Biodemography and Genetics of Aging

Anatoliy I. Yashin, Deqing Wu, Konstantin G. Arbeev, Alexander Kulminski, Liubov Arbeeva, Igor Akushevich, Eric Stallard, Svetlana Ukraintseva
Center for Population Health and Aging, Duke University, USA

Abstract

Better understanding genetic factors and mechanisms involved in regulation of human aging and life span may contribute to development of personalized medical help and improvement of population health. Numerous attempts to find genes and identify genetic mechanisms involved in regulation of human aging and longevity using genome wide association studies (GWAS) had limited success. The incompliance between information about aging and longevity accumulated in the field and data and methods used in genetic analyses might be a reason for smaller than expected progress in the field. We show that the use of demographic and longitudinal data, integrative models, and methods allows for substantial improvements of the efficiency of genetic analyses. The results of analyses of longitudinal data from the Original Framingham cohort contribute to better understanding the roles of genes in dynamics processes of aging changes and their effects on longevity. The results of genetic analyses of data from the illustrate potential of .

Introduction

The genetics of human aging and longevity became the subject of intensive analyses during last decades ranging from studies of candidate genes [1-3] [4, 5] [6] to genome-wide association studies (GWAS) [7-12] that involve hundreds-of-thousands to millions of genetic variants (SNPs, i.e., single-nucleotide polymorphisms).

The literature review on genetics of human aging and longevity indicates that the efficiency of GWAS of these traits is low: Most associations detected in these studies have not reached genome-wide level of statistical significance. They also suffer from the lack of replication in studies of independent populations. Two possible causes might contribute to this situation. One deals with the lack of comprehensive conceptual framework that would allow one to describe the roles of genes in these traits in a way appropriate for efficient statistical analyses. Another cause is related to the lack of systemic statistical approaches that would allow for integrating available information and data on aging and life span. In this paper we will discuss the possibility of improving the efficiency of statistical models and methods used in genetic analyses of aging and life span. Specifically we show that application of bio-demographic ideas, models and methods to longitudinal data has a potential to substantially improve the efficiency of genetic analyses, and address fundamental questions on dynamic aspects of aging related changes and their connection to lifespan that have never been addressed before.

Age patterns of genetic frequencies: What do they say about genetics of life span?.

Common age-associated health disorders, such as cancer, CHD, stroke, diabetes and asthma, are major contributors to old age morbidity and mortality. It seems natural to expect that genetic variants which increase risks of such diseases (also called “risk alleles” or “risk genotypes”) would negatively affect survival. Similarly, beneficial genetic variants will play protective roles against diseases and contribute to increases in lifespan. The genetic frequencies corresponding to such variants will have monotonically declining (in case of harmful alleles) or

monotonically increasing age trajectories (in case of beneficial alleles), see left panel in **Fig. 1**. Note that we deal with age trajectories of minor allele frequencies here. The results of many genetic studies of human longevity confirm these expectations, see [13-18], among others.

It turns out, however, that associations of genetic factors with diseases and lifespan are not limited by the relationships described above. More complicated connections which have surprising manifestations have been also observed. In [19, 20] the empirical frequencies of some candidate alleles/genotypes decreased until middle old ages, reached their minimum values, and then increased in the oldest old resulting in the appearance of such initially "harmful alleles" with relatively high frequency among the long-living individuals (the second panel from the left in **Fig. 1**). It looks as if initially harmful effects of such alleles on lifespan manifested earlier in life became neutralized and then transformed into beneficial effects later in life. The increasing and then declining patterns of the age trajectories of genetic frequencies were described in [21]. The "paradoxical" presence of "risk alleles" for common diseases in genomes of long-living individuals were discussed in [22-26] among others.

Several other patterns of age trajectories of genetic frequencies were also observed in our preliminary studies. Two of them deal with the frequencies that stay about the same among the adults and among the persons of early old ages, and experienced major changes only at the oldest old ages somewhat after 80-85 years of age. The one part of these trajectories increases and another part declines at the oldest old ages (the third panel from the left in **Fig. 1**). Relative stability of such frequencies among adults and the old people may indicate that these genetic variants either do not belong to the sets of risk alleles affecting major human diseases specific for these age categories, or they have opposite influences on risks of different diseases which compensate for the effects of these factors on total mortality. An increasing pattern may indicate that corresponding genes provide protective effects against lethal events at late life. A declining pattern indicates that these genes provide their carriers with increased vulnerability to mortality risk at the oldest old ages. Genes of both types are likely to be involved in regulation of aging related diseases.

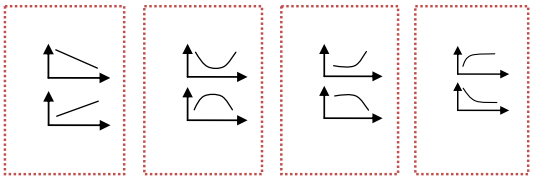


Fig. 1. Eight age patterns of genetic frequencies corresponding to different effects of genetic

The remaing two patterns are characterized by changes in genetic frequencies from the adult to old ages and relatively stable behavior after that at the oldest old ages (right panel in **Fig. 1**). The corresponding genes may belong to the set of protective (deleterious) alleles for a number of common diseases of the adult and earlier old ages. They do not influence risks of mortality and diseases at the oldest old ages. These genes are likely to have opposite effects on major human diseases.

Demographic ideas that advanced the genetic studies of aging and longevity

Researchers studying the genetics of human aging and longevity tend to underestimate the role of demographic information and bio-demographic ideas in genetic analyses. The benefits

of combining demographic and genetic information in their joint analyses were demonstrated in [19] [20] [21] [27]. These analyses showed that the use of mortality data and models of hidden heterogeneity in susceptibility to death help improve the accuracy of genetic estimates in genetic centenarian studies. The use of demographic information and models in analyses of data on genetically heterogeneous cohorts allowed researchers to compare the age patterns of mortality rates for carriers and non-carriers of candidate alleles and genotypes. Such comparisons could not be possible using data on genetic frequencies alone.

The use of such methods to genome-wide association studies (GWAS) was stimulated by the fact that useful demographic information tends to be ignored in GWAS of human aging and longevity performed during several recent years. Such an extension takes advantage of the fact that the genetic information on aging and lifespan available for researchers is contained not only in the follow up data (which researchers traditionally use when they have access to such data). In most GWAS of human aging and longevity, performed so far, the bio-specimen collection for genotyping has been performed in population of individuals of different ages (e.g., as in the Original FHS cohort). It turn out that the age distribution of genotyped individuals at the time of bio-specimen collection contains important information about genetics of longevity. This information was typically ignored in GWAS, and conclusions on genetic influence on lifespan were derived from analyses of only follow-up data. Following the ideas described in Arbeev et al., (2011) we used genetic information from both the age distribution of genotyped individuals and from the follow up data to get more accurate estimates of genetic influence on lifespan.

The analyses are based on the maximum likelihood approach which maximizes the joint likelihood function of the combined data comprised of the age structure of study participants at the time of bio-specimen collection and the follow-up data. The total likelihood of combined dataset is the product of the two likelihood functions representing each subset of available data. The benefits of such analyses stem from the fact that both likelihoods are functions of the same parameters describing the relationship between genetic factors and the phenotype of interest. These likelihoods are functions of the mortality rates for carriers, $\mu(x|G = 1)$, and non-carriers, $\mu(x|G = 0)$, and the initial proportion of the genetic variant, $p_0 = P(G = 1)$ (initial allele frequency). These mortality rates can be described parametrically (e.g., by the Gompertz, Gompertz-Makeham, or logistic curves). By maximizing the total likelihood function, one can estimate the respective quantities and test the null hypothesis on coincidence of survival functions for carriers and non-carriers of the minor allele using the likelihood ratio test.

The likelihood function for joint analyses of data

Let x_k^0, $k = 1\ldots K$, be the ages at baseline (entry to the study) of individuals from the genetic subsample of the data and let x_{m,x_k^0}, $m = 1\ldots M_k$, be their ages at the time of biospecimen collection. Denote by $N(x_{m,x_k^0}) = N_1(x_{m,x_k^0}) + N_0(x_{m,x_k^0})$ the number of individuals in the genetic subsample who were aged x_{m,x_k^0} at the time of biospecimen collection and aged x_k^0 at baseline. Here $N_g(x_{m,x_k^0})$ are the numbers of non-carriers ($g = 0$) and carriers ($g = 1$) of the respective allele/genotype. Let τ_i be the life span (it may be censored) of the i^{th} individual. Denote by $\mu(x|G = g)$ the hazard rate for carriers/non-carriers and by

