

HIDDEN FRAILTY: MYTHS AND REALITY

by

Anatoli I. Yashin^{1,2}, Ivan A. Iachine³, Alexander Z. Begun¹, and
James W. Vaupel^{1,4}

*¹Max Planck Institute for Demographic Research, Rostock
Germany; ²Duke University, Center for Demographic Studies,
USA; ³University of Southern Denmark, Department of
Statistics and Demography, Odense, Denmark; ⁴Duke
University, Sanford Institute, USA*

Correspondence should be directed to:

Prof. A. Yashin, Max Planck Institute for Demographic Research

Doberaner Str. 114, 18057 Rostock, Germany

e-mail: yashin@demogr.mpg.de; fax: +49 381 2081 169; phone: +49 381 2081 106

Acknowledgments

This research was partly supported by NIH/NIA grant PO1 AG08791-01. The authors thank Baerbel Splettstoesser and Karl Brehmer for help in preparing this paper for publication.

Synopsis. In this paper we discuss the notion of individual frailty and its interpretation. In addition, we consider the application of this concept to different areas of demography, epidemiology, and the genetics of aging.

Abstract

Hidden differences in the survival chances of individuals in a population influence the shape of the mortality rate observed in demographic and epidemiological studies. To evaluate the contribution of such hidden variations to observed hazards, frailty models have been suggested. The application of these models to the analysis of survival data in demography, epidemiology, and biostatistics has opened up new avenues for survival studies. However, along with many useful insights and ideas, several unjustified beliefs (myths) have also been generated. In this paper we critically discuss these beliefs. In particular, we discuss the notion of individual frailty and show that the interpretation thereof depends on the identifiability conditions specified for the respective frailty model. We discuss strengths and weaknesses of shared frailty models with and without observed covariates. We explain why bivariate correlated frailty models are the most appropriate for the analysis of survival data on related individuals. We discuss new bivariate survival models with non-gamma frailty distribution and potential directions for further research.

1. INTRODUCTION

Unobserved susceptibility to death, also called “hidden heterogeneity” or “frailty”, is a major concern in the demographic analysis of survival, where individual variations in survival chances cannot be ignored. To take such differences into account, frailty models have been suggested (Vaupel et al., 1979). These models are used to explain the deviant behavior of mortality rates at advanced ages (Vaupel and Yashin, 1985), to correct biased estimates of regression coefficients in Cox-type models of hazard rate (Chamberlain, 1985) and to separate compositional and biological effects in aging studies (Manton et al., 1986). Frailty models play an important role in the interpretation of the results of stress experiments (Yashin et al., 1996a) and in centenarian studies (Yashin et al., 1999). The use of survival data on related individuals opens a new avenue in frailty modeling. Genetic variation, heritability, and other properties of individual susceptibility to death can now be analyzed using correlated frailty models (Yashin and Iachine, 1995a,b; Yashin and Iachine 1997).

The use of frailty modeling has its limitations, as is the case, of course, with any kind of modeling. Unjustified beliefs (myths) arise when such limitations are ignored or not well understood. Such beliefs can promote the misuse of the models and also the misinterpretation of the results of analyses. Often, they exaggerate the ability of this approach to address real problems. In

this paper we discuss some of these unjustified beliefs. In particular, we explain why frailty models used in univariate and traditional bivariate analyses (i.e. analyses based on the idea of shared frailty) have radically different properties. We discuss why the bivariate correlated frailty models are more appropriate in the analysis of survival data on related individuals than the shared frailty models. We then evaluate the efficiency of the estimates of regression coefficients in the bivariate frailty models. Finally, we describe new bivariate survival models with non-gamma frailty distributions and discuss potential directions for further research.

2. FRAILTY MODELING

Frailty models without observed covariates. Such models are used when only survival data are available for the analysis, or when additional information is of no interest. These models are described by the stochastic hazards, $\mu(Z,x)=Z\mu_0(x)$, or

$$\mu(Z,x)=Z\mu_0(x) + \tilde{\mu}(x) \quad (1)$$

Here the non-negative random variable Z is called frailty, $\mu_0(x)$ is an underlying hazard and $\tilde{\mu}(x)$ is a “background” hazard. This model is non-identifiable from survival data, i.e., different combinations of $\mu_0(x)$, $\tilde{\mu}(x)$ and frailty distributions may produce the same marginal hazard rate $\bar{\mu}(x)$. The model becomes

identifiable when the parametric structures of $\mu_0(x)$ and $\tilde{\mu}(x)$ are fixed and Z is assumed to belong to some parametric distribution family, e.g. gamma.

Frailty modeling represents an attempt to take a deeper look at the mechanisms of aging and survival than the traditional non-parametric methods of statistical analysis of survival data allow us to do. By using these models we accept the fact that hidden internal or external factors exist that influence mortality and survival. We also assume that information about such factors is contained in the shape and structure of conditional hazards and in the form of frailty distributions. In some analyses the frailty variable may undergo further decomposition. For example, the decomposition of frailty into genetic and environmental components has been used in the analysis of twin data to better understand the roles of genes and the environment in susceptibility to death and longevity (Yashin and Iachine 1995a,b).

Frailty models with observed covariates. In frailty models with the vector of observed covariates U the individual (conditional) hazard is a function of Z and U . One of the traditionally used forms of this hazard is

$$\mu(Z, U, x) = Z\mu_0(x) e^{\beta^*U} \quad (2)$$

Here β is a vector of regression coefficients characterizing the measure of influence of U on the hazard rate, and the symbol "*" denotes transposition. Usually it is assumed that U and Z are independent. To simplify our further discussion we assume that covariate U is a scalar.

The random variable Z is often used to account for omitted covariates (Andersen et al., 1992). Another role of the random effect in model (2) is to describe the non-proportionality of the conditional hazards $\bar{\mu}(U, x)$ in order to improve the fit to the model as compared to the traditional proportional hazards models used in the Cox regression. Univariate models (2), which are frequently used in studies of unrelated individuals, are identifiable if $EZ < \infty$. In this case no assumptions about $\mu_0(x)$ or about the class of distributions of Z are needed (Elbers and Ridder, 1982). The proof of the identifiability property given by Elbers and Ridder (1982) does not provide us with statistical methods capable of estimating respective characteristics of the model from the data. The respective procedures, a version of the EM-algorithm, were later realized for the gamma frailty models (Andersen et al. 1992). Note that when $EZ = \infty$, e.g. in the case of positive stable frailty distribution (Hougaard, 1986), model (2) is not identifiable. In this case the additional information represented by the covariate is not sufficient for distinguishing between the different frailty/hazard combinations. In this case the identifiability problem can be solved by using data on related individuals.

Many studies have focused on evaluating the effects of omitted covariates on the estimate of regression coefficient β in model (2). It has been shown that the influence of the observed covariate is underestimated when the presence of omitted covariates is ignored (e.g. Chamberlain 1985, Gail 1984). This problem

can be corrected if one models the omitted covariate appropriately, e.g. by using a frailty model. It is important to point out, however, that the estimates of regression coefficients in these two situations (i.e. when the omitted covariate is included in the model, either as an observed variable or as a random effect, and when the omitted covariate is ignored) have different meanings. In the former case, the estimate represents the effect of the covariate on the individual level, i.e., when the value of the omitted variable is known. In the latter case, the estimate represents the marginal effect of the omitted covariate on survival. The distinction between the two cases might be important when dealing with forecasting, where the type of regression coefficient to be used for predicting the outcome should depend on the available information.

Bivariate models. Initially, bivariate and multivariate frailty models exploited the idea of shared frailty. These models were used in studies of the effects of dependence between life spans (or other durations) on the estimates of regression coefficients in the Cox-type hazard model with random effect (Clayton, 1978; Clayton and Cuzick, 1985; Huster, 1987; Hougaard, 1995). Although parametric specification of a frailty distribution is not necessary in models characterized by hazard (2) (Elbers and Ridder 1982), such a specification is often used as a matter of computational convenience. However, incorrect parametric specification can, in fact, bias the estimates of regression coefficients. Another appealing feature of multivariate frailty models is their identifiability in the absence of observed covariates. The information contained

in bivariate survival data along with some additional assumptions used when extending the univariate frailty model (2) to the multivariate case (e.g. the conditional independence assumption, additive frailty structure) is enough to reconstruct both the frailty distribution and the underlying hazard functions (Iachine and Yashin, 1998).

The assumption about the proportionality of a hazard makes frailty models a simple and convenient tool for analytical and computer calculations. The maximum likelihood estimation for frailty models is facilitated by the fact that the likelihood function can be expressed in terms of the Laplace transform of the frailty distribution (Aalen, 1987). Semiparametric estimates of the hazard function and other parameters of the model can be obtained in the presence of censoring by means of the EM-algorithm (Nielsen et al., 1992, Iachine and Yashin, 1995), and they have appealing asymptotic properties (Parner, 1998, Korsholm, 1999).

