Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
http://www.demogr.mpg.de

Unobserved heterogeneity
in a model with cure fraction
applied to breast cancer

Andreas Wienke (wienke@demogr.mpg.de)
Paul Lichtenstein (paul.lichtenstein@mep.ki.se)
Anatoli I.Yashin (yashin@demogr.mpg.de)

Abstract

We suggest a cure-mixture model to analyze bivariate time-to-event data, as motivated by the paper of Chatterjee and Shih (2001, *Biometrics* **57**, 779 - 786), but with a simpler estimation procedure and the correlated gamma-frailty model instead of the shared gamma-frailty model. This approach allows us to deal with left truncated and right censored lifetime data and accounts for heterogeneity as well as for an insusceptible (cure) fraction in the study population. We perform a simulation study to evaluate the properties of the estimates in the proposed model and apply it to breast cancer incidence data for 5,857 Swedish female monozygotic and dizygotic twin pairs from the so-called old cohort of the Swedish Twin Registry. This model is used to estimate the size of the susceptible fraction and the correlation between the frailties of the twin partners. Possible extensions, advantages and limitations of the proposed method are discussed.

Key words: survival analysis, breast cancer, frailty models, cure models

Corresponding author: Andreas Wienke, Max Planck Institute for Demographic Research, 18057 Rostock, Konrad-Zuse-Str. 1, Germany, e-mail: wienke@demogr.mpg.de,
Tel: (+49) 381 2081 176, Fax: (+49) 381 2081 476

1 Introduction

Models for survival analysis typically assume that everybody in the study population is susceptible to the event under study and will eventually experience this event if the follow-up is sufficiently long. This is often an unstated assumption of the widely used proportional hazard models and their extensions - frailty models (for more detailed information about the frailty concept in the univariate case, see e.g. Vaupel et al. (1979)). However, there are situations when a fraction of individuals are not expected to experience the event of interest; that is, those individuals are cured or insusceptible. For example, researchers may be interested in analyzing the recurrence of a disease. Many individuals may never experience a recurrence; therefore, a cured fraction of the population exists. Historically, cure models have been utilised to estimate the cured fraction. Cure models are survival models which allow for a cured fraction of individuals. These models extend the understanding of time-to-event data by allowing for the formulation of more accurate and informative conclusions. These conclusions are otherwise unobtainable from an analysis which fails to account for a cured or insusceptible fraction of the population. If a cured component is not present, the analysis reduces to standard approaches of survival analysis. Use of cure models has been popular for joint modeling of the overall risk of a disease and the age-at-onset distribution of the diseased individuals (e.g. Farewell 1977, 1982; Kuk and Chen 1992).

In cure models (we use 'cure fraction' and 'insusceptible fraction' as interchangeable notions) the population is divided into two sub-populations so that an individual either is cured with probability $1 - \phi$, or has a proper survival function $S_0(t)$, with probability ϕ. A model for the distribution of survival times that incorporates a cured fraction is given by

$$S(t) = (1 - \phi) + \phi S_0(t).$$

Traditional cure models assume that those individuals experiencing the event of interest are homogeneous in risk. During the last fifteen years, extensions of cure models were developed in order to allow for heterogeneity among the fraction under risk by using frailty models where the frailty distribution is a mixture of a discrete and a continuous part (e.g. Aalen 1988, 1992; Longini and Halloran 1996). The frailty mixture distribution has point mass at zero with probability $1 - \phi$ while heterogeneity among those experiencing the event of interest is

modeled via a continuous distribution with probability ϕ. In the gamma frailty mixture model the survival function is given by

$$S(t) = (1 - \phi) + \phi(1 + \sigma^2 H_0(t))^{-1/\sigma^2}.$$

Here H_0 denotes the integrated baseline hazard function related to $S_0(t)$. Price and Manatunga (2001) gave a good introduction into this area and applied leukemia remission data to different cure, frailty and frailty cure models. They conclude that frailty models are useful in modeling data with a cured fraction and found that the gamma frailty cure model provides a better fit to their remission data compared to the standard cure model.

