Aus dem
Institut für Medizinische Epidemiologie, Biometrie und Informatik
(Direktor: Prof. Dr. Johannes Haerting)

Frailty Models in Survival Analysis

Habilitation

zur Erlangung des akademischen Grades

Dr. rer. nat., rer. medic. habil.

vorgelegt der Medizinischen Fakultät
der Martin-Luther-Universität Halle-Wittenberg

von Andreas Wienke
geboren am 23. Februar 1966 in Güstrow

Gutachter:

1. Prof. Dr. Johannes Haerting, Halle
2. Prof. Dr. Ulrich Mansmann, München
3. Prof. Dr. Luc Duchateau, Ghent

verteidigt am 16. Januar 2007

urn:nbn:de:gbv:3-000011542
[http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000011542]

Kurzreferat: Die vorliegende methodische Arbeit gibt einen umfassenden Überblick zur Anwendung von Frailty-Modellen in der Lebensdaueranalyse. Zunächst wird der aktuelle Stand der Forschung auf den Gebieten der univariaten und multivariaten Frailty-Modelle dargestellt. Insbesondere wird die Bedeutung der Frailty-Modelle bei der Modellierung von Heterogeneität (univariater Fall) und abhängigen Lebensdauern (multivariater Fall) herausgearbeitet. Darauf aufbauend erfolgt die Weiterentwicklung des multivariaten Korrelierten Frailty-Modells in verschiedene Richtungen. Ein erster Schwerpunkt ist dabei die Einbeziehung beobachtbarer Kovariablen und von Interaktionen beobachtbarer und nichtbeobachtbarer Kovariablen. Andere Erweiterungen dienen der Analyse von abhängigen konkurrierenden Risiken, welche sich elegant mit Hilfe des Konzepts der korrelierten Frailty-Modelle realisieren lassen. Ein weiterer Schwerpunkt der vorliegenden Arbeit ist die Einbeziehung von Teilpopulationen, welche dem analysierten Ereignis gegenüber immun sind, sei es aufgrund einer genetischen Prädisposition, einer Impfung oder einer entsprechenden Vorerkrankung. Zahlreiche Anwendungsbeispiele insbesondere aus dem Schnittbereich von Medizin, (genetischer) Epidemiologie und Demografie illustrieren die praktische Relevanz der neu entwickelten Modelle. Dazu werden unter anderem weltweit einmalige Zwillingsdaten aus dem dänischen und schwedischen Zwillingsregister analysiert, wobei Herzkreislauf- und Krebserkrankungen im Mittelpunkt stehen. Simulationen demonstrieren die Eigenschaften der Parameterschätzer in den einzelnen Modellen für realistische Stichprobenumfänge.

Wienke, Andreas: Frailty Models in Survival Analysis. Universität Halle-Wittenberg, Medizinische Fakultät, Habilitation, 153 Seiten, 2006

Contents

Abbreviations and symbols

AFT	accelerated failure time
c.d.f.	cumulative density function
$\mathbf{E}$	expectation
f	generic symbol for a probability density function
F	generic symbol for a cumulative density function
Γ	Gamma function
$\mathbf{L}$	Laplace transform
λ	generic symbol for a (baseline) hazard function
Λ	generic symbol for a cumulative hazard function
MCMC	Monte Chain Monte Carlo
MZ, DZ twins	monozygotic and dizygotic twins
$\mathbf{P}$(A)	probability of event A
p.d.f.	probability density function
S	generic symbol for a survival function
$\mathbf{V}$	variance
$X \sim U(a, b)$	uniform distributed random variable in the interval $[a, b]$
$X \sim N(\mu, \sigma^2)$	normal distributed random variable with parameters μ, σ^2
$X \sim Exp(\lambda)$	exponential distributed random variable with parameter λ
$X \sim Weibull(a, b)$	Weibull distributed random variable with parameters a, b
$X \sim Gompertz(a, b)$	Gompertz distributed random variable with parameters a, b
$X \sim \Gamma(k, \lambda)$	Gamma distributed random variable with parameters k, λ
$X \sim cP(\gamma, k, \lambda)$	compound Poisson distributed random variable with parameters γ, k, λ

All models are wrong, some are useful. (Box, 1976)

Chapter 1

Introduction and Outline

The statistical analysis of lifetime data (or more exactly, time-to-event, event-history or duration data) plays an important role in medicine, epidemiology, biology, demography, economics, engineering and other fields. It has expanded rapidly in the last three decades, with works having been published in various disciplines in addition to statistics. But what distinguishes survival analysis from other fields of statistics? Why does survival data need a special statistical theory? The reason is that we are observing something that develops dynamically over time. There are two points related to this development. First, survival times are usually a mixture of discrete and continuous data that lend themselves to a different type of analysis than in the traditional discrete or continuous case. The mixture is the result of censoring and has an important effect on data analysis. To put it plainly, a censored observation contains only partial information about the random variable of interest. The Kaplan-Meier estimator (Kaplan and Meier, 1958) of the survival function is a major step in the development of suitable models for such kind of data. Second, most evaluations are made conditionally on what is known at the time of the analysis, and this changes over time. Usually, as the population under study is changing, we only consider the individual risk to die for those who are still alive, but this means that many standard statistical approaches cannot be applied.

Models based on the hazard function have dominated survival analysis since the construction of the proportional hazards model by Cox (1972). One of the reasons this model is so popular is because of the ease with which technical difficulties such as censoring and truncation are handled. This is due to the appealing interpretation of the hazard function as a risk that changes over time. Naturally, the concept allows for the entering of covariates in order to describe their influence and to model different levels of risk for different subgroups. However, in general it is impossible to include all relevant risk factors, perhaps because we have no information on individual values, which is of-

ten the case in demography. Furthermore, we may not know all relevant risk factors or it is impossible to measure them without great financial costs, something that is common in medical and biological studies. The neglect of covariates leads to (unobserved) heterogeneity. That is, the population consists of individuals with different risks.

The second important regression model in survival analysis is the accelerated failure time model (AFT) (Lawless, 1982). The Cox model and its various generalizations are mainly used in medical and biostatistical fields, while the AFT model is primarily applied in reliability theory and industrial experiments. We will not treat the AFT model in detail here but recommend a recent paper by Orbe et al. (2002) in which the authors compare both the Cox and AFT models and discuss the advantages and limitations of each. Especially, they consider a semi-parametric version of the AFT model.

This thesis focuses on frailty models, a specific area in survival analysis. The notion of frailty provides a convenient way of introducing unobserved heterogeneity and associations into models for survival data. In its simplest form, frailty is an unobserved random proportionality factor that modifies the hazard function of an individual, or of related individuals. In essence the concept goes back to the work of Greenwood and Yule (1920) on 'accident proneness'. The term frailty itself was introduced by Vaupel et al. (1979) in univariate survival models. Applications to problems in multivariate survival analysis date from a seminal paper by Clayton (1978).

Ordinary methods in survival analysis implicitly assume that populations are homogenous, meaning all individuals have the same risk of death. As mentioned above, it is often important to consider the population as heterogeneous, i.e. a mixture of individuals with different hazards. A frailty model is a random effects model for time-to-event data, where the random effect (the frailty) has a multiplicative effect on the baseline hazard function. It can be used for univariate (independent) lifetimes, i.e. to describe the influence of unobserved covariates in a proportional hazards model (heterogeneity). The variability of duration data is split into one part that depends on risk factors and is thus theoretically predictable, and one part that is initially unpredictable, even knowing all relevant information at that time. There are advantages in separating these sources of variability: heterogeneity can explain some unexpected results or give an alternative interpretation, for example crossing-over or levelling-off effects of hazard functions.

However, considering multivariate (dependent) duration times is more interesting. The introduction of a common random effect - the frailty - is a natural way of modelling the dependence of event times. The random effect explains the dependence in the sense that had we known the frailty, the events would have been independent. In other words, the lifetimes are conditionally independent given the frailty. This approach can be used for

survival times of related individuals such as twins or family members, where independence cannot be assumed, or for recurrent events in the same individual or for times to several events for the same individual, such as onset of different diseases, relapse or death (competing risks). Different extensions of univariate frailty models to multivariate models are also possible and will be considered later.

The standard assumption is to use a gamma distribution for the frailty, but other distributions are also possible. The relationships between individual and observed survival characteristics play a key role in the statistical analysis of duration data in heterogeneous populations.

Various frailty models have been developed. However, compared with standard random effect models, frailty models pose additional difficulties in developing inferential methods, caused by incompleteness of data due to censoring and truncation and by the requirement for a specification of a baseline hazard (or a non-parametric baseline hazard). Thus, inferential methods have been less developed here than in other random effect models.

This thesis is organized as follows: In the next chapter, a short introduction to survival analysis and a definition of common terminology is given. Chapter 3 deals with univariate frailty models. It introduces the concept of frailty, gives an overview of the research field and forms the basis for Chapter 4, which contains the main results of the thesis. Various extensions of the correlated gamma frailty model are suggested and applied to cause-specific mortality and incidence data from Danish and Swedish twins. Advantages and limitations of the proposed models are discussed and simulations show the properties of the parameter estimates for reasonable sample sizes. The appendix provides details about the application data used in this thesis and some mathematical derivatives.

I would like to take this chance to express my thanks to everyone who supported this thesis; my former Ph.D. advisor Friedrich Liese (Rostock) for waking my interest in statistical research, Renee Flibotte (Rostock), Anatoli Yashin and Konstantin Arbeev (Duke), Alexander Begun (Kiel), Nicole Giard (Gütersloh), Kaare Christensen and Ivan Iachine (Odense), Isabella Locatelli (Lausanne), Slobodan Zdravkovic, Ulf de Faire and Paul Lichtenstein (Stockholm) for many helpful discussions and fruitful collaboration, and my colleagues at the Institute for Medical Epidemiology, Biostatistics and Informatics at the University Halle, especially Johannes Haerting as head of the institute for encouragement, support and providing me with an excellent research environment. I would like to thank the Danish and Swedish Twin Registries for providing the unique twin data. I also extend my sincere thanks to the Max Planck Institute for Demographic Research and its founding Director, James Vaupel, in Rostock where I started work on frailty models.

Chapter 2

Survival analysis

2.1 Basic concepts in survival analysis

This section describes basic aspects of univariate survival data and contains notation and important results which build the basis for specific points in later chapters.
We consider a single random variable X. Specifically, let X be non-negative, representing the lifetime of an individual. In that case this has given this field its name, the event is death, but the term is also used with other events, such as the onset of disease, complications after surgery or relapses in the medical field. In demography, time to death, but also time to leaving home, pregnancy, birth or divorce is of special interest. In industrial applications, it is typically time to failure of a technical unit. In economics, it can denote the time until the acceptance of jobs by unemployed, for example. Usually, X is assumed to be continuous, and we will restrict ourselves to this case in the present thesis. All functions in the sequel are defined over the interval $[0, \infty)$. The probability density function (p.d.f.) is denoted by f. The distribution of a random variable is completely and uniquely determined by its probability density function. But there many other notions exist, which are very useful in describing a distribution in specific situations. An important one is $F(x) = \mathbf{P}(X < x) = \int_0^x f(s)\, ds$, the cumulative distribution function (c.d.f.) of X. In survival analysis, one is more interested in the probability of an individual to survive to time x, which is given by the survival function

$$S(x) = 1 - F(x) = \mathbf{P}(X \geq x) = \int_x^\infty f(s)\, ds.$$

The major notion in survival analysis is the hazard function $\lambda(\cdot)$ (also called mortality rate, incidence rate, mortality curve or force of mortality), which is defined by

$$\lambda(x) = \lim_{\Delta \to 0} \frac{\mathbf{P}(x \leq X < x + \Delta | X \geq x)}{\Delta} = \frac{f(x)}{1 - F(x)}. \tag{2.1}$$

The hazard function characterizes the risk of dying changing over time or age. It specifies the instantaneous failure rate at time x, given that the individual survives until x. Sometimes, it is useful to deal with the cumulative (or integrated) hazard function

$$\Lambda(x) = \int_0^x \lambda(s)\, ds.$$

Especially for the topic covered by this thesis the notion of the Laplace transform $\mathbf{L}$ of a random variable X is crucial to inference in this area: $\mathbf{L}(s) = \mathbf{E}e^{-sX} = \int_0^\infty e^{-sx} f(x)\, dx$. The functions $f, F, S, \lambda, \Lambda$ and $\mathbf{L}$ give equivalent specifications of the distribution of the (non-negative) random variable X. It is easy to derive relations between the different notions, for example (2.1) implies that

$$\Lambda(x) = \int_0^x \lambda(s)\, ds = \int_0^x \frac{f(s)}{1 - F(s)}\, ds = -\ln(1 - F(x))$$

and consequently

$$S(x) = 1 - F(x) = e^{-\int_0^x \lambda(s)\, ds} = e^{-\Lambda(x)}. \tag{2.2}$$

As mentioned before, the hazard function is particularly useful in survival analysis, since it describes the way in which the instantaneous probability of failure for an individual changes with time. Applications often have qualitative information about the hazard function, which can be of help in selecting a model. For example, there may be reasons to restrict the analysis to models with increasing hazards or with hazard functions that have other well-defined characteristics. The shape of a hazard function can take different forms: it can be increasing, decreasing, constant, or U-shaped. Models with these and other hazard function shapes are all useful in practice: In demography, for example, following humans from birth to death, a U-shaped hazard function is often appropriate. After an initial period in which deaths result primarily from birth defects and infant diseases, the death rate drops and remains relatively constant until the age of 30 or so, after which it increases with age. This pattern is also observed in many other populations, including those consisting of technical items. Models with increasing hazards are used the most. This is because interest often centres on periods in the life of individuals in which measurable ageing takes place (for example old ages in humans). Models with a constant hazard function are of a very simple structure, as we will see in the next section. Less common are models with a decreasing hazard, but they are sometimes used to describe failure times of electronic devices, at least over a fairly long initial period of use. The main points to remember here are that the hazard function represents an aspect of a (non-negative) distribution that has a direct physical meaning and that qualitative information about the form of the hazard function is useful in selecting an appropriate model.

2.2 Parametric models

Now we shall consider in outline some distributions that are useful in the field of survival analysis. Naturally any distribution of non-negative random variables could be used to describe durations. The distributions to be discussed here are all continuous. Throughout the literature on survival analysis, certain parametric models have been used repeatedly such as exponential and Weibull models. These distributions have closed form expressions for survival and hazard functions. Log-normal and gamma distributions are generally less convenient computationally, but are still frequently applied. To avoid the model validity issues, the non-parametric approach, supported by the well-developed Kaplan-Meier-product limit estimator and related techniques, is often regarded as the preferred course. However, this alternative is often inefficient, as noted by Miller (1983). The pros and cons of different parametric, semi-parametric and non-parametric models and methodology for statistical inference can be found in the monographs by Kalbfleisch and Prentice (1980), Miller (1981), Lawless (1982), Cox and Oakes (1984) and Klein and Moeschberger (1997).

Below we discuss some of the standard failure time models for homogeneous populations. The properties and the theoretical bases of these distributions are considered here only briefly. The distributions will be studied in the simplest case of independently and identically distributed random variables. In this and the following sections the random variable X denotes the lifetime which we are interested in making inferences about.

2.2.1 Exponential distribution

The exponential model ($X \sim Exp(\lambda)$) is the simplest parametric model and assumes a constant risk over time, which reflects the property of the distribution appropriately called 'lack of memory'. The probability to die within a particular time interval depends only on the length but not on the location of this interval. This means that the distribution of $X - x$ conditional on $X \geq x$ is the same as the original distribution. In other words, it holds that

$$\mathbf{P}(x \leq X < x + \delta | X \geq x) = \mathbf{P}(X < \delta)$$

for any positive δ. As a consequence, the exponential distribution (as the only one) is not influenced by the definition of time zero. The parameter λ attains all positive values and the distribution with $\lambda = 1$ is called the unit or standard exponential. Therefore,

the following formulae can be derived by some simple algebraic calculations:

$$\begin{aligned}
\text{probability density function} \quad & f(x) = \lambda e^{-\lambda x} \\
\text{survival function} \quad & S(x) = e^{-\lambda x} \\
\text{hazard function} \quad & \lambda(x) = \lambda, \qquad \lambda > 0 \\
\text{cumulative hazard function} \quad & \Lambda(x) = \lambda x \\
\text{mean} \quad & \mathbf{E}X = \frac{1}{\lambda} \\
\text{variance} \quad & \mathbf{V}(X) = \frac{1}{\lambda^2}
\end{aligned}$$

The exponential distribution was widely used in early work on the reliability of electronic components and technical systems. The distribution of cX with a positive constant c is again exponentially distributed with parameter λ/c. The minimum of n independent exponential random variables with parameter λ is still exponential with parameter $n\lambda$:

$$\mathbf{P}(\min\{X_1, \ldots, X_n\} \geq x) = \prod_{i=1}^{n} \mathbf{P}(X_i \geq x) = \prod_{i=1}^{n} e^{-\lambda x} = e^{-n\lambda x}.$$

The model is very sensitive to even a modest variation because it has only one adjustable parameter, the inverse of which is both mean and standard deviation. Recent works have overcome this limitation by using more flexible distributions.

2.2.2 Weibull distribution

The Weibull model (introduced by Waloddi Weibull in 1939) is an important generalization of the exponential model with two positive parameters. The second parameter in the model allows great flexibility of the model and different shapes of the hazard function. The convenience of the Weibull model for empirical work stems on the one hand from this flexibility and on the other from the simplicity of the hazard and survival function.

$$\begin{aligned}
\text{probability density function} \quad & f(x) = \lambda\gamma x^{\gamma-1} e^{-\lambda x^{\gamma}} \\
\text{survival function} \quad & S(x) = e^{-\lambda x^{\gamma}} \\
\text{hazard function} \quad & \lambda(x) = \lambda\gamma x^{\gamma-1} \\
\text{cumulative hazard function} \quad & \Lambda(x) = \lambda x^{\gamma} \\
\text{mean} \quad & \mathbf{E}X = \lambda^{-\frac{1}{\gamma}} \Gamma(1 + \frac{1}{\gamma}) \\
\text{variance} \quad & \mathbf{V}(X) = \lambda^{-\frac{2}{\gamma}} \big(\Gamma(1 + \frac{2}{\gamma}) - \Gamma(1 + \frac{1}{\gamma})^2\big)
\end{aligned}$$

where Γ denotes the Gamma function with $\Gamma(k) = \int_0^\infty s^{k-1}e^{-s}\,ds$ $(k > 0)$. We abbreviate the distribution as $Weibull(\lambda, \gamma)$. In the case of $\gamma = 1$, the exponential distribution is obtained. The hazard function decreases monotonously from ∞ at time zero to zero at time ∞ for $\gamma < 1$, constant (exponential distribution) for $\gamma = 1$ and it monotonously increases from zero at time zero to ∞ at time ∞ for $\gamma > 1$.

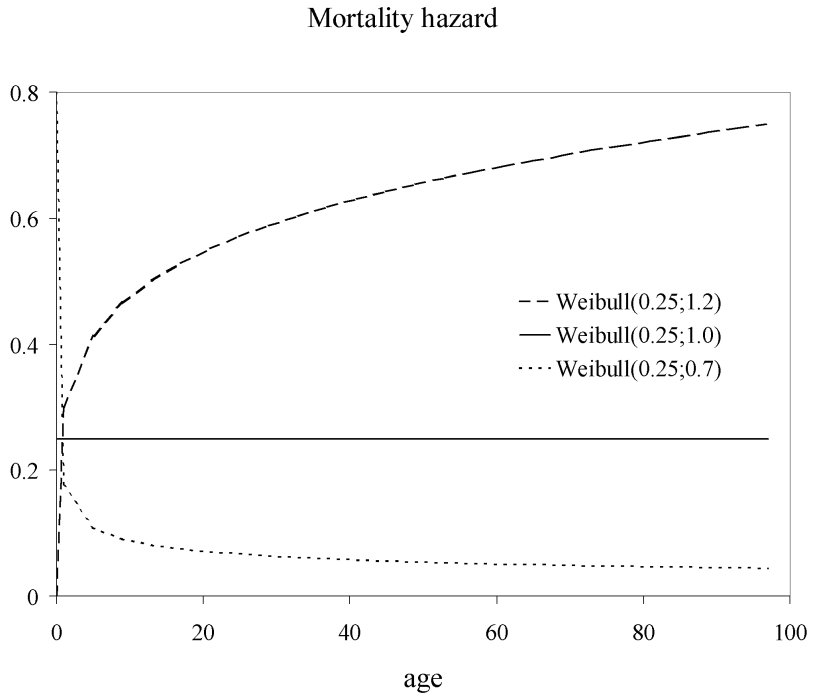

Figure 2.1: Weibull hazard functions with different shape parameters.

If $X \sim Weibull(\lambda, \gamma)$, then it holds that $cX \sim Weibull(\lambda c^{-\gamma}, \gamma)$, when c is a positive constant. Furthermore, the minimum of n i.i.d. variables from this distribution is $Weibull(n\lambda, \gamma)$ (minimum-stable distribution). The Weibull distribution can also be generated as the limiting distribution of the minimum of a sample from a continuous distribution with support on $[0, u)$ for some u $(0 < u < \infty)$. This extreme value character makes the Weibull distribution appropriate for the distribution of individual time to death, because there are different causes of death which compete with each other and the first one to strike will kill the individual.

The Weibull hazard has been theoretically derived for cancer incidence by Pike (1966), but it is unknown whether it has relevance for other diseases. The Weibull distribution is inappropriate when the hazard rate is indicated to be unimodal or bathtub-shaped. A generalization of the Weibull distribution to include such kind of shapes was proposed by Mudholkar et al. (1996).

2.2.3 Gompertz distribution

In 1825 the British actuary Benjamin Gompertz made a simple but important observation that a law of geometrical progression pervades large portions of different tables of mortality for humans. The simple formula he derived describing the exponential rise in death rates between sexual maturity and old age is commonly referred to as the Gompertz equation–a formula that remains a valuable tool in demography and in other scientific disciplines. Gompertz's observation of a mathematical regularity in human life tables led him to believe in the presence of a law of mortality that explained why common age patterns of death exist. It has been widely used, especially in actuarial and biological applications and in demography. A random variable follows a Gompertz distribution with parameters $a > 0$ and $b > 0$ ($X \sim Gompertz(a, b)$), if the following relations hold:

$$\begin{aligned} \text{probability density function} \quad & f(x) = ae^{bx}e^{-\frac{a}{b}(e^{bx}-1)} \\ \text{survival function} \quad & S(x) = e^{-\frac{a}{b}(e^{bx}-1)} \\ \text{hazard function} \quad & \lambda(x) = ae^{bx} \\ \text{cumulative hazard function} \quad & \Lambda(x) = \frac{a}{b}(e^{bx} - 1) \end{aligned}$$

The hazard function is increasing from a at time zero to ∞ at time ∞. The model can be generalized to the Gompertz-Makeham distribution by adding a constant to the hazard: $\lambda(x) = ae^{bx} + c$.

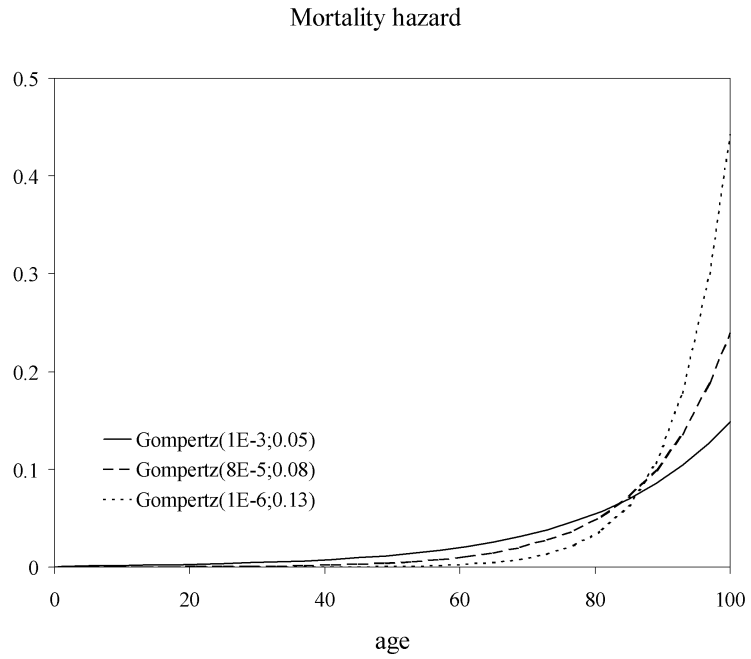

Figure 2.2: Gompertz hazard functions with different parameters.

2.2.4 Log-logistic distribution

An alternative model to the Weibull distribution is the log-logistic distribution. The log-logistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped, that is it initially increases and then decreases. The distribution imposes the following functional forms on the density, survival, hazard and cumulative hazard function:

$$\begin{aligned}
\text{probability density function} \quad & f(x) = abx^{b-1}(1+ax^b)^{-2} \\
\text{survival function} \quad & S(x) = (1+ax^b)^{-1} \\
\text{hazard function} \quad & \lambda(x) = \frac{abx^{b-1}}{1+ax^b} \\
\text{cumulative hazard function} \quad & \Lambda(x) = \ln(1+ax^b)
\end{aligned}$$

The general shape of the hazard function of a log-logistic distribution is very similar to that of the log-normal distribution considered later. The log-logistic distribution can be obtained as a mixture of Gompertz distributions with a gamma distributed mixture variable with mean and variance equal to one.

2.2.5 Gamma distribution

The gamma distribution includes the exponential distribution as a special case. The gamma distribution is of limited use in survival analysis because the gamma models do not have closed form expressions for survival and hazard functions. Both include the incomplete gamma integral

$$I_k(x) = \frac{\int_0^x s^{k-1}e^{-s}\,ds}{\Gamma(k)}.$$

Consequently, traditional maximum likelihood estimation is difficult and requires the calculation of such incomplete gamma integrals, which imposes additional numerical problems in parameter estimation. A random variable X is gamma distributed with parameter k and λ ($X \sim \Gamma(k,\lambda)$), if the following holds:

$$\begin{aligned}
\text{probability density function} \quad & f(x) = \frac{\lambda^k x^{k-1}e^{-\lambda x}}{\Gamma(k)} \qquad k,\lambda > 0 \\
\text{survival function} \quad & S(x) = 1 - I_k(\lambda x)
\end{aligned}$$

$$\begin{aligned}
\text{hazard function} \quad & \lambda(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(1 - I_k(\lambda x))\Gamma(k)} \\
\text{mean} \quad & \mathbf{E}X = \frac{k}{\lambda} \\
\text{variance} \quad & \mathbf{V}(X) = \frac{k}{\lambda^2} \\
\text{Laplace transform} \quad & \mathbf{L}(s) = \mathbf{E}e^{-Xs} = (1 + \frac{s}{\lambda})^{-k}
\end{aligned}$$

If $k = 1$, the gamma distribution is reduced to the exponential distribution. With integer k, the gamma distribution is sometimes called a special Erlangian distribution. It can be derived as the distribution of waiting time to the k-th emission from a Poisson source with intensity parameter λ. Consequently, the sum of k independent exponential variates with parameter λ has a gamma distribution with parameters k and λ (see Example 1) and can be used to model life times of technical systems with repeated repairing after failure.

Example 1 *Let $X_1, X_2, \ldots, X_k$ denote k independent exponential distributed random variables with $X_i \sim Exp(\lambda)$ $(i = 1, \ldots, k)$ and introduce X by $X = X_1 + \ldots + X_k$. Then it holds that*

$$\begin{aligned}
\mathbf{L}(s) &= \mathbf{E}e^{-Xs} = \mathbf{E}e^{-(X_1 + \ldots + X_k)s} \\
&= \prod_{i=1}^{k} \mathbf{E}e^{-X_i s} = \prod_{i=1}^{k} \mathbf{E}(1 + \frac{s}{\lambda})^{-1} = (1 + \frac{s}{\lambda})^{-k}.
\end{aligned}$$

An extension of this idea was used by Aalen (1992) dealing with the compound Poisson distribution (see Section 3.6). Despite the fact that the gamma distribution is of limited value as a life time distribution because of the problems mentioned above, the gamma distribution is a widely used frailty (mixture) distribution because of some neat mathematical features. It is mathematically tractable and readily computable. It is a flexible distribution that takes on a variety of different shapes as k varies. Furthermore, frailty cannot be negative and the gamma distribution is, along with the log-normal and Weibull distribution, one of the most commonly used distributions to model positive random variables. The assumption that frailty is gamma distributed yields some useful mathematical results, including that the frailty among survivors of any age x is again gamma distributed and frailty among those who die at any time x, too (see Vaupel et al., 1979). We will discuss other properties of the gamma distribution as frailty distribution later in more detail.

2.2.6 Log-normal distribution

In the log-normal model ($X \sim logN(m, s^2)$), the natural logarithm $\ln(X)$ of the lifetime X is assumed to be normally distributed ($ln(X) \sim N(m, s^2)$). A log normal distribution results when the variable is the product of a large number of independent, identically distributed variables in the same way that a normal distribution results when the variable is the sum of a large number of independent, identically distributed variables. The survival and hazard functions include the incomplete normal integral

$$\Phi(x) = \int_{-\infty}^{x} \phi(s)\, ds,$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ denotes the probability density function of a standard normal distribution. Consequently,

$$\begin{aligned} \text{probability density function} \quad & f(x) = \frac{1}{\sqrt{2\pi} s x} e^{-\frac{(\ln(x)-m)^2}{2s^2}} \\ \text{mean} \quad & \mathbf{E}X = e^{m+\frac{s^2}{2}} \\ \text{variance} \quad & \mathbf{V}(X) = e^{2m+s^2}(e^{s^2} - 1) \end{aligned}$$

The log-normal distribution may be convenient to use with non-censored data, but when this distribution is applied to censored data, the computations quickly become formidable. Unfortunately, the hazard function has a strange form: it has value zero at $x = 0$, increases to a maximum and then decreases, approaching zero as x heads to infinity. Because of the decreasing form of the hazard function for older ages, the distributions seem implausible as a lifetime model in most situations. Nevertheless, it makes sense if interest is focused on time periods of younger ages. Despite its unattractive features, the log-normal distribution has been widely used as failure distribution in diverse situations, such as the analysis of electrical insulation or time to occurrence of lung cancer among smokers.

Furthermore, the log-normal distribution has often been used as a frailty (mixing) distribution. Especially in the context of unobserved normal distributed covariates in the Cox model, the log-normal frailty distributions provides an appealing interpretation of the model. Unfortunately, the Laplace transform is intractable, and therefore numerical integration is needed for probability results. The log-normal distributions are in practice very close to the inverse Gaussian distributions.

2.3 Censoring

Censoring is what distinguishes survival analysis from other fields of statistics. Basically, a censored observation contains only partial information about the variable of interest. There are different types of censoring, here we consider type I right censoring only.

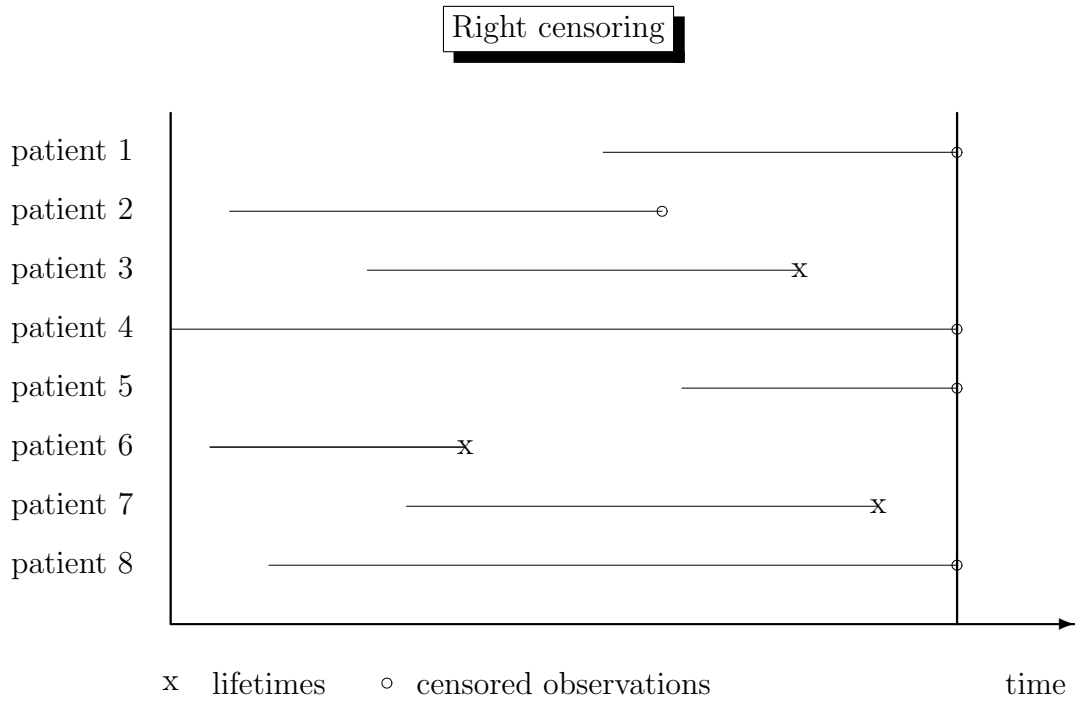

Figure 2.3: Right censored lifetimes of patients in an artificial clinical trial.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. survival times with cumulative distribution function F and let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. censoring times with cumulative distribution function G. Throughout the thesis, we assume that F and G are absolutely continuous. Furthermore, let f and g be probability density functions with respect to F and G. We can only observe $(T_1, \Delta_1), (T_2, \Delta_2), \ldots, (T_n, \Delta_n)$, where $T_i = \min\{X_i, Y_i\}$ and

$$\Delta_i = \begin{cases} 1 & : \text{ if } X_i \leq Y_i, \quad \text{that is, } T_i \text{ is not censored} \\ 0 & : \text{ if } Y_i < X_i, \quad \text{that is, } T_i \text{ is censored.} \end{cases} \tag{2.3}$$

Random censoring arises especially in medical applications, for example in clinical trials or epidemiological studies. Here, patients may enter the study at different times; then each is treated with one of several possible drugs or therapies. We are interested in observing their lifetimes, but censoring occurs in one of the following forms:

- Loss to follow up. The patient may move elsewhere; he is never seen again.
- Drop out. The treatment may have such strong side effects that it is necessary to stop the therapy. Or the patient may refuse to continue the treatment.
- Termination of the study. The study ends at a predefined point of time. This type of censoring is called administrative censoring.

We use X and Y, with no subscripts, as shorthand for all the X_i and Y_i variables.

Assumption: Lifetimes X and censoring times Y are independent. A weaker condition is to assume that censoring is non-informative.
Remark: The cumulative distribution function of the non-censored observations (while discarding the censored observations of the sample) is not F!

$$\begin{aligned}\mathbf{P}(T<t,\Delta=1) &= \mathbf{P}(X<t,X\le Y)=\int\limits_{x<t,x\le y} f(x)g(y)\,dx\,dy \\ &= \int\limits_{x<t} f(x)\Big(\int\limits_{x\le y} g(y)\,dy\Big)\,dx=\int\limits_{x<t} f(x)(1-G(x))\,dx \ne F(t) \quad (2.4)\end{aligned}$$

Theorem 1 *(Wienke (1996)) The probability density function of the data* (T,Δ) *takes the form*

$$f(t,\delta)=(f(t)(1-G(t)))^{\delta}(g(t)(1-F(t)))^{1-\delta}. \quad (2.5)$$

Proof: Denote by H_0 and H_1 sub-distribution functions and by h_0 and h_1 sub-densities. It holds (see (2.4)) $H_1(t)=\mathbf{P}(T<t,\Delta=1)=\int\limits_{x<t} f(x)(1-G(x))\,dx$. Furthermore,

$$\begin{aligned}H_0(t) &= \mathbf{P}(T<t,\Delta=0)=\mathbf{P}(Y<t,Y<X) \\ &= \int\limits_{y<t,y<x} f(x)g(y)\,dx\,dy=\int\limits_{y<t} g(y)\Big(\int\limits_{y<x} f(x)\,dx\Big)\,dy=\int\limits_{y<t} g(y)(1-F(y))\,dy\end{aligned}$$

Consequently,

$$\begin{aligned}h_0(t)&=H_0'(t)=g(t)(1-F(t)) \\ h_1(t)&=H_1'(t)=f(t)(1-G(t))\end{aligned}$$

$$\begin{aligned}f(t,\delta) &= \delta h_1(t)+(1-\delta)h_0(t)=(h_1(t))^{\delta}(h_0(t))^{1-\delta} \\ &= (f(t)(1-G(t)))^{\delta}(g(t)(1-F(t)))^{1-\delta}.\end{aligned}$$

This completes the proof. ■

Remark: If the censoring is non-informative, meaning if the censoring distribution does not contain any information about the parameters of interest, then it does not enter the likelihood function:

$$L(t,\delta) = f(t)^{\delta}(1-F(t))^{1-\delta} = \delta f(t) + (1-\delta)(1-F(t)). \tag{2.6}$$

As pointed out in Theorem 1, the density function under independent right censoring is

$$f(t,\delta) = \delta f(t)(1-G(t)) + (1-\delta)g(t)(1-F(t)).$$

The following example considers the case of dependent censoring. It turns out that the likelihood function under dependent censoring is a composition of derivatives of the joint survival function of life times and censoring times.

Example 2 *Denote by* (T,Δ), $T = \min\{X,Y\}$, $\Delta = 1(X \leq Y)$ *censored observations under the assumption of dependent censoring. Let* $S(x,y)$ *and* $f(x,y)$ *be the joint survival and probability density function of* X *and* Y*, respectively. Consequently, the sub-distribution functions can be derived as follows:*

$$\begin{aligned} H_1(t) &= \mathbf{P}(T<t, \Delta=1) = \mathbf{P}(X<t, X\leq Y) \\ &= \int\int_{\{x<t, x\leq y\}} f(x,y)\,dx\,dy = -\int_0^t S_x(x,x)\,dx. \end{aligned}$$

This implies that the sub-density of a non-censored $(\delta = 1)$ *observation is a derivative of the sub-distribution function:*

$$h_1(t) = H_1'(t) = -S_x(t,t) = -\frac{\partial S(x,y)}{\partial x}|_{x=t,y=t}.$$

Similar calculations yield the sub-distribution and sub-density functions in the case of a censored observation $(\delta = 0)$*:*

$$\begin{aligned} H_0(t) &= \mathbf{P}(T<t, \Delta=0) = \mathbf{P}(Y<t, Y<X) \\ &= \int\int_{\{y<t, y<x\}} f(x,y)\,dxdy = -\int_0^t S_y(y,y)\,dy \end{aligned}$$

and

$$h_0(t) = H_0'(t) = -S_y(t,t) = -\frac{\partial S(x,y)}{\partial y}|_{x=t,y=t}.$$

The likelihood function is a composition of the sub-density functions:

$$L(t,\delta) = -\delta S_x(t,t) - (1-\delta)S_y(t,t).$$

2.4 Truncation

Now we shall take truncation into account. We restrict our treatment to the most common type of truncation, that is left truncation. Furthermore, let us assume that truncation is non-random. Left truncation occurs when individuals come under observation only some known time after the natural origin of the event under study. That is, had the individual failed before the truncation time in question, that individual would not have been recorded. For example, the second patient in Figure 2.4 cannot be observed, because

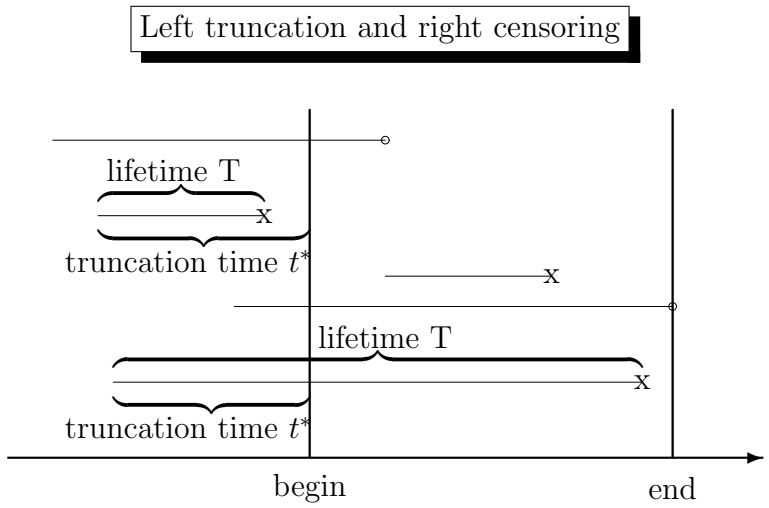

Figure 2.4: Left truncated and right censored lifetimes.

he dies before the study started. That means, the person carrying out the study would not be aware of that observation. In other words, truncation is sampling from a conditional distribution. Denote observations by (T^*, Δ^*) with $\mathcal{L}(T^*, \Delta^*)$=$\mathcal{L}((T, \Delta)|T \geq t^*)$ with t^* as known (non-random) truncation time. So we get

$$\begin{aligned} \mathbf{P}(T^* \geq t, \Delta^* = \delta) &= \mathbf{P}(T \geq t, \Delta = \delta | T \geq t^*) \\ &= \frac{\mathbf{P}(T \geq t, \Delta = \delta)}{\mathbf{P}(T \geq t^*)} = \frac{\mathbf{P}(T \geq t, \Delta = \delta)}{(1 - F(t^*))(1 - G(t^*))} \end{aligned}$$

and the density of the non-random left truncated and right censored observations:

$$\begin{aligned} f(t, \delta, t^*) &= \frac{h(t, \delta)}{(1 - F(t^*))(1 - G(t^*))} \\ &= \delta \frac{f(t)(1 - G(t))}{(1 - F(t^*))(1 - G(t^*))} + (1 - \delta) \frac{g(t)(1 - F(t))}{(1 - F(t^*))(1 - G(t^*))}. \end{aligned}$$

Hence, the likelihood function in the case of independent (non-informative) right censoring and non-random left truncation at time t^* can be written as

$$L(t, \delta, t^*) = \delta \frac{f(t)}{1 - F(t^*)} + (1 - \delta) \frac{1 - F(t)}{1 - F(t^*)}.$$

2.5 Non-parametric and semi-parametric models

For parametric inference, it is necessary to make assumptions about the distribution of failure times. In some circumstances this makes sense; for example, when additional information about the nature of the ageing or disease process is available from other experiments. When one is interested in avoiding such assumptions, it is common to use non-parametric models. The simplest non-parametric estimate of a distribution function is the empirical distribution function. That means, even in the case of continuous distributions we estimate it by a discrete distribution. Major steps in the development of appropriate methods in survival analysis (in censored observations) were the introduction of the Kaplan-Meier estimator (Kaplan and Meier (1958)) and the proportional hazards model (Cox 1972).

A key question is whether we should use a parametric model as described above or a non-parametric one. An advantage of non-parametric models is their good fit and the resulting ability to deal with any distribution without any additional assumptions. But there is a high price to pay. First, non-parametric methods need much more data to get reasonable results. Second, it is very hard to get estimates of the hazard function, which is often the most interesting and relevant information. For this, it is necessary to smooth out the discrete point masses of the Kaplan-Meier estimator, for example by kernel function smoothing. Otherwise, parametric models often allow closed form expressions of the hazard function (depending on the chosen model) and other characteristics of the failure distribution. Furthermore, parametric models can be described by the values of a few parameters and they often give good results even in the case of moderate sample size.

2.5.1 Kaplan-Meier estimator

A useful way of characterizing the survival in a homogeneous group of individuals is to compute and graph the empirical survival function. If there are no censored observations in the sample, the empirical survival function at time t is the ratio of survivors at time t and the sample size n. This step function decreases by $\frac{1}{n}$ just after each observed failure (for ease of presentation we assume no ties here). When dealing with censored data, a methodology for handling this with convenience is required. Remember that we observe the pairs $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$, where $T_i = \min\{X_i, Y_i\}$ and $\Delta = 1(X_i \leq Y_i)$. Let $T_{(1)} < T_{(2)} < \ldots < T_{(n)}$ be the order statistics of $T_1, T_2, \ldots, T_n$, and with an abuse of notation define $\Delta_{(i)}$ to be the value of Δ which is associated with $T_{(i)}$, that is, $\Delta_{(i)} = \Delta_j$ if $T_{(i)} = T_j$. Note that the $\Delta_{(1)}, \Delta_{(2)}, \ldots, \Delta_{(n)}$ are not ordered. In the sequel, we will

use capital letters to denote random variables, for example (T_i, Δ_i), and small letters for their real-value realizations: (t_i, δ_i). The Kaplan-Meier estimator (also called product limit estimator) was introduced by Kaplan and Meier (1958) as

$$\hat{S}(t) = \prod_{i:T_{(i)}<t} \Big(1 - \frac{\Delta_{(i)}}{n-i+1}\Big).$$

This function is a decreasing step function, with changes only at times of death. A slightly problematic point is that $\hat{S}$ never reduces to zero if the largest observation is a censored one, i.e. $\Delta_{(n)} = 0$. In this case, $\hat{S}$ is usually left unspecified for $t > t_{(n)}$.

