Paper presented at
The 2002 Annual Meeting of the Population Association of America
May 9-11, 2002
Atlanta

Session 152 "Early Life Conditions and Later Health Outcomes"

Early Life Conditions and Later Sex Differences in Adult Lifespan

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

Abstract

In this study we tested our hypothesis (and preliminary observations) that early-life conditions may determine in part the later sex differences in adult lifespan. We found that such variables as (1) father's age at person's conception, (2) parental lifespan, and (3) month of birth, have larger effects on adult lifespan (life expectancy at age 30) in females rather than males. Daughters born to particularly young fathers (below 25 years) or old fathers (above 45 years) live shorter lives, while sons are less affected by paternal age at conception. The response of progeny lifespan to exceptional parental longevity (lifespan above age 90) is particularly strong for female sex of the progeny. Women born in May or December live longer compared to those born in February, while male lifespan is less affected by the season of birth. Large family size (11+ siblings) increases daughter's lifespan, but decreases son's lifespan, thereby increasing the gender lifespan gap. These findings are confirmed through multivariate analysis of lifespan and hazard rates taking into account more than 20 predictor variables (including simultaneous consideration of such closely related variables as birth order, sibship size, parental ages at person's birth) as well as observed interfamilial differences and unobserved familial heterogeneity (through multilevel modeling).

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; Gavrilova, Gavrilova, 1991; 2001a; Barker, 1992; 1998; Elo & Preston, 1992; 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 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.

(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 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 14,000 adult persons (7,009 men and 6,908 women survived by age 30) and their parents, using particularly reliable and complete data on European aristocratic 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 aristocratic families (a family-linked database) was developed as a result of seven 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 the University of Hull, UK;

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

(3) relevant computerized data for European aristocratic families available in the 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 there are no birth dates for women in most cases, which makes the calculation of their life span impossible. In fact, this problem (with British aristocratic women) was first noticed by Karl Pearson a century ago (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. It was possible to trace the destiny of almost every person, even those relatively rare cases, when a person left Europe and eventually died in

other part of the World (USA, Canada, Australia, South Africa, India, Latin America, etc.).

While discussing the issue of data representativeness, it is also important to keep in mind for what purpose the data will be used. There is a significant difference in data requirements between the analytical and the descriptive studies (Levy and Lemeshow, 1999). Analytical studies that intend to test specific hypotheses (like this particular study focused on sex differences in response to some variables) are less dependent on data representativeness than the descriptive studies, which intend to describe a distribution of variables in a larger "whole population" (Levy and Lemeshow, 1999).

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) [To be on the safe side, the proportional hazard model was used in this study too, as a supplementary approach, in order to check the consistency of the findings, see later]. Data analysis was also supplemented by multilevel modeling to account for unobserved family heterogeneity.

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 24 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.

(6) birth order. This variable is represented by binary variables with the first birth order initially taken as the reference level.

(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 the 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.

(13) early marriage of a person. Cases of relatively early marriage are represented by a set of three categories (before age 20, at ages 20-24, or 25-29) with reference level for all other cases. This set of variables was included into the initial full model and then they were eliminated from the model after iterative step-wise procedure of deleting the variables with poor predictive value.

(14) proportion of boys born as siblings (characteristics of the family where the person was born). Families are categorized in the groups with different percentage of boys (below 20%, 20-39%, 40-59%, 60-79%, and above 80%).

(15) proportion of siblings died before age 30 (characteristics of the family where the person was born). Families are categorized in groups with different percentage of deceased siblings (no deaths -- 0%, 1-19%, 20-59%, and above 60%).

(16) nobility rank of the father (characteristics of the family where the person was born). Families are categorized in groups of ruling royal families (nobility rank 1), non-ruling princes and dukes (nobility rank 2), counts (nobility rank 3), barons (nobility rank 4), landed gentry (nobility rank 5) and occasional related individuals of non-aristocratic origin ("nobility rank" 6).

(17) family size (characteristics of the family where the person was born). Families are categorized in groups according to sibship sizes: 1-2, 3-4, 5-6, 7-8, 9-10, and 11+ . Later some categories were collapsed together during an iterative fitting procedure.

(18) Age of father when his first child was born (characteristics of the family where the person was born). Families are categorized in the groups according to paternal age at first childbirth: before age 25, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54 years, and above age 55.

(19) Age of mother when her first child was born (characteristics of the family where the person was born). Families are categorized in the groups according to maternal age at first childbirth: before age 20, 25-29, 30-34 years, and above age 35.

(20) Age of father when his last child was born (characteristics of the family where the person was born). Families are categorized in the groups according to paternal age at last childbirth: before age 29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64 years, and above age 65.

(21) Age of mother when her last child was born (characteristics of the family where the person was born). Families are categorized in the groups according to maternal age at last childbirth: before age 24, 25-29, 30-34, 35-39, 40-44 years, and above age 45.

(22) Extrinsic (violent) cause of paternal death (characteristics of the family where the person was born).

(23) Extrinsic (violent) cause of maternal death (characteristics of the family where the person was born).

(24) Parental age gap (characteristics of the family where the person was born). Families are categorized in the groups according to difference between paternal and maternal ages (split in five-year categories).

Table 2 presents information about distribution of these variables in our dataset:

Table 2 about here

Data analysis was performed with two complementary strategies – the analysis of restricted models with a few key variables (see tables 3 and 4 as examples) was complemented with the analysis of full models loaded with all variables (it was possible because of large sample sizes). An iterative procedure of step-wise elimination of variables with poor predictive value was applied to full models, which eventually produced the final models with the best set of predictor variables and collapsed, pooled categories for reference levels (tables 5-13).

In some cases an additional efforts were also made to increase the homogeneity of the dataset by eliminating cases with premature parental deaths -- early deaths before age 50 (tables 7, 8, 12,13), or deaths before age 60 (tables 9 and 10).

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 the 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 (this table was also discussed earlier, in Data and Methods section).

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 (7,009 males/6,908 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 (survived to age 30). 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 3 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 $p < 0.05$). The effects of the months of birth are expressed in Table 3 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 5).