Despite their simplicity these models turn out to be useful in providing insights and ideas for analyzing real-life phenomena. They have been used to explain the deceleration of the mortality rate at older ages (Vaupel et al. 1979, Vaupel and Yashin 1985, Yashin et al., 1994; Vaupel et al., 1998). They allow us to account for omitted covariates and heterogeneity when calculating the effects of observed covariates on survival in regression studies (Chamberlain 1985). They help to clarify the role of genes and the environment in the variation in human life span, (Yashin and Iachine 1997), etc. The major

criticism of these models is that the basic, underlying assumptions are insufficiently grounded in biological reality: the proportional structure of the hazard, the form of a frailty distribution, and the independence of observed and omitted covariates. Other limitations of frailty models are less well-known. They are related to some unjustified concepts and beliefs (myths) which often find expression in applications using these models but which have not as yet been sufficiently discussed in the literature.

2.THE MYTHS

Several myths associated with frailty modeling exaggerate the ability of these models to address real problems. Some of them create an illusion that the problem is solved when, in reality, basic assumptions have made a major contribution in the results. Others misinterpret the notions of individual and shared frailty. Finally, some ignore confounding problems in the estimation of the parameters of the model. The uncritical use of such unjustified beliefs and concepts may inhibit the further development of this field and make it difficult to understand the proper place of frailty modeling in the analysis of survival data.

Myth 1. The underlying hazard $\mu_0(x)$ in model (1) can be approximated by a Gompertz-Makeham curve

The models represented in (1) have been used, for example, in demographic analyses of mortality in heterogeneous populations (Vaupel et al.

1979; Lancaster and Nickell, 1980; Heckman and Singer, 1984a, b; Hougaard, 1984; Manton et al., 1986; Vaupel and Yashin, 1985a; Bretagnolle and Huber-Carol, 1988; Yashin et al., 1994). To make such models identifiable one must fix the parametric structure of the underlying hazard and the frailty distribution and then use parametric methods of statistical analysis (e.g. the maximum likelihood method). This practice is so common nowadays that nobody even tries to explain why a particular form of $\mu_0(x)$ is chosen. In demographic applications, the Gompertz or Gompertz-Makeham mortality rate are often used as an approximation for $\mu_0(x)$ (e.g. $\mu_0(x)=ae^{bx}$ and $\tilde{\mu}(x)=c$). In epidemiological studies of cause-specific mortality as well as in econometric models of duration (Manton and Stellard, 1988), the Weibull parametrization of $\mu_0(x)$ is often used. Gompertz curve for the underlying hazard is used in testing heterogeneity models of experimental data (Service, 2000). However, such assumptions are neither biologically nor empirically justified. Moreover, the indirect semiparametric estimates of the underlying hazards obtained from twin data using a gamma frailty model show that an underlying hazard may increase faster than the Gompertz curve (Yashin and Iachine, 1997). There has been no study or survival experiment, which would allow us to make a direct estimate of $\mu_0(x)$ from failure-time data and to suggest its parametric form. Weiss (1990) emphasized the importance of bringing more biological background in frailty modeling. A substantial fraction of variation in frailty has an underlying genetic

basis, which can be used in specification of both the underlying hazard and frailty distributions.

Myth 2. The “amount of hidden heterogeneity” can be estimated from survival data

It is generally believed that the variance in frailty distribution estimated in the analysis of survival data characterizes the measure of heterogeneity present in this population. Since frailty models without observed covariates) are non-identifiable unless one assumes a functional form for the underlying hazard, the estimates of frailty distribution depend on the choice of a functional form for $\mu_0(x)$. Fig. 1 shows how Gompertz and gamma-Gompertz models fit mortality data for the cohort of Swedish females born in 1862.

Fig. 1 is about here

The gamma-Gompertz model assumes that $\mu_0(x)$ follows the Gompertz curve and that frailty is gamma-distributed. The estimate of the variance in a frailty distribution in this case is about 0.5. If, however, we assume that an underlying hazard is described by a logistic curve, the estimate of frailty variance becomes zero. This means that the biological interpretation and justification of conditions, which guarantees identifiability of respective frailty models, becomes a crucial issue when the questions concerning the quantitative aspects of hidden heterogeneity are addressed: two survival models with different

“degrees” of heterogeneity describe the same data equally well. An illusion that the “amount of heterogeneity” in the population can ultimately be estimated contradicts the fact that this “amount” is determined by the conditions of identifiability (Elbers and Ridder, 1982; Heckman and Singer, 1984b; Hoem 1990; Iachine and Yashin, 1998). Different identifiability conditions result in different estimates of frailty distribution and the underlying hazards from the same data.

More generally, the identifiability of frailty models is a matter of the balance between information contained in the data and assumptions put into the model. Survival data on unrelated individuals without observed covariates require assumptions about $\mu_0(x)$ and Z . Fig. 2 shows the graph of the logarithm of the underlying hazard as a function of age and variance σ_z^2 in the case a logistic marginal mortality rate.

Fig.2 is about here

One can see from this figure that the shape of the underlying hazard depends on the value of σ_z^2 , and vice versa. This graph is a nice illustration of the non-identifiability property of the proportional hazard frailty model without observed covariates.

In a proportional hazard model with covariates the assumptions about $\mu_0(x)$ and Z can be relaxed since they can be identified from the data non-parametrically. When analyzing data on related individuals using a multivariate

shared frailty model, no covariates are required for identifiability. In this case the conditional independence assumption plays a crucial role.

Thus, there is no absolute measure of heterogeneity in a population. The “amount of heterogeneity” measured in a statistical analysis of survival data is determined by identifiability conditions for the selected model.

Myth 3. The properties of individual frailty distributions can be analyzed using shared frailty models.

The multivariate frailty models are frequently used to analyze survival data on related individuals with different types of familial relationships, e.g. identical (MZ) or fraternal (DZ) twins, parents, children, grandchildren, etc. An important concept arising in connection with this type of statistical analysis is that of individual or marginal frailty. Let us illustrate this concept using the survival of MZ and DZ twins as an example.

A crucial condition exploited in all twin studies is that characteristics of MZ twin individuals are not different from those of DZ twins, and singletons i.e. non-twins. This assumption is supported in survival studies by empirical evidence (Christensen et al, 1995, Yashin and Iachine, 1995a). In particular, both types of twins have the same marginal mortality pattern $\bar{\mu}(x)$. Suppose that one wishes to develop a bivariate frailty model that could be applicable both to MZ and to DZ twins (which should, of course, be flexible enough to deal with the potential difference in the dependence structures of MZ and DZ twin survival times). We can say that the frailty variable used in this model is

individual in nature when it is attributable to a single twin individual and not necessarily to a pair of twins. More formally, for the frailty variable to be called individual, the model has to satisfy the following natural consistency condition: it should assign the same marginal frailty distribution to MZ twins as to DZ twins.

The third myth involves the belief that shared frailty models can be used to investigate the properties of individual frailty. Although such models are identifiable without additional covariate information (Iachine and Yashin, 1998), the notion of shared frailty must be distinguished from that of individual frailty as it is used in the correlated gamma-frailty model (Yashin et al., 1995) because shared frailty models do not satisfy the consistency condition specified above.

Versions of the shared gamma frailty model have been used successfully in the analysis of bivariate survival data. Hougaard et al. (1989) and Vaupel et al. (1992) applied the gamma-frailty model to prove that MZ twins share more longevity-related genetic material than DZ twins do. In the absence of observed covariates the conditional bivariate survival function in the case of this model is:

$$S(x_1, x_2 | Z) = e^{-Z(H_1(x_1) + H_2(x_2))} \quad (3)$$

where Z is a shared frailty, $H_i(x) = \int_0^x \mu_{0i}(u) du, i = 1, 2$, and $\mu_{0i}(x), i = 1, 2$ are the underlying hazards associated with two related individuals. Note that (3) assumes the conditional independence of life spans T_1 and T_2 given Z . When frailty is

gamma-distributed with the mean 1 and variance σ^2 , the marginal bivariate survival function is (Cox and Oakes, 1984):

$$S(x_1, x_2) = \left(S_1(x_1)^{-\sigma^2} + S_2(x_2)^{-\sigma^2} - 1 \right)^{-\frac{1}{\sigma^2}} \quad (4)$$

Here $S_i(x), i = 1, 2$ are marginal univariate survival functions. One can see that the only parameter responsible for the association between T_1 and T_2 is σ^2 .