2.5.2 Proportional hazards model

The models presented above deal with the simplest case of i.i.d. data. This implies a homogeneous population. However, in most practical applications the population under study is not homogeneous. For example, individuals in epidemiological studies may differ in age (if age is not used as time scale), gender, socio-economic status, education, blood pressure, body mass index, smoking habits, nutrition, physical activity level, heart rate and so forth. Maybe some of these covariates are of special interest, such as the effect of a treatment in a clinical trial, or they are nuisance parameters which influence the variable lifetime. The proportional hazards model is a regression model with duration as dependent variable. It allows inclusion of information about known (observed) covariates in models of survival analysis and is the most applied model in this area. Statistical strategies for prediction are similar to those utilized in ordinary regression. However, the details for regression techniques in survival analysis are unique.
Let $\lambda(t, X)$ denote the hazard of an individual at time (or age) t with covariate vector $X = (X_1, \ldots, X_k)$. The proportional hazards model (Cox 1972) specifies that

$$\lambda(t, X) = \lambda_0(t) g(X) \tag{2.7}$$

where $\lambda_0(t)$ is the baseline hazard function and g some positive function. The model assumes a baseline hazard (risk of death or other event) that is common to all the individuals in the study population. The parameters of primary interest are contained in $g(X) = g(\beta, X)$, often

$$g(X) = e^{\beta^T X} = e^{\sum_{i=1}^{k} \beta_i X_i}.$$

In this model, covariates act multiplicatively on the baseline hazard, adding additional risks on an individual basis, as determined by the individuals' prognostic information. This gives the model a simple and easily understood interpretation. The main idea behind it is the separation of the age or time effect in the baseline hazard function on one

side and the effect of the covariates in an exponential term on the other side. In essence this assumption says that the hazard $\lambda(t)$ of failure at time t is related to individuals or groups of individuals by a proportionality constant which does not depend on t. The simple two-sample situation is obtained by letting X_1 be 0 or 1 ($k = 1$), depending on group membership. In this case, the method is truly non-parametric and e^{β} is the hazard ratio for mortality between the two groups. However, when X takes more than two values, a parametric form of g is required. Now inference is dependent on that form but still independent of $\lambda_0(t)$, and one speaks of a semi-parametric model. The conditional survival function for T given X is

$$S(t|X) = S_0(t)^{e^{\sum_{i=1}^{k} \beta_i X_i}},$$

where $S_0(t) = e^{-\int_0^t \lambda_0(s)\,ds}$ and the $\beta' s$ are unknown regression parameters. That means the survival function of an individual with covariates X is the baseline survival function raised to a power. The class of distributions generated by this procedure is sometimes called Lehmann alternatives.

Two different approaches are possible. In the parametric case the baseline hazard is chosen in the class of parametric lifetime distributions, for example as Weibull or Gompertz-Makeham. But the model also works without any specification of the baseline hazard function. In this second case, the model is natural and sufficiently flexible to suit many purposes. Since $e^{\sum_{i=1}^{k} \beta_i X_i}$ is always positive, the individual hazard $\lambda(t, X)$ is automatically non-negative for all t and all $\beta' s$. One additional reason for considering this model is that censoring and competing risks are relatively easily accommodated within this formulation and in particular the technical problems of statistical inference have a simple solution when the baseline hazard is arbitrary. Cox (1975) suggests using a partial likelihood approach in the case of arbitrary baseline hazard. Inference for the Cox estimator is almost exclusively based on asymptotic results (Andersen and Gill 1982). The validity of these large sample properties have been found acceptable with moderately large sample sizes, moderate amount of censoring and balanced covariate distributions. However it frequently occurs that covariates have very skew distributions, for example when only a small fraction of the individuals are exposed to the risk factor of interest. It is also very common that a large fraction of lifetimes is censored. Especially in large cohort studies analyzing the effect of a rare exposition on an event, the number of exposed cases may be very small. One may then question the validity of inference based on large sample results. Samuelsen (2003) investigates the possibilities and limitations of exact inference in the proportional hazards model.

Note that covariate X could vary with time, but this is beyond the scope of this thesis.

Chapter 3

Univariate frailty models

The structure, properties, and applicability of survival models to practical problems depend on the nature of time-to-event data, ancillary information about influential factors and processes, and the aims of the studies. This chapter focuses on the analysis of univariate data, i.e. single-spell data on unrelated individuals.

Ordinary survival models deal with the simplest case of independent and identically distributed data. This is based on the assumption that the study population is homogeneous. But it is a basic observation of medical statistics that individuals differ greatly. So do the effects of a drug, or the influence of various explanatory variables. This heterogeneity is often referred to as variability and it is generally recognized as one of the most important sources of variability in medical and biological applications. The subject of this chapter is heterogeneity in survival analysis. This heterogeneity may be difficult to assess, but it is nevertheless of great importance. In recent decades, a large amount of papers on 'frailty models' have appeared. The key idea of these models is that individuals have different frailties, and that the most frail will die earlier than the lesser frail. Consequently, systematic selection of robust individuals takes place, which distorts what is observed. When mortality rates are estimated, one may be interested in how they change over time or age. Quite often, they rise at the beginning of the observation period, reach a maximum and then decline (unimodal intensity) or level off. This, for example, is typical for death rates of cancer patients, meaning that the longer the patient lives beyond a certain time, the better his or her chances of survival are. It is likely that unimodal intensities are often the result of selection and that they do not reflect an underlying development on the individual level. The population intensity starts to decline simply because the high-risk individuals have already died. The intensity of a given individual might well continue to increase.

If covariates are known, they can be included in the analysis, for example by using the

proportional hazards model as described above. But it is nearly always impossible to include all important risk factors, perhaps because we have little or no information on the individual level (this applies for example to population studies, where often the only known risk factors are sex and age). Furthermore, we may not know the relevance of the risk factor or even that the factor exists. In other cases it may be impossible to measure the risk factor without great financial cost or time effort. In such cases, two sources of variability in duration data are useful to consider: variability accounted for by observable risk factors (which is thus theoretically predictable) and heterogeneity caused by unknown covariates and which is therefore theoretically unpredictable even when all relevant information at that time is known. It is the latter which is of specific interest here, and the subject of observable covariates is treated here only for completeness. As Hougaard (1991) pointed out, there are advantages in considering these two sources of variability separately: heterogeneity explains some 'unexpected' results or gives an alternative explanation of some results, for example non-proportional or decreasing hazards. If some individuals experience a higher risk of failure, then the remaining individuals at risk tend to form a more or less selected group with lower risk. An estimate of the individual hazard rate without taking into account the unobserved frailty will thus underestimate the hazard function to an increasingly greater extent as time goes by.

To be aware of such selection effects, mixture models could be used. That means the population is assumed to be a mixture of individuals with (at least partly unknown) different risks. The non-observable risks are described by the mixture variable, which is called frailty. It is a random variable, which follows some distribution. Furthermore, as the value is unknown, it has to be integrated out. The precise nature of the relationship between individual and population ageing depends on the distribution of frailty among individuals. Different choices of distributions for the unobserved covariates are possible, including binary, gamma and log-normal, which show both qualitative and quantitative differences. Especially, the variance of the frailty distribution determines the degree of heterogeneity in the study population.

To address the problem of heterogeneity in a population resulting from unobserved covariates, Vaupel et al. (1979) suggested a random effects model for durations. They introduced the notion of frailty and applied it to population data. The classical and mostly applied frailty model assumes a proportional hazards model that is conditional on the random effect (frailty). To be more specific, the hazard of an individual depends additionally on an unobservable, age-independent random variable Z, which acts multiplicatively on the baseline hazard function λ_0

$$\lambda(t, Z) = Z\lambda_0(t). \tag{3.1}$$

Here, Z is considered as a random mixture variable, varying across the population. Note that a scale factor common to all subjects in the population maybe absorbed into the baseline hazard function $\lambda_0(t)$, so that frailty distributions are standardized to $\mathbf{E}Z = 1$. The variance parameter $\sigma^2 = \mathbf{V}(Z)$ is interpretable as a measure of heterogeneity across the population in baseline risk. When σ^2 is small, then the values of Z are closely concentrated around one. If σ^2 is large, then values of Z are more dispersed, inducing greater heterogeneity in the individual hazards $Z\lambda_0(t)$. Frailty increases the individual risk and is sometimes called liability or susceptibility in other settings. All individuals, apart from an individual constant Z are assumed to follow the same mortality pattern. What may be observed in a population is not the individual hazard, but the net result for a number of individuals with different values of the random variable Z.
The following consideration shows that model (3.1) make sense. Let $\lambda(t, Z)$ be an individual hazard. By Taylor series expansion

$$\lambda(t, Z) = \lambda(t, 0) + Z\lambda'(t, 0) + o(Z)$$

where $o(Z)$ denotes the terms containing Z of higher order than one. By omitting these terms we get

$$\lambda(t, Z) \approx \lambda(t, 0) + Z\lambda'(t, 0).$$

Making the natural assumption that zero frailty (susceptibility) yields zero mortality (background mortality is not considered here) it holds that

$$\lambda(t, Z) \approx Z\lambda_0(t)$$

where $\lambda_0(t) = \lambda'(t, 0) = \frac{\partial\lambda(t,z)}{\partial z}|_{z=0}$. Thus the underlying hazard $\lambda_0(t)$ is a partial derivative of the individual hazard with respect to frailty taken at point $Z = 0$.
It is quite clear that a multiplicative frailty model as (3.1) represents a rather simplified view of how heterogeneity might act. Nevertheless, simple mathematical models represent one way of understanding the consequences of heterogeneity. The assumptions that the frailty is age-independent and that it acts multiplicatively on the underlying baseline hazard function are in principle arbitrary, but they have been taken as the basis for much subsequent work on unobserved heterogeneity in survival analysis.
Only for completeness do we want to mention other cases of frailty models which are not based on the proportional hazards assumption. For example the additive frailty model, where the frailty acts additively on the baseline hazard function. For more details in this very uncommon case, see Rocha (1996). Proportional odds frailty models are considered in Lam et al. (2002) and Lam and Lee (2004). Murphy et al. (1997) pointed out the

link between proportional odds models and frailty models. AFT frailty models are dealt with, for example, by Anderson and Louis (1995), Keiding et al. (1997), Klein et al. (1999), Schnier et al. (2004) and Chang (2004).
In the natural way, it is possible to introduce known covariates into the model (3.1):

$$\lambda(t, Z, X) = Z\lambda_0(t)e^{\sum_{i=1}^{k} \beta_i X_i} = Z\lambda_0(t)e^{\beta^T X} \tag{3.2}$$

with $X = (X_1, \ldots, X_k)$ and $\beta = (\beta_1, \ldots, \beta_k)$ as covariates and regression parameters, respectively. Consequently, a frailty model is a generalization of the well-known proportional hazards model. For the sake of simplicity, we will restrict our treatment to the model (3.1) sometimes in order to focus on the main ideas of frailty models. Let $S(t|Z)$ denote the survival function of an individual conditional on the frailty Z, that means

$$S(t|Z) = e^{-\int_0^t \lambda(s,Z)\,ds} = e^{-Z\int_0^t \lambda_0(s)\,ds} = e^{-Z\Lambda_0(t)}, \tag{3.3}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$ denotes the cumulative baseline hazard function and equation (3.3) is a generalization of relation (2.2). Up to now, the model has been described at the individual level. However, this individual model is not observable. Consequently, it is necessary to consider the population level. The survival function of the total population is the mean of individual survival functions with respect to the frailty distribution. It can be viewed as the survival function of a randomly drawn individual, and corresponds to what can actually be observed. The survival and density functions as well as the mean and the variance of the frailty distribution are characterized by the Laplace transform of the frailty distribution

$$\begin{aligned} S(t) &= \mathbf{E}S(t|Z) = \mathbf{E}e^{-Z\Lambda_0(t)} = \mathbf{L}(\Lambda_0(t)) \\ f(t) &= -\lambda_0(t)\mathbf{L}'(\Lambda_0(t)) \\ \mathbf{E}Z &= -\mathbf{L}'(0) \\ \mathbf{V}(Z) &= \mathbf{L}''(0) - (\mathbf{L}'(0))^2 \end{aligned} \tag{3.4}$$

which uses relation (3.3) and underlines the important role of the Laplace transform in frailty models. Thus, if the latter has a simple form, performing this calculation is easy. The connection with the Laplace transform was first pointed out and exploited by Hougaard (1984, 1986a, 1986b). It follows that, when seeking distributions for the frailty variable Z, it is natural to use those which have explicit Laplace transforms. This allows for the use of traditional maximum-likelihood methods for parameter estimation.

Theorem 2 *(Vaupel et al. (1979)) Assume a frailty model given by (3.1). The hazard function of the population* $\lambda(t) = \frac{f(t)}{S(t)}$ *is generally* $\lambda(t) = \mathbf{E}(\lambda(t, Z)|T > t)$*, or more specifically,*

$$\lambda(t) = \int_0^\infty \lambda(t, z) f(z|T > t)\, dz = \lambda_0(t) \int_0^\infty z f(z|T > t)\, dz, \tag{3.5}$$

where $f(z|T > t)$ *represents the density of frailty among the survivors of age* t*.*

Proof: Starting with relation (3.1) we get

$$\begin{aligned}
\lambda(t, z) &= \frac{f(t|z)}{S(t|z)} = z\lambda_0(x) \\
f(t|z) &= z\lambda_0(t)S(t|z) \\
f(t, z) &= z\lambda_0(t)S(t|z)f_Z(z) \\
f(t) &= \lambda_0(t) \int_0^\infty zS(t|z)f_Z(z)\, dz
\end{aligned}$$

with f_Z as p.d.f. of the frailty distribution. Hence,

$$\lambda(t) = \frac{\lambda_0(t) \int_0^\infty zS(t|z)f_Z(z)\, dz}{S(t)}.$$

Because survival at age t implies an age of death greater than t, it holds that

$$\begin{aligned}
f(z, T > t) &= f_Z(z) \int_t^\infty z\lambda_0(s)S(s|z)\, ds = f_Z(z)S(t|z) \\
f(z|T > t) &= \frac{f_Z(z)S(t|z)}{S(t)}.
\end{aligned}$$

This completes the proof. ■

The hazard of the population is thus interpreted as the mean of individual hazards among the survivors, see (3.5). Frail individuals with high values of Z will tend to die first. Thus, the average frailty of the surviving cohort $\int_0^\infty z f(z|t)\, dz$ will decline with age. Consequently, equation (3.5) implies that the force of mortality for individuals increases more rapidly than for the cohort to which the individuals belong: In this sense, then, individuals 'age faster' than their cohort, see Figure 3.1. An illustrative example is given in Manton and Stallard (1981), describing different mortality patterns of heterogeneous populations.

To prove the existence of hidden heterogeneity in a population, stress experiments with laboratory subjects (e.g. insects, worms) were suggested. The idea of using stress data

for testing the heterogeneity assumption is based on the belief that the same exposure to stress produces different survival outcomes in heterogeneous and homogeneous populations after the exposure to stress. During the exposure, the heterogeneous population will experience higher selection (since the frailer individuals die first). Hence after exposure to stress the mortality rate of the stress group will be lower than that of the control group at the respective ages.

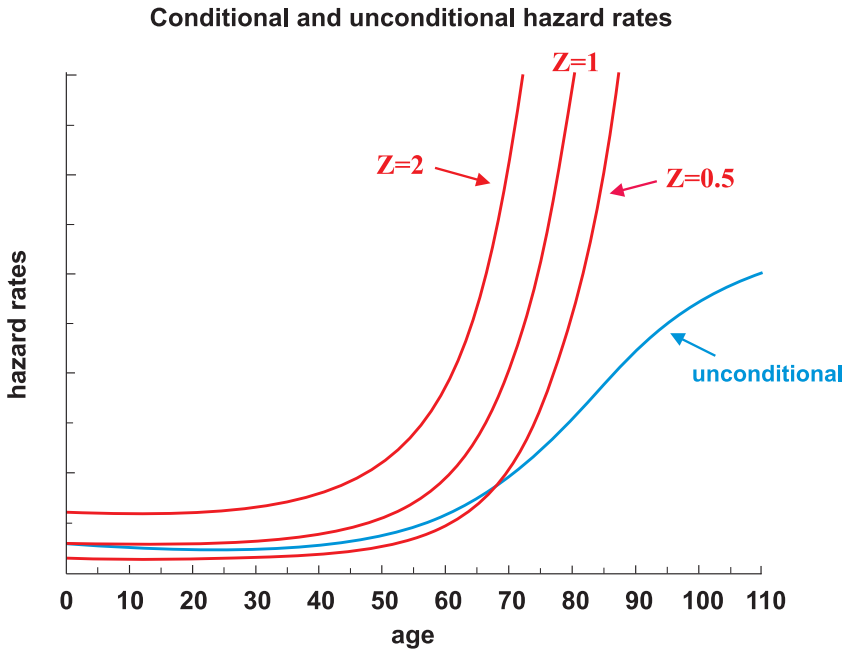

Figure 3.1: Conditional and unconditional hazard functions.

Under this model, the marginal hazard function can be expressed as the expectation of the hazard function (Nielsen et al., 1992) conditional on being at risk at time t and covariate value X; that is

$$\lambda(t|X) = \lambda_0(t)e^{\beta^T X}\mathbf{E}(Z|T \geq t, X).$$

Taking the expectation represents an averaging over all individuals in the population. Since the expectation is conditional on being at risk at time t, it constitutes averaging over a subset of the original population. An intriguing implication is that studies of human ageing based on cohort mortality data may be systematically biased or based on erroneous functional forms. The precise nature of the relationship between individual and cohort ageing depends on the distribution of frailty among individuals. Note that parametric specification of the frailty distribution is not necessary, often it is a matter of mathematical convenience. In the following sections, different frailty distributions are discussed. For a more detailed discussion of the notion of frailty from a medical and gerontological point of view see Morley et al. (2002).

3.1 Discrete frailty model

To give a simple explanation of univariate frailty models, the population under study is considered as consisting of two sub-populations with different risks. This could, for example, correspond to the presence/non-presence of a genotype.

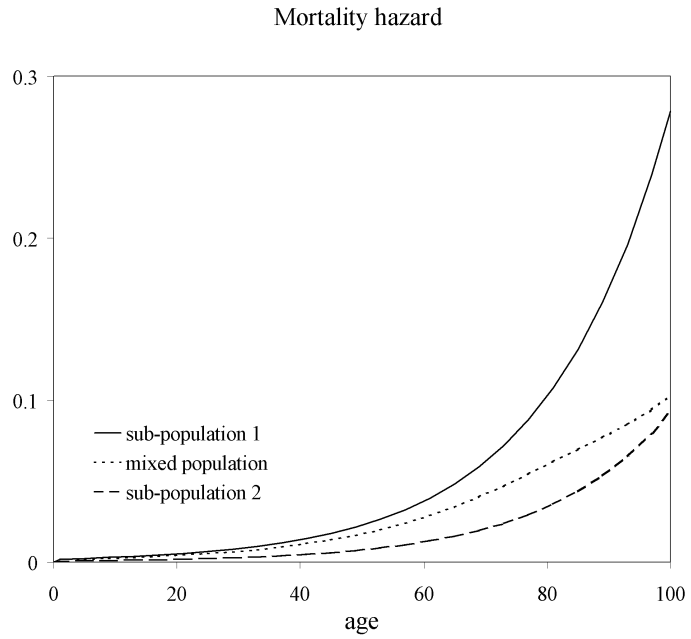

Figure 3.2: Mortality hazards of two populations with Gompertz hazard $\lambda_0(t) = ae^{bt}$ and their mixing. The values of the parameters are $a_1 = 1.875 \times 10^{-3}$, $a_2 = 0.625 \times 10^{-3}$ and $b_1 = b_2 = 0.05$ for the two sub-populations. At birth, the mixed population contains 70 % of sub-population 1 and 30 % of sub-population 2. The mortality of the mixed population is given by formula (3.6). The change over time in the proportion of the two sub-populations is shown in Figure 3.3.

If, however, no information on the genotype of each individual exists, it is impossible to include the genotype in the analysis and we instead have to consider the genotype as random. Let us assume that the proportion of individuals carrying the genotype is p. Consequently, a randomly selected person from the population carries the genotype with probability p. Naturally, this leads to an increased variation in the response time when compared to the case where the genotype is known. Assume $\mathbf{P}(Z = \eta_1) = p$ and $\mathbf{P}(Z = \eta_2) = 1 - p$. In this case, we find

$$S(t) = \mathbf{E}e^{-Z\Lambda_0(t)} = pe^{-\eta_1\Lambda_0(t)} + (1-p)e^{-\eta_2\Lambda_0(t)}$$

and

$$f(t) = p\eta_1\lambda_0(t)e^{-\eta_1\Lambda_0(t)} + (1-p)\eta_2\lambda_0(t)e^{-\eta_2\Lambda_0(t)}.$$

This implies

$$\lambda(t) = \frac{p\eta_1 e^{-\eta_1\Lambda_0(t)} + (1-p)\eta_2 e^{-\eta_2\Lambda_0(t)}}{pe^{-\eta_1\Lambda_0(t)} + (1-p)e^{-\eta_2\Lambda_0(t)}}\lambda_0(t). \tag{3.6}$$

Denote by $\pi_i(t)$ $(i = 1, 2)$ the size of the fraction of population i at time point t (e.g. $\pi_1(t)+\pi_2(t) = 1$) and $p_i = \pi_i(0)$. We are now interested in $\pi_i(t)$. These quantities change over time as a result of the selection process. Selection takes place if the life times of different subgroups follow different mortality patterns. To calculate $\pi_i(t)$, we introduce the conditional survival and hazard function, respectively. Conditioning is w.r.t. the event $\{I = i\}$, which means that the individual belongs to subpopulation i:

$$S_i(t) = \mathbf{P}(T > t|I = i) = e^{-\eta_i\Lambda_0(t)}$$

Now we can write the fraction in the following form:

$$\begin{aligned}
\pi_i(t) &= \mathbf{P}(I = i|T > t) \\
&= \frac{\mathbf{P}(T > t, I = i)}{\mathbf{P}(T > t)} \\
&= \frac{\mathbf{P}(T > t|I = i)\mathbf{P}(I = i)}{\mathbf{P}(T > t|I = 1)\mathbf{P}(I = 1) + \mathbf{P}(T > t|I = 2)\mathbf{P}(I = 2)} \\
&= \frac{p_i e^{-\eta_i\Lambda_0(t)}}{p_1 e^{-\eta_1\Lambda_0(t)} + p_2 e^{-\eta_2\Lambda_0(t)}}.
\end{aligned} \tag{3.7}$$

The binary frailty model is sometimes also called the two-point frailty model. Nickell (1979) used this model to account for heterogeneity in unemployment spell data. The study population was divided into the groups of motivated and non-motivated searchers of a new job. Furthermore, this frailty distribution was used for example by Vaupel and Yashin (1985) and Schumacher et al. (1987). A special case of binary frailty is dividing the population into a proportion who are at risk and a proportion who are never at risk. The terminology to describe the never-at-risk group varies from field to field, but includes 'long-term survivors' (Farewell, 1982) or 'cured' in epidemiology (Price and Manatunga, 2001), 'non-susceptibles' in toxicology (Pack and Morgan, 1990), 'stayers' in finite Markov transition models of occupational mobility (Blumen et al., 1955), the 'non-fecundable' in fertility models (Heckman and Walker, 1990), and 'non-recidivists' among convicted criminals (Schmidt and Witte, 1989; Maller and Zhou, 2002).

Obviously, the binary frailty model can be extended with no additional efforts to the finite discrete distribution of frailty Z (k-point distribution). In genetic studies, for example, the grouping factor is the genotype and in most human studies this is not directly observable. However, under a single-locus model with two alleles, A and a, three genotypes will occur, AA, Aa and aa. With allelic frequencies of p and $1 - p$ respectively, and under Hardy-Weinberg equilibrium the expected proportions of the

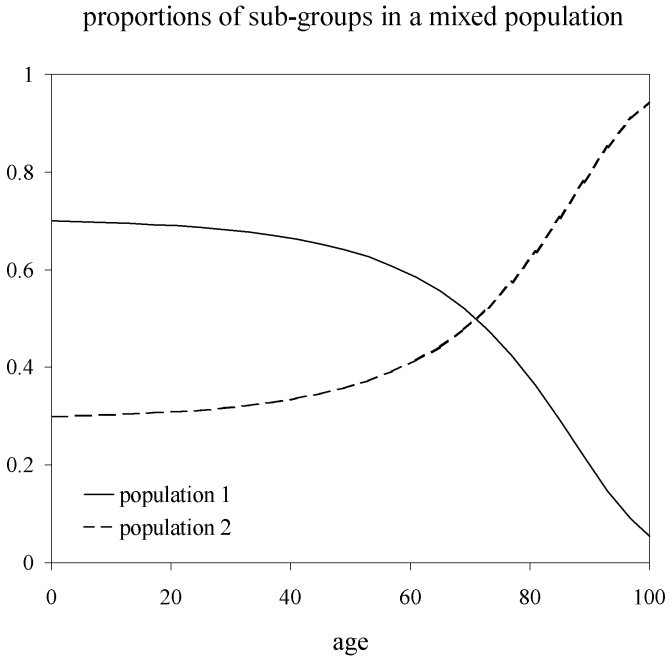

Figure 3.3: Age trajectories of the proportions of the two sub-populations in the mixing of the two Gompertz populations described in Figure 3.2. The proportions are given by (3.7). The first population (with higher mortality) dies out over time. For this reason the hazards of the mixed population and the second sub-population converge at older ages in Figure 3.2.

three unobservable genotypes are $p^2, 2p(1-p)$ and $(1-p)^2$. Thus, three latent classes with known expected proportions with parameter p to be estimated are assumed. If alleles are codominant then the location of the heterozygote Aa will be midway between AA and aa. If A is dominant, then genotypes AA and Aa will coincide and the model collapses to a two class problem. Where A is recessive, Aa and aa will coincide.

Andersen et al. (1999) found advantages of the random effect model when compared with the fixed effect model even in the case where it is known to which sub-group the individuals belong if the number of sub-groups in the total population is large and the number of individuals in each subgroup is only moderate. In this case the number of mass points of frailty is held fixed (parametric model). Macdonald (1999) applied two- and four-point frailty distributions to model the impact of genetics on insurance problems.

A more difficult case is the so-called non-parametric frailty distribution. In this case the frailty distribution is a discrete one, too. But the number of mass points is an additional parameter to be estimated (for example see Heckman and Singer, 1982b; dos Santos et al., 1995; and Lindstrom, 1996). It is common to use an iterative procedure, starting with one mass point and increasing the number until the likelihood fails to show a significant improvement. Often, two or three points of support suffice (Guo and Rodriguez, 1992).

3.2 Gamma frailty model

The gamma distribution has been widely applied as a mixture distribution (for example Greenwood and Yule, 1920; Vaupel et al., 1979; Congdon, 1995; dos Santos et al., 1995; Hougaard, 2000). From a computational and analytical point of view, it fits very well to failure data because it is easy to derive the closed form expressions of unconditional survival, cumulative density and hazard function. This is due to the simplicity of the Laplace transform. This is also the reason why this distribution has been used in most applications published to date. It is a flexible distribution that takes a variety of shapes as k varies: when $k = 1$, it is identical to the well-known exponential distribution; when k is large, it takes a bell-shaped form reminiscent of a normal distribution. Despite these advantages it is necessary to mention that no biological reason exists which makes the gamma distribution more preferable than other frailty distributions. Nearly all arguments in favor of the gamma distribution are based on mathematical and computational aspects. The paper by Abbring and van den Berg (2005) rationalizes the use of gamma distributions for frailties in time-to-event data analysis. The authors show that in a large class of (univariate and multivariate) frailty models, the distribution of the frailty among survivors converges to a gamma distribution under mild regularity assumptions.

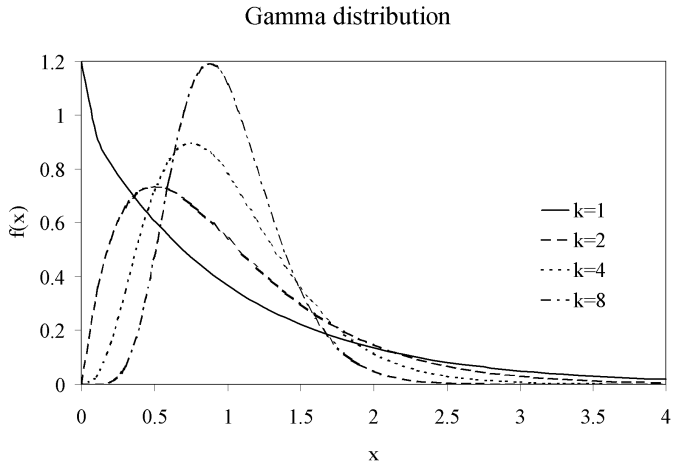

Figure 3.4: Probability density functions of different gamma distributions with mean 1 and variances 1, 0.5, 0.25, and 0.125

Frailty cannot be negative and the gamma distribution is, along with the log-normal distribution, one of the most commonly used distributions to model variables that are necessarily positive. Furthermore, it turns out that the assumption that frailty at the beginning of the follow-up is gamma distributed yields some useful mathematical results,

including

- Frailty among the survivors at any time t is gamma distributed with the same value of the shape parameter k as at birth or at begin of follow-up. The value of the second parameter, however, is now given by $\lambda + \Lambda_0(t)$, where $\Lambda_0(t)$ denotes the cumulative baseline hazard function.

- Frailty among those who die at any age t is also gamma distributed, with the same parameter $\lambda + \Lambda_0(t)$ as among those surviving to t but with shape parameter $k+1$.

In particular, it follows that the mean frailty among the deaths at age t is $\frac{\lambda+1}{\lambda+\Lambda(t)}$ compared to $\frac{\lambda}{\lambda+\Lambda(t)}$ among the survivors at the same age. This demonstrates the selection by death of the high risk individuals, e.g. the individuals with high values of frailty Z.
To make sure that the model is identifiable, it makes sense to use the parameter restriction $k = \lambda$ for the gamma distribution, which results in $\mathbf{E}Z = 1$. Denote by $\sigma^2 := \frac{1}{\lambda}$ the variance of the frailty variable. The unconditional hazard and survival function are

$$\lambda(t) = \frac{\lambda_0(t)}{1 + \sigma^2 \Lambda_0(t)} \quad \text{and } S(t) = \mathbf{L}(\Lambda_0(t)) = (1 + \sigma^2 \Lambda_0(t))^{-\frac{1}{\sigma^2}}. \tag{3.8}$$

This model was introduced by Vaupel et al. (1979).

Example 3 *Let $\lambda_0(t) = ae^{bx}$ (Gompertz) and frailty $Z \sim \Gamma(\frac{1}{\sigma^2}, \frac{1}{\sigma^2})$. Hence, the hazard and survival function on the population level are given by*

$$\lambda(t) = \frac{ae^{bt}}{1 + \sigma^2 \frac{a}{b}(e^{bt} - 1)} \text{ and } S(t) = (1 + \sigma^2 \frac{a}{b}(e^{bt} - 1))^{-\frac{1}{\sigma^2}}.$$

Figures 3.5 and 3.6 deal with one of the most important topics related to frailty models, so-called crossing over effects in the mortality hazards of different populations. Figure 3.5 shows the logarithm of two proportional baseline hazards, where the second population has a higher mortality. Assuming a higher degree of heterogeneity in this population (e.g., the variance of the gamma distributed frailty is higher in the second population) results in the crossing over effect because of selection. Higher heterogeneity implies a higher selection pressure in the second population. Consequently, frailer individuals die out faster in the second population than in the first one. If heterogeneity is on the same level in both populations, a convergence of the hazards will be observed instead of a crossing over. Such effects are described by Clayton and Kaldor (1985).
This phenomenon was also observed for example by Manton and Vaupel (1995) comparing mortality in Sweden, France, England and Japan with that in the United States.

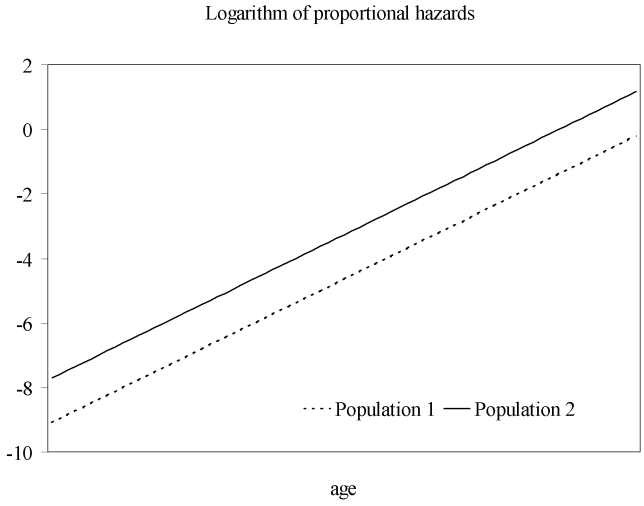

Figure 3.5: Baseline hazard functions of two populations with Gompertz hazard. The parameters are $a_1 = 10^{-4}$, $a_2 = 4 \times 10^{-4}$ and $b_1 = b_2 = 0.12$.

They noticed that mortality was much lower in all countries except the USA up to the ages of around 65. After the age of 80, the mortality was much lower in the United States. The authors pointed out that greater heterogeneity in social and economic status and health insurance coverage in the United States may account for the higher mortality at younger ages in this country. To make this more clear, consider the form of the conditional expectation for the gamma distributed random effect:

$$\mathbf{E}(Z|T \geq t, X) = \frac{1}{1 + \sigma^2 \Lambda_0(t) e^{\beta^T X}}$$

with $\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds$. Clearly, the average frailty value is a decreasing function of time. The decrease is faster in situations with greater heterogeneity in the population, as measured by σ^2, a higher cumulative baseline hazard function $\Lambda_0(t)$, and larger increases due to covariates, which are modelled by $e^{\beta^T X}$. All these factors accelerate the 'frailty selection' which results from unobserved heterogeneity in the individual risks.
Assume a single binary covariate X with values 0 (control group) and 1 (treatment group). Frailty selection also affects the marginal hazard ratio for treatment. Under the gamma frailty model, the marginal hazard ratio is (compare relation (3.8))

$$\frac{\lambda(t|X=1)}{\lambda(t|X=0)} = \frac{1 + \sigma^2 \Lambda_0(t)}{1 + \sigma^2 \Lambda_0(t) e^{\beta}} e^{\beta}. \tag{3.9}$$

A critical point here is that under the proportional hazards assumption for the conditional model, the ratio of the treatment specific marginal hazards is not generally time-invariant; unless σ^2 or β is zero, the marginal hazard ratio (3.9) is a decreasing function.

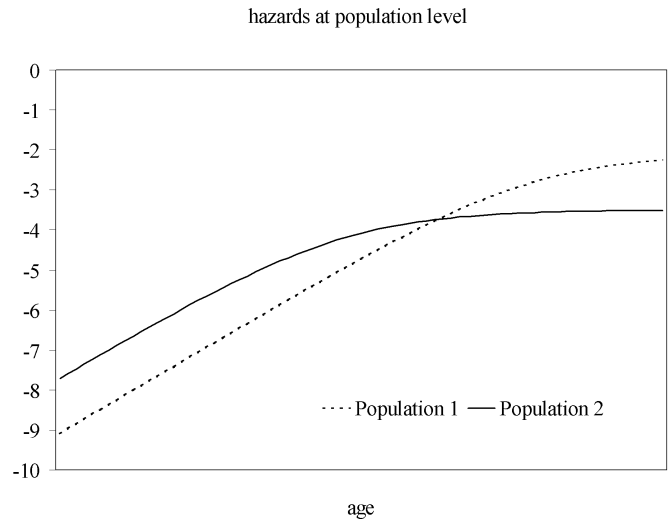

Figure 3.6: Mortality hazards of two populations with Gompertz baseline hazard and gamma distributed frailty. The parameters of the baseline hazards are the same as in Figure 3.5. The variance of the frailty is assumed to be 1 in Population 1 and 4 in Population 2.

Only at time zero is the marginal hazard ratio equal to the conditional hazard ratio e^{β}. As t increases, the ratio tends to one. Under these conditions, the time-invariant hazard ratio obtained by applying a simple marginal model is attenuated from the conditional hazard ratio e^{β}. The attenuation is increased in situations with larger values of σ^2, β and $\Lambda_0(t)$, representing the three factors which accelerate frailty selection. Conversely, the attenuation is modest if any of these factors is near zero, in which case the individual and population averaged parameters are relatively close. Formal expressions and approximations were obtained by Henderson and Oman (1999), who consider the effects of fitting a marginal model to conditional covariate effects.

There are many applications of the gamma frailty model. In a very early paper Lancaster (1979) suggested this model for the duration of unemployment spells and Vaupel et al. (1979) used it to correct life tables of heterogeneous populations. Aalen (1987) studied the expulsion of intra-uterine contraceptive devices. Manton et al. (1981) used it for comparing the mortality experience of heterogeneous populations, Manton and Stallard (1981) to explain the black/white mortality crossover in the US, Manton et al. (1986) compared the inverse normal and the gamma models, together with Gompertz and Weibull baseline hazards, in a study of survival at advanced ages, based on Medicare data. Jeong et al. (2003) used a gamma frailty to model long-term follow-up survival data from breast cancer clinical trials when the treatment effect diminishes over time as an alternative to the proportional hazards model. Jones (1998) used a gamma-Gompertz model for analyzing the impact of selective lapsation on mortality in life insurance.

Example 4 *This example deals with time from insertion of a catheter into dialysis patients until it has to be removed due to infection. It is possible that the catheters are removed for other reasons, so censoring is present. McGilchrist and Aisbett (1991) published a subset of the complete data including the first two infection times of 38 patients. Restricting the analysis to age and sex as covariates and treating the two failure times of every individual as independent, a Cox hazard model can be fitted to the data. To account for heterogeneity within the observed failure times, Hougaard (2000) applied a gamma frailty model. The results of the analysis are given in Table 3.1. As expected,*

	Cox model	gamma frailty
age	0.002 (0.009)	0.007 (0.012)
gender	-0.820 (0.300)	-1.770 (0.610)
σ^2		1.510 (0.980)

Table 3.1: Parameter estimates (s.e.) in the Cox model and the gamma frailty model applied to catheter infection data

the effect of age and gender is biased downwards in the Cox model when not corrected for unobserved heterogeneity in the study population. The gamma frailty model is able to account (at least in part) for this unobserved heterogeneity. Note that the standard deviation also increases in the gamma frailty model and that the large standard deviation of the frailty variance (σ^2) estimate does not exclude the possibility of no unobserved heterogeneity ($\sigma^2 = 0$).

Another extension of the gamma-frailty model was recently introduced by Barker and Henderson (2004). The authors introduce a mixed model which allows for both proportional and converging hazards. Two kinds of covariates are considered, X_1 and X_2. The usual gamma frailty model with covariates X_2 looks like $\lambda(t) = Z\lambda_0(t)e^{\beta_2 X_2}$, but the frailty distribution is assumed to depend on covariates X_1, namely $Z \sim \Gamma(\frac{e^{\beta_1 X_1}}{\sigma^2}, \frac{1}{\sigma^2})$. Then the unconditional survival function is obtained as

$$S(t|X_1, X_2) = \int_0^\infty e^{-Z\Lambda_0(t)e^{\beta_2 X_2}} f(z|X_1)\, dz = \left(1 + \sigma^2 \Lambda_0(t) e^{\beta_2 X_2}\right)^{-\frac{e^{\beta_1 X_1}}{\sigma^2}}. \tag{3.10}$$

When $\sigma^2 = 0$, the model reduces to the standard Cox proportional hazards model with covariates $\{X_1, X_2\}$. If $\beta_2 = 0$, the model reduces to the proportional hazards model with covariates X_1, whereas if $\beta_1 = 0$, it is clear from (3.10), that the mixed model reduces to the standard gamma frailty model with covariates X_2. The baseline hazard function is left unspecified (semi-parametric approach).
Parametric versions of the univariate gamma frailty model are included in statistical package STATA.

3.3 Positive stable frailty model

A distribution is called positive stable if the appropriately normalized sum of n independent random variables from this distribution has the same distribution. The normalization is given by $n^{\frac{1}{\gamma}}$, where the index γ must be in the range of $(0, 1]$ to get a distribution on positive numbers. Despite the fact that no closed form expressions exist for the probability density or the survival function of a random variable with positive stable distribution, the Laplace transform has a very simple form:

$$\mathbf{L}(s) = e^{-\frac{ks^\gamma}{\gamma}}.$$

For reasons of identifiability, we restrict the two-parameter frailty distribution to the case of $k = \gamma$. Consequently,

$$S(t) = \mathbf{L}(\Lambda_0(t)) = e^{-\Lambda_0(t)^\gamma},$$

$$f(t) = \gamma\lambda_0(t)\Lambda_0(t)^{\gamma-1}e^{-\Lambda_0(t)^\gamma},$$

$$\lambda(t) = \gamma\lambda_0(t)\Lambda_0(t)^{\gamma-1}.$$

This distribution was introduced as a frailty distribution by Hougaard (1986b) and applied for example by Wang et al. (1995) and Manatunga and Oakes (1999). Fine et al. (2003) and Martinussen and Pipper (2005) recently suggested new estimation procedures in the shared positive stable frailty model. It was further extended by Hougaard's power variance function distribution (Hougaard, 1986a) and Aalen's compound Poisson distribution (Aalen 1988, 1992). All moments of this distribution are infinite. This result is important with respect to identifiability issues treated by Elbers and Ridder (1982). They found that a finite mean of the frailty distribution is one condition (among others) for identifiability of univariate frailty models. This was the main reason why the positive stable distribution was introduced as a frailty distribution. Especially in bivariate/multivariate applications, much attention is given to overcoming confounding problems in the shared (gamma) frailty models (for more details see section 4.2).
The positive stable model implies some interesting features, which are given in the following examples. The most important one is the fact that the positive stable distribution is the only frailty distribution which preserves the proportional hazards assumption in the unconditional hazards after integrating out the frailty.

Example 5 *Let Z be positive stable with parameter γ. Denote by $\lambda_0(t)$ a baseline hazard function and assume a frailty model for two populations in the classical form (3.1), e.g. $\lambda_1(t,Z) = Z\lambda_0(t)$ and $\lambda_2(t,Z) = Z\rho\lambda_0(t)$. Obviously, the two individual hazards are proportional to ratio $\frac{\lambda_2(t,Z)}{\lambda_1(t,Z)} = \rho$ for two individuals with the same frailty. Using the relation above gives the unconditional hazards:*

$$\begin{aligned}\lambda_1(t) &= \gamma\lambda_0(t)\Lambda_0(t)^{\gamma-1}\\ \lambda_2(t) &= \gamma\rho\lambda_0(t)(\rho\Lambda_0(t))^{\gamma-1}\\ &= \rho^\gamma\gamma\lambda_0(t)\Lambda_0(t)^{\gamma-1}.\end{aligned}$$

Again, the two hazard functions are proportional, but now with a different ratio $\frac{\lambda_2(t)}{\lambda_1(t)} = \rho^\gamma$.

Example 6 *(Extension of Example 5) Let Z be positive stable with parameter γ and denote by $\lambda_0(t)$ a baseline hazard function and assume a frailty model for two populations in a proportional hazards model with covariates, e.g.*

$$\lambda_1(t,Z,X) = Z\lambda_0(t)e^{\sum_{i=1}^{k}\beta_i X_i}$$

and

$$\lambda_2(t,Z,X) = Z\rho\lambda_0(t)e^{\sum_{i=1}^{k}\beta_i X_i}.$$

Again, the two individual hazards are proportional with ratio $\frac{\lambda_2(t,Z,X)}{\lambda_1(t,Z,X)} = \rho$ for two individuals with the same frailty and equal covariates. Using the relation above gives the unconditional hazards:

$$\begin{aligned}\lambda_1(t,X) &= \gamma\lambda_0(t)e^{\sum_{i=1}^{k}\beta_i X_i}(\Lambda_0(t)e^{\sum_{i}^{k}\beta_i X_i})^{\gamma-1}\\ \lambda_2(t,X) &= \gamma\rho\lambda_0(t)e^{\sum_{i=1}^{k}\beta_i X_i}(\rho\Lambda_0(t)e^{\sum_{i}^{k}\beta_i X_i})^{\gamma-1}\\ &= \rho^\gamma\gamma\lambda_0(t)\Lambda_0(t)^{\gamma-1}e^{\gamma\sum_{i}^{k}\beta_i X_i}.\end{aligned}$$

Then the (unconditional) hazards are still proportional, but with regression coefficients $\gamma\beta_i$ instead of β_i. This shows that parameter estimates are biased if relevant covariates are not included.

Example 7 *Let $\Lambda_0(t) = at^b$ $(Weibull(a,b))$ and assume that the frailty variable is positive stable distributed with parameter γ. Then the resulting survival function is*

$$S(t) = e^{-\Lambda_0(t)^\gamma} = e^{-a^\gamma t^{\gamma b}}.$$

This is again a Weibull distribution, but with parameters a^γ and γb.