$\pi(x_{m,x_k^0} \mid x_k^0) = P(G=1 \mid \tau > x_{m,x_k^0}, x_k^0)$ the proportion of carriers at age x_{m,x_k^0} given that the individuals were aged x_k^0 at baseline. Denote by $S_g(x) = P(\tau > x \mid G = g)$ the survival functions for carriers/non-carriers and by $p = P(G=1)$ the initial proportion (at birth) of carriers of allele/genotype in a population, which is assumed here to be the same for different birth cohorts represented in the study. The total (population) survival function is $S(x) = pS_1(x) + (1-p)S_0(x)$. Conditional survival functions for the individuals aged x_k^0 at the baseline are $S_g(x \mid x_k^0) = P(\tau > x \mid G = g, x_k^0)$. The mortality rates for carriers/non-carriers can be of any parametric form. The proportions $\pi(x_{m,x_k^0} \mid x_k^0)$ are:

$$\pi(x_{m,x_k^0} \mid x_k^0) = \frac{P(G=1 \mid x_k^0) S_1(x_{m,x_k^0} \mid x_k^0)}{P(G=1 \mid x_k^0) S_1(x_{m,x_k^0} \mid x_k^0) + (1 - P(G=1 \mid x_k^0)) S_0(x_{m,x_k^0} \mid x_k^0)},$$

where $P(G=1 \mid x_k^0) = pS_1(x_k^0)/S(x_k^0)$. The likelihood function of the data on the ages at biospecimen collection (L_{abc}) and the likelihood function of the follow-up survival data (L_{sur}) are:

$$L_{abc} \sim \prod_{k=1}^{K} \prod_{m=1}^{M_k} \pi(x_{m,x_k^0} \mid x_k^0)^{N_1(x_{m,x_k^0})} (1 - \pi(x_{m,x_k^0} \mid x_k^0))^{N_0(x_{m,x_k^0})}$$

and

$$L_{sur} \sim \prod_{k=1}^{K} \prod_{m=1}^{M_k} \prod_{g=0}^{1} \prod_{i=1}^{N_g(x_{m,x_k^0})} \mu(\tau_i \mid G=g)^{\delta_i} S_g(\tau_i \mid x_{m,x_k^0}),$$

where δ_i is a censoring indicator. The total likelihood function of the data relevant for genetic analyses of the genetic subsample is the product of the two likelihood functions:

$$L \sim L_{abc} L_{sur}.$$

By maximizing this likelihood function one can estimate parameters of the mortality rates for carriers/non-carriers of the minor allele and the initial proportions for each genetic variant. Then the null hypothesis on coincidence of survival functions for carriers and non-carriers of the minor allele can be tested using the likelihood ratio test. The type of the effect of such genetic variants on survival (e.g., the protective effect so that the survival curve for carriers of the minor allele is shifted to the right compared to non-carriers, or the deleterious one so that the curve for carriers is shifted to the left, or the trade-off so that survival curves intersect) can be understood inspecting respective estimates of parameters and/or visualizing the estimated survival curves. The benefits of using this approach are investigated in simulation study [28]. The approach has been applied to the data collected in the Framingham Heart Study (FHS). Selected genetic variants showed highly significant negative associations with life span.

Controlling for population stratification has to account for mortality selection. If some genetic variants influence mortality risk and others are neutral (i.e., do not affect this risk) then corresponding population is genetically heterogeneous with respect to individual susceptibility to

death [29, 30]. The genetic structure of such population experiences aging related changes due to the process of mortality selection. Thus, population stratification may be caused not only by the differences in ancestry among study participants but also by mortality selection in genetically heterogeneous cohorts: individuals carrying harmful alleles or genotypes die first, and such selection changes (multidimensional) genetic distribution in the older population compared to that of the younger one. This effect is especially important if the population under study consists of left truncated birth cohorts of genotyped individuals. The left truncation in such cohorts can be induced by individual differences in ages at the time of bio-specimen collection (e.g., age differences at baseline, if the blood collection was done at baseline). Controlling for possible population stratification (e.g., using the Principal Component Analyses (PCA)[31]) may be efficient for phenotypic traits not affected by mortality selection (e.g., the eyes color). However, for longevity traits controlling for population stratification may eliminate all the effects of population structure including genetic effects on lifespan which we are interested in. Thus, the GWAS of human longevity with inappropriate use of methods of controlling for population stratification are likely to suffer from low level of statistical significance. It turns out that the effects of genetic clustering due to differences in ancestry can be separated from those induced by mortality selection of genetically vulnerable individuals.

After such separation the GWAS in which the effects of clustering due to differences in ancestry are controlled by PCs corresponding to this clustering can be performed. The separation of the two components that induce population clustering will involve modification of the covariance matrix suggested for calculation of Principal Components (PCs) in Price e al. (2006) by removing the effects of population stratification due to mortality selection until ages at biospecimen collection. The new matrix is calculated by conditioning the genetic SNP vector on the values of the variable "age at biospecimen collection." Conditioning on the age at biospecimen collection will eliminate effects of mortality selection in PCs making them dependent only on the effects due to differences in ancestry. The first several PCs (following Price et al. (2006)) are used in GWAS of human longevity.

We used methods and approaches described above in the analyses of data on aging and lifespan from the Original Framingham cOhort. Figs. 2, 3, 4 and 5 illustrate the results of GWAS of life span for males from the original FHS cohort using two different methods of controlling for possible population stratification. In one the 20 principal components were constructed using the population of individuals for whom the bio-specimen were collected at different ages (Figs. 2,3). In the second method, 20 principal components were constructed using the population of individuals whose age at the time of bio-specimen collection did not exceed 60 years (Figs. 4,5). One can see from these figures that the results of analyses shown in Figs. 4 and 5 are more statistically significant then those shown in Figs. 2 and 3.

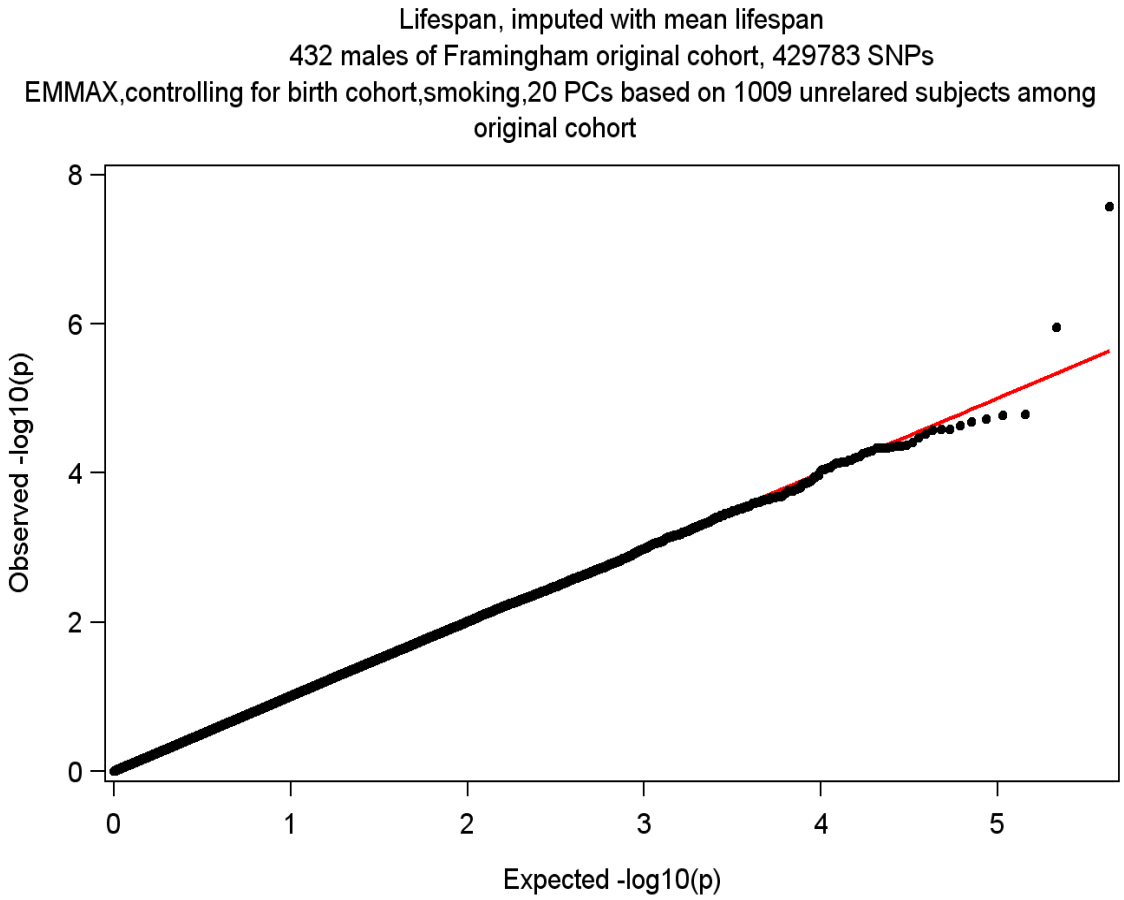


Fig. 2. The graph of the QQ plot of the results of GWAS of human lifespan for 432 males from the Original FHS cohort. Analyses have been performed using EMMAX computer program controlling for birth cohort, smoking, and 20 principal components. Principal components were used to correct for possible population stratification. They were constructed using genetic data on 1009 unrelated individuals from the Original FHS cohort.

