JOURNAL OF ANTI-AGING MEDICINE
Volume 4, Number 1, 2001
Mary Ann Liebert, Inc.

Epidemiology of Human Longevity: The Search for Appropriate Methodology

LEONID A. GAVRILOV and NATALIA S. GAVRILOVA

ABSTRACT

In contrast to the remarkable progress in genetics of aging of such animal models as yeast, nematodes, and Drosophila, little is known about mechanisms that control human longevity. The main obstacle in human studies is that the opportunities for direct experiments with humans are limited and therefore data collection through observations, i.e., epidemiological methods, are particularly important. To advance scientific knowledge in this area, it is also important to broaden the arsenal of concepts and methods for human longevity studies and to develop alternative tactics to cope with environmental and social confounding. To cope with environmental and social confounding, this paper suggests two robust exploratory tests with low risk of artifactual results, based on the analysis of two kinds of genetic influences on human longevity: (1) parental consanguinity, which increases the proportion of homozygotes in offspring, leading to the expression of recessive traits and an increased incidence of multifactorial traits (via increased variance for genetic liability distribution); and (2) advanced paternal age at conception, which is known to be one of the major sources of new mutations in human populations. This paper also describes methodologies to study the role of environmental factors (shared familial environment and early seasonal events) as determinants of human longevity.

INTRODUCTION

In contrast to the remarkable progress in genetics of aging of non-human species,[1–5] little is known about mechanisms that control human longevity. What is behind the records of extreme human longevity: just lucky chance, favorable environment, or "good" genes? How can one resolve the apparent controversy between strong familial clustering of human longevity, and poor resemblance in life span among blood relatives.[6–8] What is the nature of genetic component for such a complex quantitative trait as human longevity: special "longevity assurance genes" or just an individual variation in the burden of deleterious mutations? These fundamental problems are still unresolved.

The main obstacle in human studies is that the opportunities for direct experiments with humans are limited and therefore data collection through observations, i.e., epidemiological methods, are particularly important. To advance scientific knowledge in this area, it is also important to broaden the arsenal of concepts and methods for human longevity studies and to develop alternative tactics to cope with environmental and social confounding.

The epidemiology of human longevity is a particularly challenging area of research because this complex trait is strongly affected by

Center on Aging, NORC/University of Chicago, Chicago, Illinois.

many genetic, environmental, and social factors. Since there are many extrinsic and stochastic factors causing premature deaths, the penetrance for the longevity phenotype may be extremely low and unstable across time and space. Also, since there are many different diseases and causes of death, the longevity phenotype may be genetically highly heterogeneous with quite different modes of inheritance of this "broad survival trait"[9] in different families. Thus, before applying sophisticated statistical methods of complex segregation analysis (with numerous heavy assumptions) to the pooled highly heterogeneous data, it is important first to run several robust tests with a low risk of artifactual results, and to determine the best approaches (and data resources) to the complex problem of human longevity genetics. This paper will also describe methodologies to study the role of environmental factors (shared familial environment and early seasonal events) as determinants of human longevity.

TWO PROMISING APPROACHES TO STUDY THE GENETICS OF HUMAN LONGEVITY

This section describes two particularly promising approaches to study the genetics of human longevity, which may result in the development of new methodologies in this area of research. These two approaches (parental consanguinity test and paternal age test) proved to be effective in human genetics, but they have never been applied systematically to the studies of human longevity. The scientific background behind these two promising approaches to longevity studies is discussed below.

Parental consanguinity test: rationale

The test for parental consanguinity among affected persons has become a widely recognized instrumental approach for the validation of the autosomal recessive nature of the "condition" under study (in our case, human longevity). The rationale for using this approach is based on the knowledge that mating between close relatives (inbreeding) increases the proportion of homozygotes in offspring. Therefore, through inbreeding, recessive genes are more easily brought to the fore.[10] Thus, parental consanguinity among affected individuals is a strong indicator for the recessive nature of the condition under study.[11] For example, the percentage of first-cousin parents in children with known recessive diseases proved to be remarkably high: 33% in the case of alcaptonuria, 40% in the case of ichthyosis congenita, and 54% in the case of microcephaly.[12,13] For comparison, the percentage of first-cousin marriages in human populations is typically in the range of 0.5–5%.[14]

On the other hand, the lack of evidence for parental consanguinity among affected individuals is considered to be a strong argument against the autosomal recessive mode of trait inheritance.[15–19] Note, however, that if the trait is common, the consanguinity effect may be too small to be detectable.[11] For this reason, the parental consanguinity test is most effective for rare traits (such as exceptional longevity).

The consanguinity test is based on comparisons of the inbreeding coefficients calculated for affected and control individuals. The coefficient of inbreeding, F, of an individual is the probability that this individual receives, at a given locus, two genes that are identical by descent—that is, are copied from a single gene carried by a common ancestor.[10] For example, the inbreeding coefficient is equal to $\frac{1}{8}$ for the offspring of uncle-niece or aunt-nephew matings. In the case of the offspring of first cousins, who have two grandparents in common, the inbreeding coefficient is equal to $\frac{1}{16}$. For comparison, the average inbreeding coefficient (a) in human populations is in the range between 0.001 and 0.02 (for isolates).[10,14]

The use of the consanguinity test is particularly appealing in the case of genetic epidemiologic studies of human lifespan. In this case, the traditional approaches based on segregation analyses,[20] path models,[21] and recurrence risk analysis[22–24] encounter the fundamental problem of confounding by environmental and social variables that are often not measured. In other words, the lifespan of blood relatives may be correlated not only because of common genes but also because of shared cultural environment for which it is difficult to control. This problem is relaxed in the case of the consan-

guinity test, because one of the key variables in this analysis, the inbreeding coefficient, is fixed for each individual at the moment of his conception and is not affected by any further circumstances of his life, including education, income, and occupation. Therefore, the consanguinity test is robust to environmental and social factors. If the relationship between parental consanguinity and offspring lifespan is detected, the most plausible explanation could be found in the area of genetics, since environmental and social variables do not affect the inbreeding coefficient *post factum*. However, different socio-economic groups may have different levels of inbreeding, and this possibility should be taken into account. This can be done by using very close matched controls (i.e., the copy-pair approach).

We set out to illustrate why human longevity could be inherited as a recessive trait and may be related to parental consanguinity. Caloric restriction in mammals is known to produce a remarkable increase in lifespan,[25,26] delay in age-related pathologies,[27,28] and retardation of age-related changes in gene expression.[29] It is reasonable to hypothesize that the same effect could be produced not only by food restriction, but also by certain metabolic mutations that interfere with proper assimilation of food (caloric restriction because of food assimilation deficiency). Such conditions, caused by deficiencies in some key enzymes, are expected to be recessive since the normal enzyme coded at the homologous chromosome usually restores the wild phenotype. If this hypothesis is correct, there are good reasons to expect that long-lived people, homozygous for recessive food assimilation deficiency, come more often from consanguinous matings. It is interesting to note that long-lived people are usually very lean indeed, and, furthermore, the majority of centenarians "have never been obese."[30] Another interesting note is that such people with impaired food assimilation may be at higher risk to die in early childhood. Those of them who survive to adult age may live a very long life. It is known that parental consanguinity results in increased infant and child mortality.[31,32] The only missing step in this hypothesis is to test whether parental consanguinity is also associated with remarkable longevity. Additional support for this idea came from the preliminary study of three longevous families that revealed "a strong likelihood of family intermarriage by cousins in two of the three families."[30]

Consanguinity test and the multifactorial threshold inheritance. The increased inbreeding coefficients among affected individuals can be observed not only for recessive traits, but also for multifactorial traits with threshold inheritance.[14] This is because the variance, V_F, of the distribution of genetic liability in the population with inbreeding coefficient F, is higher than the variance, V_0, in a population with random mating:

$$V_F = V_0\ (1 + F) \tag{1}$$

The increased variance in a population with inbreeding coefficient F is attributed to increased frequency of otherwise rare homozygotes, corresponding to extreme phenotypes.[11] Equation 1 assumes no dominance.[11] With dominance, the relationship becomes more complicated, but if the rarer genes tend to be recessive (or partially so), the variance-enhancing effect is increased.[11] Thus, both autosomal recessive and multifactorial characters will increase in incidence with inbreeding.[14] Discrimination between these two models requires a careful study of the dependence between the incidence rate of the trait at different cut-off levels and the inbreeding coefficient.[14] In the case of multifactorial threshold inheritance, the relationship between the inbreeding coefficient and the trait incidence is much less pronounced, compared to a recessive mode of inheritance.