Tables 3-6 about here

Note how regular is the M-shaped dependence of women's lifespan on their month of birth (Table 3). 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 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, are challenged by the data for males presented in Tables 4 and 6. In contrast to females, the male lifespan is less dependent 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.

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 family 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 both by stratification (only aristocratic families are included into analysis) and by regression (control for nobility rank).

Paternal age at childbirth and human longevity. The dependence of female lifespan on paternal age at reproduction (when daughter was born) is presented in Tables 7 and 9. In order to avoid confounding of parental age effects by selective parental survival (short-lived parents are always young parents, because dead parents do not reproduce), the two methods are simultaneously used: (1) stratification, and (2) regression. Stratification is achieved by considering only those cases, where both parents survived by age 50 (Table 7) or age 60 (Table 9), which makes a sample more homogeneous with regard to parental lifespan. Note that there is an optimal age to father a daughter. Daughters born to older or younger fathers tend to live shorter lives on average. 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.

Tables 7-10 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 are challenged when data on males are analyzed (see Tables 8, 10). In contrast to females, the male lifespan does NOT decrease with late 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.

Response of Progeny Lifespan to Parental Longevity

Sons born to long-lived fathers (lifespan above age 90) gain additional 3.6 years in their lifespan on average, compared to sons whose fathers die before age 80 (Table 11). The gain for daughters is about 35% higher -- 4.8 years of additional lifespan on average.

Table 11 about here

The response to maternal longevity is more modest. Sons born to long-lived mothers gain only 2.7 years in their lifespan (Table 11). Again, daughters seem to benefit more (about 40% more) from maternal longevity -- 3.8 years of additional life on average.

Thus, the response of progeny lifespan to parental longevity is consistently larger for female sex of the progeny. More studies on larger samples are required to confirm the statistical significance of this observation.

Sex Differences in Response to Family Size

In this study we encountered with the following unexpected and paradoxical observation: a large family size (sibship size) has opposite effect on lifespan of males and females.

Specifically, sons born in large families (sibship size 11+) **lose** about 2.4 years of their lifespan on average (397 cases out of total 7,009; standard error = 0.87; t-value = -2.7; P-value < 0.01).

On the contrary, daughters born in large families **gain** additional 2.2 years of lifespan on average (391 cases out of total 6,908; standard error = 0.91; t-value = 2.45; P-value = 0.01). These are net effects controlled for other important predictor variables as described in the Appendix 2.

Thus, daughters born in large families have an additional advantage over sons in their lifespan of about 4.6 years on average (standard error = 1.3, $p<0.01$). More studies are required to confirm this preliminary finding on other datasets and with other methods as it is already done in this study using multilevel models (see Table 13).

Other Analytical Approaches

In order to test the validity of obtained results we tried some other analytical approaches, which may be relevant to our study. First approach, the proportional hazard modeling, is often used in the study of lifetime data with covariates. Another approach, the multilevel modeling, is suitable for family data, which represent a sample of natural clusters. Both methods have advantages and limitations. The main problem with proportional hazard models is an obvious violation of proportionality assumption, when data are analyzed over long period of time for different birth cohorts. We tried to cope with this problem using stratification by calendar year of birth (see below). With multilevel models, small size of clusters (families) may create problems with valid parameter estimates. In order to resolve this problem we used a data sample of parents who survived through the reproductive period (by age of 50) and thereby had an opportunity to realize their reproductive potential. Thus, this approach decreased the number of small families and facilitated the estimation problem for multilevel model.

Proportional Hazard Models

In this study we used the following proportional hazard model with stratification (Allison, 1995):

$$log[h(t,\mathbf{x})] = \alpha_k(t) + \boldsymbol{\beta}\mathbf{x}$$

where $h(t,\mathbf{x})$ is a hazard function, $\mathbf{x}$ – is a vector of covariates, $\alpha_k(t)$ is a baseline function for k-th strata and $\boldsymbol{\beta}$ is a vector of regression coefficients. We assumed that different birth cohorts have different baseline function, so that we used a model with calendar year of birth (grouped by 10-year intervals) as a stratifying variable. All analyses were conducted separately for males and females. All observations with extrinsic causes of death (violent deaths, acute infections, maternal deaths) were considered as censored observations. The statistical analyses have been accomplished using SAS PHREG procedure.

The results obtained with proportional hazard models are consistent with our analyses using the multivariate regression analysis with nominal variables (see Table 12).

Table 12 about here

Note that women born in November-December have lower death rates (by 13-14%, significant, $P < 0.05$), while males born in the same months are not affected by the season of birth.

Also women born to old fathers (55-59 years) have 30% higher death rates ($P < 0.05$), while males, again, are not affected by old paternal age.

Daughters born to long-lived mothers (90+ years) have 31% lower death rates ($P < 0.05$), while the sons' gains are somewhat more modest -- 25% decrease in death rates ($P < 0.05$).

Daughters born to long-lived fathers (90+ years) have 33% lower death rates ($P < 0.05$), while the sons' gains are lower -- only 18% decrease in death rates (although statistically significant too, $P < 0.05$).

These findings should be treated with some caution, because of the proportionality assumption, which is inherent in the proportional hazard model. However, the consistency of the findings, obtained by different methods is reassuring.

Multilevel Models

There is an increasing recognition of the need to account for clustering in the study of complex sample designs. Statistical procedures that ignore clustering tend to underestimate the variance of the estimated coefficients and can lead to the mistaken conclusions about statistical significance of the studied effects (Rodriguez, Goldman, 2001). In our study a family is considered as a natural cluster, which should be taken into account. Thus, in addition to the multivariate regression analysis with nominal variables, we applied 2-level hierarchical linear model with the same set of covariates:

$$\boldsymbol{y_{ij}} = \boldsymbol{\beta_{0j}} + \boldsymbol{\beta_1 x_{ij}} + \boldsymbol{\beta_2 x_j} + \boldsymbol{u_j} + \boldsymbol{r}_{ij}$$

where $\boldsymbol{y_{ij}}$ is an outcome variable (lifespan of *i*-th individual in *j*-th family); $\boldsymbol{x_{ij}}$ and $\boldsymbol{x_j}$ represent vectors of observed characteristics at the individual and family levels. Here $\boldsymbol{u_j}$ is a vector of family random effects, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are vectors of individual and family fixed effects, and r_{ij} is a random error associated with the *i*-th individual in *j*-th family. The family random effects $\boldsymbol{u_j}$ are assumed independent and normally distributed, with

$$\boldsymbol{u_j} \sim N(0, \sigma^2_2)$$

Here σ^2 is a variance of random effects at the family level. The statistical analyses have been accomplished using SAS MIXED procedure.

