Medical Infobahn for Europe
A. Hasman et al. (Eds.)
IOS Press, 2000

Multivariate Frailty Model with a Major Gene: Application to Genealogical Data

Alexander BEGUN
MPI for Demographic Research, Doberanerstrasse 114, 18057 Rostock, Germany
Bertran DESJARDINS
Université de Montréal, C.P.6128, succ.centre-ville, Montreal(Qc), Canada H3C 3J7
Ivan IACHINE
Department of Statistics and Demography, Odense University, Denmark
Anatoli YASHIN
MPI for Demographic Research, Doberanerstrasse 114, 18057 Rostock, Germany
Duke University, Center for Demographic Studies, USA

Abstract. Multivariate survival models are shown to be appropriate for the analysis of the genetic and the environmental nature of a human life-span. Models which involve continuously distributed individual frailty, play an important role in the genetic analysis of an individual's susceptibility to disease and death. These models, however, are not appropriate for the detection of the effects of separate genes on survival. For this purpose we developed a 'major gene' frailty model of multivariate survival and applied it to simulated and real pedigree data. The analysis shows that this model can be used for the detection of the presence of major genes in the population and for the evaluation of the effects of such genes on survival.

1. Introduction

One of the most important objectives in longevity studies is to identify genetic and environmental influence on the life-span. Since Fisher's time, the analysis of variances has been used for the determination of the heritability of some measurable metric traits. The hypothesis about linearity and the independence of genetic and environmental components of the trait is the basis for such an analysis.

Frailty modelling allows us to develop genetic analysis of traits, which are unobserved or non-measured and which are related by non-linear dependence with observed traits. The properties and experience of the application of univariate and bivariate frailty models for different frailty distributions are discussed in a number of papers [1-6].

To make conclusions about the genetic nature of life-span, data on related individuals are required. That is, we must deal with multivariate data using multivariate frailty models. It turns out that multivariate frailty models with discretely distributed frailty such as, for example, the 'major gene' frailty model, are often more convenient for the analysis of family data than the frailty models with continuously distributed frailty. In socalled semiparametric form, the multivariate survival function can be expressed as a function of univariate survival functions for related individuals given their genotypes.

Genealogical data, including only the birth/death dates of related individuals, contain information about the factors, characterising the global (social) and the local (familial) environment that affect the life-span. This includes information about cohort, age of child at mother's/father's death, reproductive age of the mother/father, mother's/father's life-span, et cetera. These factors may be easily included in the multivariate survival function as covariates in Cox's regression model.

In this paper, we implement the multivariate frailty model with 'major gene' to French-Canadian sibs born in Quebec between 1623 and 1705. In accordance with likelihood ratio test, the optimal model with one beneficial allele in Hardy-Weinberg equilibrium was chosen and estimates of univariate survival functions, allele frequency and coefficients in Cox's regression were calculated.

2. Material and method

From 13544 records, relating to French-Canadian children born in Quebec between 1623 and 1705, 2066 children from 793 families (1016 boys and 1050 girls) were chosen with valid birth/death dates, who survived until 30 and overlived their parents. The number of children in families fluctuates between 1 and 10 with a mean of 2.6. Using original data, we built 7 covariates; u_1 = year of birth -1650, u_2 = the age of a child at father's death, u_3 - the age of a child at mother's death, u_4 - the reproductive age of a father, u_5 - the reproductive age of a mother, u_6 - the life-span of a father and u_7 - the life-span of a mother. The second and the third covariates were categorized as a follows: 0 if $u_{2,3} \leq 5$, 1 if $5 < u_{2,3} \leq 10$, 2 if $5 < u_{2,3} \leq 10$, 3 if $10 < u_{2,3} \leq 15$, 4 - otherwise. The fourth and the fifth were put to 0 if $u_{4,5} \leq 35$ and 1 - otherwise. We assumed that the sixth and the seventh covariates are 0 if $u_{6,7} \leq 75$ and 1 otherwise.

In the model with linear influence of observed covariates on the risk of mortality, we assume that an individual's instantaneous risk of death μ at age t, as measured by the hazard of mortality, depends linearly on frailty Z, $\mu(t;Z)$=$Z\exp(\beta^*u)\mu_0(t)$, where $\mu_0(t)$ is underlying hazard, β and u are vector-columns of Cox's regression coefficients and covariates, Z is frailty. We assume, that $Z= a_{ij}= a_{ji}>0$ with probabilities $p_{ij}=p_{ji}$, $\sum_{i,j} p_{ij} = 1$, i,j=1,2 and the first allele in the index pair of the genotype (i,j) is inherited from the father and allele j is inherited from the mother, both independently. Parents are chosen independently and all persons have the same fertility. In Hardy-Weinberg equilibrium $p_{ij}= p_i\ p_j$. If frailty is conditioned by only one beneficial allele, then we have $a = a_0(1-r)^k$ in the case of k beneficial alleles in the genotype (k=0,1,2), with independent multiplicative action.