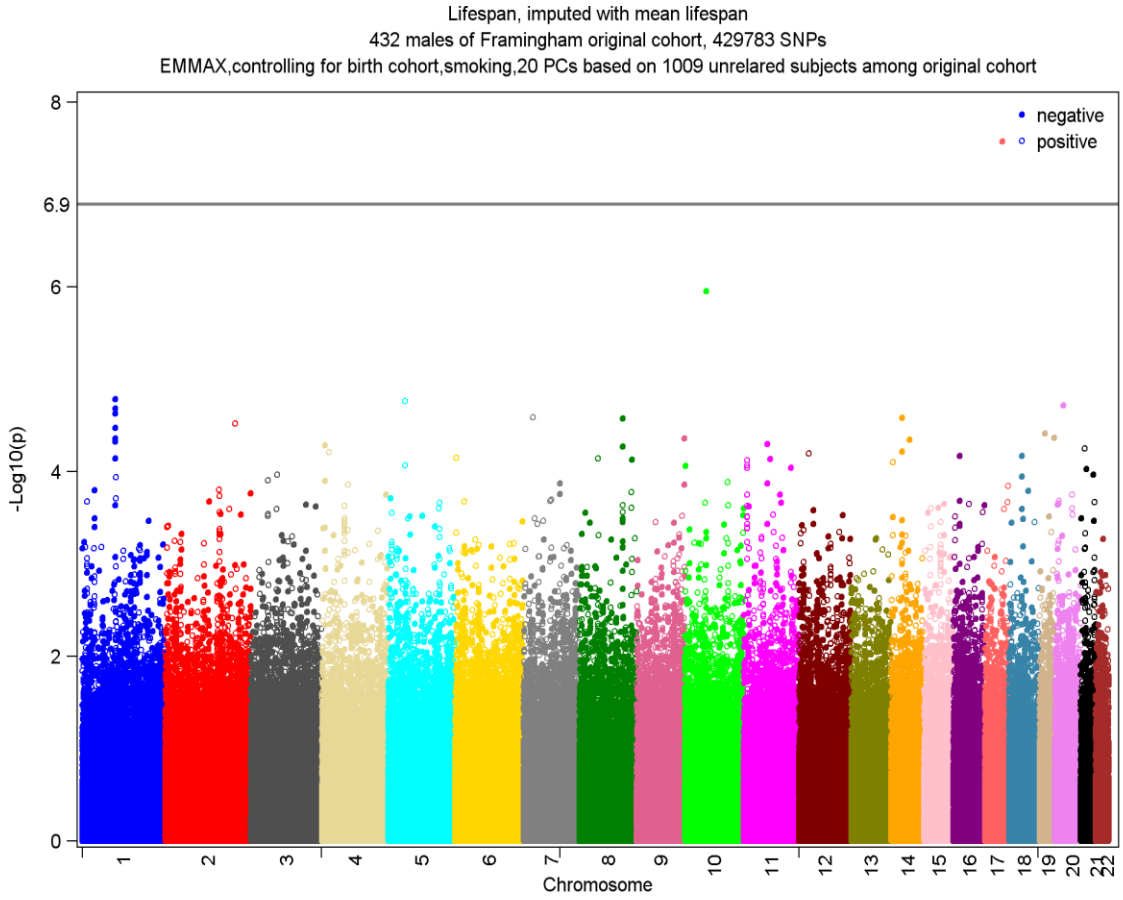


Fig. 3. The graph of the Manhattan plot of the results of GWAS of human lifespan for 432 males from the Original FHS cohort. Analyses have been performed using EMMAX computer program controlling for birth cohort, smoking, and 20 principal components. Principal components were used to correct for possible population stratification. They were constructed using genetic data on 1009 unrelated individuals from the Original FHS cohort.

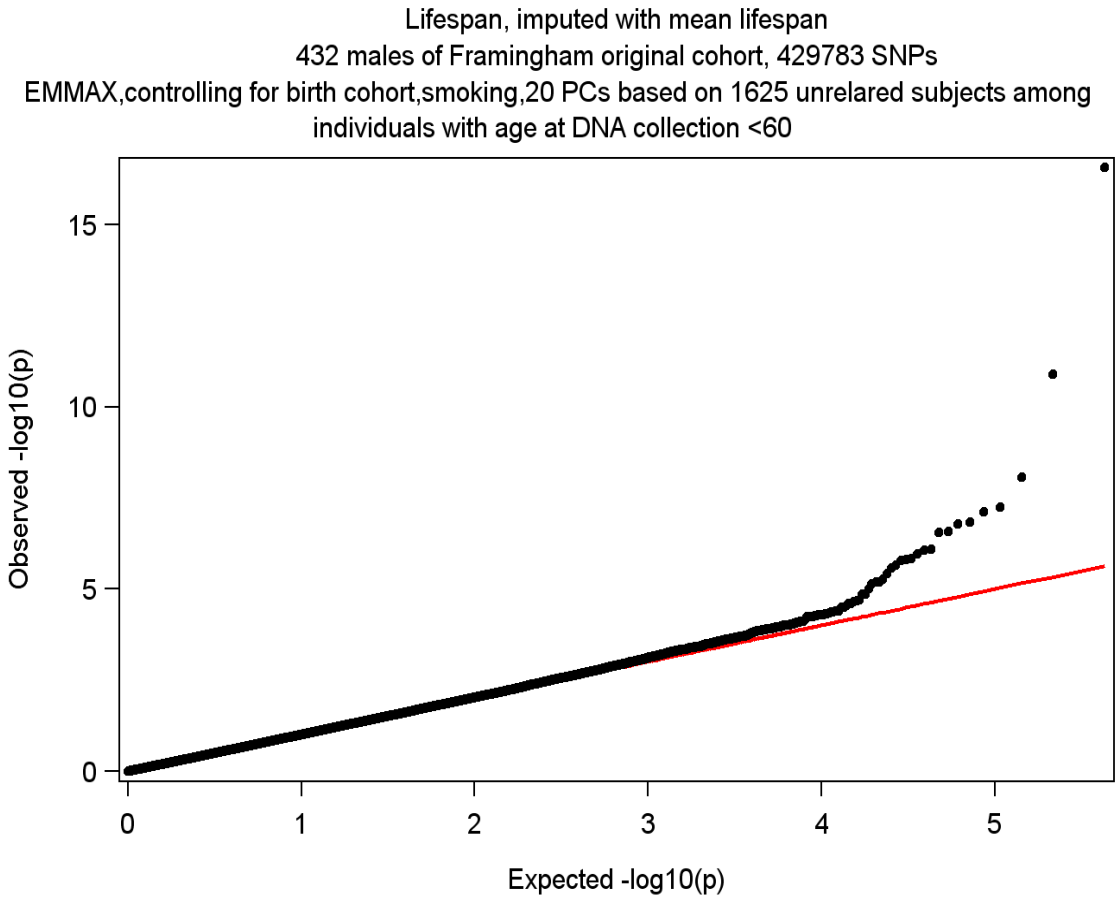


Fig. 4. The graph of the QQ plot of the results of GWAS of human lifespan for 432 males from the Original FHS cohort. Analyses have been performed using EMMAX computer program controlling for birth cohort, smoking, and 20 principal components. Principal components were used to correct for possible population stratification. They were constructed using genetic data on 1625 unrelated individuals from the Original FHS cohort whose ages at the time of bio-specimen collection did not exceed 60 years.

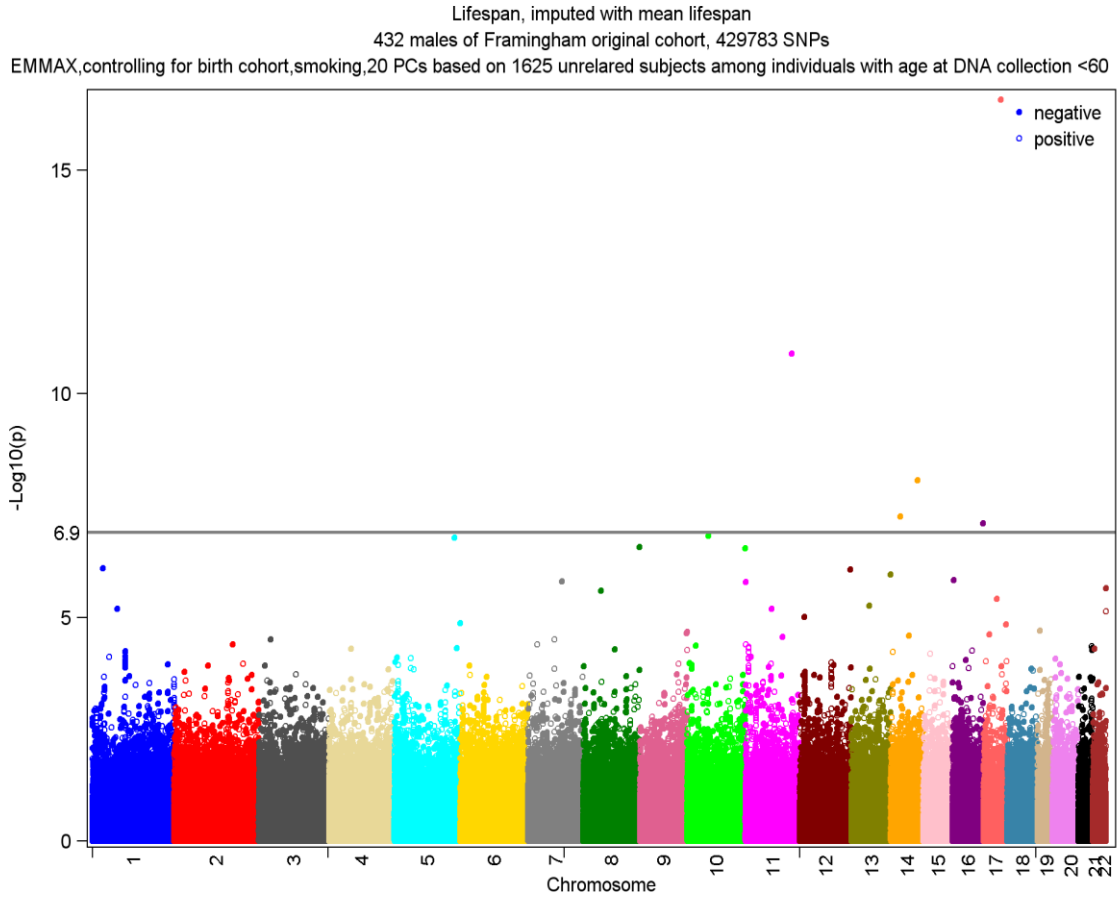


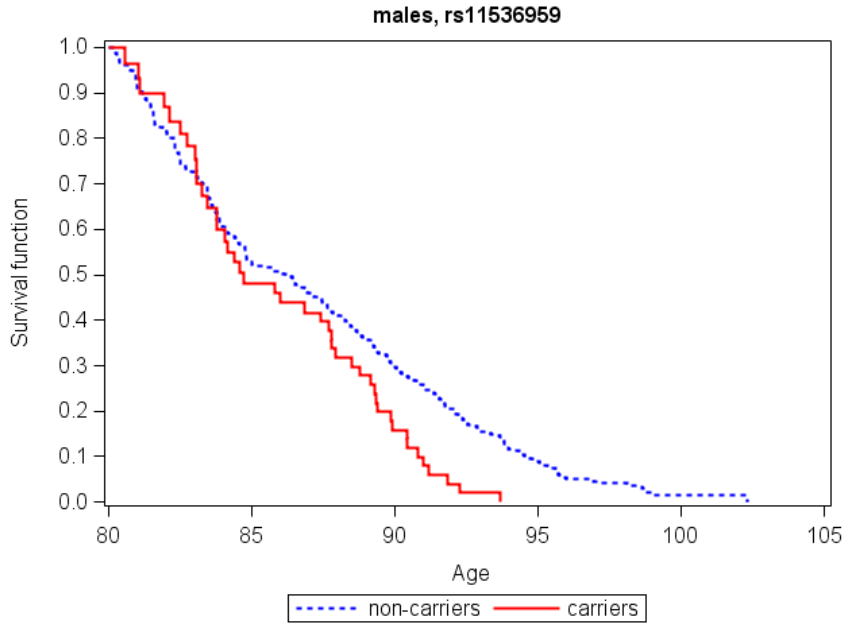
Fig. 5. The graph of the Manhattan plot of the results of GWAS of human lifespan for 432 males from the Original FHS cohort. Analyses have been performed using EMMAX computer program controlling for birth cohort, smoking, and 20 principal components. Principal components were used to correct for possible population stratification. They were constructed using genetic data on 1625 unrelated individuals from the Original FHS cohort whose ages at the time of bio-specimen collection did not exceed 60 years.