With a recessive trait, there is a relationship between the consanguinity effect and the allele frequency. In particular, the ratio, R, of the trait frequency in persons of inbreeding coefficient F to that of unrelated persons is

$$R = 1 + F(1 - q)/q$$

In this formula, q is the frequency of the recessive allele. If the consanguinity rate is higher than predicted by this formula, it suggests that more than one recessive gene is able to produce the effect.[11] This approach may be used to es-

timate the number of recessive genes involved in human longevity.

If longevity is a multifactorial threshold trait, it can be predicted that the long-lived fathers will have longer lived children than long-lived mothers. This is because the liability threshold is higher for males than for females (since fewer males survive to the highest ages). Thus, the average "genetic liability to longevity" of long-lived males is greater than for long-lived females and they should transmit more longevity genes to offspring. This prediction deserves to be tested in future studies. It might show up as a higher longevity penetrance in the progeny of long-lived fathers than of long-lived mothers. This prediction is based on a general argument known as the Carter effect,[14,33] which states that the sex with the lower incidence should have a higher proportion of affected relatives in the case of multifactorial inheritance (the classic example is pyloric stenosis[33]).

Methodology of the parental consanguinity test. For each selected long-lived individual with validated longevity and validated family history, the inbreeding coefficients should be calculated, using all known relationships in the genealogy. The calculations could be done using the program "Kinship" based on Boyce algorithm.[34,35] Also, the "PedHunter"—a new software package for the analysis of pedigrees with large genealogies based on graph theory—could be applied.[36] This software package provides computation of the inbreeding coefficients using the algorithm presented in Weir.[37] The coefficient of inbreeding, F_i, of an individual is defined as the probability that the two alleles at a locus are identical by descent.[11,38] For the pedigree with a common ancestor, the formula that is used for calculation of F_i can be generalized for each common ancestor:

$$F_i = \Sigma[(\tfrac{1}{2})^n \times (1 + F_a)],$$

where n is the number of individuals in a given path connecting one relative to the other through a common ancestor, F_a is the inbreeding coefficient of the ancestor, if applicable, and these components are summed over all possible paths between the individual and common ancestor.[11,38,39] According to this formula, there is a rapid decrease in inbreeding coefficient with adding each generation to common ancestor (reduction by one-half for each additional generation).[10] For this reason, information on three generations of ancestors is considered to be sufficient for research purposes, since the coefficient of inbreeding of individual increases only very slowly with the number of additional remote common ancestors.[14] To adjust for the remote common ancestry that is beyond the depth of known pedigrees, the isonymy method could be applied, which is based on the analysis of marriages between persons with the same surname.[40–42] Since the depth of known pedigrees may vary for different individuals, it would be useful to adjust the estimates of inbreeding coefficients for the genealogical length of known ancestry.[43]

At the first, exploratory stage of analysis, the average inbreeding coefficient (a) for long-lived individuals could be calculated and compared with the average inbreeding coefficient for typical human populations.[10,14]

Pitfalls and limitations. One of the possible sources of artifacts in these studies may be related to stratification of consanguinity in human society. For example, members of the royal family may have both better survival to the oldest ages (because of their privileged status) and a higher degree of inbreeding (because of more limited mating choices). To cope with this problem, it is important to use the matched controls for each long-lived person with known inbreeding coefficients. The closest possible matched control could not be a full sibling of the index case, because all full siblings have the same inbreeding coefficients. Half-siblings and first cousins are the closest family members that can be used as matched controls in such kind of studies (on condition that they are not long-lived, e.g., died from non-violent causes before age 75). Thus, the copy-pair approach could be applied—the inbreeding coefficients for long-lived persons and their matched controls (half-siblings or first cousins) can be calculated and compared (see Fig. 1 as an example).

The reconstruction of additional pedigrees for matched controls may take a significant

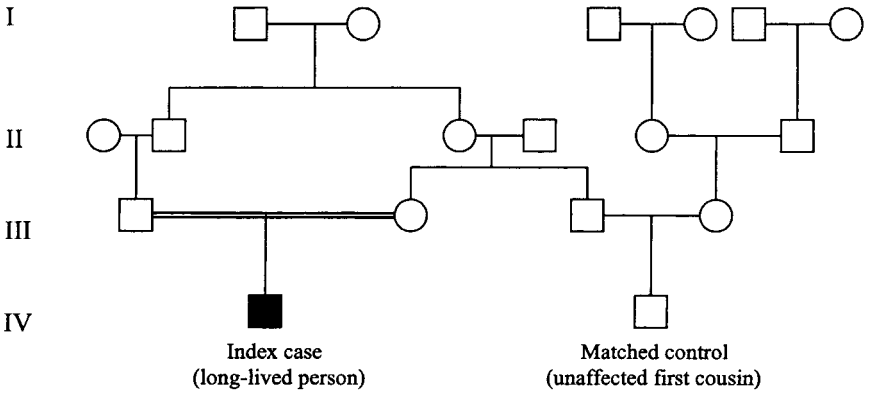

FIG. 1. In this hypothetical case, a long-lived inbred individual was born in the 4th generation as a result of consanguinous mating (symbolized by the double horizontal line). The matched control (first cousin) was unaffected and not inbred.

amount of time and effort. However, it should be done, because the data on inbreeding coefficients in matched controls are of critical importance for correct application of the consanguinity test. The resulting correlated sample could be analyzed using direct difference method and Sandler's A-Statistic.[44] Other analytical methods include conditional multiple logistic regression for matched sets,[45–47] as well as methods previously applied to study the effects of inbreeding on early childhood mortality.[31] The correlation coefficient between levels of inbreeding in long-lived persons and the matched controls could be estimated, which is important for further studies as a measure of sample stratification. If the correlation coefficient proves to be insignificant, this will justify the decision not to use the laborious matched control design in future studies. While analyzing the data, it is important to keep in mind that inbreeding may have an unusual polarizing effect on human lifespan: cases of inbreeding may be concentrated both among premature deaths[31] and among people with exceptional longevity (as expected for multifactorial traits[14]).

Testing the alternative hypothesis: heterosis and longevity. One of the possible outcomes of the consanguinity test may be that there may be no increased consanguinity among the long-lived persons. Moreover, the opposite trend may be observed: the inbreeding coefficients of long-lived persons may be lower compared to matched controls. Such a finding would suggest that the inbreeding had only deleterious effects on human health and survival (inbreeding depression). In this case, the pedigrees of long-lived individuals could be reanalyzed to test for possible heterosis or "hybrid vigor."[14,48,49] The heterosis hypothesis predicts that parents of long-lived individuals should be genetically more distant compared to the parents of matched controls. Some available pedigrees provide information on parental places of birth, which could be used as a proxy for genetic distances between parents.[50] The concordance rate for parental country of birth can be measured and compared for long-lived persons and their matched controls. The heterosis hypothesis predicts that the parents of long-lived persons should be more discordant (more often born in different countries) compared to matched controls. This hypothesis could be tested, and in the case of its support by the data it could be studied more thoroughly.

More careful tests for heterosis effects is based on calculation and comparison of the physical distances between parental places of birth—a standard approach to studying heterosis in human populations.[50] According to this approach, parents born in Sweden and Norway are genetically closer than parents born in Sweden and Spain (on average). The heterosis hypothesis predicts that physical distance between parental places of birth should be higher for long-lived persons compared to matched controls. The methods of data analysis are similar to those described earlier: the re-

sulting correlated sample could be analyzed using direct difference method and Sandler's A-Statistic.[44] Other analytical methods may include conditional multiple logistic regression for matched sets.[45–47]

Pitfalls and limitations. Testing the heterosis hypothesis is complicated by the fact that not all the distant parents necessarily produce a heterosis effect.[50,51] It may happen, for example, that mating between Spanish and Danish persons result in heterosis (longevous phenotype in our case), while mating between the Spanish and Dutch may demonstrate no heterosis. To address this issue, the ethnic combinations of mating may be studied among parents of long-lived persons (compared to matched controls) to see whether there are any particular ethnic combinations of parents (and directions of mating) favorable for offspring longevity.