We can obtain the multivariate survival function in the form

$$S(x_1,\ldots,x_n) = \sum_{k_1,\ldots,k_n} \tilde{p}_{k_l \ldots k_n} S_0^{\tilde{a}_{k_1}}(x_1) \ldots S_0^{\tilde{a}_{k_n}}(x_n),$$

$$S_0(.) = \exp(-\exp(\beta^* u)H(.)),\ H(x)=\int_0^x \mu_0(\tau)d\tau,$$

where we sum all possible combinations of n-sibs genotypes $(k_1,\ldots,k_n)$ with corresponding weights $\widetilde{p}_{k_1\ldots k_n}$. Denote genotype (1,1) as 1, genotypes (1,2) and (2,1) as 2, and genotype (2,2) as 3. Let $\widetilde{a}_{1,s} = a_{11,s}, \widetilde{a}_{2,s} = a_{12,s} = a_{21,s}, \widetilde{a}_{3,s} = a_{22,s}$, $\widetilde{p}_{1,s} = p_{11,s}, \widetilde{p}_{2,s} = p_{12,s} + p_{21,s}, \widetilde{p}_{3,s} = p_{22,s}$, sex s=m,f (males, females). It may be shown, that in the case of autosomal locus, the multivariate survival functions may be rewritten as a follows

$$\begin{aligned}
S_s(x_1,\ldots,x_n) &= \widetilde{p}_{1,m}\widetilde{p}_{1,f}\prod_{i=1}^{n} S_{0,s}{}^{\widetilde{a}_{1,s}}(x_i) + 0.5^n(\widetilde{p}_{1,m}\widetilde{p}_{2,f} + \widetilde{p}_{2,m}\widetilde{p}_{1,f})\prod_{i=1}^{n}[S_{0,s}{}^{\widetilde{a}_{1,s}}(x_i) + S_{0,s}{}^{\widetilde{a}_{2,s}}(x_i)] + \\
&+0.5^n\widetilde{p}_{2,m}\widetilde{p}_{2,f}\prod_{i=1}^{n}[S_{0,s}{}^{\widetilde{a}_{2,s}}(x_i) + 0.5S_{0,s}{}^{\widetilde{a}_{1,s}}(x_i) + 0.5S_{0,s}{}^{\widetilde{a}_{3,s}}(x_i)] + \\
&+0.5^n(\widetilde{p}_{2,m}\widetilde{p}_{3,f} + \widetilde{p}_{3,m}\widetilde{p}_{2,f})\prod_{i=1}^{n}[S_{0,s}{}^{\widetilde{a}_{2,s}}(x_i) + S_{0,s}{}^{\widetilde{a}_{3,s}}(x_i)] + \qquad (1) \\
&+(\widetilde{p}_{1,m}\widetilde{p}_{3,f} + \widetilde{p}_{3,m}\widetilde{p}_{1,f})\prod_{i=1}^{n} S_{0,s}{}^{\widetilde{a}_{2,s}}(x_i) + \widetilde{p}_{3,m}\widetilde{p}_{3,f}\prod_{i=1}^{n} S_{0,s}{}^{\widetilde{a}_{3,s}}(x_i).
\end{aligned}$$

This formula is an extension of the result derived for bivariate case in [7].
Marginal univariate survival functions were approximated with the formula

$$S(x) = \left[1 + s^2\left(a(x - x_0) + \frac{b}{c}(e^{cx} - e^{cx_0})\right)\right]^{-\frac{1}{s^2}}, \qquad (2)$$

where a, b, c, s are unknown parameters.

Given unknown parameters we can substitute the univariate survival function (2) in the left side of the equation (1) for n=1, zero covariates and find $H(.)$. Then we can calculate the multivariate survival function in the general case n≥1 and use it in the likelihood estimation procedure.