Intra-family correlations were estimated according to the formula: $\rho = \sigma^2_2 / (\sigma^2_2 + \sigma^2_1)$, where σ^2_2 is a variance of random effects at the family (2-nd) level and σ^2_1 is a variance of random effects at the individual (1^{st} level). Data for females demonstrate higher values of intra-family correlation compared to males. Comparison of intra-family correlations with corresponding intra-family correlations obtained for restricted model without fixed effects showed that our variables account for significant portion of family clustering in lifespan. Also, in the presence of control for covariates, clustering of lifespan is less marked within families.

Modeling Strategy

Our multivariate analysis is restricted to the sample of parents living over 50 years. We conducted analyses for males and females separately. The results of data analyses are presented in Table 13.

Table 13 about here

Our model includes a set of individual and family variables (see Table 2). We determined which variables to include or exclude in the model on the basis of exploratory analysis using multivariate regression analysis with nominal variables. We also tested a number of interactions among parental age at reproduction and parental lifespan, and interactions between family ethnicity; no interactions proved to be statistically significant.

The results obtained with multilevel models are consistent with our analysis using the multivariate regression analysis with nominal variables (compare Table 13 with Tables 7 and 8).

Note that women born in December live 2 years longer on average (significant, $P < 0.05$), while males born in the same month are not affected by the season of birth.

Also women born to old fathers (55-59 years) lose 4 years of their lifespan on average ($P < 0.05$), while males, again, are not affected by old paternal age.

Daughters born to long-lived mothers (90+ years) gain almost 4 years of additional lifespan on average ($P < 0.05$), while sons gains are somewhat more modest -- 3.4 years of additional life ($P < 0.05$).

Daughters born to long-lived fathers (90+ years) gain 4.2 years of additional lifespan on average ($P < 0.05$), while the sons' gains are lower (2.4 years) and statistically insignificant.

Large family size (11+ siblings) increases daughter's lifespan by 2.7 years on average ($P < 0.05$), but it decreases son's lifespan by 3.7 years ($P < 0.05$), thereby increasing the gender lifespan gap by 6.4 years on average (adjusted for effects of other variables).

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 Dr. 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 in more detail by including a new set of predictor variables (death of the sibling early in life with different cut-off points at different ages) in future data analyses.

The second research direction is related to the finding made by Dr. 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.

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 study 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 the NIA grants.

References

Allison PD (1995). *Survival Analysis Using the SAS System. A Practical Guide*. Cary, NC: SAS Institute Inc.

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.

Elo IT, Preston SH (1992). Effects of early-life conditions on adult mortality: a review. *Population Index* 58: 186-212.

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): 527-545. http://www.src.uchicago.edu/~gavr1/JTB-01.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.

Levy P.S. & Lemeshow S. (1999). *Sampling of Populations: Methods and Applications*. Third Edition. John Wiley & Sons, Inc.

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.

Rodriguez G, Goldman N (2001). Improved estimation procedures for multilevel models with binary response: a case-study. *J. R. Statist. Soc. A* 164(2): 339-355.

Singer JD (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J. Educational and Behavioral Statistics* 24: 323-355.

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.

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	66.1 ± 0.7 (445)	64.0 ± 0.7 (443)
1810-1819	66.4 ± 0.8 (472)	63.1 ± 0.7 (522)
1820-1829	66.2 ± 0.7 (590)	63.7 ± 0.6 (553)
1830-1839	67.8 ± 0.7 (620)	63.0 ± 0.6 (636)
1840-1849	70.0 ± 0.6 (673)	63.8 ± 0.5 (742)
1850-1859	71.5 ± 0.5 (872)	63.9 ± 0.5 (963)
1860-1869	74.4 ± 0.4 (1,264)	66.3 ± 0.4 (1,311)
1870-1879	76.2 ± 0.3 (1,972)	65.6 ± 0.4 (1,839)

*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 7,009 males and 6,908 females.

Table 2.

Distribution of selected variables on full sample and on sample with parents living 50 years and more

Covariates	Percentage (Number)			
	Full Sample		Sample with parents living 50+ years	
	Males	Females	Males	Females
Individual Characteristics				
Month of Birth:				
January	7.95% (557)	7.73% (534)	7.92% (421)	7.70% (390)
(February)	7.40% (519)	7.24% (500)	7.74% (411)	7.37% (373)
March	8.05% (564)	7.59% (524)	8.17% (434)	7.21% (365)
April	7.75% (543)	7.61% (526)	7.79% (414)	7.54% (382)
May	8.40% (589)	8.43% (582)	8.51% (452)	8.41% (426)
June	8.03% (563)	8.32% (575)	8.38% (445)	8.39% (425)
July	8.86% (621)	8.21% (567)	8.77% (466)	8.43% (427)
August	8.50% (596)	8.41% (581)	8.32% (442)	8.65% (438)
September	8.85% (620)	7.87% (544)	8.75% (465)	8.35% (423)
October	7.88% (552)	8.32% (575)	7.62% (405)	8.18% (414)
November	7.59% (532)	7.21% (498)	7.57% (402)	7.19% (364)
December	7.42% (520)	7.87% (544)	7.53% (400)	7.64% (387)
Unknown	3.32% (233)	5.18% (358)	2.94% (156)	4.92% (249)
Month of Death Unknown	5.81% (407)	7.50% (518)	5.18% (275)	7.01% (355)
Death from violent cause	5.44% (381)	1.52% (105)	5.46% (290)	1.48% (75)
Birth order				
(1-6)	91.98% (6447)	92.78% (6409)	91.66% (4870)	92.26% (4671)
7-9	6.04% (423)	5.73% (396)	6.21% (330)	6.14% (311)
10+	1.98% (139)	1.49% (103)	2.13% (113)	1.60% (81)
Maternal Age at Reproduction				
15-19	4.91% (344)	5.05% (349)	4.18% (222)	4.66% (236)
20-24	26.89% (1885)	27.76% (1918)	25.41% (1350)	25.34% (1283)
25-29	31.60% (2215)	31.17% (2153)	31.24% (1660)	30.54% (1546)
30-34	22.21% (1557)	21.18% (1463)	22.94% (1219)	22.10% (1119)
35-39	10.94% (767)	11.22% (775)	12.06% (641)	13.02% (659)
40+	3.44% (241)	3.62% (250)	4.16% (221)	4.35% (220)