Chatterjee and Shih (2001) considered an extension of such univariate frailty cure models to a bivariate setting. They used three different copulas in their two-step analysis procedure. We suggest the use of the copula of the correlated gamma-frailty model (an extension of Clayton's shared gamma-frailty model) and show by using simulations that all the parameters are estimable in a one-step ML estimation procedure.

In the next section we describe the proposed model, then provide an application of the model to breast cancer data from the Swedish Twin Registry in section three. This is followed by a results section. In section five we perform a simulation study to show the identifiability and the performance of the proposed method. The paper concludes with a discussion of further applications, drawbacks and advantages of the model.

2   Statistical Methods

Our approach is motivated by the paper of Chatterjee and Shih (2001). We define an individual as susceptible if he/she will eventually develop the disease if followed-up for a sufficiently long time. For a pair of individuals, $j = 1, 2$, define

$$Y_j = \begin{cases} 1 & \text{if the } j\text{ th individual is susceptible} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

and use T_j^* for the age of onset for the jth individual when $Y_j = 1$. Furthermore, let $\phi_j = \mathbf{P}(Y_j = 1)$ and $S_j(t) = \mathbf{P}(T_j^* > t | Y_j = 1)$ describe the marginal distribution of Y_j and the

failure time T_j^* for the susceptible individuals, respectively. Because of the symmetry in the twin data used as an example later on, we assume $\phi_1 = \phi_2$ and $S_1(t) = S_2(t)$.

Chatterjee and Shih (2001) used three different copulas - Clayton's model, Frank's model and the positive stable model - to specify the dependency structure between the failure times of two susceptible individuals. Here we use an extension of Clayton's model (better known as the shared gamma-frailty model). Clayton's model is given by

$$S(t_1, t_2) = \begin{cases} \left(S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1\right)^{-\frac{1}{\sigma^2}} & \text{if } \sigma^2 > 0 \\ S(t_1)S(t_2) & \text{if } \sigma^2 = 0 \end{cases} \tag{2}$$

where $S(t)$ denotes the marginal univariate survival function, assumed to be equal for both partners in a twin pair.

Shared frailty explains correlations within clusters (here twin sib-ships). However, it does have some limitations. First, it forces the unobserved factors to be the same within the cluster, which is not generally reasonable. For example, sometimes it may be inappropriate to assume that both twin partners share all their unobserved risk factors. Second, the dependence between survival times within the cluster is based on marginal distributions of survival times. To see this, when covariates are present in a proportional hazards model with gamma distributed frailty, the dependence parameter and the population heterogeneity are confounded (Clayton and Cuzick, 1985), implying that the joint distribution can be identified from the marginal distributions (Hougaard, 1986). Elbers and Ridder (1982) show that this problem exists for any univariate frailty distribution which has a finite mean. Third, in most cases univariate frailty will only induce positive association within the cluster. However, there are some situations that the survival times for subjects within the same cluster are negatively associated.

To avoid all these limitations, correlated frailty models (Yashin and Iachine, 1994; 1995; Yashin et al. 1995; Pickles et al., 1994; Commenges and Jacmin-Gadda, 1997; Petersen, 1998; Wienke et al., 2002) are developed for the analysis of multivariate failure time data, in which two associated random variables are used to characterise the frailty effect for each cluster. For example, one random variable is assigned for twin 1 and another for twin 2 so that they would no longer be constrained to have a common frailty. These two variables are associated and jointly distributed, therefore, knowing one of them does not necessarily imply the other.

In the following we apply the correlated gamma-frailty model including an insusceptible fraction to fit bivariate time-to-event (occurrence of breast cancer) data. The correlated gamma-frailty model provides a specific parameter for correlation between the two frailties. The interesting point here is that individual frailties in twin pairs could not be observed, but their correlation could be estimated by application of the correlated gamma-frailty model.

