

The Modelling of a Cure Fraction in Bivariate Time-to-Event Data

Andreas Wienke¹, Isabella Locatelli², and Anatoli I. Yashin³

¹University of Halle-Wittenberg, Halle

²Institute of Social and Preventive Medicine, Lausanne

³Duke University, Durham

Abstract: Three correlated frailty models are used to analyze bivariate time-to-event data by assuming gamma, log-normal and compound Poisson distributed frailty. All approaches allow to deal with right censored lifetime data and account for heterogeneity as well as for a non-susceptible (cure) fraction in the study population. In the gamma and compound Poisson model traditional ML estimation methods are used, whereas in the log-normal model MCMC methods are applied. Breast cancer incidence data of Swedish twin pairs illustrate the practical relevance of the models, which are used to estimate the size of the susceptible fraction and the correlation between the frailties of the twin partners. We discuss future directions of development of the methods and additional thoughts concerning their advantages and use.

Zusammenfassung: Drei korrelierte Frailty-Modelle werden benutzt um bivariate Lebensdauerdaten zu analysieren. Dabei werden die Gamma, Log-Normal und compound Poisson Verteilung für die Frailty-Variable angenommen. Alle Modelle sind auf rechts zensierte Daten anwendbar und erlauben die Modellierung einer Subpopulation, die dem interessierenden Ereignis gegenüber geschützt ist. Im Gamma und compound Poisson Modell werden traditionelle ML Schätzungen verwendet, während im Log-Normal Modell MCMC Methoden angewendet werden. Daten über Brustkrebs bei schwedischen Zwillingen illustrieren die praktische Anwendbarkeit der Modelle, die insbesondere genutzt werden, um die Größe der geschützten Subpopulation und die Korrelationen zwischen den Frailties der Zwillingspartner zu schätzen. Vorteile, Nachteile und offene Fragen der Forschung im Bereich dieser Modelle werden diskutiert.

Keywords: Breast Cancer, Frailty Model, Cure Model, Survival Analysis.

1 Introduction

The Cox proportional hazards model (Cox, 1972) is commonly used in the analysis of survival time data. An often unstated assumption of the proportional hazards model and of traditional frailty models (with the exception of those that use the compound Poisson distribution (Aalen, 1988, Aalen, 1992)) is that all individuals will experience the event of interest. However, in some situations a fraction of individuals is not expected to experience the event of interest; that is, these individuals are not at risk. The terminology to describe the never-at-risk group varies from field to field, but includes ‘long-term

survivors' or 'cured' in epidemiology, 'non-susceptibles' in toxicology, 'stayers' in finite Markov transition models of occupational mobility, the 'non-fecundable' in fertility models, and 'non-recidivists' among convicted criminals. In epidemiology and medicine, researchers may be interested in analyzing the occurrence of a disease. Many individuals may never experience that disease; therefore, there exists a fraction in the population that is protected. Cure models are survival models which allow for a cured fraction in the study population.

These models extend the understanding of time-to-event data by allowing for the formulation of more accurate and informative conclusions than previously made. These conclusions would otherwise be unobtainable from an analysis that fails to account for a cured fraction in the population. If a cured component is not present, the analysis reduces to standard approaches of survival analysis. Most cure models assume that the susceptible individuals are homogeneous in risk. This paper deals with cure models which include a frailty variable in order to allow for heterogeneity among the fraction under risk. Or, depending on the point of view, this paper deals with frailty models including a non-susceptible (or cured) fraction in the study population. In this case, the distribution of the frailty is a combination of discrete and continuous distributions.

In cure models, the population is divided into two sub-populations so that an individual is either cured with probability $1 - \phi$, or has a proper survival function $S(t)$, with probability ϕ . Here, proper survival function means $\lim_{t \rightarrow \infty} S(t) = 0$. Individuals regarded as cured will never experience the event of interest and their survival time will be defined as infinity. Therefore, the hazard and survival functions of cured individuals are set to zero and one, respectively, for all finite values of t . A univariate time-to-event model that incorporates a cure fraction is given by

$$\bar{S}(t) = (1 - \phi) + \phi S(t).$$

Longini and Halloran (1996) have proposed frailty cure models that extend standard frailty models. The frailty random variable in the former has point mass at zero with probability $1 - \phi$ while heterogeneity among those experiencing the event of interest is modelled via a continuous distribution with probability ϕ . Price and Manatunga (2001) gave an excellent introduction to this area and applied leukaemia remission data to different cure, frailty and frailty cure models. They found that frailty models are useful in modelling data with a cured fraction and that the gamma frailty cure model provides a better fit to their remission data compared to the standard cure model.

