

IZA DP No. 5748

**Unobserved Heterogeneity in
Multiple-Spell Multiple-States Duration Models**

Govert E. Bijwaard

May 2011

Unobserved Heterogeneity in Multiple-Spell Multiple-States Duration Models

Govert E. Bijwaard

*Netherlands Interdisciplinary Demographic Institute (NIDI)
and IZA*

Discussion Paper No. 5748
May 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Unobserved Heterogeneity in Multiple-Spell Multiple-States Duration Models^{*}

In survival analysis a large literature using frailty models, or models with unobserved heterogeneity, exist. In the growing literature on multiple spell multiple states duration models, or multistate models, modeling this issue is only at its infant phase. Ignoring unobserved heterogeneity can, however, produce incorrect results. This paper presents how unobserved heterogeneity can be incorporated into multistate models, with an emphasis on semi-Markov multistate models with a mixed proportional hazard structure. First, the aspects of frailty modeling in univariate (proportional hazard, Cox) duration models are addressed and some important models with unobserved heterogeneity are discussed. Second, the domain is extended to modeling of parallel/clustered multivariate duration data with unobserved heterogeneity. The implications of choosing shared or correlated unobserved heterogeneity is highlighted. The relevant differences with recurrent events data is covered next. They include the choice of the time scale and risk set which both have important implications for the way unobserved heterogeneity influence the model. Multistate duration models can have both parallel and recurrent events. Incorporating unobserved heterogeneity in multistate models, therefore, brings all the previously addressed issues together. Although some estimation procedures are covered the emphasis is on conceptual issues. The importance of including unobserved heterogeneity in multistate duration models is illustrated with data on labour market and migration dynamics of recent immigrants to The Netherlands.

JEL Classification: C41, J61

Keywords: multiple spell multiple state duration, mixed proportional hazard, multistate model, unobserved heterogeneity, frailty

Corresponding author:

Govert E. Bijwaard
Netherlands Interdisciplinary Demographic Institute (NIDI)
PO Box 11650
NL-2502 AR The Hague
The Netherlands
E-mail: bijwaard@nidi.nl

^{*} Financial support from the NORFACE research programme on Migration in Europe – Social, Economic, Cultural and Policy Dynamics is gratefully acknowledged. This paper is forthcoming in *Demographic Research* under a slightly different title *Multistate event history analysis with frailty*.

1 Introduction

Demographers are increasingly interested in understanding the life histories or the individual life course with a focus on events, their sequence, ordering and transitions that people make from one state of life to another. A multistate model describes the transitions people experience as life unfolds. When people may change among a set of multiple states and/or may experience repeated changes through time a multistate event history model (also known as multistate lifetable and increment-decrement life tables) is a proper choice. Typical examples of such processes in demography include migration, Rogers (1975, 1995), changes in marital status and other life course processes, Courgeau and Lelièvre (1992) and (Willekens 1999). Many other demographic applications of the multistate models exist. Multistate models are also common in medicine and economics. In medicine (biostatistics), the states can describe conditions like healthy, diseased and death. For an overview of the use of multistate models in biostatistics see a.o. Commenges (1999), Hougaard (2000), and Putter et al. (2007)). In economics the main application of multistate models has been labour force dynamics, see Flinn and Heckman (1983), Van den Berg (2001) and, Fougère and Kamionka (2008). Poverty dynamics and recidivism are other important applications of multistate models. The methodology of multistate models is discussed in several books, the most important are Andersen et al. (1993), Hougaard (2000), and Aalen et al. (2008).

The basic parameters of a multistate model are the transition hazard rates or intensities. These intensities may depend on the time spend in a particular state (so called semi-Markov models) and on observed characteristics. Many multistate models assume that the intensities are homogeneous conditional on these observed factors. Unfortunately, it is hardly ever possible to include all relevant factors, either because the researcher does not know all the relevant factors or because it is not possible to measure all the relevant factors. Ignoring such unobserved heterogeneity or frailty may have a huge impact on inference in multistate models. For univariate event history data, also called survival data or duration data, a large literature on models with frailty exits, e.g. Van den Berg (2001), Duchateau and Janssen (2008) and Wienke (2011). In the multistate literature the issue of including frailty is only at its infant phase.

The purpose of this article is to provide an overview of frailty modeling for multistate event history models. We assume that the frailty, just as the effect of observed characteristics, enters the intensity multiplicatively. Thus, we only consider the Cox model in continuous time with frailty, in econometrics called the Mixed Proportional Hazard model, and its multivariate extensions.

The outline of the paper is as follows. In Section 2 we start with a discussion on the issues involved with unobserved heterogeneity in univariate survival models. Multistate models extend the univariate survival models in two dimensions: (1) People may exit to different competing states or several people that may experience an event may be grouped in clusters. In both cases we have parallel events. (2) People may experience multiple spells of the same type or recurrent events. In both dimensions unobserved heterogeneity can be independent, shared or correlated. In Section 3 we discuss these issues separately for parallel and for recurrent event data. In Section 4 unobserved heterogeneity in a multistate setting are addressed, combining the knowledge of the preceding section on incorporating frailties in models for parallel data and in models for recurrent data. In Section 5 we illustrate the importance of incorporating frailty in a semi-Markov multistate model with data on labour market and migration dynamics of recent immigrants to The Netherlands. In Section 6 some important aspects of multistate frailty modeling not yet covered are briefly discussed. Section 7 summarizes the findings.