3.4 Inverse Gaussian frailty model

The inverse Gaussian (inverse normal) distribution was introduced as an alternative to the gamma distribution by Hougaard (1984) and has been used for example by Manton et al. (1986), Klein et al. (1992), Keiding et al. (1997) and Price and Manatunga (2001). The probability density function of an inverse normal distributed random variable with mean one and variance σ^2 is

$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2 z^3}} e^{-\frac{1}{2\sigma^2 z}(z-1)^2}.$$

Consequently, the Laplace transform of the inverse normal distribution is given by

$$\mathbf{L}(s) = e^{\frac{1}{\sigma^2}(1-\sqrt{1+2\sigma^2 s})}.$$

Hence, the conditional survival and hazard function take the forms

$$S(t) = e^{\frac{1}{\sigma^2}(1-\sqrt{1+2\sigma^2\Lambda_0(t))}}$$

and

$$\lambda(t) = \frac{\lambda_0(t)}{(1+2\sigma^2\Lambda_0(t))^{1/2}}.$$

Parametric versions of the inverse Gaussian frailty model are available in STATA.

3.5 PVF frailty model

An extended family of frailty distributions, including gamma, inverse Gaussian as well as positive stable distributions, is the family of power variance function distributions, suggested by Tweedy (1984) and later derived independently by Hougaard (1986a). This is a three parameter family denoted by $PVF(\gamma, k, \lambda)$. The Laplace transform is

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

Expectation and variance of a PVF distributed random variable Z are

$$\mathbf{E}Z = k\lambda^{\gamma-1} \text{ and } \mathbf{V}(Z) = k(1-\gamma)\lambda^{\gamma-2}. \tag{3.11}$$

The resulting survival function is given by

$$S(t) = e^{-\frac{k}{\gamma}\left((\lambda+\Lambda_0(t))^\gamma - \lambda^\gamma\right)}$$

and the unconditional hazard function is

$$\lambda(t) = k\lambda_0(t)(\lambda+\Lambda_0(t))^{\gamma-1}.$$

Using the constraint $\mathbf{E}Z = 1$ and relation $\sigma^2 = \frac{1-\gamma}{\lambda}$ (see (3.11)) it holds that

$$\lambda(t) = \frac{\lambda_0(t)}{(1 + \frac{\sigma^2}{1-\gamma}\Lambda_0(t))^{1-\gamma}}.$$

For certain values of the parameters, the analytic expression above is not well defined immediately, but should be defined by continuity. If $\gamma = 0$, the gamma distributions $\Gamma(k, \lambda)$ are obtained with the same parameterization. For $\gamma = 0.5$, the inverse Gaussian distributions and for $\lambda = 0$ the positive stable distributions are obtained. The parameter set is $0 \leq \gamma \leq 1, k > 0$, with $\lambda \geq 0$ for $\gamma > 0$, and $\lambda > 0$ for $\gamma \geq 0$. The distribution among survivors at age t is $PVF(\gamma, k, \lambda + \Lambda_0(t))$. Hougaard et al. (1992) applied the model to lifetimes of Danish twins.

3.6 Compound Poisson frailty model

The 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 zero frailty, which survives forever. This is a model relevant to medicine and demography. 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. This can be viewed as a hit model, where each individual experiences a random number of hits, each of a random size.

$$Z = \begin{cases} X_1 + X_2 + \ldots + X_N & : \text{ if } N > 0, \\ 0 & : \text{ if } N = 0, \end{cases} \tag{3.12}$$

where N is Poisson distributed with expectation ρ, while $X_1, X_2, \ldots$ are independent and gamma distributed with $X_i \sim \Gamma(k, \lambda)$. The Laplace transforms of gamma and Poisson distributions are given by $\mathbf{L}_X(s) = (1 + \frac{s}{\lambda})^{-k}$ and $\mathbf{L}_N(s) = e^{-\rho+\rho e^{-s}}$, respectively. The following standard derivation can now be applied:

$$\mathbf{L}(s) = \mathbf{E}e^{-sZ} = \mathbf{E}e^{-s(X_1+\ldots+X_N)} = \mathbf{E}\mathbf{L}_X(s)^N = \mathbf{L}_N(-\ln(\mathbf{L}_X(s))).$$

Inserting the previous expressions results in the following Laplace transform of Z:

$$\mathbf{L}(s) = e^{-\rho+\rho(1+\frac{s}{\lambda})^{-k}}.$$

In the following we want to use another parameterization:

$$\rho = -\frac{k\lambda^\gamma}{\gamma}, \qquad \lambda = \lambda, \qquad k = -\gamma.$$

The Laplace transform of the compound Poisson distribution is now given by

$$\mathbf{L}(s) = e^{-\frac{k}{\gamma}\left((\lambda+s)^{\gamma}-\lambda^{\gamma}\right)}. \tag{3.13}$$

Parameter γ divides the class of distributions in two major subfamilies: For $\gamma \geq 0$, the distribution is a power variance function distribution (PVF). The extension to $\gamma < 0$ was suggested by Aalen (1988) and yields the compound Poisson distribution. The two subclasses are separated by the gamma distribution ($\gamma = 0$). Note that a different parameterization was used by Aalen. The notation $cP(\gamma, k, \lambda)$ is used for a compound Poisson distribution.

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

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

This results in

$$S(t) = e^{-\frac{1-\gamma}{\gamma\sigma^2}((1+\frac{\sigma^2}{1-\gamma}\Lambda_0(t))^{\gamma}-1)} \quad \text{and} \quad \lambda(t) = \frac{\lambda_0(t)}{(1+\frac{\sigma^2}{1-\gamma}\Lambda_0(t))^{1-\gamma}},$$

similar to the PVF frailty model. It should be noted that the integral of $\lambda(t)$ (cumulative hazard function) over $[0, \infty)$ is finite when $\gamma < 0$. Consequently, the survival function is incomplete because a fraction of individuals has zero frailty who will never experience the event under study. The model was used by Aalen (1992) to model the incidence of marriage of women born in Denmark. Marriage is an example of an event that does not happen to everybody. A certain percentage of individuals never marry and models of marriage incidence have to account for this. The fact that the observed incidence peaks around age 23 and becomes rather low after age 30 is therefore interpreted to be a selection phenomenon due to heterogeneity, meaning that those who are most prone to marriage will marry quite early, and the reminder will have a lower tendency to marry.

A second application of Aalen (1992) deals with fertility data in Norwegian woman. It is a well-known fact that around 5% to 10% of all couples are unable to conceive children. Naturally, $\mathbf{P}(Z = 0)$ (where Z denotes the frailty to conceive a child) will be the probability of infertility, and the variation of the frailty variable Z expresses the varying fecundabilities among fertile couples.

Hougaard et al. (1994) applied the model to diabetic nephropathy onset data, a serious complication experienced by some diabetic patients. In a recent paper, Aalen and Tretli (1999) applied the compound Poisson distribution to testicular cancer. Testicular cancer has two striking epidemiological features. First, its incidence has increased rapidly over the past few decades. Secondly, the incidence is greatest among younger men, and

then declines from a certain age. The idea of the model is that a subgroup of men is particularly susceptible to testicular cancer, which results in selection over time. The model is fit to incidence data from the Norwegian Cancer Registry collected between 1953 and 1993. This work is being continued by Moger et al. (2004a). Haukka et al. (2003) applied the model to schizophrenia data from the Finnish population born 1950 - 1968. They concluded that only a small part of the population is susceptible to schizophrenia and found increasing individual risk with higher age among the susceptible part of the population.

To illustrate the many interesting features of the model above, the malignant melanoma data set is considered, given in Andersen et al. (1993). It includes records of 205 patients, all of whom had a radical surgery for malignant melanoma (skin cancer) at the University Hospital of Odense in Denmark. The time scale is the time since operation and the end of the follow-up was 1977. There were 57 deaths related to malignant melanoma and 14 deaths to other causes (considered as censored). The covariates age, sex and thickness of the tumor were included. Three models were applied to the data by Hougaard (2000), the Cox regression model (2.7), the gamma frailty model and the PVF frailty model. The results are given in Table 3.2.

	Cox model	gamma frailty	PVF frailty
age	0.012 (0.008)	0.007 (0.015)	0.006 (0.008)
sex	-0.550 (0.270)	-0.920 (0.510)	-0.910 (0.490)
thickness	0.151 (0.033)	0.513 (0.187)	0.520 (0.140)
λ		0.207 (0.091)	0.286 (0.159)
γ			-0.130 (0.220)

Table 3.2: Parameter estimates (s.e.) in the Cox model, gamma frailty model and PVF frailty model applied to malignant melanoma data.

As expected, the Cox analysis suggests that males face a higher risk of death caused by malignant melanoma than women. Increasing age and increasing size of the tumor are related to higher mortality. Using the same covariates and gamma and PVF frailty models results in an increase in the effect of the covariates thickness and sex, where the effect of age is decreased. The estimate of γ is negative, consequently this is a compound Poisson frailty model. Consequently, some patients will never die from this disease. This is possible because only death to malignant melanoma was studied and not total mortality. The estimated proportion with zero risk is 11.6 %.

Further applications are considered in Price and Manatunga (2001).

3.7 Log-normal frailty models

Log-normal frailty models are especially useful in modelling dependence structures in multivariate frailty models, for example in McGilchrist and Aisbett (1991), McGilchrist (1993), Lillard (1993), Lillard et al. (1995), Xue and Brookmeyer (1996), Sastry (1997), Gustafson (1997), Ripatti and Palmgren (2000); Ripatti et al. (2002), Huang and Wolfe (2002).

However, the log-normal distribution has also been applied in univariate cases, for example by Flinn and Heckman (1982). Two variants of the log-normal model exist. We assume a normally distributed random variable W to generate frailty as $Z = e^W$. The two variants of the model are given by the restrictions $\mathbf{E}W = 0$ and $\mathbf{E}Z = 1$, where the first one is much more popular in the literature. Unfortunately, no explicit form of the unconditional likelihood exists. Consequently, estimation strategies based on numerical integration in the maximum likelihood approach are required.

3.8 Univariate frailty cure models

The Cox proportional hazards model is commonly used in the analysis of survival time data. An often unstated assumption of this model (and of frailty models with the exception of those that use the compound poisson distribution) 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 cured. For example, researchers may be interested in analyzing the recurrence of a disease. Many individuals may never experience a recurrence of that disease; therefore, a fraction in the population exists that has recovered from disease. Historically, cure models have been utilized to estimate the healed population fraction. 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 fraction component is not present, the analysis is reduced to standard approaches of survival analysis. Cure models assume that the individuals experiencing the event of interest are homogeneous in risk. This section deals with extensions of cure models in order to allow for heterogeneity among the fraction under risk by using frailty models. Or, depending on the point of view, it deals with extensions of frailty models to allow for a cured fraction in the study population. In this case, the distribution of the frailty is a combination of discrete and continuous distributions. For example, when the omitted variable is alcohol consumption in which a discrete proportion of non-drinkers occurs.

Spilerman (1972) considered the 'spiked-gamma' as an example of such a distribution. 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 $S_0(t)$, with probability ϕ. Individuals regarded as cured will never experience the event of interest and their survival time will be defined as infinity. Therefore, hazard and survival functions of cured individuals are set to zero and one, respectively, for all finite values of t. A model of survival times that incorporates a cured fraction is given by

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

Longini and Halloran (1996) have proposed frailty cure (cure-mixture) 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. In the gamma frailty cure model, the survival function is given by

$$S^*(t) = (1-\phi) + \phi(1+\sigma^2\Lambda_0(t))^{-1/\sigma^2}. \tag{3.14}$$

The idea behind this model is similar but not equivalent to the compound Poisson frailty model suggested by Aalen (1988, 1992). Price and Manatunga (2001) gave a comprehensive introduction to this area and applied different cure, frailty and frailty cure models to leukaemia remission data. They conclude that frailty models are useful in modelling 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.

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 followed up for a sufficiently long period of time. It is possible that these models 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.

The application of mixture models for joint modelling of the overall risk of a disease and the age-at-onset distribution of the diseased individuals is popular (Farewell 1977; Kuk and Chen 1992; Lam et al. 2005). We define an individual to be susceptible if he/she

will eventually develop the disease if followed for a sufficiently long time. Define

$$Y = \begin{cases} 1 & : \text{ if the individual is susceptible} \\ 0 & : \text{ if otherwise} \end{cases} \tag{3.15}$$

and let T denote the age at onset when $Y = 1$. With the above notion let $\phi = \mathbf{P}(Y = 1)$ and $S(t) = \mathbf{P}(T > t|Y = 1)$ describe the distribution of Y and the failure time T. It is impossible to observe Y, but it is possible to observe whether or not a subject has experienced the event during his/her follow-up time.

- For those observations which experience the event it holds that $\Delta = 1$. Obviously, $Y = 1$ and the survival function for the uncensored observations is of the form $\mathbf{P}(Y = 1)\mathbf{P}(T \leq C|Y = 1) = \phi(1 - S(C))$, where C denotes the censoring time.
- For the other observations a failure is not observed ($\Delta = 0$). This may occur either because $Y = 0$ or because the observation is really censored. Therefore $\mathbf{P}(Y = 0) + \mathbf{P}(Y = 1)\mathbf{P}(T > C|Y = 1) = (1 - \phi) + \phi S(C)$.

Combining these results the likelihood function takes the form

$$L(t, \delta) = \delta\phi f(t) + (1 - \delta)\big(1 - \phi + \phi S(t)\big), \tag{3.16}$$

which is a generalization of (2.6) in the case of independent censoring. This model is applied to time-to-onset of breast cancer in Swedish twins. The data set contains records of 11,714 female twins followed up from 1959/61 to 27 October 2000. Altogether, 715 cases of breast cancer were identified. More details about the Swedish Twin Register can be found in Lichtenstein et al. (2002) and the Appendix A.1.

	gamma	gamma cure	compound Poisson
γ	0		-0.089 (0.132)
σ	7.685 (0.499)	3.030 (1.887)	6.714 (1.141)
ϕ	1.000	0.221 (0.178)	0.237*
likelihood	5236.5914	5236.2517	5236.2751

Table 3.3: Time-to-onset of breast cancer in 11,714 Swedish twins. * calculated by $\phi = 1 - e^{\frac{1-\gamma}{\gamma\sigma^2}}$

In the first column of Table 3.3 the gamma frailty model is applied to account for heterogeneity in the population. This model is described in more detail in Section 3.2. The model is extended to the gamma frailty cure model (second column, see (3.14)), which allows for a fraction of non-susceptible individuals, e.g. patients who are not at risk of suffering from breast cancer. The third model is the compound Poisson frailty model (Section 3.6). All models use a Gompertz baseline hazard function.

In the gamma frailty model the size of the susceptible fraction is 100 % per definition.The gamma frailty cure model gives an estimate of 22.1% of the size of the susceptible fraction, e.g. only 22.1% of all woman are at risk for breast cancer. The respective number calculated from the compound Poisson frailty model is similar with 23.7%. The gamma frailty model is a special case of the gamma frailty cure model (when $\phi = 1$) and the compound Poisson frailty model (when $\gamma = 0$). These estimates are in line with results found by Chatterjee and Shih (2001) and Wienke et al. (2003a) in bivariate analysis and estimates of a lifetime risk of breast cancer of around 8-12 % (Harris et al. 1992; Feuer et al. 1993; Rosenthal and Puck 1999; Ries et al. 1999) in current western populations, which give a lower boundary for the size of the susceptible fraction. A large estimate of σ^2 in the gamma model indicates the existence of large heterogeneity in the population, which is partially accounted for by the introduction of a non-susceptible fraction in the gamma frailty cure model, where the heterogeneity is smaller. In the compound Poisson frailty model the parameter estimate γ is negative, indicating the existence of a non-susceptible fraction. The good fit of the model is demonstrated in Figure 3.7.

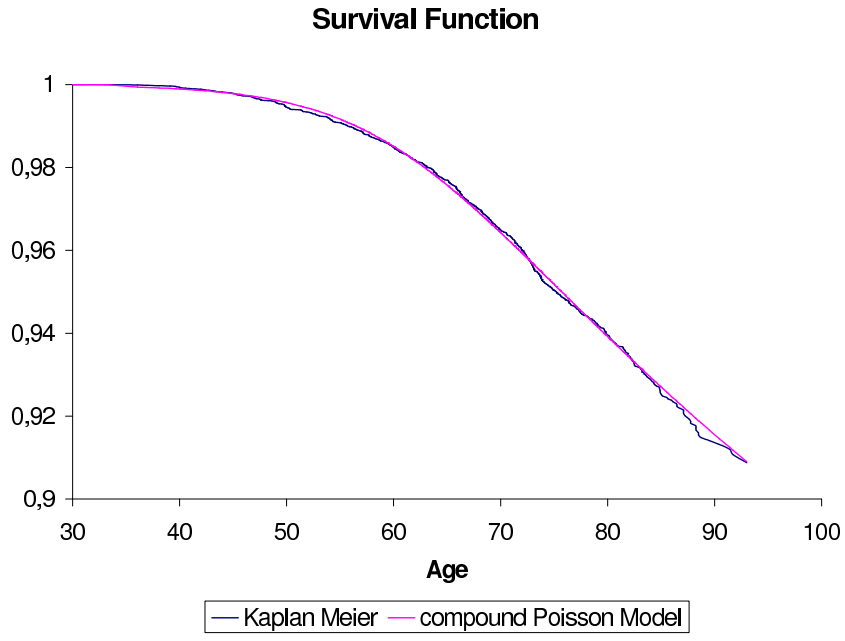

Figure 3.7: Kaplan-Meier and compound Poisson model for Swedish breast cancer data.

The correlation between the times to onset of breast cancer in the twin pairs is neglected in these univariate models. Cure models suffer from an inherent identifiability problem with right censored 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 to observe the event. For a discussion of the pro's and con's of univariate frailty models as cure models see Schumacher (1989).

Chapter 4

Multivariate frailty models

So far, we have focused on the frailty model as a way of dealing with possible heterogeneity due to unobserved covariates. This is the main interpretation of frailty in the application to univariate time-to-event data. This results in selection effects over time, for example shown as levelling-off or crossing-over effects in population hazards.
Another, completely different aspect of this approach is to use it to model statistical dependence, e.g., Clayton (1978). Implicitly, most of the statistical models and methods for failure time data (and here especially the Cox proportional hazards model) were developed under the assumption that the observations from subjects are statistically independent of each other. While this is sensible in many applications, it has become obvious that this assumption does not hold in other situations which are not as uncommon as originally thought. The following three examples given in Liang et al. (1995) serve to illustrate this point:

- Diabetic retinopathy is one of the leading causes of visual loss and blindness. Given the relatively high prevalence of this disease among the population, new treatments to delay the onset of severe visual loss are critical. In 1971, the Diabetic Retinopathy Study was initiated to study the effectiveness of laser photocoagulation. This randomized, controlled clinical trial involved more than 1,700 patients enrolled at 15 medical centers in the US. Patients with diabetic retinopathy and visual acuity of 20/100 or higher in both eyes were eligible to participate in the study. The design protocol randomly selected one of each participant's eyes that was treated, and the other eye was observed without having received treatment. The event was the occurrence of visual acuity less than 5/200 at two consecutively completed follow-ups of four months. This design differs from conventional randomized trials in that each patient served as his own control. Consequently, each patient contributed two correlated observations to the analysis, one from each eye.

- Family studies are critical in assessing the role genetics play in the disease process. Statistical methods such as variance component models and path analysis have been developed and adopted to analyze family data with quantitative traits, such as cholesterol level. Variance component models aim to measure the extent to which the total variation in the quantitative trait is due to a correlation between relatives, as well as the degree of correlation among full siblings and other relatives. When the considered trait is the onset age of the disease, conventional methods are not appropriate. This is partially due to the censoring and truncation feature of this variable but also to the need for a measure of within-family correlation which incorporates time, a feature not shared by the conventional measures such as the correlation coefficient.

- The trial of gamma interferon as detailed in Fleming and Harrington (1991) is a typical example of recurrent events. In their study, patients with chronic granulomatous disease (CGD) were randomly assigned to either gamma interferon or placebo, with the event of interest being the diagnosis of a serious infection. It is clear that infections can recur. Thus both the time and the number of infection occurrence may be informative about the effectiveness of the treatment.

These three examples share an important feature, namely, that the failure times for observations from the same cluster correlate with one another. In Examples 1 and 3 the cluster is an individual, in Example 2 it is a family. Typically, the cluster sizes are small relative to the number of clusters. These three examples, however, differ from one another in terms of their scientific objectives. The main objective for Examples 1 and 3 is to examine the effectiveness of a new treatment, which presumably can be characterized through regression modelling. The within-cluster correlation in these examples is usually of secondary interest, although ignoring it could lead to erroneous conclusions. The within-family association for Example 2 is of primary interest, although regression adjustment for each related subject is critical in order to minimize the potential that the observed association is mainly due to environmental factors shared by the family. Furthermore, the mechanisms for the within-cluster associations may vary so that different statistical models to describe the associations may be needed. It is clear, for instance, that the mechanism leading to the correlation between two fellow eyes is different from that attributed to the correlation of observations measured over time from the same eye. Multivariate survival analysis may provide an effective tool for information from multiple/recurrent events in situations as described above. Available statistical models fall into two broad classes - marginal and frailty models (Wei and Glidden 1997). Marginal

methods of analysis specify models for the effect of covariates on the hazards of the individual events (the margins), taking into account the fact that observed event times are correlated but without the need for explicitly modelling this correlation (Wei et al. 1989; Lee et al. 1992; Cai and Prentice 1995). The association between the events is considered a nuisance parameter. The marginal baseline hazards can be modelled differently (Wei et al. 1989) or with a common functional form (Lee et al. 1992). As with the analysis of longitudinal data, regression parameters are estimated from generalized estimating equations, and the corresponding variance-covariance estimators are corrected properly to account for the dependence structure. An excellent review of the robust and well developed marginal approach is given in Lin (1994).

The marginal approach is ideal for making inferences on the population average effect of risk factors on failure time. However, it provides no insight into the multivariate relationship among failure times. These type of questions are answered by frailty models, explicitly considering the association between various events. In general, frailty models have an intuitive appeal and provide insight into the relationship between failures, and we will focus on this approach in the present thesis.

An approach occupying an intermediate position between the two aforementioned models to estimate both regression coefficients with traditional interpretation as well as correlations is described by Mahé and Chevret (1999).

A commonly used and very general approach to the problem of modelling multivariate data is to specify independence among observed data items conditional on a set of unobserved or latent variables. A multivariate model for the observed data is then induced by averaging over an assumed distribution for the latent variables. The dependence structure in the multivariate model arises when dependent latent variables enter into the conditional models for multiple observed data items, and the dependence parameters may often be interpreted as variance components. Frailty models for multivariate survival data are derived under a conditional independence assumption by specifying latent variables which act multiplicatively on the baseline hazard.

Let us now focus on multivariate models with dependent random hazards as described above. This concept provides a multivariate extension of the traditional univariate frailty model (Vaupel et al. 1979), and it allows us to take mutual dependence of life times of related individuals into account in the analysis of survival data. Survival models for dependent life times are useful because they allow us to address more sophisticated questions about the nature of ageing, disease and the mortality processes.

The first important approach is the shared frailty concept. In a shared frailty model, frailty is defined as a measure of the relative risk which individuals in a group share.

Thus, the frailty variable is associated with groups of individuals rather than individuals as such. The hazard model for each individual, however, looks exactly the same as in the standard univariate frailty model:

$$\lambda(t, Z) = Z\lambda_0(t).$$

The model assumes that all failure times are independent given the frailties. In other words, the lifetimes are conditionally independent. The value of the frailty Z is constant over time and common to the individuals in the group, and thus it is responsible for creating dependence. This dependence is always positive. The conditional survival function in the bivariate case is

$$S(t_1, t_2|Z) = S(t_1|Z)S(t_2|Z) = e^{-Z\Lambda_0(t_1)}e^{-Z\Lambda_0(t_2)}.$$

The correlated frailty model is the second important concept in multivariate frailty models. This is a natural extension of the shared frailty model. In the correlated frailty model, the individuals in a group share only parts of the frailty Z. It enables the explicit inclusion of an additional correlation parameter which then allows questions about genetic and environmental effects on individual frailty. The conditional survival function in the bivariate case looks like

$$S(t_1, t_2|Z_1, Z_2) = S(t_1|Z_1)S(t_2|Z_2) = e^{-Z_1\Lambda_0(t_1)}e^{-Z_2\Lambda_0(t_2)},$$

where Z_1 and Z_2 are two correlated random variables. The assumption of conditional independence plays an important role in genetic analysis of survival. Under this assumption the frailty variable becomes the only carrier of genetic influence on longevity. It means that the underlying hazard represents only non-genetic influence on life span. Univariate frailty models do not have such a property.

In the case of a degenerated frailty Z there exists no dependence between the lifetimes within a group. Different groups are considered as independent. The number of individuals in a group is assumed to be known. We will consider the bivariate case of pairs of individuals in more detail in order to explain the main ideas. Twin studies provide typical examples of bivariate event data and will be considered in detail later. Another example are the times to failure for several similar human organs, like time to blindness of the right and the left eye as in studies on diabetic retinopathy. Extensions to higher-dimensional cases are given where necessary.

In all cases, it is necessary to assign a distribution to the frailty. Assuming a random frailty means that we can integrate the frailty out of the expressions, and thus evaluate the multivariate survival times. Almost all calculations can be made based on the Laplace transform of the respective frailty distribution.

4.1 Shared frailty model

A shared frailty model in survival analysis is defined as follows. Suppose there are n clusters and that the i-th cluster has n_i individuals and associates with an unobserved frailty Z_i, $(1 \leq i \leq n)$. A vector X_{ij} $(1 \leq i \leq n, 1 \leq j \leq n_i)$ is associated with the ij-th complete survival time T_{ij} of the j-th individual in the i-th cluster. Conditional on frailties Z_i, the survival times are assumed to be independent and their hazard functions to be of the form

$$\lambda(t) = Z_i \lambda_{0j}(t) e^{\beta^T X_{ij}},$$

where $\lambda_{0j}(t)$ are the baseline hazard functions and β is a vector of fixed effect parameters to be estimated. The frailties Z_i are assumed to be identically and independently distributed random variables with a common density function $f(z, \theta)$, where θ is the parameter of the frailty distribution. A semi-parametric shared frailty model is a frailty model with a non-parametric baseline hazard function $\lambda_{0j}(t)$.

For simplicity, we restrict our treatment of frailty models to the bivariate case ($n_i = 2$), because extensions to the multivariate case are straightforward. The assumption of a shared frailty model is that both individuals in a pair share the same frailty Z, and this is why the model is called the shared frailty model. It was introduced by Clayton (1978) (who did not use the notion of 'frailty') and extensively studied in Hougaard (2000), Therneau and Grambsch (2000), Duchateau et al. (2002), (2003) and Duchateau and Janssen (2004). The two lifetimes are assumed to be conditionally independent with respect to the shared (common) frailty. We derive the quantities based on this conditional formulation below. Conditionally on Z, the hazard function of an individual in a pair is of the form $Z\lambda_0(t)$, where the value of Z is common to both individuals in the pair, and thus is the cause for dependence between life times within pairs. Independence of the life times within a pair corresponds to a degenerate frailty distribution (no variability in Z). In all other cases, the dependence is positive. It is assumed that there is independence between different pairs. If $\mathbf{P}(Z > 0) = 1$ holds, the shared frailty model leads to absolute continuous distributions and thus cannot model dependence due to common events. Consequently, it is not appropriate for event-related dependence (shock models), because an event in one individual is not relevant for the partner, it only changes the information available on the frailty.

Using an argument similar to equation (3.4) we can derive the bivariate survival function. Conditional on Z, it is

$$S(t_1, t_2|Z) = S_1(t_1)^Z S_2(t_2)^Z = e^{-Z\Lambda_{01}(t_1)} e^{-Z\Lambda_{02}(t_2)} = e^{-Z(\Lambda_{01}(t_1)+\Lambda_{02}(t_2))}, \tag{4.1}$$

where $\Lambda_{0i}(t) = \int_0^t \lambda_{0i}(s)\, ds$, $i = 1, 2$ are the cumulative baseline hazard functions.

Averaging (4.1) with respect to Z produces the marginal bivariate survival function

$$\begin{aligned} S(t_1, t_2) &= \mathbf{E}S(t_1, t_2|Z) \\ &= \mathbf{E}S_1(t_1)^Z S_2(t_2)^Z \\ &= \mathbf{E}e^{-Z(\Lambda_{01}(t_1)+\Lambda_{02}(t_2))} \\ &= \mathbf{L}(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)), \end{aligned}$$

where $\mathbf{L}$ denotes the Laplace transform of Z. Thus, the bivariate survival function is expressed as the Laplace transform of the frailty distribution, evaluated at the cumulative baseline hazard. For the marginal survival functions the following holds:

$$S_i(t_i) = \mathbf{E}S_i(t_i|Z) = \mathbf{E}S_i(t_i)^Z = \mathbf{E}e^{-Z(\Lambda_{0i}(t_i))} = \mathbf{L}(\Lambda_{0i}(t_i)) = p(\Lambda_{0i}(t_i)).$$

with $p = \mathbf{L}$ and $i = 1, 2$. Consequently, $\Lambda_{0i}(t_i) = q(S_i(t_i))$, where q denotes the inverse function of p and the bivariate unconditional survival function is given by

$$S(t_1, t_2) = p(q(S_1(t_1)) + q(S_2(t_2))),$$

the archimedian copula family of Genest and MacKay (1986), where p is a function that can be differentiated twice with $p(0) = 1$, $p'(\cdot) < 0$ and $p''(\cdot) > 0$.
The standard assumption about the frailty distribution is that it is a gamma distribution with mean 1 and variance σ^2. Averaging (4.1) with respect to Z produces the marginal bivariate survival function

$$\begin{aligned} S(t_1, t_2) &= \mathbf{L}(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)) \\ &= (1 + \sigma^2(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)))^{-1/\sigma^2} \\ &= (S_1(t_1)^{-\sigma^2} + S_2(t_2)^{-\sigma^2} - 1)^{-1/\sigma^2}, \end{aligned} \tag{4.2}$$

where the last relation is a consequence of equation (3.8). The notion of shared frailty is different from the definition of individual frailty introduced by Vaupel et al. (1979) in their analysis of univariate duration data. This difference has gone largely unrecognized, perhaps because of the superficial similarity of the individual hazards in the two approaches. The frailty in the bivariate shared frailty model is only a part of the individual frailty, capturing only the components of frailty that both individuals share.
Clayton (1978), Cox and Oakes (1984) and Yashin and Iachine (1999a) pointed out that the bivariate survival function calculated above in the shared gamma frailty model (4.2) can also be derived using a radically different approach. Denote dependent life spans by T_1, T_2 , and let $S(t_1, t_2) = \mathbf{P}(T_1 > t_1, T_2 > t_2)$ be a bivariate survival function that is absolutely continuous with marginals $S_1(t_1) = S(t_1, 0)$ and $S_2(t_2) = S(0, t_2)$.

Consequently, the conditional survival function of T_1 given $T_2 > t_2$ is $S(t_1|T_2 > t_2) = \frac{S(t_1,t_2)}{S(t_2)}$ and that of T_1 given $T_2 = t_2$ is

$$S(t_1|T_2 = t_2) = \frac{\frac{\partial S(t_1,t_2)}{\partial t_2}}{\frac{\partial S(t_2)}{\partial t_2}}.$$

The respective conditional hazards are important for the following considerations. Using the relation $\lambda(t) = -\frac{S'(t)}{S(t)}$ implies

$$\lambda(t_1|T_2 > t_2) = -\frac{\partial}{\partial t_1} \ln(S(t_1, t_2)) \tag{4.3}$$

and

$$\lambda(t_1|T_2 = t_2) = -\frac{\partial}{\partial t_1} \ln\big(-\frac{\partial}{\partial t_2} S(t_1, t_2)\big). \tag{4.4}$$

These hazards describe the risk of failure at age t_i for the i-th individual, given the information about the status of the second individual. The first hazard (4.3) uses the condition $\{T_j > t_j\}$ and the second one (4.4) is conditional on $\{T_j = t_j\}$. The deviation of the ratio of these hazards from 1 was used by Oakes (1989) as a measure of mutual dependence of respective marginal life times. The shared gamma frailty model can now be based on the introduction of the following relation between the above hazards:

$$\lambda(t_1|T_2 = t_2) = (1 + \sigma^2)\lambda(t_1|T_2 > t_2) \tag{4.5}$$

Clearly, (4.5) is equivalent to a similar condition with the roles of T_1 and T_2 being interchanged. This relation defines the bivariate survival function (4.2) uniquely (up to the marginal distributions):

$$\begin{aligned}
\lambda(t_1|T_2 = t_2) &= (1 + \sigma^2)\lambda(t_1|T_2 > t_2) \\
\frac{\partial}{\partial t_1} \ln\big(-\frac{\partial}{\partial t_2} S(t_1, t_2)\big) &= (1 + \sigma^2)\frac{\partial}{\partial t_1} \ln(S(t_1, t_2)) \\
\int_0^{t_1} \frac{\partial}{\partial t} \ln\big(-\frac{\partial}{\partial t_2} S(t, t_2)\big)\, dt &= (1 + \sigma^2) \int_0^{t_1} \frac{\partial}{\partial t} \ln(S(t, t_2))\, dt \\
\ln\big(-\frac{\partial}{\partial t_2} S(t_1, t_2)\big) - \ln\big(-\frac{\partial}{\partial t_2} S_2(t_2)\big) &= (1 + \sigma^2)\big(\ln(S(t_1, t_2)) - \ln(S_2(t_2))\big) \\
\ln\big(-\frac{\partial}{\partial t_2} S(t_1, t_2)\big) - (1 + \sigma^2)\ln(S(t_1, t_2)) &= \ln\big(-\frac{\partial}{\partial t_2} S_2(t_2)\big) - (1 + \sigma^2)\ln(S_2(t_2)) \\
\frac{\frac{\partial}{\partial t_2} S(t_1, t_2)}{S(t_1, t_2)^{1+\sigma^2}} &= \frac{\frac{\partial}{\partial t_2} S_2(t_2)}{S_2(t_2)^{1+\sigma^2}} \\
\int_0^{t_2} \frac{\frac{\partial}{\partial t} S(t_1, t)}{S(t_1, t)^{1+\sigma^2}}\, dt &= \int_0^{t_2} \frac{\frac{\partial}{\partial t} S_2(t)}{S_2(t)^{1+\sigma^2}}\, dt \\
S(t_1, t_2)^{-\sigma^2} - S_1(t_1)^{-\sigma^2} &= S_2(t_2)^{-\sigma^2} - 1 \\
S(t_1, t_2) &= \big(S_1(t_1)^{-\sigma^2} + S_2(t_2)^{-\sigma^2} - 1\big)^{-\frac{1}{\sigma^2}}
\end{aligned}$$

The advantage of model (4.5) is the nice interpretation of $(1 + \sigma^2)$ as the relative risk associated with a non-surviving relative. Thus, there are two ways of deriving (4.2) based on radically different concepts: one uses the assumption (4.5), which concerns proportionality of conditional hazards; the other uses the concept of random hazards with a gamma distributed shared frailty.

The bivariate shared frailty model can be extended to the multivariate case with p related failure times, which results in the case of gamma distributed frailty in the following unconditional multivariate survival function:

$$S(t_1, \ldots, t_p) = (\sum_{i=1}^{p} S_i(t_i)^{-\sigma^2} - p + 1)^{-1/\sigma^2},$$

which was used for example by Cook and Johnson (1981). The characterization (4.5) was extended to the multivariate case by Guo and Rodriguez (1992) and Guo (1993). The correlation between life times of randomly selected pairs is always the same, which makes the model less useful for modelling correlations in family studies with groups of different relatives (mother-father, mother-daughter, grandfather-son, brother-brother etc.).

In the shared gamma frailty model with observed covariates the frailty Z_i, $(i = 1, \ldots, n)$ in each cluster can be estimated (Nielsen et al., 1992) by

$$\hat{Z}_i = \frac{1/\sigma^2 + \sum_{j=1}^{n_i} \delta_{ij}}{1/\sigma^2 + \sum_{j=1}^{n_i} e^{\beta X_{ij}} \Lambda(t_{ij})}.$$

This is possible because of the repeated observations in each cluster, where all observations in one cluster are based on the same value of the frailty variable. This approach was used by Carvalho et al. (2003) to investigate the quality of dialysis centers (represented by the unobserved frailties) in Brazil.

Asymptotic properties of the non-parametric maximum likelihood estimates in the shared gamma frailty model are well established. Murphy shows consistency (Murphy, 1994) and asymptotic normality (Murphy, 1995) in the shared gamma frailty model without covariates. These results were generalized to the correlated gamma frailty model with observed covariates by Parner (1998). Shih and Louis (1995) proposed a graphical method for assessing the gamma distribution assumption when the baseline hazard is parametric and in the absence of covariates. Glidden (1999) suggested a test for the gamma frailty model without parameterizing the baseline hazard and without covariates in the model. A graphical as well as a numerical method for checking the adequacy of the gamma distribution in a shared frailty model is suggested by Cui and Sun (2004). Their test is based on the posterior expectation of the frailties given the observable data over time, extending the work by Glidden (1999).

4.2 Limitations of the shared frailty model

Shared (or common) frailty explains correlations within clusters (here, a cluster can consist of individuals from the same group, such as a family, litter, clinic, community; or of multiple or recurrent events from the same individual). However, it does have some limitations. In the following, we use some of the arguments by Xue and Brookmeyer (1996).

First, it forces the unobserved factors to be the same within the cluster, which is generally not appropriate. For example, in general it may be inappropriate to assume that both partners in a twin pair share all of their unobserved risk factors; for a multi-staged disease, usually heterogeneity is higher in the earlier stages and people tend to be more homogeneous in the later stages.

Second, in most cases shared frailty will only induce positive associations within the cluster (for exceptions, see Joe 1993). However, in some situations the survival times for subjects within the same cluster are negatively associated. For example, growth rates of animals living in the same litter with limited food supply are probably negatively associated. Another example are transplantation studies, which found out that generally the longer an individual must wait for a transplantation, the shorter the chances of survival after the transplantation are. Thus, the waiting and the survival time may be negatively associated. As another example, suppose a patient is repeatedly admitted to hospital for the same disease. The sicker the patient is, the higher the risk of re-admission is and the lower the 'risk' (chance) of discharge. Therefore, the duration of the stays inside hospital, on one hand, and outside hospital, on the other, are expected to be negatively associated. An additional example is provided by competing risk scenarios, where the reduction of the risk of dying from one disease increases the risk to die from another disease.

Third, 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 a gamma distributed frailty, the dependence parameter and the population heterogeneity are confounded (Clayton and Cuzick 1985), meaning that the joint distribution can be identified from the marginal distributions (Hougaard 1986a). Elbers and Ridder (1982) show that this problem exists for any univariate frailty distribution with a finite mean. However, 'shared frailty' in bivariate models differs from 'individual frailty' used in the case of univariate data. Initially this difference in the notions of frailty was not clearly understood. It is worth noting that the value of σ^2 estimated from the univariate data may, in fact, have nothing to do with association. Indeed, consider two hypothetical bivariate data sets with different associations between life spans

of respective related individuals, for example the lifetimes of men and their sons and for the same men and their grandsons. In both situations the value of the parameter σ^2 will be the same estimated from the data on the group of men (grandfathers), despite the fact that associations between life spans of grandfathers and their grandsons and grandfathers and their sons are different.

This last and maybe most important limitation of shared frailty models is a consequence of identifiability of the univariate frailty model with observed covariates. Hence, it is an inherent feature to all shared frailty models with a finite mean of frailty distribution. To overcome this problem, Hougaard (1986a, 1987) suggests the shared positive stable frailty model. In this case the univariate model with observed covariates is not identifiable because the mean of the positive stable distribution is infinite. So one can expect more flexibility from a shared frailty model with positive stable distribution than from models with gamma frailty. The bivariate survival function in the shared positive stable frailty model is

$$S(t_1, t_2) = \exp\{-\big((-\ln(S_1(t_1)))^{1/\gamma} + (-\ln(S_2(t_2)))^{1/\gamma}\big)^{\gamma}\}. \tag{4.6}$$

When covariates X_i, $(i = 1, 2)$ are observed, and the conditional hazard is (3.2) then the univariate survival functions are

$$S_i(t) = \exp\{-\Lambda_i^{\gamma}(t)e^{\gamma\beta_i X_i}\} = \exp\{-\Lambda_i^{*}(t)e^{\beta_i^{*} X_i}\},$$

where $\beta_i^* = \gamma\beta_i$ and $\Lambda_i^*(t) = \Lambda_i^{\gamma}(t)$. Using a shared positive stable frailty model one can estimate both the association parameter γ and regression coefficients β_i from data on related individuals. One problem remains yet unsolved: the interpretation of regression parameters β_i. To illustrate this problem let $\beta = \beta_1 = \beta_2$ for two groups of relatives, for example MZ and DZ twins. It is clear that the association parameter γ is different for MZ and DZ twins. Hence the values of parameter $\beta^* = \gamma\beta$ and $H^*(t)$ in equation (4.6) are also different for MZ and DZ twins, which contradicts the natural assumption that the survival of these individuals conditioned on observed covariates follow the same Cox type model. If the parameters $\beta^* = \gamma\beta$ are assumed to be the same for MZ and DZ twins then the parameters β and the baseline hazards $\lambda_0(t)$ should be different for these individuals, which creates a problem for the interpretation of the conditional hazard (3.2) for this model.

To avoid such methodological problems, correlated frailty models have been developed for the analysis of multivariate failure time data. These models have two associated random variables that characterize the frailty effect for each cluster. One random variable, for example, is assigned to twin 1 and one to twin 2 so that they are no longer

constrained by having a common frailty. The two variables are associated and jointly distributed, therefore knowing one of them does not necessarily imply the other. Also, the two variables can certainly be negatively associated, and this would induce a negative association between survival times.

Between shared and correlated frailty models (sometimes also called univariate and bivariate frailty models with respect to the dimension of the frailty, but both models deal with bivariate time-to-event data), there is an intermediate approach in which two variables are assigned for each cluster to count for heterogeneity; but they are generated from one common random variable. For example, a bivariate survival time (T_1, T_2) with Z_1 to account for heterogeneity of T_1 and Z_2 for T_2, where Z_1 and Z_2 are defined as follows:

$$Z_1 = e^{\alpha W} \quad \text{and} \quad Z_2 = e^{\beta W} \tag{4.7}$$

in which W is a random variable and α and β are parameters. This one factor error specification has been used frequently (Flinn and Heckman, 1982; Clayton and Cuzick, 1985; Heckman and Walker, 1990; Huang and Wolfe, 2002). This approach is more flexible than that of assuming shared frailty for T_1 and T_2, and to some extent it allows for a negative association by allowing different signs of α and β. However, it still imposes a linkage between the variance and the correlation, and also the mean and the variance. The overall hazard for the population (rather than the individual hazard) in consequence is forced to move between restricted ranges. Therefore, Lindeboom and Van Den Berg (1994) concluded that it is hazardous to estimate bivariate survival models in which the mixing distribution is parameterized univariately, in that a univariate random variable may not be able to account for both the dependence of the survival times and for the change in the composition of the sample due to the unobserved heterogeneity.

Clearly, such kinds of problems do not arise in a genuine bivariate approach in which the dependence of T_1 and T_2 can be changed without changing the marginal distribution of T_1 and T_2. Aalen (1987) discussed multivariate mixing distributions applied on a Markov Chain. Marshall and Olkin (1988) discussed various multivariate correlated frailty distributions. Their work was continued by Yashin et al. (1993, 1995), who considered the bivariate correlated gamma frailty model and applied it to twin survival data. This approach forms the starting point for different extensions analyzed in this thesis which will be discussed in more detail after the following section.

4.3 Correlated frailty model

Consider some bivariate observations, e.g., the life times of twins, or age at onset of a disease in spouses, time to disease in paired organs like kidneys or eyes etc. In the (bivariate) correlated frailty model the frailty of each individual in a pair is defined by a measure of relative risk, i.e., exactly as it was defined in the univariate case. For two individuals in a pair, frailties are not necessarily the same, as they are in the shared frailty model. We are assuming that the frailties are acting multiplicatively on the baseline hazard function and that the observations in a pair are conditionally independent, given the frailties. Hence, the hazard of individual j ($j = 1, 2$) in pair i ($i = 1, \ldots, n$) has the form

$$\lambda(t) = Z_{ij}\lambda_{0j}(t)e^{\beta^T X_{ij}}, \tag{4.8}$$

where t denotes age or time, X_{ij} is a vector of observable covariates, β is a vector of unknown regression coefficients describing the effect of the covariates X_{ij}, $\lambda_{0j}(t)$ are baseline hazard functions, and Z_{ij} are unobserved (random) effect or frailty. Bivariate correlated frailty models are characterized by the joint distribution of a two-dimensional vector of frailties (Z_{i1}, Z_{i2}).