Paternal Age at Reproduction				
15-24	3.50% (245)	3.69% (255)	2.86% (152)	3.16% (160)
25-29	17.06% (1196)	17.70% (1223)	15.55% (826)	16.22% (821)
30-34	25.42% (1782)	24.71% (1707)	24.34% (1293)	22.99% (1164)
35-39	23.46% (1644)	23.02% (1590)	23.64% (1256)	23.27% (1178)
40-44	15.22% (1067)	15.62% (1079)	16.09% (855)	16.75% (848)
45-49	8.83% (619)	8.28% (572)	9.98% (530)	9.28% (470)
50-54	4.09% (287)	4.28% (296)	4.72% (251)	5.04% (255)
55-59	1.57% (110)	1.66% (115)	1.86% (99)	2.03% (103)
60+	0.84% (59)	1.01% (70)	0.96% (51)	1.26% (64)
Calendar Year of Birth				
1800-1804	3.24% (227)	3.14% (217)	2.67% (142)	2.80% (142)
1805-1809	3.08% (216)	3.30% (228)	2.82% (150)	3.00% (152)
1810-1814	3.52% (247)	3.11% (215)	2.92% (155)	2.73% (138)
1815-1819	3.92% (275)	3.72% (257)	3.63% (193)	3.32% (168)
1820-1824	3.88% (272)	4.50% (311)	3.76% (200)	4.23% (214)
1825-1829	4.01% (281)	4.04% (279)	3.73% (198)	3.69% (187)
1830-1834	4.82% (338)	4.10% (283)	4.52% (240)	3.77% (191)
1835-1839	4.25% (298)	4.88% (337)	3.76% (200)	4.72% (239)
1840-1844	5.06% (355)	4.75% (328)	4.99% (265)	4.82% (244)
1845-1849	5.52% (387)	4.99% (345)	5.67% (301)	4.90% (248)
1850-1854	6.23% (437)	6.09% (421)	6.36% (338)	6.32% (320)
1855-1859	7.50% (526)	6.53% (451)	7.96% (423)	7.09% (359)
1860-1864	8.57% (601)	8.27% (571)	9.07% (482)	8.30% (420)
1865-1869	10.13% (710)	10.03% (693)	10.80% (574)	10.27% (520)
1870-1874	11.84% (830)	12.35% (853)	12.70% (675)	12.76% (646)
1875-1879	12.36% (866)	13.58% (938)	12.50% (664)	14.44% (731)
Loss of Father Before Age 15	13.38% (938)	15.74% (1087)	7.81% (415)	8.16% (413)
Loss of Mother Before Age 15	11.61% (814)	12.49% (863)	0.51% (27)	0.45% (23)
Family/Individual Characteristics				
Ethnicity				
British	3.10% (217)	3.40% (235)	3.12% (166)	3.61% (183)
Russian	15.64% (1096)	15.66% (1082)	14.55% (773)	14.77% (748)
(Other)	81.27% (5696)	80.94% (5591)	82.33% (4374)	81.61% (4132)
Parental Age Gap				
(Below 5 years)	34.97% (2451)	36.09% (2493)	33.63% (1787)	35.16% (1780)
5-10 years	32.50% (2278)	31.41% (2170)	32.11% (1706)	31.44% (1592)
10-15 years	19.45% (1363)	18.80% (1299)	20.27% (1077)	18.57% (940)
15 years and more	13.08% (917)	13.69% (946)	13.98% (743)	14.83% (751)

Maternal Lifespan				
Below 80	77.39% (5424)	76.87% (5310)	72.97% (3877)	71.62% (3626)
80-84	12.33% (864)	12.22% (844)	14.51% (771)	14.77% (748)
85-89	7.22% (506)	7.67% (530)	8.71% (463)	9.50% (481)
90+	3.07% (215)	3.24% (224)	3.80% (202)	4.11% (208)
Paternal Lifespan				
Below 80	82.39% (5775)	83.24% (5750	80.84% (4295)	81.08% (4105)
80-84	10.60% (743)	9.94% (687)	11.42% (607)	11.16% (565)
85-89	5.09% (357)	4.94% (341)	5.65% (300)	5.59% (283)
90+	1.91% (134)	1.88% (130)	2.09% (111)	2.17% (110)
Sibship size				
Below 9	85.89% (6020)	87.83% (6067)	85.32% (4533)	86.71% (4390)
9-10	8.45% (592)	6.51% (450)	8.56% (455)	7.07% (358)
11+	5.66% (397)	5.66% (391)	6.12% (325)	6.22% (315)
Nobility Rank of Father				
1-2	25.00% (1752)	24.62% (1701)	24.51% (1302)	24.41% (1236)
(3-5)	67.56% (4735)	67.56% (4667)	68.25% (3626)	68.02% (3444)
Unknown	7.45% (522)	7.82% (540)	7.25% (385)	7.56% (383)
Death of Sibling Before Age 30	50.44% (3535)	44.83% (3097)	49.22% (2615)	44.85% (2271)
Paternal Age at First Childbirth				
Below 45	96.11% (6736)	95.25% (6580)	95.37% (5067)	94.43% (4781)
45-49	2.58% (181)	2.79% (193)	3.03% (161)	3.22% (163)
50+	1.31% (92)	1.95% (135)	1.60% (85)	2.35% (119)
Maternal Age at First Childbirth				
30+	8.96% (628)	11.48% (793)	9.64% (512)	12.58% (637)
Extrinsic Death of Mother	0.68% (48)	1.03% (71)	0.09% (5)	0.18% (9)
Extrinsic Death of Father	0.98% (69)	0.81% (56)	0.64% (34)	0.32% (16)
Number of individuals	7,009	6,908	5,313	5,063