2 Unobserved heterogeneity in univariate duration models

The most simple multistate model is a univariate survival model, which considers the transition from ‘alive’ to ‘dead’. The observation for a given individual will in this case consist of a random variable T , representing the time from a given origin (time 0) to the occurrence of the event ‘death’. The

distribution of T may be characterized by the survival function $S(t) = \Pr(T > t)$. We can also characterize the distribution of T by its hazard rate

$$\lambda(t) = \frac{\partial \ln(S(t))}{\partial t} = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (1)$$

Thus, $\lambda(\cdot)$ is the transition intensity from state ‘alive’ to state ‘dead’, i.e. the instantaneous probability per time unit of going from ‘alive’ to ‘dead’. The hazard rate provides a full characterization of the distribution of T , just like the distribution function, the survival function, and the density function.

A typical feature of survival analysis is the inability to observe complete durations. A common problem is that by the end of the observation period some individuals are still alive. This kind of incomplete observation is known as right-censoring. The hazard function is usually the focal point of analysis. A major advantage of using the hazard function as a basic building block is that it is invariant to (independent) censoring. The most common model for the hazard rate is the Cox or proportional hazard (PH) model, with hazard rate

$$\lambda(t|X) = \lambda_0(t) \exp(\beta' X), \quad (2)$$

where $\lambda_0(t)$ the called the baseline hazard or duration dependence and it is a function of t alone. This permits coefficients β to be easily interpretable. Suppose that the j th regressor x_j increases by one unit and the other regressors remain unchanged, then $\lambda(t|x_{\text{new}}) = \exp(\beta_j)\lambda(t|x)$. Thus, the elasticity of the the hazard, $\partial \ln(\lambda(t|x))/\partial x_j$ is equal to β_j .

In a Mixed Proportional Hazard (MPH) model it is assumed that all unmeasured factors and measurement error can be captured in a multiplicative random term V . The hazard rate becomes

$$\lambda(t|X, V) = V \lambda_0(t) \exp(\beta' X), \quad (3)$$

This model was independently developed by Vaupel et al. (1979) and by Lancaster (1979). The (random) frailty $V > 0$ is time-independent and independent of the observed characteristics x .

An important feature of MPH model is that the unconditional survival function has the Laplace transform, \mathcal{L}

$$S(t|X) = E_V \left[e^{-v\Lambda(t, X)} \right] = \mathcal{L} \left[\Lambda(t, X) \right] \quad (4)$$

with $\Lambda(t, X) = \int_0^t \lambda_0(s) e^{\beta' X} ds$, the integrated hazard. The derivatives of the Laplace transform can be used to obtain general results about unconditional survival distributions. For example, the unconditional hazard function can be characterized by the Laplace transform of the frailty distribution and their derivatives. This is helpful in clarifying that the impact of frailty in event history models differs substantially from the impact of frailty in linear regression models. In ordinary regression models unobserved heterogeneity leads to more variability of the response compared to the case when the variables are included. In event history data, however, the increased variability implies a change in hazard function. This is clear when we realize that in an MPH model the observed hazard (with the frailty distribution integrated out using the Laplace transform) is

$$\lambda(t|X) = \lambda_0(t) \exp(\beta' X) E(V|T > t, X) \quad (5)$$

with the last term both changing in t and in x . A consequence of this is that ignoring the unobserved heterogeneity (i.e. if one adopts a PH model whereas the data are generated by an MPH model) would make the duration dependence more negative

$$\frac{\partial \ln \lambda(t|X)}{\partial t} = \frac{\lambda'_0(t)}{\lambda_0(t)} - \frac{\text{Var}(V|T > t, X)}{E(V|T > t, X)} \lambda_0(t) \exp(\beta' X) \quad (6)$$

and the effect of covariates biased towards zero

$$\frac{\partial \ln \lambda(t|X)}{\partial x} = \beta - \beta \frac{\text{Var}(V|T > t, X)}{\text{E}(V|T > t, X)} \int_0^t \lambda_0(s) ds \quad (7)$$

The intuition behind this is that individuals with the highest frailty value v (and thus the highest hazard) on average leave the alive state the quickest, so that individuals who are still alive at high durations tend to have lower values of v and thus lower hazards. This sorting phenomenon occurs in all survival models with unobserved heterogeneity, and is not restricted to the MPH model.

We discuss briefly the most commonly used frailty distributions and its implication for the hazard, the survival and other properties of the duration distribution. We focus on the Gamma frailty distribution, the log-normal frailty distribution and, the discrete frailty distribution. For more details on these and other frailty distributions, like the Power Variance Function family of frailty distributions that includes the important Inverse Gaussian and Stable frailty distributions, see Hougaard (2000) and Wienke (2011). Often the distribution of V is standardized at $\text{E}(V) = 1$ (except for positive Stable model). The variance $\sigma^2 = \text{Var}(V)$ is a measure of heterogeneity.

2.1 Gamma frailty model

The Gamma distribution is the most widely applied frailty distribution. From an analytical and computational view it is a very convenient distribution. The closed form expressions for the unconditional survival and hazard are easy to derive. The density of the Gamma distribution $\Gamma(k, \theta)$ is $f(v) = \frac{1}{\Gamma(k)} \theta^k v^{k-1} e^{-\theta v}$ with $\text{E}(V) = \frac{k}{\theta}$ and $\text{Var}(V) = \frac{k}{\theta^2}$. When $k = 1$ it is identical to the exponential distribution. A common normalization is $k = \theta$, such that $\text{E}(V) = 1$ and $\sigma_v^2 = \frac{1}{\theta}$. The unconditional survival (with the frailty integrated out) is

$$S(t|X) = \left[1 + \sigma^2 \Lambda(t, X)\right]^{-1/\sigma_v^2}$$

and the unconditional hazard

$$\lambda(t|X) = \frac{\lambda_0(t) \exp(\beta' X)}{1 + \sigma_v^2 \Lambda(t, X)}$$