The bivariate survival function is given by

$$S(t_1, t_2) = \begin{cases} S(t_1)^{1-\rho} S(t_2)^{1-\rho} (S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1)^{-\frac{\rho}{\sigma^2}} & \text{if } \sigma^2, \rho > 0 \\ S(t_1) S(t_2) & \text{if } \sigma^2 = 0 \text{ or } \rho = 0 \end{cases} \tag{3}$$

where $S(t)$ denotes the marginal univariate survival function, assumed to be equal for both partners in a twin pair and $0 \leq \rho \leq 1$ holds. Obviously, the shared gamma-frailty model (2) is a special case of (3) when $\rho = 1$ holds.

We use a parametric approach by fitting a Gamma-Gompertz model to the data, e.g.

$$S(t) = \left(1 + \sigma^2 \frac{\alpha}{\beta} (e^{\beta t} - 1)\right)^{-\frac{1}{\sigma^2}},$$

where $\alpha, \beta, \sigma^2, \rho$ are the parameters to be estimated.

Let $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ be independent and identically distributed (i.i.d.) non-negative two-dimensional random vectors (pairs of lifetimes). The lifetimes (X_{i1}, X_{i2}) are assumed to be independently censored from the right by i.i.d. pairs of non-negative random variables $(C_{11}, C_{12}), \ldots, (C_{n1}, C_{n2})$. Thus, instead of (X_{i1}, X_{i2}) we only observe

$$(T_{i1}, T_{i2}, \Delta_{i1}, \Delta_{i2}) \tag{4}$$

with $T_{ij} = \min\{X_{ij}, C_{ij}\}$, $\Delta_{ij} = 1(X_{ij} \leq C_{ij})$ $(i = 1, \ldots, n; j = 1, 2)$, where $1(\cdot)$ denotes the indicator function of the event in the brackets. The likelihood function of the data in (4) is given by Chatterjee and Shih (2001):

$$\begin{aligned} L(t_1, t_2, \delta_1, \delta_2) &= \delta_1 \delta_2 p_{11} S_{t_1 t_2}(t_1, t_2) + \delta_1(1 - \delta_2)\Big(p_{11} S_{t_1}(t_1, t_2) + p_{10} S_{t_1}(t_1)\Big) \\ &+ (1 - \delta_1)\delta_2 \Big(p_{11} S_{t_2}(t_1, t_2) + p_{01} S_{t_2}(t_2)\Big) \\ &+ (1 - \delta_1)(1 - \delta_2)\Big(p_{11} S(t_1, t_2) + p_{10} S(t_1) + p_{01} S(t_2) + p_{00}\Big) \end{aligned}$$

where $p_{11} = \mathbf{P}(Y_1 = 1, Y_2 = 1)$, $p_{10} = \mathbf{P}(Y_1 = 1, Y_2 = 0)$, $p_{01} = \mathbf{P}(Y_1 = 0, Y_2 = 1)$, $p_{00} = \mathbf{P}(Y_1 = 0, Y_2 = 0)$ and derivatives $S_{t_j}(t_1, t_2) = \frac{\partial S(t_1, t_2)}{\partial t_j}$, $S_{t_1 t_2}(t_1, t_2) = \frac{\partial^2 S(t_1, t_2)}{\partial t_1 \partial t_2}$ and $S_{t_j}(t_j) = \frac{dS(t_j)}{dt_j}$ $(j = 1, 2)$. Here $(t_1, t_2, \delta_1, \delta_2)$ denotes a realisation of the random vector $(T_1, T_2, \Delta_1, \Delta_2)$.