Table 3.
Female lifespan as a function of month-of-birth (restricted model)

Month-of-birth	Net effect* (point estimate)	Standard Error	P value
February	0.00	Reference level	
March	0.67	0.94	0.4812
April	1.51	0.94	0.1096
May	**2.35**	**0.92**	**0.0108**
June	1.58	0.92	0.0875
July	1.81	0.93	0.0506
August	1.41	0.92	0.1259
September	1.48	0.94	0.1132
October	1.70	0.92	0.0656
November	**2.03**	**0.96**	**0.0339**
December	**2.65**	**0.94**	**0.0047**
January	0.91	0.94	0.3333
February	0.00	Reference level	

Table 4. Male lifespan as a function of month-of-birth (restricted model)

Month-of-birth	Net effect* (point estimate)	Standard Error	P value
February	0.00	Reference level	
March	0.30	0.90	0.7382
April	-0.74	0.91	0.4142
May	1.36	0.89	0.1280
June	1.52	0.90	0.0906
July	-0.92	0.88	0.2961
August	-0.70	0.89	0.4315
September	0.12	0.88	0.8901
October	-0.35	0.91	0.7009
November	-0.75	0.91	0.4136
December	-0.21	0.92	0.8179
January	0.22	0.90	0.8048
February	0.00	Reference level	

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category (see Appendix 1 for more details).

Table 5.
Female lifespan as a function of month-of-birth (final model)

Month-of-birth	Net effect* (point estimate)	Standard Error	P value
February	0.00	Reference level	
March	1.10	0.92	0.2331
April	1.72	0.92	0.0619
May	**2.35**	**0.90**	**0.0090**
June	1.66	0.90	0.0665
July	**1.86**	**0.91**	**0.0404**
August	1.49	0.90	0.0978
September	1.51	0.92	0.0986
October	**1.95**	**0.90**	**0.0308**
November	**2.13**	**0.93**	**0.0229**
December	**3.04**	**0.91**	**0.0009**
January	0.94	0.92	0.3086
February	0.00	Reference level	

Table 6. Male lifespan as a function of month-of-birth (final model)

Month-of-birth	Net effect* (point estimate)	Standard Error	P value
February	0.00	Reference level	
March	-0.03	0.87	0.9755
April	-1.16	0.87	0.1833
May	1.00	0.86	0.2436
June	1.37	0.87	0.1139
July	-0.94	0.85	0.2688
August	-0.80	0.86	0.3489
September	-0.01	0.85	0.9869
October	-0.59	0.87	0.4990
November	-0.81	0.88	0.3577
December	-0.36	0.88	0.6831
January	0.37	0.87	0.6691
February	0.00	Reference level	

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category (see Appendix 2 for more details).

Table 7. Female lifespan as a function of paternal age at person's birth
(5,063 cases, both parents lived 50+ years)

Paternal Age	Net effect* (point estimate)	Standard Error	P value
15-24	-2.26	1.28	0.078
25-29	-0.36	0.71	0.615
30-34	-0.53	0.62	0.388
35-39	0	Reference level	
40-44	-0.11	0.67	0.868
45-49	-0.95	0.88	0.282
50-54	-1.90	1.16	0.101
55-59	**-5.37**	**1.65**	**0.001**

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category. The data are point estimates of the differential intercept coefficients adjusted for other explanatory variables using multivariate regression with nominal variables (see Appendix 3).

Table 8. Male lifespan as a function of paternal age at person's birth
(5,313 cases, both parents lived 50+ years)

Paternal Age	Net effect* (point estimate)	Standard Error	P value
15-24	**-2.77**	**1.28**	**0.031**
25-29	0.06	0.68	0.935
30-34	-1.05	0.58	0.069
35-39	0	Reference level	
40-44	-0.85	0.64	0.185
45-49	0.01	0.81	0.990
50-54	0.17	1.13	0.878
55-59	-1.69	1.64	0.304

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category. The data are point estimates of the differential intercept coefficients adjusted for other explanatory variables using multivariate regression with nominal variables (see Appendix 3).

Table 9. Female lifespan as a function of paternal age at person's birth
(3,622 cases, both parents lived 60+ years)

Paternal Age	Net effect* (point estimate)	Standard Error	P value
15-24	**-3.33**	**1.53**	**0.029**
25-29	0	Reference level	
30-34	-0.14	0.85	0.871
35-39	-0.67	0.94	0.475
40-44	-0.86	1.12	0.442
45-49	-1.04	1.34	0.436
50-54	-2.36	1.63	0.148
55-59	**-4.88**	**2.18**	**0.025**

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category. The data are point estimates of the differential intercept coefficients adjusted for other explanatory variables using multivariate regression with nominal variables (see Appendix 4).

Table 10. Male lifespan as a function of paternal age at person's birth
(3.776 cases, both parents lived 60+ years)

Paternal Age	Net effect* (point estimate)	Standard Error	P value
15-24	-1.51	1.60	0.345
25-29	0	Reference level	
30-34	-0.40	0.83	0.631
35-39	0.97	0.94	0.303
40-44	0.20	1.13	0.859
45-49	0.71	1.31	0.587
50-54	0.31	1.70	0.856
55-59	-1.20	2.16	0.578

*Net effect corresponds to additional years of life gained (or lost) compared to the reference category. The data are point estimates of the differential intercept coefficients adjusted for other explanatory variables using multivariate regression with nominal variables (see Appendix 4).

Table 11.