Any method in this context is based on likelihood functions. In order to derive a marginal likelihood function, the assumption of conditional independence of life spans given the frailty is always used. Let δ_{ij} be a censoring indicator for an individual j ($j = 1, 2$) in pair i ($i = 1, \ldots, n$). Indicator δ_{ij} is 1 if the individual has experienced the event of interest, and 0 otherwise. According to (4.8), the conditional survival function of the j-th individual in the i-th pair is

$$S(t|Z_{ij}, X_{ij}) = e^{Z_{ij}\Lambda_{0j}(t)e^{\beta^T X_{ij}}}, \tag{4.9}$$

where $\Lambda_{0j}(t)$ is the cumulative baseline hazard function. Here and in the following, S is used as a generic symbol for a survival function. The contribution of the j-th individual in the i-th pair of the conditional likelihood is given by

$$L(t_{ij}, \delta_{ij}|Z_{ij}, X_{ij}) = \left(Z_{ij}\lambda_{0j}(t_{ij})e^{\beta^T X_{ij}}\right)^{\delta_{ij}} e^{Z_{ij}\Lambda_{0j}(t_{ij})e^{\beta^T X_{ij}}}, \tag{4.10}$$

where t_{ij} stands for age at death or the censoring time of individual j from pair i. Then, assuming the conditional independence of life spans given the frailty and integrating out the frailty, we obtain the marginal likelihood function:

$$L(t,\delta|X)=\prod_{i=1}^{n}\iint\limits_{R^+\times R^+}\Big(z_{i1}\lambda_{01}(t_{i1})e^{\beta^T X_{i1}}\Big)^{\delta_{i1}}e^{z_{i1}\Lambda_{01}(t_{i1})e^{\beta^T X_{i1}}} \tag{4.11}$$
$$*\Big(z_{i2}\lambda_{02}(t_{i2})e^{\beta^T X_{i2}}\Big)^{\delta_{i2}}e^{z_{i2}\Lambda_{02}(t_{i2})e^{\beta^T X_{i2}}}f_Z(z_{i1},z_{i2})\,dz_{i1}\,dz_{i2},$$

where $t=(t_1,\ldots,t_n)$, $t_i=(t_{i1},t_{i2})$, $\delta=(\delta_1,\ldots,\delta_n)$, $\delta_i=(\delta_{i1},\delta_{i2})$, $X=(X_1,\ldots,X_n)$, $X_i=(X_{i1},X_{i2})$, and $f_Z(\cdot,\cdot)$ is the probability density function of the corresponding frailty distribution.

A different approach to bivariate frailty modelling was used by Bandyopadhyay and Basu (1990) and Gupta and Gupta (1990). The key idea of their (more specific) model is a bivariate hazard model:

$$\lambda(t_1,t_2,Z)=Z\lambda_0(t_1,t_2).$$

Here the two lifetimes are not conditionally independent given the frailty. Dependence is caused by the bivariate model $\lambda_0(t_1,t_2)$ as well as by integrating out the frailty Z.

As in univariate frailty models, the choice of the frailty distribution is important for modelling. Below three correlated frailty models are considered: the correlated gamma frailty model and the correlated log-normal frailty model as the two most popular. Furthermore, a new correlated frailty model based on the compound Poisson distribution is suggested.

4.4 Correlated gamma frailty model

This model was introduced by Yashin et al. (1993, 1995) and applied to related lifetimes in many different settings, for example by Pickles et al. (1994), Yashin and Iachine (1995a,b, 1997, 1999a,b), Yashin et al. (1996), Iachine et al. (1998), Iachine (2002), Petersen (1998), Wienke et al. (2000, 2001, 2002, 2003a,b, 2004, 2005a), Zdravkovic et al. (2002, 2004).

Let k_0,k_1,k_2 be some real positive variables. Set $\lambda_1=k_0+k_1$ and $\lambda_2=k_0+k_2$. Let Y_0,Y_1,Y_2 be independently gamma distributed random variables with $Y_0\sim\Gamma(k_0,\lambda_0)$, $Y_1\sim\Gamma(k_1,\lambda_1)$, $Y_2\sim\Gamma(k_2,\lambda_2)$. Consequently,

$$Z_1=\frac{\lambda_0}{\lambda_1}Y_0+Y_1\sim\Gamma(k_0+k_1,\lambda_1) \tag{4.12}$$
$$Z_2=\frac{\lambda_0}{\lambda_2}Y_0+Y_2\sim\Gamma(k_0+k_2,\lambda_2) \tag{4.13}$$

and $\mathbf{E}Z_1=\mathbf{E}Z_2=1,\mathbf{V}(Z_1)=\frac{1}{\lambda_1}:=\sigma_1^2$, $\mathbf{V}(Z_2)=\frac{1}{\lambda_2}:=\sigma_2^2$.

The following relations hold:

$$\mathbf{E}Y_0^2 = \mathbf{V}(Y_0) + (\mathbf{E}Y_0)^2 = \frac{k_0}{\lambda_0^2} + \big(\frac{k_0}{\lambda_0}\big)^2 = \frac{k_0^2 + k_0}{\lambda_0^2}$$

$$\begin{aligned}
\mathbf{E}Z_1Z_2 &= \mathbf{E}(\frac{\lambda_0}{\lambda_1}Y_0 + Y_1)(\frac{\lambda_0}{\lambda_2}Y_0 + Y_2) \\
&= \mathbf{E}(\frac{\lambda_0^2}{\lambda_1\lambda_2}Y_0^2 + \frac{\lambda_0}{\lambda_1}Y_0Y_2 + \frac{\lambda_0}{\lambda_2}Y_0Y_1 + Y_1Y_2) \\
&= \frac{\lambda_0^2}{\lambda_1\lambda_2}\frac{k_0^2+k_0}{\lambda_0^2} + \frac{\lambda_0}{\lambda_1}\frac{k_0k_2}{\lambda_0\lambda_2} + \frac{\lambda_0}{\lambda_2}\frac{k_0k_1}{\lambda_0\lambda_1} + \frac{k_1k_2}{\lambda_1\lambda_2} \\
&= \frac{k_0 + (k_0+k_1)(k_0+k_2)}{\lambda_1\lambda_2} \\
&= \frac{k_0}{(k_0+k_1)(k_0+k_2)} + 1
\end{aligned}$$

$$\mathbf{cov}(Z_1, Z_2) = \mathbf{E}Z_1Z_2 - \mathbf{E}Z_1\mathbf{E}Z_2 = \frac{k_0}{(k_0+k_1)(k_0+k_2)}.$$

This leads to the correlation

$$\rho = \frac{\mathbf{cov}(Z_1, Z_2)}{\sqrt{\mathbf{V}(Z_1)\mathbf{V}(Z_2)}} = \frac{k_0}{\sqrt{(k_0+k_1)(k_0+k_2)}}. \tag{4.14}$$

Consequently, because of relation $k_0 + k_i = \lambda_i = \frac{1}{\sigma_i^2}$, $(i = 1, 2)$ it holds that $k_0 = \frac{\rho}{\sigma_1\sigma_2}$ and $k_i = \frac{1}{\sigma_i^2} - k_0 = \frac{1-\frac{\sigma_i}{\sigma_j}\rho}{\sigma_i^2}$ $(i, j = 1, 2;\ i \neq j)$.
Now we can derive the unconditional model, applying the Laplace transform of gamma distributed random variables. Hence,

$$\begin{aligned}
S(t_1, t_2) &= \mathbf{E}S(t_1, t_2|Z_1, Z_2) \qquad (4.15) \\
&= \mathbf{E}S_1(t_1|Z_1)S_2(t_2|Z_2) \\
&= \mathbf{E}e^{-Z_1\Lambda_1(t_1)}e^{-Z_2\Lambda_2(t_2)} \\
&= \mathbf{E}e^{-(\frac{\lambda_0}{\lambda_1}Y_0+Y_1)\Lambda_1(t_1)}e^{-(\frac{\lambda_0}{\lambda_2}Y_0+Y_2)\Lambda_2(t_2)} \\
&= \mathbf{E}e^{-Y_0(\frac{\lambda_0}{\lambda_1}\Lambda_1(t_1)+\frac{\lambda_0}{\lambda_2}\Lambda_2(t_2))-Y_1\Lambda_1(t_1)-Y_2\Lambda_2(t_2)} \\
&= (1 + \frac{1}{\lambda_0}(\frac{\lambda_0}{\lambda_1}\Lambda_1(t_1) + \frac{\lambda_0}{\lambda_2}\Lambda_2(t_2)))^{-k_0}(1 + \frac{1}{\lambda_1}\Lambda_1(t_1))^{-k_1}(1 + \frac{1}{\lambda_2}\Lambda_2(t_2))^{-k_2} \\
&= (1 + \sigma_1^2\Lambda_1(t_1) + \sigma_2^2\Lambda_2(t_2))^{\frac{-\rho}{\sigma_1\sigma_2}}(1 + \sigma_1^2\Lambda_1(t_1))^{\frac{-1+\frac{\sigma_1}{\sigma_2}\rho}{\sigma_1^2}}(1 + \sigma_2^2\Lambda_1(t_2))^{\frac{-1+\frac{\sigma_2}{\sigma_1}\rho}{\sigma_2^2}}
\end{aligned}$$

which results in the following representation of the correlated gamma frailty model:

$$S(t_1, t_2) = \frac{S_1(t_1)^{1-\frac{\sigma_1}{\sigma_2}\rho}S_2(t_2)^{1-\frac{\sigma_2}{\sigma_1}\rho}}{(S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{\frac{\rho}{\sigma_1\sigma_2}}}, \tag{4.16}$$

using the independence of the gamma distributed random variables Y_0, Y_1, Y_2 and (3.8). The range of the correlation between frailties depends on the values of σ_1 and σ_2:

$$0 \leq \rho \leq \min\{\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\}.$$

Hence, if $\sigma_1 \neq \sigma_2$, it is always less than one. This property can be a serious limitation when the values of σ_1 and σ_2 differ strongly.

Partial derivatives of the bivariate survival function (which are necessary for the likelihood) are given in Appendix C. The log-likelihood in case of non-censored observations is given by:

$$\begin{aligned}
\ln L &= \frac{\rho}{\sigma_1\sigma_2}\ln(S_1(t_1)^{-\sigma_1^2}) + \frac{\rho}{\sigma_1\sigma_2}\ln(S_2(t_2)^{-\sigma_2^2}) \\
&- (\frac{\rho}{\sigma_1\sigma_2}+2)\ln(S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1) \\
&+ \ln\big(1 - \frac{\sigma_1}{\sigma_2}\rho - \frac{\sigma_2}{\sigma_1}\rho + \rho^2 + (1-\frac{\sigma_2}{\sigma_1}\rho)S_1(t_1)^{-2\sigma_1^2} + (1-\frac{\sigma_1}{\sigma_2}\rho)S_2(t_2)^{-2\sigma_2^2} \\
&+ (2 - \frac{\sigma_1}{\sigma_2}\rho - \frac{\sigma_2}{\sigma_1}\rho + \sigma_1\sigma_2\rho + \rho^2)S_1(t_1)^{-\sigma_1^2}S_2(t_2)^{-\sigma_2^2} \\
&+ (-2 + \frac{\sigma_1}{\sigma_2}\rho + 2\frac{\sigma_2}{\sigma_1}\rho - \rho^2)S_1(t_1)^{-\sigma_1^2} + (-2 + \frac{\sigma_2}{\sigma_1}\rho + 2\frac{\sigma_1}{\sigma_2}\rho - \rho^2)S_2(t_2)^{-\sigma_2^2}\big)
\end{aligned}$$

In the following we apply the correlated gamma frailty model to cause-specific mortality data of Danish twins, described in more detail in Appendix A.2. All cause-mortality, death by cancer (all cases of cancer combined), coronary heart disease (CHD), stroke and respiratory diseases are considered. Because of the symmetric structure of the twin data we use the simplification $S(t) = S_1(t) = S_2(t)$ and $\sigma^2 = \sigma_1^2 = \sigma_2^2$. We use both a parametric and a semi-parametric approach. In the parametric approach a Gamma-Gompertz model is applied, e.g.

$$S(t) = (1 + s^2\frac{a}{b}(e^{bt} - 1))^{-\frac{1}{s^2}},$$

in the semi-parametric model the univariate marginal survival is left unspecified. The data is right censored, which has to be included into the likelihood. The results of the maximum likelihood parameter estimation procedure are given in Tables 4.1 and 4.2 for males and females, respectively.

For each cause of death a separate analysis of the same data was performed, treating all other causes of death as independent censored observations. In all considered cases of cause-specific mortality the estimates of correlations of frailty for MZ twins (ρ_{MZ}) tend to be higher than for DZ twins (ρ_{DZ}) with one exception - the correlation for respiratory diseases is nearly equal for MZ and DZ twins in males. Higher correlations in MZ twins

model	a	b	s	σ	ρ_{MZ}	ρ_{DZ}
all causes (4877 deaths)						
parametric model	5.31e-5	0.094	0.187	1.333	0.573	0.340
	(6.73e-6)	(0.002)	(0.088)	(0.177)	(0.071)	(0.056)
semi-parametric model				1.318	0.563	0.337
				(0.177)	(0.073)	(0.057)
cancer (1249 deaths)						
parametric model	4.71e-6	0.117	1.806	2.431	0.275	0.148
	(1.61e-6)	(0.006)	(0.160)	(0.647)	(0.090)	(0.061)
semi-parametric model				2.411	0.267	0.147
				(0.637)	(0.090)	(0.061)
CHD (1403 deaths)						
parametric model	3.34e-6	0.117	0.812	2.315	0.538	0.242
	(9.70e-7)	(0.005)	(0.149)	(0.493)	(0.106)	(0.074)
semi-parametric model				2.433	0.510	0.228
				(0.489)	(0.099)	(0.068)
stroke (439 deaths)						
parametric model	8.41e-8	0.154	2.068	5.202	0.422	0.098
	(5.52e-8)	(0.010)	(0.318)	(1.353)	(0.104)	(0.063)
semi-parametric model				4.404	0.453	0.106
				(1.014)	(0.113)	(0.066)
resp. diseases (346 deaths)						
parametric model	2.44e-7	0.133	1.130	3.059	0.415	0.425
	(1.44e-7)	(0.009)	(0.548)	(1.557)	(0.291)	(0.255)
semi-parametric model				3.155	0.384	0.402
				(1.484)	(0.253)	(0.232)

Table 4.1: Parameter estimates (s.e.) in the correlated gamma frailty model applied to 3755 Danish male twin pairs.

compared to DZ twins indicate the influence of genetic factors in frailty to cause-specific and total mortality. Estimates of correlations range between 0.01 (respiratory diseases, DZ female twins) and 0.62 (CHD, MZ female twins). Large differences in the estimates of the standard deviation σ indicate different levels of heterogeneity. Especially for stroke and respiratory diseases in females, heterogeneity seems to be huge. Another important aspect is the similarity of the results in parametric and semi-parametric analysis. The estimates are nearly identical, which supports the choice of the Gompertz baseline hazard in the parametric model. A re-analysis of this data to get quantitative results about the heritability of cause-specific mortality can be found in Wienke (2004).

It is necessary to keep in mind that the parameter ρ describes the correlation between the frailties and not the correlation of the respective lifetimes. Lindeboom and Van Den Berg (1994) analyzed the relation of correlation between frailties and between lifetimes.

model	a	b	s	σ	ρ_{MZ}	ρ_{DZ}
all causes (4493 deaths)						
parametric model	3.17e-5	0.095	0.002	1.168	0.612	0.276
	(2.58e-5)	(0.001)	(0.000)	(0.228)	(0.122)	(0.073)
semi-parametric model				1.208	0.566	0.258
				(0.224)	(0.106)	(0.066)
cancer (1246 deaths)						
parametric model	7.98e-6	0.110	2.642	5.088	0.163	0.054
	(3.07e-6)	(0.007)	(0.223)	(1.214)	(0.049)	(0.026)
semi-parametric model				5.432	0.148	0.051
				(1.293)	(0.046)	(0.024)
CHD (1073 deaths)						
parametric model	2.10e-7	0.144	0.841	1.906	0.618	0.230
	(8.00e-8)	(0.005)	(0.156)	(0.404)	(0.141)	(0.089)
semi-parametric model				1.996	0.600	0.220
				(0.412)	(0.134)	(0.084)
stroke (521 deaths)						
parametric model	1.30e-7	0.141	1.226	9.149	0.239	0.046
	(7.00e-8)	(0.008)	(0.322)	(2.006)	(0.072)	(0.039)
semi-parametric model				9.751	0.237	0.041
				(2.019)	(0.069)	(0.034)
resp. diseases (294 deaths)						
parametric model	4.90e-7	0.114	0.041	10.385	0.209	0.016
	(1.70e-7)	(0.004)	(0.033)	(3.680)	(0.097)	(0.038)
semi-parametric model				10.684	0.189	0.014
				(3.814)	(0.091)	(0.034)

Table 4.2: Parameter estimates (s.e.) in the correlated gamma frailty model applied to 4200 Danish female twin pairs.

The bivariate correlated gamma frailty model can be extended to the case with p related life spans, which results in the gamma distributed frailty case in

$$S(t_1, \ldots, t_p) = \prod_{i=1}^{p} S_i(t_i)^{1-\rho} \Big(\sum_{i=1}^{p} S_i(t_i)^{-\sigma^2} - p + 1\Big)^{-\rho/\sigma^2}.$$

In this model, identical marginal frailty distributions are assumed, and all possible correlations between frailties in a group of individuals are described by parameter ρ. Further extensions of this model can be found in Yashin and Iachine (1999a), but they need additional requirements, which may be rather restrictive for real applications.

Parner (1998) proved consistency and asymptotic normality of the non-parametric maximum likelihood estimator in the multivariate correlated gamma frailty model with observed covariates.

Paik et al. (1994) provided an extension of the correlated gamma frailty model. It uses a decomposition of the frailty into a shared and unique part, e.g. $Z_1 = Y_0 + Y_1$ and $Z_2 = Y_0 + Y_2$ with independent random variables $Y_0 \sim \Gamma(k_0, \lambda_0), Y_1 \sim \Gamma(k_1, \lambda_1)$ and $Y_2 \sim \Gamma(k_2, \lambda_2)$. In the correlated gamma frailty model, the assumption $\lambda_0 = \lambda_1 = \lambda_2$ is essential, because it yields the gamma distribution of Z_1 and Z_2. Paik et al. allow the λ's to differ, which on the one hand complicates the computation of the likelihood function, yet on the other hand generates a more flexible model. In both models, the restriction $\mathbf{E}Z_i = 1\,(i = 1, 2)$ is imposed. The model is only identifiable with repeated (at least two) observations per individual. The main feature of the model is to allow the frailty (and consequently the dependence function) to vary over time intervals. Restricting the model to two related individuals with two recurrent observation times each gives the following structure of frailties, shown in Figure 4.1: Note that Paik et al. (1994) considered a non-

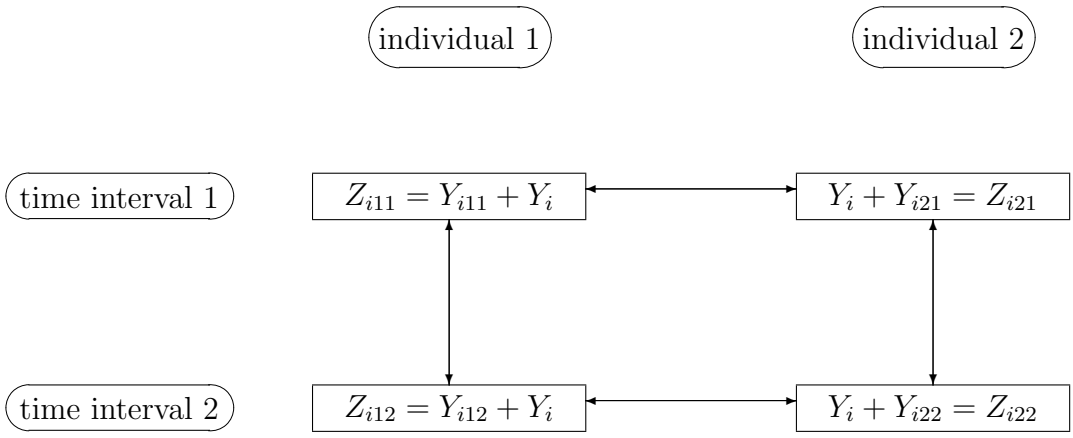

Figure 4.1: Time-varying frailties for two relatives from family i in the Paik model

competing risk situation, that means it is assumed that all four life times are observable (with independent censoring). Different extensions to the Paik model are discussed in Wintrebert et al. (2004).

In most applications of frailty models, a parametric approach is used, which means that the baseline hazard function is specified up to a finite dimensional parameter (e.g. Weibull or Gompertz). The main advantage of multivariate frailty models when compared with univariate frailty models is that it is possible to relax the parametric assumption about the baseline hazard function in a way similar to the Cox regression model. In semi-parametric frailty models no parametric assumption about the form of the baseline hazard function is necessary. This causes problems in the calculation of the exact standard errors of the estimates. Andersen et al. (1997) analyzed this problem in detail.

4.5 Correlated compound Poisson frailty model

This model is an extension of the correlated PVF frailty model (correlated three-parameter frailty model) suggested by Yashin et al. (1999a) and has not yet been considered in bivariate survival studies. It is based on a bivariate extension of the compound Poisson frailty model introduced to univariate survival analysis by Aalen (1988, 1992). It is also related to the correlated gamma frailty cure model by Wienke et al. (2003a), which allows for a non-susceptible fraction in the population, as well as the compound Poisson frailty models with a random scale considered by Moger et al. (2004b) and Moger and Aalen (2005).

Let k_0, k_1 be some real positive variables and let Y_0, Y_1, Y_2 be independently compound Poisson distributed random variables with $Y_0 \sim cP(\gamma, k_0, \lambda)$, $Y_1 \sim cP(\gamma, k_1, \lambda)$ and $Y_2 \sim cP(\gamma, k_1, \lambda)$. Consequently, using a similar additive structure for the frailties as in (4.12) it holds that

$$Z_1 = Y_0 + Y_1 \sim cP(\gamma, k_0 + k_1, \lambda)$$
$$Z_2 = Y_0 + Y_2 \sim cP(\gamma, k_0 + k_1, \lambda).$$

For simplicity, we shall present only the symmetric case here, where the two life times are interchangeable. An extension to the non-symmetric case is straightforward.
Furthermore, the following relations are assumed:

$$\mathbf{E}Z_1 = \mathbf{E}Z_2 = 1, \mathbf{V}(Z_1) = \mathbf{V}(Z_2) = \sigma^2.$$

This implies (see 3.11) $(k_0 + k_1)\lambda^{\gamma-1} = 1$ and $(k_0 + k_1)(1-\gamma)\lambda^{\gamma-2} = \sigma^2$. Consequently, $(k_0 + k_1)\lambda^{\gamma-2} = 1/\lambda$ and $(k_0 + k_1)\lambda^{\gamma-2} = \frac{\sigma^2}{1-\gamma} = \frac{1}{\lambda}$. Hence, $\lambda = \frac{1-\gamma}{\sigma^2}$, which results in

$$(k_0 + k_1)\lambda^{\gamma} = \lambda = \frac{1-\gamma}{\sigma^2} \tag{4.17}$$

It holds that

$$\mathbf{E}Y_0^2 = \mathbf{V}(Y_0) + (\mathbf{E}Y_0)^2 = k_0(1-\gamma)\lambda^{\gamma-2} + (k_0\lambda^{\gamma-1})^2$$

$$\begin{aligned}
\mathbf{E}Z_1Z_2 &= \mathbf{E}(Y_0 + Y_1)(Y_0 + Y_2) \\
&= \mathbf{E}(Y_0^2 + Y_0Y_1 + Y_0Y_2 + Y_1Y_2) \\
&= k_0(1-\gamma)\lambda^{\gamma-2} + k_0^2\lambda^{2\gamma-2} + k_0k_1\lambda^{2\gamma-2} + k_0k_1\lambda^{2\gamma-2} + k_1^2\lambda^{2\gamma-2} \\
&= k_0(1-\gamma)\lambda^{\gamma-2} + (k_0 + k_1)^2\lambda^{2\gamma-2} \\
&= k_0(1-\gamma)\lambda^{\gamma-2} + 1
\end{aligned}$$

$$\mathbf{cov}(Z_1, Z_2) = \mathbf{E}Z_1Z_2 - \mathbf{E}Z_1\mathbf{E}Z_2 = k_0(1-\gamma)\lambda^{\gamma-2}.$$

This leads to the correlation

$$\rho = \frac{\mathbf{cov}(Z_1, Z_2)}{\sqrt{\mathbf{V}(Z_1)\mathbf{V}(Z_2)}} = \frac{k_0(1-\gamma)\lambda^{\gamma-2}}{(k_0+k_1)(1-\gamma)\lambda^{\gamma-2}} = \frac{k_0}{k_0+k_1}. \tag{4.18}$$

Consequently, because of (4.17) and (4.18)

$$k_0\lambda^\gamma = \frac{k_0}{k_0+k_1}(k_0+k_1)\lambda^\gamma = \rho\frac{1-\gamma}{\sigma^2}. \tag{4.19}$$

Now we can derive the unconditional model, applying the Laplace transform of compound Poisson distributed random variables (3.13). Hence,

$$\begin{aligned}
S(t_1,t_2) &= \mathbf{E}S(t_1,t_2|Z_1,Z_2) \\
&= \mathbf{E}S(t_1|Z_1)S(t_2|Z_2) \\
&= \mathbf{E}e^{-Z_1\Lambda_0(t_1)}e^{-Z_2\Lambda_0(t_2)} \\
&= \mathbf{E}e^{-(Y_0+Y_1)\Lambda_0(t_1)}e^{-(Y_0+Y_2)\Lambda_0(t_2)} \\
&= \mathbf{E}e^{-Y_0(\Lambda_0(t_1)+\Lambda_0(t_2))-Y_1\Lambda_0(t_1)-Y_2\Lambda_0(t_2)} \\
&= e^{-\frac{k_0}{\gamma}((\lambda+\Lambda_0(t_1)+\Lambda_0(t_2))^\gamma-\lambda^\gamma)}e^{-\frac{k_1}{\gamma}((\lambda+\Lambda_0(t_1))^\gamma-\lambda^\gamma)}e^{-\frac{k_1}{\gamma}((\lambda+\Lambda_0(t_2))^\gamma-\lambda^\gamma)}.
\end{aligned} \tag{4.20}$$

The three terms are considered in detail. For the marginal survival function holds that

$$S(t) = e^{-\frac{k_0+k_1}{\gamma}((\lambda+\Lambda_0(t))^\gamma-\lambda^\gamma)}, \tag{4.21}$$

which implies

$$\lambda + \Lambda_0(t) = (\lambda^\gamma - \frac{\gamma}{k_0+k_1}\ln S(t))^{1/\gamma}. \tag{4.22}$$

Hence, using (4.18) and (4.21)

$$\begin{aligned}
e^{-\frac{k_1}{\gamma}((\lambda+\Lambda_0(t))^\gamma-\lambda^\gamma)} &= e^{-\frac{k_1}{k_0+k_1}\frac{k_0+k_1}{\gamma}((\lambda+\Lambda_0(t))^\gamma-\lambda^\gamma)} \\
&= e^{-(1-\rho)\frac{k_0+k_1}{\gamma}((\lambda+\Lambda_0(t))^\gamma-\lambda^\gamma)} \\
&= (e^{-\frac{k_0+k_1}{\gamma}((\lambda+\Lambda_0(t))^\gamma-\lambda^\gamma)})^{1-\rho} \\
&= S(t)^{1-\rho}.
\end{aligned}$$

For the first term in (4.20) holds because of (4.22)

$$\begin{aligned}
e^{-\frac{k_0}{\gamma}((\lambda+\Lambda_0(t_1)+\Lambda_0(t_2))^\gamma-\lambda^\gamma)} &= e^{-\frac{k_0}{\gamma}((\lambda+\Lambda_0(t_1)+\lambda+\Lambda_0(t_2)-\lambda)^\gamma-\lambda^\gamma)} \\
&= e^{-\frac{k_0}{\gamma}(((\lambda^\gamma-\frac{\gamma}{k_0+k_1}\ln(S(t_1)))^{1/\gamma}+(\lambda^\gamma-\frac{\gamma}{k_0+k_1}\ln(S(t_2)))^{1/\gamma}-\lambda)^\gamma-\lambda^\gamma)} \\
&= e^{\frac{k_0\lambda^\gamma}{\gamma}(1-((1-\frac{\gamma}{(k_0+k_1)\lambda^\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma}{(k_0+k_1)\lambda^\gamma}\ln(S(t_2)))^{1/\gamma}-1)^\gamma)},
\end{aligned}$$

which results because of (4.17) and (4.19) in the following representation of the correlated compound Poisson frailty model:

$$S(t_1,t_2) = S(t_1)^{1-\rho}S(t_2)^{1-\rho} \\ * \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\}.$$

This model includes both gamma and inverse Gaussian correlated frailty models as special cases, the gamma model corresponds to $\gamma = 0$, the inverse Gaussian model has $\gamma = 0.5$. For $\gamma \geq 0$ the correlated PVF frailty model is obtained, which was introduced by Yashin et al. (1999a). For $\gamma < 0$ the correlated compound Poisson frailty model is obtained, including a fraction of individuals in the population with zero frailty. The model is successfully applied to breast cancer incidence data by Wienke et al. (2006a,c) for identical (MZ) and fraternal (DZ) female Swedish twins born 1886 - 1925. This data set contains records of 2003 MZ twin pairs and 3854 DZ twin pairs. Individuals were followed up from 1959/61 to 27 October 2000, and 715 cases of breast cancer were identified during the follow-up (see Appendix A.1). Age at onset of breast cancer ranges from 36 years to 93 years. More detailed information about the construction of the Swedish Twin Register can be found in Lichtenstein et al. (2002). The results of the analysis are given in Table 4.3:

	gamma frailty	inverse Gaussian frailty	compound Poisson frailty
a	2.50e-7 (1.60e-7)	6.25e-6 (1.66e-6)	3.20e-7 (2.30e-07)
b	0.161 (0.012)	0.097 (0.005)	0.155 (0.015)
γ	0	0.5	-0.052 (0.096)
σ	7.612 (0.470)	7.406 (1.084)	7.034 (0.994)
ρ_{MZ}	0.124 (0.040)	0.199 (0.064)	0.124 (0.040)
ρ_{DZ}	0.097 (0.029)	0.161 (0.047)	0.099 (0.030)
susceptible	100%	100%	33.6%
likelihood	-5218.6397	-5265.7593	-5218.4581

Table 4.3: Parameter estimates (s.e.) in the analysis of time to onset of breast cancer in 5857 Swedish twin pairs. The size of the susceptible fraction is calculated by $1 - e^{\frac{1-\gamma}{\gamma\sigma^2}}$.

Parameters a and b denote the parameters of the Gompertz baseline hazard. The model in the first column is the correlated gamma frailty model (4.4). Parameter σ^2 is a measure of heterogeneity, which is large in these data, and all individuals of the population are assumed to be susceptible to breast cancer. The correlated gamma frailty model is a special case of the correlated compound Poisson/PVF frailty model with $\gamma = 0$.

In the second model inverse Gaussian distributed frailty is used to account for heterogeneity in the population and to model dependence between times to onset of breast cancer in twin pairs. Parameter estimates are different from those in the gamma case with smaller heterogeneity and larger correlations. As in the first model, all women are assumed to be susceptible to breast cancer. The correlated inverse Gaussian frailty model is a special case of the correlated compound Poisson/PVF frailty model with $\gamma = 0.5$.

The most interesting parameter in the compound Poisson model is γ, which is negative ($\gamma = -0.052$) and indicates the existence of a fraction of individuals who are nonsusceptible to breast cancer. The size of the susceptible fraction depends on γ and σ^2 and is around 0.336. The estimate of the size of a susceptible fraction (due to breast cancer) is larger than the estimate 0.22 found by Chatterjee and Shih (2001) in a study population that is completely different and by Wienke et al. (2003a) applying a correlated gamma frailty cure model to the same data. The compound Poisson model fits the data better than the two submodels, but on the costs of an additional parameter. Compared to the gamma model the improvement is not significant.

Additionally, the estimate of the size of the susceptible fraction in the compound Poisson frailty model is not far from the range of the figures obtained by Farewell et al. (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.

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 %). A disadvantage of the model is that it is not able to handle negative dependencies.

The newly introduced correlated compound Poisson/PVF frailty model offers a very elegant approach to integrating the concept of cure models into frailty modelling. The likelihood function is explicitly available in a very simple form and is the most important advantage of the proposed model compared to the model suggested by Moger and Aalen (2005). Popular frailty models like the correlated and shared gamma ($\gamma = 0$) and inverse Gaussian model ($\gamma = 0.5$) are included in this model family and provide a great flexibility of the model. Because of the extension to negative values of γ the gamma distribution ($\gamma = 0$) as the most popular frailty distribution is no longer on the border of the parameter space. Consequently, traditional tests can be applied to test hypotheses about the frailty distribution (for example $H_0 : \gamma = 0$ versus $H_A : \gamma \neq 0$). Simulation studies show a good performance of the parameter estimates in this model with nearly no bias.

4.6 Correlated log-normal frailty model

The log-normal model is much more flexible than the gamma model, because it is not based on the additive composition of the two frailties as used in (4.12) and (4.13). However, the log-normal distribution does not allow for an explicit representation of the likelihood function, which requires more sophisticated estimation strategies. The distribution can be obtained by assuming a bivariate normal distribution on the logarithm of the frailty vector

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \ln \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \text{with} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim \mathrm{N} \left(\begin{pmatrix} m \\ m \end{pmatrix}, \begin{pmatrix} s^2 & rs^2 \\ rs^2 & s^2 \end{pmatrix} \right), \tag{4.23}$$

with N denoting the bivariate normal distribution whose parameters are some functions of the frailty parameters σ^2 and ρ (see for example Hutchinson and Lai, 1991):

$$\mu = \mathbf{E} Z_j = e^{m + \frac{s^2}{2}} \tag{4.24}$$

$$\sigma^2 = \mathbf{V}(Z_j) = e^{2m+s^2}(e^{s^2} - 1) \tag{4.25}$$

$$\rho = \mathbf{corr}(Z_1, Z_2) = \frac{e^{rs^2} - 1}{e^{s^2} - 1}. \tag{4.26}$$

Two different types of log-normal frailty models arise from two restrictions on the parameters of frailty distribution. First, one can use the restriction $m = 0$. This means that the logarithm of frailty has a mean of zero. In this case, a 'standard' individual has the logarithm of the hazard rate which is equal to $\ln \lambda_0(t)$. Any individual in a population has the logarithm of the hazard rate distorted by some random variables $W_j = \ln Z_j$. This value is added to the 'true' logarithm of the hazard rate $\ln \lambda_0(t)$ to provide the logarithm of hazard rate of the individual. In this interpretation, it is natural to assume that the distortions W_j have a normal distribution with a mean of zero. Second, following the usual definition of frailty used in demography (Vaupel et al., 1979; Clayton, 1978), one can use $\mu = 1$. It follows from (4.24) - (4.26) that

$$m = \mathbf{E} \ln Z_j = -\frac{1}{2} s^2 \tag{4.27}$$

$$s^2 = \mathbf{V}(\ln Z_j) = \ln(1 + \sigma^2)\,. \tag{4.28}$$

In this model, a 'standard' individual has the hazard rate $\lambda_0(t)$. Individual j in the i-th pair has the hazard rate of a 'standard' individual multiplied by the frailty Z_{ij}. The above restriction on μ means that the average frailty in a population equals 1 (at the beginning of the follow-up).

The shared log-normal model was applied by, for example, McGilchrist and Aisbett (1991), McGilchrist (1993), Gustafson (1997) and Bellamy et al. (2004). In the latter paper, the case of interval censored data is considered assuming an underlying Weibull baseline hazard function. The model was extended to allow for heterogeneity in the frailty distribution (dispersed frailty) by Lee and Lee (2003).
Xue and Brookmeyer (1996) were the first to consider the correlated log-normal frailty model and applied it to mental health data to evaluate the health policy effects for in-patient psychiatric care. Yau and McGilchrist (1997) used a GLMM approach to analyze data from litter matched tumorigenesis experiments in rats with help of the correlated log-normal model. Cook et al. (1999) used a correlated log-normal frailty in two-state mixed renewal processes for chronic disease. Other examples can be found in Ripatti and Palmgren (2000) and Ripatti et al. (2002). Using correlated log-normal frailty models Pankratz et al. (2005) perform genetic analysis on age at onset in breast cancer in a large familial cohort. Their method is based on a Laplace approximation similar to Ripatti and Palmgren (2000).

4.7 Other correlated frailty models

A general correlated frailty model can be introduced with the help of frailties generated by non-negative Lévy processes. Aalen and Hjort (2002) considered such frailty distributions in the univariate case. Let $Z_1 = D(u_0) + D(u_1)$ and $Z_2 = D(u_0) + D(u_2)$ be frailties with respect to two correlated lifetimes.
The univariate unconditional survival function (for simplicity it is assumed to be equal for both partners) is given by $S(t) = \mathbf{E}S(t|Z_i) = \mathbf{E}e^{-(D(u_0)+D(u_i))\Lambda_0(t)} = e^{-(u_0+u_i)\psi(\Lambda_0(t))}$. Consequently, after some simple algebra we arrive at

$$\Lambda_0(t) = \psi^{-1}(\frac{-1}{u_0+u_i}\ln S(t)).$$

Hence,

$$\begin{aligned}
S(t_1,t_2) &= \mathbf{E}e^{-Z(u_0)(\Lambda_0(t_1)+\Lambda_0(t_2))-Z(u_1)\Lambda_0(t_1)-Z(u_2)\Lambda_0(t_2)} \\
&= e^{-u_0\psi(\Lambda_0(t_1)+\Lambda_0(t_2))}e^{-u_1\psi(\Lambda_0(t_1))}e^{-u_2\psi(\Lambda_0(t_2))} \\
&= e^{-u_0\psi(\Lambda_0(t_1)+\Lambda_0(t_2))}e^{-\frac{u_1}{u_0+u_1}(u_0+u_1)\psi(\Lambda_0(t_1))}e^{-\frac{u_2}{u_0+u_2}(u_0+u_2)\psi(\Lambda_0(t_2))} \\
&= e^{-u_0\psi(\psi^{-1}(\frac{-1}{u_0+u_1}\ln S(t_1))+\psi^{-1}(\frac{-1}{u_0+u_2}\ln S(t_2)))}S(t_1)^{\frac{u_1}{u_0+u_1}}S(t_2)^{\frac{u_2}{u_0+u_2}}
\end{aligned}$$

Assuming the symmetric case ($u_1 = u_2$) and introducing the parameter $\rho = \frac{u_0}{u_0+u_i}$ as the

correlation between Z_1 and Z_2, we get the final form in the general correlated frailty model:

$$S(t_1,t_2) = S(t_1)^{1-\rho}S(t_2)^{1-\rho}e^{-u_0\psi(\psi^{-1}(\frac{-1}{u_0+u_1}\ln S(t_1))+\psi^{-1}(\frac{-1}{u_0+u_2}\ln S(t_2)))}$$

Special cases are the correlated gamma frailty model and the correlated compound Poisson frailty model. Another special case is the shared positive stable frailty model, considered by Hougaard (1986a,b, 1995), Lam and Kuk (1997) and Qiou et al. (1999).
Another approach is suggested by Henderson and Shimakura (2003). They used their correlated frailty model with gamma distributed frailty to model longitudinal count data. Applying their concept of gamma distributed frailty to bivariate time-to-event data, the bivariate survival function is

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

which is another correlated gamma frailty model, not based on the additive decomposition of the gamma distributed frailties as is the model in section 4.4. But it is also an extension of the shared gamma frailty model ($\rho = 1$).

4.8 Correlated gamma frailty model with covariates

One reason for the popularity of frailty models is that (observed) covariates can easily be included into the model. We demonstrate the use of the correlated gamma frailty model with covariates by applying it to cause-specific mortality data on Danish monozygotic and dizygotic twins. We analyze the influence of smoking and body mass index (BMI) on heritability estimates of susceptibility to coronary heart disease (CHD).
Twin studies are one of the most widely used methods for quantifying the influence of genetic and environmental factors on specific diseases. In the case of binary traits (where the disease is either present or not), concordance analysis provides a powerful and widely accepted method in genetic epidemiology. Concordance rates are easily calculated and allow for a clear interpretation (McGue 1993; Gatz et al. 2000). In practical applications, time-to-event data (time of onset of disease, age at death) is often available, but usually in a censored form. Censoring of bivariate observations can be a complex problem, as either or both individuals of a pair may be subject to censoring, and the censoring times need not be the same for both individuals. Furthermore, covariates are available in many cases. Unfortunately, it is difficult to manage censored time-to-event data and covariates within the context of concordance analysis. A large part of the motivation for the methodology of this section is exploring the potential for censored data and

the inclusion of measured covariates. One key question here is whether the inclusion of observed covariates changes the heritability estimates of susceptibility to the disease under study.

We apply the bivariate correlated gamma frailty model to twin data on mortality due to coronary heart disease (CHD). The role of family aggregation of CHD is well established. Questions about the nature of the genetic effects (additive versus non-additive) can be addressed. Former studies from the Danish (Harvald and Hauge 1970) and Swedish (de Faire 1975; Marenberg et al. 1994) twin registries found a genetic component in the risk of death from coronary heart disease. Recent studies (Wienke et al. 2001, 2006b; Zdravkovic et al. 2002, 2004) used approaches from survival analysis to account for censoring present in the data. Such kinds of methods have to be combined with methods from genetic epidemiology. Heterogeneity of individuals with susceptibility to CHD as well as important covariates are included in the present model. For such a combined analysis, we apply the correlated gamma frailty model with observed covariates (Yashin et al. 1996; Yashin and Iachine 1997), which takes into account the dependence of life spans of relatives (twins). This allows the estimation of the effect of genetic factors in susceptibility to CHD and the evaluation as to what extent smoking, BMI, and susceptibility to CHD are all influenced by common genetic factors. This approach allows the combination of data on age at death with data on cause of death, smoking, and BMI; and further, dealing with censored observations. For each individual we assume two independent underlying competing risks of latent times (lifetime with respect to death due to CHD and lifetime with respect to death due to all other diseases (including censoring)). In addition, we assume that these competing risks are independent. An extension of the model to more than two competing risks is possible, but it is beyond the scope of the present thesis. We empirically demonstrate the advantages of the model in the statistical analysis of lifetime data from Danish twins, which were already used in Herskind et al. (1996a), but now with a special focus on mortality caused by CHD and applying a frailty model. The model allows us to check hypotheses about genetic confounding in the relationship between the covariates (smoking and BMI) and susceptibility to CHD. The concept of genetic confounding was originally introduced by R.A. Fisher, who suggested that the association between smoking and lung cancer was caused by common genetic factors (Fisher, 1958). For further references related to the discussion in this field, see for example the paper by Herskind et al. (1996a).

4.8.1 Material and methods

Mortality data of twins were provided by the Danish Twin Registry, founded in 1954 as the world's first nation-wide twin registry. This population-based registry includes twins born in Denmark during the period 1870 - 1910 and all like-sex pairs born between 1911 and 1930. For more detailed information about the data from the Danish Twin Registry, see Appendix A.2.

In 1966, a questionnaire including questions about smoking, height, and weight was mailed to all twins born 1890-1920 who were alive and traceable on 1 January 1966. 3709 individuals answered the questionnaire (response rate 65 %). Excluded from the study were 813 twins with non-responding partners, four pairs with unknown zygosity and 212 pairs with incomplete or uncertain information on height and weight. 23 pairs were excluded because of incomplete information about cause of death, resulting in a study population of 1209 twin pairs.

Individuals were followed from 1 January 1966 to 31 December 1993. Those persons identified as deceased after the end of the follow up period are classified for our purposes as 'living'. At the end of follow-up, approximately 40 % of the twins were still alive, resulting in right censored data. Altogether, there are 210 male monozygotic twin pairs and 316 dizygotic twin pairs, 273 female monozygotic twin pairs and 410 dizygotic twin pairs. In addition to age at death, there is also information on cause of death available for all individuals who died during the follow-up. For the present study, only the underlying cause of death was considered. Detailed information about death status, gender, zygosity, smoking, and BMI of the study population is given in Table 4.4 and 4.5.

	Non-smokers	Other (ex-smokers etc.)	pipe/cigar smokers	cigarette smokers	Total
males					
BMI < 22	11	12	45	31	99 (9.4%)
BMI 22- 28	78	141	333	208	760 (72.2%)
BMI > 28	24	35	91	43	193 (18.3%)
Total	113 (10.7%)	188 (17.9%)	469 (44.6%)	282 (26.8%)	
females					
BMI < 22	105	46	47	99	297 (21.7%)
BMI 22- 28	350	138	134	171	793 (58.1%)
BMI > 28	157	40	37	42	276 (20.2%)
Total	612 (44.8%)	224 (16.4%)	218 (16.0%)	312 (22.8%)	

Table 4.4: Study population (number of individuals) by sex, BMI and smoking

	males		females	
	MZ twins	DZ twins	MZ twins	DZ twins
deaths				
–CHD	96	153	76	110
–other	206	280	204	312
all causes	302	433	280	422
alive	118	199	266	398
total	420	632	546	820

Table 4.5: Study population (number of individuals) by gender, zygosity and cause of death

4.8.2 Mortality

After the age of six, death rates for Danish twins born between 1870 and 1900 are almost the same as those for the same cohorts of the Danish population. The distributions of age at death for monozygotic twins are close to those of dizygotic twins for both sexes (Christensen et al., 1995). Recent papers dealing with twin cohorts born during the period of 1870 - 1930 found similar mortality patterns for Danish twins and the general Danish population with respect to CHD (Wienke et al., 2001; Christensen et al., 2001). This similarity suggests that it is possible to generalize genetic results from survival analysis of twins to the total population with respect to mortality due to CHD.