The frailty distribution among survivors is still Gamma distributed but with expectation and variance

$$\text{E}(V|T > t, X) = \frac{1}{1 + \sigma_v^2 \Lambda(t, X)} < 1 \quad \text{Var}(V|T > t, X) = \frac{\sigma_v^2}{[1 + \sigma_v^2 \Lambda(t, X)]^2} \rightarrow 0$$

Note that nearly all arguments in favour of the gamma distribution are based on mathematical and computational aspects. However, Abbring and van den Berg (2007) rationalise the preference for the gamma distribution, by showing that, in a large class of univariate frailty distributions, the distribution of the frailty among the survivors converges to a gamma distribution as time goes to infinity under mild regularity conditions.

2.2 Log-normal frailty model

The link with random effects or mixed models makes the log-normal model very attractive. A disadvantage is the lack of closed form expressions. But with the increasing computer power the numerical solution of the integrals involved is not an issue anymore. In the log-normal frailty model the frailty $V = e^W$ with $W \sim \mathcal{N}(m, s^2)$, then we have

$$\text{E}(V) = e^{m + \frac{s^2}{2}} \quad \sigma^2 = \text{Var}(V) = e^{2m + s^2} (e^{s^2} - 1) \quad (8)$$

In the literature two types of constraints on the model are considered. Either impose that the frailty has expectation one, which implies that $m = -\frac{1}{2}s^2$ and $\text{Var}(V) = (e^{s^2} - 1)$, or impose $m = 0$, which implies that the logarithm of the frailty has expectation zero. The latter implies that mixed models software can be used.

2.3 Discrete frailty model

A simple explanation for univariate frailty models is to assume that the population consists of two (or more) latent sub populations (latent classes), which are homogeneous within. For example we may have (1) a high risk subpopulation that leaves fast, and (2) a low risk subpopulation that leaves slowly, but the class identification for each individual is unknown. Let us assume that the proportion of individuals belonging to sub-population 1 is p , then

$$V = \begin{cases} v_1 > 1 & \Pr(V = v_1) = p \\ v_2 < 1 & \Pr(V = v_2) = 1 - p \end{cases} \quad (9)$$

Usually a transformed version is used with $v_j = e^{\alpha_j}$, securing that the frailty is non-negative and, $p = e^\gamma / (1 + e^\gamma)$, securing that the probability is between zero and one. The expectation and variance of the discrete frailty are $E(V) = pv_1 + (1-p)v_2$ and $\text{Var}(V) = p(1-p)(v_1 - v_2)^2$ and the unconditional survival is (using the Laplace transform)

$$S(t|X) = p \exp(-v_1 \Lambda(t, X)) + (1-p) \exp(-v_2 \Lambda(t, X)) \quad (10)$$

and the unconditional hazard

$$\lambda(t|X) = \lambda_0(t) \exp(\beta' X) \frac{pv_1 \exp(-v_1 \Lambda(t, X)) + (1-p)v_2 \exp(-v_2 \Lambda(t, X))}{p \exp(-v_1 \Lambda(t, X)) + (1-p) \exp(-v_2 \Lambda(t, X))} \quad (11)$$

Note that the proportion of individuals in subpopulation 1 decline over time as

$$\begin{aligned} p(t) &= \Pr(\text{type 1} | T > t) \\ &= \frac{p \exp(-v_1 \Lambda(t, X))}{p \exp(-v_1 \Lambda(t, X)) + (1-p) \exp(-v_2 \Lambda(t, X))} < p \end{aligned} \quad (12)$$

Sometimes the restriction $E(V) = 1$ is imposed then $v_2 = (1 - pv_1)/(1 - p)$. The discrete frailty model is a finite mixture model. In econometrics such models are the most commonly applied frailty models. The estimation of the finite mixture model may be carried out under the assumption of either known or unknown number of mixtures. Heckman and Singer (1984c) derive the nonparametric maximum likelihood estimation for the latter case. The two-point mixture model was used by Vaupel and Yashin (1985) to discuss ideas of heterogeneity and selection in more detail. Schumacher et al. (1987) use this frailty distribution to model heterogeneity in clinical trials. A special case of a discrete frailty distribution is when a subpopulation will never experience the event. This so-called mover-stayer, also called cure model or split population model, includes long-term survivors who never experience the event, see a.o. Boag (1949), Schmidt and Witte (1989) and Maller and Zhou (1996). Finally, the discrete frailty model is also related to latent class models.

3 Unobserved heterogeneity in multivariate duration models

There are two typical ways multivariate event history data can arise. The first situation is *parallel duration data*, also termed *clustered duration data*, in which several individuals that may experience

an event are grouped in a cluster. Examples include twin and family studies, but also simultaneous events of one individual. A typical clustered data example is the competing risks model of different causes of death. The second situation is *recurrent/repeated events* which arises when several events of the same type are registered for each individual, for instance child birth to a woman. In univariate event history models frailty captures the possible heterogeneity due to unobserved covariates. In a multivariate setting unobserved heterogeneity can also be used to model associations between event times.

A key point for an MPH model is conditional independence, that is conditional on the frailty v the survival times are independent. In our multivariate setting we continue to assume the MPH structure and the conditional independence. In principle the frailty might be independent for each event time. Then the analysis does not differ from the analysis in a univariate setting. Here we consider the more interesting cases of (1) shared frailty and (2) correlated frailty. First we focus on parallel data.