The following example provides an extension of the above model to include censored observations. Consider two types of expressions for a disease, the incidence and the age of disease onset. Risk models for overall susceptibility (lifetime risk) that consider only the first expression by treating the disease as a binary trait of being affected or not can give wrong results because, for individuals without the disease, due to censoring, it is often not known whether they will eventually develop the disease. On the other hand, models from survival analysis typically assume that everyone has the same susceptibility to the disease and will eventually be effected if the follow-up is sufficiently long. These models possibly do not accurately describe the disease risk factors. In models dealing with both types of expressions, the effect of a covariate can act on either the overall susceptibility or the age at onset or both. In general, cure models are special cases of the binary (two-point) frailty

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. Wienke et al. (2003b) suggested the use of a correlated gamma frailty model (Yashin and Iachine, 1995; Yashin et al., 1995; Wienke et al., 2003a).

The main task of the present paper is the estimation of the size of the susceptible fraction with respect to breast cancer by comparing the results of the correlated gamma cure, the correlated log-normal cure and the correlated compound Poisson frailty model.

In the next section we describe the proposed models, then provide an application of the models to breast cancer data from the Swedish Twin Registry in section three. The paper concludes with a discussion of further applications, drawbacks and advantages of the models.

2 Statistical Models

2.1 Correlated Gamma Frailty Cure Model

A bivariate shared frailty cure model for familiar association in diseases was established by Chatterjee and Shih (2001). We use their notation here and define for a pair of individuals

$$V_j = \begin{cases} 1 & \text{if the } j\text{-th individual is susceptible,} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and let T_j denote the age at onset for the j -th individual when $V_j = 1$ ($j = 1, 2$). In terms of hazard rates, the model is given by the relation

$$\lambda_j(t|V_j, Z_j) = V_j Z_j \lambda_0(t),$$

with Z_j denoting the frailty of the j -th individual in the pair.

Chatterjee and Shih (2001) used three different copulas in their approach: the shared gamma frailty model (Clayton, 1978), Frank's copula and Hougaard's shared positive stable frailty model (Hougaard, 1986). A two step estimation procedure was applied. Their model was partly extended by Wienke et al. (2003b), who substituted the shared gamma frailty model by the correlated gamma frailty model. Furthermore, all parameters were estimated in a one-step maximum likelihood procedure. The form of the baseline hazard is important because all methods described below are parametric. In principle, any parametric formula for a hazard rate is possible (e.g., Gompertz, Gompertz-Makeham, Weibull, exponential, piecewise constant). A vast literature on human mortality and time to onset of specific diseases suggests using the Gompertz hazard rate. For that reason and to save space, we investigate only bivariate frailty models that have the Gompertz baseline hazard rate. Let $\phi_{11} = P(\{V_1 = 1, V_2 = 1\})$, $\phi_{10} = P(\{V_1 = 1, V_2 = 0\})$, $\phi_{01} = P(\{V_1 = 0, V_2 = 1\})$, and $\phi_{00} = P(\{V_1 = 0, V_2 = 0\})$ describe the joint distribution of (V_1, V_2) . The general bivariate survival function after integrating out the random variables V_1, V_2, Z_1, Z_2 looks like

$$\bar{S}(t_1, t_2) = \phi_{11}S(t_1, t_2) + \phi_{10}S(t_1) + \phi_{01}S(t_2) + \phi_{00}.$$