Paternal age test: rationale

The test for advanced paternal age at conception of the affected individuals is an important instrumental approach for suggesting a sporadic autosomal dominant nature of the condition under study.[15–19] The rationale for using this test is based on the knowledge that advanced paternal age at reproduction is the major determinant for new dominant mutations including DNA base substitutions and other copy-errors produced during numerous DNA replications in male sperm cells.[14,52–55]

Advanced paternal age is known to be associated with an increase in new dominant mutations in offspring that result in congenital anomalies.[56–68] In particular, paternal age is responsible for new dominant autosomal mutations that cause different malformations, including achondroplasia,[56,57,60] Apert syndrome,[56,57] Marfan syndrome,[56,57] osteogenesis imperfecta,[67,68] and other inherited diseases. Older paternal age was observed among patients with Costello syndrome,[69] chondrodysplasia punctata,[70] fibrodysplasia ossificans progressiva,[71,72] and thanatophoric displasia.[73]

The evidence of advanced paternal age in the case of Hutchinson-Gilford syndrome ("progeria" or "progeria of childhood") was considered an important argument that this disease is most likely a sporadic autosomal dominant condition.[15–19]

Recent studies on Alzheimer's disease revealed that, in the case of the sporadic form of this disease, the paternal age at conception of affected individuals was significantly higher than in the control group, while maternal age differences were statistically insignificant.[74]

At first sight, it may seem counter-intuitive to expect that long-lived individuals were conceived to older fathers, since so many diseases are associated with late fatherhood. However, this contradiction is more apparent than real. After all, the advanced paternal age is merely a proxy indicator for new mutations. It is not surprising that most of the mutations are deleterious. However, a few of them may be advantageous. Otherwise, biological evolution could never take place. Therefore, advanced paternal age could be observed both in the cases of premature deaths,[75] as well as in the cases of remarkable longevity. It is really surprising that this test has never been applied to the studies on the genetics of exceptional longevity.

One particularly appealing feature of the paternal age test is its robustness to environmental and social confounding. This is because the parental age at conception is a kind of variable that is fixed forever for each particular individual (by definition, at the age of his conception). It is not affected by any further events and by variables, like education, income, and occupation. Therefore, if the relationship between advanced paternal age and human longevity is detected, the first possible explanation for it should be searched in the field of genetics. However, different socio-economic groups may have different typical paternal ages, and this possibility should be also taken into account by using matched controls within the same families.

Methodology of the paternal age test. The paternal age test is based on the prediction of advanced paternal age at conception for affected individuals, if the condition under study is caused by new dominant mutations. In this case, the closest matched controls are unaffected siblings of the affected individuals born

by the same couple. The difference (contrast) between paternal age at conception of long-lived individual and the mean father's age at conception for all other unaffected siblings could be calculated within each particular family. The mean of these differences calculated across all the families with long-lived persons is expected to be positive (and statistically significant), if the longevous phenotype is related to new dominant mutations. Statistical methods of data analysis could be similar to those described earlier.[44–47]

In the case of the paternal age test, it is particularly important to analyze the data for long-lived daughters and sons separately. Our previous studies revealed that paternal age effects are sex-specific (only daughters are affected), perhaps because of the specific inheritance of paternal X chromosome by daughters.[76–78]

Pitfalls and limitations. The true biological fathers may be different in some cases from the official claims. Also, in such cases the biological fathers tend to be younger than the husbands to whom the child was assigned. For this reason, the observed paternal age may be smaller than would be expected in the case of perfect data (without paternity mistakes). These concerns should be taken into account and the results of data analysis should be treated cautiously with the understanding that the paternal age effect is probably understated. Also, all cases of extremely late fatherhood (after 60 years) should be considered as suspicious cases, and a sensitivity analysis should be made with these cases excluded. The same procedure should be done for other doubtful paternity cases. For the reasons described above, the paternal age effects are likely to be underestimates, so if one finds a significant paternal effect it will certainly be important.

ENVIRONMENTAL EXPOSURES AND HUMAN LONGEVITY

Two exploratory tests will be described in this section that may help to understand the importance and the nature of environmental exposures on human longevity. The Spousal Lifespan Test could be used to explore the role of shared familial environmental exposures in adult ages (for married persons). The Month-of-Birth Test could be applied to estimate the role of early seasonal environmental impacts during critical periods of fetus/child development on subsequent adult lifespan. The scientific background behind these two research approaches is described below.

Spousal Lifespan Test: rationale

The Spousal Lifespan Test is based on studies of the lifespan resemblance between spouses in an attempt to understand the importance of environmental exposures on human longevity.[79–82] This test is based on the knowledge that spouses from typical non-consanguinous marriages do not have common genes, identical by descent, but they share many environmental exposures in common, such as the following:

1. Diet—the amount of food consumed, its quality and composition, including carcinogens produced during some cooking procedures, food pollution by pesticides, and contamination with bacteria and viruses.
2. Common infectious agents, including common exposure to tuberculosis in previous years and infections from other family members and pets.
3. Quality of water supply (pollution with heavy metals, contamination with infectious agents, content of calcium, magnesium, and microelements).
4. Air pollution levels (in-house radon levels, allergens, dust, heavy metal, and asbestos exposures from in-house paint and construction materials in the previous years; passive smoking when the other spouse smokes).
5. Common exposure to local climate and meteorological factors (air temperature and humidity, heat waves, cold weather, levels of solar radiation, temperature fluctuations, etc.).
6. Local industrial pollution, including radiation, noise levels, etc.
7. Socio-economic characteristics of the households are identical for both spouses, and they tend to have similar access to medical care of the same quality.

8. In addition to the risk factors, protective factors also exist,[9] and the spouses may share some of these protective factors during their cohabitation.

The Spousal Lifespan Test has one important advantage over such traditional approaches as twin studies[83,84] and adoption studies,[85,86] which are routine procedures to estimate the relative contributions of genes and environment in human lifespan. These two standard procedures cannot be applied to study rare cases of exceptional longevity for one simple reason: it is virtually impossible to find enough cases to study centenarian monozygotic twins or to find adopted centenarians. The spousal test, on the contrary, is a more feasible approach in this case, since most of the centenarians have or had a spouse. Since marriages are much more common in human populations than twinning or adoptions, the spousal lifespan test is probably the best way to explore the role of shared familial environment in exceptional longevity. This is an important issue, because exceptional longevity may have a higher familial component compared to "normal" lifespan.[7,87,88]

There are some controversies in previous studies of spousal lifespan resemblance. Wyshak[80] and Philippe[79] found significant correlation between spousal lifespans, which was used, according to Philippe, "as an environmental component index in longevity." However, "after age 50, the environmental correlation between spouses vanishes and so does the phenotypic correlation in relatives."[79] The following conclusion was made: "parent–offspring correlations as well as sib correlations are of the same order of magnitude as that between spouses for various age groups at death. It is suggested that heritability of survival is nearly zero."[79] The opposite trend was observed by other authors, who found no resemblance in lifespan between shorter-lived spouses and significant resemblance after 50 years of age that was attributed to cohabitation effect.[81]

Contrary to all previous studies, Desjardins and Charbonneau[82] found that "there is no relationship between ages at death of spouses, who are related by marriage, but not by blood." Similar conclusion was made by Perls and Silver[30]: "In fact, the people who would be most likely to share our centenarians' lifelong habits and benefit from their example were not necessarily aging well. These were the centenarians' spouses. If a shared diet or shared attitudes were responsible for the sibling's longevity, we should have seen at least some statistical difference in the spouses' length of life. But there was no visible correlation between being the spouse of a centenarian and living to extreme old age. As a group, the spouses' life spans were just average."[30] To resolve these controversies, it is important to re-evaluate the previous findings using larger samples of long-lived persons and carefully chosen matched controls.