3.1 Shared frailty

The shared frailty approach assumes that within a cluster the value of the frailty term is constant over time and common to all individuals in the cluster. This common term, thus, creates dependence between event times within a cluster. This dependence is always positive. The shared frailty model, first introduced by Clayton (1978), dominates the literature on multivariate survival models, see a.o. Hougaard (2000), Therneau and Grambsch (2000) and Duchateau and Janssen (2008).

Hence frailty measures the specific risk level for a cluster and given V the event times are independent. Consider n event times T_1, \dots, T_n related to one cluster with observed characteristics X_{ij} and frailty V . Note that the number of members in a cluster can vary between the clusters. We assume an MPH structure. Then, conditional on frailty V , the joint survival in that cluster is

$$S(t_j, \dots, t_{jn}|X, V) = \Pr(T_1 > t_1, \dots, T_n > t_n) = \exp\left(-V \sum_{i=1}^n \Lambda(t_i, X_i)\right) \quad (13)$$

The joint unconditional survival is obtained by integrating out V

$$S(t_1, \dots, t_n|X) = \mathbb{E}\left[\exp\left(-V \sum_{i=1}^n \Lambda(t_i, X_i)\right)\right] = \mathcal{L}\left(\sum_{i=1}^n \Lambda(t_i, X_i)\right) \quad (14)$$

Thus, for a shared frailty model the multivariate survival function is also expressed as the Laplace transform of the frailty distribution, now evaluated at the sum of the integrated hazard functions. Similar to the univariate case many possible alternative choices for the frailty distribution in the shared frailty model exist. Since the unconditional survival can still be expressed in the Laplace form (compare equation (14) and equation (4)) the hazard and other properties of the (unconditional) joint event times for the shared frailty case are related to the event time distribution in the univariate case with the same frailty distribution.

Despite the similarity between individual frailty and shared frailty conceptually they are different. In the univariate case the frailty variance σ^2 is a measure of unobserved heterogeneity, while in a shared frailty multivariate case the frailty variance is a measure of correlation between event times within a cluster. Event times from different clusters are considered to be independent.

Similar to the univariate case is that the most popular choice is the Gamma frailty distribution. If V follows Gamma distribution with mean 1 and variance σ^2 , then the unconditional joint survival is

$$\begin{aligned} S(t_1, \dots, t_n|X) &= \left(1 + \sigma^2 \sum_{i=1}^n \Lambda(t_i, X_i)\right)^{-\frac{1}{\sigma^2}} \\ &= \left(\sum_{i=1}^n S(t_i|X_i)^{-\sigma^2} - (n-1)\right)^{-\frac{1}{\sigma^2}} \end{aligned} \quad (15)$$

The correlations between event history times of members of the same cluster are always the same, implying a symmetric situation, and these correlations are always positive. This makes the model less useful for modeling correlations in family studies with groups of relatives.

After the Gamma distribution the log-normal distribution is the most important frailty distribution. Normally distributed random effects allow for more flexibility, especially in modeling multivariate correlation structures, see Section 3.2. A drawback of the log-normal frailty distribution is the lack of analytical solutions to the hazard and survival functions, see Section 2.2. The development of user-friendly programs that incorporate integral approximation techniques and/or Bayesian Markov Chain Monte Carlo procedure in connection with the increasing computer power has reduced the computational burden and the flexibility of normal distributed frailties outweighs its computational disadvantages.

The shared discrete frailty model also allows for more flexible correlation structures. In contrast to a log-normal distribution a discrete frailty distribution has analytical solutions to the (unconditional) hazard, survival and density functions. For example, for a two-point discrete shared frailty that takes the values v_1 and v_2 with $\Pr(V = v_1) = p$, the unconditional survival within a cluster is

$$S(t_1, \dots, t_n | X) = p \exp\left(-v_1 \sum_{i=1}^n \Lambda(t_i, X)\right) + (1 - p) \exp\left(-v_2 \sum_{i=1}^n \Lambda(t_i, X)\right) \quad (16)$$

Many of the PVF family of frailty distributions also have an analytical solution in the shared frailty situation. Hougaard (2000) discuss the shared PVF frailty model and its special cases in detail.

The shared frailty model has some important limitations (see Xue and Brookmeyer (1996) for an extensive discussion). First, the assumption that the frailty is the same for all members in the cluster is often inappropriate. For example, in a family study it hard to defend that all relatives in a family share all their unobserved risk factors. Second, shared frailty models only induce positive association within clusters. However, in some situations the event times for individuals within the same cluster are negatively associated. For example, the reduction in the risk of dying from one disease may increase the risk of dying from another disease. Third, the dependence between survival times within a cluster is based on marginal distributions of event times. This leads to a symmetric relationship between all possible pairs within a cluster. It also limits the interpretation of the variance of shared frailty model as a measure of association between event times within a cluster and not as a measure of unobserved heterogeneity. Correlated frailty models allow for more flexibility.

3.2 Correlated frailty

The correlated frailty model combines both the shared frailty approach and the univariate frailty model. In the correlated frailty model the frailties of individuals within a cluster are correlated but not necessarily shared. It enables the inclusion of additional correlation parameters and associations are no longer forced to be the same for all pairs of individuals within a cluster. We consider three different ways of generating correlated frailties: (i) additive frailty in which the frailty is the sum of a cluster-specific and an individual-specific component; (ii) nested frailty, in which the frailty is the multiplication of a cluster-specific and an individual-specific component; and (iii) joint modeling of the member specific frailties within a cluster. In all three cases the conditional survival still has an MPH structure.

$$S(t_1, \dots, t_n | V_1, \dots, V_n, X) = \prod_{j=1}^n \exp\left(-V_j \Lambda_j(t_j, X_j)\right) \quad (17)$$

where V_1, \dots, V_n are n correlated frailties.