The bivariate survival function $S(t_1, t_2)$ is the conditional survival function of pairs with both individuals susceptible, e.g. given the event $\{V_1 = 1, V_2 = 1\}$. The likelihood function of bivariate right censored lifetime data in this model is given by

$$L(t_1, t_2, \delta_1, \delta_2) = \left(\delta_1 \delta_2 \phi_{11} S_{t_1, t_2}(t_1, t_2) \right. \\ \left. + \delta_1 (1 - \delta_2) (\phi_{11} S_{t_1}(t_1, t_2) + \phi_{10} S_{t_1}(t_1)) \right. \\ \left. + (1 - \delta_1) \delta_2 (\phi_{11} S_{t_2}(t_1, t_2) + \phi_{01} S_{t_2}(t_2)) \right. \\ \left. + (1 - \delta_1) (1 - \delta_2) (\phi_{11} S(t_1, t_2) + \phi_{10} S(t_1) + \phi_{01} S(t_2) + \phi_{00}) \right).$$

The form of $S(t_1, t_2)$ depends on the frailty distribution. The correlated gamma frailty model (Pickles et al., 1994; Yashin and Iachine, 1995; Yashin et al., 1995) is developed for the analysis of multivariate failure time data, in which two associated random variables are used to characterize the frailty effect for each cluster. To be more specific, let k_0, k_1 be some real positive variables. Set $\lambda = k_0 + k_1$ and let Y_0, Y_1, Y_2 be independent gamma distributed random variables with $Y_0 \sim \Gamma(k_0, \lambda)$, $Y_1 \sim \Gamma(k_1, \lambda)$, and $Y_2 \sim \Gamma(k_1, \lambda)$. Consequently,

$$Z_1 = Y_0 + Y_1 \sim \Gamma(k_0 + k_1, \lambda) = \Gamma(\lambda, \lambda) \quad (2)$$

$$Z_2 = Y_0 + Y_2 \sim \Gamma(k_0 + k_1, \lambda) = \Gamma(\lambda, \lambda) \quad (3)$$

are the frailties of individual 1 and 2 in a pair. The bivariate survival function of this model is given by

$$S(t_1, t_2) = S(t_1)^{1-\rho} S(t_2)^{1-\rho} \left(S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1 \right)^{-\rho/\sigma^2}, \quad (4)$$

where $S(t)$ denotes the marginal univariate survival function, assumed to be equal for both twin partners and $0 \leq \rho \leq 1$ holds. Furthermore, it holds that $\rho = \text{corr}(Z_1, Z_2)$, $E(Z_j) = 1$ and $\sigma^2 = \text{var}(Z_j)$. Obviously, the shared gamma frailty model by Clayton (1978) is a special case of (4) when $\rho = 1$.

2.2 Correlated Log-Normal Frailty Cure Model

The correlated log-normal model is much more flexible than the correlated gamma one, because it is not based on the additive composition of the two frailties as used in (2) and (3). However, the log-normal distribution does not allow an explicit representation of the likelihood function, which requires more sophisticated estimation strategies. We assume that the two frailties of individuals in a pair are given by

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \text{LogN} \left(\begin{pmatrix} m \\ m \end{pmatrix}, \begin{pmatrix} s^2 & r s^2 \\ r s^2 & s^2 \end{pmatrix} \right), \quad (5)$$

where LogN denotes the (bivariate) log-normal distribution. Here, m, s^2 , and r denote the mean, variance and correlation of the respective normal distribution. This distribution can be obtained by assuming a bivariate normal distribution on the logarithm of the frailty vector

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \log \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

with

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim \mathbf{N} \left(\begin{pmatrix} m \\ m \end{pmatrix}, \begin{pmatrix} s^2 & rs^2 \\ rs^2 & s^2 \end{pmatrix} \right), \quad (6)$$

and \mathbf{N} denotes the bivariate normal distribution whose parameters are some functions of the frailty parameters μ , σ^2 , and ρ (see, e.g., Hutchinson and Lai, 1991)

$$\mu = \mathbf{E}(Z_j) = \exp(m + s^2/2) \quad (7)$$

$$\sigma^2 = \text{var}(Z_j) = \exp(2m + s^2)(\exp(s^2) - 1) \quad (8)$$

$$\rho = \text{corr}(Z_1, Z_2) = (\exp(rs^2) - 1)/(\exp(s^2) - 1). \quad (9)$$

For reasons of identifiability of the model let $\mu = 1$. It follows from (7) to (9) that

$$\begin{aligned} m &= \mathbf{E}(\log Z_j) = -s^2/2 \\ s^2 &= \text{var}(\log Z_j) = \log(1 + \sigma^2). \end{aligned}$$

There exists no closed form expression for the bivariate survival function $S(t_1, t_2)$.

