Genetic model for longitudinal studies of aging, health, and longevity and its application to Framingham Heart Study data

Konstantin G. Arbeev, Igor Akushevich, Alexander M. Kulminski, Liubov S. Arbeeva, Lucy Akushevich, Svetlana V. Ukraintseva, Irina V. Culminskaya, Anatoli I. Yashin

Center for Population Health and Aging, Duke University, Trent Hall, Room 002, Box 90408, Durham, NC, 27708-0408, USA.

The influence of genes on aging, health and longevity is mediated by thousands of biological and physiological variables which are also affected by environmental, behavioral and other factors. Some of such variables are measured in longitudinal studies of aging, health and longevity. That is why the data on genetic markers collected for participants of a longitudinal study are probably most appropriate for evaluating the genetic contribution to the aging-related decline in the health/well-being status and the life span. Such data, however, often cannot be collected for all participants of the study. This is because: (i) the large-scale collection of genetic data is a relatively new business, thus, some individuals, who initially participated in a longitudinal study, have already died or dropped out of a population; (ii) obtaining genetic information is still an expensive business and cannot be performed at the same scale as medical examinations or a sociological survey; (iii) not all individuals who agreed to participate in a medical examination or to respond to the survey's questionnaire agree to participate in a genetic analysis. Thus, the presence of genetic information divide participants of a longitudinal study into two groups: one (the genetic group) includes those for whom genetic data were also collected. The other (the non-genetic group) consists of those for whom longitudinal data are available but genetic information was not collected.

Such a situation when information on covariates essential for analyses of risks is missing for some sub-sample of individuals (either due to cost limitations or by the study design) is typical in epidemiological studies. For example, two-stage designs are routinely used in epidemiology when a disease status (or other general information) is ascertained for a large group of individuals at the first stage and information on covariates essential for analyses of their relation to the risk of the disease is collected at the second stage for smaller sub-samples of individuals. Statistical methods for analyses of such data are well developed for regression models. One of the main advantages of such methods is that they use information from the first and second stages to estimate regression parameters. This can lead to a considerable improvement in the efficiency of estimates compared to the estimates based on the second stage data alone. Applications of such designs and methods in genetic epidemiology are also discussed in the literature.

A traditional way of evaluating effects of genes on individuals' health/well-being/survival status is to directly estimate respective hazards (e.g., incidence or mortality rate) for carriers of a selected allele (genotype). Such practice is completely justified in the absence of data about other factors and processes affecting these characteristics. The advantage of longitudinal data for the genetic studies of aging and longevity is in the

opportunity to estimate not only direct genetic effects on morbidity and mortality but also indirect genetic effects mediated by age trajectories of physiological variables collected in the longitudinal study (which may modulate mechanisms of aging not directly measured in longitudinal data).

The purpose of this study is to elaborate a genetic model for studying longitudinal data on aging, health, and longevity which would permit: 1) joint analyses of genetic and non-genetic data to make use of all available information and increase the accuracy of estimates compared to analyses of genetic data alone; 2) evaluation of indirect genetic effects mediated by age trajectories of physiological variables collected in a longitudinal study; and 3) incorporation of essential mechanisms of aging-related changes in organisms that are not directly measured in longitudinal data but can be estimated from individual age trajectories of physiological indices and data on mortality or morbidity. The stochastic process model (SPM) of human mortality and aging (Manton and Yashin 2000; Woodbury and Manton 1977; Yashin 1985; Yashin and Manton 1997) is the conceptual approach in this study and its extension presented here has all three above-mentioned properties. The important feature of the SPM is a biologically-justified U- or J- shaped risks as functions of respective indices. Such shapes of the risk functions are observed for different physiological indices. The original SPM was recently modified (Yashin et al. 2007a) to include major concepts of aging known to date: *age-specific physiological norms* (Lewington et al. 2002; Palatini 1999; Westin and Heath 2005), *allostasis* and *allostatic load* (Karlamangla et al. 2006; Seeman et al. 2001), the decline in *adaptive capacity* with age (*homeostenosis*) (Lund et al. 2002; Troncale 1996), the decline in *stress resistance* with age (Hall et al. 2000; Ukraintseva and Yashin 2003; Yashin et al. 2006), and *stochasticity* (Goldberger et al. 2002). The one- and two-dimensional versions of the model were successfully applied to different data sets to reveal complicated interplay among different components of aging-related changes in humans (Yashin et al. 2007b; Yashin et al. 2007c; Yashin et al. 2008). The model presented in this study is a step forward in analyzing contribution of genes to dynamic regularities in aging-related changes in a human organism. This model incorporates information on genetic markers collected for a sub-sample of participants of a longitudinal study and permits evaluation of all above-mentioned characteristics (age-specific norms, decline in stress resistance, etc.), as well as respective hazard rates, for carriers and non-carriers of a selected allele (genotype) to address questions concerning genetic influence on these aging-related characteristics (here we formulated the model for two types of individuals: carriers and non-carriers of some selected allele/genotype, however, its extension to the case of many alleles/genotypes is straightforward). The method is based on extracting genetic information from the entire sample of longitudinal data consisting of genetic (those with available genetic information) and non-genetic (those for whom genetic information was not collected) sub-samples. The group of individuals with genetic data becomes automatically divided into subgroups of carriers and non-carriers of respective alleles or genotypes. The non-genetic group consists of carriers of the same genotypes identified in the genetic group and, hence, non-genetic data contain information about genetic influence on all phenotypes observed in a longitudinal study. We developed statistical methods for extracting genetic information from the entire sample of longitudinal data consisting of genetic and non-genetic sub-