This model has several remarkable properties. It has a semiparametric copula structure, which allows us to take into account a difference in univariate survival functions of related individuals from different generations, as well as individuals of different sex. This model is identifiable without our having to make assumptions about the parametric specification of the conditional survival function: this function and the parameters of other models can be identified from bivariate survival data (Iachine and Yashin, 1998). When using this model, one can easily establish the difference in association parameters, σ^2 , for populations of pairs of individuals with different degrees of familial relationship (e.g. MZ and DZ twins, Vaupel et al., 1992). However, since in the shared frailty model the marginal distribution of frailty coincides with the distribution of the shared frailty itself, the marginal frailty distributions for MZ and DZ twins will be different. Thus, shared frailty does not characterize individual frailty.

Note that the application of the shared gamma-frailty model to studies of genetic aspects of aging and survival is subject to certain limitations. In particular, although characteristics of this model can always be estimated from bivariate

survival data, the genetic interpretation of these characteristics is often problematic. For example, the large-sample estimates of parameter σ^2 will always be different for MZ and DZ twins, ($\sigma_{MZ}^2 > \sigma_{DZ}^2$), which means that these twins share different amounts of genetic material. However, without additional assumptions this finding is only of limited use for evaluating of the influence of genes on human mortality and longevity. Thus, the shared frailty model cannot be used for evaluating properties of individual frailty.

Myth 4. The association parameter in shared gamma-frailty models with observed covariates can be estimated from univariate data.

The fourth myth involves the belief that, in shared frailty models with finite mean of a frailty distribution and observed covariates, an association parameter (which in the case of gamma-frailty coincides with the variance of the frailty distribution) can be estimated from univariate data. This belief is based on a result of Elbers and Ridder (1982): if a frailty distribution has a finite mean, then the presence of observed covariates (with the covariate effects provided by model (2)) makes the univariate proportional hazard frailty model identifiable. The source of confusion is equation (2), which specifies the same functional form for the hazards used in univariate and shared frailty models. However, “shared frailty” in bivariate survival models differs from “individual frailty” as it is used in the univariate case. Initially, this difference in the different notions of frailty was not clearly understood. Clayton and Cuzik (1985) notice that in case of bivariate proportional hazards model with gamma-

distributed shared frailty and observed covariate "the extent to which interpretation of σ^2 is confounded by population heterogeneity, which is not shared by the two coordinates is in need for further studies". Wei et al. (1989) found, that in this model "the dependence parameter and the regression parameter are confounded". Hougaard pointed out a strange property of this model in the discussion of Clayton and Cuzick (1985). He claims that "the joint distribution of (T_1, T_2) can be identified from the marginal distributions and thus the association parameter σ^2 describes more than just association". Similar statements about the opportunity to identify the variance of frailty in bivariate survival model from univariate data have been repeatedly cited in biometrical and statistical literature (e.g. Hougaard 1986; 1987; 1989; Klein 1992 (p.102); Hougaard et al. 1989; Andersen et al. (p.674); Pickles et al. 1994). It is worth noting that this conclusion is only true when the shared frailty model is indeed a correct model for the data. In general, however, when the data is generated by a different mechanism, the value of σ^2 estimated from univariate data may, in fact, have nothing to do with a measure of the association between life spans. Indeed, in accordance with this statement one expects to identify different association parameters for MZ and DZ twins analyzing respective univariate life-span distributions. The identifiability property (Elbers and Ridder, 1982) implies, however, that univariate analysis gives the same values of a variance of a frailty distribution σ^2 for MZ and DZ twins who presumably have the same univariate life-spans distributions given observed covariates. Similar situation

corresponds to survival data for a group of women and their daughters and for the same group of women and their granddaughters. In both situations the value of parameter σ^2 estimated from the data on mothers will be the same despite the fact that the life span associations between related individuals in these two data sets are different.

The reason for this contradiction lies in what might be called the “overidentifiability” of the shared frailty models with finite mean and covariate effects provided by (2). Given the large amount of information present in the data (bivariate observations with observed covariate information), these models impose too many assumptions on the structure of the survival time distribution (proportional hazard covariates, conditional independence, finite mean of frailty), thereby permitting the estimation of the “association” parameter from univariate data. Thus, there is nothing wrong with the shared frailty model itself – it is just that “overidentifiable” models easily become misspecified when applied to real data. As a result, an uncritical interpretation of the results may lead to misleading conclusions.

Myth 5. The shared gamma-frailty model with observed covariates is a convenient tool for the analysis of twin data

This myth has to do with the following version of the shared frailty model:

$$S(x_1, x_2 | Z, U_1, U_2) = e^{-Ze^{\beta U_1} H(x_1)} e^{-Ze^{\beta U_2} H(x_2)} \quad (5)$$

Here Z is the shared frailty variable, which has a finite mean, $H(x)$ is the cumulative hazard function as in (3), and β is the regression coefficient describing the influence of the covariates U_1, U_2 on the underlying hazard. The variance of shared frailty in model (5) is often used as a measure of dependence between the two survival times. Clearly, when the frailty variance is zero, the survival times are conditionally independent given the covariates. On the other hand, conditional variance of shared frailty induces dependence between survival times (Yashin and Iachine, 1999). So, why might it be wrong to use model (5) to analyze the degree of dependence between individuals who are related to each other in various different ways? Again, let us consider identical (MZ) and fraternal (DZ) twins as an illustrative example.

Note that univariate sub-models of model (5) have the following form:

$$S(x|Z, U) = e^{-Ze^{\beta U} H(x)} \quad (6)$$

which clearly coincides with model (2). It is crucial that this model be identifiable when the frailty variable has a finite mean (Elbers and Ridder, 1982). It is thus possible to apply model (6) to data on MZ and DZ twins, treating them as singletons (i.e. as unrelated individuals), and to recover the parameters of the models separately for MZ and DZ twins. The question is whether the parameters for MZ twins are similar to those for DZ twins?

On the one hand, MZ and DZ twins are considered univariately equivalent in twin studies, and model (6) can be viewed as a straightforward univariate

model for the entire general population, including the twins taken separately. In this case univariate analysis of MZ and DZ twin data should yield similar results.

On the other hand, model (6) is an identifiable univariate submodel of the shared frailty model (5) and should, if the shared frailty model is a correct model for the data, yield consistent estimates of the parameters of the shared frailty model itself since they coincide with the parameters of the univariate model. But if variance in shared frailty or the distribution of shared frailty in general is a measure of the dependence between the twins' life spans, such an analysis should yield different frailty distributions for MZ and DZ twins, assuming that MZ twins exhibit a different association between their survival times than DZ twins.

The reason for this apparent contradiction is quite simple – a shared frailty model with finite mean can never be a correct model for both MZ and DZ twins in the presence of observed Cox-type covariates. Fig.3 shows the graphs of the estimates of the correlation of life spans, the standard deviation of shared frailty σ , and the regression coefficient β as functions of correlation coefficient of frailty ρ_z , obtained using a shared frailty model with observed covariates.

Fig.3 is about here

The data sets are generated using a correlated frailty model with observed covariates for different values of ρ_z . The estimates of σ and β are obtained using a shared frailty model with observed covariates. The correlation of life spans

was estimated directly from the data. One can see that for $\rho_z=0$, i.e., for the independent individuals the estimate of σ (which is supposed to characterize an association between life spans along the lines of a shared frailty model) is about 0.55. It is clear that this association is spurious and that it does not characterize the dependence between the respective life spans.

As was already noted above, in the presence of observed Cox-type covariates the shared gamma-frailty model fails to satisfy some natural assumptions. For example, in the analysis of several bivariate data sets (e.g. for MZ and DZ twins) one cannot assume that the marginal univariate survival functions are the same, since in this case all parameters of bivariate models should also be the same, including the association parameters. The reason for this contradiction is that shared frailty does not have an individual interpretation. This property limits the possibility of a joint analysis of MZ and DZ twin data, e.g. in the evaluation of genetic characteristics of frailty and longevity using methods of survival analysis and quantitative genetics (Yashin and Iachine, 1995a,b), where the individual interpretation of frailty plays a crucial role.

Moreover, some additional limitations of the shared-gamma-frailty model become evident in the presence of observed covariates. If, for example, the true conditional hazards are:

$$\mu_i(Z, U_i, x) = Z\mu_{0i}(x)e^{\beta_i U_i}, \quad i = 1, 2 \quad (7)$$

with U_i and β_i , $i=1,2$ specified as observed covariates and their respective regression coefficients, then the bivariate survival function given observed covariates is:

$$S(x_1, x_2 | U_1, U_2) = \left(S_1(x_1 | U_1)^{-\sigma^2} + S_2(x_2 | U_2)^{-\sigma^2} - 1 \right)^{\frac{1}{\sigma^2}} \quad (8)$$

where $S_i(x | U) = \left(1 + \sigma^2 H_i(x) e^{\beta_i U} \right)^{-\frac{1}{\sigma^2}}$. In this model the estimates of the association parameter σ^2 and the regression coefficients β_i in the Cox-like hazards with random effects are always confounded if data on related individuals are used. This property has important methodological consequences for the statistical analysis of bivariate data.

