlapping pieces of information. This cross-checking procedure allowed us to increase the completeness of the database by complementation of information taken from different sources.

The Structure of the Genealogical Database. The database approach used in this study is similar to the approach used for existing family-linked databases, such as the Utah Population Database (Skolnick et al., 1979), Laredo Epidemiological Project (Buchanan et al., 1984) or other historical databases (Gutmann et al., 1989).

Each record in the database represents an individual’s event data (birth and death dates and places) and individual’s descriptive information, that is, identification number, sex, first and last names, nobility rank, occupation, birth order, cause of death (violent/nonviolent), ethnicity, marital status, data source code number, data source year of publication. Individual information is supplemented by data for parents (identification numbers, first and last names, birth, death and marriage dates, cause of death) and spouses. Thus, the database that is used in this project is organized in the form of triplets (referred to as the "ego" and two parents). This structure of records is widely used in human genetics and is adequate for studies of parent-child relationships. Similar database structure was used in the recent study of kinship networks (Post et al., 1997).

Data Quality Control. Data quality control was an important part of our study designed to develop high quality family-linked database and to use it for scientific research.

The genealogical data sources were checked for the following: (1) *completeness* -- in reporting birth and death dates, which is crucial for calculating individual life span, the variable of particular interest in our study; (2) *accuracy* -- whether the percentage of mistakes and inconsistencies between reported dates (such as, for example, birth by the dead mother) is low enough to be acceptable; (3) *representativeness* -- whether the characteristics of investigated data sets (distribution by age, sex, marital status, age at death, etc.) is close to demographic characteristics of populations in similar geographic areas, historical periods and social groups. In our study we referred to the well-known publication by Thomas Hollingsworth (1962) on British peerage as a standard for European aristocracy, to check for data representativeness.

The *completeness* in birth and death dates reporting in the New Gotha Almanac was very high: dates of all vital events were reported for nearly 95% of all persons. Such high completeness is not common for many other genealogical data sources. For example, for British Peerage data published in Burke almanac in most cases there are no birth dates for women, which makes the calculation of their life span impossible. In fact, this problem (with British aristocratic women) was first noticed a century ago by Karl Pearson (Beeton and Pearson, 1899, 1901). He used the British Peerage data to study the longevity inheritance and had to exclude women from his consideration for the following reason: "*The limitation to the male line was enforced upon us partly by the practice of tracing pedigrees only through the male line, partly by the habitual reticence as to the age of women, even at death, observed by the compilers of peerages and family histories*" (Beeton and Pearson, 1901, pp. 50-51).

The *accuracy* of data published in the New Gotha Almanac is also very high: the frequency of inconsistent records is less than 1 per 1000 records while for many other genealogical data sources it falls within 1 per 300-400 records.

As for *representativeness*, the comparison of our data with Hollingsworth's analysis of British peerage (Hollingsworth, 1962) revealed good agreement between his findings and our data on mortality patterns, including male/female gap in life expectancy (7-10 years of female advantage in lifespan).

The genealogies for the members of European aristocratic families presented in the "Gotha Almanac" are of descending type, tracing almost all the descendants of relatively few founders. This is an important advantage of this data source over other genealogies that are often of ascending type (pedigrees). It is known in historical demography that the ascending genealogies are biased, over-representing more fertile and longer-lived persons who succeed to become ancestors, and for this reason such genealogies should be treated with particular caution (Jetté and Charbonneau, 1984; Fogel, 1993).

Another important advantage of this dataset is that the data are not spoiled by selective emigration (common problem for local registers), because every person is traced until his/her death in this dataset.

Thus, the genealogical data published in the Gotha Almanac are characterized by high quality and accuracy. We have, however, encountered two problems regarding the data completeness, which are discussed below, along with proposed solutions.

Censored, truncated observations and missing death dates. Our study revealed that the percentage of cases with unreported death dates is rather small in our main data sources (Gotha Almanac), and is caused mainly by right censoring of long-lived persons who were still alive by the date of data collection and publication. The percentage of non-reported death dates varies from 0 to 7% in extinct birth cohorts (1800-1880), while it is higher in later birth cohorts (1880-1899) - 23% for women and 8% for men, since some individuals were still alive by the date of data collection and volume publication. Note that women, who live longer, have a higher proportion of right-censored observations. The high proportion of censored observations in genealogies is not desirable, since the exact dates of censoring are often unknown. This uncertainty creates problems for data analysis, so the researchers working with genealogies prefer to use non-censored, extinct birth cohorts in their studies (Mayer, 1991; Pope, 1992; Kasakoff and Adams, 1995). We also used extinct (non-censored) birth cohorts in our study. For this purpose only those birth cohorts were used in the study that were born at least 100 years before the year of data publication (to be sure that the birth cohort under study is almost extinct).