One of the astonishing features of centenarians is a surprisingly high proportion of married persons among them. Theoretically, almost all people with exceptional longevity should be already widowed if they were ever married before. In reality, however, about 4% of centenarian-women and 25% of centenarian-men are still married.[89] This is much higher than expected on the basis of cohort life tables provided by the Social Security Administration[90]: about 0.4% for centenarian women and 2% for centenarian men. Two explanations of this paradox are possible: (1) centenarians' spouses do live longer than other spouses and/or (2) centenarians' spouses are much younger than their long-lived partner. The latter explanation is also extremely interesting because it provides a clue as to how a person could survive to extreme old age thanks to the care provided by the younger spouse. This indicates that cases of married long-livers are probably more "environmental" than "genetic," compared to cases of widowed long-lived persons, who should have much higher genetic potential for long life. This idea may be useful while selecting the longevous families for genetic analysis.

The resemblance between spouses may also be caused by assortative mating,[11,38] and this limitation of the spousal lifespan test is discussed later.

Methodology of the Spousal Lifespan Test. The Spousal Lifespan Test is based on testing the hypothesis that spouses of long-lived persons

are expected to live longer life if shared familial environment is important. Two kinds of pairs of long-lived persons and their spouses should be identified: (1) pairs with long-lived female partners and (2) pairs with long-lived male partners. For each spouse, the closest possible control is his/her same-sex sibling who was married to another non-affected person (not long-lived, that is, died before 85 years). For example, for a centenarian-man, his spouse (as index case) and the sister of the spouse (as control) should be considered. For a centenarian-woman, her spouse (as the index case) and the brother of the spouse (as control) should be considered. In those cases when data on several same-sex siblings of the spouse are available, the married same-sex sibling with the closest birth year (to spousal birth year) should be selected as a control. The hypothesis should be tested that spouses of long-lived individuals tend to live longer compared to their same-sex siblings, married to non-affected individuals. The resulting correlated sample (spouses of long-lived individuals matched with the siblings of the spouses) may be analyzed using the direct difference method and Sandler's A-Statistic.[44] Other analytical methods may include conditional multiple logistic regression for matched sets.[45–47] The triplets (long-lived person, spouse, spousal same-sex sibling) should be analyzed separately for long-lived men and women, and then the two groups should be compared to see whether there are any gender differences depending on the sex of long-lived person.

Pitfalls and limitations. The resemblance between spouses may be caused not only by the shared environmental exposures during spousal cohabitation, but also by the assortative mating.[11,38] Assortative mating occurs when spouses resemble each other phenotypically, because they select each other rather than mate at random. As a result of such mutual selection for marriage, husbands and wives are correlated for skin color, height, IQ, personality types, smoking habits, disabilities (deafness, blindness, etc.) and specific diseases (tuberculosis, diabetes, etc.). However, it seems unlikely that spouses could select each other for marriage on the basis of future lifespan, which is not known in advance. Only indirect selection through other traits is possible.

The difference between the assortative mating hypothesis and the cohabitation hypothesis is that the resemblance between spouses does *not* depend on the duration of cohabitation in the former case and increases with cohabitation exposure in the latter case. Therefore, spousal pairs with different ages at marriage should be compared in order to discriminate between these two possible explanations. The lifespan of those spouses who married the long-lived partner at younger ages is expected to be higher if the duration of cohabitation is important.

Test for early seasonal environmental impacts

The test for early seasonal environmental impacts allows us to validate the hypothesis that early seasonal environmental exposures in the past (such as seasonal vitamin deficiency) may affect human survival in later life. The rationale for this approach is based on the Ames theory[91] that micronutrient deficiencies play a major role in DNA damaging, human aging, and premature deaths from cancer and heart disease. Deficiencies of vitamins B_{12}, folic acid, B_6, niacin, and vitamins C and E appear to mimic radiation in damaging DNA by causing single- and double-strand breaks, oxidative lesions, or both.[91] These health hazards are highly significant because even now in such a developed country as the United States half of the population may be deficient in at least one of these micronutrients.[91] In previous years, when the people who are now elderly were born, vitamin deficiencies were even more acute, particularly in the late-winter season, just before vegetation starts anew (February in the northern hemisphere).

It is reasonable to hypothesize (and to test this hypothesis) that vitamin deficiencies during critical periods of fetus and infant development may affect the later health and longevity of the deficiency-exposed birth cohorts. For example, preceding vitamin deficiencies in February in the past may produce a subsequent lifespan-shortening effect in February birth cohorts among adults. The same February avitaminosis during the third month of

pregnancy may produce another fragile birth cohort born in August. The third month of pregnancy is known to be a critical period when the brain is vulnerable, when the nervous system and sense organs develop, when all of the major organs have been established, and when the embryo becomes a fetus.[92,93] Preliminary studies have confirmed that there are two seasonal minimums in adult life span for those born in February and in August.[94] Adult lifespan minimum in August birth cohorts was also found in the earlier studies.[95] In general, all previous studies found statistically significant seasonality in adult lifespan according to month of birth, but there is controversy over the exact seasonal pattern of lifespan fluctuations.[94–97] Further studies are required in order to validate the previous findings, address the existing controversies, and explore the possible mechanisms of lifespan seasonality.

Early seasonal impacts on subsequent adult lifespan may include not only seasonal vitamin deficiency, but also other seasonal impacts, such as infectious diseases. Seasonal peaks of disease occurrence are typical for many conditions,[98] including tularemia and Rocky mountain spotted fever (spring-early summer), the St. Louis encephalitis and other viral encephalitides (late summer-early fall), influenza (mid-winter), measles (rubeola, late winter-early spring), enteric bacterial infections (summer), poliomyelitis (peak in July-August, minimum in March), and infectious virus hepatitis (late winter). Some diseases have additional cyclic variation with a periodicity of longer than 1 year,[98] such as, for example, measles (rubeola, 3-year cycle) and meningococcal meningitis (7–9-year cycle). The most drastic effects of infectious agents in pregnancy, which probably represent the tip of the iceberg of the damage to progeny,[92] include the following:

1. For the rubella virus (German measles): cardiac malformation, deafness, cataracts, glaucoma, and tooth defects.
2. For cytomegalovirus: growth retardation, blindness, mental retardation, and deafness.
3. For the herpes simplex virus: microcephaly and mental retardation.
4. For varicella (chickenpox): skin scarring, muscle atrophy, and mental retardation.
5. For poliovirus: adult schizophrenia.[99] Poliovirus epidemics peak in July-August, and exposure to this virus in the second trimester of gestation seems to produce subsequent adult schizophrenia in February birth cohorts.[99] Adult schizophrenia is also associated with neonatal meningitis caused by another enterovirus, Coxackie B5.[100]

Thus, both infectious agent exposures and vitamin-deficiency exposures should be considered for possible explanation and study of the early seasonal environmental impacts on adult lifespan.

Methodology of the test for early seasonal environmental impacts. The method is based on testing the hypothesis that there is an association between individual month of birth and longevity. Seasonal distribution by month-of-birth for long-lived individuals should be studied and compared with a control distribution. The closest possible controls in this case are non-affected siblings (died before 85 years) of long-lived individuals born in the same families and having the closest possible birth year (but not twins). Seasonal distribution by month-of-birth for long-lived individuals and for matched controls may be compared by standard statistical methods including the χ^2 test[44] to see whether the difference between these two distributions is statistically significant. Also, the two seasonal distributions by month-of-birth for long-lived individuals of each sex could be compared using the χ^2 test to see whether there are any significant gender differences and whether these two samples could be pooled together to increase the statistical power.

If persons born in specific months have shorter adult lifespan, as the preliminary data suggest for February and August birth cohorts,[94] their frequencies among long-lived persons will be lower, compared to the frequencies in other months-of-birth, and also compared to relative frequencies in the control group of non-affected siblings. This hypothesis could be tested for each month of birth separately using the χ^2 test, relative risk, and odds ratio estimates.[44,98]

Pitfalls and limitations. While comparing monthly frequencies, it is important to adjust

the data for the different durations of different months (28 days in February versus 31 days in August, etc.). The frequencies should be adjusted proportionally to the number of days in each month.

Another possible source of artifacts is related to the seasonal differences in birth rates in the past, which affect the current month-of-birth distribution of long-lived persons. This problem could be addressed using closely matched controls (non-affected siblings of long-lived individuals born in the same families and having the closest possible birth year).