In the case of non-linear influence of observed covariates on the risk of mortality, we introduce them in the univariate survival function as follows. We calculate the intermediate mortality risk as a function on unknown parameters a, b, c, s

$$\widetilde{\mu}(x) = \frac{a + be^{cx}}{1 + s^2\left(a(x - x_0) + \frac{b}{c}(e^{cx} - e^{cx_0})\right)}.$$

Then using unknown parameters g and q we construct fictitious genotype frequencies $g_{1,s} = g^2, g_{2,s} = 2g(1-g), g_{3,s} = (1-g)^2$, genotype risks $q_{1,s} = q^2, q_{2,s} = q, q_{3,s} = 1$ and find univariate survival function $S_s(x) = \sum_k g_{k,s}\exp(-q_{k,s}\exp(\beta^* u)\widetilde{h}_s(x))$, where $\widetilde{h}_s(x) = \int_{x0}^{x}\widetilde{\mu}(t)dt$. We substitute this univariate survival function in the left side of the equation (1) for n=1 and zero covariates and find $H(.)$. The multivariate survival function we calculate from formula (1) with zero covariates.

3. Results and discussions

The computer program for our model was written in MATLAB and tested on simulated data. Then it was used to estimate unknown parameters on the data of French-Canadians. All estimates were obtained through a maximization of the likelihood function. The likelihood ratio test was used to choose the set of parameters. Since the models with linear and non-linear influence of observed covariates on the risk of mortality are not nested, we used AKAIKE criteria to make a choice between them. In accordance with this criteria, we chose a non-linear model. All other choices were made using the likelihood ratio test. In exception of the univariate fit parameters, there were no significant differences between males and females. It was found that we can use the model with one beneficial allele 'a' in Hardy-Weinberg equilibrium with allele frequency $p = 0.406$ and multiplicative action $1-r = 0.485$. We have found no cohort effect and no effect of age of a child at parental death for males or females. Only two significant coefficients of Cox's regression were found: $\beta_4 = \beta_5 = 0.188$, $\beta_6 = \beta_7 = -0.451$. That is, we can find the beneficial allele in about 41% of the cases and the presence of each beneficial allele in the genotype decreases the mortality risk by about 2.1 times (we have set risk for the genotype with no beneficial alleles at $a_0 = 1$). The greater a parent's life-span is, the less a child's mortality risk will be. On the contrary, the higher a reproductive age of a parent is, the greater the mortality risk is of an offspring. All these dependencies are shown in Table 1. One can see from this table, that the life expectancy at 30 for females is greater in all cases with the exception of the first two columns for the worse genotype. Each beneficial allele in genotype increases the life expectancy by about 6-7 years. Father's (mother's) reproductive age after 35 decreases the life expectancy by approximately 1.5 years. The parent's life-span of more than 75 increases the offspring's longevity by about 3-4 years, compared to the case with a parent's life-span of less than 75.

Table 1. Life expectancy at age 30

Genotype	Sex	$u_4 = u_5 = 0$ $u_6 = u_7 = 0$	u_4 (or u_5)=1 $u_6 = u_7 = 0$	$u_4 = u_5 = 0$ u_6 (or u_7)=1
AA	male	30.77	29.33	34.25
	female	30.72	29.09	34.59
aA or	male	37.44	35.99	40.93
Aa	female	38.00	36.44	41.70
aa	male	44.30	42.85	47.75
	female	45.16	43.67	48.65

We can explain these results as a follows. Familial environment may have profound effects not only on infant/childhood mortality, but also on adult mortality. The most important factors of this environment are the parental longevity and the parental reproduction age. The genetic material, which a parent transmits to its offspring might be essentially damaged in the reproductive age after 35, which leads to a shorter child's longevity. But the effect of the parental longevity is stronger. It does not mean that only genetic factors play the crucial role in child's longevity. Familial habits and the life-style can affect the life-span to some degree as well. Further studies are needed to confirm and explain these findings.

References

[1] J.Vaupel, K.Manton and E.Stallard, The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**(1979) 439-454.

[2] J.Vaupel, A.Yashin, Deviant dynamics of death in heterogeneous populations. In: N.Tuma (ed)., *Sociological Methodology,* San Francisco, Jossey-Bass 1985, pp.179-211.

[3] P.Hougaard, Frailty Models Derived from the stable Distributions. Preprint N7. Institute of Mathematical Statistics, University of Copenhagen, 1984.

[4] D.Clayton, J.Cuzick, Multivariate generalization of the proportional hazards model. *Journal of Royal Statistical Society, Ser., A* **148**(1985) 82-117.

[5] A.Yashin, I.Iachine, Environment determines 50% of variability in individual frailty: Results from Danish twin study. Research report, population studies of aging, Odense University, Denmark, 1994.

[6] J.Vaupel and Q.Tan, How Many Longevity Genes Are There? In: Population association of America. 1998 Annual meeting. Chicago, Illinois, April 2-4, 1998.

[7] A.Begun, I.Iachine and A.Yashin, Genetic Nature of Individual Frailty: Comparison of Two Approaches. *Twin Research* **3**(2000) 51-57.