Underreporting of women and children. In many genealogical books and databases non-married women as well as children died in infancy are often missed or reported with less completeness. Since genealogical records are focused on family names which are transmitted by males only, women could be lost in genealogies when they marry and change their family names (Hollingsworth, 1976). Also, in many cases data for women do not contain information on their birth and death dates resulting in biased sex ratio in the sample with complete dates. We have also encountered this problem in our studies although for somewhat different reason. Our analysis revealed that the main cause of the sex bias in the New Gotha Almanac is related to the manner of data representation: more recent generations are presented completely, while the earlier generations are limited mainly to the male ancestors (in order to avoid repetitive publication of individuals already presented in previous volumes). That is why, the sex ratio among early birth cohorts (1800-1860) is biased in favor of males while for more recent birth cohorts (1880-1899) it is within normal range. Since in our previous studies the most recent volumes of the New Gotha Almanac (published after 1980) were computerized and analyzed (in order to avoid censoring), the proportion of males in extinct birth cohorts (early generations) was substantially higher than expected (Gavrilov, Gavrilova, 1997a; b; 1999a; 2000a; 2001a; Gavrilov et al., 1997; Gavrilova et al., 1998)). Sex bias is an important issue, particularly when gender differences are studied (as it is done in the present study). Therefore, every effort is made to ensure that the dataset used in this particular study is sex-balanced (see Table 1).

Table 1 about here

To our knowledge, this new database is the only genealogical database on European aristocratic families, where there is no sex bias.

The underreporting of children who died in infancy may be also a serious problem, especially for studies that include fertility analysis. Fortunately, in the Gotha Almanac the families that belong to the higher nobility rank (kings, princes, earls) are described with

remarkable completeness. In particular, all ever born children are recorded, including those who died the same day when they were born. Another indicator of data completeness is the normal sex ratio at birth (101 to 108) observed among these families (based on analysis of our sample). In our database, over 90 aristocratic genuses belonged to the upper nobility were recorded completely, although data for lower rank nobility were not yet completed. Underreporting of children is not a problem for this particular study that is focused on adult life span for those who survived by age 30 years.

Analytical Methods

Since the data collected for this study are characterized by remarkable accuracy and completeness, it was possible to apply simple and straightforward methods of data analysis without making heavy assumptions. In particular, since the length of life is known for every person (there were no right censored observations) it was possible to analyze the duration of life directly as a dependent, outcome variable in multivariate regression model. There was no need to apply the Cox proportional hazard model and to make a heavy assumption about multiplicative effects of covariates on hazard rate. Instead, the persons' lifespan is studied directly as a dependent outcome variable and a function of other explanatory and potentially confounding predictor variables (see below).

In this study we applied a multivariate regression analysis with nominal variables, which is a very flexible tool to control for effects of both quantitative and qualitative (categorized) variables. This method also allows researchers to accommodate for complex non-linear and non-monotonic effects of predictor variables. The beauty of this method is that it does not require any assumptions about the analytical function describing the effects of predictor variables. Instead, the model allows researchers to calculate directly a conditional mean lifespan in a group of individuals with a particular set of predictor variables values. The regression coefficients obtained in this model (named as differential intercept coefficients) have a clear interpretation of additional years of life gained (or lost) due to change in a particular predictor variable.

We applied the methodology of prospective historical study to the data for extinct birth cohorts (born in 1800-1880), free of censored observations. We also tested a long list of explanatory and potentially confounding variables (described below) to avoid possible artifacts.

Life span of adult (30+) progeny (sons and daughters separately) was considered as a dependent outcome variable in multivariate regression with dummy (0-1) variables using the SAS statistical package (procedure REG). The independent predictor variables included 12 types of binary variables:

(1) calendar year of birth (to control for historical increase in life expectancy as well as for complex secular fluctuations in lifespan). The whole birth year period of 1800-1879 was split into 5-year intervals (16 intervals) presented by 15 binary (0-1) variables with reference level set at 1875-1879 birth years.

(2) maternal lifespan (to study maternal lifespan effects through combined genetic effects and shared environment). The maternal lifespan data were grouped into 5-year intervals (14 intervals) with the exception of the first (15-29 years) and the last (90+

years) longer intervals with small number of observations. The data were coded with 13 dummy variables with reference level set at 75-80 years for maternal lifespan.

(3) paternal lifespan (to study paternal lifespan effects through combined genetic effects and shared environment). The data were grouped and coded in a way similar to maternal lifespan (see above).

(4) maternal age when a person (proband) was born. This variable is used to control for possible confounding effects of maternal age on offspring lifespan. The data for mother's age were grouped in 5-year intervals (6 intervals to cover the age range of 15-60 years) with the exception of the last longer interval of 40+ years with small number of observations. Maternal age of 25-29 years is selected as a reference category.