The additive frailty model based on a correlated gamma distribution was introduced by Yashin et al. (1995). The model has a very convenient representation of the survival function in closed form.

Consider a bivariate model clustered event history model with additive gamma frailty. Each frailty is constructed by adding two components, one common to both, W_0 and one individual specific:

$$V_1 = \sigma_1^2(W_0 + W_1) \quad (18)$$

$$V_2 = \sigma_2^2(W_0 + W_2) \quad (19)$$

with $W_0 \sim \Gamma(1, \frac{1}{k_0})$, $W_1 \sim \Gamma(1, \frac{1}{k_1})$, and $W_2 \sim \Gamma(1, \frac{1}{k_2})$, three gamma distributed with mean 1 and variance $1/k_j$. This implies that the two frailties are also gamma distributed with $V_1 \sim \Gamma(1, \frac{1}{k_0+k_1})$ and $V_2 \sim \Gamma(1, \frac{1}{k_0+k_2})$. Thus, both frailties have mean one and variance

$$\text{Var}(V_1) = \sigma_1^2 = (k_0 + k_1)^{-1} \quad (20)$$

$$\text{Var}(V_2) = \sigma_2^2 = (k_0 + k_2)^{-1} \quad (21)$$

The correlation between the frailties implied by this model is

$$\rho = \frac{k_0}{\sqrt{(k_0 + k_1)(k_0 + k_2)}} \quad (22)$$

The unconditional joint survival function is

$$S(t_1, t_2 | X_1, X_2) = \frac{S_1(t_1 | X_1)^{1 - \frac{\sigma_1}{\sigma_2} \rho} S_2(t_2 | X_2)^{1 - \frac{\sigma_2}{\sigma_1} \rho}}{\left[S_1(t_1 | X_1)^{-\sigma_1^2} + S_2(t_2 | X_2)^{-\sigma_2^2} - 1 \right]^{\frac{\rho}{\sigma_1 \sigma_2}}} \quad (23)$$

with $S_j(t_j | X_j) = (1 + \sigma_j^2 \Lambda_j(t_j | X_j))^{-1/\sigma_j^2}$. An important limitation of the additive gamma frailty model is that the correlation between frailties is for $\sigma_1 \neq \sigma_2$ always less than one

$$0 \leq \rho \leq \min \left[\frac{(k_0 + k_1)}{(k_0 + k_2)}, \frac{(k_0 + k_2)}{(k_0 + k_1)} \right] < 1 \quad (24)$$

Hence, when the values of σ_1 and σ_2 differ a lot the correlation cannot be very large. Another disadvantage of the additive correlated gamma frailty is that the model becomes very complex with increasing cluster size.

One solution to this problem is to assume an additive discrete correlated frailty. This distribution does not restrict the correlation between two members of a cluster. For a two-point additive discrete frailty we have $V_1 = W_0 + W_1$ and $V_2 = W_0 + W_2$ where for $j = 0, 1, 2$ $W_j = w_{j1}$ with probability p_j and $W_j = w_{j2}$ with probability $1 - p_j$. Thus,

$$V_1 = \begin{cases} v_{11} = w_{01} + w_{11} & \Pr(V_1 = v_{11}) = p_{11} = p_0 p_1 \\ v_{12} = w_{01} + w_{12} & \Pr(V_1 = v_{12}) = p_{12} = p_0 (1 - p_1) \\ v_{13} = w_{02} + w_{11} & \Pr(V_1 = v_{13}) = p_{13} = (1 - p_0) p_1 \\ v_{14} = w_{02} + w_{12} & \Pr(V_1 = v_{14}) = p_{14} = (1 - p_0) (1 - p_1) \end{cases} \quad (25)$$

and similar for V_2 . The unconditional joint survival for the additive discrete frailty is

$$S(t_1, t_2 | X_1, X_2) = \sum_{k=1}^4 \sum_{l=1}^4 p_{1k} p_{2l} \exp\left(-v_{1k} \Lambda_1(t_1, X_1)\right) \exp\left(-v_{2l} \Lambda_2(t_2, X_2)\right) \quad (26)$$

A disadvantage of the additive discrete correlated frailty distribution is that the number of parameters grows very fast with the number of members in a cluster and with the number of support points. If the clusters have n members and we use k support points the additive correlated frailty adds $(n+1)(2k-1)$

more parameters to estimate. For other frailty distributions the additive correlated frailty model leads to very complex survival and hazard functions.

The *nested frailty* model assumes that the clustering of the event times occurs at multiple levels. In family studies where we have a hierarchical clustering, by family and individual, this models seems appropriate. Sastry (1997a) suggested a nested frailty model with two hierarchical levels in which the frailty of member j in a particular cluster is $V_j = W_0 \cdot W_j$ with W_0 and W_j are mutually independent unit-mean gamma distributed random variables with variance η_0 and η_1 . Thus, within each cluster the frailty is composed of a cluster-specific component common to all cluster members times an individual-specific component that are mutually independent. The unconditional survival for this nested gamma frailty has a complicated form, but estimation is possible using an EM-algorithm (Sastry 1997a; Sastry 1997b), a Bayesian procedure (Manda 2001) or penalized likelihood methods (Rondeau et al. 2003).

It is also possible to base the nested frailty model on discrete frailty distributions. For a two-point nested discrete frailty we have $V_j = W_0 \cdot W_j$ where for $j = 0, 1, \dots, n$ $W_j = w_{j1}$ with probability p_j and $W_j = w_{j2}$ with probability $1 - p_j$. The unconditional survival for the bivariate nested frailty model with two-point discrete is very similar to (26) and therefore not presented.