SELECTION OF DATA RESOURCES AND PROBLEMS OF DATA QUALITY

Criteria for selection of the data sources. The methods described earlier require high-quality data meeting the following criteria:

1. The data should contain a sufficient number of records (several thousands) for long-lived individuals (90+ years), since the main focus is on long-lived persons and their relatives.
2. The data for long-lived persons should be initially highly reliable, because checking thousands of dates with the original archival documents is too laborious and time-consuming.
3. An extensive family history for each long-lived person should already exist, to allow reconstruction of the genealogical trees for up to five ancestral generations for computation of the inbreeding coefficients.
4. The data should represent the population where consanguinity is (or was) a quite common phenomenon (in order to study its effects).
5. To apply the paternal age test it is important that there be significant variation in the parental age at conception in the study sample, with a sufficient number of cases of late fatherhood.
6. It is desirable to have an opportunity for subsequent extension of the exploratory research via inclusion of living descendants of long-lived individuals into further full-scale longitudinal studies of the "rate-of-change traits" and "survival traits" (including development of repositories of cells and/or DNA for linkage analysis.[9]

Recommended data sources. Data on European nobility families seems to be particularly promising because these data meet all the selection criteria described above:

1. These data contain a sufficient number of long-lived individuals to study. For example, we have already identified 1,565 long-lived persons (90+ years). Some of them (333 persons) lived 95 years and more. Others (51 persons) were centenarians. Furthermore, this data resource provides researchers with opportunities for significant extension of the database.
2. The data are initially highly reliable with detailed information on exact birth date (day, month, year) and birth place, taken from primary documents in archives and published in reputable professional editions (dozens of volumes of the "Genealogisches Handbuch des Adels," 1951–1994[101–104] and other sources listed in Gavrilova and Gavrilov[105]).
3. The data on long-lived individuals could be taken from genealogical records, so the major work for reconstruction of the pedigrees has been done already by previous researchers.
4. European aristocratic families are notorious for their marriages among close relatives (consanguinity). In fact, inbreeding in these families is often mentioned as a limitation of this data source. However, in the studies focused on possible effects of consanguinity, this remarkable feature of European aristocratic families is of critical importance.
5. Another remarkable feature of this data set is a significant proportion of cases with fatherhood in later life, since old kings and aristocrats often married young fertile women. For example, Queen Victoria of England, who passed the hemophilic gene to further royal generations, was born to an older father (Duke of Kent, 52 years) and received the hemophilic gene most probably as a result of paternal sperm mutation.[48,49] Such cases of late fatherhood provide researchers with an opportunity to apply the paternal age test to the selected dataset on European aristocratic families.

6. An important advantage of this data source is the public availability of the postal addresses of the living descendents, since their addresses are published in the professional genealogical books (see *Genealogisches Handbuch des Adels*, 1954–1994[101–104]). This ensures opportunities for subsequent recruitment of living descendents of long-lived individuals for further full-scale longitudinal studies of the "rate-of-change traits" and "survival traits" (including development of repositories of cells and/or DNA for linkage analysis[9]).

Data quality. We recommend to use the database on European royal and noble families that was developed in our previous studies. To develop this database, we have used one of the best professional sources of genealogical data available—the reputable German edition of the *Genealogisches Handbuch des Adels (Genealogical Yearbook of Nobility*[101–104]). This edition is known worldwide as the "Gotha Almanac" ("Old Gotha" published in Gotha in 1763–1944 and "New Gotha" published in Marburg since 1951). Data from the Gotha Almanac were often used in early biodemographic studies of fertility[106] and are used now in the studies of human longevity.[6–8,76,77]

Each volume of the "New Gotha Almanac" 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 Almanac." 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 one per 1,000 records), so this source of data is very accurate compared to other published genealogies. The sample distribution of long-lived persons from this data source is presented in Table 1.

Table 1. Sample Distribution of Long-Lived Persons With Completed Family Data (Known Birth Dates for Both Parents) According to the Year of Birth, Lifespan, and Gender

	Lifespan, years		
Birth year interval	90^+	95^+	100^+
1600–1699	2	0	0
1700–1799	75	15	1
1800–1849	153	19	6
1850–1900	1,335	299	44
Total number	1,565	333	51
Total number of males	468	78	9
Total number of females	1,097	255	42

Note that the majority of long-livers is concentrated in the recent birth cohorts, and that the nonagenarians/centenarians ratio as well as female/male ratio at advanced ages are very high, in accord with known demographic behavior for data of high quality.[107]

Pitfalls and limitations. There has been a lot of criticism concerning the value of genealogical datasets for longevity studies.[108] Our recent data analysis confirms that such a criticism is still valid in certain cases.[105,109] Incompleteness and selection bias are common for many genealogical datasets and include underreporting of children, women (resulting in sex bias), unmarried and/or infertile persons and migrants.[108–110] Since the distributions by age at death are truncated by dates of data collection that often are not specified, the data for censored, non-extinct birth cohorts are biased.[108] Another source of artifacts is related to ignoring the time trends in mortality rates in earlier studies.[108] Therefore, the special efforts should be made to cope with these limitations, and we believe, based on our previous experience,[7,109] that these problems could be avoided: (1) by using recent professional quality data on particularly well-studied aristocratic families; (2) by multiple cross-checking of the data from dozens of independent sources[105] and by complementation of pieces of information obtained from different sources; (3) by using complete descending genealogies for non-censored, extinct birth cohorts[7]; (4) by applying robust methods of data analysis with low risk of artifactual results (described earlier); and (5) most

important, by using very closely matched controls within the same families (copy-pair approach).

Data on aristocracy are sometimes criticized as not being representative, but the same can be said of widely used laboratory animals and their relationship to animals in the wild. For the purpose of this study, the data on socially elite aristocratic families are in fact preferable since environmental and social confounding is minimized in this more homogeneous group, which is not affected by poverty and starvation. Another unpublished criticism of aristocratic data is based on the speculation that a particularly strong desire for an heir among the aristocracy might lead, in cases of infertility, to illegitimate births or concealed adoptions. Even if we accept this assumption, it will result in underestimating the true effects of consanguinity and paternal age, rather than in producing spurious relationships (artifacts). To address the criticism of aristocratic data, it is important to study also non-aristocratic families as potentially supporting data from other sources such as Amish genealogies.[36]

Validation of longevity

It is always preferable to use highly reliable data, which have already been validated by previous researchers before publication. However, we recommend to make two additional steps in data validation to exclude any mistakes, that may be caused by publication misprints or by occasional errors in our data computerization:

1. Each record should be double-checked with additional data sources (another published volume of "Gotha Almanac" or other reputable professional genealogical edition). Our previous experience revealed that such cross-checking is both feasible and an effective way for validation of the record.
2. Each record should be also tested for consistency in vital dates. For example, if the birth date of some woman was recorded 20 years before the true date, this could be revealed by (a) exceptionally late birth of her first and last children, (b) very unusual spousal age gap, or (c) extremely early birth date before the parents even married. The data should contain the dates of all vital events that allow researchers to detect easily any data inconsistencies.

Important note. To ensure data reliability we recommend to use longevity records in their modest range (90–105 years). We are fortunate that there are no records in our database with claims of exceptional longevity beyond 110 years where the questionable data are usually concentrated. Also, the previous experience of other researchers on longevity validation[107] should be used in such kind of studies whenever possible.

Reconstruction and validation of family histories

For each selected long-lived person with validated longevity, the detailed family tree should be reconstructed for up to five ancestral generations. The data resources that could be used for this purpose are summarized in our special publication on this topic[105] and are also made available at the web site: *http://www.demographic-research.org/Volumes/Vol1/4/default.htm*.

The validation of the reconstructed family histories could be achieved through cross-checking with additional data sources and by testing their consistency in the dates of vital events (approach similar to validation of longevity, see previous section for more details). The methods of data quality control that could be used in such kind of studies should be based on the techniques and ideas of Louis Henry,[111] John Knodel,[112] Thomas Hollingsworth,[106,113] Jim Oeppen,[114,115] and other scientists.[116–122] The validated family histories could be computerized for further data analysis in Gedcom format[105,123] and also in the form of a relational database (with personal and familial identification numbers). Based on our previous experience, the expected success rate for family reconstitution and longevity validation is about 80%. In other words, from 1,000 long-lived persons initially selected for data validation and family reconstitution, about 800 validated pedigrees of long-lived individuals could be obtained for further analysis.