2.3 Correlated Compound Poisson Frailty Model

The third bivariate model which we consider here is the correlated compound Poisson frailty model. The univariate compound Poisson distribution was introduced by Aalen (1988, 1992) as a frailty distribution. An interesting property of the model is that it yields a subgroup of individuals with zero frailty, who survive forever. Despite the fact that the density of the continuous part is only given as an infinite series which has to be calculated numerically, the distribution is mathematically convenient. It may also be seen as a natural choice. The distribution can be constructed as the sum of a Poisson distributed number of independent and identical gamma distributed random variables. The Laplace transform of the compound Poisson distribution is given by

$$L(s) = \exp \left(-\frac{k}{\gamma} \left((\lambda + s)^\gamma - \lambda^\gamma \right) \right).$$

The notation $\text{cP}(\gamma, k, \lambda)$ is used for a compound Poisson distribution. Expectation and variance of a compound Poisson distributed random variable Z are

$$\mathbf{E}(Z) = k\lambda^{\gamma-1}, \quad \text{var}(Z) = k(1 - \gamma)\lambda^{\gamma-2}. \quad (10)$$

The Laplace transform given above implies the marginal survival and hazard function in case of a compound Poisson frailty model

$$S(t) = \exp \left(-\frac{k}{\gamma} \left((\lambda + \Lambda_0(t))^\gamma - \lambda^\gamma \right) \right), \quad \lambda(t) = k\lambda_0(t)(\lambda + \Lambda_0(t))^{\gamma-1}.$$

Using the constraint $\mathbf{E}(Z) = 1$ and relationship $\sigma^2 = (1 - \gamma)/\lambda$ (see (10)) it holds

$$S(t) = \exp \left(-\frac{1 - \gamma}{\gamma\sigma^2} \left(\left(1 + \frac{\sigma^2}{1 - \gamma} \Lambda_0(t) \right)^\gamma - 1 \right) \right), \quad \lambda(t) = \frac{\lambda_0(t)}{\left(1 + \frac{\sigma^2}{1 - \gamma} \Lambda_0(t) \right)^{1-\gamma}}.$$

It should be noted that the integral of $\lambda(t)$ over $[0, \infty)$ is finite when $\gamma < 0$. Consequently, the survival function is incomplete because a fraction of individuals have zero frailty, who could never experience the event under study.

To introduce the bivariate model, let Y_0, Y_1, Y_2 be independent compound Poisson distributed random variables with $Y_0 \sim \text{cP}(\gamma, k_0, \lambda)$, $Y_1 \sim \text{cP}(\gamma, k_1, \lambda)$, $Y_2 \sim \text{cP}(\gamma, k_1, \lambda)$. Consequently,

$$Z_1 = Y_0 + Y_1 \sim \text{cP}(\gamma, k_0 + k_1, \lambda) \quad (11)$$

$$Z_2 = Y_0 + Y_2 \sim \text{cP}(\gamma, k_0 + k_1, \lambda) \quad (12)$$

are the frailties of individual 1 and 2 in a pair. The bivariate survival function of this model is given by

$$S(t_1, t_2) = S(t_1)^{1-\rho} S(t_2)^{1-\rho} \exp \left(\frac{\rho(1-\gamma)}{\gamma\sigma^2} \left(1 - \left(\left(1 - \frac{\gamma\sigma^2}{1-\gamma} \log(S(t_1)) \right)^{1/\gamma} + \left(1 - \frac{\gamma\sigma^2}{1-\gamma} \log(S(t_2)) \right)^{1/\gamma} - 1 \right)^\gamma \right) \right), \quad (13)$$

where $S(t)$ denotes the marginal univariate survival function, assumed to be equal for both twin partners and $0 \leq \rho \leq 1$ holds. Furthermore, it holds that $\rho = \text{corr}(Z_1, Z_2)$ and $\sigma^2 = \text{var}(Z_j)$.