(5) father's age when a person was born. This explanatory variable is used to study paternal age effects on offspring lifespan. The data were grouped and coded in 5-year intervals (9 intervals to cover the age range of 15-80 years) with the exception of the first (15-24 years) and the last (60-79 years) longer intervals with small number of observations. Paternal age of 40-44 years is selected as a reference category.

(6) birth order (first-born status). This variable is represented by binary variable coded as 1 when individual was a first-born child and coded as zero otherwise.

(7) nationality. The nationality of individual is represented by a set of 6 categories - Germans, British, Italians, Poles, Russians and 'others'. Germans (the largest group in our sample) is selected as a reference group.

(8) cause of death ('extrinsic' versus 'natural'). The death is coded as extrinsic or premature in following cases: (1) violent cause of death (war losses, accidents, etc.), (2) death in prison and other unfavorable conditions (concentration camp, etc.), (3) death from acute infections (cholera, etc.) and (4) maternal death (for women only). Deaths from all other causes combined were considered as a reference outcome. The proportion of reported ‘extrinsic’ deaths in our dataset was about 5% for males and about 1% for females.

(9) loss of the father in the formative years of life (before age 15). This is a binary variable coded as 1 when father was lost before the age 15 and coded as zero otherwise.

(10) loss of the mother before age 15. This binary variable is coded as 1 in those cases when mother was lost before the age 15 and coded as zero otherwise.

(11) loss of both parents (orphanhood) before age 15. This binary variable is coded as 1 in those cases when *both* parents were lost before the age 15 and coded as zero otherwise.

(12) month of birth. This variable was included into analysis, because previous studies indicated that month of birth may be an important predictor of adult lifespan (Gavrilov, Gavrilova, 1999a; Doblhammer, Vaupel, 2001), particularly for daughters (Gavrilov, Gavrilova, 1999a). This variable was represented as a set of 11 dummy variables with those born in February considered as a reference group. The main focus of this particular study is on sex-differences in the month-of-birth effects that were not well studied before.

Sensitivity analysis. In order to determine how robust are our findings, the sensitivity analysis was made. Specifically, the data were re-analyzed in several different ways,

when either the initial dataset was partially changed, or the set of predictor variables was modified. Changes in the dataset included the deletion of data for disadvantaged ethnicity with low lifespan (Russians), or deletion of the data for most recent birth cohorts (born in 1860-1880). Changes in predictor variables included consideration of such additional variables as nobility rank, sibship size, and reproductive lifespans (ages at last childbirth) both for mother and father.

Results and Discussion

Sex ratio and lifespan values. The characteristics of analyzed dataset are presented at Table 1:

Table 1 about here

There are several notable features to mention here:

First, the numbers of males and females are rather similar in all studied birth cohorts (no apparent sex bias). The sex ratio in the entire dataset is 1.02 (6,635 males/6,488 females). This is close to the normal sex ratio, in contrast to the sex ratio of 1.42, observed in the British peerage database with many missing records for women (Gavrilov, Gavrilova, 1999b). Thus, it seems to be possible to study sex differences in response to early-life conditions without concerns about selective sex bias in our dataset.

Second, the values for mean lifespan are rather high - more than 62 years for males and 66 years for females. It indicates that lifespan in this socially elite population is comparable with modern lifespan values observed now in some countries of the world. Thus, observations made on these historical data may perhaps be applicable, with some caution and certain reservations, to contemporary populations.

Third, there is a significant increase in lifespan over studied historical period, in particular for females (10 years gain). Therefore, the data should be adjusted for secular trends in lifespan (which has been done in this study). Finally, the temporal changes in lifespan are clearly not linear (no improvement in lifespan during the first 30 years), and sometimes even not monotonic which justifies the method of analysis used in this study (multivariate regression with nominal variables and treating the year of birth as categorized predictor variable).

Season of birth and human longevity. Table 2 presents striking data that the month of birth is an important predictor for the life expectancy of adult women (30 years and above). In particular, women born in May and December tend to live 2-3 years longer on average compared to those born in February (significant at $p < 0.01$). The effects of the months of birth are expressed in Table 2 as a difference from the reference level in February and are point estimates of the differential intercept coefficients adjusted for effects of other variables.

It is important to emphasize that the month of birth continues to be an important predictor for women's lifespan, even after adjustment for the effects of all other explanatory variables mentioned earlier in the "Data and Methods" section.

Table 2 about here

Note how regular is the M-shaped dependence of women's lifespan on their month of birth (Table 2). Starting with February "ground zero", the lifespan is increasing monotonically through March and April, reaching its first peak in May. Then lifespan starts to decline monotonically through June and July, reaching the local minimum in August. Then lifespan starts to increase again in a regular way through September, October and November, reaching its second peak in December. After that, it drops down through January to February forming the M-shaped pattern (bimodal distribution) with February and August as "bad" months to be born.