A very flexible way to allow for correlated frailties is by modeling the *joint frailty* distribution directly. The correlated log-normal frailty model, first applied by Xue and Brookmeyer (1996), is especially useful in modeling dependence structures. The distribution can be obtained by assuming a multivariate normal distribution on the logarithm of the frailty vector, which is in the bivariate case,

$$\begin{pmatrix} \ln V_1 \\ \ln V_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s_1^2 & r s_1 s_2 \\ r s_1 s_2 & s_2^2 \end{pmatrix}\right) \quad (27)$$

Similar to the univariate log-normal frailty model we have $EV_j = e^{0.5s_j^2}$ and $\text{Var}(V_j) = \sigma_j^2 = e^{s_j^2}(e^{s_j^2} - 1)$. The correlation between the two frailties is

$$\rho = \text{Corr}(V_1, V_2) = \frac{e^{rs_1s_2} - 1}{\sqrt{e^{s_1^2} - 1}\sqrt{e^{s_2^2} - 1}} \quad (28)$$

It is rather easy to generalize the correlated log-normal frailty model to more than two correlated frailties. However, as already mentioned in Section 2.2, the log-normal distribution does not have analytical solutions for the unconditional joint survival and hazards and the number of integrals to evaluate for calculating them increases with the dimension of the multivariate normal distribution.

It is also possible to use a joint discrete frailty distribution. If the cluster has only two members and we use a two-point discrete frailty the joint frailty distribution is

$$\begin{aligned} \Pr(V_1 = v_{11}, V_2 = v_{21}) &= p_1 \\ \Pr(V_1 = v_{11}, V_2 = v_{22}) &= p_2 \\ \Pr(V_1 = v_{12}, V_2 = v_{21}) &= p_3 \\ \Pr(V_1 = v_{12}, V_2 = v_{22}) &= p_4 = 1 - p_1 - p_2 - p_3 \end{aligned}$$

Thus, we need to estimate seven additional parameters. When the number of support points and the number of members in a cluster increase the number of parameters to estimate, which is $(2nk - 1)$, increases very fast. A one-factor loading specification has been a popular solution to reduce the number of parameters. It assumes that there is a univariate random variable W such that

$$V_j = \exp(\alpha_j W) \quad (29)$$

It is straightforward to generalize this specification to higher dimensions of the random variable W . If W is two-dimensional then we obtain a two-factor loading specification, i.e. $V_j = \exp(\alpha_{j1}W_1 + \alpha_{j2}W_2)$. Note that if the distribution of W is (multivariate) normal we obtain the log-normal (correlated)

frailty. Thus, the factor loading approach can also be used to reduce the computational burden of the correlated log-normal frailty, see for example Bonnal et al. (1997). A common approach for a discrete frailty is to assume that the distribution(s) of W are mutually independent on $(-1, 1)$ with $p_k = Pr(W_k = 1)$. If A is the matrix of factor loading then the variance-covariance matrix of the log-frailty is given by $Var(\ln(V)) = AVar(W)A'$ with $Var(W_k) = 4p_k(1 - p_k)$. The main disadvantage of the discrete frailty one-factor loading specification is that it imposes a perfect positive correlation between the frailties. The two-factor loading specification (or a factor loading specification of higher dimension) does not impose such a strong correlation. It is also possible to use a factor loading model based on other distributional assumptions for W . Note that the computational burden is less a problem in the case of discrete frailties.

3.3 Unobserved heterogeneity in recurrent events models

So far we have only considered parallel events, that includes the case of multiple simultaneous events for one individual. It is also possible that a single individual can experience the same event several times. Reviews of models for such recurrent event data appeared in Cook and Lawless (2007). An example of recurrent event data are the fertility histories of women. There is a state corresponding to the number of children a woman has given birth to at any time (or age). For each birth she moves to the next state. Recurrent data can be represented in different ways depending on the timescale that is used; gap time or total time (see e.g. Kelly and Lim (2000)). Related to the choice of the time scale are the *risk-interval* and the *risk set*. The risk interval corresponds to the time interval where an individual is at risk of experiencing the event. The risk set is the collection of individuals which are at risk at a certain point in time. In the gap-time representation, time at risk starts at 0 after an event ends at the time of the next event. Hence, time is reset to zero after each event. In the total-time formulation, the length of the time at risk is the same as in the gap-time representation. The difference is that the starting time of the at-risk period is not reset to zero after an event but it is put equal to the actual time since the beginning of the observation period.

From the risk intervals for each individual we can derive the risk sets for each occurrence. The risk set contains all the individuals who are at risk for a particular event. In survival data, the risk set at a particular time typically consists of all the individuals that have entered the study and that are still observed at that time. For recurrent events we can distinguish between restricted and unrestricted risk sets, depending on whether we view each occurrence as a separate event conditional on the previous occurrence or as a sequence of similar events. In the latter case we use unrestricted risk sets which allow that all individuals' risk intervals may contribute to the risk set for any given occurrence, regardless of the number of occurrences experienced by each individual. This is the recurrent events analog of the risk sets for survival data. That is, at each point in time the risk set for any occurrence at that time consists of all individuals currently observed. If the order of the events is considered to be important a restricted risk set is used in which contributions to the j^{th} risk set is restricted to include only the j^{th} occurrence risk intervals. Thus, an individual who has experienced the event 3 times after 10 months is with an unrestricted risk set in the risk set for any occurrence at month 10 of its own and the other individuals including the first and second occurrence of any other individual, while with a restricted risk set it is only in the risk set for the occurrences at month 10 of those individuals who have experienced the event exactly 3 times before 10 months.