Parameter γ divides the class of distributions in two major subfamilies. For $\gamma \geq 0$, the distribution is a power variance function distribution (PVF) or three parameter distribution. This bivariate correlated frailty model was proposed by Yashin and Iachine (1995) and applied to Danish twin mortality data. The inverse Gaussian model is included as a special case by $\gamma = 1/2$. The case $\gamma > 0$ is not considered further here, because we expect parameter estimates $\gamma < 0$ in our real data example.

The extension to $\gamma < 0$ in the univariate case was suggested by Aalen (1988, 1992) and shown to yield the compound Poisson distribution. Parameter values of $\gamma < 0$ imply the existence of a non-susceptible fraction in the population. The extended bivariate model is applied to breast cancer data of Swedish twins to estimate the size of the susceptible fraction and results are compared to the analysis by the correlated gamma frailty cure model.

3 Example

The data set from the Swedish Twin Register contains records of 5857 female twin pairs with both partners being alive in 1959–61. Individuals were followed up from 1959/61 to 27 October 2000. Altogether, we have 2003 monozygotic and 3854 dizygotic twin pairs, and 715 cases of breast cancer were identified during the follow-up. More detailed information about the Swedish Twin Register can be found in Lichtenstein et al. (2002). The results are given in Table 1. We consider the three different correlated frailty cure models described in the last section. In the first and second model the susceptible status of the individuals in a pair is assumed to be independent of each other, e.g.,

$$P(\{V_1 = p_1, V_2 = p_2\}) = P(\{V_1 = p_1\})P(\{V_2 = p_2\})$$

Table 1: Analysis of time to onset of breast cancer in 5857 Swedish twin pairs. In the correlated gamma and log-normal model estimates are constrained by $\phi_{11} = \phi^2$, $\phi_{10} = \phi_{01} = \phi(1 - \phi)$, $\phi_{00} = (1 - \phi)^2$.

	correlated gamma	correlated log-normal	correlated compound Poisson
γ	–	–	-0.052 (0.096)
σ	2.780 (0.931)	4.865 (3.791)	7.034 (0.994)
ρ_{MZ}	0.900 (0.552)	0.717 (0.196)	0.124 (0.040)
ρ_{DZ}	0.725 (0.465)	0.645 (0.223)	0.099 (0.030)
ϕ	0.201 (0.081)	0.102 (0.007)	0.336*

* number calculated by $\phi = 1 - \exp((1 - \gamma)/(\gamma\sigma^2))$.

with $p_1, p_2 \in \{0, 1\}$. The size of the susceptible fraction is uniquely described by the probability $\phi = P(\{V_1 = 1\}) = P(\{V_2 = 1\})$, which results in $\phi_{11} = \phi^2$, $\phi_{10} = \phi_{01} = \phi(1 - \phi)$, $\phi_{00} = (1 - \phi)^2$. The case of dependence between the susceptible status of the twin partners was already considered in Wienke et al. (2003b) analyzing the same data.

In the first model gamma distributed frailty is used to account for heterogeneity in the population and a cure fraction is included into the model. Parameter σ^2 is a measure of heterogeneity, which is large in this population, and ϕ as the size of the susceptible fraction, which is found to be around 20%. The gamma frailty model without cured fraction is a special case of the compound Poisson frailty model with $\gamma = 0$. A MCMC approach was used to estimate the model parameters. More details about this case can be found in Locatelli et al. (2004).

A log-normal distributed frailty is used and a cured fraction is included in the second model. Parameter estimates are different from that in the gamma case with higher heterogeneity ($\sigma^2 = 4.865$) and a smaller susceptible fraction ($\phi = 0.102$).

In the compound Poisson model (third column) the parameter estimates are larger than in the gamma and log-normal model ($\phi = 0.336$, $\sigma^2 = 7.034$).