There have been several attempts to fix the “overidentifiability” problem of the shared frailty models. In short, they were driven by the desire to get rid of the identifiability property of the univariate model (2) while keeping the bivariate model identifiable. This can be done either by using a frailty distribution with infinite mean (Hougaard, 1987) or by relaxing the proportionality assumption for the covariate effects.

Myth 6. The use of “positive stable” shared frailty models allows us to avoid a confounding problem.

The sixth myth is associated with the attempt to avoid a confounding problem by using a positive stable distribution (Hougaard 1984) of shared frailty instead of gamma frailty. Although this replacement helps us to avoid problems induced by the identifiability of the univariate gamma-frailty model

(the univariate positive stable-frailty model with observed covariates is non-identifiable) it still results in the confounding of association parameter and regression coefficient in the Cox-type hazard, as the following analysis shows.

The positive stable distribution with the parameter $0 < \alpha < 1$ is characterized by its Laplace transform $L(s) = e^{-s^\alpha}$. So, if frailty Z is positive-stable distributed, the univariate marginal survival function is:

$$S(x) = \exp(-H(x)^\alpha) \quad (9)$$

where $H(x) = \int_0^x Q(u)du$. In the case of shared frailty the bivariate marginal survival function is (Hougaard 1987):

$$S(x_1, x_2) = \exp \left\{ - \left((-\ln S_1(x_1))^{\frac{1}{\alpha}} + (-\ln S_2(x_2))^{\frac{1}{\alpha}} \right)^\alpha \right\} \quad (10)$$

When covariates $U_i, i=1,2$ are observed, and the conditional hazards are given by (5), then the univariate survival functions conditional on observed covariates U_i are:

$$S_i(x|U) = \exp(-H_i(x)^\alpha e^{\alpha\beta_i U}) = \exp(-\tilde{H}_i(x)e^{\tilde{\beta}_i U}) \quad (11)$$

where $\tilde{\beta}_i = \alpha\beta_i$, and $\tilde{H}_i(x) = H_i(x)^\alpha$. A remarkable property of this model is that the proportional hazards structure of the hazard with the Cox-type regression term is preserved after integration with respect to unobserved frailty. Since $|\tilde{\beta}| < |\beta|$, the effect of hidden frailty (omitted covariates) on the regression coefficient is obvious, and its value is measured by the value of parameter α .

Using a positive stable shared frailty model one can estimate both an association parameter α and regression coefficients β_i from the data on related individuals. One problem, however, remains unsolved: the interpretation of regression coefficients β_i . To illustrate this problem let us assume first that $\beta = \beta_1 = \beta_2$ is the same for both members of MZ and DZ twin-pairs. Let us consider two hypothetical bivariate data sets: one for MZ twins, the other for DZ twins. It is clear that the values of parameter α estimated for MZ and DZ twins will be different since they characterize the associations between the respective life spans. Hence the values of parameter $\tilde{\beta} = \alpha\beta$ and $\tilde{H}(x)$ in equation (9) are also different for the members of MZ and DZ pairs, which contradicts the natural assumption that the survival of these individuals conditioned on observed covariates follows the same (Cox-type) hazards. If we assume that parameters $\tilde{\beta} = \alpha\beta$ are the same for members of MZ and DZ pairs then parameters β and the underlying hazards $\mu_0(x)$ should be different for these individuals.

The values of β are frequently interpreted as the “strength” of the covariates’ influence on survival. So, if the estimates of β are different for MZ and DZ twins, does it mean that the same covariate has a different degree of influence on these two types of individuals? Not necessarily, since we are considering the effects of the covariate on the shared frailty level, which are different for MZ and DZ twins. Indeed, it does not make any sense to compare

the estimates of β for MZ and DZ twins, since we only observe the effect of confounding modulated by the differences in α_{MZ} and α_{DZ} . This is not a characteristic feature of the positive stable frailty model alone. The difference in the regression coefficient estimated for MZ and DZ twins will be obtained in other shared frailty models (i.e. gamma) when analyzing data with the same marginal hazards but different levels of association.

Thus, confounding estimates of association parameter and regression coefficient is a common feature of the shared frailty models with observed covariates.

3. CORRELATED FRAILTY MODELS

To avoid the methodological problems associated with shared frailty models, both with finite and infinite mean of a frailty distribution, the idea of correlated frailty has been suggested (Yashin and Iachine, 1994; Yashin et al., 1995). In this model one more parameter – the correlation coefficient of frailty ρ – is introduced to describe the association between life spans of related individuals. In the presence of observed covariates such a model is internally consistent: if bivariate data correspond to the correlated frailty model, then the large-sample parameter estimates obtained from univariate and bivariate data coincide. This model is convenient for the analysis of genetic aspects of susceptibility to death and longevity since the estimates of heritability in frailty have a one-to-one correspondence to the estimates of correlation coefficients of

frailty for MZ and DZ twins. The correlated gamma-frailty model is characterized by the bivariate survival function

$$S(x_1, x_2) = \frac{S(x_1)^{1-\rho} S(x_2)^{1-\rho}}{(S(x_1)^{-\sigma^2} + S(x_2)^{-\sigma^2} - 1)^{\rho/\sigma^2}} \quad (12)$$

where σ^2 and ρ are the variance and correlation coefficient of bivariate frailty distribution, respectively, and $S(x) = P(T > x)$ is the univariate survival function (Yashin et al., 1995). A statistically significant difference in correlation coefficients of frailty estimated for MZ and DZ twins, ρ_{MZ} and ρ_{DZ} , may indicate the presence of a genetic influence on susceptibility to death. When such an influence is established, the genetic characteristics of frailty (e.g. a narrow-sense heritability) can be estimated directly from the data. For this purpose the genetic model of frailty has to be incorporated into a survival model (Yashin and Iachine 1995a,b).

The use of the correlated frailty models allows us to estimate the age pattern of an underlying hazard without making preliminary assumptions about its functional form. Fig. 4 shows in the logarithmic scale the graphs of $\mu_0(x)$ and $\bar{\mu}(x)$ obtained in the studies of Danish, Finnish, and Swedish twins (Iachine et al., 1998).

Fig. 4 is about here.

One can see that in all six cases (two sexes and three countries) the underlying hazards increase faster than a Gompertz hazard (see Myth 1).

Genetic models of frailty. Here we consider the application of the correlated frailty model to the statistical analysis of genetic factors influencing the survival of MZ and DZ twins. In particular, we calculate semiparametric estimates for the six *genetic models of frailty*, which correspond to six different assumptions about its structure. We refer to the notations used in McGue et al. (1993) and in Yashin and Iachine (1994) for such models. To be more specific, let A , D , I , C , E , and H be the five components of frailty which represent additive genetic effects, dominant genetic effects, epistatic genetic effects, common environmental effects, uncommon environmental effects, and total genetic effects, respectively, in the additive decomposition of frailty. From the estimation point of view not more than three components should be represented in the model when MZ and DZ data are analyzed (more components can be considered if, in addition, data about adopted children and twins reared apart are available). In these notations an ACE model refers to the decomposition of frailty $Z = A + C + E$. An AE model refers to the decomposition $Z = A + E$. ADE, DE, DCE, and HE models are defined similarly. We use small letters a^2 , d^2 , i^2 , c^2 , e^2 to refer to the respective proportions of variance. For example, the relationship

$$1 = a^2 + c^2 + e^2$$

corresponds to the decomposition of variance in the ACE model of frailty. Yashin and Iachine (1994) use the assumption that the correlation coefficient of the bivariate frailty distribution for this model admits decomposition:

$$\rho = \rho_1 a^2 + \rho_4 c^2 + \rho_5 e^2$$

where a^2 is the proportion of the variance associated with additive genetic effects, called narrow-sense heritability; c^2 is the proportion of variance associated with shared environmental factors; e^2 is the proportion of variance with non-shared environmental factors; and ρ_1 , ρ_4 , and ρ_5 are correlations between additive genetic, shared environmental, and non-shared environmental components for related individuals. Similarly, ρ_2 and ρ_3 are correlation coefficients between dominant genetic and epistatic genetic components, respectively. H^2 is used for the broad-sense heritability coefficient $H^2 = a^2 + d^2 + i^2$. The model for broad-sense heritability includes a genetic component, H, a common environmental component, C, and an independent environmental, E (the HCE-model). The respective equations for proportions of variance and correlation coefficient are:

$$1 = H^2 + c^2 + e^2$$

$$\rho = RH^2 + \rho_4 c^2 + \rho_5 e^2$$

where R is the correlation coefficient between the total genetic components of the phenotype. We use a semiparametric representation for the marginal bivariate survival function with respective decompositions of the correlation coefficients of frailty. Standard assumptions of quantitative genetics specify different values of ρ_i ($i = 1, 2, \dots, 5$) and R for MZ and DZ twins. For MZ twins

$\rho_i = 1$ ($i = 1, 2, 3, 4$), $\rho_5 = 0$, $R = 1$. For DZ twins $\rho_i = 0.5$, $\rho_2 = 0.25$, $\rho_3 = m$, $\rho_4 = 1$, $\rho_5 = 0$, $R = k$. Here $0 \leq m \leq 0.25$ and $0 \leq k \leq 0.5$ are unknown parameters. Another important assumption is that the variances of the phenotypic traits for MZ and DZ twins are the same. This hypothesis was tested and confirmed using Danish twin data for both sexes (Yashin and Iachine 1995).