Pitfalls and limitations. A common concern with any family history is that the true biological father may be different from the official claims. This uncertainty may result both in false positive consanguinity (when it is in fact absent) and in false negative consanguinity (when it is in fact present). To cope with this problem, the analysis should be based on a comparison of the consanguinity levels in the affected (long-lived) group and matched controls, rather than on absolute numbers of consanguineous matings per se. To be on the safe side, all cases of extremely late fatherhood (after 60 years) should be considered as suspicious cases, and a sensitivity analysis should be made with these cases excluded. Also, the analysis could be limited to five generations of ancestors with the understanding that paternal uncertainty makes reconstruction of the deeper family history somewhat doubtful. Since remote ancestors contribute so little to the inbreeding, very little information is lost by truncation of pedigrees at 5 generations.

Another limitation of the data is that pedigrees for early ancestors are often limited to males only (women are missing). For this reason, the reconstruction of complete pedigrees (including women ancestors) may take a significant amount of time and effort. The completeness of pedigrees is of vital importance for correct estimation of the inbreeding coefficients.

Selection of candidate families for further studies

It is important to find particularly interesting candidate families with suggested specific modes of longevity inheritance in order to use them in further more targeted and detailed studies.

The search could be done for the following candidate families:

1. Candidate families where longevity tends to be inherited as a recessive trait, or as a multifactorial trait. If the consanguinity test reveals higher inbreeding coefficients for long-lived persons, all the families with confirmed consanguinity and longevity could be considered as candidate families to test the recessive mode of longevity inheritance in these families (against the alternative of multifactorial inheritance) by statistical methods of genetic epidemiology.
2. Families where longevity may be related to new dominant mutations. If the paternal age test reveals advanced paternal age at conception of long-lived persons (indicator of *new* mutations), all the families with long-livers born to old fathers could be considered as candidate families to test autosomal dominant mode of longevity inheritance among the descendants of long-lived persons.
3. Families where longevity may be related to heterosis. If long-lived persons tend to be born to very distant parents (or by some specific ethnic pairs of parents) the core of such families could be selected for further testing of the heterosis hypothesis.
4. Families where longevity clustering seems to be related to shared familial environment. If spouses of long-lived persons also tend to live significantly longer, those families where both spouses are long-lived could be selected to explore the role of common environment in human longevity.

These candidate families could be used for further, more targeted, and detailed studies on genetic epidemiology of human longevity.

ACKNOWLEDGMENTS

We are most grateful to Professor James F. Crow, Laboratory of Genetics, University of Wisconsin, Madison, for useful comments. We would also like to acknowledge the support from the National Institute on Aging grants.

REFERENCES

1. Johnson TE, Shook DR. Identification and mapping of genes determining longevity. In: Wachter KW, Finch CF, eds. *Between Zeus and the salmon. The biodemography of longevity*. Washington, DC: National Academy Press, 1997:108–126.
2. Shook DR, Johnson TE. Quantitative trait loci affecting survival and fertility-related traits in *Caenorhabditis elegans* show genotype-environment interactions, pleiotropy and epistasis. *Genetics* 1999; 153:1233–1243.

3. Jazwinski SM. Molecular mechanisms of yeast longevity. *Trends Microbiol* 1999;7:247–252.
4. Curtsinger JW, Fukui HH, Resler AS, Kelly K, Khazaeli AA. Genetic analysis of extended life span in *Drosophila melanogaster*. I. RAPD screen for genetic divergence between selected and control lines. *Genetica* 1998;104:21–32.
5. Vieira C, Pasyukova EG, Zeng ZB, Hackett JB, Lyman RF, Mackay TF. Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. Genetics 2000;154:213–227.
6. Gavrilova NS, Gavrilov LA, Evdokushkina GN, Semyonova VG. Mechanisms of familial transmission of human longevity: comparison of maternal and paternal contributions into offspring lifespan [Abstract 20596]. In: *Population Association of America. 1998 annual meeting. Final program and abstracts*. 1998:142.
7. Gavrilova NS, Gavrilov LA, Evdokushkina GN, Semyonova VG, Gavrilova AL, Evdokushkina NN, Kushnareva YuE, Kroutko VN, Andreyev AYu. Evolution, mutations and human longevity: European royal and noble families. *Hum Biol* 1998;70:799–804.
8. Gavrilova NS, Gavrilov LA, Evdokushkina GN, Semyonova VG. Non-linear transmission of human longevity from mother to daughter. *Gerontologist* 1998;38:8.
9. Hadley EC, Albert SM, Bailey-Wilson J, Baron J, Cawthon R, Christian JC, Corder EH, Franceschi C, Kestenbaum B, Kruglyak L, Lauderdale D, Lubitz J, Martin GM, McClearn GE, McGue M, Miles T, Mineau G, Ouellette G, Pedersen NL, Preston SH, Page WF, Province M, Schächter F, Schork NJ, Vaupel JW, Vijg J, Wallace R, Wang E, Wijsman EM. Genetic epidemiologic studies on age-specified traits. NIA Aging and Genetic Epidemiology Working Group report, 1999. Online document: http://www.nih.gov/nia/conferences/GeneticReport111199.htm.
10. Bodmer WF, Cavalli-Sforza LL. *Genetics, evolution, and man*. San Francisco: W. H. Freeman, 1976.
11. Crow JF, Kimura M. *An introduction to population genetics theory*. New York: Harper & Row, 1970.
12. Reed S. *Counseling in medical genetics*, 3rd ed. New York: Alan Liss, 1980.
13. Cummings M. *Human heredity. Principles and issues*, 3rd ed. St. Paul: West Publishing Company, 1994.
14. Vogel F, Motulsky AG. *Human genetics. Problems and approaches*. Berlin: Springer-Verlag, 1997.
15. Brown WT, Zebrower M, Kieras FJ. Progeria, a model disease for the study of accelerated aging. In: Woodhead A, Blackett AD, Hollaender A, eds. *Molecular biology of aging*. New York: Plenum Press, 1985:375–396.
16. Brown WT, Zebrower M, Kieras FJ. Progeria: a genetic disease model of premature aging. In: Harrison D, ed. *Genetic effects on aging II*. Caldwell, NJ: Telford Press, 1990:521–542.
17. Brown WT. Progeria: a human-disease model of accelerated aging. *Am J Clin Nutr* 1992;55(suppl 6): 1222S–1224S.
18. DeBusk FL, Brown WT. Progeria. In: *Nelson's Textbook of Pediatrics*, 15th ed. Philadelphia: W.B. Saunders, 1996:1997.
19. Martin GM. A means to an end: the biological basis of aging and death [Book Review]. *Am J Hum Genet* 1999;65:1477–1478.
20. Khoury MJ, Beath TH, Cohen BH. *Fundamentals of genetic epidemiology*. New York: Oxford University Press, 1993.
21. Neale MC, Cardon LR. *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers, 1989.
22. Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 1990;46: 222–228.
23. Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 1990;46:229–241.
24. Sadovnick AD, Ebers GC, Dyment DA, Risch N. Evidence for genetic basis of multiple sclerosis. *Lancet* 1996;347:1728–1730.
25. Finch CE. *Longevity, senescence and the genome*. Chicago: University of Chicago Press, 1990.
26. Sohal RS, Weindruch R. Oxidative stress, caloric restriction, and aging. *Science* 1996;273:59–63.
27. Weindruch R, Walford RL. *The retardation of aging and disease by dietary restriction*. Springfield, IL: Thomas, 1988.
28. Keenan KP, Smith PF, Hertzog P, Soper K, Ballam GC, Clark RL. The effects of overfeeding and dietary restriction on Sprague-Dawley rat survival and early pathology biomarkers of aging. *Toxicol Pathol* 1994; 22:300–305.
29. Lee CK, Klopp RG, Weindruch R, Prolla TA. Gene expression profile of aging and its retardation by caloric restriction. *Science* 1999;285:1390–1393.
30. Perls TT, Silver MH. *Living to 100*. New York: Basic Books, 1999.
31. Dorsten LE, Hotchkiss L, King TM. The effect of inbreeding on early childhood mortality: twelve generations of an Amish settlement. *Demography* 1999; 36:263–271.
32. Stoltenberg C, Magnus P, Skrondal A, Lie RT. Consanguinity and recurrence risk of stillbirth and infant death. *Am J Public Health* 1999;89:517–523.
33. Carter CO. Genetics of common disorders. *Br Med Bull* 1969;25:52–57.
34. Boyce AJ. Computation of inbreeding and kinship coefficients on extended pedigrees. *J Hered* 1983;74: 400–404.
35. Boyce AJ. *Kinship*. San Antonio: Southwest Foundation for Biomedical Research, 1988.
36. Agarwala R, Biesecker LG, Hopkins KA, Francomano CA, Schaffer AA. Software for constructing and verifying pedigrees within large genealogies and an application to the Old Order Amish of Lancaster County. *Genome Res* 1998;8:211–221.
37. Weir BS. *Genetic data analysis II*. Sunderland, MA: Sinauer Associates, 1996.
38. Crow JF. *Basic concepts in population quantitative, and evolutionary genetics*. New York: W. H. Freeman, 1986.