As mentioned above, the twin pair data set used is not randomly selected from the total twin population. Since both members of a twin pair had to remain alive in 1958 to be included in the study population, the survival times are sampled from specific conditional distributions. If a twin pair was born in year y (where y=1886, ... ,1925), the condition of survival of both twins until the year 1958 implies that both twins had to survive until 1958-y in order to be included in the sample. If the survival times are denoted by T_1 and T_2 with survival function $S(t_1, t_2)$, then the conditional survival function for a twin pair born in year y is:

$$S(t_1, t_2 | T_1 > 1958 - y, T_2 > 1958 - y) = \frac{S(t_1, t_2)}{S(1958 - y, 1958 - y)}.$$

Consequently, the likelihood function of bivariate left truncated and right censored lifetime data is given by

$$
\begin{aligned}
L(t_1, t_2, \delta_1, \delta_2, y) &= \Big(\delta_1 \delta_2 p_{11} S_{t_1 t_2}(t_1, t_2) + \delta_1 (1 - \delta_2) \big(p_{11} S_{t_1}(t_1, t_2) + p_{10} S_{t_1}(t_1) \big) \\
&+ (1 - \delta_1) \delta_2 \big(p_{11} S_{t_2}(t_1, t_2) + p_{01} S_{t_2}(t_2) \big) + (1 - \delta_1)(1 - \delta_2) \big(p_{11} S(t_1, t_2) \\
&+ p_{10} S(t_1) + p_{01} S(t_2) + p_{00} \big) \Big) \Big/ \big(p_{11} S(1958 - y, 1958 - y) \\
&+ p_{10} S(1958 - y) + p_{01} S(1958 - y) + p_{00} \big).
\end{aligned}
$$

For a combined analysis of monozygotic and dizygotic twins we include two correlation coefficients, ρ_{MZ} and ρ_{DZ}, respectively. These correlations between monozygotic and dizygotic twins provide information about genetic and environmental influences on frailty within individuals.

3 Breast cancer data of Swedish twins

Breast cancer incidence data of identical (MZ) and fraternal (DZ) female twins were provided by the Swedish Twin Registry. This was founded in the years 1959-61 as the world's largest

nation-wide twin registry and has been continually developed ever since. This population-based registry includes all (traced) twins born in Sweden in the period 1886-1990. We restrict our analysis to the so-called old cohort (born 1886-1925) because of small numbers in the middle cohort. The old cohort consists of all same-sex twin pairs with both individuals alive when the registry was established in 1959-61. The data set was created by merging the Swedish Twin Registry with the Swedish Cancer Registry maintained by the National Board of Health and Welfare. At the time of record linkage, the Swedish Cancer Registry contained all cases of cancer that were diagnosed during the period from 1959 through 2000.

The church registers from all parishes of the relevant time period were manually checked to identify all twin births. In 1959-61 a questionnaire was sent to all twins including a question about phenotypic similarities to assess the zygosity (Where you as children as alike as two peas in a pod?). When both partners agreed, they were defined as MZ twins. This zygosity classification was compared with laboratory methods. The misclassification rate for this method was found to be very low (Cederlöf et al., 1961).

The data set provided by the Swedish Twin Registry contains records of 5,857 female twin pairs with both partners alive in 1959-61. Consequently, lifetimes are bivariate left truncated. Individuals were followed from 1959-61 to October 27, 2000. Altogether, we have 2,003 monozygotic twin pairs and 3,854 dizygotic twin pairs and 715 cases of breast cancer were identified during the follow-up. More detailed information about the composition of the Swedish Twin Registry can be found in Lichtenstein et al. (2002). Mortality in the study population was determined by linkage to the Mortality Registry of Statistics, Sweden. Analysis was made with SPSS and GAUSS.

	concordant pairs	discordant pairs	concordance rate
MZ twins	18	218	0.14
DZ twins	27	407	0.12