For the present analysis, CHD is grouped as ICD 420 in the sixth and seventh revision and as ICD 410 - 414 in the eighth ICD revision.

4.8.3 Statistical methods

Now we make more specific assumptions about the structure of the lifetimes. To include heterogeneity in our model, we assume a correlated gamma frailty model. Let Z_j $(j = 1, 2)$ be the frailties, and X_j $(j = 1, 2)$ vectors of observable covariates of the two individuals of a twin pair. Assume that their individual hazards are represented by the proportional hazards model $\lambda(t) = Z_j \lambda_0(t) e^{\beta^T X_j}$ $(j = 1, 2)$ with a baseline hazard function $\lambda_0(t)$ describing the risk of dying as a function of age and β denotes the vector of regression parameters. Let the lifetimes of the two twin partners be conditionally independent given their frailties Z_1 and Z_2. Because frailties Z_j $(j = 1, 2)$ are usually unobservable, their correlation coefficient used in the methods of quantitative genetics cannot be estimated directly from the empirical data. So a bivariate lifetime model which allows indirect calculation of the parameters is needed. The unconditional bivariate survival function of the correlated gamma frailty model with observed covariates is

given by:

$$S(t_1,t_2|X_1,X_2) = S(t_1|X_1)^{1-\rho}S(t_2|X_2)^{1-\rho}(S(t_1|X_1)^{-\sigma^2} + S(t_2|X_2)^{-\sigma^2} - 1)^{-\frac{\rho}{\sigma^2}}, \quad (4.29)$$

where $S(t|X)$ denotes the marginal univariate survival function, assumed to be equal for both partners in a twin pair. Using a parametric approach we fit a Gamma-Gompertz model to the data, e.g. $S(t|X) = \left(1 + [(1 + s^2\frac{a}{b}(e^{bt} - 1))^{\frac{\sigma^2}{s^2}} - 1]e^{\beta X}\right)^{-\frac{1}{s^2}}$, where $a, b, s^2, \beta, \sigma^2, \rho$, are parameters to be estimated.

The lifetimes are assumed to be independently censored from the right by i.i.d. pairs of non-negative random variables, which are independent of the lifetimes. Thus, we observe

$$(T_{i1}, T_{i2}, \Delta_{i1}, \Delta_{i2}, X_{i1}, X_{i2}) \quad (4.30)$$

with $\Delta_{ij}\,(i = 1,\ldots,n;\, j = 1,2)$ as a binary variable with values 1 (event) and 0 (no event). Let the lifetimes follow a distribution (dependent on covariates X_1, X_2) given by the bivariate survival function $S(t_1,t_2|X_1,X_2) = \mathbf{P}(T_{i1} > t_1, T_{i2} > t_2|X_1,X_2)$. Starting from this model, we are able to derive the likelihood function of the data given by (4.30) (see Appendix E):

$$\begin{aligned} L(t_1,t_2,\delta_1,\delta_2,X_1,X_2) = {} & \delta_1\delta_2 S_{t_1t_2}(t_1,t_2|X_1,X_2) - \delta_1(1-\delta_2)S_{t_1}(t_1,t_2|X_1,X_2) \quad (4.31)\\ & - (1-\delta_1)\delta_2 S_{t_2}(t_1,t_2|X_1,X_2) + (1-\delta_1)(1-\delta_2)S(t_1,t_2|X_1,X_2). \end{aligned}$$

Partial derivatives of the marginal survival functions are given by $S_{t_j}(t_1,t_2) = \frac{\partial S(t_1,t_2)}{\partial t_j}$ $(j = 1,2)$ and $S_{t_1t_2}(t_1,t_2) = \frac{\partial S(t_1,t_2)}{\partial t_1 \partial t_2}$. Because of the independence assumption between lifetimes and censoring times, the distribution of the censoring times does not enter the likelihood function.

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 (see Appendix B).

4.8.4 Results

Because not all models are nested, the likelihood ratio test can not be applied to compare all genetic models. Applying the correlated gamma frailty model with and without observed covariates the Akaike Information Criterion (AIC) prefers the AE and DE model. Using these two models, heritability changes from 0.45 (0.11) without covariates, to 0.55 (0.13) with covariates (see Table 4.6). Standard errors for the ACE model are not shown since $c^2 = 0$ is the boundary of the parameter space.

Using the best fitting DE model, the likelihood ratio test indicates a significant influence of BMI on CHD mortality ($\beta_1 = 0.53$ (0.25) and $\beta_2 = 0.47$ (0.24)). Here β_1 and β_2 describe individuals with BMI less than $22\,kg/m^2$ and more than $28\,kg/m^2$, respectively. The reference group are individuals with BMI between $22kg/m^2$ and $28\,kg/m^2$. Cigarette smoking shows a significant influence on CHD mortality ($\beta_3 = 0.57$ (0.24), $\beta_4 = 0.32$ (0.23) and $\beta_5 = 0.48$ (0.26)). Here β_3, β_4 and β_5 denote cigarette smokers, pipe/cigar smokers and former smokers, respectively. The reference group is the non-smokers. $\beta_6 = 1.44$ (0.28) denotes the log hazard of males compared to females. The results with respect to the (second best) AE model are similar.

σ^2	a^2	d^2	c^2	e^2	BMI < 22	BMI > 28	cigarette smokers	other smokers	former smokers	males	AIC
ACE	**model**										
7.74	0.45		0.00	0.55							
(-)	(-)		(-)	(-)							4566.1
4.54	0.48		0.00	0.52	0.54	0.47	0.56	0.32	0.49	1.47	
(-)	(-)		(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	4494.6
AE	**model**										
7.74	0.45			0.55							
(4.02)	(0.11)			(0.11)							4564.1
4.54	0.48			0.52	0.54	0.47	0.57	0.32	0.49	1.47	
(1.71)	(0.12)			(0.12)	(0.26)	(0.25)	(0.24)	(0.24)	(0.26)	(0.29)	4492.6
ADE	**model**										
7.73	0.44	0.01		0.55							
(4.05)	(0.33)	(0.37)		(0.13)							4566.1
4.38	0.17	0.36		0.47	0.53	0.47	0.57	0.32	0.49	1.45	
(1.66)	(0.41)	(0.47)		(0.14)	(0.26)	(0.25)	(0.24)	(0.23)	(0.26)	(0.28)	4494.0
DE	**model**										
7.16		0.50		0.50							
(4.07)		(0.13)		(0.13)							4565.9
4.28		0.55		0.45	0.53	0.47	0.57	0.32	0.48	1.44	
(1.61)		(0.13)		(0.13)	(0.25)	(0.24)	(0.24)	(0.23)	(0.26)	(0.28)	4492.2
CE	**model**										
8.77			0.30	0.70							
(4.36)			(0.08)	(0.08)							4567.4
4.83			0.31	0.69	0.55	0.45	0.56	0.32	0.50	1.50	
(1.80)			(0.08)	(0.08)	(0.27)	(0.25)	(0.25)	(0.24)	(0.27)	(0.30)	4497.2

Table 4.6: Estimates (s.e.) of variance components in frailty to mortality from CHD (n=1209). σ^2 - variance of frailty, a^2 - additive genetic effects, d^2 - genetic effects as a result of dominance, c^2 - common environment, e^2 - non-shared (individual) environment (including measurement errors), AIC - Akaike Information Criterion, other smokers - pipe/cigar smokers

4.8.5 Discussion

The method in this section with its suitability for censored data and the possibility to include observed covariates allows us to overcome the well-known drawbacks of the traditional concordance analysis in twin studies with time-to-event data. An important question arising in genetic analysis of models with observed covariates (here smoking and BMI) is whether genes that are responsible for variation in observed covariates also contribute to a variation in susceptibility to CHD. In that case traditional methods of regression analysis can lead to spurious effects of covariates. In extreme cases it is not covariates, but common genes which may be responsible for the variation in life span.

Both smoking and BMI are influenced by genes with heritability estimates $0.35 - 0.75$ (smoking) and $0.5 - 0.8$ (BMI) (Bouchard, 1994; Heath and Madden, 1995; Herskind et al., 1996b). However, whether common genes influence these phenotypic traits, as well as susceptibility to CHD, is an open question. In 1958 R.A. Fisher suggested that the association between smoking and lung cancer is spurious and reflects only the circumstance that the same genes influence both smoking habits and lung cancer. This was the starting point for a long debate on genetic confounding.

The main result of the present analysis was that the inclusion of smoking and BMI do not cause any substantial decrease in the heritability estimates. Hence, no evidence was found for common genetic factors acting on smoking and susceptibility to CHD or BMI and susceptibility to CHD. This study confirms the earlier finding that the genetic influence on susceptibility to CHD is not mediated through genetic influence on smoking and BMI. Similar results were found by Zdravkovic et al. (2004) in Swedish twins.

As expected the inclusion of observable covariates decreases the heterogeneity in the population, which can be seen in the decline of the variance in frailty from $\sigma^2 = 7.74$ to $\sigma^2 = 4.54$ (AE model) and from $\sigma^2 = 7.16$ to $\sigma^2 = 4.28$ in the DE model. This makes clear that frailty is not a phenotypic trait. Frailty depends on the model, and it describes factors not included in the model.

When observed covariates are included in the model, the relative importance of environmental factors (shared environment, C, and non-shared environment, E) is reduced, leading to an increase in the heritability estimates observed in the present analysis. This looks a bit strange, although use of the heritability coefficient when computed as a proportion of variance may be misleading. The variance in the trait (frailty) can be decomposed as follows: $\sigma^2 = \sigma^2_{genes} + \sigma^2_{environment}$. One does not know whether the heritability increases due to an increase in the genetic variance or due to a decrease in the environmental variance. In the present case, the (moderate) increase was largely due to the influences of the latter. By including smoking and BMI, the genetic variance was

reduced from $\sigma^2_{genes} = 3.48$ to $\sigma^2_{genes} = 2.18$ in the AE model and from $\sigma^2_{genes} = 3.58$ to $\sigma^2_{genes} = 2.35$ in the DE model, respectively. The reduction of the environmental heterogeneity in frailty is more pronounced, e.g. from $\sigma^2_{environment} = 4.26$ to $\sigma^2_{environment} = 2.36$ in the AE model and from $\sigma^2_{environment} = 3.58$ to $\sigma^2_{environment} = 1.93$ in the DE model. Thus, the focus should not be on the increase in heritability but rather on the decrease in the environmental variance. These results help us understand that the risk factors studied primarily represent environmental sources of variation for CHD-death, despite the role that genetic influences may play for the specific risk factors. For a more detailed discussion of variance components see Hopper (1993).

This point is also important when discussing why the AE model could be replaced by the DE model when observed covariates are taken into account. For example, this could occur when the additive genetic component of observed covariates contains a substantial portion of the additive genetic component of a trait (as well as a part of independent environmental component). The resulting genetic and environmental variances of the trait may be reduced substantially as in this case. The sizes of these parts will determine the best fitting model and the new value of heritability estimate. It may well be then that the residual proportion of genetic variation in the trait (frailty) in the presence of observed covariates is mostly due to dominant genetic effects.

The analysis underlines the importance of genetic factors on individual susceptibility to CHD. Otherwise, the heritability estimate in our study (0.45 in the AE model without covariates) is lower than those found in a previous analysis of an extension of the presented data (without covariates) with heritability estimates of 0.53 and 0.58 for males and females, respectively (Wienke et al., 2001). The lower estimates in the subsample analyzed here may be a consequence of a decline in the heritability of CHD with increasing age, as found in Marenberg et al. (1994) and Zdravkovic et al. (2002). The twin population in this study is much older compared with the one in Wienke et al. (2001).

The main limitation of the present study is that the information on risk factors was obtained from the self-report questionnaire in 1966. It can not be ruled out that self-reported information, especially regarding weight, is biased downwards. Smoking information appears reliable, as smoking was widely accepted in the 1960s. Alternatively, people reported as smokers in the 1960s may have since stopped smoking, but may still be in the smoking category, which could result in a downward risk bias for smokers. Detailed information about tobacco smoked per day over the follow-up time is not available. Consequently, the group of smokers consists of both heavy and light smokers, which may result in an additional downward risk bias for smokers. More details in a slightly different analysis can be found in Wienke et al. (2005a).

4.9 MCMC methods for the log-normal frailty model

Two important assumptions in frailty models are related to the shape of the underlying hazard and the distribution of the frailty variables. Frailty models have been estimated both parametrically and semi-parametrically. In parametric models, the baseline hazard is assumed to follow a known distribution family up to a few finite-dimensional unknown parameters. The parametric hypothesis that is most commonly used is the Gompertz baseline hazard (Vaupel et al. 1992; Iachine et al. 1998; Wienke et al., 2001, 2002, 2005b) but other shapes are also possible, for example the Weibull (Sahu et al. 1997) or exponential shape (Xue and Ding 1999).

Every distribution of a positive random variable can be adapted to model frailty. The gamma distribution has been widely applied in the literature (Sections 3.2 and 4.4). The gamma choice is convenient from a mathematical point of view, because of the simplicity of the Laplace transform, which allows for the use of traditional maximum likelihood procedures in parameter estimation. Another possibility is to assume that frailty is log-normally distributed (Sections 3.7 and 4.6). The log-normal approach is much more flexible than the gamma model in creating correlated but different frailties as required in the correlated frailty model. Unfortunately, with a log-normal assumption it is impossible to derive the marginal likelihood function in an explicit form and parameter estimation has to be performed with the help of more sophisticated estimation strategies.

Methods for estimation in frailty models are put into the two categories: maximum likelihood and Markov Chain Monte Carlo (MCMC) methods. Yashin et al. (1995), Wienke et al. (2003a,b) applied procedures based on maximum likelihood methods in the gamma context, where an explicit representation of the likelihood function is always available. The maximum likelihood method has also been adopted in the log-normal framework with the help of different numerical algorithms (McGilchrist and Aisbett 1991; McGilchrist 1993; Lillard 1993; Lillard et al. 1995; Sastry 1997; Ripatti and Palmgren 2000; Ripatti et al. 2002). These methods are implemented in the aML software package (Lillard and Panis 2000).

In the present section (based on work together with Isabella Locatelli, see Locatelli 2003, Locatelli et al., 2004), we consider an example of how to apply MCMC methods to the correlated log normal frailty model as described in section 4.6 with $\mathbf{E}Z = \mu = 1$. We assume that the two frailties in each pair have the same variance σ^2 because of the symmetry of twin data, which is the object of applications here.

Bayesian MCMC methods have been applied as estimation procedures especially in

shared frailty models (Clayton, 1991; Spiegelhalter et al. 1996; Sahu et al. 1997; Sinha and Dey 1997) but also in correlated frailty models (Xue and Ding 1999). The Bayesian framework is in fact natural when we are dealing with conditionally independent observations and when we are working with hierarchical models, with the frailty variables at an intermediate stage between the observations and the so-called hyperparameters. In the Bayesian context, the frailty distribution represents a 'prior' of the model, and its parameters (hyperparameters) are also considered as random variables following some non-informative distribution. An MCMC method consists in generating a set of Markov chains whose joint stationary distribution corresponds to the joint posterior of the model, this one being in the Bayesian framework the distribution of random parameters given observed data. In a hierarchical model, the posterior distribution is often very difficult to work with and almost always impossible to integrate out in order to find the marginal posterior of each random parameter. The MCMC methods enable us to circumvent this problem. The posterior of each parameter is approximated by the empirical distribution of the values of the corresponding Markov chain and empirical summary statistics calculated along each chain can be used to make inferences about the true value of the corresponding parameter (see Gilks et al. 1996 for a review). The Gibbs sampling (Geman and Geman 1984) is one of the algorithms that have been created in order to obtain Markov chains with the desired stationary distribution. The basic idea behind the Gibbs sampling is to successively sample from the conditional distribution of each random node, given all the others in the model. These distributions are known as 'full conditional distributions'. It can be shown that, under broad conditions, this process eventually provides samples from the joint posterior distribution of the unknown quantities.

In this study, Bayesian MCMC methods have been adopted to estimate the correlated log-normal frailty model. Calculations are performed within the software WinBUGS 1.4 (Spiegelhalter et al. 1999). The correlated log-normal frailty model applied here can be represented as a Bayesian hierarchical (3 - levels) model in the following way:

1. Likelihood function:

$$L(t,\delta|W,a,b)=\prod_{i=1}^{n}\prod_{j=1}^{2}(\exp(W_{ij})a\exp(bt_{ij}))^{\delta_{ij}}\exp(-\exp(W_{ij})\frac{a}{b}(\exp(bt_{ij})-1))$$

2. Priors:

$$(i)\quad \begin{pmatrix} W_{i1} \\ W_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} -\frac{1}{2}\ln(\sigma^2+1) \\ -\frac{1}{2}\ln(\sigma^2+1) \end{pmatrix}, \begin{pmatrix} \ln(\sigma^2+1) & \ln(\rho\sigma^2+1) \\ \ln(\rho\sigma^2+1) & \ln(\sigma^2+1) \end{pmatrix}\right)$$

$$(ii)\quad a \sim \Gamma(0.01,0.01)$$

$$(iii)\quad b \sim \Gamma(0.01,0.01)$$

3. Hyperpriors:

$$\begin{aligned} (i) \qquad \sigma^2 &\sim \Gamma(0.01, 0.01) \\ (ii) \qquad \rho &\sim U(-1, 1) \end{aligned}$$

where $W = (W_1, ..., W_n)$, $W_i = (W_{i1}, W_{i2})$; $t = (t_1, ..., t_n)$, $t_i = (t_{i1}, t_{i2})$; Γ and U denote the gamma and uniform distribution, respectively, and a and b are parameters of the Gompertz baseline hazard. The prior (i), assigned to the vector (W_{i1}, W_{i2}), is chosen in order to have, according to the traditional definition of frailty, a vector of log-normal distributed frailties $(Z_{i1}, Z_{i2}) = \exp(W_{i1}, W_{i2})$ whose mean is equal to one. Finally, non-informative priors are assigned to the parameters of the Gompertz curve and to the frailty parameters (hyperparameters). The full conditional distributions can be obtained because they are proportional to the joint distribution of all the random quantities of the model. In our case, this joint distribution takes the form:

$$\pi(t, \delta, W, a, b, \sigma^2, \rho) = L(t, \delta | W, a, b) \prod_{i=1}^{n} [\prod_{j=1}^{2} \pi(W_{ij} | \sigma^2, \rho)] \pi(a) \pi(b) \pi(\sigma^2) \pi(\rho) \qquad (4.32)$$

where $\pi(\cdot)$ indicates the density function of the corresponding argument.
Often, the full conditional distributions have a complicated form, which makes it impossible to sample from them directly. That is why a slice-sampler algorithm (in version 1.4 of the software WinBUGS) is used for non log-concave densities defined on a restricted range (Neal 1997). This has an adaptive phase of 500 iterations, which are discarded from all summary statistics. A Metropolis within Gibbs algorithm based on a symmetric normal proposal distribution is applied in the case of non log-concave densities defined on an unrestricted range (Metropolis et al. 1953; Hastings 1970; Besag and Green 1993). In this case, the adaptive phase is of 5,000 iterations. The Metropolis within Gibbs procedure is applied in the log-normal case. The results of applying the

	a	b	σ^2	ρ_{MZ}	ρ_{DZ}
Mean	2.54e-5	0.072	45.19	0.311	0.104
Median	2.52e-5	0.072	41.50	0.299	0.097
sdv	3.24e-6	0.003	17.05	0.046	0.108
MC error	7.92e-8	8.94e-5	0.824	0.005	0.002
CSRF	1.002	1.006	1.055	1.008	1.005

Table 4.7: Correlated log-normal frailty model. Convergence achieved after 50,000 iterations.

correlated log-normal frailty model to the Swedish breast cancer data are presented in Table 4.7. The study population consists of 12,568 female twin pairs, e.g. the old cohort (born between 1886 and 1925 with both partners alive in 1959) and the middle cohort

(born between 1926 and 1967 with both partners alive in 1970). 1096 cases of breast cancer were identified through 27 October 2000. Estimated values include the Gompertz parameters a and b, the variance of the frailty distribution σ^2, which can be seen as the extent of population heterogeneity with respect to breast cancer, and estimates of the correlation coefficient for both monozygotic twins (ρ_{MZ}) and dizygotic twins (ρ_{DZ}). Two estimates for each parameter are given in terms of the mean and the median of the correspondent Markov chain. In all cases, the two values are very close to one another. This means that empirical estimates of the marginal posterior densities (kernel density estimates) are approximately symmetric. For each parameter, the sample standard deviation and an estimate of the standard error of the mean are also given. This one is obtained following Roberts' (1996) batch means method. In the last row, we reported the value of the *Corrected Scale Reduction Factor* (CSRF) for each parameter. This value corresponds to the Gelman-Rubin convergence statistic (Gelman and Rubin 1992), as modified by Brooks and Gelman (1998), and is based on a comparison of the inter and intra chain variance for each variable. When values of this diagnostic are approximately equal to one, the sample must have arisen from the stationary distribution. In this case, descriptive statistics are valid estimates of unknown parameters.

According to the above model, the population would be largely heterogenous (σ^2) in terms of susceptibility towards breast cancer. The estimated correlation between frailties is larger for monozygotic than for dizygotic twins. This means that monozygotic individuals who genetically are more similar than dizygotic twins also present a larger correlation in terms of frailty towards breast cancer. This finding suggests that there is a genetic influence on breast cancer propensity. We estimate the extent of such an influence with the help of three different genetic models (see Appendix B).

In Table 4.8, we compare an ACE, AE and ADE model. Estimates of each parameter are given in terms of the sample mean. Sample median values are omitted because they are very close to the mean as in Table 4.7. We give the posterior standard deviation of each parameter in parentheses. This quantity is a measure of the dispersion of the posterior density estimate, giving an idea of a parameter's significance.

The last column of Table 4.8 shows the values of the Deviance Information Criterion (DIC) for the three models. This is a statistic introduced by Spiegelhalter et al. (2002) in order to compare Bayesian models in terms of adequacy and complexity. DIC is defined as: $DIC = \overline{D(\theta)} + p_D$ where $\overline{D(\theta)}$ represents an estimate (in terms of the posterior mean) of the model's deviance and is suggested to be a Bayesian measure of fit or adequacy; p_D is the difference between the posterior mean of the deviance and the deviance of the posterior mean of the parameters of interest and is proposed as a measure

of the effective number of the model's parameters (complexity). The deviance $D(\theta)$ is defined as equal to $-2\ln p(y|\theta)$ where y comprises all stochastic nodes giving values (i.e. data), and θ comprises the stochastic nodes upon which the distribution of y depends. Spiegelhalter et al. (2002) have shown that DIC is related to other information criteria and that, particularly, in models with negligible prior information, DIC is approximately equivalent to Akaike's criterion. The model with the smallest DIC possibly best predicts a replicate data set of the same structure as that currently observed. In Table 4.8, the model which presents the lowest value of DIC is the ADE model.

Thus, from the comparison between the three models we conclude that genetic effects explain globally almost 30% of the variability of propensity to breast cancer. Environmental effects would be predominant in breast cancer susceptibility and these would be primarily individual-specific, that is non-shared effects. Finally, a model including dominance genetic effects should preferably be used for genetic and statistical reasons.

	σ^2	a^2	c^2	d^2	e^2	DIC
ACE	45.21	0.176	0.053		0.771	15138.6
	(17.7)	(0.094)	(0.046)		(0.089)	
AE	47.31	0.230			0.770	15102.3
	(18.3)	(0.091)			(0.091)	
ADE	48.30	0.127		0.149	0.724	15091.8
	(16.7)	(0.086)		(0.100)	(0.084)	

Table 4.8: Three genetic models. Convergence achieved after 50,000 iterations

The WinBUGS package proved to be extremely useful and flexible enough to estimate correlated frailty models and to add to them equations typical of genetic models. With the same software it is easy to modify the hypothesis on the frailty distribution. Different assumptions about the frailty distribution and the shape of the baseline hazard function can be compared using the same software with a Bayesian information criterion (DIC).

A disadvantage of using WinBUGS is the time required for estimation. In fact, we are working with models that include a very large number of parameters, especially when we deal with large data sets. This means that every MCMC algorithm which updates parameters one by one (like the Gibbs Sampling used in WinBUGS) is very time consuming. To overcome this problem, an algorithm that enables the updating of parameters all together (or groups of parameters) at the same time should be adopted.

More details about the MCMC approach can be found in Locatelli (2003) and Locatelli et al. (2004).

4.10 Comparison of different estimation strategies

In this section, we examine bivariate correlated frailty models, and especially the behavior of the parameter estimates when using different estimation strategies. We consider three different correlated frailty models: the gamma model and two versions of the log-normal model. The traditional maximum likelihood procedure of parameter estimation in the gamma case with an explicitly available likelihood function is compared with maximum likelihood methods based on numerical integration and a Bayesian approach using MCMC methods by means of a comprehensive simulation study. We detected a strong dependence between the two parameter estimates (variance and correlation of frailties) in the bivariate correlated frailty model and analyzed it in detail.

The most common frailty distribution is the gamma distribution. It has been widely applied as a mixture distribution (sections 3.2 and 4.4). From a computational and analytical point of view, the gamma distribution fits well into the proportional hazards framework, because it leads to closed form expressions of survival, density and hazard function. This is due to the simplicity of the Laplace transform. Throughout this section we will refer to the correlated gamma frailty model as Model 1.

The second frailty model considered in detail in this section is the log-normal model (sections 3.7 and 4.6). Again, the frailty is acting multiplicatively on the baseline hazard following a log-normal distribution. Especially in multivariate modelling, the log-normal approach is much more flexible than the gamma model in creating correlated but different frailties. Two variants of the log-normal model are analyzed here. We assume a normally distributed random variable W to generate frailty as $Z = e^W$. The two variants of the model are given by the constraints $\mathbf{E}W = 0$ (Model 2) and $\mathbf{E}Z = 1$ (Model 3). Unfortunately, no explicit form of the unconditional likelihood exists. Consequently, estimation strategies based on numerical integration are required.

To see whether the dependence is related to the estimation strategy or the choice of frailty distribution we use the bivariate models mentioned above and apply three different estimation strategies. First, we perform a traditional maximum likelihood estimation procedure (only possible in the gamma model); second, we use a maximum likelihood approach based on numerical integration; and finally we apply MCMC methods.

4.10.1 Estimation strategies

Parameter estimation in the gamma model is straightforward. The frailty term can be integrated out and an explicit representation of the unconditional bivariate survival function exists (4.29), which can be used to derive the likelihood function.

Unfortunately, the integrals in (4.11) have no explicit solution in the log-normal model. Several estimation methods for bivariate log-normal frailty models in consequence have been suggested to be used within a non-Bayesian framework. Various modifications of the maximum likelihood procedure are applicable to the bivariate frailty models. Ripatti and Palmgren (2000) derived an estimating algorithm based on the penalized partial likelihood (PPL). Xue and Brookmeyer (1996) suggested a modified EM algorithm for the log-normal frailty models. Sastry (1997) developed the modified EM algorithm for the multiplicative two-level gamma frailty model. Ripatti et al. (2002) present yet another method to deal with EM-like algorithms in a log-normal frailty model.

In the present section we use numerical integration procedures. Integrals over univariate and multivariate normal distributions can be approximated in different ways. One possibility is to use Gauss-Hermite quadratures (Naylor and Smith 1982; Smith et al. 1987). Similar ideas are employed in various applications of random effect models in survival analysis (Lillard 1993; Lillard et al. 1995; Panis and Lillard 1995; among others). The methods are implemented in the aML software package (Lillard and Panis 2000).

Several studies on the application of Bayesian methods to multivariate frailty models exist. Bolstad and Manda (2001) considered the gamma frailty model. Gibbs' sampling scheme for the bivariate log-normal frailty model with an exponential baseline hazard is given in Xue and Ding (1999). Korsgaard et al. (1998) present a Bayesian inference in the log-normal frailty model with a semi-parametric hazard.

In the Bayesian framework the correlated frailty model takes the form of a *hierarchical* model, with the frailty variables at an intermediate stage between observations and so-called hyperparameters. The conditional likelihood, that is the distribution of observed data given all the random parameters, represents the first level of the model:

$$L(t,\delta|Z,a,b,\beta)=\prod_{i=1}^{n}\Big(Z_{i1}\lambda_0(t_{i1})e^{\beta X_{i1}}\Big)^{\delta_{i1}}e^{-Z_{i1}\Lambda_0(t_{i1})e^{\beta X_{i1}}}\Big(Z_{i2}\lambda_0(t_{i2})e^{\beta X_{i2}}\Big)^{\delta_{i2}}e^{-Z_{i2}\Lambda_0(t_{i2})e^{\beta X_{i2}}},$$

with $t=(t_1,\ldots,t_n)$, $t_i=(t_{i1},t_{i2})$, $\delta=(\delta_1,\ldots,\delta_n)$, $\delta_i=(\delta_{i1},\delta_{i2})$, $X=(X_1,\ldots,X_n)$, $X_i=(X_{i1},X_{i2})$, $Z=(Z_1,\ldots,Z_n)$, $Z_i=(Z_{i1},Z_{i2})$, and a, b representing parameters of the Gompertz baseline hazard $\lambda_0(t)$.

By definition, the vector of frailties (Z_{i1},Z_{i2}) is assumed to follow a bivariate log-normal distribution, with variances σ^2 and correlation coefficient ρ. Parameters of the baseline hazard, regression coefficients β, σ^2 and ρ (the latter two being the hyperparameters) are presumed to follow a non-informative distribution. We adopt uniform priors over the intervals [1e-7, 0.005], [0.05, 0.15] and [-1, 1] for a, b and ρ, respectively; log-normal priors with mean 0.5 and variance 0.25 for σ^2; multivariate normal priors for β. More details about the MCMC approach can be found in section 4.9.

4.10.2 Simulations

We estimated Model 1 following a maximization procedure and Models 2 and 3 using a numerical integration procedure (Gauss-Hermite quadrature). MCMC methods are employed in all three models. We generated data sets with different frailty distributions. First, we used $\sigma^2 = 1$ and $\rho = 0.7$. Second, we used $\sigma^2 = 0.3$ and $\rho = 0.2$. In both cases $a = 0.003$, $b = 0.07$, $\beta = (\beta_1, \beta_2)$, $\beta_1 = 0.1$, and $\beta_2 = -0.2$. The observed covariates were generated as

$$X_{ij1} = \begin{cases} 1 & \text{if } i \leq \frac{n}{2} \\ 0 & \text{if } i > \frac{n}{2} \end{cases} \tag{4.33}$$

and $X_{ij2} \sim N(0, 1)$. Consequently, the first covariate is pair-specific where the second covariate is individual-specific. We used sample sizes of 500 and 5,000 pairs and simulated 500 data sets in each case (50 data sets only in Bayes methods because of time constraints). Censoring was not allowed. Results are shown in Tables 4.9 - 4.11.

Method	Sample size	a	b	σ^2	ρ	β_1	β_2
	true values	3.00e-3	0.070	0.300	0.200	0.100	-0.200
ML	500	3.01e-3	0.070	0.294	0.222	0.105	-0.198
		(3.70e-4)	(0.004)	(0.087)	(0.194)	(0.082)	(0.042)
ML	5000	3.01e-3	0.070	0.299	0.197	0.099	-0.200
		(1.25e-4)	(0.001)	(0.027)	(0.069)	(0.026)	(0.013)
Bayes	5000	3.04e-3	0.070	0.292	0.208	0.093	-0.196
		(1.16e-4)	(0.001)	(0.024)	(0.059)	(0.027)	(0.013)
	true values	3.00e-3	0.070	1.000	0.700	0.100	-0.200
ML	500	2.99e-3	0.070	1.001	0.699	0.107	-0.202
		(4.03e-4)	(0.005)	(0.141)	(0.080)	(0.105)	(0.054)
ML	5000	3.01e-3	0.070	1.000	0.700	0.098	-0.199
		(1.34e-4)	(0.002)	(0.044)	(0.023)	(0.034)	(0.017)

Table 4.9: Model 1 with Gompertz baseline hazard and two covariates, simulated data, 500 data sets (Bayes 50 data sets), means of estimates, standard errors in parentheses

The three models show the same pattern. As expected, the estimations for the larger sample size are far more accurate. The most striking effect is the strong negative correlation between estimates of ρ and σ^2, independently of the model and the estimation procedure (see Table 4.12).

As Bayesian methods proved to be very time-consuming, we generated only 50 data sets with 5,000 pairs each. We run two parallel chains from different starting points and considered the first 4,000 iterations for each chain as a 'burn-in' interval.

Method	Sample size	a	b	σ^2	ρ	β_1	β_2
	true values	3.00e-3	0.070	0.300	0.200	0.100	-0.200
ML	500	3.02e-3	0.071	0.372	0.237	0.097	-0.203
(num. int.)		(8.23e-4)	(0.008)	(0.378)	(0.352)	(0.080)	(0.045)
ML	5000	2.99e-3	0.070	0.308	0.204	0.099	-0.200
(num. int.)		(2.04e-4)	(0.002)	(0.077)	(0.083)	(0.025)	(0.013)
Bayes	5000	3.04e-3	0.070	0.300	0.218	0.095	-0.199
		(2.01e-4)	(0.002)	(0.067)	(0.089)	(0.022)	(0.012)
	true values	3.00e-3	0.070	1.000	0.700	0.100	-0.200
ML	500	2.81e-3	0.075	1.283	0.689	0.113	-0.211
(num. int.)		(1.06e-3)	(0.014)	(0.731)	(0.186)	(0.118)	(0.059)
ML	5000	3.07e-3	0.069	0.977	0.720	0.098	-0.199
(num. int.)		(3.44e-4)	(0.004)	(0.173)	(0.072)	(0.034)	(0.017)
Bayes	5000	3.12e-3	0.069	0.981	0.726	0.091	-0.198
		(3.93e-4)	(0.004)	(0.193)	(0.074)	(0.031)	(0.015)

Table 4.10: Model 2 with Gompertz baseline hazard and two covariates, simulated data, 500 data sets (Bayes 50 data sets), means of estimates, standard errors in parentheses

The simulated values of parameters of random effects in our case have auto-correlations close to unity. Convergence is very slow. Altogether 10,000-60,000 iterations per chain were generated after a 'burn-in' interval for each data set. The values of the Gelman-Rubin statistics are quite close to one (see Section 4.9 for details).

4.10.3 Discussion

Because of their simplicity, multivariate frailty models have become very popular over the last decade. A wide range of papers have been published, dealing with different structures of multivariate models (shared vs. correlated frailty models), different distributions of frailty (gamma, log-normal, stable etc.), different assumptions about the baseline hazard (parametric vs. semi-parametric models), and different estimation strategies (traditional maximum likelihood procedures, maximum likelihood procedures based on numerical integration, EM algorithm, MCMC methods). After dealing with correlated frailty models for a long time, we recognized a striking correlation between the variance and the correlation estimates. The present study is the first to analyze this kind of correlation. Our aim has been to draw attention to the problem, to elaborate possible reasons for this effect and to present (very preliminary) suggestions on how to overcome the problem.

Method	Sample size	a	b	σ^2	ρ	β_1	β_2
	true values	3.00e-3	0.070	0.300	0.200	0.100	-0.200
ML	500	3.01e-3	0.071	0.361	0.243	0.095	-0.204
(num. int.)		(4.22e-4)	(0.007)	(0.297)	(0.358)	(0.078)	(0.043)
ML	5000	2.99e-3	0.070	0.308	0.204	0.099	-0.200
(num. int.)		(1.38e-4)	(0.002)	(0.075)	(0.082)	(0.025)	(0.013)
Bayes	5000	3.03e-3	0.070	0.302	0.218	0.095	-0.199
		(1.36e-4)	(0.002)	(0.070)	(0.092)	(0.022)	(0.012)
	true values	3.00e-3	0.070	1.000	0.700	0.100	-0.200
ML	500	3.00e-3	0.075	1.323	0.683	0.107	-0.212
(num. int.)		(4.35e-4)	(0.015)	(0.998)	(0.160)	(0.117)	(0.064)
ML	5000	3.00e-3	0.070	1.022	0.701	0.099	-0.201
(num. int.)		(1.46e-4)	(0.004)	(0.179)	(0.067)	(0.034)	(0.018)
Bayes	5000	3.02e-3	0.070	1.000	0.713	0.102	-0.199
		(1.30e-4)	(0.003)	(0.134)	(0.058)	(0.034)	(0.015)

Table 4.11: Model 3 with Gompertz baseline hazard and two covariates, simulated data, 500 data sets (Bayes 50 data sets), means of estimates, standard errors in parentheses

First we tested whether the correlation of the estimates is dependent on the distribution of the frailty. We used three very popular frailty distributions to answer this question: the gamma distribution and two log-normal distributions (Models 1 - 3).

Second we tested whether the observed effect was caused by the estimation strategy. This is why we used three different estimation strategies: traditional maximum likelihood estimation (using a self-written GAUSS code), maximum likelihood estimation based on numerical integration (using routines in aML and a self-written code in Matlab), and MCMC methods in WinBUGS.

The results of the simulations are very clear. The observed effect is stable over different frailty distributions and different estimation strategies. Moreover, different choices of parameters and sample sizes did not change the correlation.

A high correlation of parameter estimates could be a sign of identifiability problems in the model. Correlated frailty models were investigated in order to overcome the problems of the shared frailty models, which provide only one parameter to model variance and correlation. One idea was to include observed covariates into the models to improve identifiability characteristics. This is why all models were run both with and without observed covariates. The results in both cases are very similar; consequently, we dropped the results for models without observed covariates. Two covariates were used, one dichotomous and one continuous. No effect of the covariates was detected.

Model	Method	Sample size	$\mathbf{corr}(\rho, \sigma^2)$	Parameter
1	ML	500	-0.229**	$\sigma^2 = 0.3$, $\rho = 0.2$
1	ML	5000	-0.227**	
1	Bayes	5000	-0.331*	
1	ML	500	-0.431**	$\sigma^2 = 1$, $\rho = 0.7$
1	ML	5000	-0.396**	
2	ML (num. int.)	500	-0.241**	$\sigma^2 = 0.3$, $\rho = 0.2$
2	ML (num. int.)	5000	-0.414**	
2	Bayes	5000	-0.382**	
2	ML (num. int.)	500	-0.789**	$\sigma^2 = 1$, $\rho = 0.7$
2	ML (num. int.)	5000	-0.875**	
2	Bayes	5000	-0.921**	
3	ML (num. int.)	500	-0.242**	$\sigma^2 = 0.3$, $\rho = 0.2$
3	ML (num. int.)	5000	-0.409**	
3	Bayes	5000	-0.438**	
3	ML (num. int.)	500	-0.717**	$\sigma^2 = 1$, $\rho = 0.7$
3	ML (num. int.)	5000	-0.862**	
3	Bayes	5000	-0.854**	

Table 4.12: Models 1-3 with Gompertz baseline hazard and 2 covariates: correlation between parameter estimates, 500 simulated data sets (Bayes 50 data sets). * $p < 0.05$, ** $p < 0.01$.

The present study focuses on parametric models, which implies the parametric specification of the baseline hazard. In a separate simulation of the correlated gamma frailty model with an unspecified baseline hazard (semi-parametric model), we found the correlation between the estimates similar to that in the parametric models.

Regarding identifiability aspects, note that heterogeneity and correlation between frailties are not completely independent in a frailty model based on conditional independence. To see this, assume that the variance of the frailty tends to zero. This implies zero correlation. The assumption about the conditional independence of lifetimes given the frailty could be the reason for the correlation linking the estimates of variance (heterogeneity) and correlation between frailties. This would explain why the observed effect is stable over different models and estimation procedures.

The conclusion to draw is that researchers should be cautious, and be aware of the problem presented in applying these models. Nevertheless, this study shows that the models perform well and that there is nearly no bias in the estimates. No correlation between the estimates of the regression coefficients β_1 and β_2 was found. This supports the use of correlated frailty models for obtaining accurate estimates of covariate effects.

4.11 Bivariate frailty cure models

A bivariate frailty cure approach for modelling familiar association in diseases was established by Chatterjee and Shih (2001), which is an extension of the univariate frailty cure model in (3.16). For a pair of individuals we define

$$Y_j = \begin{cases} 1 & : \text{ if the j-th individual is susceptible} \\ 0 & : \text{ if otherwise} \end{cases} \tag{4.34}$$

and let T_j denote the age at onset for the j-th individual when $Y_j = 1$ (j=1,2). In that case, the likelihood function is:

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

where $\phi_{11} = \mathbf{P}(Y_1 = 1, Y_2 = 1)$, $\phi_{10} = \mathbf{P}(Y_1 = 1, Y_2 = 0)$, $\phi_{01} = \mathbf{P}(Y_1 = 0, Y_2 = 1)$ $\phi_{00} = \mathbf{P}(Y_1 = 0, Y_2 = 0)$. $S(t_1, t_2)$ denotes the bivariate survival function of pairs with both individuals susceptible to the event under study. This bivariate survival is often given in form of a copula. Chatterjee and Shih (2001) used three different copulas in their approach: the shared gamma frailty model (Claytons model), Frank's copula, and Hougaard's shared positive stable frailty model. The authors applied a two step estimation procedure to breast cancer using the kinship data from the Washington Ashkenazi Study by ignoring the dependency among different pairs within the same family.

Their model was extended by Wienke et al. (2003a), who substituted the shared gamma frailty model by the correlated gamma frailty model (see 4.16):

$$S(t_1, t_2) = S(t_1)^{1-\rho}S(t_2)^{1-\rho}(S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1)^{\frac{\rho}{\sigma^2}}.$$

Furthermore, all parameters were estimated in a one-step maximum likelihood procedure. A parametric model with Gompertz baseline hazard was used, e.g.

$$S(t) = S(0, t) = S(t, 0) = \big(1 + s^2\frac{a}{b}(e^{bt} - 1)\big)^{-\frac{1}{s^2}}.$$

The likelihood function of bivariate right censored lifetime data in this model is given by the expression

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

The model was applied to breast cancer incidence data of identical and fraternal female twins provided by the Swedish Twin Registry. This data set contains records of 5857 female twin pairs with both partners being alive in 1959-61 (old cohort). 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. For more details see Appendix A.1. The results are given in Table 4.13:

	gamma frailty without cure fraction	gamma frailty with cure fraction[1]	gamma frailty with cure fraction[2]	log-normal frailty with cure fraction[1]
a	2.50e-7 (1.60e-7)	2.22e-6 (2.00e-6)	1.98e-6 (2.16e-6)	6.27e-6 (2.85e-06)
b	0.161 (0.012)	0.148 (0.013)	0.148 (0.013)	0.147 (0.010)
σ	7.612 (0.470)	2.780 (0.931)	2.963 (1.391)	4.865 (3.791)
ρ_{MZ}	0.124 (0.040)	0.900 (0.552)	0.934 (0.609)	0.717 (0.196)
ρ_{DZ}	0.097 (0.029)	0.725 (0.465)	0.761 (0.527)	0.645 (0.223)
ϕ_{11}	1.000 (-)	0.040 (-)	0.043 (0.040)	0.010 (-)
ϕ_{10}	0.000 (-)	0.160 (-)	0.176 (0.100)	0.092 (-)
ϕ_{00}	0.000 (-)	0.639 (-)	0.605 (0.235)	0.806 (-)
ϕ	1.000 (-)	0.201 (0.081)	0.219[3] (-)	0.102 (0.007)
likelihood	-5218.6397	-5217.3336	-5217.3102	

Table 4.13: Results of breast cancer data with the correlated gamma frailty model without and with a cure 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 $\phi = \phi_{11} + \phi_{10}$

We consider three different cases of cure models. In the first case, it is assumed that the susceptible status of the individuals in a pair is independent, 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 $p_1, p_2 \in \{0,1\}$. 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, the restriction of independence between the susceptibility status of the two partners in a pair is relaxed and substituted by the weaker constraints $\phi_{10} = \phi_{01}$, $\phi_{11} + \phi_{10} + \phi_{01} + \phi_{00} = 1$. When comparing the likelihoods, it turns out that the cure model with an independent

susceptible status of the twin partners shows a non-significantly better fit than the model without a cure fraction ($\chi^2_1 = 2.61$, $p = 0.11$). The more complicated cure model without an independence assumption between the susceptible status of the twin partners shows no significant improvement compared to the cure model assuming independence ($\chi^2_1 = 0.05$, $p = 0.83$). Interestingly, the estimate of the size of a susceptible fraction (due to breast cancer) with $\phi = 0.201\,(0.081)$ is 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. In the third case shown in the fourth column of Table 4.13 it is again assumed that the susceptible status of the individuals in a pair is independent of each other. But the frailty distribution was chosen to be log-normal (see section 4.6) and a MCMC approach was used to estimate the model parameter (see section 4.9). Interestingly, parameter estimates are quite different depending on whether the frailty distribution among the susceptible individuals is chosen as gamma or log-normal. But even with log-normal distributed frailty, only a small fraction ($\phi = 0.102$) of all women is indicated to be susceptible to breast cancer. Nevertheless, the estimates of the susceptible fraction in all three models in Table 4.13 are perfectly in the range of the figures obtained by Farewell et al. (1977) for different combinations of four risk factors. 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.