Effects of selected variants on survival and age trajectories of physiological indices

To show that selected genetic variants are associated with human survival we constructed survival functions for carriers and non-carriers of minor alleles of selected SNPs. Using available longitudinal data on repeated measurements of physiological variables we also evaluated average age trajectories of physiological indices for groups of individuals with different genetic background. The results for the four SNPs: rs115536959, rs10845099, rs5743998, and rs9971555 are shown in Figs. 2, 3, 4, and 5.

One can see from these figures that selected SNPs have different influence on survival and on average age trajectories of physiological indices. The carriers and non-carriers of minor allele of the rs115536959 SNP have different average age trajectories of systolic and diastolic

blood pressure. The corresponding difference for carriers and non-carriers of minor allele of the rs10845099 SNP is much smaller. One can also see the difference in patterns of influence of corresponding SNPs on body mass index. There is also difference in the effects of rs5743998 and rs9971555 SNPs on serum cholesterol. The molecular biological mechanisms responsible for these differences require further analyses.



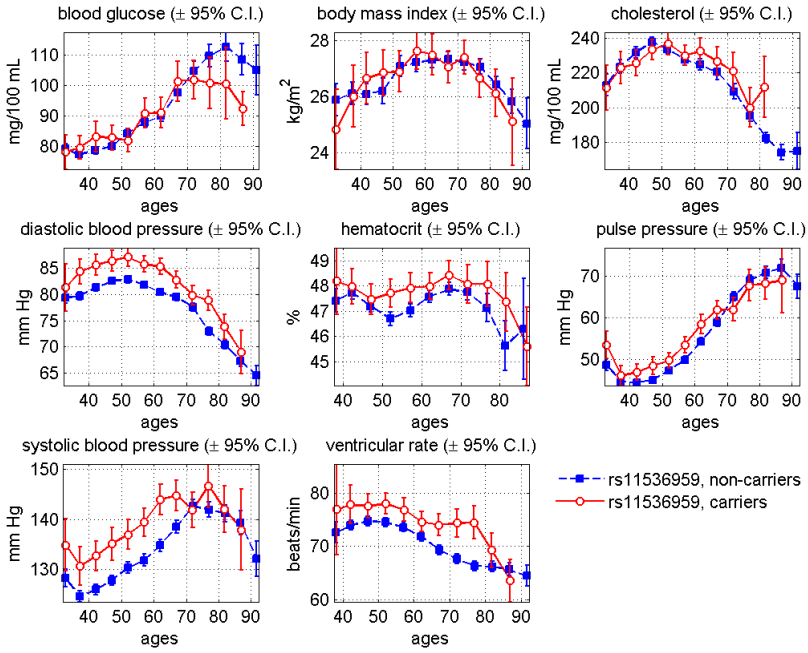
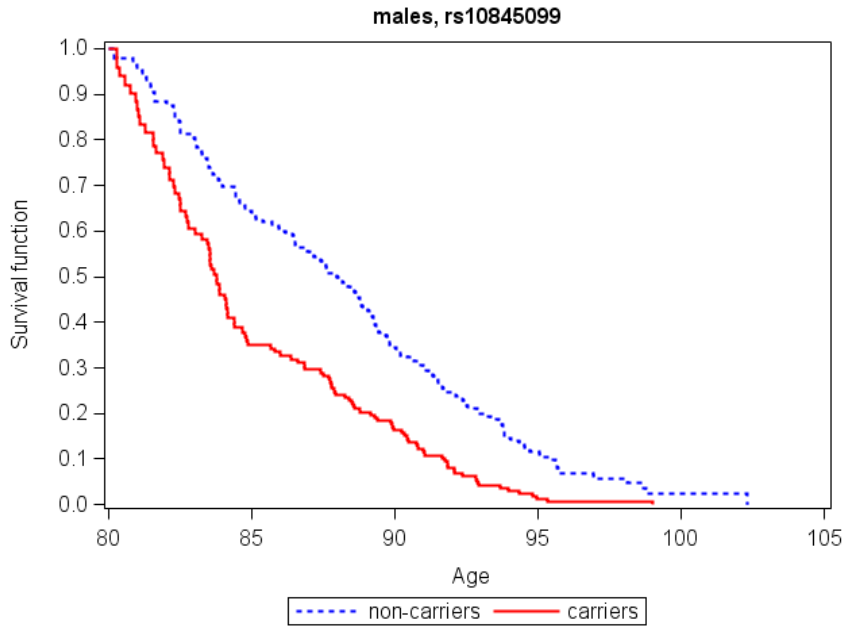


Fig. 6. Age patterns of survival functions (top panel) and average age trajectories of pysiological indices (bottom panel) for carriers and noncarriers of minor allele of the rs11536959 SNP.



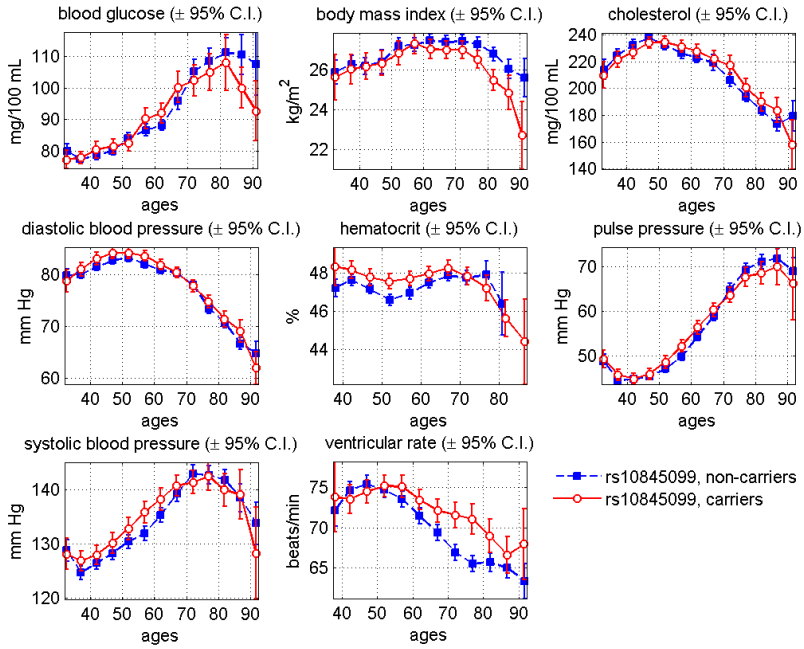
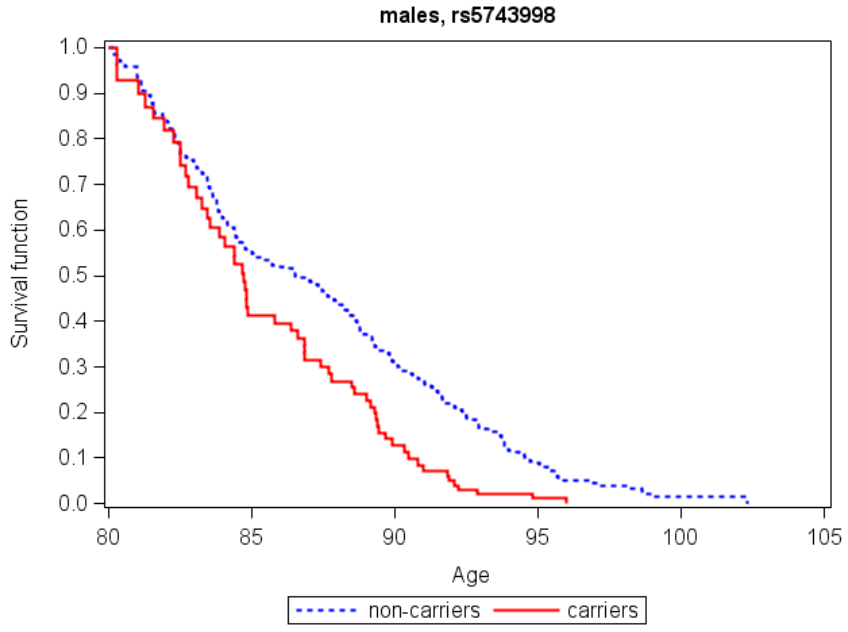


Fig. 7. Age patterns of survival functions (top panel) and average age trajectories of pysiological indices (bottom panel) for carriers and noncarriers of minor allele of the rs10845099 SNP.