Correlated non-gamma-frailty models. To test how the results of such analyses depend on the type of frailty distribution, we compare the results of a statistical analysis of Danish twin data using three bivariate survival models: the gamma-, inverse Gaussian, and general “three parameter” correlated frailty models (Yashin et al., 1998). The latter model is based on a bivariate extension of “the three-parameter” distribution of frailty $P(z; \alpha, \delta, \theta)$ (with parameters α, δ, θ), which was introduced in a univariate survival analysis by Hougaard (1984).

The most general case corresponds to the correlated frailty model with three-parameter frailty distribution (Yashin et al., 1999). This model is characterized by the bivariate survival function of the form

$$S(x_1, x_2) = S(x_1)^{1-\rho} S(x_2)^{1-\rho} \times \exp \left\{ \frac{\rho(1-\alpha)}{\alpha\sigma^2} \left[1 - \left(\left(1 - \frac{\alpha\sigma^2}{1-\alpha} \ln S(x_1) \right)^{\frac{1}{\alpha}} + \left(1 - \frac{\alpha\sigma^2}{1-\alpha} \ln S(x_2) \right)^{\frac{1}{\alpha}} - 1 \right)^\alpha \right] \right\} \quad (13)$$

The gamma-frailty model (12) can be obtained from (13) by allowing α to approach zero. The inverse Gaussian correlated frailty model is characterized by a marginal bivariate survival function obtained from (13) by taking $\alpha=0.5$:

$$S(x_1, x_2) = S(x_1)^{1-\rho} S(x_2)^{1-\rho} \exp \left\{ \frac{\rho}{\sigma^2} (1 - ((1 - \sigma^2 \ln S(x_1))^2 + (1 - \sigma^2 \ln S(x_2))^2 - 1)^{1/2}) \right\} \quad (14)$$

The marginal univariate survival functions were approximated as

$$S(x) = \left[1 + s^2 \left(a(x - x_0) + \frac{b}{c} (e^{cx} - e^{cx_0}) \right) \right]^{-\frac{1}{s^2}}, \quad x_0=30 \quad (15)$$

The parameters σ^2, ρ in equations (13) and (14) are the variance and the correlation coefficient of the bivariate frailty distributions. Parameter α in (13) characterizes the three-parameter distribution.

Although no significant difference was detected between the general three-parameter model and its special cases for this particular data set, the ability to carry out such comparisons in the case of general data on duration presents an important advantage over previous approaches based on the use of gamma-distributed frailty alone. The application of the three parameter correlated frailty model may serve as a convenient goodness-of-fit assessment procedure for the less complicated models from this family (Yashin et al 1999).

Frailty and liability models. The notion of liability was first used in quantitative genetics. According to Falconer (1990) liability Z is a standard, normally distributed random variable, which is related to the discontinuous trait Y by a

threshold. For example, the trait may be associated with the absence ($Y=0$) or presence ($Y=1$) of a disease. In this case if $Z < a$, where a is the threshold value, then $Y=0$, otherwise $Y=1$. Later, this definition was adjusted to describe more sophisticated liability-trait relationships. For example, Eaves (1978) and Kendler and Eaves (1988) considered the conditional probability of disease as a logistic function of liability. Models of more complicated quantitative traits, including duration, have been also suggested and analyzed using multivariate survival data (Meyer and Eaves 1988, Meyer et al., 1991). In these models, multidimensional normal distribution is traditionally used to describe the association between liabilities of related individuals. Contrary to the distribution of liability (which is always assumed to be normal), the form of the conditional survival function depends on the particular problem one is analyzing. For example, Meyer et al. (1991) use gamma density, with a scale parameter as a function of liability, to specify the conditional distribution of the age at menarche. For a given value of liability z_i , $i=1,2$ this density distribution function is:

$$f_i(t|z_i) = \frac{\mu_i(\mu_i t)^{\gamma-1} e^{-\mu_i t}}{\Gamma(\gamma)}, i = 1,2$$

with $\mu_i = e^{\alpha + \beta z_i}$, i.e., duration is conditionally gamma distributed. Under the assumption of conditional independence, the joint marginal density distribution function for the age at menarche for two related individuals can be obtained by

averaging the product $f_1(t_1|z_1)f_2(t_2|z_2)$ with respect to the bivariate normal distribution of z_1, z_2 .

Liability models provide a convenient methodological framework for studying genetic and environmental aspects of survival. They can describe multiple correlations, which is important in the analysis of multivariate (pedigree) data. Correlation coefficients of liability are among the parameters of these models. Once estimated, they can be useful in further genetic studies of durations. Standard methods of quantitative genetics can infer genetic characteristics, such as heritability, from correlations (Falconer, 1990). Note that the association parameters in the shared frailty models discussed below do not have such a natural connection with quantitative genetics. Unfortunately, some technical limitations restrict the broader use of this important family of random effect models in survival studies.

One such limitation is the need to provide a parametric specification of the univariate conditional survival functions given liability. Such parametrization is often difficult to justify biologically. In the case of bivariate frailty models, this specification is not necessary, since the function can be estimated semiparametrically (Yashin et al., 1995). Thus, the use of frailty- instead of liability models in bivariate survival studies allows us to avoid one hardly justifiable technical assumption.

Although widely used in multivariate studies of quantitative genetics and genetic epidemiology, liability methods have not yet become popular in demography and biostatistics. In contrast, frailty (heterogeneity) modeling has a relatively long tradition in demographic and survival studies. Its contribution to the research area is recognized, and the methodology is widely used. This approach provides a better fit to the univariate survival data than traditional methods that do not take unobserved heterogeneity in frailty into account. For this reason the demographic interpretation of bivariate findings can be easier when frailty models of multivariate survival are used.

In the case of traditional liability models, the likelihood function of survival data is not represented in a closed form. It is often the case that an averaging with respect to unobserved liability cannot be achieved analytically. This means that the maximum likelihood procedure may require significant computational efforts and can be accompanied by undesirable complications. Such efforts are not required in the case of the frailty models discussed here: the likelihood functions of the data can be written in closed analytical forms. For liability, this problem can be solved by combining the best features of liability and frailty modeling in a new approach suggested by Yashin and Iachine (1996). Finally, frailty models usually exploit the proportional hazards assumption. Despite the fact that additional efforts are often needed to justify its use, this assumption is still very popular in epidemiology and biostatistics when analyzing the effects of covariates on survival. Therefore, it seems easier to compare, say,

the results of bivariate and traditional epidemiological or biostatistical univariate survival studies when frailty models are used. Frailty models are convenient for the evaluation of longevity limits. The additive decomposition of individual frailty, which is used in genetic models, generates a competing risks structure in the survival models, where respective risks are associated with the genetic and environmental components of mortality. The survival function associated with the genetic component can be further used in the evaluation of the biological limits of human longevity (Yashin and Iachine 1995a).

The concept of liability is used in genetic epidemiology and quantitative genetics to describe unobserved susceptibility to disease and death. It provides a convenient methodological framework for studying genetic and environmental aspects of survival. Liability models have a capacity to describe multiple correlations, which is important in the genetic analysis of multivariate (pedigree) data. The correlation coefficients of liability play the key role in the estimation of heritability indices, i.e., parameters which characterize the contribution of genes to the variability in a respective trait (Falconer, 1990). Note that the parameters in the shared frailty models do not have a natural connection with the parameters used in quantitative genetics. Unfortunately, some technical and methodological limitations restrict the broader use of this important family of random effect models in survival studies.