On using the risk intervals and risk sets, we can formulate a stochastic counting process that describes the number of occurrences of the event. Despite that the counting process formulation has increasingly become the standard framework for analyzing event history data, especially in biostatistics, and it facilitates the analysis we only use the traditional approach. We refer the reader interested in counting process theory to Andersen et al. (1993) or Aalen et al. (2008). Less technical books on counting processes include Klein and Moeschberger (2003) and Therneau and Grambsch (2000).

Based on the choice of the risk set the three most common approaches to recurrent events are the

independent increment model of Andersen and Gill (1982), the marginal model of Wei et al. (1989), and the conditional model of Prentice et al. (1981), see Kelly and Lim (2000) for a comparison. For the marginal and conditional models, each occurrence of the event is modeled as a separate event, while the independent increment model assumes that all events of one individual are identical. The independent increments model is usually defined in total time, but it can also be formulated in gap time. This model assumes that the gap times are generated from a renewal process. In essence, the marginal model treats the consecutive event times as if they come from an unordered competing risk setting, with the number of occurrences at the number of competing events. The marginal model can only be formulated in total time. The conditional model assumes that an individual cannot be at risk for the second occurrence of an event until the event has occurred for the first time.

Nielsen et al. (1992) discuss how to include unobserved heterogeneity in the independent increments model. In Chapter 9 of his seminal book Hougaard (2000) also discusses shared (over time) frailty models for recurrent events. Frailty models specially designed for recurrence data are considered in detail in Oakes (1992), Duchateau et al. (2003) and Bijwaard et al. (2006). For recurrent events the conditional joint survival is

$$S(t_1, \dots, t_n | X, V_1, \dots, V_k) = \prod_{j=1}^n \exp\left(-V_k \Lambda_k(t_j, X_k)\right) \quad (30)$$

which is very similar to the conditional joint survival for parallel data. Therefore, we only mention the relevant differences. For recurrent events the frailty variation is not a group variation, but a variation between individuals, and the variation described by the hazard function is not an individual variation but a variation within individuals. The interpretation of the frailty variance also depends on the time scale and risk sets used. For example, in a gap time recurrent events model with restricted risk sets the risk set decreases at each time point, just as for parallel data, whereas in a total time recurrent events model with unrestricted risk sets, the risk set is constant over the whole observation period. As a consequence in the former case it is crucial to observe the event times, whereas in the latter case the frailty approach leads to variation in the number of events, even though the observation period is the same for all individuals. If we are only interested in the number of events instead of when these events occur count models, like the Poisson model and negative binomial model, are more appropriate. Poisson model with random effect are then the obvious extension for including frailties. We refer to Cook and Lawless (2007) for more details on Poisson models. Here, we assume the event timing is important.

Thus, for recurrent events we can choose whether we have independent frailties, shared frailties or correlated frailties. If we assume independent frailties, that is for each individual and each occurrence the random frailty is independent, we are basically back to the univariate case in Section 2. The shared frailty (individual specific) recurrent events is statistically equal to the shared frailty models for parallel data, see Section 3. But the frailty variance now leads to additional variation between individuals instead of within group correlation. All the mentioned frailty distribution in Section 2 and Section 3 can be used. The choice of the time scale and the risk set also influence the interpretation of the frailty variance. Duchateau et al. (2003) address this issue and conclude that the choice of the timescale mainly depends on the question you want to answer.

The arguments for correlated frailty models also apply for recurrent events. An extension of the correlated frailty model that is particularly relevant for recurrent events is time-dependent frailty models. Often re-occurrence times close to each other are highly correlated, while times further apart are less correlated. To model such kind of serial dependence Yau and McGilchrist (1998) define a dynamic frailty model that assumes that the frailties on subsequent intervals follow an autocorrelation process of order one.

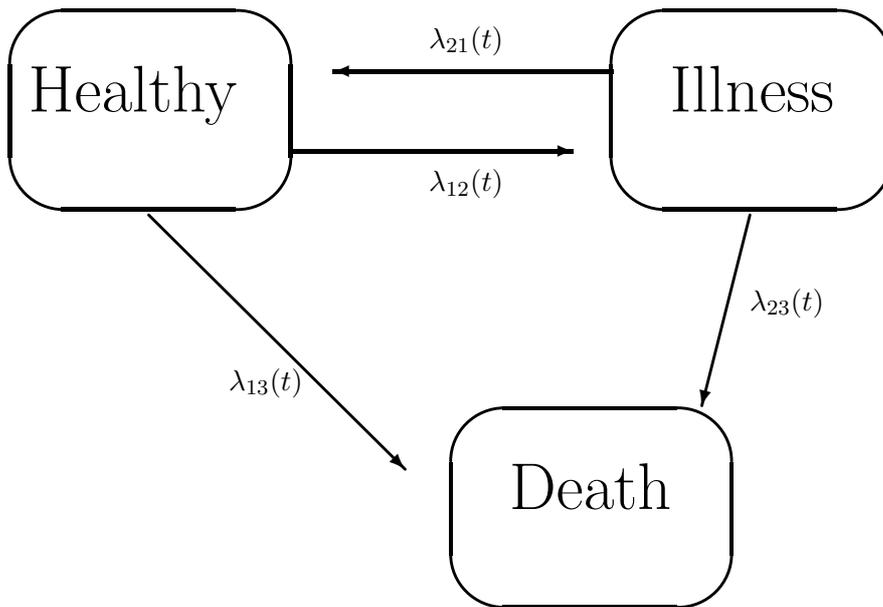


Figure 1: Illness-death model

4 Unobserved heterogeneity in multistate models

A multistate model is defined as a stochastic process, which at any point in time occupies one of a set of discrete states. The class of multistate models includes both recurrent and clustered multivariate duration data. In that respect special cases of a multistate model are the multivariate parallel and recurrent models in the previous section. Thus, including frailty in multistate models follows the lines of the previous sections. Multistate models can be very complex. Before introducing multistate models with unobserved heterogeneity we explain the main concepts of multistate models.