Table 1: Probandwise concordances for breast cancer

4 Results

Applying the correlated frailty model with and without the cure fraction to the breast cancer data described above yields the results given in Table 2. We consider two different cases of cure models. In the first case we assume that the susceptible status of the individuals in a twin pair is independent of each other, e.g. $\mathbf{P}(Y_1 = p_1, Y_2 = p_2) = \mathbf{P}(Y_1 = p_1)\mathbf{P}(Y_2 = p_2)$ with $p_1, p_2 \in \{0, 1\}$. In that case the cure fraction is uniquely described by the univariate probability $\phi = \mathbf{P}(Y_1 = 1) = \mathbf{P}(Y_2 = 1)$, which results in $\phi_{11} = \phi^2$, $\phi_{10} = \phi_{01} = \phi(1-\phi)$, $\phi_{00} = (1-\phi)^2$. In the second case, which is an extension of the first one, we relax the above restriction of independence between the susceptibility status of the two partners in a pair and use the weaker constraints $\phi_{10} = \phi_{01}$, $\phi_{11} + \phi_{10} + \phi_{01} + \phi_{00} = 1$. Comparing the likelihood of the two it turns out that the cure model with the independent susceptible status of the twin partners shows a significantly better fit than the model without the cure fraction ($\chi_1^2 = 4.32$, $p = 0.04$). The more complicated cure model without the independence assumption between the susceptible status of the twin partners shows no significant improvement compared with the cure model assuming independence ($\chi_1^2 = 0.23$, $p = 0.63$).

Parameter	without cure fraction estimates (std)	with cure fraction[1] estimates (std)	with cure fraction[2] estimates (std)
α	$1.31{\cdot}10^{-5}$ $(1.04{\cdot}10^{-5})$	$7.64{\cdot}10^{-5}$ $(4.84{\cdot}10^{-5})$	$1.175{\cdot}10^{-5}$ $(1.170{\cdot}10^{-5})$
β	0.099 (0.016)	0.091 (0.012)	0.086 (0.015)
σ	5.736 (0.680)	2.107 (0.406)	1.576 (0.951)
ρ_{MZ}	0.154 (0.052)	1.000 (-)	1.000 (-)
ρ_{DZ}	0.125 (0.040)	0.934 (0.361)	0.962 (0.457)
ϕ_{11}	1.000 (-)	0.049 (-)	0.038 (0.021)
ϕ_{10}	0.000 (-)	0.173 (-)	0.133 (0.058)
ϕ_{00}	0.000 (-)	0.605 (-)	0.696 (0.136)
ϕ	1.000 (-)	0.222 (0.045)	0.171^3 (-)
likelihood	5122.7020	5120.5408	5120.4237

Table 2: Results of breast cancer data with correlated gamma-frailty model without and with cured fraction. [1] constrained by $\phi_{11} = \phi^2$, $\phi_{10} = \phi_{01} = \phi(1 - \phi)$, $\phi_{00} = (1 - \phi)^2$, [2] constrained by $\phi_{10} = \phi_{01}$, $\phi_{11} + \phi_{10} + \phi_{01} + \phi_{00} = 1$, [3] calculated by $\phi = \phi_{11} + \phi_{10}$

5   Simulations

All simulations involve generating gamma-distributed frailties, bivariate lifetimes, censoring and truncation times as well as the inclusion of a cured fraction in the study population. We will try to mimic the characteristics of the Swedish twin data which we analyzed in the previous example. A total of 6,500 twin pairs are simulated, a number which is reduced by the truncation process to a final sample size of around 5,700 twin pairs. Samples are generated as follows:

- Generate frailty variables using independent gamma-distributed random variables.

- Generate lifetimes given the frailties using $S(t|Z) = e^{-Z\frac{\alpha}{\beta}(e^{\beta t}-1)}$.

- Define cured individuals by using a random variable.

- The censored (bivariate) lifetimes are generated by using the year 2000 as the end of the study.

- Birth years are generated by using a uniform distribution on [1886,1925] to mimic the truncation pattern. The year of truncation is 1958.