4 Discussion

In the present paper three different models are applied to the Swedish breast cancer data set. In all three bivariate models the association between the lifetimes of the twins is accounted for. Furthermore, it is possible to estimate the correlation between the frailties of the twin partners, which is large in the correlated gamma frailty cure and the correlated log-normal frailty cure model. It should be noted, that in the gamma and log-normal model this correlation is estimated from pairs only, where both partners are susceptible, which is the reason for the high correlations. In the correlated compound Poisson frailty model the frailty variable is a mixture of a continuous and a discrete part. Consequently, in this model the frailty Z plays the role of VZ in the gamma and the log-normal model. That is why the correlation in the compound Poisson model is much lower, because it is the correlation in all pairs including non-susceptible individuals. This problem is already discussed by Chatterjee and Shih (2001) in detail. The correlated gamma and log-normal

frailty cure model incorporate two types of association, one between the lifetime risk or the overall susceptibility of two individuals ($\text{corr}(V_1, V_2)$) and one between the frailties of onset between susceptible individuals ($\text{corr}(Z_1, Z_2)$). In the correlated compound Poisson frailty model these two types of correlation are combined and can not be separated. In all models correlations in monozygotic pairs are higher than in dizygotic pairs, but the differences are small. This is in line with the well known fact that the influence of genetic factors on susceptibility to breast cancer is small (5 - 10 %).

The estimates of the size of a susceptible fraction (due to breast cancer) are close to the estimate $\phi = 0.22$ (0.0093) in the parametric model found by Chatterjee and Shih (2001) in a study population that is completely different.

Interestingly, parameter estimates are quite different depending on whether the frailty distribution among the susceptible individuals is chosen as gamma, log-normal or compound Poisson. In all considered cases only a relatively small fraction (10% - 34%) of all women is indicated to be susceptible to breast cancer. Nevertheless, the estimates of the susceptible fraction in all three models in Table 1 are close to the figures obtained by Farewell (1977) for different combinations of four risk factors. The authors found that, if none of the risk factors is present the susceptible fraction is around 0.015, if all risk factors are present, the estimate increases to 0.272.

The size of the fraction of woman susceptible to breast cancer is given by the model parameter ϕ in the gamma and log-normal frailty cure model. This parameter does not exist in the compound Poisson frailty model. Here in this model the size of the susceptible fraction can be calculated by $\phi = 1 - \exp((1 - \gamma)/(\gamma\sigma^2))$. The estimates for the size of the susceptible fraction have to be compared with the overall lifetime risk of breast cancer, which is around 8% - 12% in current western populations (Harris et al., 1992; Feuer et al., 1993; Rosenthal and Puck, 1999; Ries et al., 1999). The twin population considered here was born between 1886 and 1925 when the lifetime risk for breast cancer was lower because of competing causes of death like infections, which are much less important today.

The newly introduced correlated compound Poisson frailty model offers a very elegant approach to integrate the concept of cure models into frailty modelling. The survival function is explicitly available and of easy form which allows traditional maximum likelihood parameter estimation. This is the most important advantage of the suggested model compared to the recently introduced model by Moger and Aalen (2005). Important frailty models like the gamma model ($\gamma = 0$) and the inverse Gaussian model ($\gamma = 1/2$) are included in this model family and provide a great flexibility of the model. Simulation studies are needed to analyze the behavior of the parameter estimates in this model. Different parameter estimation strategies (ML, ML with numerical integration, MCMC) are already analyzed in correlated frailty models without a cure fraction by Wienke et al. (2005). Of great interest would be a non-parametric version of the correlated compound Poisson frailty model, where the baseline hazard functions are not specified. This will be part of future research in this direction.

Cure models suffer from an inherent identifiability problem with right censored observations: The event under study has not occurred either because the person is unsusceptible or the person is susceptible, but follow-up was not long enough to observe the event. The identifiability problem is growing with increasing censoring, but is reduced by the

parametric modelling of the baseline hazard. In cure models with fixed censoring times (caused by ending the study) censoring is no longer non-informative even in case where the censoring and the survival times are independent. The proportion of censored observations contains important information about the parameters in the model. For example, in the (usual ideal) case of no censoring, it holds $\phi = 1$.

Acknowledgements

The authors wish to thank the Swedish Twin Register for providing the twin data and Susann Backer for help in preparing the paper for publication. A. Wienke and A.I. Yashin would like to thank the Max Planck Institute for Demographic Research in Rostock (Germany) for providing excellent research facilities during their stay at the institute in summer 2005. A. Wienke was supported by a grant of the Bundesministerium für Bildung und Forschung (NBL3-programme, 01ZZ0404).