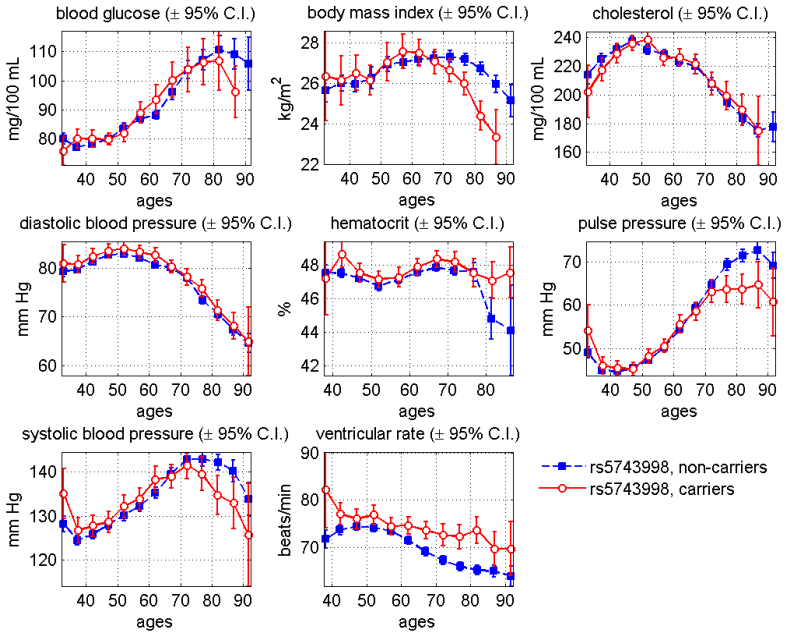


Fig. 8. Age patterns of survival functions (top panel) and average age trajectories of pysiological indices (bottom panel) for carriers and noncarriers of minor allele of the rs5743998 SNP.

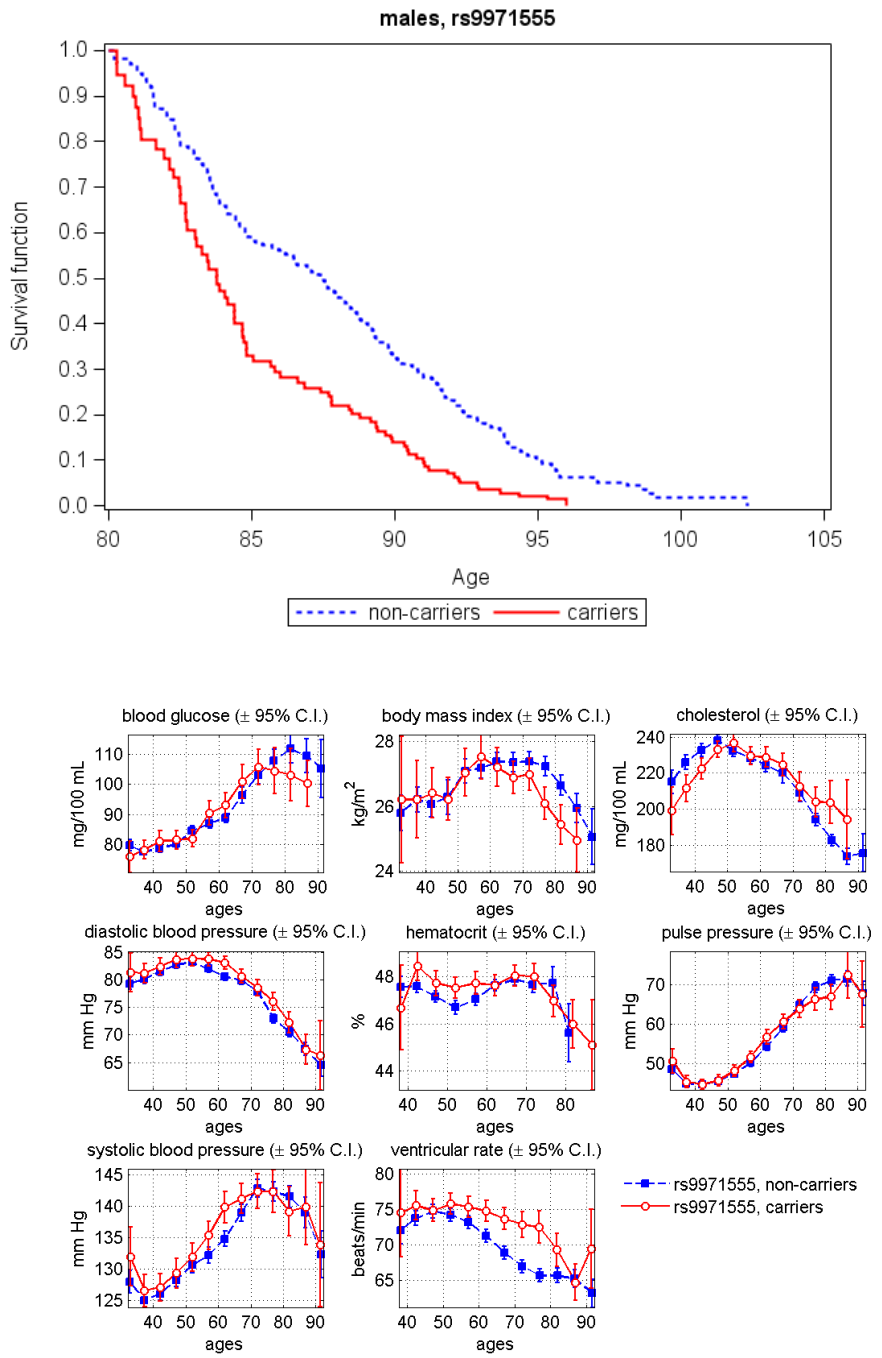


Fig. 9. Age patterns of survival functions (top panel) and average age trajectories of pysiological indices (bottom panel) for carriers and noncarriers of minor allele of the rs9971555 SNP.

Genetics of hidden biomarkers of aging: The need for dynamic modeling of physiological changes. More sophisticated analyses of longitudinal data can be performed using joint modeling of longitudinal dynamics of covariates and risks of time-to-event outcomes (e.g., [32-

51]). Although these methods address many important questions about behavior of repeatedly measured variables in the presence of informative censoring, they do not allow for evaluating a number of essential components of aging-related processes which are not directly measured in most longitudinal studies but play fundamental roles in longitudinal dynamics of physiological variables and other biomarkers. These *essential components* include *age-specific physiological norms* [52-54], *allostasis* and *allostatic load* [55, 56], *stress resistance* [57-60], *adaptive capacity* (*homeostenosis*) [61, 62], and *short-term stochasticity* [63]. The information about these components is a part of knowledge about aging, health, and longevity accumulated in the research field which is typically ignored in statistical analyses of longitudinal data.

To evaluate genetic influence on these hidden components from the data using methods of statistical modeling they have to be included into the model of longitudinal data and properly linked with measured variables as well as with health and survival outcomes. Such integrative analysis can be performed within a special methodological framework presented in the genetic version of the stochastic process model of human mortality and aging (GenSPM)[64]. As a result, the variables measured in a longitudinal study become dynamically linked together, with genetic and other E-factors, and with the essential aging processes. Altogether these variables provide a comprehensive description of the mechanisms of aging-related changes in human organisms. An important point is that the age trajectories of all aging-related variables can be efficiently estimated from the longitudinal data using SPM [65-67].

In contrast to the standard approaches for analyzing the effects of observed covariates on ages at onset of health or survival outcomes using the proportional hazard Cox-type regression models, SPM explicitly recognizes that such risks often are U- or J-shaped. The evidence for U- or J- shaped risks as functions of different physiological markers is abundant in epidemiological studies [68-76]. Biological studies of longevity also confirm the U-shape of dose-response curves (longevity hormesis effect [77, 78]). Therefore, the use of quadratic (U- or J- shaped) hazards in analysis is biologically meaningful.

In short, GenSPM uses stochastic differential equations to describe age-dynamics of individual changes in physiological markers until death or end of follow-up. The coefficients of these equations are specified in terms of variables characterizing allostatic load, homeostenosis, stress resistance, and stochasticity for carriers of and non-carriers of selected genotypes. Our prior studies showed that this model permits identification of all these characteristics for carriers of selected genotypes [79, 80]. This model can incorporate "static" covariates to evaluate joint effects of genetic and non-genetic (environmental) factors on quality of death [64, 81].

To understand the research results we briefly describe the version of the GenSPM used in this analyses. The evolution of physiological variables Y_t over age t is described by stochastic differential equation

$$dY_t = a(t,G)(Y_t - f_1(t,G))dt + B(t,G)dW_t,$$

with the normally distributed initial condition Y_{t_0}. Here G (G = 0, 1; $P(G=1) = p_1$) is a discrete random variable characterizing differences in genetic backgrounds among the groups of individuals, W_t is a Wiener process independent of Y_{t_0} and G. The coefficient $B(t,G)$ was considered constant in these applications.

The effect of allostatic adaptation $f_1(t,G)$ [64, 82] is described as quadratic function of t. This choice comes from the empirical observations of the average trajectories of the physiological variables in the FHS, which generally have a quadratic form, although, of course,

these average trajectories do not necessary have to follow $f_1(t,G)$.

The negative feedback coefficient $a(t,G)$ characterizes by strength of homeostatic forces. The decline in the absolute value of this coefficient with age represents the decline in the adaptive (homeostatic) capacity with age ("homeostenosis") which has been shown to be an important characteristic of aging [57, 61, 62, 83]. We used a linear approximation of this coefficient as function of age. The U- or J- shapes of the mortality and morbidity risks as functions of various physiological variables and other risk factors were confirmed in a number of studies. This indicates that a quadratic function can capture dependence of the risk on deviations of trajectories of a physiological variable Y_t from its "optimal" values [64, 84-88]. Such function has been used to describe mortality rate conditional on Y_t and G:

$$\mu(t,Y_t,G) = \mu_0(t,G) + (Y_t - f_0(t,G))^2 \mu_1(t,G).$$

Here $\mu_0(t,G)$ is the baseline hazard, $f_0(t,G)$ are "optimal" trajectories ("physiological norms"). We used the gamma-Gompertz (logistic) baseline hazards $\mu_0(t,G)$ [89].

The coefficient $\mu_1(t,G)$ characterizes stress resistance. Its increase with age corresponds to the decline in stress resistance because it narrows U-shape of the risk, i.e., making an organism more vulnerable to deviations from the "optimal" values. which can be considered as a manifestation of the senescence process [90, 91]. In our analyses $\mu_1(t,G)$ was approximated by a linear function of age.