The simulated data were generated assuming independence between the susceptibility status of the partners (second column in Table 2), but in the estimation procedure the more general model with a dependent susceptibility status was applied (third column in Table 2). We simulated 1,000 data sets.

Parameter	true value	Mean of estimates	standard deviation
α	$1 \cdot 10^{-5}$	$1.16 \cdot 10^{-5}$	$6.70 \cdot 10^{-6}$
β	0.120	0.120	0.010
σ	2.000	2.016	0.270
ρ	0.600	0.606	0.132
ϕ_{11}	0.160	0.164	0.027
ϕ_{10}	0.240	0.241	0.014
ϕ_{00}	0.360	0.354	0.046

Table 3: Parameter estimation in the simulation study.

The mean parameter estimates of the model are shown in Table 3, in comparison with the true values used for simulation. There appears to be nearly no bias in the parameter estimates, and the overall performance is quite accurate.

6 Discussion

In this paper we have suggested a cure-mixture model for the modeling of correlations in bivariate time-to-event data. This model extends the approach outlined in the paper of Chatterjee and Shih (2001) in various ways. First, instead of the shared gamma-frailty model we use the much more flexible correlated gamma-frailty model, which includes the shared gamma-frailty model as a special case. Second, we propose to use a direct estimation procedure in the parametric model instead of the two-step estimation procedure used by Chatterjee and Shih. Third, we think that our twin data are more appropriate as an illustrative example than the family data of Chatterjee and Shih (who ignored higher order correlations in their family data) for such bivariate models. Nevertheless, our estimate of the size of a susceptible fraction (due to breast cancer) with 0.222 (0.045) is very close to the estimate 0.22 (0.0093) in the parametric model found by Chatterjee and Shih in a completely different study population. Furthermore, this finding is in line with the results of Peto and Mack (2002). Fourth, we allow the lifetimes to be truncated in our model.

Cure models with the right censored observations suffer from an inherent identifiability problem. For such observations the event under study has not occurred either because the person is insusceptible, or because the person is susceptible but follow-up was not long enough for the event to be observed. The identifiability problem increased with increasing censoring, but is lessened by the parametric modeling of the baseline hazard. The simulation study shows that the estimation procedure works well under the given truncation and censoring scheme in our sample data set. Stronger right censoring causes strong identifiability problems. For example, in an additional simulation study (not shown here) using the same parameters as described in the simulation section but using the birth years 1926-1958 (which is the situation in the so-called middle cohort of the Swedish Twin Registry) resulted in a complete breakdown of the estimation procedure. In cure models with fixed censoring times (caused by the end of the study) censoring is no longer non-informative even when the censoring times and the survival times are independent. The proportion of censored observations contains important informa-

tion about parameters in the model. For example, in the (usually ideal) case of no censoring, it holds $\phi = 1$.

The present paper is restricted to the parametric case, meaning in our case the marginal survival function is specified parametrically up to a few (one-dimensional) parameters, $S(t) = \left(1 + \sigma^2 \frac{\alpha}{\beta}(e^{\beta t} - 1)\right)^{-\frac{1}{\sigma^2}}$ with Gompertz parameters α and β. From a statistical point of view such a parametric assumption is unsatisfactory, because it is non-justifiable. Frailty models of univariate data have been strongly criticised because assumptions have to be made about both the shape of the underlying mortality trajectory and the distribution of the frailty: different pairs of assumptions can result in equally good fits to the data. Without an insusceptible fraction in the population ($\phi = 1$) this problem can be solved by using the non-parametric correlated gamma-frailty model (Yashin and Iachine, 1995). Applying the (true) parametric and semi-parametric estimation procedures to the same (simulated) data generated from the correlated gamma-frailty model, the semi-parametric estimation procedure shows good performance, despite the fact that it does not make use of the additional information about the parametric structure of the marginal survival functions. The estimates of σ^2 and ρ are similar in both cases (results are not shown here). Nevertheless, using the wrong parametric model may result in biased parameter estimates.