Not all frailty models are appropriate for addressing the problems involved in the genetic analysis of durations. Multivariate frailty models have additional

constraints on correlation coefficients (Yashin and Iachine 1999b), which restricts their use in the analysis of pedigree data. The idea of quadratic hazards (or: of a quadratic hazard) combines positive features of the liability and correlated frailty models. As a result, one can assume different correlations of frailty variables for different pairs of related individuals in pedigree data. One can assume dynamic changes in the frailty variable with age. One can also introduce dependence between observed and unobserved covariates into the survival model and test hypotheses about this dependence. In addition, the quadratic hazard model preserves the quadratic structure during the averaging with respect to unobserved (normally distributed) covariates. It can also be regarded as a second-order approximation of the nonlinear conditional hazard rate. Thus, quadratic hazard models combine the multivariate flexibility of the liability models with the analytical power of frailty modeling.

4. DISCUSSION

In this paper we have presented several examples of how an uncritical application of frailty modeling can lead to misleading and sometimes incorrect conclusions about the problem under study. These include an incorrect assessment and interpretation of the heterogeneity distribution in the population, biased results on the strength of the association between survival times, and erroneously qualified significance of observed covariates for survival. It is clear that these problems are not merely the result of careless

application of “standard” frailty models to arbitrary survival data without worrying about the questions of fit or the appropriateness of the models used.

Such problems arise naturally when analyzing complex phenomena of hidden population heterogeneity. Obviously, when dealing with unobserved quantities some assumptions are necessary in order to relate the observed data with unobserved information. That is why the concept of model identifiability is so important in frailty modeling, which is a process of finding a delicate balance between the available data and the assumptions put into the model. On the one hand, it should be possible to recover the parameters of the model from the available data, so a number of assumptions to facilitate identifiability should be made. On the other hand, making too many assumptions may lead to paradoxes, such as the one we saw with the shared frailty model, where the association between the mothers and their daughters can be estimated without even collecting the data on the daughters.

Is the “overidentifiability” of some frailty models a problem – or is it a benefit for the statistician? For example, when applying a shared frailty model to bivariate survival data with covariates, one might compare the results of the bivariate analysis with those of the univariate analysis obtained using a univariate frailty submodel. If the results are different (if, for example, a different variance of frailty is estimated in the two analyses) then this is an indication that a shared frailty model is not a correct model for the data. This will be the case in all examples presented in this paper, which concern survival

data on relatives with different association levels (e.g., MZ and DZ twins, mothers, daughters and granddaughters). The trouble is that this simple procedure for testing consistency is rarely applied in the statistical analysis of data. The point here is that using a model “blindly”, as a “standard” tool for data analysis is the wrong way to go about things.

It is in the light of this important property, which might be called the “falsifiability” of a model, that we should view the historical attempts to extend the shared frailty model to get rid of the confounding problem. Simply making the univariate model non-identifiable (i.e. for the positive stable frailty distribution) corrects the confounding problem, but at what price? First, it becomes impossible to check for consistency in the shared frailty model and second, the stable frailty assumption effectively limits the modeling possibilities for the influence of covariates on survival to the case of a simple proportional hazard Cox-regression model. The role of frailty models as a tool for the correction of the non-proportionality of marginal hazards given the covariates is lost, too. This does not mean that frailty cannot follow a positive stable distribution. However, restricting ourselves exclusively to this class of distributions severely limits the flexibility of the marginal model.

There do exist more sensible approaches for solving the confounding problem of shared frailty models. One of them is based on extending the dimension of the frailty variable and making it equal the dimension of the survival data, i.e., introducing multivariate frailty. Yashin et al. (1995)

suggested a correlated frailty model which used an additive model of gamma frailties to define a bivariate frailty distribution with a given variance and correlation coefficient as parameters. The idea of this approach is to allow for the identifiability of a marginal frailty distribution from univariate data, while limiting the identifiability of association parameters (i.e. the correlation coefficient) to the case of multivariate data analysis. In this case, the pairs of individuals with different levels of association between life spans will be distinguished by the fact that respective correlation coefficients are different. An additional advantage is that components of such a multivariate frailty have an individual interpretation that allows for interesting applications, e.g., genetic analysis of frailty by using the methods of quantitative genetics.

Another approach, which one could call “copula modeling”, is based on relaxing the proportional hazard assumption for the covariate effects, i.e. an arbitrary dependence of the underlying hazard on the covariates is assumed. Using a parametric or a non-parametric approach, an appropriate parametrization of the univariate survival functions given the covariates is obtained. Then, this parametrization is substituted into the so-called semiparametric or “copula” representation of a particular frailty model. In the case of the shared gamma frailty model, this representation takes the form of (8) and allows for the estimation of the association parameter from bivariate data. Here, the confounding problem does not exist, since the result of Elbers and Ridder (1982) does not apply to the arbitrary dependence of the underlying

hazard on survival and the univariate submodel is not identifiable. The representation for the correlated gamma frailty is given by (12).

Since the shared frailty model is nested within the correlated frailty model, it is always possible to test for whether or not the shared frailty model is appropriate for the data. It seems, however, that the concept of individual frailty, which is used in the multivariate frailty models (e.g. in the correlated gamma frailty model), provides a greater potential in most applications for building models with biologically justified constructs.

Legends to the figures

Fig. 1. Graphs of conditional probability of death for the Swedish female cohort born in 1862: empirical (solid line), estimated using the Gompertz-Makeham model (short dashed line), and estimated using a gamma-Makeham model (long dashed line).

Fig. 2. Graphs of the logarithm of the underlying hazard as a function of age and σ^2 in the case of a fixed logistic mortality rate.

Fig. 3. Graphs of the estimates of correlation of life span, standard deviation of shared frailty σ and the regression coefficient of frailty ρ_z . The data were generated using a correlated gamma frailty model with different values of ρ_z . The estimates of σ and β were obtained using a shared frailty model. The life-span correlation was estimated using a correlated frailty model.

Fig. 4. Graphs of marginal and underlying hazards for males and females calculated from data on three Scandinavian twin registers, using a correlated frailty model.

References

- Akaike, H. 1987. Factor Analysis and AIC. *Psychometrika*. 52 (3):317-322.
- Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding 1993. *Statistical Models Based on Counting Processes*. Berlin: Springer.
- Bretagnolle, J., Huber-Carol, C. Effects of omitting covariates in Cox's model for survival data. *Scandinavian Journal of Statistics* 1988,15:125-38.
- Begun A.Z., Iachine, A.Z and A.I. Yashin 2000. Genetic Nature of Individual Frailty: Comparison of Two Approaches. *Twin Research*, 3, 51-57.
- Chamberlain, G. 1985. Heterogeneity, Omitted Variable Bias, and Duration Dependence. pp. 3-38 in *Longitudinal Analysis of Labour Market Data*. Eds. J. Heckman and B. Singer, Cambridge University Press.
- Clayton, D. G., and J. Cuzick. 1985. Multivariate Generalizations of the Proportional Hazards Model (with discussion). *Journal of Royal Statistical Society Ser. A*, 148:82-117.
- Clayton, D .G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 64(1):141-151.
- Christensen, K., J. Vaupel, N. V. Holm, A. I. Yashin. 1995. Mortality among twins after age 6: fetal origins hypothesis versus twin method. *British Medical Journal* 310:432-436.
- Eaves, L. J. 1987. Survival Analysis: Lessons from Quantitative Genetics. In *Evolution of Longevity in Animals*, A. D. Woodhead and K. H. Thompson, 339-343. New York: Plenum.
- Elbers, C. and Ridder, G. (1982) True and spurious Duration Dependence: The Identifiability of Proportional Hazard Model. *Review of Economic Studies* XLIX: 403-409.
- Falconer, D.S. (1990). *Introduction to Quantitative Genetics*, 3rd edition, New York: Longman Group Ltd. Assessment, and Prevention. *Journal of National Cancer Institute*, Vol. 88, No.8: 496-509.
- Gail, M. H., S. Wieand, and S. Piantadosi. 1984. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 71:431-444.
- Guo, G., and G. Rodriguez. 1992. Estimating a Multivariate Proportional Hazards Model for Clustered Data Using EM Algorithm, with an Application to Child Survival in Guatemala. *JASA* 87:969-976.