Effect of parental exceptional longevity (90+ years) on adult lifespan of sons and daughters

Parental net effect in years	Net effect for sons Effect, in years ± S.E. (sample size)	Net effect for daughters Effect, in years ± S.E. (sample size)
Paternal effect	**3.58 ± 1.25**** (134 cases out of total 7,009)	**4.83 ± 1.32***** (130 cases out of total 6,908)
Maternal effect	**2.71 ± 1.00**** (215 cases out of total 7,009)	**3.78 ± 1.01***** (224 cases out of total 6,908)

** Significant at $p<0.01$

*** Significant at $p<0.001$

Note: Net effect corresponds to additional years of life gained compared to the reference category (parental lifespan below 80 years). The reason why it was possible to collapse all cases of parental lifespan below age 80 into one reference category (separate for maternal and paternal lifespan variables) is explained elsewhere (Gavrilova, Gavrilov, 2001). In short, the effects of parental lifespan on progeny lifespan are negligible if a parent lives less than 80 years, in this particular dataset (Gavrilova, Gavrilov, 2001). The data are point estimates of the differential intercept coefficients adjusted for other explanatory variables using multivariate regression with nominal variables (see Appendix 2).

Table 12.

Parameter estimates for stratified proportional hazard model of lifespan among the offspring of parents living 50+ years

Covariates	Females		Males	
	Parameter Estimates	Hazard Ratio	Parameter Estimates	Hazard Ratio
Month of Birth:				
January	-0.08	0.926	-0.07	0.934
(February)				
March	-0.06	0.939	-0.12	0.889
April	-0.09	0.913	0.04	1.035
May	**-0.16**	**0.854**	**-0.15**	**0.864**
June	-0.12	0.888	-0.15	0.861
July	-0.11	0.895	0.01	1.011
August	-0.07	0.930	0.02	1.024
September	**-0.15**	**0.865**	-0.09	0.917
October	-0.08	0.920	-0.08	0.920
November	**-0.15**	**0.860**	-0.03	0.971
December	**-0.14**	**0.866**	-0.05	0.950
Unknown	0.07	1.076	0.09	1.095
Month of Death Unknown	-0.07	0.930	0.02	1.016
Birth order				
(1-6)	0	1	0	1
7-9	0.10	1.105	0.02	1.016
10+	0.17	1.186	-0.08	0.925
Maternal Age at Reproduction				
15-19	-0.04	0.963	0.04	1.044
20-24	0.07	1.067	0.08	1.080
(25-29)	0	1	0	1
30-34	0.01	1.013	0.02	1.020
35-39	-0.03	0.968	0.08	1.087
40+	0.02	1.017	0.08	1.080
Paternal Age at Reproduction				
15-24	0.13	1.135	0.06	1.058
25-29	0.05	1.053	-0.06	0.942
30-34	0.05	1.056	-0.01	0.992
35-39	0.04	1.043	-0.06	0.938
(40-44)	0	1	0	1
45-49	0.10	1.099	-0.03	0.969
50-54	0.10	1.101	-0.06	0.936
55-59	**0.26**	**1.30**	-0.04	0.955

Loss of Father Before Age 15	0.09	1.092	-0.001	1.000
Loss of Mother Before Age 15	0.01	1.011	-0.21	0.812
Ethnicity				
British	**-0.24**	**0.79**	**-0.23**	**0.793**
Russian	**0.29**	**1.34**	**0.35**	**1.415**
(Other)	0	1		
Death of Sibling Before Age 30	0.05	1.052	0.05	1.048
Nobility Rank of Father				
1-2	**0.12**	**1.13**	0.02	1.020
(3-5)	0	1	0	1
Unknown	0.11	1.12	-0.02	0.978
Parental Age Gap				
(Below 5 years)				
5-10 years	-0.08	0.926	-0.02	0.983
10-15 years	-0.02	0.977	-0.02	0.981
15 years and more	-0.04	0.959	-0.02	0.977
Maternal Lifespan				
(Below 80)	0	1	0	1
80-84	**-0.12**	**0.891**	-0.05	0.948
85-89	**-0.18**	**0.836**	**-0.11**	**0.893**
90+	**-0.37**	**0.693**	**-0.29**	**0.749**
Paternal Lifespan				
(Below 80)	0	1	0	1
80-84	**-0.10**	**0.906**	**-0.20**	**0.822**
85-89	**-0.07**	**0.928**	**-0.18**	**0.834**
90+	**-0.40**	**0.668**	**-0.20**	**0.816**
Sibship size				
(Below 11)	0	1	0	1
11+	-0.09	0.92	**0.18**	**1.198**
Paternal Age at First Childbirth				
(Below 45)	0	1	0	1
45-49	-0.14	0.87	0.02	1.019
50+	**-0.27**	**0.77**	0.17	1.186
Maternal Age at First Childbirth				
30+	**-0.13**	**0.88**	0.02	1.024

Note: Deaths from extrinsic causes were considered as censored observations. Strata – calendar year of birth (10-year groups). Parameter estimates significant at $p \leq .05$ are printed in bold. Omitted categories are shown in parentheses.

Table 13.
Parameter estimates (fixed effects) for multilevel model of lifespan among the offspring of parents living 50+ years