samples. This joint analysis results in a substantial increase in the accuracy of statistical estimates of genetic parameters (without collecting additional genetic data) compared to methods that use only information from a genetic sub-sample. Simulation studies illustrated the increase in the accuracy in different scenarios for datasets structurally similar to the Framingham Heart Study (FHS) (Dawber et al. 1951). The model was applied to the FHS data (exams 1-26) that contain information on: Angiotensin I converting enzyme (ACE) and Apolipoprotein E (APOE) common polymorphisms for a sub-sample of the participants ("genetic sub-sample"), detailed longitudinal information on different physiological indices (such as blood pressure, pulse pressure, pulse rate, serum cholesterol, blood glucose, hematocrit, and body mass index) and risks of chronic degenerative diseases (such as cardiovascular disease, stroke, cancer, diabetes, hypertension) and mortality risk for the entire sample ("genetic" and "non-genetic" sub-samples). The proposed innovative advanced tool for statistical analyses of such data allows for capturing systemic regularities of changes in health/well-being/survival status in an aging human organism.

References

Dawber T.R., Meadors G.F., Moore F.E. (1951) Epidemiological approaches to heart disease - The Framingham Study. *American Journal of Public Health* 41 (3): 279-286.

Goldberger A.L., Peng C.K., Lipsitz L.A. (2002) What is physiologic complexity and how does it change with aging and disease? *Neurobiology of Aging* 23 (1): 23-26.

Hall D.M., Xu L., Drake V.J., Oberley L.W., Oberley T.D., Moseley P.L., Kregel K.C. (2000) Aging reduces adaptive capacity and stress protein expression in the liver after heat stress. *Journal of Applied Physiology* 89 (2): 749-759.

Karlamangla A.S., Singer B.H., Seeman T.E. (2006) Reduction in allostatic load in older adults is associated with lower all-cause mortality risk: MacArthur studies of successful aging. *Psychosomatic Medicine* 68 (3): 500-507.

Lewington S., Clarke R., Qizilbash N., Peto R., Collins R., Prospective Studies C. (2002) 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* 360 (9349): 1903-1913.

Lund J., Tedesco P., Duke K., Wang J., Kim S.K., Johnson T.E. (2002) Transcriptional profile of aging in C-elegans. *Current Biology* 12 (18): 1566-1573.

Manton K.G., Yashin A.I. (2000) *Mechanisms of Aging and Mortality: A Search for New Paradigms. Odense Monograph on Population Aging No. 7*. Odense University Press: Odense, Denmark.

Palatini P. (1999) Need for a revision of the normal limits of resting heart rate. *Hypertension* 33 (2): 622-625.

Seeman T.E., McEwen B.S., Rowe J.W., Singer B.H. (2001) 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* 98 (8): 4770-4775.

Troncale J.A. (1996) The aging process - Physiologic changes and pharmacologic implications. *Postgraduate Medicine* 99 (5): 111-114, 120-122.

Ukraintseva S.V., Yashin A.I. (2003) Individual aging and cancer risk: How are they related? *Demographic Research* 9 (8): 163-196.

Westin S., Heath I. (2005) Thresholds for normal blood pressure and serum cholesterol. *British Medical Journal* 330 (7506): 1461-1462.

Woodbury M.A., Manton K.G. (1977) Random-walk model of human mortality and aging. *Theoretical Population Biology* 11 (1): 37-48.

Yashin A.I. (1985) Dynamics in survival analysis: Conditional Gaussian property vs. Cameron-Martin formula. In: Krylov N.V., Lipster R.S., Novikov A.A. (Eds.) "*Statistics and Control of Stochastic Processes*." Springer: New York, pp. 446–475.

Yashin A.I., Akushevich I.V., Arbeev K.G., Akushevich L., Ukraintseva S.V., Kulminski A. (2006) Insights on aging and exceptional longevity from longitudinal data: novel findings from the Framingham Heart Study. *Age* 28 (4): 363-374.

Yashin A.I., Arbeev K.G., Akushevich I., Kulminski A., Akushevich L., Ukraintseva S.V. (2007a) Stochastic model for analysis of longitudinal data on aging and mortality. *Mathematical Biosciences* 208 (2): 538-551.

Yashin A.I., Arbeev K.G., Kulminski A., Akushevich I., Akushevich L., Ukraintseva S.V. (2007b) Cumulative index of elderly disorders and its dynamic contribution to mortality and longevity. *Rejuvenation Research* 10 (1): 75-86.

Yashin A.I., Arbeev K.G., Kulminski A., Akushevich I., Akushevich L., Ukraintseva S.V. (2007c) Health decline, aging and mortality: how are they related? *Biogerontology* 8 (3): 291-302.

Yashin A.I., Arbeev K.G., Kulminski A., Akushevich I., Akushevich L., Ukraintseva S.V. (2008) 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* 129 (4): 191-200.

Yashin A.I., Manton K.G. (1997) Effects of unobserved and partially observed covariate processes on system failure: A review of models and estimation strategies. *Statistical Science* 12 (1): 20-34.