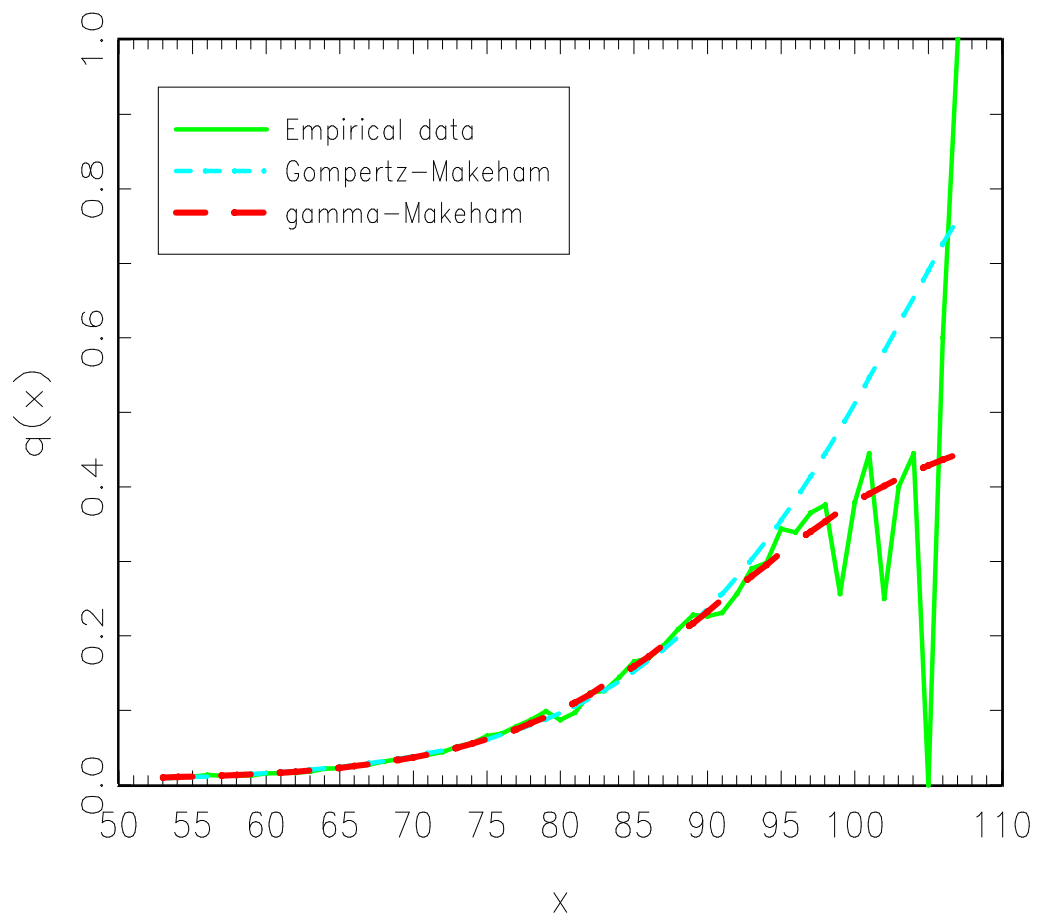
- Hauge, M. 1981. The Danish twin register. In "Prospective Longitudinal Research, An Empirical Basis for the Primary Prevention of Psychosocial Disorders", ed. S. A., Mednick, A. E. Baert, B. P. Backmann, 218-221. London: Oxford University Press.
- Heckman, J.J. and Singer, B. (1984b). The Identifiability of the Proportional Hazard Model. *Review of Economic Studies*:231-241.
- Heckman, J.J. and Singer, B. (1984b). A method for minimizing the impact of distributional assumptions in econometric models of duration data. *Econometrica*,71:435-44.
- Hoem, J.M. 1990. Identifiability of Hazard Models with Unobserved Heterogeneity: The Compatibility of Two Apparently Contradictory Results. *Theoretical Population Biology*, 37: 124-128.
- Hougaard, P. 1995. Frailty Models for Survival Data. *Lifetime Data Analysis* 1:255-273.
- Hougaard, P., B. Harvald, and N. V. Holm. 1992. Assessment of Dependence in the Life Times of Twins. In *Survival Analysis: State of the Art*, ed. G. P. Klein and P. K. Goel, 77-97. Printed in the Netherlands: Kluwer Academic Publishers.
- Hougaard, P. 1984. Life Table Methods for Heterogeneous Populations: Distributions Describing Heterogeneity. *Biometrika*, 71:75-83.
- Hougaard, P. Modeling multivariate survival. 1987. *Scandinavian Journal of Statistics* 14:291-304.
- Iachine, I. A., and A. I. Yashin. 1998. Identifiability of bivariate frailty models based on additive independent components. Research Report 8, Department of Statistics and Demography, Odense University.
- Iachine, I. A., N. V. Holm, J. R. Harris, A. Z. Begun, M. K. Iachina, M. Laitinen, J. Kaprio, and A. I. Yashin. 1998. How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finnish twins. *Twin Research* 1 (4):196-205.
- Kendler, K.S. and L.J.Eaves 1985. Models for the joint effect of Genotype and environment on liability to psychiatric illness. *The American Journal of Psychiatry*, 143,3:279-289.
- Korsholm, L. 1999. Some Semiparametric Models. Ph. D. thesis, Department of Theoretical Statistics, University of Aarhus, Denmark. Chapter 3, 93-109.
- Lancaster, T., Nickel, S. The analysis of re-employment probabilities for the unemployed. *Journal of the Royal Statistical Society, Series A* (1980) 143:141-65.
- Manton K. G., E. Stallard. and J. W. Vaupel. 1986. Alternative Models of Heterogeneity in Mortality Risks Among the Aged. *Journal of American Statistical Association* 81:635-644.
- Manton K.G. and A. I. Yashin 1999. *Mechanisms of Aging and Mortality: The Search for New Paradigms*. Odense monograph on Population Aging #7. Odense University Press. Odense Denmark.

- Manton K.G. and Stallard, E. (1988). Chronic Disease Modeling: Measurement and Evaluation of the Risks of Chronic Disease Processes. Charles Griffin Ltd. London, 1988.
- Meyer, J. M., and L. J. Eaves. 1988. Estimating Genetic Parameters of Survival Distributions: A Multifactorial Model. *Genetic Epidemiology* 5:265-275.
- Meyer, J. M., L. J. Eaves, A.C.Heath, N.G.Martin 1991. Estimating Genetic Influence on the Age at Menarche: A Survival Analysis Approach. *Am.J. Med. Genet.* 39:148-154
- Nielsen, G. G., R. D. Gill, P. K. Andersen, and T. I. A. Sørensen. 1992. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* 19 (1):25-43.
- Oakes, D. 1989. Bivariate Survival Models Induced by Frailties. *JASA*, 84, 487-493.
- Parner, E. 1998. Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics* 26 (1):183-214.
- Perera, F. P. 1996. Molecular Epidemiology: Insights Into Cancer Susceptibility, Risk Assessment, and Prevention. *Journal of National Cancer Institute* 88 (8):496-509.
- Petersen J.H., Andersen, P. K. and R.D.Gill (1996). Variance components for survival data. *Statistica Nederlandica*, **50(1)**, 193-196.
- Pickles, A. and R. Crouchley. 1995. A Comparison of Frailty Models for Multivariate Survival Data. *Statistics in Medicine* 14:1447-1461.
- Pickles, A., Crouchley, R. and Simonoff E. et al. Survival Methods for Developmental Genetic Data: Age-of-Onset of Puberty and Antisocial Behavior in Twins. *Genetic Epidemiology*. 11(1994); 155-170
- Pickles, A.(1994). Generalizations and applications of frailty models for survival and event data. *Statistical Methods in Medical Research* 3:263-278.
- Service, P.M. (2000). Heterogeneity in Individual Mortality Risk and its Importance for Evolutionary Studies of Senescence. *American Naturalist* 156,1:1-13.
- Vaupel, J. W., K. G. Manton, and E. Stallard. 1979. The Impact of Heterogeneity in Individual frailty on the Dynamics of Mortality. *Demography* 16:439-454.
- Vaupel, J.W. and Yashin, A.I. 1985. Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics. *American Statistician*, 39: 176-185.
- Vaupel, J. W. Inherited frailty and longevity. 1988. *Demography*; 25:277-287.
- Vaupel, J. W. 1991. Relative's Risk: Frailty Models of Life History data. *Theoretical Population Biology* 37 (1):220-234.
- Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., Iachine, I. A., Kannisto, V., Khazaeli, A. A., Liedo, P., Longo, V. D., Zeng, Y., Manton, K. G., and Curtsinger, J. W. (1998) Biodemographic trajectories of longevity. *Science*, 1998, 280: 855-860.
- Weiss, K.M. (1990). The Biodemography of Variation in Human Frailty. *Demography* vol. 27, no2:185-206.
- Xue, X., and R. Brookmeyer. 1996. Bivariate Frailty Models for the analysis of Multivariate Survival Time. *Lifetime Data Analysis* 2:277-289.