It is interesting to note that the months of February and August are already known in scientific literature as 'bad' months to be born. For example, a similar bimodal month-of-birth distribution was found for birth frequencies of cystic fibrosis disease with peak births in February and August (Brackenridge, 1980). Further studies are required to find out whether this just a coincidence of findings or a general seasonal pattern.

The fact that such an early circumstance of human life as the month of birth may have a significant effect 30 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.).

It is known that the deficiency of vitamins B_{12}, folic acid, B_6, niacin, C, or E, appears to mimic radiation in damaging DNA by causing single- and double-strand breaks, oxidative lesions, or both, and may contribute to premature aging (Ames, 1998). The seasonal lack of these vitamins in late winter/early spring, in coincidence with one of the two critical periods in fetus or child development (the third critical month of pregnancy and the first months after birth), may explain a dramatic life span shortening among those born in August and February. Our finding is also consistent with the reliability theory of aging, which emphasizes the importance of the initial level of damage that determines the future length of human life (Gavrilov, Gavrilova, 1991; 2001b).

These general explanations, however, do not match with data for males presented in Table 3. In contrast to females, the male lifespan does NOT depend on month of birth, at least in this particular dataset. This observation is the first example in our study when sex differences in response to early-life conditions are observed.

Table 3 about here

The sex specificity of the month-of-birth effects on adult lifespan is a puzzling observation, but it is also a reassuring one from the methodological point of view. Indeed, the data for men and women are taken from the same sources and are represented by the same set of parental variables (because they are brothers and sisters to each other).

Therefore, any possible flaws in data collection and analysis (such as omission of important predictor variable, for example) should produce very similar artifacts both in males and females data. Instead we observe a clear-cut sex-specific effect, which is reassuring from the methodological perspective.

While discussing the greater response of female lifespan to the season of birth, it is interesting to see whether other traits such as female childlessness are also affected by the month of birth. Indeed, studies on Dutch women found that the birth distribution of childless women, as compared with fecunds, was best represented with bimodal curve with zeniths in January and July (Smits et al., 1997). It is interesting to note that the two peaks for childlessness (January and July) seems to correspond well with the two observed minimums for female adult lifespan observed in our study (February and August – just only one month shift compared to childlessness findings).

Our finding that the month of February is "bad" month to be born for female corresponds well with schizophrenia studies. The risk of schizophrenia is higher for persons, whose birth date is close to February, and this seasonal effect is more marked among females (Dassa et al., 1995). It was also found that pre-natal exposure to influenza epidemic is associated with later development of schizophrenia in females but not in males (Takei et al., 1993; 1994).

Finally, we would like to comment on the importance to control for socio-economic status while studying the effects of month of birth. This is very important issue because there are significant differences in birth seasonality between different social classes (Smithers, Cooper, 1984; Bobak, Gjonca, 2001). Therefore, studies of aggregated data for whole countries (Doblhammer, Vaupel, 2001) may simply reflect the well-known differences in procreation habits of different socio-economic classes. In our study we control for socio-economic status by stratification (only aristocratic families are included into analysis).

Paternal age at childbirth and human longevity. The dependence of female lifespan on paternal age at reproduction (when daughter was born) is presented in Table 4. Note that there is an optimal age to father a daughter, which is rather late - about 40-44 years (considered as reference level in this study). Daughters born to older or younger fathers tend to live about 2-year shorter lives on average (significant at $P<0.05$, see Table 4). These are the net effects of paternal age, when all other covariates (see "Data and Methods" section) are taken into account, including maternal age effects that surprisingly proved to be less important.

Table 4 about here

Shorter lifespan of daughters conceived to older fathers could be explained by age-related accumulation of mutations in DNA of paternal germ cells (Crow, 1997; Gavrilov, Gavrilova, 2000a; 2001a). Advanced paternal age at person's conception is an important risk factor for such disease of adult age as schizophrenia (Malaspina, 2001; Malaspina et al., 2001), and such disease of old age as sporadic (non-familial) Alzheimer disease (Bertram et al., 1998).

It is more difficult to explain, why daughters born to particularly young fathers also live shorter lives. Standard social explanation, that low-income males without education start reproducing earlier seems not to be easily applicable to this socially elite group of royal and noble families.

Analysis of the scientific literature suggests that there may be a fundamental biological explanation of the "young father - short daughters' lifespan" paradox. It was found that the risk of congenital heart defects (ventricular septal defects, VSD, and atrial septal defects, ASD) is increased not only among the offspring of the older fathers, but also among the offspring of particularly young fathers - below 20 years (Olshan et al., 1994). Children born to younger fathers (under 20 years) have increased risk of neural tube defects, hypospadias, cystic kidney, and Down syndrome (McIntosh et al., 1995).