A simulation study was performed to evaluate the properties of the estimates in the proposed gamma frailty model. All simulations involve generating gamma-distributed frailties, bivariate lifetimes, censoring times as well as the inclusion of a cured fraction in the study population. A total of 5000 twin pairs are simulated. Samples are generated as follows:

- Generate frailty variables using independent gamma-distributed random variables.
- Generate lifetimes given the frailties using $S(t|Z) = e^{-Z\frac{a}{b}(e^{bt}-1)}$.
- Define cured individuals by using a random variable.
- Censored 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 censoring pattern.

The simulated data are generated assuming dependence between the susceptibility status of the partners (second column in Table 4.13), but in the estimation procedure the more general model with independent susceptibility status was applied (see third column in Table 4.13). 1000 data sets were simulated.

The mean parameter estimates of the model are shown in Table 4.14, 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 very good.

Parameter	true value	Mean of estimates	standard deviation
a	0.100	0.116	0.067
b	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 4.14: Parameter estimation in the simulation study.

Multivariate cure models suffer from the same inherent identifiability problem with the right censored observations as univariate cure models (compare last paragraph of section 3.8). For such observations the event under study has not occurred, either because the person is insusceptible or the person is susceptible, but the follow-up did not last long enough to observe the event. The identifiability problem grows with increasing censoring, but is less of a problem with parametric modelling of the baseline hazard. The simulation study shows that the estimation procedures work well under the given censoring scheme in the sample data set. Stronger right censoring causes strong identifiability problems. For example, another simulation study (not shown here) that uses the same parameters as described in the simulation section but uses the birth years 1926-1958 (describing the middle cohort of the Swedish Twin Registry) results in a complete breakdown of the estimation procedure. In cure models with fixed censoring times (caused by ending the study) censoring is no longer non-informative, even if censoring times and the survival times are independent. The proportion of censored observations contains important information about parameters in the model. For example, in what would be an ideal case of no censoring it holds $\phi = 1$. More information about this study is given in Wienke et al. (2003a).

Another approach to include a cure fraction into a bivariate frailty model is the correlated compound Poisson model in section 4.5.

4.12 Dependent competing risks in frailty models

In section 4.4 we analyzed cause-specific mortality data using the correlated gamma-frailty model, assuming independence among causes of death in a 'competing risk' scenario (Wienke et al., 2000, 2001, 2003b). In this section, we investigate the effect of removing this limitation. The model allows us to test the hypothesis on dependence between the competing risks. The class of multivariate distributions presented is characterized by the association parameters, using arbitrary marginal distributions. The multivariate distribution is specified in full by the association and variance parameters and the marginal distribution functions.

We can empirically demonstrate the advantages of this new model, having revisited the statistical analysis of the lifespan data previously explored in Wienke et al. (2000, 2001, 2003b). In the present analysis, we focus on the mortality rates of coronary heart disease (CHD). To simplify description, in this thesis we consider models limited to two competing risks (death as a result of CHD and death arising from other causes). The model can be extended to the case of multiple competing risks or multivariate lifetimes. Results of a simulation study are included. Both limitations and future uses for this model are discussed.

4.12.1 The statistical model

Identifying correlations of durations is a requirement for successfully analyzing genetic factors. In survival analysis there is a recurring problem of censored data, which complicates the statistical analysis far more than does complete data. Using a survival model to estimate correlations among lifetimes can solve this problem. In this thesis, instead of treating life spans directly, we wish to analyze both genetic and environmental factors acting on frailty for cause-specific mortality. The correlated gamma frailty model can be used to fit the lifetime data and provide a specific parameter for the correlation among frailties. This model was used to describe total mortality in twins by Yashin and Iachine (1995a) and cause-specific mortality in twins under the assumption of independence between competing risks by Wienke et al. (2000, 2001). However, the assumption of independence between competing risks is questionable. Typically, in clinical and epidemiological studies two different types of censoring occur. The observations of certain individuals are censored due to the fact that they are still alive at the end of the study. Other individuals drop from the follow-up for reasons not associated with the disease under study, but through life-events beyond the control of the researcher, such as migration.

If censoring can be assumed to be non-informative with regard to all different causes, then the traditional model may be applied with the censoring times taken as the minimum of the hypothetical censoring times arising from the different causes of censoring. For estimating the marginal survival function S the Kaplan-Meier estimator is appropriate. However, the situation becomes much more difficult if the censoring arising from at least one of the different causes can be assumed to be informative.

Various approaches have been proposed to account for informative censoring. Link (1989) proposed a model for informative censoring in which censoring only occurs in a subpopulation defined by the frailty distribution. Emoto and Matthews (1990) assumed a bivariate Weibull model for failure and censoring times. Zheng and Klein (1995) used a copula to study dependent competing risks. Lin et al. (1996) used data collected after non-fatal failure events to model dependent censoring by death or selective patient withdrawal. Lee and Wolfe (1998) proposed a test for independent censoring. Their method involves further follow-up of a subpopulation of lost-to-follow-up censored subjects. Asymptotic results in a specific situation are obtained by Wienke (1998). All these methods consider uncorrelated subjects only. The method described in this section is to be applied to correlated subjects and was published by Wienke et al. (2002). Essentially it views correlated subjects as partial replicates for each other, and hence gains identifiability for informative censoring.

We consider a case where two types of censoring occur, one non-informative and the other informative. Let $(X_{i1}, Y_{i1}, C_{i1}, X_{i2}, Y_{i2}, C_{i2})$ $(i = 1, \ldots, n)$ be i.i.d. vectors of non-negative random variables. The variables (X_{i1}, X_{i2}) denote the (usually non-observable) lifetimes (with respect to the cause of death of interest) of pairs of individuals. The (Y_{i1}, Y_{i2}) are informative censoring times (which may be lifetimes with respect to causes of death not under study) and (C_{i1}, C_{i2}) are non-informative censoring times (for example caused by end of study). Again, for $i = 1, \ldots, n$ and $j = 1, 2$ we observe $T_{ij} = \min\{X_{ij}, Y_{ij}, C_{ij}\}$ and

$$\Delta_{ij} = \begin{cases} 1 & : \text{ if } X_{ij} \leq \min\{C_{ij}, Y_{ij}\} \\ 0 & : \text{ if } C_{ij} < \min\{X_{ij}, Y_{ij}\} \\ -1 & : \text{ if } Y_{ij} < \min\{X_{ij}, C_{ij}\} \end{cases} \tag{4.35}$$

where $\Delta_{ij} = 1$ means no censoring, $\Delta_{ij} = 0$ is non-informative censoring and $\Delta_{ij} = -1$ is informative censoring. Now we derive the four-dimensional survival function of the data. Suppose that we use (X_1, Y_1, X_2, Y_2) as a shorthand for $(X_{i1}, Y_{i1}, X_{i2}, Y_{i2})$ $(i = 1, \ldots, n)$. Let (X_1, Y_1, X_2, Y_2) and (Z_1, Z_2, Z_3, Z_4) be the survival times of life- and (informative) censoring times and the frailties of the two individuals with respect to two different

causes of death; let their individual hazards be given by the proportional hazards model

$$\begin{aligned} X_1 \sim \lambda_1(x_1, Z_1) &= Z_1\lambda_1(x_1) \qquad X_2 \sim \lambda_1(x_2, Z_3) &= Z_3\lambda_1(x_2) \\ Y_1 \sim \lambda_2(y_1, Z_2) &= Z_2\lambda_2(y_1) \qquad Y_2 \sim \lambda_2(y_2, Z_4) &= Z_4\lambda_2(y_2), \end{aligned} \tag{4.36}$$

where $X \sim \lambda$ means that λ denotes the hazard function of X. Hence, the baseline hazard of the lifetime of the first (X_1) and second twin (X_2) with respect to the first cause of death are assumed to be equal (given by λ_1). The same is true for the lifetime of the first (Y_1) and second twin (Y_2) with respect to the second cause of death (λ_2). We assume that X_1, Y_1, X_2, and Y_2 are independent given the vector of frailties (Z_1, Z_2, Z_3, Z_4). Let $V_1, V_8 \sim \Gamma(k_1, \lambda_0)$, $V_2 \sim \Gamma(k_2, \lambda_1)$, $V_3 \sim \Gamma(k_3, \lambda_2)$, $V_4, V_7 \sim \Gamma(k_4, \lambda_2)$, $V_5, V_6 \sim \Gamma(k_5, \lambda_1)$ independent gamma distributed random variables with parameters $k_1+k_2+k_5 := \lambda_1 = \frac{1}{\sigma_1^2}$ and $k_1 + k_3 + k_4 := \lambda_2 = \frac{1}{\sigma_2^2}$. Now the frailties are given by the following construction:

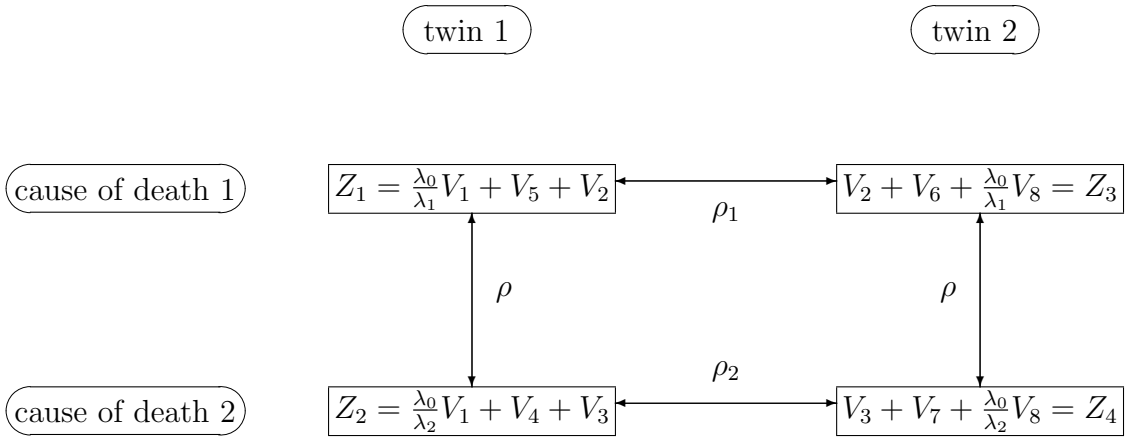

Figure 4.2: Cause-specific frailties and their correlations in a twin pair

Here, Z_1, Z_3 denote frailties with respect to the first cause of death (cause under study) and Z_2, Z_4 denote frailties with respect to the second cause of death of both individuals. Furthermore, the parameter ρ describes the correlations between the frailties: $\rho_1 = \mathbf{corr}(Z_1, Z_3)$, $\rho_2 = \mathbf{corr}(Z_2, Z_4)$ and $\rho = \mathbf{corr}(Z_1, Z_2) = \mathbf{corr}(Z_3, Z_4)$. Now the four-dimensional survival function can be derived by averaging over the conditional lifetimes, using relation (4.36) and applying the Laplace transform of gamma distributed random variables (for more detailed calculations see Appendix F):

$$\begin{aligned} S(x_1, y_1, x_2, y_2) &= \mathbf{E}S_1(x_1)^{Z_1}S_2(y_1)^{Z_2}S_1(x_2)^{Z_3}S_2(y_2)^{Z_4} \\ &= (S_1(x_1)^{-\sigma_1^2} + S_1(x_2)^{-\sigma_1^2} - 1)^{-\frac{\rho_1}{\sigma_1^2}}(S_2(y_1)^{-\sigma_2^2} + S_2(y_2)^{-\sigma_2^2} - 1)^{-\frac{\rho_2}{\sigma_2^2}} \\ &* \quad (S_1(x_1)^{-\sigma_1^2} + S_2(y_1)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}}(S_1(x_2)^{-\sigma_1^2} + S_2(y_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\ &* \quad S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_1(x_2)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_2(y_1)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \end{aligned} \tag{4.37}$$

with $0 \leq \rho \leq \min\{\frac{\sigma_2}{\sigma_1}(1-\rho_1), \frac{\sigma_1}{\sigma_2}(1-\rho_2)\}$. In this model ρ_1 denotes the correlation between Z_1 and Z_3, the main parameter of interest. This parameter describes the correlation of frailties of individuals in a pair with respect to the cause of death under study and is the key figure for genetic analysis of susceptibility to death from cause-specific mortality. The second parameter ρ_2 models the correlation between frailties with respect to all other causes of death (combined to the second cause of death or informative censoring). The parameter ρ is responsible for the association between causes of death in each individual. With this parameter, it is possible to test the hypothesis of dependence between competing risks in the above model.
S_1 and S_2 are the marginal survival functions with respect to the first and second cause of death. Note that it is impossible to use the Kaplan-Meier estimator to get non-parametric estimates of the marginal survival functions because of the assumed dependence between the two competing risks. To overcome this problem we use a parametric approach by fitting a Gamma-Gompertz model to the data, so that $S_j(t) = \left(1+s_j^2\frac{a_j}{b_j}(e^{b_j t}-1)\right)^{-\frac{1}{s_j^2}}$, $(j = 1, 2)$, where a_j, b_j, s_j^2 are parameters to be estimated. It is necessary to account for left truncation in the data. The likelihood function of this model is given in Appendix F. Due to the assumption concerning non-informative censoring with respect to (C_{1j}, C_{2j}), the function C does not enter the likelihood function.

4.12.2 Simulation study

To evaluate the performance of the estimates in the model described above a simulation study was carried out. All simulations involve generating gamma-distributed frailties, bivariate lifetimes, dependent and independent censoring times and truncation times. We will try to mimic the characteristics of the Danish twin data which we analyze in the next section. A total of 8000 twin pairs (3000 MZ and 5000 DZ pairs) are simulated, a number which is reduced by the truncation process to a final sample size of around 4200-4300 twin pairs. Samples are generated as follows:

- Generate frailty variables Z_1, Z_2, Z_3, Z_4 using independent gamma-distributed random variables $V_1, \ldots, V_8$.
- Generate lifetimes with respect to the first (X_1, X_2) and second disease (Y_1, Y_2) given the frailties using $S_i(t|Z) = e^{-Z\frac{a_i}{b_i}(e^{b_i t}-1)}$, $(i = 1, 2)$.
- The censored (bivariate) lifetimes $T_{ij} = \min\{X_{ij}, Y_{ij}, C_{ij}\}$ are generated by using the lifetimes with respect to the second cause of death as dependent censoring times and uniform distributed random variables on [40,100] as independent censoring times.

- Birth years are generated by using a uniform distribution on [1870,1930] to mimic the truncation pattern.
- The year of truncation is 1943.

For the simulations 1,000 data sets were generated. The mean parameter estimates of the model are shown in Table 4.15, in comparison with the true values used for simulation. Although there appears to be some bias in certain parameter estimates, the magnitude does not appear to be of any practical significance and the overall performance is quite good.

Parameter	true value	Mean of estimates	standard deviation
a_1	1.000	1.008	0.278
b_1	0.120	0.121	0.008
s_1	2.000	1.991	0.200
a_2	1.000	0.974	0.303
b_2	0.120	0.123	0.009
s_2	2.000	2.061	0.254
σ_1	2.000	1.964	0.288
σ_2	2.000	2.174	0.689
$\rho_1(MZ)$	0.400	0.408	0.073
$\rho_1(DZ)$	0.200	0.205	0.051
$\rho_2(MZ)$	0.100	0.107	0.064
$\rho_2(DZ)$	0.060	0.067	0.047
ρ	0.500	0.539	0.237

Table 4.15: Parameter estimation in the simulation study.

4.12.3 Example

In our example, we investigated how well the method performed when used to analyze the respective influence of genetic and environmental factors affecting risk of mortality from coronary heart disease (CHD). In this example the second cause of death is all other causes combined. The data we use for our analysis are the survival times of MZ and DZ female twins sampled from the Danish Twin Registry, described in more detail in Appendix A.2

Based on methods of quantitative genetics described in detail in Appendix B, we compared the ADE and DE as well as the ACE and CE models. The likelihood ratio test prefers the DE and the ACE model (Table 4.16). The ACE model converges to the AE

model. Standard errors are not given in the ACE model since 0 is the boundary of the parameter space. A comparison of the AE and the DE model cannot be done using the likelihood ratio test because the models are not nested. According to the Akaike Information Criterion (AIC), the AE model fits the data best and gives a inheritance estimate of 0.29, with standard error 0.09. Using the sub-model of independent causes of death ($\rho = 0$, standard correlated gamma frailty model), the inheritance estimate was found to be 0.58 (0.14).

	σ	a^2	d^2	c^2	e^2	ρ	Log-L
ACE	1.35 (—)	0.29 (—)		0.00 (—)	0.71 (—)	1.00 (—)	-22266.97
AE	1.35 (0.18)	0.29 (0.09)			0.71 (0.09)	1.00 (—)	-22266.97
ADE	1.34 (0.18)	0.25 (0.18)	0.05 (0.19)		0.71 (0.09)	1.00 (—)	-22266.97
DE	1.34 (0.27)		0.32 (0.09)		0.68 (0.09)	1.00 (—)	-22268.06
CE	1.34 (0.19)			0.19 (0.06)	0.81 (0.06)	1.00 (—)	-22270.21
AE*	1.87 (0.41)	0.58 (0.14)			0.42 (0.14)	0.00 (—)	-22269.24

Table 4.16: Parameter estimates (s.e.) σ^2 - variance of frailty, a^2 - additive genetic effects, d^2 - genetic effects due to dominance, c^2 - shared environment, e^2 - non-shared environment, ρ - correlation between frailties associated with competing risks, Log-L - value of the Log-Likelihood function, AE* - AE model with $\rho = 0$ (independent model)

4.12.4 Discussion

Frailty models are mixture models within survival analysis. In survival analysis, one typically has to deal with censored observations. In most applications censoring is assumed to be simply non-informative. In clinical studies, for example, where patients contribute censored observations because they are still alive at the preassigned termination point of study, this assumption is realistic. Some others get lost during the time of follow-up for reasons that are not related to the event under study. In such cases censoring can be assumed to be non-informative. However, in some cases this assumption is questionable, especially in cases where there are competing causes of death. This thesis suggests that in such cases, where only a part of the censored observations is assumed to be non-informatively censored an extension of the bivariate correlated gamma frailty model (Yashin et al., 1993; Pickles et al., 1994) be used. Because competing risks can also be correlated within families and may share unobserved dependencies with the cause of interest, the standard approach, which treats competing risks as independent, could lead to biased estimates of the variance components associated with the cause of interest. Here, the frailties are modelled in terms of standard variance components for additive

and dominance genetic effects and shared and unique environmental effects. This thus provides a rich class of models for analyzing this complex pattern of dependencies between family members and between causes of death. Furthermore, frailty models are well suited for inclusion of observed covariates in the analysis (Wienke et al., 2005a).

Using cause-specific mortality data of relatives (here twins) it is possible to overcome problems due to identifiability in univariate censored lifetimes as stated in Tsiatis (1975). The model we have developed allows for dependencies among competing risks, and makes it possible to test for such dependencies. Furthermore, combining methods from survival analysis (especially from frailty models) and genetic analysis as we did, improves the genetic analysis of time-to-event data in the case of informative and non-informative censoring together as well as accounting for heterogeneity in the population. Our example is an extension of the analysis in the case of independent causes of death (Wienke et al., 2000, 2001, 2003b, 2004), where deaths from other causes than the cause under study are treated as non-informative and collapsed with censored observations caused by end of study. In both cases (called here dependent and independent) the AE model is the best fitting model for CHD. This shows a certain degree of consistency in the model. Comparing both cases, it turns out that the heritability of frailty on mortality due to CHD changes substantially. Fixing the correlation between frailty on mortality from CHD and frailty on mortality from other causes to zero has an impact on the heritability estimate - changing it from 0.29 (0.09) to 0.58 (0.14). Both models detected the significant influence of genetic factors. The parameter ρ can be used to test the hypothesis of dependence between the competing risks. In the present case $\rho = 1$ holds, which means that a shared frailty model is the best fitting model for the competing risks. The likelihood ratio test indicates that the simpler independent model is sufficient to describe the data.

Mortality patterns in twins and in the general population are very similar (Christensen et al., 1995, 2001), which is an important argument for generalizing the results of twin studies to the general population.

The proof of consistency and asymptotic normality of the maximum likelihood estimators is still an open problem, but our simulation results support the asymptotic validity of the proposed method.

One important limitation of the presented model should be kept in mind; the correlation coefficients between the frailties are always non-negative by construction. This restriction makes sense when comparing the lifetimes of relatives, but it is not clear that the same holds for the competing risks in an individual. On one hand, many major diseases have risk factors in common and consequently, the presence of any one of these risk

factors will increase the risk of death with respect to all diseases. On the other hand, everyone dies eventually, so it is only logical that if the risk of death from one cause is decreased then the risk from another cause is increased. Furthermore, the parameter ρ is only identifiable in a 'real' multivariate case. Having pairs of unrelated individuals (e.g. $\rho_1 = \rho_2 = 0$), implying the univariate case, make the parameter ρ non-identifiable. The nature of dependencies among competing risks deserves further study.

Classical twin studies are based on the important assumption that MZ and DZ twins have the same correlation with respect to environmental factors (equal environment assumption). This standard assumption is necessary for the identifiability of heritability, i.e. so as to be able to interpret the difference in concordance between MZ and DZ twins as being explained in full by their difference in genetic concordance. However, without doubt, the assumption is also questionable: MZ twins are generally treated the same by their parents to a much greater extend than are DZ twins. This implies an overestimation of heritability. The equal environment assumption seems to be acceptable with respect to environmental factors related to CHD.

The suggested model gives a clear illustration of how the methods of survival analysis and genetic epidemiology may be merged to improve the genetic investigation of time-to-event data. Further extensions of the model to multiple causes of death and/or multiple related lifetimes will be important in elucidating the properties of this strategy.

A similar model for current state data was established in Giard (2001) and Giard et al. (2002). A more general approach compared to our model (4.37) was investigated by Bandeen-Roche and Liang (1996). Their hierarchical model is given by

$$S(t_1, \ldots, t_p) = \mathbf{E} e^{-\sum_k^N Z_k \sum_{j \in I_k} \Lambda(t_j)},$$

where p denotes the dimension of the multivariate observations in the cluster, N is the number of sub-clusters, Z_k the shared frailty in each sub-cluster, and I_k contains the indices of all observations belonging to sub-cluster k. In our case $p = 4$ holds and there exist four sub-clusters each with one single observation. The Z_k are given in picture 4.2. The difference in the models is in the observed data. In the model presented above only the minimum of two competing lifetimes in each individual is observed, whereas Bandeen-Roche and Liang (1996) assume that all (four) lifetimes are observable.

A similar approach was used by Huang and Wolfe (2002). Their model is based on log-normal frailty and relation (4.7), which means a model somewhere between shared and correlated frailty models. In a recent paper Huang et al. (2004) suggested a test procedure to test the hypothesis of dependence between survival and censoring times in their model.

Chapter 5

Different aspects of frailty modelling

5.1 Omitted covariates in hazard models

The result of omitted covariates in simple regression models is known: Parameter estimates are unbiased unless the omitted risk factors are correlated with the included observed covariates. In hazard models even this result does not hold. Many univariate biostatistical studies have focused on evaluating the effect of omitted covariates on the estimates of regression parameters β, see for example Gail et al. (1984), Schumacher et al. (1987), Schumacher (1989), Bretagnolle and Huber-Carol (1988), Chamberlain (1992), Hougaard et al. (1994), Schmoor and Schumacher (1997), and Keiding et al. (1997). In hazard models, neglecting a subset of the important covariates leads to biased estimates of both regression coefficients and the hazard rate. The reason for such bias is the time-dependent hazard rate, a unique element in survival analysis. To understand this result, consider the example of a population initially consisting of two sub-populations, a high-risk group and a low-risk group, each subject to a constant risk (binary frailty model). If this heterogeneity is not accounted for and a common hazard is assumed for the whole population (as a step function for example), the estimated hazard will not be a simple average of the constant hazards of the two groups, but it will fall over time.

Consequently, in a simple proportional hazards model with observed covariates, an omitted covariate results in distorted parameter estimates for the included factors, even when the distribution of omitted characteristics is initially the same in all categories of included covariates. The problem is that the distribution of characteristics can not remain the same across categories over time. This holds because high-risk individuals with certain values of the omitted covariate will die earlier, so that the distribution of omitted characteristics within each category of an included covariate must also change over time. As consequence, all parameter estimates must be contaminated by the effect of the omitted

risk factor. This is quite disturbing since one can never know the true model.

Bretagnolle and Huber-Carol (1988) predict the sign of the bias in the two following cases: first, when there is only one covariate left in the used model, whatever the number of omitted covariates is. Then the effect on the survival of the covariate under study is always underestimated. Second, in case of several covariates remaining in the analyzed model the authors prove that the same result of underestimation holds for each of them at least up to some fixed time which, in practical cases, is reasonably long. The asymptotic bias resulting from such omissions is not negligible, as shown by simulations. In the case of left truncated and right censored data it may even happen that the estimates crosses 0, that is a beneficial covariate ($\beta < 0$) results, through the biased estimate in an unfavorable one ($\beta > 0$). Nevertheless this crossing 0 phenomenon does not happen with only right censoring if one assumes that censoring is independent on the included covariates.

Univariate frailty models are one (but not the only) way to account for the effect of omitted covariates in a proportional hazards model with observed covariates. For reasons of convenience, analysts frequently choose parametric representations of the frailty distribution that are mathematically tractable. With such parametric assumptions about heterogeneity, the hazard can be represented parametrically or semi-parametrically. There could be the issue of misspecifying the frailty distribution. Heckman and Singer (1982b) demonstrate, however, for a given parametric baseline hazard function results can be very sensitive to the choice of the parametric form of the frailty distribution, even when a flexible form is chosen (see also Heckman and Singer (1984), Keiding et al. (1997)). For this purpose, Heckman and Singer (1982b) use a simple parametric family of distributions to investigate this sensitivity (log normal, normal, gamma distribution). The authors propose a non-parametric representation of the frailty distribution and through simulations show, that parameter estimates of the covariate effects can be estimated with great precision. The frailty distribution is approximated by a discrete mixture with finitely many mass points, where the number, location and probability mass associated with each point are to be estimated (see last part of section 3.1). The results of Heckman and Singer (1982b) indicate that the distribution of heterogeneity cannot be approximated well. Thus their procedure corrects for heterogeneity without clearly identifying its distribution.

It should be noted that Heckman and Singer's (1982b) arguments against parametric assumptions about the frailty distribution are points of discussion and controversies by authors who claim that Heckman and Singer misspecified the duration dependence in the example data set used to demonstrate the potentials of their nonparametric approach.

Newer studies conclude that parameter estimates are not that sensitive with respect to the choice of the frailty distribution (see Manton et al. 1986; Klein et al. 1992; Guo and Rodriguez 1992; Guo 1993).

To get any results at all with a non-parametric frailty distribution, however, one must impose a parametric form on the baseline hazard function like Weibull or Gompertz distribution. What is not mentioned in Heckman and Singer (1982b) but is demonstrated, though not emphasized, by Manton et al. (1981) is that results can also be sensitive to the functional form of the assumed baseline hazard. Trussell and Richards (1985) show that parameter estimates could be very sensitive with respect to the assumed parametric form of the baseline hazard function when using the non-parametric frailty distribution suggested by Heckman and Singer (1982b).

Henderson and Oman (1999) intend to quantify the bias which may occur in estimated covariate effects and fitted marginal distributions when frailty effects are present in survival data but ignored in a wrongly specified proportional hazards analysis. They considered this problem for positive stable frailty models (see Example 6). In general, however, no explicit solutions were found. Henderson and Oman (1999) used first order approximations to estimate the bias in different situations. It turns out that the amount of bias depends on the form of frailty distribution, but to the first order approximation the bias is independent of the covariate distribution. Censoring in the Koziol-Green model (the censoring survival function is a power of the baseline survival function) and type I censoring show that the asymptotic bias is reduced by censoring, particularly with log-normal frailty. For non-censored survival and gamma-distributed frailty with variance parameter σ^2, the following approximation holds:

$$\beta = (1 + \sigma^2)\beta^* \tag{5.1}$$

where β, β^* denote the unbiased and biased parameter, respectively, in the Cox model, assuming that the covariates are centered. Consequently, the bias due to omitted covariates becomes more important with increasing heterogeneity (frailty variance) and relation (5.1) could be used to compensate for the frailty effect.

Congdon (1995) used Weibull and Gompertz hazard specifications to investigate the influence of different frailty distributions (gamma, inverse Gaussian, stable, binary) on total and cause-specific mortality data from London area 1988 - 1990.

5.2 Dependence between frailty and observed covariates

Nearly all contributions and applications on frailty models in a regression setting like (3.2) consider the frailty Z as independent of the observed explanatory variables X. This strong assumption is made only for mathematical convenience. Of course, in some cases this independence assumption is inappropriate. In a competing risk setting this question was analyzed by di Serio (1997) to explain the effect of unexpected protectivity. This problem occurs when one covariate shows a protective impact not expected from a medical point of view. A simulation study was conducted in a two competing risks scenario with two frailties. Each frailty is related to one competing event. It turns out that dependence between the two frailties alone does not cause false protectivity; conversely, false protectivity may occur according to the magnitude and the sign of the dependence between frailty and the observed covariate. Here, dependence between the two frailties means dependence on the two competing event times, resulting in a univariate lifetime model with dependent censoring. The problem is that an assumption about the distribution of the frailty is not enough, since what we need are actually (non-testable) assumptions about the direction and the intensity of the dependence between the frailty and the observable covariates, which makes analysis quite arbitrary. Thus, survival models where observed covariates depend on unobserved frailty may not be identifiable from univariate data.

It turns out that such models can be identified from bivariate data when they are imbedded into a correlated gamma frailty model by using the relation $\sigma(X) = \sigma e^{\gamma X}$, where γ is a parameter to be estimated and σ^2 denotes the variance of the frailty variable Z. Thus, the distribution of Z (or more specificly their variance) depends on the observed covariate values X.

An example of such kind of models (but only in the multivariate case) can be found in Wassell and Moeschberger (1993), who study the impact of interventions in the Framingham Heart Study by introducing the notion of modified gamma frailty. Yashin et al. (1999b) investigate the heritability of susceptibility to death after accounting for the dependence between frailty and observed covariates (BMI and smoking) in Danish twins. Noh et al. (2006) suggested the notion of dispersed frailty for such kind of models and applied it to the well known kidney infection data from McGilchrist and Aisbett (1991) to verify the hypothesis of heterogeneity in the frailty distribution in the study population.

5.3 Tests for heterogeneity

In practical application of methods from survival analysis the researcher is usually confronted with the problem of whether unobserved heterogeneity is present in the data or not. For this problem a number of tests for heterogeneity have been proposed during the last 15 years. The relevant testing problem for the null hypothesis of no unobserved heterogeneity is

$$H_0 : \sigma^2 = 0 \quad \text{versus} \quad H_A : \sigma^2 > 0,$$

where σ^2 denotes the variance of the frailty (unobserved heterogeneity). In this situation a one-sided testing hypothesis is natural. The main problem here is that the heterogeneity parameter is on the boundary of its parameter space under H_0 and the classical likelihood ratio asymptotic chi-squared distribution theory is no longer valid. It turns out that in many cases the likelihood ratio statistic has an asymptotic null distribution which is an equal mixture of a point mass at zero and a chi-square distribution with one degree of freedom, often denoted $0.5\ \chi_0^2 + 0.5\ \chi_1^2$. Rigourously established limiting distributions of the likelihood ratio statistic are provided by Maller and Zhou (2003) and Nguti et al. (2004). Other, generally more applied results about tests for heterogeneity in survival data can be found in Oakes (1982), Blossfeld und Hamerle (1989, 1992), Hamerle (1992), Commenges and Andersen (1995), Gray (1995), Commenges and Jacqmin-Gadda (1997), Verweij et al. (1998), Bordes and Commenges (2001), Dunson and Chen (2004). A test procedure for proportional hazards in the presence of nonobserved heterogeneity is suggested by McCall (1994).

5.4 Log-rank test in frailty models

In analyzing time-to-event data of two groups with the log-rank test, a proportional hazards model is usually assumed. Nevertheless, this hypothesis is no longer valid in the case of unobserved covariates. Omission of a balanced covariate from a proportional hazards model generally leads to a model with non-proportional hazards. As a consequence, the simple log-rank test is no longer optimal. A general treatment of this problem is given in Oakes and Jeong (1998). The authors establish connections between theories of weighted log-rank tests and of frailty models and analyze parametric as well as non-parametric models. Furthermore, the authors consider different frailty distributions: positive stable, inverse Gaussian, gamma and two-point distribution and the models are extended to include random censoring. The optimal limiting weight function is

$$w(t) = 1 + \Lambda_0(t)\Big(\frac{\mathbf{L}''(\Lambda_0(t))}{\mathbf{L}'(\Lambda_0(t))} - \frac{\mathbf{L}'(\Lambda_0(t))}{\mathbf{L}(\Lambda_0(t))}\Big),$$

where $\mathbf{L}'$ and $\mathbf{L}''$ are the first and second derivative of the Laplace transform $\mathbf{L}$ of the frailty distribution. For example, in the gamma frailty model Oakes and Jeong (1998) derived the expression

$$w(t) = (1 + \sigma^2 \Lambda_0(t))^{-1}$$

(corrected for an error in their formula and transformed to our notation). The asymptotic relative efficiency of the simple log-rank test and of the optimally weighted log-rank test relative to the adjusted test that would be used if the covariate values were known are given in terms of the Laplace transform of the frailty distributions. These frailties represent the effect of the missing covariates. Two main conclusions are derived by Oakes and Jeong (1998) from their analysis. First, the loss of efficiency from omitting a covariate is generally more important than the additional loss due to the misspecification of the resulting non-proportional hazards model. Furthermore, the loss of efficiency increases with higher variance of the frailty distribution. However, binary frailty distributions can provide exceptions to this rule, reflecting the fact that for a very extreme frailty distribution, one can identify the values of the frailties from the ordering of the failures - the individuals who experience the event of interest first are almost certain to have the higher (of the only two possible) values of frailty. Consequently, the knowledge of the values of the frailties (which means the values of missing covariates) becomes less important. Second, censoring tends to remove the late events and, consequently, reduces the loss of efficiency in the simple log-rank test compared to the adjusted log-rank test. The above results depend on the knowledge of the unknown frailty variance σ^2, consequently, their practical applicability is limited.

Jeong (2003) extended the aforementioned results to quantify and visualize the loss of efficiency of the log-rank test when a dependence structure between survival and censoring times is being ignored. The assumed dependence structure is based on a correlated gamma frailty model (see section 4.4). In the given situation the loss of efficiency is minimal under the proportional hazards model, even when the correlation between potential survival and censoring times is strong unless the dependent censorship causes a severe non-proportionality.

Broët et al. (1999) suggest a more practical approach for taking into account unobserved covariates in a weighted log-rank test. The construction of the test is based on a gamma distribution of the frailty. For the frailty parameter σ^2 a range is assumed. Simulations investigate the power of the test for different frailty distributions.

5.5 Time-dependent frailty models

Standard frailty models assume that the individual frailty Z is determined at time zero (begin of follow-up) and follows the individual throughout the rest of life, resulting in the individual risk later in life being perfectly correlated with that at the beginning. It is clearly of interest to overcome this restrictive assumption of fixed frailty in order to develop more flexible models.

There are different approaches to dealing with time-dependent frailty models by regarding the hazard of an individual as a stochastic process. This provides individual flexibility. However, since what is observable is still the average hazard function, the main problem is to find models which show a tractable mathematical connection between the individual hazard functions and the average one. One attempt to tackle this problem was suggested by Yashin and Manton (1997), defining the hazard function as a diffusion process by means of a stochastic differential equation. By using Ornstein-Uhlenbeck-type processes such kind of models reflect the fact that many biological parameters tend to stabilize around certain values, which is called homeostasis.

In a second approach McGilchrist and Yau (1996) and Yau and McGilchrist (1998) focus on recurrent event time data. For the inter-recurrence time data, it is conceived that inter-recurrence times closed to each other on the times scale are highly associated, while times which are further apart from each other on the time scale are less correlated. To model such kind of serial dependence structure, a dynamic frailty model can be constructed by assuming the frailties of subsequent time intervals following an auto-regressive process of order one. Yau and McGilchrist (1998) adopted AR(1) frailty models to analyze chronic granulomatous disease data. A similar but simpler idea without using an AR(1) process was suggested by Wintrebert et al. (2004). A Bayesian approach can be found in Manda and Meyer (2005).

The third approach is based on Lévy processes (Gjessing et al., 2003). Like diffusion processes, hazard functions driven by Lévy processes also yield some degree of tractability. It is possible to get explicit formulae for the relationship between conditional and unconditional hazard, and the frailty and hazard of survivors may be calculated. A basic difference between this and the first approach is the jump nature of the Lévy processes. However, it could well be imagined that the individual hazard may increase in jumps, for example with the onset of an acute disease. The model is an extension of the fixed frailty model in Aalen and Hjort (2002).

5.6 Identifiability of frailty models

One of the aims of frailty modelling is to study the properties of the hazard function and the frailty distribution in real life applications. Such analysis is possible if the frailty distribution as well as the hazard function can be reconstructed (identified) from the survival data. Let $S(t)$ represent some survival function. It turns out that $S(t)$ can always be represented as a marginal survival function in some model with unobserved frailty Z and cumulative baseline hazard $\Lambda_0(t)$. In other words univariate frailty models are not identifiable if only the marginal distribution of the failure times are known. Indeed, let $\mathbf{L}$ be the Laplace transform of some arbitrary distribution of the non-negative random variable Z, and let $\mathbf{L}^{-1}$ be the inverse function of $\mathbf{L}$. For any $0 \leq t < \infty$ the cumulative hazard is defined as

$$\Lambda(t) = \mathbf{L}^{-1}(S(t)).$$

To prove that $\Lambda(t)$ is a cumulative hazard function of a survival distribution it is sufficient to show that $\Lambda(t)$ is a non-decreasing function. This follows from the fact that both $S(t)$ and $\mathbf{L}^{-1}(t)$ are non-increasing functions of t. Thus, for any given survival function $S(t)$ and probability distribution on the positive half-line (given by Laplace transform $\mathbf{L}$), one can construct a frailty model with marginal survival function $S(t)$. The non-identifiability aspect for particular distributions of frailty was discussed by Heckman and Singer (1982a, 1984) and Hoem (1990) among others. Elbers and Ridder (1982) show that if frailty distribution has a finite mean then the presence of observed covariates makes the univariate proportional hazards frailty model identifiable. Other conditions are given by Heckman and Honoré (1989). These findings stimulated the use of frailty models with observed covariates (regression models) in the analysis of univariate survival data (Andersen et al., 1993).

Bivariate data present an opportunity to identify the frailty distribution and the baseline hazards. For example, all shared frailty models are identifiable from bivariate data without additional information such as observed covariates or parametric assumptions about the baseline hazards. It turns out that the identifiability property holds for a broader class of frailty models, including correlated frailty models. Furthermore, the condition of a finite mean required for the identifiability in a univariate frailty model with observed covariates by Elbers and Ridder (1982) is not needed in this case. For more details see Yashin and Iachine (1999b). Identifiability aspects of the correlated frailty model are analyzed in detail by Iachine and Yashin (1998) and further extended by Iachine (2004), where the latter paper contains an interesting counterexample.

Chapter 6

Summary

In the present thesis we discuss several aspects of univariate and bivariate frailty models. The proportionality of hazards make frailty models a simple tool for analytical and computer calculations. Despite their simplicity these models turn out to be useful in providing insights and ideas for analyzing reality.

In the univariate case they explain the deceleration of mortality rates at older ages (Vaupel et al., 1998) and allow for the evaluation of bias in regression coefficients when calculating the effect of observed covariates on survival in the presence of heterogeneity. This heterogeneity may be difficult to assess, but it is nevertheless of great importance. The key idea of univariate frailty models is that individuals have different frailties, and that the most frail will die earlier than the lesser frail. Consequently, systematic selection of robust individuals takes place, which distorts what is observed. When mortality rates are estimated, one may be interested in how they change over time or age. Quite often, they rise at the beginning of the observation period, reach a maximum and then decline (unimodal intensity) or level off. This, for example, is typical for death rates of cancer patients, meaning that the longer the patient lives beyond a certain time, the better his or her chances of survival are. It is likely that unimodal intensities are often the result of selection and that they do not reflect an underlying development on the individual level. The population intensity starts to decline simply because the high-risk individuals have already died. The intensity of a given individual might well continue to increase. An estimate of the individual hazard rate without taking into account the unobserved frailty will thus underestimate the hazard function to an increasingly greater extent as time goes by.

In general, two sources of variability in duration data are useful to consider: variability accounted for by observable risk factors (which is thus theoretically predictable) and heterogeneity caused by unknown covariates, and which is therefore theoretically unpre-

dictable even when all relevant information at that time is known. It is the latter which is of specific interest in the present thesis, and the subject of observable covariates is treated here only for completeness. As Hougaard (1991) pointed out, there are advantages in considering these two sources of variability separately: heterogeneity explains some 'unexpected' results or gives an alternative explanation of some results, for example non-proportional or decreasing hazards.

The precise nature of the relationship between individual and population ageing depends on the distribution of frailty among individuals. Different choices of distributions for the unobserved covariates are possible, including binary, gamma and log-normal, which show both qualitative and quantitative differences. The compound Poisson frailty model includes interesting special cases (gamma and log-normal frailty model) and is very flexible. Even cure models can be integrated in this framework in an elegant and natural way. The variance of the frailty distribution determines the degree of heterogeneity in the study population.

Bivariate frailty models help to clarify the roles of genes and environment in life span, time to onset of diseases, and other durations. Some of the bivariate frailty models admit semi-parametric representations (copulae). Such representation allows us to avoid unjustifiable technical assumptions about the parametric structure of the conditional survival function given frailty. It also facilitates the estimation procedure and improves the interpretation of the results of data analysis. The structure of the correlated frailty models is appropriate for addressing questions about the genetic nature of individual susceptibility to disease and death. The bivariate models are identifiable even without observed covariates. The dependent competing risks problem (which is non-identifiable in the univariate case) becomes identifiable from bivariate data about cause-specific mortality for related individuals. The methodology of heterogeneity analysis is thereby extended to the multivariate case. As a result, new heterogeneity models arise. For example, the newly proposed correlated compound Poisson frailty model is an interesting extension of the correlated gamma frailty model, which allows us to consider a much richer class of models in the data analysis. As in the univariate case, the problem of a possible cure fraction is also solved by this model.

Despite this clear advantage of frailty modelling, further work in improving the methods of analysis of univariate and bivariate survival data is needed. This is because the assumptions of fixed frailty and the proportionality of hazard, used in these models, are not always realistic. The interpretation of frailty often depends on identifiability conditions. For example, the characteristics of frailty identified by parametric demographic (e.g. gamma-Makeham) models differs from characteristics of frailty in biostatistical mod-

els with Elbers and Ridder (1982) identifiability conditions, which in turn differ from characteristics of frailty identified from bivariate data using a correlated gamma frailty model. The assumption that frailty is independent of observed covariates is questionable. Shared frailty models for related individuals should be used with care. Correlated frailty models may be not flexible enough in describing multiple correlations in clusters of larger size. To overcome current problems with the genetic analysis of durations, extended versions of susceptibility models have to be developed. For example, dynamic models of frailty need to be introduced and analyzed. Possible dependence between frailty and observed covariates should be taken into account. Models for pedigree data with observed covariates need to be developed. Quadratic hazards models of liability should be tested.

The major criticism of frailty models is concerned with the lack of sufficient biological background for their basic assumptions: the proportional structure of the hazards, the form of a frailty distribution, and the independence of observed and omitted covariates. Furthermore, the inability to measure frailty directly often complicates the interpretation of the results of a frailty based analysis. Another limitation is that frailty models are usually formulated in terms of the survival function. Consequently, their application to p-variate survival data requires taking partial derivatives of the order p. Hence, generalization from the bivariate to the multivariate case is not trivial.

Despite these limitations, frailty models provide a valuable tool for the analysis of univariate and bivariate time-to-event data especially in medicine, (genetic) epidemiology, demography, biology and other sciences. They form an interesting alternative to traditional models in survival analysis.

Appendix A

Data

To emphasize the practical purpose of the frailty models considered, we provide brief descriptions of data sets from the Swedish and Danish twin registries serving as illustrations for the chapters above.