The average age trajectories of respective physiological variables in long-lived (life span ≥ 90 for females; life span ≥ 85 for males) female and male carriers and non-carriers of the APOE e4 allele were considered as "optimal" trajectories $f_0(t,G)$ in the model. The likelihood optimization and the statistical tests have been performed using corresponding Toolboxes in MATLAB. It is important to note that maximization of the likelihood function of the genetic SPM is computationally extensive. Generally, it involves solution of the systems of ordinary differential equations (ODE) for each measurement and at each step of the likelihood optimization procedure. However, as our experience with these models shows, the calculations are feasible on modern computers and using modern statistical and technical software, e.g., MATLAB's Optimization Toolbox and ODE solvers, or SAS/OR PROC OPTMODEL, implementing different optimization algorithms (such as the Newton-Raphson or trust-region methods) and the Runge-Kutta method for the ODE solution. Therefore, the use of such models in analyses of *candidate SNPs* is feasible, but the computational burden prohibits their applications to analyses of *all SNPs* in GWAS data. Taking this limitation and advantages of the model into consideration, the GenSPM was used for studying *initially pre-selected* SNPs.

Figs. 10, 11, and 12 show age patterns of a number of hidden biomarkers of aging for groups of individuals carrying different numbers of longevity alleles.

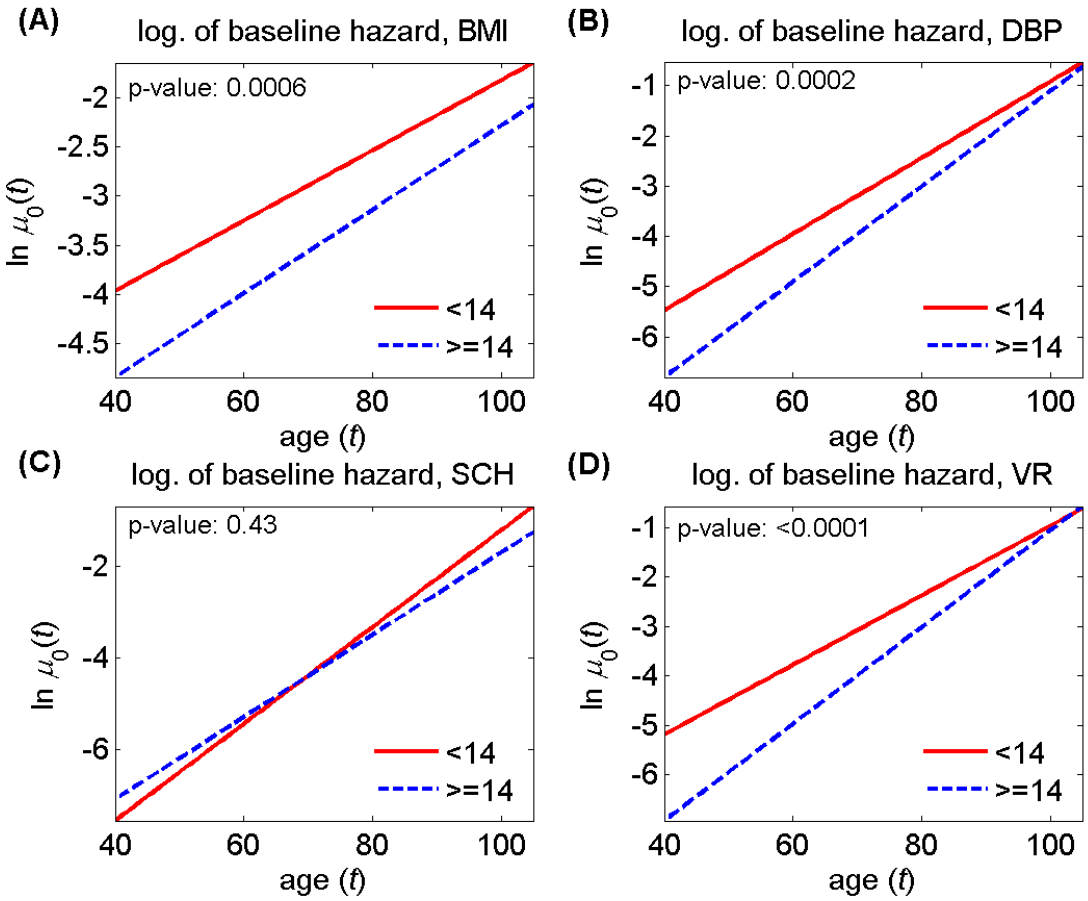


Fig. 10: Estimates of the logarithm of the baseline hazard rates in the stochastic process model (Yashin et al., 2007a) applied to data on longitudinal measurements of four physiological indices and total mortality in individuals carrying different number of longevity alleles (<14 and >=14) out of the 27 such alleles identified in Yashin et al. (2012c): **(A)** estimates for body mass index (BMI); **(B)** estimates for diastolic blood pressure (DBP); **(C)** estimates for cholesterol (SCH); **(D)** estimates for ventricular rate (VR). P-values are for the null hypotheses on the equality of baseline hazards in the two groups.

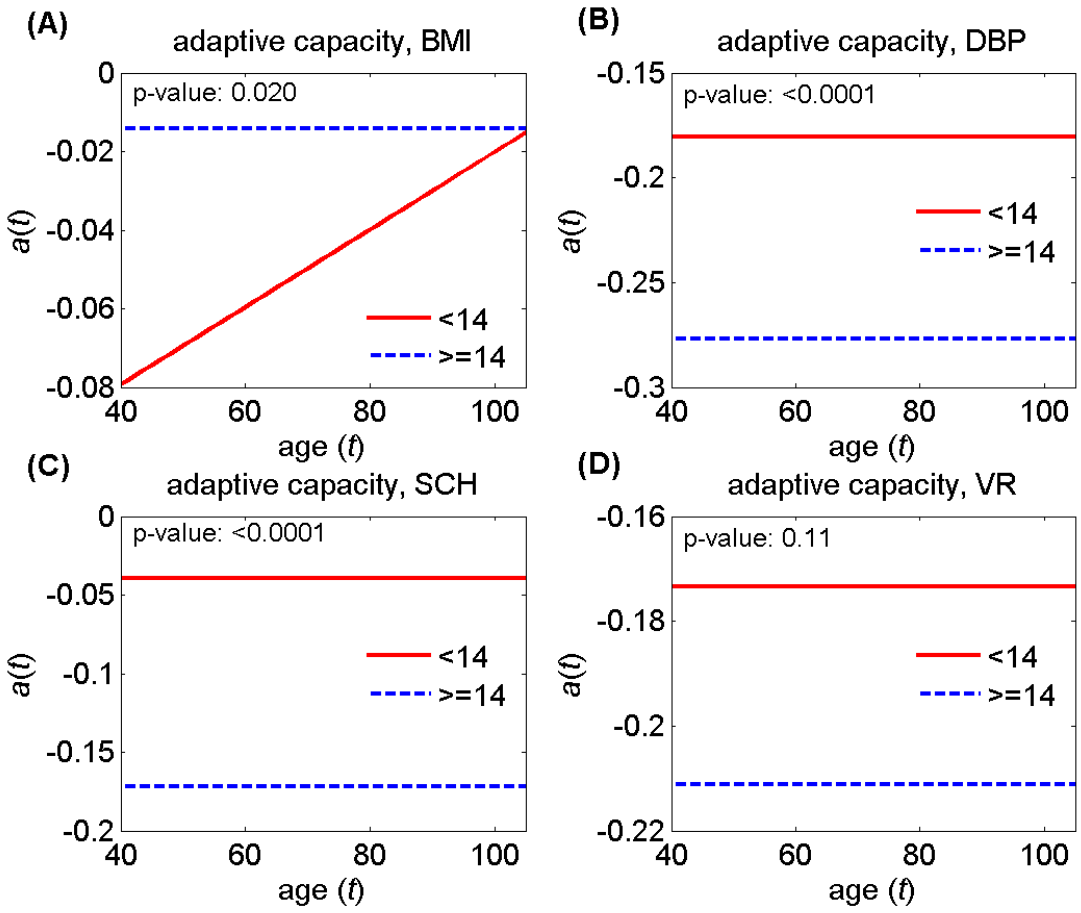


Fig. 11. Estimates of adaptive capacity in the stochastic process model (Yashin et al., 2007a) applied to data on longitudinal measurements of four physiological indices and total mortality in individuals carrying different number of longevity alleles (<14 and >=14) out of the 27 such alleles identified in Yashin et al. (2012c): **(A)** estimates for body mass index (BMI); **(B)** estimates for diastolic blood pressure (DBP); **(C)** estimates for cholesterol (SCH); **(D)** estimates for ventricular rate (VR). P-values are for the null hypotheses on the equality of adaptive capacities in the two groups.