Covariates	Females		Males	
	Parameter Estimate	t-value	Parameter Estimate	t-value
Fixed Effects				
Individual Characteristics				
Month of Birth:				
January	0.34	0.33	0.29	0.29
(February)	0		0	
March	0.58	0.55	1.29	1.330
April	1.06	1.00	-0.79	-0.80
May	**2.52**	**2.45**	1.34	1.38
June	1.80	1.76	**2.00**	**2.07**
July	1.52	1.48	-0.63	-0.65
August	1.31	1.29	-0.45	-0.46
September	1.34	1.30	1.02	1.07
October	1.20	1.16	0.40	0.40
November	1.70	1.59	0.23	0.23
December	**2.09**	**1.99**	0.07	0.07
Unknown	**-3.01**	**-2.21**	-2.46	-1.60
Month of Death Unknown	**2.06**	**2.22**	0.32	0.31
Death from violent cause	**-15.50**	**-9.17**	**-15.45**	**-17.63**
Birth order				
(1-6)	0		0	
7-9	**-2.17**	**-2.17**	-0.46	-0.48
10+	-3.29	-1.70	1.16	0.70
Maternal Age at Reproduction				
15-19	-0.11	-0.10	-0.42	-0.36
20-24	-1.03	-1.61	-0.85	-1.42
(25-29)	0		0	
30-34	-0.73	-1.06	-0.53	-0.82
35-39	-0.12	-0.12	-0.35	-0.36
40+	0.13	0.09	-0.37	-0.25
Paternal Age at Reproduction				
15-24	-2.86	-1.62	-1.35	-0.76
25-29	-0.82	-0.67	1.51	1.26
30-34	-0.98	-1.03	0.21	0.23
35-39	-0.17	-0.23	1.03	1.42
(40-44)	0		0	
45-49	-0.41	-0.44	0.53	0.61
50-54	-1.05	-0.79	0.29	0.23
55-59	**-4.14**	**-2.19**	-1.57	-0.85
Loss of Father Before Age 15	**-1.78**	**-1.96**	0.40	0.47
Loss of Mother Before Age 15	-1.49	-0.47	2.43	0.86

Family/Individual Characteristics				
Ethnicity				
British	2.31	1.87	**2.49**	**2.00**
Russian	**-3.99**	**-5.52**	**-4.46**	**-6.45**
(Other)	0		0	
Death of Sibling Before Age 30	**-1.00**	**-2.08**	-0.58	-1.24
Nobility Rank of Father				
1-2	**-2.21**	**-3.99**	-0.54	-1.02
(3-5)	0		0	
Unknown	**-2.25**	**-2.50**	-0.26	-0.30
Parental Age Gap				
(Below 5 years)	0		0	
5-10 years	0.51	0.76	0.28	0.43
10-15 years	-0.69	-0.72	0.15	0.16
15 years and more	-0.21	-0.15	0.89	0.66
Maternal Lifespan				
(Below 80)	0		0	
80-84	1.03	1.64	0.76	1.27
85-89	**2.00**	**2.61**	**2.24**	**2.97**
90+	**3.89**	**3.47**	**3.37**	**3.08**
Paternal Lifespan				
(Below 80)	0		0	
80-84	1.22	1.73	**3.03**	**4.52**
85-89	**2.12**	**2.18**	**2.35**	**2.56**
90+	**4.16**	**2.72**	2.41	1.65
Sibship size				
(Below 11)	0		0	
9-10	1.50	1.52	0.11	0.13
11+	**2.70**	**2.32**	**-3.66**	**-3.38**
Paternal Age at First Childbirth				
(Below 45)	0		0	
45-49	**2.61**	**1.89**	0.20	0.15
50+	**3.57**	**2.10**	-2.39	-1.25
Random effects		z-value		z-value
Standard deviations (restricted model)				
Family	**47.07**	**9.02**	**25.28**	**6.48**
Residual	**191.17**	**34.13**	**194.73**	**39.08**
Intraclass correlation ρ (restr.model)	0.20		0.14	
Standard deviations (full model)				
Family	**29.09**	**6.66**	**18.49**	**5.35**
Residual	**182.15**	**35.14**	**182.38**	**39.36**
Residual ρ (full model)	0.12		0.09	
% decrease in family variance component	46.29		36.44	

Note: Estimates for nominal variables corresponding to the calendar year of birth (all statistically significant) are not presented here. Omitted categories are shown in parentheses. Parameter estimates significant at $p \leq .05$ are printed in bold. Net effect corresponds to additional years of life gained (or lost) compared to the reference category. The data are point estimates of the differential intercept coefficients adjusted for other explanatory variables.

Appendix 1

Results for **Table 3** "**Female** lifespan as a function of month-of-birth are obtained through analysis of lifespan data (outcome variable) for **6,908** women born in 1800-1880 (extinct birth cohorts), who survived by age 30 (to focus on analysis of adult lifespan). In this initial restricted model the data are controlled for calendar year of birth (15 five-year categories), and data incompleteness – cases of unknown month-of-birth (358 cases, 5.2%), and unknown month of death (518 cases, 7.5%). The F-value for this restricted regression model is 18.69 ($p<0.0001$).

Results for **Table 4** "**Male** lifespan as a function of month-of-birth are obtained through analysis of lifespan data (outcome variable) for **7,009** men born in 1800-1880 (extinct birth cohorts), who survived by age 30 (to focus on analysis of adult lifespan). In this initial restricted model the data are controlled for calendar year of birth (15 five-year categories), and data incompleteness – cases of unknown month-of-birth (233 cases, 3.2%), and unknown month of death (407 cases, 5.8%). The F-value for this restricted regression model is 4.98 ($p<0.0001$).

Appendix 2

Results for **Table 5** "**Female** lifespan as a function of month-of-birth" are obtained through analysis of lifespan data (outcome variable) for **6,908** women born in 1800-1880 (extinct birth cohorts), who survived by age 30 (to focus on analysis of adult lifespan). The following additional predictor variables are also included in the final model because of their predictive value: (1) Calendar year of birth, (2) Ethnicity (Russian, British and others), (3)Lloss of father during formative years of childhood (before age 15), (4) Loss of mother during formative years of childhood (before age 15), (5) Cause of death (violent vs non-violent), (6) Early death of at least one sibling (before age 30), (7) High birth order (7+), (8) Nobility rank of the father (indicator of social status), (9) Large family size (number of siblings 9+), (10) Maternal lifespan, (11) Paternal lifespan, (12) Paternal age at person's birth, (13) Late paternal age at first childbirth (50+ years), (14) Birth of the first child by mother after age 30, (15) Death of mother from violent cause of death. The F-value for the final regression model is 18.12 ($p<0.0001$).