A.1 The Swedish Twin Registry

First established in the late 1950s to study the importance of smoking and alcohol consumption on cancer and cardiovascular diseases while controlling for genetic propensity to disease, the Swedish Twin Registry has today developed into a unique source. Since its establishment, the registry has been expanded and updated on several occasions, and the focus has similarly been broadened to include most common complex diseases. Today it contains around 70,000 twin pairs born between 1886-1990. Nearly 400 papers have been published based on data from the Swedish Twin Registry.

When the twin registry was initiated, the church registers from all parishes of the time period 1886-1925 were checked manually to identify all twin births. Between 1959 and 1961, a questionnaire was sent to all twins, which included a question about phenotypic similarities to assess the zygosity: 'Were you as children as alike as two peas in a pod?' When both partners agreed, they were defined as MZ twins. If both responded 'not alike' they were classified as dizygotic. If the twins did not agree, or if only one member of the pair answered the question, the pair was classified as unknown zygosity. This zygosity classification was compared with laboratory methods (serological markers). The misclassification rate for this method was found to be very low (Cederlöf et al., 1961).

At present, the Swedish Twin Registry contains information about three cohorts of Swedish twins referred to as the 'old', 'middle' and 'young' cohort. The old cohort consists of all same-sexed pairs born between 1886 and 1925 where both members in

a pair were living in Sweden in 1959. In 1970 a new cohort of twins born between 1926 and 1958, the middle cohort, was compiled, this time by use of nationalized birth registrations. Twins born later than 1958 are not considered in the present analysis because of the small number of cases. The data are described in Table A.1 and A.2, categorized according to the censoring status. The event under study is the onset of breast cancer. If a woman did not develop breast cancer or if she died from other causes during the follow-up, the corresponding observation is censored.

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.

Data about coronary heart disease in the old and middle cohorts (including information about observed covariates which are important with respect to coronary heart disease) are also available and were analyzed in detail by Zdravkovic et al. (2002, 2004).

For a comprehensive description of the Swedish Twin Registry database, with a focus on the recent data collection efforts and a review of the principal findings that have come from the registry, see Lichtenstein et al. (2002).

	both censored	one censored	none censored	total
MZ	1767	218	18	2003
DZ	3420	407	27	3854
total	5187	625	45	5857

Table A.1: Composition of the old cohort (born 1886 - 1925) by zygosity and censoring status. Swedish Twin Registry. Number of twin pairs. End of follow-up October 2000.

	both censored	one censored	none censored	total
MZ	2537	117	15	2669
DZ	3816	218	8	4042
total	6353	335	23	6711

Table A.2: Composition of the middle cohort (born 1926 - 1967) by zygosity and censoring status. Swedish Twin Registry. Number of twin pairs. End of follow-up October 2000.

A.2 The Danish Twin Registry

The Danish Twin Registry was the world's first nation-wide twin registry, established in 1954 by Bent Harvald and Mogens Hauge. The older part of this population-based registry includes all twins born in Denmark during the period 1870-1910 and all like-sex pairs born between 1911 and 1930 in which both partners survived to the age of 6 years. The birth registers from all 2,200 parishes of the relevant calendar years were manually scrutinized to identify all multiple births. A search was then carried out for twins, or whenever needed, for their closest relatives in regional population registers (in operation since 1924) or other public sources, especially the archives of probate courts and censuses. As soon as a twin was traced, a questionnaire was sent to the twin (if he/she was alive) or to the closest relatives (if not). Questions about phenotypic similarities were included in the questionnaires, so as to assess the zygosity by self-reported similarities. The reliability of this method was validated by comparison with laboratory methods based on blood, serum enzyme group determination. The misclassification rate was found to be less than 5% (Holm, 1983). The follow-up procedure traced nearly all twins who did not die or emigrate before the age of 6. For further, detailed information about the construction and the composition of the Danish Twin Registry see Hauge (1981). The data provided by the Danish twin registry contains records of 8,201 MZ and DZ twin pairs who were born between 1 January 1870 and 31 December 1930 and who were both still alive on 1 January 1943. 246 pairs with incomplete cause of death information were excluded, leaving a study population of 7,955 pairs. Individuals were followed up through 31 December 1993, and those identified as deceased after that date have been classified here as 'living'. Altogether, we have 1,344 male MZ twin pairs and 2,411 DZ twin pairs, 1,470 female MZ twin pairs and 2,730 DZ twin pairs. In addition to the lifetimes, there is information about cause of death for all non-censored lifetimes, i.e. for all individuals in the study population who died before 31 December 1993. For more detailed information about cause of death, gender, and zygosity of the study population see Table A.3.

Death status, age at death and cause of death were obtained from the Central Person Register, the Danish Cause-of-Death Register, the Danish Cancer Register, and other public registries in Denmark. The main source for obtaining information on cause of death was the Death Register at the National Institute of Public Health. Information about cause of death is available from this register for individuals who died after 1942 (Juel and Helweg-Larsen, 1999). Consequently, cause of death information was included in the twin register only for twins who died after this year. The validity of the twin

cause of death	males		females	
	MZ twins	DZ twins	MZ twins	DZ twins
cancer	440	809	423	823
coronary heart disease	506	897	369	704
stroke	161	278	186	335
respiratory diseases	143	203	89	205
other causes	496	944	509	850
all causes together	1746	3131	1576	2917
alive (censored)	942	1691	1364	2543

Table A.3: Danish twin data by sex, zygosity and cause of death. Number of individuals.

register was established on the basis of a comparison of information about year of death with the nation-wide Danish Cancer Register. There was 99 % agreement, although both registries were independent. Further data corrections increased this level of agreement to almost 100 %.

Information about cause of death was coded following the International Classification of Diseases (ICD, sixth, seventh and eighth editions). Four different groups of causes of death are considered here: cancer, coronary heart disease (CHD), stroke, and diseases of the respiratory system. Codes in three revisions of the ICD for these broad cause-of-death groups are given in the following table:

cause of death	ICD revision 6 & 7	ICD revision 8
cancer	140 - 205	140 - 209
coronary heart disease	420	410 - 414
stroke	330 - 334	430 - 439
respiratory diseases	470 - 527	460 - 519

Table A.4: Cause of death groups by ICD.

Appendix B

Quantitative Genetics

The aim of genetic analysis is to determine whether and to which extent genetic variation may account for the variation of a specific phenotype (e.g. lifetime, susceptibility to disease). To address this question, measurements of the phenotype must be combined with genetic information. In studies of related individuals (families, twins, litter, etc.) genetic information is usually available as pedigree. Based on methods of quantitative genetics developed by Falconer (1990) and Neale and Cardon (1992) the relative role of genetic factors is determined by estimating heritability.

The classic twin method, attributed to Francis Galton (1887), argues that whereas in MZ twins the differences in a pair must be due to environmental differences, for DZ twins it includes effects associated with their genetic difference. Through making a critical assumption, that differences with respect to environmental factors between identical pairs are of the same (average) strength as those between DZ twins (equal environment assumption), inference can be made about the role of genetic factors simply by comparing MZ and DZ pair correlations.

Typical models of quantitative genetics can be incorporated with ease into the correlated frailty model described above. Models of quantitative genetics (Falconer, 1990) are based on the decomposition of a phenotypic trait in a sum of different components, which are assumed to be independent. Using this approach, it is possible to estimate the proportion of the total variability of the phenotype which is related to genetic factors. In particular, a heritability estimate can be calculated for human longevity by identifying the phenotype with the life span variable (McGue et al. 1993). Usually, heritability is defined as the percentage of variation of the trait explained by the variation of genetic factors.

Yashin and Iachine (1995a) suggested an approach based on the frailty variable Z instead of the life span T. It is interesting to find out the relative importance of genes and the environment in determining individual susceptibility towards mortality (overall or cause

specific). An advantage of this approach is that, through the additive decomposition of frailty into a genetic and an environmental component, one can obtain a competing risk structure for the respective survival model. In other words, observed mortality is represented as a sum of two terms: one depends on genetic and another on environmental parameters, both estimated from bivariate data.
In more detail, let the frailty be represented by:

$$Z = genes + environment = A + D + I + C + E \tag{B.1}$$

where A represents additive genetic effects, D corresponds to genetic effects caused by dominance, I denotes epistatic genetic effects, and C and E stand for shared and non-shared environmental effects, respectively. All factors are assumed to be independent. The associated variance proportions of the components are defined as follows:

$$a^2 = \frac{V(A)}{V(Z)},\ d^2 = \frac{V(D)}{V(Z)},\ i^2 = \frac{V(I)}{V(Z)},\ c^2 = \frac{V(C)}{V(Z)},\ e^2 = \frac{V(E)}{V(Z)}$$

The following additive decomposition of the frailty variance and the correlation coefficient between co-twins' frailty hold that:

$$1 = a^2 + d^2 + i^2 + c^2 + e^2 \tag{B.2}$$

$$\rho = \rho_1 a^2 + \rho_2 d^2 + \rho_3 i^2 + \rho_4 c^2 + \rho_5 e^2 \tag{B.3}$$

where lowercase letters a^2, d^2, i^2, c^2, e^2 indicate the proportions of the total variance σ^2 associated with the correspondent components of frailty, and ρ_i $(i = 1, ..., 5)$ are correlations between respective components within a twin pair. In this case broad sense heritability can be expressed as

$$H^2 = a^2 + d^2 + i^2,$$

where a^2 denotes small sense heritability. Standard assumptions of quantitative genetics models specify different values of ρ_i $(i = 1, ..., 5)$ for monozygotic and dizygotic twins. In the case of monozygotic twins $\rho_i = 1$, $i = 1, ..., 4$ and $\rho_5 = 0$, while for dizygotic twins $\rho_1 = 0.5$, $\rho_2 = 0.25$, $\rho_3 = m$, $\rho_4 = 1$, $\rho_5 = 0$ and $0 \leq m \leq 0.25$ is an unknown parameter. Not all parameters of the genetic decomposition of frailty can be estimated simultaneously, even under the assumption of no epistasis ($i^2 = 0$). In this case it is only possible to conclude that the true heritability H^2 is in the interval (Iachine, 2002)

$$\frac{4}{3}(\rho_{MZ} - \rho_{DZ}) \leq H^2 \leq \min\{\rho_{MZ}, 2(\rho_{MZ} - \rho_{DZ})\}.$$

The model in fact reduces to three equations (two relationships (B.3) for monozygotic and dizygotic twins and one constraint (B.2)) allowing estimations of no more than three

parameters at the same time. One possibility is to consider an ACE (additive genetic - common environmental - uncommon environmental) model. In this case, equations (B.2) and (B.3) lead to the following:

$$\begin{aligned} 1 &= a^2 + c^2 + e^2 \\ \rho_{MZ} &= a^2 + c^2 \\ \rho_{DZ} &= 0.5a^2 + c^2 \end{aligned} \tag{B.4}$$

This system can be integrated into the correlated frailty model giving place to a re-parameterization of the original model. The only difference is that, if we are interested in estimating parameters of a genetic model, data for monozygotic and dizygotic twins have to be analyzed simultaneously and a likelihood function for combined data has to be drawn. Equally, other genetic models can be obtained combining no more than three components of frailty (Yashin and Iachine, 1995a).

It is necessary to note that heritability estimation requires strong assumptions which are often difficult to verify in practice. For example, the trait must be represented as an additive combination of uncorrelated genetic and environmental factors and the variances of phenotypic traits associated with related individuals must be the same. This does not decrease the statistical attractiveness of this direction of research, however the interpretation of heritability estimates must be used with care as pointed out by Feldman and Lewontin (1975).

Appendix C

Correlated gamma frailty model

$$S(t_1,t_2) = S_1(t_1)^{1-\frac{\sigma_1}{\sigma_2}\rho} S_2(t_2)^{1-\frac{\sigma_2}{\sigma_1}\rho} (S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}}$$

$$\begin{aligned}
S_{t_1}(t_1,t_2) &= -(1-\frac{\sigma_1}{\sigma_2}\rho) S_1(t_1)^{-\frac{\sigma_1}{\sigma_2}\rho} f_1(t_1) S_2(t_2)^{1-\frac{\sigma_2}{\sigma_1}\rho} (S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&- \frac{\sigma_1}{\sigma_2}\rho S_1(t_1)^{-\frac{\sigma_1}{\sigma_2}\rho-\sigma_1^2} f_1(t_1) S_2(t_2)^{1-\frac{\sigma_2}{\sigma_1}\rho} (S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}-1} \\
&= S_1(t_1)^{-\frac{\sigma_1}{\sigma_2}\rho} f_1(t_1) S_2(t_2)^{1-\frac{\sigma_2}{\sigma_1}\rho} (S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}-1} \\
&* \ \big(- S_1(t_1)^{-\sigma_1^2} - (1-\frac{\sigma_1}{\sigma_2}\rho) S_2(t_2)^{-\sigma_2^2} + (1-\frac{\sigma_1}{\sigma_2}\rho)\big)
\end{aligned}$$

$$\begin{aligned}
S_{t_2}(t_1,t_2) &= -S_1(t_1)^{1-\frac{\sigma_1}{\sigma_2}\rho} (1-\frac{\sigma_2}{\sigma_1}\rho) S_2(t_2)^{-\frac{\sigma_2}{\sigma_1}\rho} f_2(t_2) (S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&- S_1(t_1)^{1-\frac{\sigma_1}{\sigma_2}\rho} \frac{\sigma_2}{\sigma_1}\rho S_2(t_2)^{-\frac{\sigma_2}{\sigma_1}\rho-\sigma_2^2} f_2(t_2) (S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}-1} \\
&= S_1(t_1)^{1-\frac{\sigma_1}{\sigma_2}\rho} S_2(t_2)^{-\frac{\sigma_2}{\sigma_1}\rho} f_2(t_2) (S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}-1} \\
&* \ \big(- (1-\frac{\sigma_2}{\sigma_1}\rho) S_1(t_1)^{-\sigma_1^2} - S_2(t_2)^{-\sigma_2^2} + (1-\frac{\sigma_2}{\sigma_1}\rho)\big)
\end{aligned}$$

$$
\begin{aligned}
S_{t_1t_2}(t_1,t_2) &= (1-\frac{\sigma_1}{\sigma_2}\rho)(1-\frac{\sigma_2}{\sigma_1}\rho)S_1(t_1)^{-\frac{\sigma_1}{\sigma_2}\rho}f_1(t_1)S_2(t_2)^{-\frac{\sigma_2}{\sigma_1}\rho}f_2(t_2)\\
&* \ (S_1(t_1)^{-\sigma_1^2}+S_2(t_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}}\\
&+ \ (1-\frac{\sigma_1}{\sigma_2}\rho)\frac{\sigma_2}{\sigma_1}\rho S_1(t_1)^{-\frac{\sigma_1}{\sigma_2}\rho}f_1(t_1)S_2(t_2)^{-\frac{\sigma_2}{\sigma_1}\rho-\sigma_2^2}f_2(t_2)\\
&* \ (S_1(t_1)^{-\sigma_1^2}+S_2(t_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}-1}\\
&+ \ (1-\frac{\sigma_2}{\sigma_1}\rho)\frac{\sigma_1}{\sigma_2}\rho S_1(t_1)^{-\frac{\sigma_1}{\sigma_2}\rho-\sigma_1^2}f_1(t_1)S_2(t_2)^{-\frac{\sigma_2}{\sigma_1}\rho}f_2(t_2)\\
&* \ (S_1(t_1)^{-\sigma_1^2}+S_2(t_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}-1}\\
&+ \ \rho(\rho+\sigma_1\sigma_2)S_1(t_1)^{-\frac{\sigma_1}{\sigma_2}\rho-\sigma_1^2}f_1(t_1)S_2(t_2)^{-\frac{\sigma_2}{\sigma_1}\rho-\sigma_2^2}f_2(t_2)\\
&* \ (S_1(t_1)^{-\sigma_1^2}+S_2(t_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}-2}\\
&= \ \big(S_1(t_1)^{-\frac{\sigma_1}{\sigma_2}\rho}f_1(t_1)S_2(t_2)^{-\frac{\sigma_2}{\sigma_1}\rho}f_2(t_2)(S_1(t_1)^{-\sigma_1^2}+S_2(t_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}-2}\big)\\
&\quad \big((1-\frac{\sigma_1}{\sigma_2}\rho)(1-\frac{\sigma_2}{\sigma_1}\rho)(S_1(t_1)^{-\sigma_1^2}+S_2(t_2)^{-\sigma_2^2}-1)^2\\
&+ \ (1-\frac{\sigma_1}{\sigma_2}\rho)\frac{\sigma_2}{\sigma_1}\rho S_2(t_2)^{-\sigma_2^2}(S_1(t_1)^{-\sigma_1^2}+S_2(t_2)^{-\sigma_2^2}-1)\\
&+ \ (1-\frac{\sigma_2}{\sigma_1}\rho)\frac{\sigma_1}{\sigma_2}\rho S_1(t_1)^{-\sigma_1^2}(S_1(t_1)^{-\sigma_1^2}+S_2(t_2)^{-\sigma_2^2}-1)\\
&+ \ \rho(\rho+\sigma_1\sigma_2)S_1(t_1)^{-\sigma_1^2}S_2(t_2)^{-\sigma_2^2}\big)
\end{aligned}
$$

Appendix D

Correlated compound Poisson frailty model

$$S(t_1,t_2) = S(t_1)^{1-\rho}S(t_2)^{1-\rho}$$
$$* \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\}$$

$$S_{t_1}(t_1,t_2) = -(1-\rho)f(t_1)S(t_1)^{-\rho}S(t_2)^{1-\rho}$$
$$* \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\}$$
$$-\rho f(t_1)S(t_1)^{-\rho}S(t_2)^{1-\rho}(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma-1}$$
$$* \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\}$$
$$* ((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma-1}$$

$$S_{t_2}(t_1,t_2) = -(1-\rho)f(t_2)S(t_1)^{1-\rho}S(t_2)^{-\rho}$$
$$* \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\}$$
$$-\rho f(t_2)S(t_1)^{1-\rho}S(t_2)^{-\rho}(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma-1}$$
$$* \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\}$$
$$* ((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma-1}$$

$$
\begin{aligned}
S_{t_1t_2}(t_1,t_2) &= (1-\rho)^2 f(t_1)S(t_1)^{-\rho} f(t_2)S(t_2)^{-\rho} \\
&\quad * \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\} \\
&\quad + \rho(1-\rho) f(t_1)S(t_1)^{-\rho} f(t_2)S(t_2)^{-\rho}(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma-1} \\
&\quad * \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\} \\
&\quad * ((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma-1} \\
&\quad + \rho(1-\rho) f(t_1)S(t_1)^{-\rho} f(t_2)S(t_2)^{-\rho}(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma-1} \\
&\quad * \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\} \\
&\quad * ((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma-1} \\
&\quad + \rho^2 f(t_1)S(t_1)^{-\rho} f(t_2)S(t_2)^{-\rho} \\
&\quad * \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\} \\
&\quad * ((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{2\gamma-2} \\
&\quad * (1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma-1}(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma-1} \\
&\quad + \rho\sigma^2 f(t_1)S(t_1)^{-\rho} f(t_2)S(t_2)^{-\rho} \\
&\quad * \exp\{\frac{\rho(1-\gamma)}{\gamma\sigma^2}\big[1-((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma}\big]\} \\
&\quad * ((1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma}+(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma}-1)^{\gamma-2} \\
&\quad * (1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1)))^{1/\gamma-1}(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2)))^{1/\gamma-1}
\end{aligned}
$$

Appendix E

Bivariate lifetime models with censoring

As mentioned in the introduction, a unique feature of survival data is censoring. The present section deals with the problem of censoring in bivariate failure data. For this purpose, let $(T_1, T_2, \Delta_1, \Delta_2)$ be the censored observations with

$$T_i = \min\{X_i, Y_i\}, \quad \Delta_i = 1(X_i \leq Y_i), \quad i = 1, 2$$

and denote the bivariate survival function of (T_1, T_2) by $S(t_1, t_2)$. Furthermore, denote the density of (X_1, X_2) and (Y_1, Y_2), respectively, by $f(x_1, x_2)$ and $g(y_1, y_2)$. Assume that the lifetimes (X_1, X_2) and censoring times (Y_1, Y_2) are independent.

$$\begin{aligned} H(t_1, t_2, 1, 1) &= P(T_1 > t_1, T_2 > t_2, \Delta_1 = 1, \Delta_2 = 1) \\ &= P(X_1 > t_1, X_2 > t_2, X_1 \leq Y_1, X_2 \leq Y_2) \\ &= \int\int\int\int_{\{x_1 > t_1, x_2 > t_2, x_1 \leq y_1, x_2 \leq y_2\}} f(x_1, x_2) g(y_1, y_2)\, dx_1 dx_2 dy_1 dy_2 \\ &= \int_{t_2}^{\infty} \int_{t_1}^{\infty} f(x_1, x_2) \Big(\int_{x_2}^{\infty} \int_{x_1}^{\infty} g(y_1, y_2)\, dy_1 dy_2 \Big)\, dx_1 dx_2 \end{aligned}$$

Consequently, the sub-density for non-censored pairs of lifetimes is

$$h(t_1, t_2, 1, 1) = \Lambda_{t_1, t_2}(t_1, t_2, 1, 1) = f(t_1, t_2) \int_{t_2}^{\infty} \int_{t_1}^{\infty} g(y_1, y_2)\, dy_1 dy_2.$$

Analogously, the calculation of the other sub-densities is:

$$\begin{aligned} H(t_1, t_2, 1, 0) &= P(T_1 > t_1, T_2 > t_2, \Delta_1 = 1, \Delta_2 = 0) \\ &= P(X_1 > t_1, Y_2 > y_2, X_1 \leq Y_1, X_2 > Y_2) \\ &= \int\int\int\int_{\{x_1 > t_1, y_2 > t_2, x_1 \leq y_1, x_2 > y_2\}} f(x_1, x_2) g(y_1, y_2)\, dx_1 dx_2 dy_1 dy_2 \\ &= \int_{t_2}^{\infty} \int_{t_1}^{\infty} \Big(\int_{y_2}^{\infty} f(x_1, x_2)\, dx_2 \int_{x_1}^{\infty} g(y_1, y_2)\, dy_1 \Big) dx_1 dy_2 \end{aligned}$$

Consequently,

$$h(t_1, t_2, 1, 0) = \Lambda_{t_1,t_2}(t_1, t_2, 1, 0) = -S_{t_1}(t_1, t_2) \int_{t_1}^{\infty} g(y_1, t_2)\, dy_1.$$

$$\begin{aligned} H(t_1, t_2, 0, 1) &= P(T_1 > t_1, T_2 > t_2, \Delta_1 = 0, \Delta_2 = 1) \\ &= P(Y_1 > t_1, X_2 > t_2, X_1 > Y_1, X_2 \le Y_2) \\ &= \int\int\int\int_{\{y_1 > t_1, x_2 > t_2, x_1 > y_1, x_2 \le y_2\}} f(x_1, x_2) g(y_1, y_2)\, dx_1 dx_2 dy_1 dy_2 \\ &= \int_{t_2}^{\infty} \int_{t_1}^{\infty} \Big(\int_{y_1}^{\infty} f(x_1, x_2)\, dx_1 \int_{x_2}^{\infty} g(y_1, y_2)\, dy_2 \Big)\, dx_2 dy_1 \end{aligned}$$

Hence,

$$h(t_1, t_2, 0, 1) = \Lambda_{t_1,t_2}(t_1, t_2, 0, 1) = -S_{t_2}(t_1, t_2) \int_{t_2}^{\infty} g(t_1, y_2)\, dy_2.$$

$$\begin{aligned} H(t_1, t_2, 0, 0) &= P(T_1 > t_1, T_2 > t_2, \Delta_1 = 0, \Delta_2 = 0) \\ &= P(Y_1 > t_1, Y_2 > t_2, X_1 > Y_1, X_2 > Y_2) \\ &= \int\int\int\int_{\{y_1 > t_1, y_2 > t_2, x_1 > y_1, x_2 > y_2\}} f(x_1, x_2) g(y_1, y_2)\, dx_1 dx_2 dy_1 dy_2 \\ &= \int_{t_1}^{\infty} \int_{t_2}^{\infty} \Big(\int_{y_1}^{\infty} \int_{y_2}^{\infty} f(x_1, x_2)\, dx_1 dx_2 \Big) g(y_1, y_2)\, dy_1 dy_2 \end{aligned}$$

Consequently,

$$h(t_1, t_2, 0, 0) = \Lambda_{t_1,t_2}(t_1, t_2, 0, 1) = S(t_1, t_2) g(t_1, t_2).$$

Hence, the likelihood function under independent censoring is given by

$$\begin{aligned} L(t_1, t_2, \delta_1, \delta_2) &= \delta_1 \delta_2 h(t_1, t_2, 1, 1) + \delta_1 (1 - \delta_2) h(t_1, t_2, 1, 0) \\ &\quad + (1 - \delta_1) \delta_2 h(t_1, t_2, 0, 1) + (1 - \delta_1)(1 - \delta_2) h(t_1, t_2, 0, 0) \\ &= \delta_1 \delta_2 S_{t_1 t_2}(t_1, t_2) - \delta_1 (1 - \delta_2) S_{t_1}(t_1, t_2) \\ &\quad - (1 - \delta_1) \delta_2 S_{t_2}(t_1, t_2) + (1 - \delta_1)(1 - \delta_2) S(t_1, t_2). \end{aligned}$$

Appendix F

Dependent competing risks model

The following relations are used in the calculations: $\mathbf{E}Z_1 = \mathbf{E}Z_2 = \mathbf{E}Z_3 = \mathbf{E}Z_4 = 1$, $\mathbf{V}(Z_1) = \mathbf{V}(Z_3) = \frac{1}{k_1+k_2+k_5} = \sigma_1^2$, $\mathbf{V}(Z_2) = \mathbf{V}(Z_4) = \frac{1}{k_1+k_3+k_4} = \sigma_2^2$.
The second moment of V_2 is needed to calculate the mixed moment $\mathbf{E}Z_1Z_3$:

$$\mathbf{E}V_2^2 = \mathbf{V}(V_2) + (\mathbf{E}V_2)^2 = \frac{k_2}{\lambda_1^2} + \big(\frac{k_2}{\lambda_1}\big)^2 = \frac{k_2^2 + k_2}{\lambda_1^2}$$

Hence,

$$\begin{aligned}
\mathbf{E}Z_1Z_3 &= \mathbf{E}(\frac{\lambda_0}{\lambda_1}V_1 + V_2 + V_5)(V_2 + V_6 + \frac{\lambda_0}{\lambda_1}V_8) \\
&= \mathbf{E}\big(\frac{\lambda_0}{\lambda_1}V_1V_2 + \frac{\lambda_0}{\lambda_1}V_1V_6 + \frac{\lambda_0^2}{\lambda_1^2}V_1V_8 \\
&+ V_2^2 + V_2V_6 + \frac{\lambda_0}{\lambda_1}V_2V_8 + V_2V_5 + V_5V_6 + \frac{\lambda_0}{\lambda_1}V_5V_8\big) \\
&= \frac{k_1k_2}{\lambda_1^2} + \frac{k_1k_5}{\lambda_1^2} + \frac{k_1^2}{\lambda_1^2} + \frac{k_2^2+k_2}{\lambda_1^2} + \frac{k_2k_5}{\lambda_1^2} + \frac{k_1k_2}{\lambda_1^2} + \frac{k_2k_5}{\lambda_1^2} + \frac{k_5^2}{\lambda_1^2} + \frac{k_1k_5}{\lambda_1^2} = \frac{k_2}{\lambda_1^2} + 1
\end{aligned}$$

and consequently

$$\mathbf{cov}(Z_1, Z_3) = \mathbf{E}Z_1Z_2 - \mathbf{E}Z_1\mathbf{E}Z_2 = \frac{k_2}{\lambda_1^2}.$$

Now we are able to derive the correlation:

$$\rho_1 = \frac{\mathbf{cov}(Z_1, Z_3)}{\sqrt{V(Z_1)V(Z_3)}} = \frac{k_2}{\lambda_1} = k_2\sigma_1^2 \tag{F.1}$$

Similar calculations imply $\rho_2 = k_3\sigma_2^2$ and $\rho = k_1\sigma_1\sigma_2$. Consequently, $k_1 + k_2 + k_5 = \frac{1}{\sigma_1^2}$ and $k_1 + k_3 + k_4 = \frac{1}{\sigma_2^2}$ imply the following relations:

$$k_5 = \frac{1}{\sigma_1^2} - k_2 - k_1 = \frac{1}{\sigma_1^2} - \frac{\rho_1}{\sigma_1^2} - \frac{\rho}{\sigma_1\sigma_2} \qquad \text{and} \qquad k_4 = \frac{1}{\sigma_2^2} - k_3 - k_4 = \frac{1}{\sigma_2^2} - \frac{\rho_2}{\sigma_2^2} - \frac{\rho}{\sigma_1\sigma_2}.$$

If $Y \sim \Gamma(k, \lambda)$, then $\mathbf{E}e^{-sY} = (1 + \frac{s}{\lambda})^{-k}$. Now we are able to derive the survival function in (4.37):

$$\begin{aligned}
&S(x_1, y_1, x_2, y_2) \\
&\quad = \mathbf{E}S_1(x_1)^{Z_1} S_2(y_1)^{Z_2} S_1(x_2)^{Z_3} S_2(y_2)^{Z_4} \\
&\quad = \mathbf{E}e^{-V_1(\frac{\lambda_0}{\lambda_1}H_1(x_1)+\frac{\lambda_0}{\lambda_2}H_2(y_1))} e^{-V_2(H_1(x_1)+H_1(x_2))} \\
&\quad * \ e^{-V_3(H_2(y_1)+H_2(y_2))} e^{-V_8(\frac{\lambda_0}{\lambda_1}H_1(x_2)+\frac{\lambda_0}{\lambda_2}H_2(y_2))} e^{-V_4H_2(y_1)} e^{-V_5H_1(x_1)} e^{-V_6H_1(x_2)} e^{-V_7H_2(y_2)} \\
&\quad = (1 + \frac{1}{\lambda_0}(\frac{\lambda_0}{\lambda_1}H_1(x_1) + \frac{\lambda_0}{\lambda_2}H_2(y_1)))^{-k_1}(1 + \frac{1}{\lambda_1}H_1(x_1) + \frac{1}{\lambda_1}H_1(x_2))^{-k_2} \\
&\quad * \ (1 + \frac{1}{\lambda_2}H_2(y_1) + \frac{1}{\lambda_2}H_2(y_2))^{-k_3}(1 + \frac{1}{\lambda_0}(\frac{\lambda_0}{\lambda_1}H_1(x_2) + \frac{\lambda_0}{\lambda_2}H_2(y_2)))^{-k_1} \\
&\quad * \ (1 + \frac{1}{\lambda_2}H_2(y_1))^{-k_4}(1 + \frac{1}{\lambda_1}H_1(x_1))^{-k_5}(1 + \frac{1}{\lambda_1}H_1(x_2))^{-k_5}(1 + \frac{1}{\lambda_2}H_2(y_2))^{-k_4} \\
&\quad = (S_1(x_1)^{-\sigma_1^2} + S_1(x_2)^{-\sigma_1^2} - 1)^{-\rho_1/\sigma_1^2}(S_2(y_1)^{-\sigma_2^2} + S_2(y_2)^{-\sigma_2^2} - 1)^{-\rho_2/\sigma_2^2} \\
&\quad * \ (S_1(x_1)^{-\sigma_1^2} + S_2(y_1)^{-\sigma_2^2} - 1)^{-\rho/\sigma_1\sigma_2}(S_1(x_2)^{-\sigma_1^2} + S_2(y_2)^{-\sigma_2^2} - 1)^{-\rho/\sigma_1\sigma_2} \\
&\quad * \ S_2(y_1)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} S_1(x_2)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}
\end{aligned}$$

The likelihood function (of the truncated data) takes the following form:

$$\begin{aligned}
L(t_1, t_2, \delta_1, \delta_2, t^*) \ = \ &\Big(1(\delta_1 = 1, \delta_2 = 1)S_{x_1x_2}(t_1, t_1, t_2, t_2) \\
&+ \ 1(\delta_1 = 1, \delta_2 = 0)S_{x_1}(t_1, t_1, t_2, t_2) \\
&+ \ 1(\delta_1 = 0, \delta_2 = 1)S_{x_2}(t_1, t_1, t_2, t_2) \\
&+ \ 1(\delta_1 = 0, \delta_2 = 0)S(t_1, t_1, t_2, t_2) \\
&+ \ 1(\delta_1 = -1, \delta_2 = -1)S_{y_1y_2}(t_1, t_1, t_2, t_2) \\
&+ \ 1(\delta_1 = -1, \delta_2 = 0)S_{y_1}(t_1, t_1, t_2, t_2) \\
&+ \ 1(\delta_1 = 0, \delta_2 = -1)S_{y_2}(t_1, t_1, t_2, t_2) \\
&+ \ 1(\delta_1 = 1, \delta_2 = -1)S_{x_1y_2}(t_1, t_1, t_2, t_2) \\
&+ \ 1(\delta_1 = -1, \delta_2 = 1)S_{y_1x_2}(t_1, t_1, t_2, t_2)\Big)/S(t^*, t^*, t^*, t^*)
\end{aligned}$$

The derivatives are given by

$$
\begin{aligned}
S_{x_1}(x_1,y_1,x_2,y_2) &= -\rho_1(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}-1} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho-\sigma_1^2}\mu_1(x_1) \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&- \ \frac{\sigma_1}{\sigma_2}\rho(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}-1} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho-\sigma_1^2}\mu_1(x_1) \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&- \ (1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)\sigma_1^2(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}\mu_1(x_1)S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}
\end{aligned}
$$

$$
\begin{aligned}
S_{y_1}(x_1,y_1,x_2,y_2) &= -\rho_1(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}-1} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho-\sigma_1^2}\mu_1(y_1) \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&- \ \frac{\sigma_1}{\sigma_2}\rho(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}-1} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho-\sigma_1^2}\mu_1(y_1) \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&- \ (1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)\sigma_1^2(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}\mu_1(y_1) \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}
\end{aligned}
$$

$$
\begin{aligned}
S_{x_2}(x_1,y_1,x_2,y_2) &= -\rho_2(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}-1} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho-\sigma_2^2}\mu_2(x_2) \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&- \ \frac{\sigma_2}{\sigma_1}\rho(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}-1} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho-\sigma_2^2}\mu_2(x_2) \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&- \ (1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)\sigma_2^2(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}\mu_2(x_2) \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}
\end{aligned}
$$

$$
\begin{aligned}
S_{y_2}(x_1,y_1,x_2,y_2) &= -\rho_2(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}-1} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho-\sigma_2^2}\mu_2(y_2) \\
&- \ \frac{\sigma_1}{\sigma_2}\rho(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}-1} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho-\sigma_2^2}\mu_2(y_2) \\
&- \ (1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)\sigma_2^2(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\frac{\rho_1}{\sigma_1^2}} \\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho_2}{\sigma_2^2}} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ (S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\frac{\rho}{\sigma_1\sigma_2}} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho} \\
&* \ S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}\mu_2(y_2)
\end{aligned}
$$

$$
\begin{aligned}
S_{x_1,y_1}(x_1,y_1,x_2,y_2) &= \mu_1(x_1)\mu_1(y_1)(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\rho_1/\sigma_1^2-2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\rho_2/\sigma_2^2} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\rho/\sigma_1\sigma_2-1}(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\rho/\sigma_1\sigma_2-1} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ \Big[\rho^2\frac{\sigma_1^2}{\sigma_2^2}S_1(x_1)^{-\sigma_1^2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^2 \\
&+ \ \rho\rho_1\frac{\sigma_1}{\sigma_2}S_1(x_1)^{-\sigma_1^2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1) \\
&+ \ \rho(1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)\frac{\sigma_1}{\sigma_2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^2(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1) \\
&+ \ \rho\rho_1\frac{\sigma_1}{\sigma_2}S_1(x_1)^{-\sigma_1^2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ \rho(1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)\frac{\sigma_1}{\sigma_2}S_1(x_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^2(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ \rho_1(\rho_1+\sigma_1^2)S_1(x_1)^{-\sigma_1^2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ \rho_1(1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)S_1(x_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1) \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ \rho_1(1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1) \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ (1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)^2(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^2(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\Big]
\end{aligned}
$$

$$
\begin{aligned}
S_{x_1,y_2}(x_1,y_1,x_2,y_2) &= \mu_1(x_1)\mu_2(y_2)(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\rho_1/\sigma_1^2-1}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\rho_2/\sigma_2^2-1} \\
&* (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\rho/\sigma_1\sigma_2-1}(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\rho/\sigma_1\sigma_2-1} \\
&* S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \Big[\rho\rho_2\frac{\sigma_1}{\sigma_2}S_1(x_1)^{-\sigma_1^2}S_2(y_2)^{-\sigma_2^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \rho_1\rho_2 S_1(x_1)^{-\sigma_1^2}S_2(y_2)^{-\sigma_2^2}(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \rho_2(1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)S_2(y_2)^{-\sigma_2^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \rho^2 S_1(x_1)^{-\sigma_1^2}S_2(y_2)^{-\sigma_2^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \rho\rho_1\frac{\sigma_2}{\sigma_1}S_1(x_1)^{-\sigma_1^2}S_2(y_2)^{-\sigma_2^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1) \\
&+ \rho(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)\frac{\sigma_1}{\sigma_2}S_1(x_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1) \\
&* (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \rho(1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)\frac{\sigma_2}{\sigma_1}S_2(y_2)^{-\sigma_2^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1) \\
&* (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1) \\
&+ \rho_1(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)S_1(x_1)^{-\sigma_1^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&* (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ (1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&* (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\Big]
\end{aligned}
$$

$$
\begin{aligned}
S_{x_2,y_1}(x_1,y_1,x_2,y_2) &= \mu_2(x_2)\mu_1(y_1)(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\rho_1/\sigma_1^2-1}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\rho_2/\sigma_2^2-1}\\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\rho/\sigma_1\sigma_2-1}(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\rho/\sigma_1\sigma_2-1}\\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}\\
&* \ \Big[\rho\rho_2\frac{\sigma_1}{\sigma_2}S_2(x_2)^{-\sigma_2^2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)\\
&+ \ \rho_1\rho_2S_2(x_2)^{-\sigma_2^2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\\
&+ \ \rho_2(1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)S_2(x_2)^{-\sigma_2^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)\\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\\
&+ \ \rho^2S_2(x_2)^{-\sigma_2^2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)\\
&+ \ \rho(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)\frac{\sigma_1}{\sigma_2}S_1(y_1)^{-\sigma_1^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)\\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)\\
&+ \ \rho\rho_1\frac{\sigma_2}{\sigma_1}S_2(x_2)^{-\sigma_2^2}S_1(y_1)^{-\sigma_1^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\\
&+ \ \rho(1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)\frac{\sigma_2}{\sigma_1}S_2(x_2)^{-\sigma_2^2}(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)\\
&* \ (S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\\
&+ \ \rho_1(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)S_1(y_1)^{-\sigma_1^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\\
&+ \ (1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho)(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)\\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\Big]
\end{aligned}
$$

$$
\begin{aligned}
S_{x_2,y_2}(x_1,y_1,x_2,y_2) &= \mu_2(x_2)\mu_2(y_2)(S_1(x_1)^{-\sigma_1^2}+S_1(y_1)^{-\sigma_1^2}-1)^{-\rho_1/\sigma_1^2}S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\rho_2/\sigma_2^2-2} \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)^{-\rho/\sigma_1\sigma_2-1}(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)^{-\rho/\sigma_1\sigma_2-1} \\
&* \ S_1(x_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_1(y_1)^{1-\rho_1-\frac{\sigma_1}{\sigma_2}\rho}S_2(x_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho}S_2(y_2)^{1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho} \\
&* \ \Big[\rho^2\frac{\sigma_2^2}{\sigma_1^2}S_2(x_2)^{-\sigma_2^2}S_2(y_2)^{-\sigma_2^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^2 \\
&+ \ \rho\rho_2\frac{\sigma_2}{\sigma_1}S_2(x_2)^{-\sigma_2^2}S_2(y_2)^{-\sigma_2^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1) \\
&+ \ \rho(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)\frac{\sigma_2}{\sigma_1}S_2(y_2)^{-\sigma_2^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^2(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1) \\
&+ \ \rho\rho_2\frac{\sigma_2}{\sigma_1}S_2(x_2)^{-\sigma_2^2}S_2(y_2)^{-\sigma_2^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ \rho(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)\frac{\sigma_2}{\sigma_1}S_2(x_2)^{-\sigma_2^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^2(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ \rho_2(\sigma_2^2+\rho_2)S_2(x_2)^{-\sigma_2^2}S_2(y_2)^{-\sigma_2^2}(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ \rho_2(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)S_2(x_2)^{-\sigma_2^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ \rho_2(1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)S_2(y_2)^{-\sigma_2^2}(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&* \ (S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1) \\
&+ \ (1-\rho_2-\frac{\sigma_2}{\sigma_1}\rho)^2(S_2(x_2)^{-\sigma_2^2}+S_2(y_2)^{-\sigma_2^2}-1)^2(S_1(x_1)^{-\sigma_1^2}+S_2(x_2)^{-\sigma_2^2}-1)(S_1(y_1)^{-\sigma_1^2}+S_2(y_2)^{-\sigma_2^2}-1)\Big]
\end{aligned}
$$

Bibliography

[1] Aalen, O.O. (1987) Two examples of modelling heterogeneity in survival analysis. Scandinavian Journal of Statistics 14, 19 - 25

[2] Aalen, O.O. (1988) Heterogeneity in survival analysis. Statistics in Medicine 7, 1121-1137

[3] Aalen, O.O. (1992) Modelling heterogeneity in survival analysis by the compound Poisson distribution. Annals of Applied Probability 4, 951 - 972

[4] Aalen, O.O., Tretli, S. (1999) Analyzing incidence of testis cancer by means of a frailty model. Cancer Causes and Control 10, 285 - 292

[5] Aalen, O., Hjort, N.L. (2002) Frailty models that yield proportional hazards. Statistics & Probability Letters 58, 335 - 342

[6] Abbring, J., van den Berg, G.J. (2005) The unobserved heterogeneity distribution in duration analysis. Working Paper, Free University Amsterdam

[7] Andersen, P.K., Gill, R.D. (1982) Cox's regression model for counting processes: a large sample study. Annals of Statistics 10, 1100 - 1120

[8] Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N. (1993) Statistical models based on counting processes. Springer Verlag, New York

[9] Andersen, P.K., Klein, J.P., Knudsen, K.M., y Palacios, R.T. (1997) Estimation of variance in Cox's regression model with shared gamma frailties. Biometrics 53, 1465 - 1475

[10] Andersen, P.K., Klein, J.P., Zhang, M.-J. (1999) Testing for centre effects in multi-centre survival studies: a Monte Carlo comparison of fixed and random effects Tests. Statistics in Medicine 18, 1489 - 1500

[11] Anderson, J.E., Louis, T.A. (1995) Survival analysis using a scale change random effects model. Journal of the American Statistical Association 90, 669 - 679

[12] Bandeen-Roche, K.J., Liang, K.-Y. (1996) Modelling failure-time associations in data with multiple levels of clustering. Biometrika 83, 29 - 39

[13] Barker, P., Henderson, R. (2004) Modelling converging hazards in survival analysis. Lifetime Data Analysis 10, 263 - 281

[14] Bandyopadhyay, D., Basu, A.P. (1990) On a generalization of a model by Lindley and Singpurwalla. Advances in Applied Probability 22, 498 - 500

[15] Bellamy, S.L., Li, Y., Ryan, L.M., Lipsitz, S., Canner, M.J., Wright, R. (2004) Analysis of clustered and interval censored data from a community-based study in asthma. Statistics in Medicine 23, 3607 - 2621

[16] Besag, J., Green, P.J. (1993) Spatial statistics and Bayesian computation. Journal of the Royal Statistical Society (B) 55, 25 - 37

[17] Blumen, I., Kogan, M., McCarthy, P.J. (1955) The industrial mobility of labor as a probability process. Ithaca, New York: Cornell University Press

[18] Blossfeld, H.P., Hamerle, A. (1989) Unobserved heterogeneity in hazard rate models - A test and an illustration from a study of career mobility. Quality & Quantity 23, 129 - 141

[19] Blossfeld, H.P., Hamerle, A. (1992) Unobserved heterogeneity in event history models. Quality & Quantity 26, 157 - 168

[20] Bolstad, W.M., Manda, S.O. (2001) Investigating child mortality in Malawi using family and community random effects: a Bayesian analysis. Journal of the American Statistical Association 96, 12 - 19

[21] Bordes, L., Commenges, D. (2001) Asymptotics for homogeneity tests based on a multivariate random effects proportional hazards model. Journal of Multivariate Analysis 78, 83 - 102

[22] Bouchard, C., (1994) Genetics of obesity: Overview and research directions. In: Claude Bouchard (ed.) The genetics of obesity. 1st edn. CRC Press, Boca Raton, pp. 223 - 233

[23] Bretagnolle, J., Huber-Carol, C. (1988) Effects of omitting covariates in Cox's model for survival data. Scandinavian Journal of Statistics 15, 125 - 138

[24] Broët, P., Moreau, T., Lellouch, J., Asselain, B. (1999) Unobserved covariates in the two-sample comparison of survival times: a maxmin efficiency robust test. Statistics in Medicine 18, 1791-1800

[25] Brooks S.P., Gelman A. (1998) Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434 - 455

[26] Cai, J., Prentice, R.L. (1995) Estimating equations for hazard ratio parameters based on correlated failure time data. Biometrika 82, 151 - 164

[27] Carvalho, M., Henderson, R., Shimakura, S., Sousa, I.P.S.C. (2003) Survival of hemodialysis patients: modeling differences in risk of dialysis centers. International Journal for Quality in Health Care 15, 189 - 196

[28] 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

[29] Chamberlain, G. (1992) Heterogeneity, omitted variable bias, and duration dependence. In: J. Heckman, B. Singer (eds.) Longitudinal analysis of labour market data, pp. 3 - 38, Cambridge University Press.