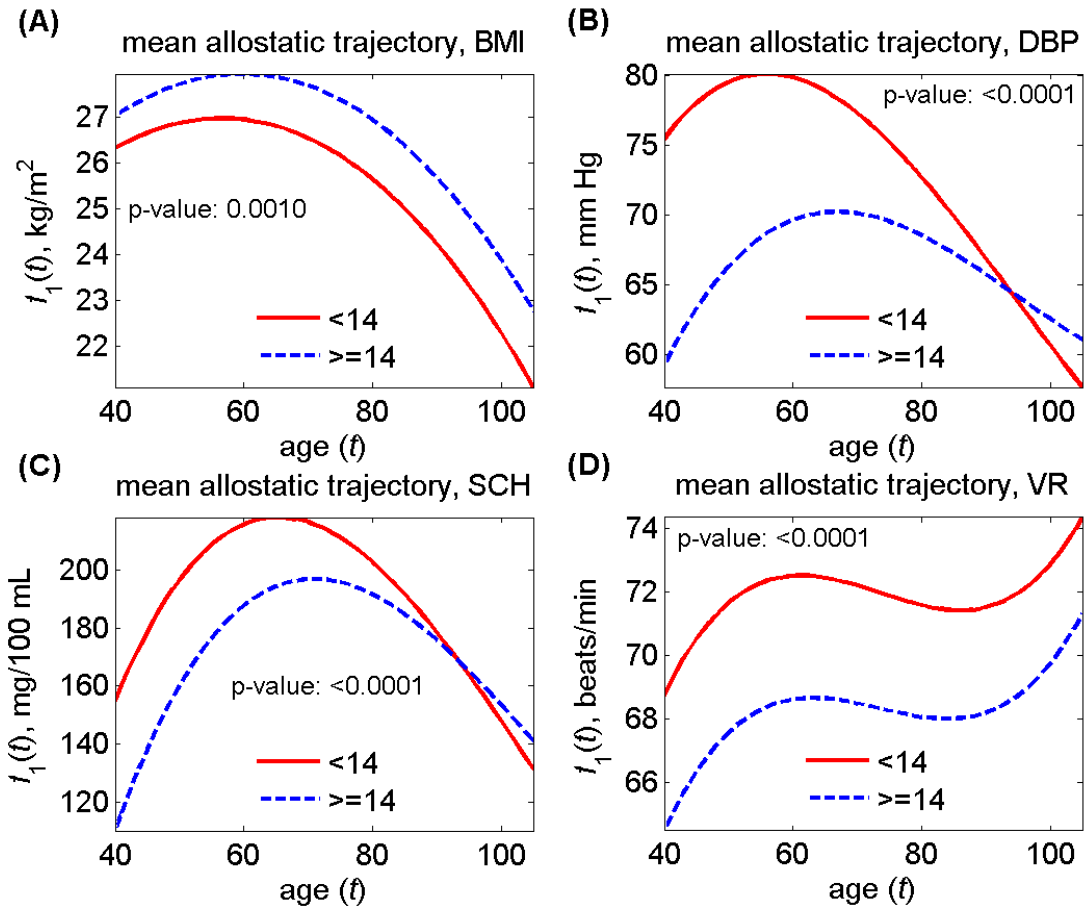


Fig. 12: Estimates of mean allostatic trajectories in the stochastic process model (Yashin et al., 2007a) applied to data on longitudinal measurements of four physiological indices and total mortality in individuals carrying different number of longevity alleles (<14 and >=14) out of the 27 such alleles identified in Yashin et al. (2012c): **(A)** estimates for body mass index (BMI); **(B)** estimates for diastolic blood pressure (DBP); **(C)** estimates for cholesterol (SCH); **(D)** estimates for ventricular rate (VR). P-values are for the null hypotheses on the equality of mean allostatic trajectories in the two groups.

Conclusion

The results of these analyses indicate that the use of demographic and longitudinal data in genetic analyses of aging and longevity may substantially improve our understanding of the roles of genetic factors in regulation of aging and life span. These results also show that control for population stratification in genetic studies of longevity related traits has to be used with care. The low efficiency of genetic studies of these traits might be related to the fact that the estimates of the effects of genetic factors on such traits can be reduced or eliminated when traditional methods of controlling for population stratification are used. The results show that involvement of longitudinal data in genetic analyses of aging and life span enriches our understanding of how genetic factors affect survival and how physiological indices mediate these effects. The use of

GenSPM tool allowed for evaluating genetic effects on hidden biomarkers of aging. Altogether the results of these analyses indicate that statistical modeling could be an efficient tool for integrated analyses of genetic and phenotypic data on aging and longevity. More efforts are needed to integrate biological information into genetic analyses of longitudinal data.

Acknowledgements. The FHS project is conducted and supported by the NHLBI in collaboration with Boston University (N01 HC25195). The FHS data used for the analyses were obtained through dbGaP. The authors acknowledge the investigators that contributed the phenotype and genotype data for this study. This manuscript was not prepared in collaboration with investigators of the FHS and does not necessarily reflect the opinions or views of the FHS, Boston University, or the NHLBI. This work was partly supported by NIH/NIA grants R01AG030612, R01AG030198, R01AG032319, and 1R01AG46860-01.

References

1. Franceschi, C., et al., *The network and the remodeling theories of aging: historical background and new perspectives.* Exp Gerontol, 2000. **35**(6-7): p. 879-96.
2. Tan, Q., et al., *A logistic regression model for measuring gene-longevity associations.* Clinical Genetics, 2001. **60**(6): p. 463-469.
3. De Benedictis, G., et al., *Recent advances in human gene-longevity association studies.* Mech Ageing Dev, 2001. **122**(9): p. 909-20.
4. Christensen, K., *Human biodemography: Some challenges and possibilities for aging research.* Demographic Research, 2008. **19**: p. 1575-1586.
5. Christensen, K., T.E. Johnson, and J.W. Vaupel, *The quest for genetic determinants of human longevity: challenges and insights.* Nat Rev Genet, 2006. **7**(6): p. 436-48.
6. Willcox, B.J., et al., *FOXO3A genotype is strongly associated with human longevity.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(37): p. 13987-92.
7. Newman, A.B., et al., *A Meta-analysis of Four Genome-Wide Association Studies of Survival to Age 90 Years or Older: The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium.* Journals of Gerontology Series A Biological Sciences and Medical Sciences, 2010. **65**(5): p. 478-487.
8. Lunetta, K.L., et al., *Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study.* BMC Med Genet, 2007. **8 Suppl 1**: p. S13.
9. Nebel, A., et al., *A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals.* Mech Ageing Dev, 2011. **132**(6-7): p. 324-30.
10. Deelen, J., et al., *Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited.* Aging Cell, 2011. **10**(4): p. 686-698.
11. Deelen, J., et al., *Identifying the genomic determinants of aging and longevity in human population studies: progress and challenges.* Bioessays, 2013. **35**(4): p. 386-96.
12. Yashin, A.I., et al., *Joint influence of small-effect genetic variants on human longevity.* Aging (Albany NY), 2010. **2**(9): p. 612-20.
13. Benes, P., et al., *The C766T low-density lipoprotein receptor related protein polymorphism and coronary artery disease, plasma lipoproteins, and longevity in the Czech population.* Journal of Molecular Medicine, 2001. **79**(2-3): p. 116-120.

14. Lescai, F., F. Marchegiani, and C. Franceschi, *PON1 is a longevity gene: Results of a meta-analysis.* Ageing Research Reviews, 2009. **8**(4): p. 277-284.
15. Park, J.W., et al., *Candidate gene polymorphisms for diabetes mellitus, cardiovascular disease and cancer are associated with longevity in Koreans.* Experimental and Molecular Medicine, 2009. **41**(11): p. 772-781.
16. Nebel, A., et al., *A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals.* Mechanisms of Ageing and Development, 2011. **132**(6-7): p. 324-330.
17. De Benedictis, G., et al., *Age-related changes of the 3 ' APOB-VNTR genotype pool in ageing cohorts.* Annals of Human Genetics, 1998. **62**: p. 115-122.
18. Ruiz, J.R., et al., *Are centenarians genetically predisposed to lower disease risk?* Age, 2012. **34**(5): p. 1269-1283.
19. Yashin, A.I., et al., *Genes, demography, and life span: The contribution of demographic data in genetic studies on aging and longevity.* American Journal of Human Genetics, 1999. **65**(4): p. 1178-1193.
20. Yashin, A.I., et al., *Genes and longevity: Lessons from studies of centenarians.* Journals of Gerontology Series A Biological Sciences and Medical Sciences, 2000. **55**(7): p. B319-B328.
21. Tan, Q.H., et al., *Logistic regression models for polymorphic and antagonistic pleiotropic gene action on human aging and longevity.* Annals of Human Genetics, 2003. **67**: p. 598-607.
22. Beekman, M., et al., *Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(42): p. 18046-18049.
23. Holstege, H., et al., *A longevity reference genome generated from the world's oldest woman*, in *International Congress of Human Genetics*2011: Montreal, Canada.
24. Shi, H., et al., *Genetic variants influencing human aging from late-onset Alzheimer's disease (LOAD) genome-wide association studies (GWAS).* Neurobiol Aging, 2012. **33**(8): p. 1849 e5-18.
25. Atzmon, G., et al., *Lipoprotein genotype and conserved pathway for exceptional longevity in humans.* PLoS Biol, 2006. **4**(4): p. e113.
26. Bergman, A., et al., *Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging.* PLoS Comput Biol, 2007. **3**(8): p. e170.
27. Yashin, A.I., K.G. Arbeev, and S.V. Ukraintseva, *The accuracy of statistical estimates in genetic studies of aging can be significantly improved.* Biogerontology, 2007. **8**(3): p. 243-55.
28. Yashin, A.I., et al., *How the quality of GWAS of human lifespan and health span can be improved.* Front Genet, 2013. **4**: p. 125.
29. Vaupel, J.W., K.G. Manton, and E. Stallard, *The impact of heterogeneity in individual frailty on the dynamics of mortality.* Demography, 1979. **16**(3): p. 439-54.
30. Vaupel, J.W. and A.I. Yashin, *Heterogeneity's ruses: some surprising effects of selection on population dynamics.* Am Stat, 1985. **39**(3): p. 176-85.
31. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies.* Nature Genetics, 2006. **38**(8): p. 904-909.
32. Tsiatis, A.A. and M. Davidian, *Joint modeling of longitudinal and time-to-event data: An overview.* Statistica Sinica, 2004. **14**(3): p. 809-834.
33. Yu, M.G., et al., *Joint longitudinal-survival-cure models and their application to prostate cancer.* Statistica Sinica, 2004. **14**(3): p. 835-862.
34. Ibrahim, J.G., H. Chu, and L.M. Chen, *Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data.* Journal of Clinical Oncology, 2010. **28**(16): p. 2796-2801.
35. Sousa, I., *A review on joint modelling of longitudinal measurements and time-to-event.* Revstat-Statistical Journal, 2011. **9**(1): p. 57-81.