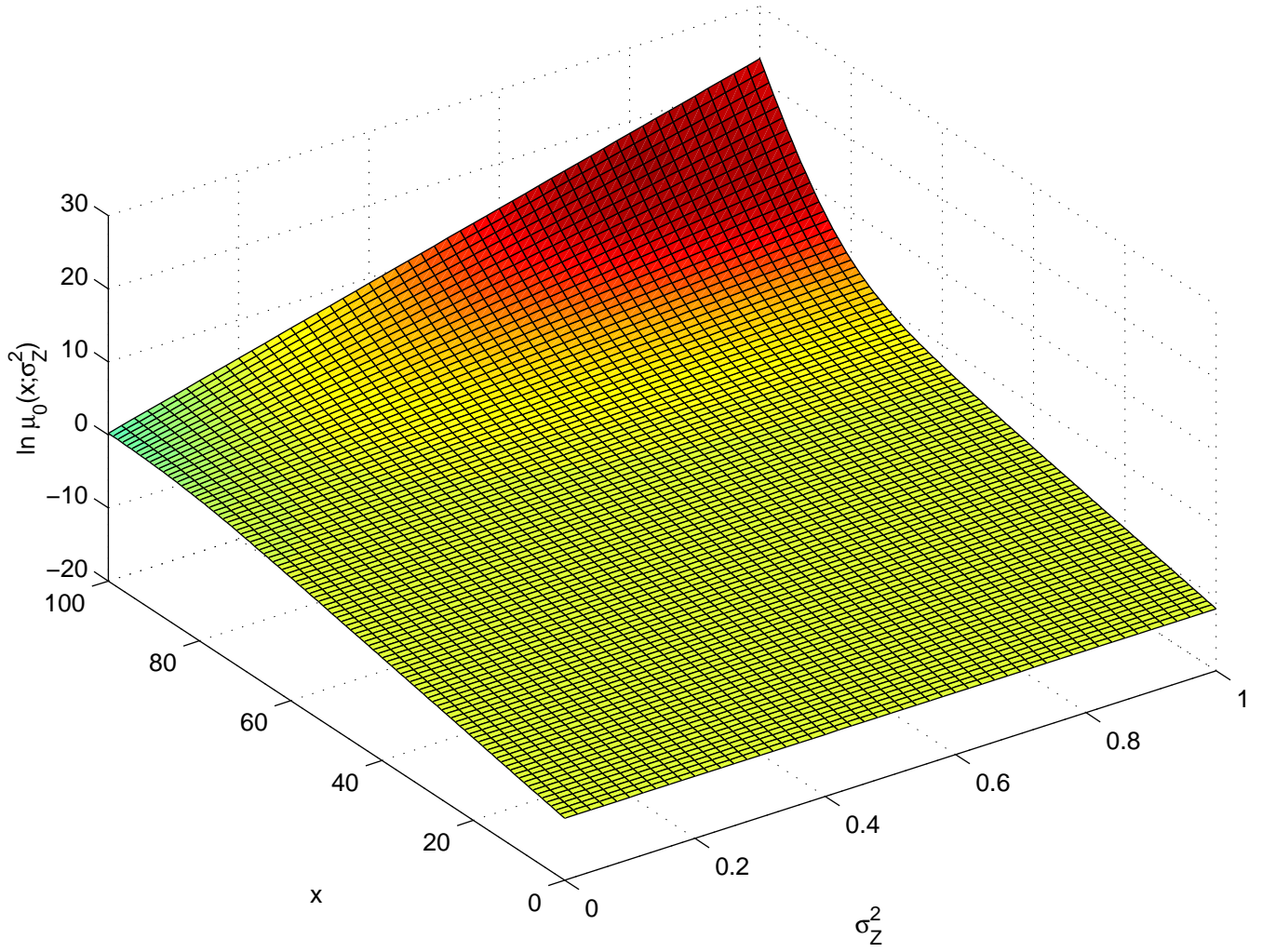
- Yashin, A.I., Vaupel, J.W., and I.A.Iachine 1993. Correlated Individual Frailty: An Advantageous Approach to Survival Analysis of Bivariate Data. RR #7. Population Studies of Aging. Centre for health and Social Policy Dec. 1993, Odense University, Denmark
- Yashin ,A.I. and I.A.Iachine 1994. Mortality models with application to twin survival data. In *CISS-First Joint Conference of International Simulation Societies Proceedings* (Halin J, Karplus W, and Rimane R., eds.). Zurich, Switzerland, pp. 567-571.
- Yashin, A. I., Vaupel, J. W. and I.A. Iachine 1994. A duality of aging: The equivalence of mortality models based on radically different concepts. *Mechanisms of Aging and Development*, 74:1-14.
- Yashin ,A.I. and I.A.Iachine 1995a. How long can humans live? Lower bound for biological limit of human longevity calculated from Danish twin data using correlated frailty model. *Mechanisms of Ageing and Development* 80:147-159.
- Yashin, A.I. and I.A.Iachine 1995b. Genetic analysis of durations: Correlated frailty model applied to survival of Danish Twins. *Genetic Epidemiology* 12:529-538.
- Yashin, A.I., Vaupel, J.W., and I.A.Iachine 1995. Correlated Individual Frailty: An Advantageous Approach to Survival Analysis of Bivariate Data. *Mathematical Population Studies* Vol. 5(2): 145-159.
- Yashin, A. I., Manton, K. G., Woodbury, M. A. and E. Stallard 1995. The Effects of Health Histories on Stochastic Process Models of Aging and Mortality. *Journal of Mathematical Biology*, 34:1-16.
- Yashin, A.I. and I.A.Iachine 1996. Random effect models of bivariate survival: Quadratic hazard as a new alternative. In *Symposium in Anvendt Statistik* (Kristensen G, ed.). Odense, Denmark: Odense University, pp. 87-101.
- Yashin, A.I., Andreev, K.F., Khazeli, A., Vaupel, J.W., and J.W.Curtsinger 1996a. Death-after-stress-data in the analysis of heterogeneous mortality. Transactions of Symposium I Anvendt Statistik, 22-24 January. Redigeret af Gustav Kristensen. Okonomisk Institut, Odense Universitet, pp. 24-35.
- Yashin, A.I., Manton, K.G. and I.A.Iachine 1996b. Genetic and environmental factors in duration studies. Multivariate frailty models and estimation strategies. *Journal of Epidemiology and Biostatistics* 1(2): 115-120.
- Yashin A.I. and K.G.Manton 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.
- Yashin, A. I. and I. A. Iachine 1997. How Frailty Models Can be Used in the Analysis of Mortality And Longevity Limits. *Demography* 34:31-48.
- Yashin A. I., A. Z. Begun, I. A. Iachine. 1999a. Genetic factors in susceptibility to death: a comparative analysis of bivariate survival models. *Journal of Epidemiology and Biostatistics* 4 (1):53-60.

- Yashin, A.I. I.A. Iachine and J. R. Harris 1999b. Half of Variation in Susceptibility to Mortality is Genetic: Findings from Swedish Twin Survival Data. *Behavior Genetics*, Vol. 29, No.: 11-20.
- Yashin, A.I., A.Z. Begun and I.A. Iachine 1999. Genetic Factors in Susceptibility to Death: Comparative Analysis of Bivariate Survival Models. *Journal of Epidemiology and Biostatistics*, Vol.4, No 1: 53-60.
- Yashin, A.I., I. A. Iachine and J. R. Harris 1999c. Half of Variation in Susceptibility to Mortality is Genetic: Findings from Swedish Twin Survival Data. *Behavior Genetics*, Vol. 29, No.: 11-20.
- Yashin, A.I. G. De Benedictis, J.W. Vaupel, Q. Tan, K.F. Andreev, I.A. Iachine, M. Bonafe, M. De Luca, S. Valensin, L. Carotenuto, and C. Franceschi 1999d. Gene, Demography and Life Span: The Contribution of Demographic Data in Genetic Studies of Aging and Longevity. *American Journal of Human Genetics*, 65: 1178-1193
- Yashin, A.I. and I.A. Iachine (1999a) What Difference Does the Dependence Between Durations Make? *Life Time Data Analysis*, 5: 5-22.
- Yashin, A.I. and I. A. Iachine (1999b). Dependent Hazards in the Problem of Multivariate Survival. *Journal of Multivariate Analysis*, 71:241-261.
- Yashin, A.I. G. De Benedictis, J.W. Vaupel, Q. Tan, K.F. Andreev, I.A. Iachine, M. Bonafe, M. De Luca, S. Valensin, L. Carotenuto, and C. Franceschi 2000. Genes and Longevity: Lessons from Studies of Centenarians. *Journal of Gerontology. Biological Sciences*. 55A. B1-B10.

Sweden (Lund.) Females Cohort:1862



Underlying hazard trajectories for logistic marginal mortality



Confounding in shared gamma frailty model
True model: correlated gamma frailty, $\sigma=2$, $\beta=8$