[30] Chang, S.-H. (2004) Estimating marginal effects in accelerated failure time models for serial sojurn times among repeated events. Lifetime Data Analysis 10, 175 - 190

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

[32] Christensen, K., Vaupel, J., Holm, N., Yashin, A. (1995) Mortality among twins after age 6: fetal origins hypothesis versus twin method. British Medical Journal 310, 432-436

[33] Christensen, K., Wienke, A., Skytthe, A., Holm, N., Vaupel, J., Yashin, A. (2001) Cardiovascular mortality in twins and the fetal origins hypothesis. Twin Research 4, 344-49

[34] 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 65, 141 - 151

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

[36] Clayton, D., Kaldor, J.M. (1985) Heterogeneity models as an alternative to proportional hazards in cohort studies. Bulletin of the International Statistical Institute 51, 3.2 1 - 16

[37] Clayton, D. (1991) A Monte Carlo method for Bayesian inference in frailty models. Biometrics 47, 467 - 485

[38] Commenges, D., Andersen P.K. (1995) Score test of homogeneity for survival data. Lifetime Data Analysis 1, 145 - 160

[39] 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, 157-71

[40] Congdon, P. (1995) Modelling frailty in area mortality. Statistics in Medicine 14, 1859-74

[41] Cook, R.J., Ng, E.T.M., Mukherjee, J., Vaughan, D. (1999) Two-state mixed renewal processes for chronic disease. Statistics in Medicine 18, 175 - 188

[42] Cook, R.D., Johnson, M.E. (1981) A family of distributions for modeling non-elliptically symmetric multivariate data. The Journal of the Royal Statistical Society (B) 43, 210 - 18

[43] Cox, D.R. (1972) Regression models and life-tables. The Journal of the Royal Statistical Society (B) 34, 187 - 220

[44] Cox, D.R. (1975) Partial likelihood. Biometrika 62, 269 - 276

[45] Cox, D.R., Oakes, D. (1984) Analysis of survival data. Chapman & Hall, London

[46] Cui, S., Sun, Y., (2004) Checking for the gamma frailty distribution under the marginal proportional hazards frailty model. Statistica Sinica 14, 249 - 267

[47] de Faire, U., Friberg, L., Lundman, T. (1975) Concordance for mortality with special reference to ischaemic heart disease and cerebrovascular disease: a study on the Swedish Twin Registry. Preventive Medicine 4, 509 - 517

[48] di Serio, C. (1997) The protective impact of a vovariate on competing failures with an example from bone marrow transplantation. Lifetime Data Analysis 3, 99 - 122

[49] dos Santos, D.M., Davies, R.B., Francis, B. (1995) Nonparametric hazard versus nonparametric frailty distribution in modelling recurrence of breast cancer. Journal of Statistical Planning and Inference 47, 111 - 127

[50] Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R., Sylvester, R. (2002) The shared frailty model and the power for heterogeneity tests in multicenter trials. Computational Statistics & Data Analysis 40, 603 - 620

[51] Duchateau, L., Janssen, P., Kezic, I., Fortpied, C. (2003) Evolution of recurrent asthma event rate over time in frailty models. Journal of the Royal Statistical Society (B) 52, 355 - 363

[52] Duchateau, L., Janssen, P. (2004) Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. Biometrics 60, 608 - 614

[53] Dunson, D.B., Chen, Z. (2004) Selecting factors predictive of heterogeneity in multivariate event time data. Biometrics 60, 352 - 358

[54] Elbers, C., Ridder, G. (1982) True and spurious duration dependence: The identifiability of the proportional hazard model. Review of Economic Studies XLIX, 403 - 409

[55] Emoto, S.E., Matthews, P.C. (1990) A Weibull model for dependent censoring. The Annals of Statistics 18, 1556 - 1577

[56] Falconer, D.S. (1990) Introduction to quantitative genetics. Longman Group, New York

[57] Farewell, V.T. (1977) A model for a binary variable with time-censored observations. Biometrika 64, 43 - 46

[58] Farewell, V.T., Math, B., Math, M. (1977) The combined effect of breast cancer risk factors. Cancer 40, 931 - 936

[59] Farewell, V.T. (1982) The use of mixture models for the analysis of survival data with long-term survivors. Biometrics 38, 1041 - 1046

[60] Feldman, M., Lewontin, R. (1975) The heritability hang-up. Science 190, 1163 - 1168

[61] Feuer, E.J., Wun, L.-M., Boring, C.C., Flanders, W.D., Timmel, M.J., Tong, T. (1993) The lifetime risk of developing breast cancer. Journal of the National Cancer Institute 85, 892 - 897

[62] Fine, J.P., Glidden, D.V., Lee, K.E. (2003) A simple estimator for a shared frailty regression model. Journal of the Royal Statistical Society (B) 65, 317 - 329

[63] Fisher R.A. (1958) Cancer and smoking. Nature 182, 596

[64] Fleming, T., Harrington, D. (1991) Counting processes and survival analysis. Wiley, Chichester

[65] Flinn, C., Heckman, J. (1982) New methods for analyzing structural models of labor force dynamics. Journal of Econometrics 18, 115 - 168

[66] Gail, M.H., Wieand, S., Piantadosi, S. (1984) Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika 71, 431 - 444

[67] Galton, F. (1887) Natural inheritance. Macmillan, London

[68] Gatz, M., Pedersen, N.L., Crowe, M., Fiske, A. (2000) Defining discordance in twin studies of risk and protective factors for late life disorders. Twin Research 3, 159 - 64

[69] Gelman, A., Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7, 457 - 511

[70] Geman, S., Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restauration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721 - 741

[71] Genest, C., MacKay, J. (1986) The joy of copulas: bivariate distributions with uniform marginals. The American Statistician 40, 280 - 283

[72] Giard, N. (2001) A four-dimensional correlated frailty model. GCA-Verlag, Herdecke

[73] Giard, N., Lichtenstein, P., Yashin, A.I. (2002) A multistate model for the genetic analysis of the ageing process. Statistics in Medicine 21, 2511 - 2526

[74] Gilks, W.R., Richardson, S., Spiegelhalter, D.G. (1996) Markov Chain Monte Carlo in practice. Chapman and Hall, London

[75] Gjessing, H.K., Aalen, O.O., Hjort, N.L. (2003) Frailty models based on Lévy processes. Advances in Applied Probability 35, 532 - 550

[76] Glidden, D.V. (1999) Checking the adequacy of the gamma frailty model for multivariate failure times. Biometrika 86, 381 - 393

[77] Gompertz, B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philiosphical Transactions of the Royal Society of London 115, 513 - 585

[78] Gray, R.J. (1995) Tests for variation over groups in survival data. Journal of the American Statistical Association 90, 198 - 203

[79] Greenwood, M., Yule, G.U. (1920) An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurence of multiple attacks of disease or of repeated accidents. Journal of the Royal Statistical Society 83, 255 - 279

[80] Gustafson, P. (1997) Large hierarchical Baysian analysis of multivariate survival data. Biometrics 53, 230 - 242

[81] Guo, G., Rodriguez, G. (1992) Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala. Journal of the American Statistical Association 87, 969 - 976

[82] Guo, G. (1993) Use of sibling data to estimate family mortality effects in Guatemala. Demography 30, 15 - 32

[83] Gupta, P. L., Gupta, R.D. (1990) A bivariate random environmental stress model. Advances in Applied Probability 22, 501 - 503

[84] Hamerle, A. (1992) A test for detecting overdispersion in multiple spell failure time models. Biometrical Journal 34, 81 - 90

[85] Hastings, W.K. (1970) Monte Carlo sampling methods using Markov Chains and their applications. Biometrika 57, 97 - 109

[86] Harvald, B., Hauge, M. (1970) Coronary occlusion in twins. Acta Geneticae Medicae et Gemellologiae 19, 248 - 250

[87] Harris, J.R., Lippman, M.E., Veronesi, U., Willett, W. (1992) Breast cancer (i). New England Journal of Medicine 327, 319 - 328

[88] Hauge, M. (1981) The Danish twin register. In: Prospective longitudinal research. Mednich, S., Baert, A., Bachmann, B. (eds.) Oxford Medical Publisher, Oxford, pp. 217 - 222

[89] Haukka, J., Suvisaari, J., Lönnqvist, J. (2003) Increasing age does not decrease risk of schizophrenia up to age 40. Schizophrenia Research 61, 105 - 110

[90] Heath, A.C., Madden, P.A.F. (1995) Genetic influences on smoking behavior. In: Turner, J.R., Cardon, L.R., Hewitt, J.K., (eds.) Behavior: Genetic approaches in behavioral medicine. Plenum Press, New York, London, pp. 45 - 63

[91] Heckman, J.J., Honoré B.E. (1989) The identifiability of the competing risks model. Biometrika 76, 325 - 330

[92] Heckman, J.J., Singer, B. (1982a) The identification problem in econometric models for duration data. In: Advances in Econometrics (ed. W. Hildenbrandt), Cambridge, 39 - 77

[93] Heckman, J.J., Singer, B. (1982b) Population heterogeneity in demographic models. In K. Land and A. Rogers (eds.) Multidimensional Mathematical Demography. New York, Academic Press

[94] Heckman, J.J., Singer, B. (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52, 271 - 320

[95] Heckman, J.J., Walker J.R. (1990) Estimating fecundability from data on waiting-times to 1st conception. Journal of the American Statistical Association 85, 283 - 294

[96] Henderson, R., Oman, P. (1999) Effect of frailty on marginal regression estimates in survival analysis. The Journal of the Royal Statistical Society (B) 61, 367 - 379

[97] Henderson, R., Shimakura, S. (2003) A serially correlated gamma frailty model for longitudinal count data. Biometrika 90, 355 - 366

[98] Herskind, A.M., McGue, M., Iachine, I.A., Holm, N., Sørensen, T.I.A., Harvald, B., Vaupel, J.W. (1996a) Untangling genetic influences on smoking, body mass index and longevity: a multivariate study of 2464 Danish twins followed for 28 years. Human Genetics 98, 467 - 475

[99] Herskind, A.M., McGue, M., Sørensen, T.I.A., Harvald, B. (1996b) Sex and age specific assessment of genetic and environmental influences on body mass index in twins. International Journal of Obesity 20, 106 - 113

[100] Hoem, J.M. (1990) Identifiability in hazard models with unobserved heterogeneity: the compatibility of two apparently contradictionary results. Theoretical Population Biology 37, 124 -128

[101] Holm, N.V. (1983) The use of twin studies to investigate causes of diseases with complex etiology with a focus on cancer. Ph.D. thesis (in Danish), Odense University, Odense

[102] Hopper, J.L. (1993) Variance components for statistical genetics: applications in medical research to characteristics related to human diseases and health. Statistical Methods in Medical Research 2; 199 - 223

[103] Hougaard, P. (1984) Life table methods for heterogeneous populations. Biometrika 71, 75 - 83

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

[105] Hougaard, P. (1986b) A class of multivariate failure time distributions. Biometrika 73, 671 - 678

[106] Hougaard, P. (1987) Modeling multivariate survival. Scandinavian Journal of Statistics 14, 291-304

[107] Hougaard, P. (1991) Modeling heterogeneity in survival data. J. of Applied Probability 28, 695-701

[108] Hougaard, P., Harvald, B., Holm, N.V. (1992) Measuring the similarities between the lifetimes of adult Danish twins born 1881 - 1930. Journal of the American Statistical Association 87, 17 - 24

[109] Hougaard, P., Myglegaard, P., Borch-Johnsen, K. (1994) Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. Biometrics 50, 1178 - 1188

[110] Hougaard, P. (1995) Frailty models for survival data. Lifetime Data Analysis 1, 255 - 273

[111] Hougaard, P. (2000) Analysis of multivariate survival data. Springer, New York

[112] Huang, X., Wolfe, R.A. (2002) A frailty model for informative censoring. Biometrics 58, 510 - 520

[113] Huang, X., Wolfe, R.A., Hu, C. (2004) A test for informative censoring in clustered survival data. Statistics in Medicine 23, 2089 - 2107

[114] Hutchinson, T.P., Lai, C.D. (1991) The engineering statistician's guide to continuous bivariate distributions. Rumsby, Adelaide.

[115] Iachine, I. (2002) The Use of Twin and Family Survival Data in the Population Studies of Aging: Statistical Methods Based on Multivariate Survival Models. Ph.D. Thesis. Monograph 8, Department of Statistics and Demography, University of Southern Denmark

[116] Iachine, I. (2004) Identifiability of bivariate frailty models. Preprint 5, Department of Statistics, University of Southern Denmark, Odense

[117] Iachine, I., Holm, N., Harris, J., Begun, A., Iachina, M., Laitinen, M., Kaprio, J., Yashin, A. (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, 196 - 205

[118] Iachine, I.A., Yashin, A.I. (1998) Identifiability of bivariate frailty models based on additive independent components. Research Report 8, Department of Statistics and Demography, University of Southern Denmark, Odense

[119] Jeong, J.-H. (2003) Efficiency of log-rank test under dependent censoring. Communications in Statistics, Theory and Methods 32, 1197 - 1211

[120] Jeong, J.-H., Jung, S.-H., Wieand, S. (2003) A parametric model for long-term follow-up data from phase III breast cancer clinical trials. Statistics in Medicine 22, 339 - 352

[121] Joe, H. (1993) Parametric families of multivariate distributions with given margins. Journal of Multivariate Analysis 46, 262 - 282

[122] Jones, B.L. (1998) A model for analyzing the impact of selective lapsation on mortality. North American Actuarial Journal 2, 79 - 86

[123] Juel, K., Helweg-Larsen, K. (1999) The Danish registers of causes of death. Danish Medical Bulletin 46, 354 - 357

[124] Kalbfleisch, J.D., Prentice, R.L. (1980) The statistical analysis of failure time data. John Wiley & Sons, New York

[125] Kaplan, E.L., Meier, P. (1958) Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457 - 481

[126] Keiding, N., Andersen, P., Klein, J. (1997) The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. Statistics in Medicine 16, 215 - 224

[127] Klein, J.P. (1992) Semiparametric estimation of random effects using the Cox model based on the EM algorithm. Biometrics 48, 795 - 806

[128] Klein, J.P., Moeschberger, M.L. (1997) Survival analysis - techniques for censored and truncated data. Springer, New York

[129] Klein, J.P., Moeschberger, M., Li, Y.H., Wang, S.T. (1992) Estimating random effects in the Framingham Heart Study. In: Survival Analysis : State of the Art, pp. 99 - 120

[130] Klein, J.P., Pelz, C., M.-J. Zhang (1999) Random effects for censored data by a multivariate normal regression model. Biometrics 55, 497 - 506

[131] Korsgaard, I.R., Andersen, A.H. (1998) The additive genetic gamma frailty model. Scandinavian Journal of Statistics 25, 255 - 270

[132] Korsgaard, I.R., Madsen, P., Jensen, J. (1998) Bayesian inference in the semiparametric log normal frailty model using Gibbs sampling. Genetics, Selection, Evolution 30, 241-256

[133] Kuk, A.Y.C., Chen, C.-H. (1992) A mixture model combining logistic regression with proportional hazards regression. Biometrika 79, 531 - 541

[134] Lancaster, T. (1979) Econometric methods for the duration of unemployment. Econometrica 47, 939 - 956

[135] Lam, K.F., Kuk, Y.C. (1997) A marginal likelihood approach to estimation in frailty models. Journal of the American Statistical Association 92, 985 - 990

[136] Lam, K.F., Lee, Y.W., Leung, T.L. (2002) Modeling multivariate survival data by a semiparametric random effects proportional odds model. Biometrics 58, 316 - 323

[137] Lam, K.F., Lee, Y.W. (2004) Merits of modelling multivariate survival data using random effects proportional odds model. Biometrical Journal 46, 331 - 342

[138] Lam, K.F., Fong, D.Y.T., Tang, O.Y. (2005) Estimating the proportion of cured patients in a censored sample. Statistics in Medicine 24, 1865 - 1879

[139] Lawless, J.F. (1982) Statistical models and methods for lifetime data. Wiley & Sons, New York

[140] Lee, S., Lee, S. (2003) Testing heterogeneity for frailty distribution in shared frailty model. Communications in Statistics, Theory and Methods 32, 2245 - 2253

[141] Lee, E.W., Wei, L.J., Amato, D.A. (1992) Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Survival analysis: state of the art. Klein J., Goel P. (eds). Dordrecht, Boston, London: Kluwer Academic, pp. 237-247

[142] Lee, S., Wolfe, R.A. (1998) A simple test for independent censoring under the proportional hazards model. Biometrics 54, 1176 - 1182

[143] Liang, K.-Y., Self, S., Bandeen-Roche, K., Zeger, S. (1995) Some recent developments for regression analysis of multivariate failure time data. Lifetime Data Analysis 1, 403 - 415

[144] 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

[145] Lillard, L.A. (1993) Simultaneous equations for hazards: marriage duration and fertility timing. Journal of Econometrics 56, 189 - 217

[146] Lillard, L.A., Brian, M.J., Waite, M.J. (1995) Premarital cohabitation and subsequent marital dissolution: a matter of self-selection? Demography 32, 437 - 457

[147] Lillard, L.A., Panis, C. (2000) aML user's guide and reference manual. Los Angeles: Econ-Ware

[148] Lin, D.Y. (1994) Cox regression analysis of multivariate failure time data: the marginal approach. Statistics in Medicine 13, 2233 - 2247

[149] Lin, D.Y., Robins, J.M., Wei, L.J. (1996) Comparing two failure time distributions in the presence of depending censoring. Biometrika 83, 381 - 393

[150] Lindeboom, M., Van Den Berg, G.J. (1994) Heterogeneity in models for bivariate survival: the importance of the mixing distribution. Journal of the Royal Statistical Society (B) 56, 49 - 60

[151] Lindstrom, D.P. (1996) Economic opportunity in Mexico and return migration from the United States. Demography 33, 357 - 374

[152] Link, W.A. (1989) A model for informative censoring. Journal of the American Statistical Association 84, 749 - 752

[153] Locatelli, I. (2003) Frailty models for twins duration data: a Bayesian approach. Ph.D. thesis, University Luigi Bocconi, Milano

[154] Locatelli, I., Lichtenstein, P., Yashin, A.I. (2004) The heritability of breast cancer: A Bayesian correlated frailty model applied to Swedish twins data. Twin Research 7, 182 - 91

[155] Longini, I.M., Halloran, M.E. (1996) A frailty mixture model for estimating vaccine efficacy. Applied Statistics 45, 165 - 173

[156] Macdonald, A.S. (1999) Modeling the impact of genetics on insurance. North American Actuarial Journal 3, 83 - 105

[157] Mahé, C., Chevret, S. (1999) Estimating regression parameters and degree of dependence for multivariate failure time data. Biometrics 55, 1078 - 1084

[158] Maller, R.A., Zhou, X. (2002) Analysis of parametric models for competing risks. Statistica Sinica 12, 725 - 750

[159] Maller, R.A., Zhou, X. (2003) Testing for individual heterogeneity in parameter models for event-history data. Mathematical Methods of Statistics 12, 276 - 304

[160] Manatunga, A.K., Oakes, D. (1999) Parametric analysis of matched pair survival data. Lifetime Data Analysis 5, 371 - 387

[161] Manda, S.O.M., Meyer, R. (2005) Bayesian inference for recurrent event data using time-dependent frailty. Statistics in Medicine 24, 1263 - 1274

[162] Manton, K.G., Stallard, E. (1981) Methods for evaluating the heterogeneity of aging processes in human populations using vital statistics data: Explaining the black/white mortality crossover by a model of mortality selection. Human Biology 53, 47 - 67

[163] Manton, K.G., Stallard, E., Vaupel, J.W. (1981) Methods for comparing the mortality experience of heterogeneous populations. Demography 18, 389 - 410

[164] Manton, K., Stallard, E., Vaupel, J. (1986) Alternative models for heterogeneity of mortality risks among the aged. Journal of the American Statistical Association 81, 635 - 644

[165] Manton, K.G., Vaupel, J.W. (1995) Survival after the age of 80 in the United States, Sweden, France, England, and Japan. New England Journal of Medicine 333, 1232 - 1235

[166] Marenberg, M.E., Risch, N., Berkman, L.F., Floderus, B., de Faire, U. (1994) Genetic susceptibility to death from coronary heart disease in a study of twins. The New England Journal of Medicine 330, 1041 - 1046

[167] Marshall, A.W., Olkin, I. (1988) Families of multivariate distributions. Journal of the American Statistical Association 83, 834 - 841

[168] Martinussen, T., Pipper, C.B. (2005) Estimation in the positive stable shared frailty Cox proportional hazards model. Lifetime Data Analysis 11, 99 - 115

[169] McCall, B.P. (1994) Testing the proportional hazards assumption in the presence of unmeasured heterogeneity. Journal of Applied Econometrics 9, 321 - 334

[170] McGilchrist, C.A., Aisbett, C.W. (1991) Regression with frailty in survival analysis. Biometrics 47, 461 - 466

[171] McGilchrist, C.A. (1993) REML estimation for survival models with frailty. Biometrics 49, 221-225

[172] McGilchrist, C.A., Yau, K.K.W. (1996) Survival analyis with time dependent frailty using a longitudinal model. Australian Journal of Statistics 38, 53 - 60

[173] McGue, M., Vaupel, J.W., Holm, N., Harvald, B. (1993) Longevity is moderately heritable in a sample of Danish twins born 1870-1880. Journal of Gerontology: Biological Sciences B 48, B237-44

[174] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller A., Teller, E. (1953) Equations of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087-1091

[175] Miller, R.G. (1981) Survival analysis. John Wiley & Sons, New York

[176] Miller, R.G. (1983) What price Kaplan-Meier? Biometrics 39, 1077 - 1081

[177] Moger, T.A., Aalen, O.O., Halvorsen, T.O., Storm, H.H., Tretli, S. (2004a) Frailty modelling of testicular cancer incidence using Scandinavian data. Biostatistics 5, 1 - 14

[178] Moger, T.A., Aalen, O.O., Heimdal, K., Gjessing, H.K. (2004b) Analysis of testicular cancer data using a frailty model with familial dependence. Statistics in Medicine 23, 617 - 632

[179] Moger, T.A., Aalen, O.O. (2005) A distribution for multivariate frailty based on the compound Poisson distribution with random scale. Lifetime Data Analysis 11, 41 - 59

[180] Morley, E., Perry, H.M., Miller, D.K. (2002) Something about frailty. Journal of Gerontology: MEDICAL SCIENCES 57A, M698 - M704

[181] Mudholkar, G.S., Srivastava, D.K., Kollia, G.D. (1996) A generalisation of the Weibull with application to the analysis of survival data. Journal of the American Statistical Association 91, 1575 - 1583

[182] Murphy, S.A. (1994) Consistency in a proportional hazards models incorporating a random effect. The Annals of Statistics 22, 712 - 731

[183] Murphy, S.A. (1995) Asymptotic theory for the frailty model. The Annals of Statistics 23, 182-198

[184] Murphy, S.A., Rossini, A., van der Vaart, A.W. (1997) Maximum likelihood estimation in the proportional odds model. Journal of the American Statistical Association 92, 968 - 976

[185] Naylor, J.C., Smith, A.F.M. (1982) Applications of a method for the efficient computation of posterior distribution. Applied Statistics 31, 214 - 225

[186] Neal, R. (1997) Markov chain Monte Carlo methods based on 'slicing' the density function. Technical Report 9722, Department of Statistics, University of Toronto, Canada

[187] Neale, M.C., Cardon, L.R. (1992) Methodology for genetic studies of twins and families. Kluwer, Dordrecht

[188] Nguti, R., Claeskens, G., Janssen, P. (2004) One-sided tests in shared frailty models. Report

[189] Nickell, S. (1979) Estimating the probability of leaving unemployment. Econometrica 47, 1249-66

[190] Nielsen, G.G., Gill, R.D., Andersen, P.K., Sørensen, T.I.A. (1992) A counting process approach to maximum likelihood estimation in frailty models. Scandinavian Journal of Statistics 19, 25-43

[191] Noh, M., Ha, I.D., Lee, Y. (2006) Dispersion frailty models via HGLMs. Statistics in Medicine, in press

[192] Oakes, D. (1982) A concordance test for independence in the presence of censoring. Biometrics 38, 451 - 455

[193] Oakes, D. (1989) Bivariate survival models induced by frailties. Journal of the American Statistical Association 84, 487 - 493

[194] Oakes, D., Jeong, J.-H. (1998) Frailty models and rank tests. Lifetime Data Analysis 4, 209 - 228

[195] Orbe, J., Ferreira, E., Nùñez-Antòn, V. (2002) Comparing proportional hazards and accelerated failure time models for survival analysis. Statistics in Medicine 21, 3493- 3510

[196] Panis, C.W.A., Lillard, L.A. (1995) Child mortality in Malaysia: explaining ethnic differences and the recent decline. Population Studies 49, 463 - 479

[197] Pack, S.E., Morgan, B.J.T. (1990) A mixture model for interval-censored time-to-response quantal assay data. Biometrics 46, 749 - 757

[198] Paik, M.C., Tsai, W.-Y., Ottman, R. (1994) Multivariate survival analysis using piecewise gamma frailty. Biometrics 50, 975 - 988

[199] Pankratz, V.S., de Andrade, M., Therneau, T. M. (2005) Random-effects Cox proportional hazards model: general variance components methods for time-to-event data.
Genetic Epidemiology 28, 97 - 109

[200] Pareto, V. (1897) Cours d'Economie Politique. Paris, Rouge

[201] Parner, E. (1998) Asymptotic theory for the correlated gamma-frailty model. The Annals of Statistics 26, 183 - 214

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

[203] 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

[204] Pike, M.C. (1966) A method of analysis of a certain class of experiments in carcinogesis. Biometrics 22, 142 - 161

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

[206] Qiou, Z., Ravishanker, N., Dey, D. (1999) Multivariate survival analysis with positive stable frailties. Biometrics 55, 637 - 644

[207] Ries, L.A.G., Kosary, C.L., Hankey, B.F. (eds). SEER cancer statistics review 1973-1999. Bethesda, MD. National Cancer Institute, 1999

[208] Ripatti, S., Palmgren, J. (2000) Estimation of multivariate frailty models using penalised partial likelihood. Biometrics 56, 1016 - 1022

[209] Ripatti, S., Larsen, K., Palmgren, J. (2002) Maximum likelihood inference for multivariate frailty models using an automated MCEM algorithm. Lifetime Data Analysis 8, 349 - 360

[210] Roberts, G.O. (1996). Markov chain concepts related to sampling algorithms. In W.R. Gilks, S. Richardson, D.J. Spiegelhalter (eds.) Markov Chain Monte Carlo in practice. Chapman and Hall, London.

[211] Rocha, C.S. (1996) Survival models for heterogeneity using the non-central chi-squared distribution with zero degrees of freedom. In Lifetime data: Models in reliability and survival analysis (eds. N. Jewell et al.), Dordrecht: Kluwer Academic Publishers, 275-79

[212] Rosenthal, T.C., Puck, S.M. (1999) Screening for Genetic Risk of Breast Cancer. American Family Physician 59, 99 - 104

[213] Sahu, K.S., Dey, D.K., Aslanidou, H., Sinha, D. (1997) A Weibull regression model with Gamma frailties for multivariate survival data. Lifetime Data Analysis 3, 123 - 137

[214] Samuelsen, S.O. (2003) Exact inference in the proportional hazard model: possibilities and limitations. Lifetime Data Analysis 9, 239 - 260

[215] Sastry, N. (1997) A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. Journal of the American Statistical Association 92, 426 - 435

[216] Schmidt, P., Witte, A.D. (1989) Predicting criminal recidivism using split population survival time models. Journal of Econometrics 40, 141 - 159

[217] Schnier, C., Hielm, S., Saloniemi, H.S. (2004) Comparison of the breeding performance of cows in cold and warm loose-housing systems in Finland. Preventive Veterinary Medicine 62, 135 - 151

[218] Schmoor, C., Schumacher, M. (1997) Effect of covariate omission and categorization when analyzing randomized trials with the Cox model. Statistics in Medicine 16, 225 - 237

[219] Schumacher, M., Olschewski, M., Schmoor, C. (1987) The impact of heterogeneity on the comparison of survival times. Statistics in Medicine 6, 773 - 784

[220] Schumacher, M. (1989) Frailty-Modelle und ihre Anwendungen in der Medizin. unpublished draft

[221] Shih, J.H., Louis, T.A. (1995) Assessing gamma frailty models for clustered failure time data. Lifetime Data Analysis 1, 205 - 220

[222] Sinha, D., Dey, K.D. (1997) Semiparametric Bayesian analysis of survival data. Journal of the American Statistical Association 92, 1195 - 1212

[223] Smith, A.F.M., Skene, A.M., Shaw, J.E.H., Naylor, J.C. (1987) Progress with numerical and graphical methods for practical Bayesian statistics. The Statistician 36, 75 - 82

[224] Spiegelhalter, D.J., Thomas, A., Best, N.G., Gilks, W.R. (1996) BUGS Examples Volume 1

[225] Spiegelhalter, D.J., Thomas, A., Best, N.G. (1999) WinBUGS Version 1.2 User Manual.

[226] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society (B) 64, 583 - 639

[227] Spielerman, S. (1972) Extensions of the mover-stayer model. American Journal of Sociology 78, 599 - 626

[228] Therneau, T.M., Grambsch, P.M. (2000) Modeling Survival Data. New York: Springer.

[229] Trussell, J., Richards, T. (1985) Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. Sociological Methodology 15, 242 - 276

[230] Tsiatis, A.A. (1975) A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Sciences (USA) 72, 20 - 22

[231] Tweedy, M. (1984) An index which distinguishes between some important exponential families. Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. J. Ghosh, J. Roy (eds), 579-604

[232] 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., Curtsinger, J.W. (1998). Biodemographic trajectories of Longevity. Science 280, 855 - 860

[233] Vaupel, J.W., Harvald, B., Holm, N.V., Yashin, A.I., Xiu L. (1992) Survival analysis in genetics: Danish twin data applied to a gerontological question. Kluwer Academic Publishers, Netherlands

[234] 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

[235] Vaupel, J.W., Yashin, A.I. (1985) Heterogeneity's ruses: Some surprising effects of selection on population dynamics. The American Statistician 39, 176 - 185

[236] Verweij, P.J., van Houwelingen, H.C., Stijnen, T. (1998) A goodness-of-fit test for Cox's proportional hazards model based on martingale residuals. Biometrics 54, 1517 - 1526

[237] Wang, S.T., Klein, J.P., Moeschberger, M.L. (1995) Semi-parametric estimation of covariate effects using the positive stable frailty model. Applied Stochastic Models and Data Analysis 11, 121-33

[238] Wassell, J.T., Moeschberger, M.L. (1993) A bivariate survival model with modified gamma frailty for assessing the impact of interventions. Statistics in Medicine 12, 241-248

[239] Wei, L.J., Glidden, D.V. (1997) An overview of statisical methods for multiple failure time data in clinical trials. Statistics in Medicine 16, 831 - 839

[240] Wei, L.J., Lin, D.Y., Weissfeld, L. (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 84, 1065 - 1073

[241] Weibull, W. (1939) A statistical theory of the strength of materials. Ingenior Ventenskaps Akademien Handlinger 151, Stockholm: Generalstabens Litografiska Anstalts Förlag

[242] Wienke, A. (1996) Asymptotisch optimale adaptive Auswahlverfahren in semiparametrischen Modellen. Hans Jacobs, Lage (Ph.D. thesis)

[243] Wienke, A. (1998) An asymptotically optimal adaptive selection procedure in a proportional hazards model with conditionally independent censoring. Biometrical Journal 40, 963 - 978

[244] Wienke, A., Christensen, K., Holm, N., Yashin, A. (2000) Heritability of death from respiratory diseases: an analysis of Danish twin survival data using a correlated frailty model. Medical Infobahn for Europe. A. Hasman et al., IOS Press, Amsterdam, 407-411

[245] Wienke, A., Holm, N., Skytthe, A., Yashin, A.I. (2001) The heritability of mortality due to heart diseases: a correlated frailty model applied to Danish twins. Twin Research 4, 266 - 274

[246] 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

[247] Wienke, A., Lichtenstein, P., Yashin, A.I. (2003a) A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. Biometrics 59, 1178 - 1183

[248] Wienke, A., Holm, N., Christensen, K., Skytthe, A., Vaupel, J., Yashin, A.I. (2003b) 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

[249] Wienke, A. (2004) Die Vererbbarkeit der Todesursache: ein correlated-frailty Modell angewendet auf Dänische Zwillinge, geboren 1870 - 1930. In: R. Scholz und J. Flöthmann (Hrsg): Lebenserwartung und Mortalität. Materialien zur Bevölkerungswissenschaft Heft 111, Wiesbaden, 81-98

[250] Wienke, A., Herskind, A.M., Christensen, K., Skytthe, A., Yashin, A.I. (2005a) The heritability of CHD mortality in Danish twins after controlling for smoking and BMI. Twin Research and Human Genetics 8, 53 - 59

[251] Wienke, A., Arbeev, K., Locatelli, I., Yashin, A.I. (2005b) A comparison of different correlated frailty models and estimation strategies. Mathematical Biosciences 198, 1 - 13

[252] Wienke, A., Locatelli, I., Yashin, I.A. (2006a) The modelling of a cure fraction in bivariate time-to-event data. Austrian Journal of Statistics 35, 67 - 76

[253] Wienke, A., Lichtenstein, P., Czene, K., Yashin, A.I. (2006b) The role of correlated frailty models in studies of human health, ageing and longevity. In: Applications to Cancer and AIDS Studies, Genome Sequence Analysis, and Survival Analysis. Balakrishnan, N., Auget, J.-L., Mesbah, M., Molenberghs, G. (eds.) Birkhäuser 151 - 166

[254] Wienke, A., Ripatti, S., Palmgren, J., Yashin, A.I. (2006c) A bivariate survival model with compound Poisson frailty. (Journal of the Royal Statistical Society B, submitted)

[255] Wintrebert, C.M.A., Putter, H., Zwinderman, A.H., van Houwelingen, J.C. (2004) Centre-effect on survival after bone marrow transplantation: application of time-dependent frailty models. Biometrical Journal 46, 512 - 525

[256] Xue, X., Brookmeyer, R. (1996) Bivariate frailty model for the analysis of multivariate survival time. Lifetime Data Analysis 2, 277 - 290

[257] Xue, X., Ding, Y. (1999) Assessing heterogeneity and correlation of paired failure times with the bivariate frailty model. Statistics in Medicine 18, 907 - 918

[258] Yashin, A.I., Begun, A., Iachine, I.A. (1999a) Genetic factors in susceptibility to death: a comparative analysis of bivariate survival models. Journal of Epidemiology and Biostatistics 4, 53-60

[259] Yashin, A.I., Herskind, A.-M., Begun, A.Z., Iachine, I.A. (1999b) Survival models for genetic analysis with random effects depending on observed covariates. Unpublished draft

[260] Yashin, A.I., Iachine, I.A. (1995a) Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. Genetic Epidemiology 12, 529 - 538

[261] Yashin, A.I., Iachine, I.A. (1995b) Survival of related individuals: an extension of some fundamental results of heterogeneity analysis. Mathematical Population Studies 5, 321-39

[262] Yashin, A.I., Iachine, I.A. (1997) How frailty models can be used for evaluating longevity limits: Taking advantage of an interdisciplinary approach. Demography 34, 31 - 48

[263] Yashin, A.I., Iachine, I. (1999a) Dependent hazards in multivariate survival problems. Journal of Multivariate Analysis 71, 241 - 261

[264] Yashin, A.I., Iachine, I. (1999b) What difference does the dependence between durations make? Insights for population studies of aging. Lifetime Data Analysis 5, 5 - 22

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

[266] Yashin, A.I., Manton, K.G., Iachine, I.A. (1996) Genetic and environmental factors in duration studies: multivariate frailty models and estimation strategies. Journal of Epidemiology and Biostatistics 1, 115 - 120

[267] Yashin, A.I., Vaupel, J.W., Iachine, I.A. (1993) Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. Working Paper Series: Population Studies of Aging 7, CHS, Odense University

[268] 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

[269] Yau, K.K.W., McGilchrist, C.A. (1997) Use of generalised linear mixed models for the analysis of clustered survival data. Biometrical Journal 39, 3 - 11

[270] Yau, K.K.W., McGilchrist, C.A. (1998) ML and REML estimation in survival analysis with time dependent correlated frailty. Statistics in Medicine 17, 1201 - 1213

[271] Zdravkovic, S., Wienke, A., Pedersen, N.L., Marenberg, M.E., Yashin, A.I., de Faire, U. (2002) Heritability of death from coronary heart disease: a 36 years follow-up of 20,966 Swedish twins. Journal of Internal Medicine 252, 247 - 54

[272] Zdravkovic, S., Wienke, A., Pedersen, N.L., Marenberg, M.E., Yashin, A.I., de Faire, U. (2004) Genetic influences on CHD-death and the impact of known risk factors: Comparison of two frailty models. Behavior Genetics 34, 585 - 591

[273] Zheng, M., Klein, J.P. (1995) Estimates of marginal survival for dependent competing risks based on a assumed copula. Biometrika 82, 127 - 138

Thesen

1. Frailty-Modelle sind statistische Modelle in der Lebensdaueranalyse mit zufälligen Effekten. Univariate Frailty-Modelle dienen der Modellierung unbeobachteter Kovariablen in Lebensdauerdaten, d.h. von unbeobachteter Heterogenität. Bei multivariaten Frailty-Modellen steht die Modellierung von Korrelationsstrukturen bei Lebensdauern in Clustern (Familie, Studienzentrum, Patient, Arztpraxis) im Vordergrund. Dazu wird das Konzept der bedingten Unabhängigkeit verwendet.

2. Der Schwerpunkt der vorliegenden methodischen Arbeit liegt auf bivariaten Frailty-Modellen. Das Ziel ist es, das etablierte Shared-Frailty-Modell zu erweitern und neue Modelle einzuführen, welche eine flexiblere Korrelationsstruktur zwischen den Lebensdauern und eine bessere Interpretierbarkeit der Parameter erlauben. Dabei spielt die Verteilung der zufälligen Effekte (Frailty) eine wichtige Rolle.

3. Ausgangspunkt ist das Korrelierte Gamma-Frailty-Modell mit dem Shared-Gamma-Frailty-Modell als Spezialfall. Zusätzlich zum Heterogenitätsparameter σ^2 beinhaltet es den Parameter ρ, der die Korrelation zwischen den Frailtyvariablen beschreibt.

4. Dieser Korrelationsparameter ist zum Beispiel von zentralem Interesse bei der Analyse des Einflusses von genetischen und Umweltfaktoren auf ein Merkmal.

5. Das Korrelierte Gamma-Frailty-Modell eignet sich hervorragend, um todesursachenspezifische Lebensdauern von Familienangehörigen zu analysieren. Angewendet wird das Modell auf die Lebensdauern von 7955 dänischen Zwillingspaaren der Geburtsjahrgänge 1870 - 1930 bezüglich totaler Mortalität, Krebs, KHK, Schlaganfall und Atemwegserkrankungen in einem follow-up von 1943 - 1993. Mit Hilfe einer Modellvariante wurde die Heritabilität der Todesursachen aus den Zwillingsdaten heraus geschätzt. So zeigt sich eine hohe Heritabilität (ca. 50-60%) bei totaler und KHK-Mortalität, während die Mortalität bedingt durch Atemwegserkrankungen nur in geringem Umfang dem Einfluss genetischer Faktoren unterliegt. Krebsmortalität wird moderat (ca. 20-25%) durch genetische Faktoren beeinflusst.

6. In das Korrelierte Gamma-Frailty Modell können beobachtete Kovariablen wie im proportionalen Hazardmodell auf integriert werden. Dies führt in der Regel zur Verringerung der Heterogenität in der Population, die in kleineren Schätzwerten für den Parameter σ^2 ihren Ausdruck findet. Dies wird für KHK-Sterblichkeit bei 1209 dänischen Zwillingspaaren unter Betrachtung der Kovariablen Rauchen, BMI und Geschlecht demonstriert, wobei sich die Varianz der Frailty nahezu halbiert.

Der Einfluss der beobachtbaren Kovariablen liegt im erwarteten Bereich und ist unabhängig von den betrachteten genetischen Modellen. So ist das relative Risiko für Tod durch KHK eines Rauchers verglichen mit einem Nichtraucher bei 1.77 (95% KI 1.11 - 2.83), für Männer verglichen mit Frauen 4.35 (95% KI 2.46 - 7.69).

7. Eine wichtige Erweiterung der bisher existierenden Modelle stellt die Einführung des Korrelierten compound-Poisson-Frailty-Modells durch den Autor dar. Dieses Modell enthält als Spezialfälle das Korrelierte Gamma-Frailty und das Korrelierte inverse-Gauss-Frailty Modell. Entscheidender Vorteil ist die einfache Darstellung der bivariaten Lebensdauerfunktion. Damit können Maximum-Likelihood-Schätzverfahren angewendet und technische Schwierigkeiten vermieden werden.

8. Das Korrelierte compound-Poisson-Frailty-Modell beinhaltet die Existenz einer gegenüber dem interessierenden Ereignis resistenten Subpopulation, sei es durch Immunisierung oder genetische Prädisposition. Dieses Modell zeigt gute Eigenschaften bei der Analyse von Brustkrebs bei 5857 schwedischen Zwillingspaaren der Geburtsjahrgänge 1886 - 1925 in einem follow-up von 1958 bis 2000. Lediglich 34% der Frauen erweisen sich als anfällig gegenüber Brustkrebs.

9. Moderne MCMC Methoden erlauben eine hohe Flexibilität bei Frailty-Modellen. Insbesondere Parameter in log-Normal-Frailty-Modellen lassen sich damit hervorragend schätzen. Diese Modelle bieten keine explizite Darstellung der Lebensdauerfunktion und Maximum-Likelihood-Schätzungen erweisen sich als technisch schwierig.

10. In umfangreichen Simulationsstudien wurden die Eigenschaften unterschiedlicher Schätzverfahren in verschiedenen Modellen analysiert. Die Parameter lassen sich in den betrachteten Modellklassen ohne Verzerrung schätzen.

11. Frailty-Modelle eignen sich hervorragend zur Modellierung von konkurrierenden Risiken. Hierfür wurde vom Autor ein vierdimensionales Korreliertes Gamma-Frailty-Modell eingeführt, welches eine Korrelation zwischen verschiedenen Todesursachen erlaubt. Während sich diese Korrelation im univariaten Fall aufgrund fehlender Identifizierbarkeit nicht schätzen lässt, ist dies im multivariaten Fall möglich.

12. Die vorliegende Arbeit zeigt anhand zahlreicher praktischer Anwendungen und Simulationen die hervorragenden statistischen Eigenschaften korrelierter Frailty-Modelle und deren Potential für die Analyse von multivariaten Lebensdauerdaten in der medizinischen Forschung.

Lebenslauf

Name:	Andreas Matthias Wienke
Geburtsdatum:	23. Februar 1966
Geburtsort:	Güstrow
Staatsangehörigkeit:	deutsch
Familienstand:	verheiratet, drei Kinder
Schulbildung:	1972 - 1980 Polytechnische Oberschule "Richard Wossidlo", Güstrow
	1980 - 1984 Erweiterte Oberschule "John Brinckman", Güstrow
	Abitur (mit Auszeichnung)
Wehrdienst:	1984 - 1986 Grundwehrdienst
Studium:	1986 - 1991 Universität Rostock
	Diplommathematik mit Spezialisierung in Biometrie
	Nebenfächer Physik, Biologie, Informatik
	1991 Diplom (sehr gut)
Werdegang:	1991 - 1993 Stipendiat am FB Mathematik, Universität Rostock
	1993 - 1996 wiss. Mitarbeiter am FB Mathematik, Universität Rostock
	Promotion zum Dr. rer. nat. (magna cum laude)
	1996 - 1998 wiss. Mitarbeiter am Institut für Epidemiologie
	und Sozialmedizin, Universität Münster
	1998 - 2003 wiss. Mitarbeiter am Max Planck Institut für
	demographische Forschung, Rostock
	seit 2003 wiss. Mitarbeiter am Institut für Medizinische Epidemiologie,
	Biometrie und Informatik und am Koordinierungszentrum für
	Klinische Studien, Universität Halle

Halle, den 12. April 2006

Selbständigkeitserklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Habilitation selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Halle, den 12. April 2006 Andreas Wienke