39. Falconer DS. *Introduction to Quantitative Genetics*. London: Longman, 1989.
40. Crow JF, Mange AP. Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen Q* 1965;12:199–203.
41. Crow JF. The estimation of inbreeding from isonymy. *Hum Biol* 1980;52:1–14.
42. Crow JF. Isonymy: a thirty-year retrospective. *Riv Antropol (Roma)* 1996;74:25–34.
43. Cazes M-H, Cazes P. Comment mesurer la profondeur généalogique d'une ascendance? *Population* 1996;51:117–140.
44. Christensen LB, Stoup CM. *Introduction to statistics for the social and behavioral sciences*. Monterey: Brooks/Cole Publishing Co., 1986.
45. Breslow NE, Day NE. *Statistical methods in cancer research. Vol. 1. The analysis of case-control studies*. IARC scientific publication no. 32. Lyon: IARC, 1980.
46. Aldrich JH, Nelson FD. *Linear probability, logit, and probit models*. Beverly Hills: Sage, 1984.
47. Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley & Sons, 1989.
48. Strickberger MW. *Genetics*, 2nd ed. New York: Macmillan, 1976.
49. Strickberger MW. *Genetics*, 3rd ed. New York: Macmillan, 1985.
50. Schmidt HD, Glavce C, Wischer S. The heterosis effect in newborn infants. *Anthropol Anz* 1996;54:125–134.
51. Comfort A. *The biology of senescence*. New York: Elsevier, 1984.
52. Crow JF. How much do we know about spontaneous human mutation rates? *Environ Mol Mutagenet* 1993;21:122–129.
53. Crow JF. Spontaneous mutation as a risk factor. *Exp Clin Immunogenet* 1995;12:121–128.
54. Crow JF. The high spontaneous mutation rate: is it a health risk? *Proc Natl Acad USA* 1997;94:8380–8386.
55. Crow JF. Spontaneous mutation in man. *Mutat Res* 1999;437:5–9.
56. Auroux M. Age du pere et developpement. *Contracept Fertil Sex* 1993;21:382–385.
57. Auroux M. La qualite du conceptus en fonction de l'age du pere. *J Urol (Paris)* 1993;99:29–34.
58. Auroux M. La viellissement testiculaire et ses consequences. *J Gynecol Obstet Biol Reprod (Paris)* 1983;12:11–17.
59. Risch N, Reich EW, Wishnick MM, McCarthy JG. Spontaneous mutation and parental age in humans. *Am J Hum Genet* 1987;41:218–248.
60. Lian ZH, Zack MM, Erickson JD. Paternal age and the occurrence of birth defects. *Am J Hum Genet* 1986;39:648–660.
61. McIntosh GC, Olshan AF, Baird PA. Paternal age and the risk of birth defects in offspring. *Epidemiology* 1995;6:282–288.
62. Meacham RB, Murray MJ. Reproductive function in the aging male. *Urol Clin North Am* 1994;21:549–556.
63. Savitz DA, Schwingl PJ, Keels MA. Influence of paternal age, smoking, and alcohol consumption on congenital anomalies. *Teratology* 1991;44:429–440.
64. Friedman JM. Genetic disease in the offspring of older fathers. *Obstet Gynecol* 1981;57:745–749.
65. Bordson BL, Leonardo VS. The appropriate upper limit for semen donors: a review of the genetic effects of paternal age. *Fertil Steril* 1991;56:397–401.
66. Vogel F. Mutation in man. In: Emery AEH, Rimoin D, eds. *Principles and practices of medical genetics*. London: Churchill Livingstone, 1983:397–401.
67. Carothers AD, McAllion SJ, Paterson CR. Risk of dominant mutation in older fathers: evidence from osteogenesis imperfecta. *J Med Genet* 1986;23:227–230.
68. Young ID, Thompson EM, Hall CM, Pembrey ME. Osteogenesis imperfecta type IIA: evidence for dominant inheritance. *J Med Genet* 1987;24:386–389.
69. Lurie IW. Genetics of the Costello syndrome. *Am J Med Genet* 1994;52:358–359.
70. Sheffield LJ, Danks DM, Mayne V, Hutchinson AL. Chondrodysplasia punctata—23 cases of a mild and relatively common variety. *J Pediatr* 1976;89:916–923.
71. Connor JM, Evans DA. Genetic aspects of fibrodysplasia ossificans progressiva. *J Med Genet* 1982;19:35–39.
72. Rogers JG, Chase GA. Paternal age effect in fibrodysplasia ossificans progressiva. *J Med Genet* 1979;16:147–148.
73. Martinez-Frias ML, Ramos-Arroyo MA, Salvador J. Thanatophoric displasia: an autosomal dominant condition? *Am J Med Genet* 1988;31:815–820.
74. Bertram L, Busch R, Spiegl M, Lautenschlager NT, Müller U, Kurz A. Paternal age is a risk factor for Alzheimer disease in the absence of a major gene. *Neurogenetics* 1998;1:277–280.
75. Gourbin C, Wunsch G. Paternal age and infant mortality. *Genus* 1999;55:61–72.
76. Gavrilov LA, Gavrilova NS. Parental age at conception and offspring longevity. *Rev Clin Gerontol* 1997;7:5–12.
77. Gavrilov LA, Gavrilova NS. When should fatherhood stop? [Letter]. *Science* 1997;277:17–18.
78. Gavrilov LA, Gavrilova NS, Kroutko VN, Evdokushkina GN, Semyonova VG, Gavrilova AL, Lapshin EV, Evdokushkina NN, Kushnareva YuE. Mutation load and human longevity. *Mutat Res* 1997;377:61–62.
79. Philippe P. Familial correlations of longevity: an isolate-based study. *Am J Med Genet* 1978;2:121–129.
80. Wyshak G. Fertility and longevity of twins, sibs, and parents of twins. *Soc Biol* 1978;25:315–330.
81. Bocquet-Appel JP, Jakobi L. Familial transmission of longevity. *Ann Hum Biol* 1990;17:81–95.
82. Desjardins B, Charbonneau H. L'héritabilité de la longévité. *Population* 1990;45:603–615.
83. McGue M, Vaupel JW, Holm N, Harvald B. Longevity is moderately heritable in a sample of Danish twins born 1870–1880. *J Gerontol* 1993;48:B237–B244.
84. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. The heritability of human

longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. *Hum Genet* 1996;97:319–323.
85. Sorensen TI. Genetic epidemiology utilizing the adoption method: studies of obesity and of premature death in adults. *Scand J Soc Med* 1991;19:14–19.
86. Nielsen GG, Gill RD, Andersen PK, Sorensen TIA. A counting process approach to maximum likelihood estimation in frailty models. *Scand J Stat* 1992;19: 25–43.
87. Perls TT, Alpert L, Wager CG, Vijg J, Kruglyak L. Siblings of centenarians live longer. *Lancet* 1998;351: 1560.
88. Gavrilov LA, Gavrilova NS. Biodemographic study of familial determinants of human longevity. *Population: An English Selection* 2001;13:197–222.
89. Krach CA, Velkoff VA. *Centenarians in the United States*. U.S. Bureau of the Census, Current Population Reports, series P23-199RV. Washington, DC: U.S. Government Printing Office, 1999.
90. Bell FC, Wade AH, Goss SC. *Life tables for the United States Social Security area 1900–2080*. Actuarial study no. 107. Baltimore: U.S. Department of Health and Human Services, 1992.
91. Ames BN. Micronutrients prevent cancer and delay aging. *Toxicol Lett* 1998;102–103:5–18.
92. Moore KL. *The developing human. Clinically oriented embriology*, 4th ed. Philadelphia: W.B. Saunders, 1988.
93. Raven PH, Johnson GB. *Biology*, 3rd ed. St. Louis, MO: Mosby–Year Book, Inc., 1992.
94. Gavrilov LA, Gavrilova NS. Season of birth and human longevity. *J Anti-Aging Med* 1999;2:365–366.
95. Huntington E. *Season of birth*. New York: Wiley & Sons, 1938.
96. Miura T, Shimura M. Longevity and season of birth. *Jpn J Biometeorol* 1980;17:27–31.
97. Doblhammer G. Longevity and month of birth: evidence from Austria and Denmark. *Demogr Res* 1999; 1:1–22. Online document: http://www.demographic-research.org/Volumes/Vol1/3/default.htm .
98. Fox JP, Hall CE, Elveback LR. *Epidemiology. Man and disease*. London: Macmillan, 1970.
99. Suvisaari J. *Incidence and risk factors of schizophrenia in Finland* [Dissertation]. Helsinki: University of Helsinki, 1999.
100. Rantakallio P, Jones P, Moring J, von Wendt L. Association between central nervous system infections during childhood and adult onset schizophrenia and other psychosis: a 28-year follow-up. *Int J Epidemiol* 1997;26:837–843.
101. *Genealogisches Handbuch der Adeligen Hauser* (1953). *Vol. 5. Genealogisches Handbuch des Adels [Genealogical Handbook of Nobility]*. Van Hueck W, ed. Limburg an der Lahn: C.A. Starke Verlag. Ibid. (1955) 9, 11; Ibid. (1956) 12; Ibid. (1957) 15; Ibid. (1958) 17; Ibid. (1959) 20; Ibid. (1960) 22; 24; Ibid. (1961) 26; Ibid. (1962) 29; Ibid. (1964) 32; Ibid. (1965) 34, 36; Ibid. (1966) 38; Ibid. (1968) 41; Ibid. (1969) 45; Ibid. (1970) 46; Ibid. (1971) 49; Ibid. (1972) 52; Ibid. (1974) 57; Ibid. (1975) 60; Ibid. (1977) 64; Ibid. (1979) 71; Ibid. (1980) 73; Ibid. (1981) 76; 78; Ibid. (1983) 81; Ibid. (1984) 83; Ibid. (1985) 86, 87; Ibid. (1986) 89; Ibid. (1987) 89; Ibid. (1988) 93, Ibid. (1992) 103.
102. *Genealogisches Handbuch der Freiherrlihen Hauser* (1952) *Vol. 4. Genealogisches Handbuch des Adels [Genealogical Handbook of Nobility]* (1994). Van Hueck W, ed. Limburg an der Lahn: C.A. Starke Verlag. Ibid. (1954) 7; Ibid. (1956) 13; Ibid. (1957) 16; Ibid. (1959) 21; Ibid. (1962) 27; Ibid. (1963) 30, 31; Ibid. (1966) 37; Ibid. (1967) 39; Ibid. (1969) 44; Ibid. (1970) 48; Ibid. (1971) 51; Ibid. (1975) 59; Ibid. (1976) 62; Ibid. (1977) 65; Ibid. (1978) 68; Ibid. (1979) 69; Ibid. (1982) 79, 80; Ibid. (1986) 88; Ibid. (1992) 102; Ibid. (1994) 106, 107.
103. *Genealogisches Handbuch der Furstlichen Hauser. Genealogisches Handbuch des Adels* (1951). *Vol. 1.* Van Hueck W, ed. Limburg an der Lahn: C.A. Starke Verlag. Ibid. (1953) 3; Ibid. (1955) 8; Ibid. (1956) 14; Ibid. (1959) 19; Ibid. (1961) 25; Ibid. (1964) 33; Ibid. (1968) 42; Ibid. (1971) 50; Ibid. (1978) 70; Ibid. (1980) 74, 75; Ibid. (1984) 85; Ibid. (1987) 90; Ibid. (1991) 100.
104. *Genealogisches Handbuch der Graflichen Hauser. Genealogisches Handbuch des Adels* (1952). *Vol. 2*. Van Hueck W, ed. Limburg an der Lahn: C.A. Starke Verlag. Ibid. (1953) 6; Ibid. (1955) 10; Ibid. (1958) 18; Ibid. (1960) 23; Ibid. (1962) 28; Ibid. (1965) 35; Ibid. (1967) 40; Ibid. (1970) 47; Ibid. (1973) 54, 55, 56; Ibid. (1976) 63; Ibid. (1979) 72; Ibid. (1983) 82; Ibid. (1991) 101; Ibid. (1993) 104, 105.
105. Gavrilova NS, Gavrilov LA. Data resources for biodemographic studies on familial clustering of human longevity. *Demogr Res* 1999;1:1–48. Online document: http://www.demographic-research.org/Volumes/Vol1/4/default.htm .
106. Hollingsworth TH. *Historical demography*. Ithaca: Cornell University Press, 1969.
107. Jeune B, Vaupel J, eds. *Validation of exceptional longevity. Odense monographs on population aging 6*. Odense: Odense University Press, 1999.
108. Cohen BH. Family patterns of mortality and life span: a critical review. *Q Rev Biol* 1964;39:130–191.
109. Gavrilov LA, Gavrilova NS. Is there a reproductive cost for human longevity? *J Anti-Aging Med* 1999;2 :121–123.
110. Post W, Van Poppel F, Van Imhoff E, Kruse E. Reconstructing the extended kin-network in the Netherlands with genealogical data: methods, problems, and results. *Popul Studies* 1997;51:263–278.
111. Henry L. *Techniques d'analyse en démographie historique*. Paris: INED, 1980.
112. Knodel JE. *Demographic behavior in the past: a study of fourteen German village populations in the eighteenth and nineteenth centuries*. New York: Cambridge University Press, 1988.
113. Hollingsworth TH. Genealogy and historical demography. *Ann Demogr Historique* 1976;167–170.
114. Oeppen JE. Mortality and marital status in the British peerage: 1603–1938. Presented at the S.S.H.A. Conference, Chicago, 1998.

115. Smith JE, Oeppen J. Estimating number of kin in historical England using demographic microsimulation. In: Reher DS, Schofield R, eds. *Old and new methods in historical demography*. Oxford: Clarendon Press, 1993:280–317.
116. Fogel RW. New sources and new techniques for the study of secular trends in nutritional status, health, mortality, and the process of aging. *Historical Methods* 1993;26:5–43.
117. Willigan JD, Lynch KA. *Sources and Methods for Historical Demography*. New York: Academic Press, 1982.
118. Gutmann M, Alter G. Family reconstitution as event-history analysis. In: Reher DS, Schofield R, eds. *Old and new methods in historical demography*. Oxford: Clarendon Press, 1993:159–177.
119. Hammel EA. Incomplete histories in family reconstitution: a sensitivity test of alternative strategies with historical Croatian data. In: Reher DS, Schofield R, eds. *Old and new methods in historical demography*. Oxford: Clarendon Press, 1993:125–144.
120. Bean LL. The Utah Population Database: demographic and genetic convergence and divergence. In: *Convergent issues in genetics and demography*. New York: Oxford University Press, 1990:231–244.
121. Mineau GP, Anderton DL, Bean LL. Description and evaluation of linkage of the 1880 manuscript census to family genealogies with implications for Utah fertility research. *Historical Methods* 1989;22:144–157.
122. Skolnick M, Bean LL, Dintelman SM, Mineau G. A computerized family history data base system. *Sociol Soc Res* 1979;63:506–523.
123. Family History Department. *The GEDCOM standard release 5.5*. Salt Lake City: The Church of Jesus Christ of Latter-Day Saints, 1996.

Address reprint requests to:
Dr. Leonid A. Gavrilov
Center on Aging
NORC/University of Chicago
1155 East 60th Street
Chicago, IL 60637

E-mail: lagavril@midway.uchicago.edu