To what extent this method is applicable in the much more complicated semi-parametric model with cure fraction is still an open question, one that needs further careful consideration. Dealing with a disease with late age of onset resulting in heavily censored data may lead to problems in estimating the (infinite dimensional) nuisance parameter - the marginal survival function - and, consequently, in estimating the parameters of interest, σ^2 and ρ.

Our study points to the existence of an important insusceptible fraction. The suggested model gives a clear illustration of how survival analysis and cure models could be merged for analysis of time-to-event data of related individuals.

Acknowledgments

References

[1] Aalen, O.O. (1988) Heterogeneity in Survival Analysis.
Statistics in Medicine 7, 1121 - 1137

[2] Aalen, O.O. (1992) Modelling Heterogeneity in Survival Analysis by the Compound Poisson Distribution. Annals of Applied Probability 4 (2), 951 - 972

[3] Cederlöf, R., Friberg, L., Jonsson, E., Kaij, L. (1961) Studies on similarity diagnosis in twins with the aid of mailed questionnaires. Acta Genetica 11, 338 - 62

[4] Chatterjee, N., Shih, J. (2001) A bivariate cure-mixture approach for modeling familial association in diseases. Biometrics 57, 779 - 786

[5] Clayton, D., Cuzick, J. (1985) Multivariate Generalizations of the Proportional Hazards Model. Journal of the Royal Statistical Society A 148, 82 - 117

[6] Commenges, D., Jacmin-Gadda, H. (1997) Generalized score test of homogeneity based on correlated random effects models.
Journal of the Royal Statistical Society (B) 59, 157 - 171

[7] Elbers, C., Ridder, G. (1982) True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model. Review of Economic Studies XLIX, 403 - 409

[8] Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. Biometrika 73, 387 - 396

[9] Lichtenstein, P., de Faire, U., Floderus, B., Svartengren, M., Svedberg, P., Pedersen, N.L. (2002) The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. Journal of Internal Medicine 252, 184 - 205

[10] Longini, I.M., Halloran, M.E. (1996) A Frailty Mixture Model for Estimating Vaccine Efficacy. Applied Statistics 45, 165 - 173

[11] Pickles, A., Crouchley, R., Simonoff, E., Eaves, L., Meyer, J., Rutter, M., Hewitt, J., Silberg, J. (1994) Survival Models for Developmental Genetic Data: Age of Onset of Puberty and Antisocial Behavior in Twins. Genetic Epidemiology 11, 155 - 170

[12] Petersen, J.H. (1998) An additive frailty model for correlated lifetimes.
Biometrics 54, 646 - 661

[13] Peto, J., Mack, T.M. (2000) High constant incidence in twins and other relatives of women with breast cancer. Nature Genetics 26, 411 - 414

[14] Price, D.L., Manatunga, A.K. (2001) Modelling survival data with a cured fraction using frailty models. Statistics in Medicine 20, 1515 - 1527

[15] Vaupel, J.W., Manton, K.G., Stallard, E. (1979) The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. Demography 16, 439 - 454

[16] Wienke, A., Christensen, K., Skytthe, A., Yashin, A.I. (2002) Genetic analysis of cause of death in a mixture model with bivariate lifetime data. Statistical Modelling 2, 89 - 102

[17] Yashin, A.I., Iachine, I.A. (1994) Environment determines 50 % of variability in individual frailty: results from Danish twin study.
Research Report, Population Studies of Aging 10, Odense University, Odense

[18] Yashin, A.I., Iachine, I.A. (1995) Genetic analysis of durations: correlated frailty model applied to the survival of Danish twins. Genetic Epidemiology, 12, 529 - 38

[19] Yashin, A.I., Vaupel, J.W., Iachine, I.A. (1995) Correlated individual frailty: An advantageous approach to survival analysis of bivariate data.
Mathematical Population Studies 5, 145 - 159