Results for **Table 6** "**Male** lifespan as a function of month-of-birth" are obtained through analysis of lifespan data (outcome variable) for **7,009** men born in 1800-1880 (extinct birth cohorts), who survived by age 30 (to focus on analysis of adult lifespan). The following additional predictor variables are also included in the final model because of their predictive value: (1) Calendar year of birth, (2) Ethnicity (Russian, British and others), (3) Loss of father during formative years of childhood (before age 15), (4) Loss of mother during formative years of childhood (before age 15), (5) Cause of death (violent vs non-violent), (6) Early death of at least one sibling (before age 30), (7) High birth order (7+), (8) Nobility rank of the father (indicator of social status), (9) Large family size (number of siblings 9+), (10) Maternal lifespan, (11) Paternal lifespan, (12) Paternal age at person's birth, (13) Late paternal age at first childbirth (50+ years), (14) Birth of the first child by mother after age 30, (15) Death of mother from violent cause of death. The F-value for the final regression model is 14.90 ($p<0.0001$).

Appendix 3

Results for **Table 7** "**Female** lifespan as a function of paternal age at reproduction" are obtained through analysis of lifespan data (outcome variable) for **5,063** women born in 1800-1880 (extinct birth cohorts), who survived by age 30 (to focus on analysis of adult lifespan). In order to avoid confounding of parental age effects by selective parental survival (short-lived parents are always young parents, because dead parents do not reproduce), the two methods are simultaneously used: (1) stratification, and (2) regression. Stratification is achieved by considering only those cases, where both parents survived by age 50, which makes a sample more homogeneous with regard to parental lifespan. In addition to this, the data on parental lifespan above age 50 are also included in the multivariate regression model as categorized predictor variables (grouped into five-year intervals). The following additional predictor variables are also included in the final model because of their predictive value: (1) calendar year of birth, (2) ethnicity (Russian British and others), (3) cause of death (violent vs non-violent), (4) nobility rank of the father (indicator of social status), (5) maternal age at childbirth (in order to discriminate between maternal and paternal age effects), (6) large family size (number of siblings 11+), (7) late paternal age at first childbirth (45-49 years and 50+ years), (8) young maternal age at last childbirth (before 25 years). The F-value for the final regression model is 10.12 ($p<0.0001$). Note that there were 103 cases of women born to father at ages 55-59 years and 160 cases of women born to young father (15-24 years).

Results for **Table 8** "**Male** lifespan as a function of paternal age at reproduction" are obtained through analysis of lifespan data (outcome variable) for **5,313** men born in 1800-1880 (extinct birth cohorts), who survived by age 30 (to focus on analysis of adult lifespan). In order to avoid confounding of parental age effects by selective parental survival (short-lived parents are always young parents, because dead parents do not reproduce), the two methods are simultaneously used: (1) stratification, and (2) regression. Stratification is achieved by considering only those cases, where both parents survived by age 50, which makes a sample more homogeneous with regard to parental lifespan. In addition to this, the data on parental lifespan above age 50 are also included in the multivariate regression model as categorized predictor variables (grouped into five-year intervals). The following additional predictor variables are also included in the final model because of their predictive value: (1) calendar year of birth, (2) ethnicity (Russian British and others), (3) cause of death (violent vs non-violent), (4) nobility rank of the father (indicator of social status), (5) maternal age at childbirth (in order to discriminate between maternal and paternal age effects), (6) large family size (number of siblings 11+), (7) late paternal age at first childbirth (45-49 years and 50+ years), (8) young maternal age at last childbirth (before 25 years). The F-value for the final regression model is 8.46 ($p<0.0001$). Note that there were 99 cases of men born to father at ages 55-59 years and 152 cases of men born to young father (15-24 years).

Appendix 4

Results for **Table 9** "**Female** lifespan as a function of paternal age at reproduction" are obtained through analysis of lifespan data (outcome variable) for **3,622** women born in 1800-1880 (extinct birth cohorts), who survived by age 30 (to focus on analysis of adult lifespan). In order to avoid confounding of parental age effects by selective parental survival (short-lived parents are always young parents, because dead parents do not reproduce), the two methods are simultaneously used: (1) stratification, and (2) regression. Stratification is achieved by considering only those cases, where both parents survived by age **60**, which makes a sample more homogeneous with regard to parental lifespan. In addition to this, the data on parental lifespan above age 60 are also included in the multivariate regression model as categorized predictor variables (grouped into five-year intervals). The following additional predictor variables are also included in the final model because of their predictive value: (1) calendar year of birth, (2) ethnicity (Russian British and others), (3) loss of father during formative years of childhood (before age 15), (4) cause of death (violent vs non-violent), (5) nobility rank of the father (indicator of social status), (6) parental age gap (5 years and above), (7) maternal age at childbirth (in order to discriminate between maternal and paternal age effects), (8) large family size (number of siblings 11+), (9) late paternal age at first childbirth (50+ years). The F-value for the final regression model is 11.3 ($p<0.0001$). Note that there were 87 cases of women born to father at ages 55-59 years and 114 cases of women born to young father (15-24 years).

Results for **Table 10** "**Male** lifespan as a function of paternal age at reproduction" are obtained through analysis of lifespan data (outcome variable) for **3,776** men born in 1800-1880 (extinct birth cohorts), who survived by age 30 (to focus on analysis of adult lifespan). In order to avoid confounding of parental age effects by selective parental survival (short-lived parents are always young parents, because dead parents do not reproduce), the two methods are simultaneously used: (1) stratification, and (2) regression. Stratification is achieved by considering only those cases, where both parents survived by age **60**, which makes a sample more homogeneous with regard to parental lifespan. In addition to this, the data on parental lifespan above age 60 are also included in the multivariate regression model as categorized predictor variables (grouped into five-year intervals The following additional predictor variables are also included in the final model because of their predictive value: (1) calendar year of birth, (2) ethnicity (Russian British and others), (3) loss of father during formative years of childhood (before age 15), (4) cause of death (violent vs non-violent), (5) nobility rank of the father (indicator of social status), (6) parental age gap (5 years and above), (7) maternal age at childbirth (in order to discriminate between maternal and paternal age effects), (8) large family size (number of siblings 11+), (9) late paternal age at first childbirth (50+ years). The F-value for the final regression model is 8.77 ($p<0.0001$). Note that there were 80 cases of men born to father at ages 55-59 years and 99 cases of men born to young father (15-24 years).