In laboratory mouse, offspring born from older mature fathers exhibit better behavioural performances (for spontaneous activity in both sex and learning capacity in males) than those born from particularly young post-pubescent fathers (Auroux et al., 1998). Similar results were obtained for humans in the study that involved the distribution of scores obtained in psychometric tests by 18-year-old male subjects, according to their father's age at the time of their birth. The curve of such scores produced an inverted U-shape, with poor scores for those conceived to particularly young or old fathers. Maternal age did not appear to play a part in this event. These results pose the problem of identifying genetic and/or biosocial factors associated with young fathers, which might have an impact on the quality of the conceptus (Auroux et al., 1989).

The practical importance of these findings is obvious: the age constrains for the donors of sperm cells in the case of IVF (in vitro fertilization) should be probably revised to exclude not only the old donors, but also those donors who are too young. Of course, more detailed studies are required, before such important practical recommendation could be made.

Again, all these interesting ideas and suggestions fail when data on males are analyzed (see Table 5).

Table 5 about here

In contrast to females, the male lifespan does NOT depend on paternal age at person's birth, at least in this particular dataset. This observation is the SECOND example in our study when sex differences in response to early-life conditions are observed.

Birth order (first-born status) and human longevity. The third example of sex differences in response to early-life conditions refers to effects of birth order on adult lifespan (Table 6). Specifically, the first-born daughters tend to live 1.6 years longer compared to next-born daughters (statistically significant at $p<0.005$, see Table 6), while sons again demonstrate no response at all.

This is kind of strange, because the first-born sons are most likely to inherit parental property in the studied aristocratic families and, therefore, should benefit most from their first-born status. From this perspective, the results seem to be completely opposite to expectations based on socio-economic explanations.

Table 6 about here

As for possible biological explanations, it is interesting to note the results of 17-year demographic study of 315 yellow baboons at Mikumi National Park, Tansania (Wasser, Norton, 1993). This study found that female, but not male, infant survival was inversely correlated with birth order, that matches to some extent with our findings on aristocratic families.

We suggest here the following working hypothesis for future testing. After the birth of the first child, a woman may be immunologically sensibilized to paternal antigens present in the first fetus. This immunologic sensibilization may be particularly strong to female fetuses (daughters), which carry paternal X chromosome that is particularly rich in paternal antigens compared to small Y chromosome carried by sons. Therefore, next fetus of female sex may experience particularly debilitating immunological attack from their mother during *in utero* development. According to this hypothesis, an increasing risk of hemolytic disease with increase in birth order (caused by maternal immunologic attack on paternal rhesus-antigens in fetus) is just a small known tip of the iceberg, while many other debilitating conditions of the similar nature are still less studied, because of their delayed consequences. We plan to test a number of predictions from this hypothesis in the future.

We also suggest here the following second working hypothesis to explain why women rather than men are affected by such early-life conditions as month-of-birth. We believe that it is the high reproductive load (numerous pregnancies in the past) that may be responsible for women's vulnerability and sensitivity to early-life injuries, and we also plan to test in the future a number of predictions that follow from this hypothesis.

Prospects for future research. There are several interesting directions for further development of these studies.

The first research direction is related to the findings by Tommy Bengtsson and his colleagues that it is the disease load in early life (estimated through infant mortality rate), which is the key early predictor for mortality in later life (Bengtsson, Lindstrom, 2000; 2001). Our dataset allows us to elaborate on this issue by including a new predictor variable (death of the sibling early in life) in future data analyses.

The second research direction is related to the finding made by Frans van Poppel and his colleagues that women's fecundability is associated with month of birth (Smits et al., 1997). Our dataset allows us to test this finding and to include fecundability variable in the future data analyses as the outcome variable, as well as the predictor/confounding variable for adult lifespan.

The third possible research direction is related to findings of George Alter and his colleagues on the importance of interfamily differences in adult mortality (Alter et al., 2001). It would be interesting to take into consideration the "family effects" using the

random effects model and to see how it affects our preliminary findings made in this study.

Finally, we believe that the findings presented in this study should be interpreted with caution and need to be replicated on other datasets. However, the results of this pilot exploratory study clearly indicate the need for separate analysis of data for males and females when late-life consequences of early-life conditions and events are explored. There is a definite need for subsequent full-scale studies of the effects of early-life conditions on sex-specific health outcomes in later life, and our pilot study presented here justifies the need of further work in this direction.

Acknowledgments

This pilot exploratory study was supported in part by NIA grants.

References

Alter G, Brostrom G, Edvinsson S (2001). Family effects on mortality in nineteenth-century Northern Sweden. Paper for the SSHA conference in Chicago, 15-18 November, 2001.

Ames BN (1998). Micronutrients prevent cancer and delay aging. *Toxicology Letters* 102-103: 5-18.

Auroux M, Nawar NN, Naguib M, Baud M, Lapaquellerie N (1998). Post-pubescent to mature fathers: increase in progeny quality? *Hum. Reprod.* 13: 55-59.

Barker DJP (1992). *Fetal and Infant Origins of Adult Disease*. London: BMJ Publishing Group.