36. Laird, N.M. and J.H. Ware, *Random-effects models for longitudinal data.* Biometrics, 1982. **38**(4): p. 963-74.
37. Faucett, C.L. and D.C. Thomas, *Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach.* Statistics in Medicine, 1996. **15**(15): p. 1663-1685.
38. Wulfsohn, M.S. and A.A. Tsiatis, *A joint model for survival and longitudinal data measured with error.* Biometrics, 1997. **53**(1): p. 330-339.
39. Ibrahim, J.G., M.H. Chen, and D. Sinha, *Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials.* Statistica Sinica, 2004. **14**(3): p. 863-883.
40. Xu, J. and S.L. Zeger, *The evaluation of multiple surrogate endpoints.* Biometrics, 2001. **57**(1): p. 81-87.
41. Tsiatis, A.A. and M. Davidian, *A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error.* Biometrika, 2001. **88**(2): p. 447-458.
42. Song, X., M. Davidian, and A.A. Tsiatis, *A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data.* Biometrics, 2002. **58**(4): p. 742-753.
43. Song, X.A., M. Davidian, and A.A. Tsiatis, *An estimator for the proportional hazards model with multiple longitudinal covariates measured with error.* Biostatistics, 2002. **3**(4): p. 511-528.
44. Brown, E.R. and J.G. Ibrahim, *A Bayesian semiparametric joint hierarchical model for longitudinal and survival data.* Biometrics, 2003. **59**(2): p. 221-228.
45. Song, X. and C.Y. Wang, *Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients.* Biometrics, 2008. **64**(2): p. 557-566.
46. Taylor, J.M.G., W.G. Cumberland, and J.P. Sy, *A stochastic model for analysis of longitudinal AIDS data.* Journal of the American Statistical Association, 1994. **89**(427): p. 727-736.
47. LaValley, M.P. and V. DeGruttola, *Models for empirical Bayes estimators of longitudinal CD4 counts.* Statistics in Medicine, 1996. **15**(21-22): p. 2289-2305.
48. Wang, Y. and J.M.G. Taylor, *Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome.* Journal of the American Statistical Association, 2001. **96**(455): p. 895-905.
49. Henderson, R., P. Diggle, and A. Dobson, *Joint modelling of longitudinal measurements and event time data.* Biostatistics, 2000. **1**(4): p. 465-480.
50. Ye, W., X. Lin, and J.M.G. Taylor, *Semiparametric Modeling of Longitudinal Measurements and Time-to-Event Data - A Two-Stage Regression Calibration Approach.* Biometrics, 2008. **64**(4): p. 1238-1246.
51. Xu, J. and S.L. Zeger, *Joint analysis of longitudinal data comprising repeated measures and times to events.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 2001. **50**: p. 375-387.
52. Palatini, P., *Need for a revision of the normal limits of resting heart rate.* Hypertension, 1999. **33**(2): p. 622-625.
53. Lewington, S., et al., *Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies.* Lancet, 2002. **360**(9349): p. 1903-1913.
54. Westin, S. and I. Heath, *Thresholds for normal blood pressure and serum cholesterol.* British Medical Journal, 2005. **330**(7506): p. 1461-1462.
55. Seeman, T.E., et al., *Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(8): p. 4770-4775.

56. Karlamangla, A.S., B.H. Singer, and T.E. Seeman, *Reduction in allostatic load in older adults is associated with lower all-cause mortality risk: MacArthur studies of successful aging.* Psychosomatic Medicine, 2006. **68**(3): p. 500-507.
57. Hall, D.M., et al., *Aging reduces adaptive capacity and stress protein expression in the liver after heat stress.* Journal of Applied Physiology, 2000. **89**(2): p. 749-759.
58. Ukraintseva, S.V. and A.I. Yashin, *Individual aging and cancer risk: How are they related?* Demographic Research, 2003. **9**(8): p. 163-196.
59. Yashin, A.I., et al., *Insights on aging and exceptional longevity from longitudinal data: novel findings from the Framingham Heart Study.* Age, 2006. **28**(4): p. 363-374.
60. Yashin, A.I., et al., *Stochastic model for analysis of longitudinal data on aging and mortality.* Math Biosci, 2007. **208**(2): p. 538-51.
61. Troncale, J.A., *The aging process: Physiologic changes and pharmacologic implications.* Postgraduate Medicine, 1996. **99**(5): p. 111-114, 120-122.
62. Lund, J., et al., *Transcriptional profile of aging in C-elegans.* Current Biology, 2002. **12**(18): p. 1566-1573.
63. Goldberger, A.L., C.K. Peng, and L.A. Lipsitz, *What is physiologic complexity and how does it change with aging and disease?* Neurobiology of Aging, 2002. **23**(1): p. 23-26.
64. Arbeev, K.G., et al., *Genetic model for longitudinal studies of aging, health, and longevity and its potential application to incomplete data.* Journal of Theoretical Biology, 2009. **258**(1): p. 103-111.
65. Yashin, A.I., et al., *The quadratic hazard model for analyzing longitudinal data on aging, health, and the life span.* Phys Life Rev, 2012. **9**(2): p. 177-88; discussion 195-7.
66. Arbeev, K.G., et al., *Effect of the APOE Polymorphism and Age Trajectories of Physiological Variables on Mortality: Application of Genetic Stochastic Process Model of Aging.* Scientifica (Cairo), 2012. **2012**: p. Article ID 568628.
67. Yashin, A.I., et al., *How lifespan associated genes modulate aging changes: lessons from analysis of longitudinal data.* Front Genet, 2013. **4**: p. 3.
68. Witteman, J.C.M., et al., *J-shaped relation between change in diastolic blood pressure and progression of aortic atherosclerosis.* Lancet, 1994. **343**(8896): p. 504-507.
69. Allison, D.B., et al., *Hypothesis concerning the U-shaped relation between body mass index and mortality.* American Journal of Epidemiology, 1997. **146**(4): p. 339-349.
70. Boutitie, F., et al., *J-shaped relationship between blood pressure and mortality in hypertensive patients: New insights from a meta-analysis of individual-patient data.* Annals of Internal Medicine, 2002. **136**(6): p. 438-448.
71. Kulminski, A.M., et al., *Body mass index and nine-year mortality in disabled and nondisabled older U.S. Individuals.* Journal of the American Geriatrics Society, 2008. **56**(1): p. 105-110.
72. Kuzuya, M., et al., *J-shaped relationship between resting pulse rate and all-cause mortality in community-dwelling older people with disabilities.* Journal of the American Geriatrics Society, 2008. **56**(2): p. 367-368.
73. Mazza, A., et al., *Serum uric acid shows a J-shaped trend with coronary mortality in non-insulin-dependent diabetic elderly people. The CArdiovascular STudy in the ELderly (CASTEL).* Acta Diabetologica, 2007. **44**(3): p. 99-105.
74. Okumiya, K., et al., *A U-shaped association between home systolic blood pressure and four-year mortality in community-dwelling older men.* Journal of the American Geriatrics Society, 1999. **47**(12): p. 1415-1421.
75. Protogerou, A.D., et al., *Diastolic blood pressure and mortality in the elderly with cardiovascular disease.* Hypertension, 2007. **50**(1): p. 172-180.

76. Troiano, R.P., et al., *The relationship between body weight and mortality: A quantitative analysis of combined information from existing studies.* International Journal of Obesity, 1996. **20**(1): p. 63-75.
77. Neafsey, P.J., *Longevity hormesis. A review.* Mech Ageing Dev, 1990. **51**(1): p. 1-31.
78. Wu, D., et al., *The U-shaped response of initial mortality in Caenorhabditis elegans to mild heat shock: does it explain recent trends in human mortality?* J Gerontol A Biol Sci Med Sci, 2008. **63**(7): p. 660-8.
79. Arbeev, K.G., et al., *Genetic model for longitudinal studies of aging, health, and longevity and its potential application to incomplete data.* J Theor Biol, 2009. **258**(1): p. 103-11.
80. Arbeev, K.G., et al., *Evaluation of genotype-specific survival using joint analysis of genetic and non-genetic subsamples of longitudinal data.* Biogerontology, 2011. **12**(2): p. 157-66.
81. Yashin, A.I., et al., *The quadratic hazard model for analyzing longitudinal data on aging, health, and the life span.* Physics of Life Reviews, 2012. **9**(2): p. 177-188.
82. Yashin, A.I., et al., *Stochastic model for analysis of longitudinal data on aging and mortality.* Mathematical Biosciences, 2007. **208**(2): p. 538-551.
83. Rankin, M.M. and J.A. Kushner, *Adaptive beta-cell Proliferation Is Severely Restricted With Advanced Age.* Diabetes, 2009. **58**(6): p. 1365-1372.
84. Yashin, A.I., et al., *Maintaining physiological state for exceptional survival: What is the normal level of blood glucose and does it change with age?* Mechanisms of Ageing and Development, 2009. **130**(9): p. 611-618.
85. Yashin, A.I., et al., *What age trajectories of cumulative deficits and medical costs tell us about individual aging and mortality risk: Findings from the NLTCS-Medicare data.* Mechanisms of Ageing and Development, 2008. **129**(4): p. 191-200.
86. Yashin, A.I., et al., *Health decline, aging and mortality: how are they related?* Biogerontology, 2007. **8**(3): p. 291-302.
87. Yashin, A.I., et al., *Exceptional survivors have lower age trajectories of blood glucose: lessons from longitudinal data.* Biogerontology, 2010. **11**(3): p. 257-265.
88. Arbeev, K.G., et al., *Age trajectories of physiological indices in relation to healthy life course.* Mechanisms of Ageing and Development, 2011. **132**(3): p. 93-102.
89. Vaupel, J.W., et al., *Biodemographic trajectories of longevity.* Science, 1998. **280**(5365): p. 855-860.
90. Semenchenko, G.V., et al., *Stress resistance declines with age: analysis of data from a survival experiment with Drosophila melanogaster.* Biogerontology, 2004. **5**(1): p. 17-30.
91. Robb, E.L., M.M. Page, and J.A. Stuart, *Mitochondria, cellular stress resistance, somatic cell depletion and lifespan.* Current Aging Science, 2009. **2**(1): p. 12-27.