References

- Aalen, O. O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, 7, 1121-1137.
- Aalen, O. O. (1992). Modelling heterogeneity in survival analysis by the compound poisson distribution. *Annals of Applied Probability*, 4, 951-972.
- Chatterjee, N., and Shih, J. (2001). A bivariate cure-mixture approach for modeling familial association in diseases. *Biometrics*, 57, 779-786.
- Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141-151.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society (B)*, 34, 187-220.
- Farewell, V. T. (1977). A model for a binary variable with time-censored observations. *Biometrika*, 64, 43-46.
- Feuer, E. J., Wun, L. M., Boring, C. C., Flanders, W. D., Timmel, M. J., and Tong, T. (1993). The lifetime risk of developing breast cancer. *Journal of the National Cancer Institute*, 85, 892-897.
- Harris, J. R., Lippman, M. E., Veronesi, U., and Willett, W. (1992). Breast cancer (i). *New England Journal of Medicine*, 327, 319-328.
- Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73, 387-396.
- Hutchinson, T. P., and Lai, C. D. (1991). *The engineering statistician's guide to continuous bivariate distributions*. Adelaide: Rumsby.
- Lichtenstein, P., de Faire, U., Floderus, B., Svartengren, M., Svedberg, P., and Pedersen, N. L. (2002). The swedish twin registry: a unique resource for clinical, epidemiological and genetic studies. *Journal of Internal Medicine*, 252, 184-205.
- Locatelli, I., Lichtenstein, P., and Yashin, A. I. (2004). The heritability of breast cancer: A bayesian correlated frailty model applied to swedish twins data. *Twin Research*, 7, 182-191.

- Longini, I. M., and Halloran, M. E. (1996). A frailty mixture model for estimating vaccine efficacy. *Applied Statistics*, 45, 165-173.
- Moger, T. A., and Aalen, O. O. (2005). A distribution for multivariate frailty based on the compound poisson distribution with random scale. *Lifetime Data Analysis*, 11, 41-59.
- Pickles, A., Crouchley, R., Simonoff, E., Eaves, L., Meyer, J., Rutter, M., et al. (1994). Survival models for developmental genetic data: age of onset of puberty and anti-social behavior in twins. *Genetic Epidemiology*, 11, 155-170.
- Price, D. L., and Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*, 20, 1515-1527.
- Ries, L. A. G., Kosary, C. L., and Hankey, B. F. (1999). *Seer cancer statistics review 1973-1999*. Bethesda, MD: National Cancer Institute.
- Rosenthal, T. C., and Puck, S. M. (1999). Screening for genetic risk of breast cancer. *American Family Physician*, 59, 99-104.
- Wienke, A., Arbeev, K., Locatelli, I., and Yashin, A. I. (2005). A comparison of different correlated frailty models and estimation strategies. *Mathematical Biosciences*, 198, 1-13.
- Wienke, A., Holm, N. V., Christensen, K., Skytthe, A., Vaupel, J. W., and Yashin, A. I. (2003a). The heritability of cause-specific mortality: a correlated gamma-frailty model applied to mortality due to respiratory diseases in danish twins born 1870 - 1930. *Statistics in Medicine*, 22, 3873-3887.
- Wienke, A., Lichtenstein, P., and Yashin, A. I. (2003b). A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics*, 59, 1178-1183.
- Yashin, A. I., and Iachine, I. A. (1995). Genetic analysis of durations: correlated frailty model applied to the survival of danish twins. *Genetic Epidemiology*, 12, 529-538.
- Yashin, A. I., Vaupel, J. W., and Iachine, I. A. (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, 5, 145-159.

Authors' addresses:

Andreas Wienke
Institute for Medical Epidemiology, Biometrics and Computer Science
Martin-Luther-University Halle-Wittenberg
Magdeburger Straße 27
D-06112 Halle (Saale)

Isabella Locatelli
Institute of Social and Preventive Medicine
University of Lausanne
Rue du Bugnon 17
CH-1005 Lausanne

Anatoli I. Yashin
Center for Demographic Studies
Duke University
Box 90408 Durham, NC 27708-0408