Barker DJP (1998) *Mothers, Babies, and Disease in Later Life*. 2nd edition. London: Churchill Livingstone.

Beeton M, Pearson K (1899). Data for the problem of evolution in man, II: A first study of the inheritance of longevity and the selective death rate in man. *Proceedings of the Royal Society of London* 65: 290-305.

Beeton M, Pearson K (1901). On the inheritance of the duration of life and the intensity of natural selection in man. *Biometrika* 1: 50-89.

Bengtsson T, Lindstrom M (2000). Childhood misery and disease in later life: The effects on mortality in old age of hazards experienced in early life, southern Sweden, 1760-1894. *Population Studies* 54: 263-277.

Bengtsson T, Lindstrom M (2001). Early-life conditions and mortality in later life: Southern Sweden 1765-1894. Paper for the SSHA conference in Chicago, 15-18 November, 2001.

Bertram L, Busch R, Spiegl M, Lautenschlager NT, Muller U, Kurz A (1998). Paternal age is a risk factor for Alzheimer disease in the absence of a major gene. *Neurogenetics* 1: 277-280.

Blackwell DL, Hayward MD & Crimmins EM (2001). Does childhood health affect chronic morbidity in later life? *Social Science & Medicine* 52: 1269-1284.

Bobak M, Gjonca A (2001). The seasonality of live birth is strongly influenced by socio-demographic factors. *Human Reproduction* 16: 1512-1517.

Brackenridge CJ (1980). Bimodal month of birth distribution in cystic fibrosis. *Am. J. Med. Genet.* 5: 295-301.

Buchanan AV, Weiss KM, Schwartz RJ, MacNaughton NL, McCartan MA, Bates SS (1984). Reconstruction of genealogies from vital records: The Laredo Epidemiology Project. *Comput. Biomed. Res.* 17: 326-351.

Crow JF (1997). The high spontaneous mutation rate: Is it a health risk? *Proc. Natl. Acad. USA* 94: 8380-86.

Dassa D, Azorin JM, Ledoray V, Sambuc R, Giudicelli S (1996). Season of birth and schizophrenia: sex difference. *Prog. Neuro-Psychopharmacol. & Biol. Psychiat.* 20: 243-251.

Doblhammer G, Vaupel JW (2001). Lifespan depends on month of birth. *Proc. Natl. Acad. USA* 98: 2934-2939.

Fogel RW (1993). New sources and new techniques for the study of secular trends in nutritional status, health, mortality, and the process of aging. *Historical Methods* 26(1): 5-43.

Fogel RW (1997). Economic and social structure for an ageing population. *Phil. Trans. Royal Soc. London* **B 352**: 1905-1917.

Fogel RW (1999). Catching up with the economy. *Am. Economic Review* **89**: 1-21.

Fogel RW, Costa DL (1997). A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs. *Demography* **34**: 49-66.

Gavrilov LA, Gavrilova NS (1991). *The Biology of Life Span: A Quantitative Approach*, NY, Chur: Harwood Academic Publisher.
http://www.amazon.com/exec/obidos/ASIN/3718649837/107-2298796-3386931

Gavrilov LA, Gavrilova NS (1997a). Parental age at conception and offspring longevity. *Reviews in Clinical Gerontology* 7: 5-12.

Gavrilov LA, Gavrilova NS (1997b) When fatherhood should stop? *Science* 277: 17-18. http://www.sciencemag.org/cgi/content/full/277/5322/17b

Gavrilov LA & Gavrilova NS (1999a). Season of birth and human longevity. *Journal of Anti-Aging Medicine* **2**, 365-366. http://www.src.uchicago.edu/~gavr1/Season-of-Birth.pdf

Gavrilov LA & Gavrilova NS (1999b). Is there a reproductive cost for human longevity? *Journal of Anti-Aging Medicine* **2**, 121-123. http://www.src.uchicago.edu/~gavr1/JAAM-Reproductive-Cost.pdf

Gavrilov LA & Gavrilova NS (2000a). 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. http://www.amazon.com/exec/obidos/ASIN/3540677402/107-2298796-3386931

Gavrilov LA, Gavrilova NS (2000b). Life expectancy and the month of birth. In: *Healthy Life Expectancy*. REVES 12 Annual Meeting, March 20-22, Los Angeles, 2000, p.34. http://www.usc.edu/dept/gero/reves12/abstracts.html

Gavrilov LA, Gavrilova NS (2001a). The reliability theory of aging and longevity. *Journal of Theoretical Biology* **213**(4) (in press). http://www.src.uchicago.edu/~gavr1/JTB-Abstract.pdf

Gavrilov LA, Gavrilova NS (2001b). Biodemographic study of familial determinants of human longevity. *Population, English Selection* **13**(1) 197-222. http://www.ined.fr/englishversion/publications/population/englishselection/

Gavrilov LA, Gavrilova NS, Evdokushkina GN, Semyonova VG, Gavrilova AL, Evdokushkina NN, Lapshin EV (1996). Determinants of human longevity: parental age at reproduction and offspring longevity. *Longevity Report* (ISSN 0964-5659), 10(54): 7-15. http://www.geocities.com/HotSprings/Sauna/3748/lr54.htm

Gavrilov LA, Gavrilova NS, Kroutko VN, Evdokushkina GN, Semyonova VG, Gavrilova AL, Lapshin EV, Evdokushkina NN, Kushnareva YuE (1997). Mutation load and human longevity. *Mutation Research*, 377: 61-62. http://www.src.uchicago.edu/~gavr1/MutationResearch-97.pdf

Gavrilova NS, Gavrilov LA (1999). Data resources for biodemographic studies on familial clustering of human longevity. *Demographic Research* [Online], vol.1(4): 1-48. Available: http://www.demographic-research.org/Volumes/Vol1/4/default.htm.

Gavrilova NS, Gavrilov LA (2001) When does human longevity start?: Demarcation of the boundaries for human longevity. *Journal of Anti-Aging Medicine*, **4**(2): 115-124. http://www.src.uchicago.edu/~gavr1/JAAM-Boundaries-for-Human-Longevity.pdf

Gavrilova NS, Gavrilov LA, Evdokushkina GN, Semyonova VG, Gavrilova AL, Evdokushkina NN, Kushnareva YuE, Kroutko VN, Andreyev AYu (1998). Evolution, mutations and human longevity: European royal and noble families. *Human Biology* 70: 799-804. http://www.src.uchicago.edu/~gavr1/HumanBiology.pdf

Gutmann M, Fliess KH, Holmes AE, Fairchild AL, Teas WA (1989). Keeping track of our treasures: managing historical data with relational database software. *Historical Methods* 22(4), 128-143.

Hollingsworth TH (1962). The demography of the British Peerage. *Population Studies*, suppl., 18: 3-107.

Hollingsworth TH (1969). *Historical Demography*. Ithaca, N.Y.: Cornell University Press.

Jetté R, Charbonneau H (1984). Généalogies déscendantes et analyse démographique. *Annales de Démographie Historique* 45-54.

Kasakoff AB, Adams JW (1995). The effect of migration on ages at vital events: a critique of family reconstitution in historical demography. *Eur. J. Pop.* 11: 199-242.

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

Leon DA, Lithell HO, Vågerö D, Koupilová I, Mohsen R, Berglund L, Lithell U-B & McKeigue PM (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.

Lucas A (1991). Programming by early nutrition in man. In: *The Childhood Environment and Adult Disease* (Bock, G.R. & Whelan, J., eds), pp.38-55. Chichester: Whiley.

Lucas A, Fewtrell MS & Cole TJ (1999). Fetal origins of adult disease - the hypothesis revisited. *Br. Med. J.* 319: 245-249.

Malaspina D (2001) Paternal factors and schizophrenia risk: de novo mutations and imprinting. *Schizophr Bull* 27: 379-393.

Malaspina D, Harlap S, Fennig S, Heiman D, Nahon D, Feldman D, Susser ES (2001) Advancing paternal age and the risk of schizophrenia. *Arch Gen Psychiatry* 58: 361-367.

Mayer PJ (1991). Inheritance of longevity evinces no secular trend among members of six New England families born 1650-1874. *Am. J. Hum. Biol.* 3: 49-58.

McIntosh GC, Olshan AF, Baird PA (1995). Paternal age and the risk of birth defects in offspring. *Epidemiology* 6: 282-8.

Olshan AF, Schnitzer PG, Baird PA (1994). Paternal age and the risk of congenital heart defects. *Teratology* 50: 80-84.

Pope CL (1992). Adult mortality in America before 1900. A view from family histories. In: C.Goldin and H.Rockoff (eds.), *Strategic Factors in Nineteenth Century American Economic History. A Volume to Honor Robert W. Fogel.* Chicago and London: Univ. Chicago Press, 267-296.

Post W, Van Poppel F, Van Imhoff E, Kruse E (1997). Reconstructing the extended kin-network in the Netherlands with genealogical data: Methods, problems, and results. *Pop. Studies* 51: 263-278.

Skolnick M, Bean LL, Dintelman SM, Mineau G (1979). A computerized family history data base system. *Sociology and Social Research* 63: 506-523.

Smithers AG, Cooper HJ (1984). Social-class and season of birth. *Journal of Social Psychology* 124: 79-84.

Smits LJ, Van Poppel FW, Verduin JA, Jongbloet PH, Straatman H, Zielhuis GA (1997) Is fecundability associated with month of birth? An analysis of 19th and early 20th century family reconstitution data from The Netherlands. *Hum Reprod* **12**: 2572-2578.

Takei N, O'Callaghan E, Sham PC, Glover G, Murray RM (1993) Does prenatal influenza divert susceptible females from later affective psychosis to schizophrenia? *Acta Psychiatr Scand* 88: 328-336.

Takei N, Sham P, O'Callaghan E, Murray GK, Glover G, Murray RM (1994). Prenatal exposure to influenza and the development of schizophrenia: is the effect confined to females? *Am J Psychiatry* 151: 117-119.

Van Poppel F (2000) Long-term trends in relative health differences between men and women. *European Journal of Obstetrics & Gynecology* **93**: 119-122.

Wasser SK, Norton G (1993). Baboons adjust secondary sex-ratio in response to predictors of sex-specific offspring survival. *Behavioral Ecology and Sociobiology* 32: 273-281.

Wyshak G (1978). Fertility and longevity of twins, sibs, and parents of twins. *Soc. Biol.* 25: 315-30.

Table 1. Characteristics of the dataset.

Birth cohort (year of birth)	Mean Age at Death* ± Standard Error (years)	
	Daughters (sample size)	Sons (sample size)
1800-1809	66.2 ± 0.8 (419)	64.2 ± 0.7 (408)
1810-1819	66.2 ± 0.8 (444)	63.1 ± 0.7 (486)
1820-1829	66.1 ± 0.7 (563)	63.6 ± 0.7 (522)
1830-1839	67.8 ± 0.7 (579)	62.9 ± 0.6 (587)
1840-1849	70.1 ± 0.6 (634)	63.8 ± 0.5 (697)
1850-1859	71.7 ± 0.5 (810)	63.9 ± 0.5 (925)
1860-1869	74.3 ± 0.4 (1,168)	66.2 ± 0.4 (1,235)
1870-1879	76.3 ± 0.3 (1,871)	65.7 ± 0.4 (1,775)

*Mean age at death is calculated for those persons who survived by age 30. This variable refers to 'adult lifespan' in this study. The study dataset consists of 6,635 males and 6,488 females.

Table 2. Female lifespan as a function of month-of-birth.

Month-of-birth	Net effect* (point estimate)	Standard Error	P value
February	0.00	Reference level	
March	1.04	0.96	0.2792
April	1.62	0.96	0.0901
May	**2.71**	**0.94**	**0.0039**
June	1.83	0.94	0.0506
July	1.65	0.94	0.0806
August	1.34	0.94	0.1541
September	1.58	0.95	0.0981
October	**2.14**	**0.94**	**0.0230**
November	**2.32**	**0.98**	**0.0175**
December	**3.46**	**0.95**	**0.0003**
January	0.95	0.96	0.3195
February	0.00	Reference level	

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category.

Table 3. Male lifespan as a function of month-of-birth.

Month-of-birth	Net effect* (point estimate)	Standard Error	P value
February	0.00	Reference level	
March	-0.46	0.89	0.6052
April	-1.03	0.90	0.2539
May	1.03	0.88	0.2430
June	1.26	0.89	0.1595
July	-0.97	0.87	0.2649
August	-0.63	0.88	0.4729
September	0.06	0.87	0.9455
October	-0.47	0.90	0.6049
November	-1.19	0.90	0.1869
December	-0.33	0.91	0.7165
January	0.31	0.89	0.7323
February	0.00	Reference level	

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category.

Table 4. Female lifespan as a function of paternal age at reproduction.

Paternal Age	Net effect* (point estimate)	Standard Error	P value
15-24	-2.45	1.26	0.052
25-29	**-1.70**	**0.81**	**0.035**
30-34	-0.98	0.71	0.169
35-39	-0.13	0.69	0.854
40-44	0.00	Reference level	
45-49	-0.65	0.89	0.470
50-54	**-2.36**	**1.14**	**0.039**
55-59	-2.15	1.74	0.218

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category. Data on progeny of long-lived fathers (80+ years) are excluded from this particular analysis for the reasons explained elsewhere (Gavrilov, Gavrilova, 2000a).

Table 5. Male lifespan as a function of paternal age at reproduction.

Paternal Age	Net effect* (point estimate)	Standard Error	P value
15-24	-1.15	1.20	0.341
25-29	0.41	0.76	0.588
30-34	-0.37	0.66	0.580
35-39	0.38	0.65	0.557
40-44	0.00	Reference level	
45-49	0.90	0.83	0.278
50-54	1.19	1.15	0.301
55-59	-0.31	1.74	0.861

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category. Data on progeny of long-lived fathers (80+ years) are excluded from this particular analysis for the reasons explained elsewhere (Gavrilov, Gavrilova, 2000a).

Table 6. Effect of the birth order (first-born status) on adult lifespan of males and females.

Sex	Net effect of first-born status* (point estimate)	Standard Error	P value
Males	-0.03	0.48	0.957
Females	**1.58**	**0.55**	**0.004**

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category (not first-born status).