4.1 Multistate model concepts

The most commonly applied multistate model in biostatistics is the illness-death model, see Putter et al. (2007). This model is depicted in Figure 1. In this class of models individuals start out healthy, the initial state 1. From healthy they may become ill (state 2) or they may die (state 3). Ill individuals may die or recover and become healthy again. Most concepts of multistate duration analysis can be explained using this simple model.

Multistate modeling is closely related to Markov chain theory and many of its terminology originate from the theory of Markov chains and processes. Most multistate models have three types of states: the initial state(s), the states an individual can enter the study; absorbing states, states that represent an endpoint from which the individual cannot leave or one is not interested in what happens after this state has been reached; intermediate or transient states are all other states. In an illness-death model death is an absorbing state and illness is an intermediate state. The multistate event history model is defined in hazard/transitions rates. We denote the hazard to make a transition from state i to state j ($i \neq j$) at t by $\lambda_{ij}(t)$.

Just as for recurrent events the choice of the time scale in a multistate model has important implications for the analysis. In a total time representation the event times, t , correspond to the time since the individual entered the initial state. The time keeps moving forward, also when intermediate events occur or when the individual returns to the initial state. In a gap time representation the event times correspond to the time since the entry in state i . The time is reset to zero each time an individual makes a transition. Gap time is also called sojourn time, clock reset time and backward recurrence time. The time scale chosen has implications for the risk set, that defines who are at risk for a particular event, for a transition within a multistate framework.

Another important choice in relation to the risk set is whether a Markov model is assumed. In a Markov model the transition rate only depends on the state an individual is in, not on the time an individual has been in that state nor on any other events that occurred before entering that state. Thus, multistate models in gap time representation cannot be Markov models as the time scale itself depends on the history before the current state has been reached. If it is assumed the gap times depend on the history of the process only through the present state, the resulting multistate model is a Markov renewal model. In a semi-Markov model the transition rate from one state to another state also depends on the time the individual has spent in that particular state. A semi-Markov model in total time representation is also possible, but only with an additional time scale, measuring the sojourn time. In a Markov model the transition rate may depend on the time since entry. This is called a time-inhomogeneous Markov model. Including such time dependence in a semi-Markov model implies an additional time scale, measuring the time since entry. We restrict to uni-time scale multistate models with a semi-Markov model in gap time representation. Thus, we assume duration dependence in the transition rates. The duration dependence measures the effect of the length of stay in a particular state. It is rather straightforward to extend this model to let the transition rates depend on the order or occurrence of the particular state, recurrent events effects, by allowing for occurrence specific duration dependence and for occurrence specific covariate effects. We restrict to mixed proportional hazard type transition rates that assume that, conditionally on the value of the frailty, the semi-Markov property holds.

4.2 Unobserved heterogeneity in illness-death model

We illustrate the choices involved in including unobserved heterogeneity in multistate models by using the illness-death model. In the illness-death model we have, assuming for the moment that the transition rates from illness to healthy and vice versa do not depend on the number of times an individual has been ill, four transitions rates

$$\lambda_{12}(t|X_{12}, v_{12}) = v_{12}\lambda_{012}(t)\exp(\beta_{12}X_{12}) \quad \text{from healthy to ill} \quad (31)$$

$$\lambda_{13}(t|X_{13}, v_{13}) = v_{13}\lambda_{013}(t)\exp(\beta_{13}X_{13}) \quad \text{from healthy to death} \quad (32)$$

$$\lambda_{21}(t|X_{21}, v_{21}) = v_{21}\lambda_{021}(t)\exp(\beta_{21}X_{21}) \quad \text{from ill to healthy} \quad (33)$$

$$\lambda_{23}(t|X_{23}, v_{23}) = v_{23}\lambda_{023}(t)\exp(\beta_{23}X_{23}) \quad \text{from ill to death} \quad (34)$$

When all the frailties are mutually independent the model reduces to two independent competing risks model, and we can use one of the frailty distributions mentioned in Section 2. From the healthy state the competing states the individual can move to are the ill state and the death state. From the illness state the individual can either move to healthy (again) or to death. In both cases the competing risks are uncorrelated and the frailty variance is a measure of unobserved heterogeneity within the origin (healthy or ill) -destination (ill or death) combination.

For both a shared and a correlated frailty model we have three possible linkages. The frailties can be linked over the origin states, over the destination states or over both origin and destination states. An illness-death model with shared frailty model over destination states (by origin) implies equal frailties from the healthy state $v_1 = v_{12} = v_{13}$ and equal frailties from the illness state $v_2 = v_{21} = v_{23}$. Then the competing risk models from these two origin states have joint unconditional survivals as in equation (15) or (16). The frailty variance is in this case a measure of correlation between events times from either healthy to illness or to death or from illness to healthy or to death. When the frailties are shared over the origin states (by destination) we only have that the frailties to death from either healthy or illness are equal, $v_3 = v_{13} = v_{23}$. Because there is only one transition to illness and only one to healthy the frailties v_{12} and v_{21} are still independent. Sharing over the origin states to death implies that the joint unconditional survival functions to death are linked using one of the shared frailty distributions in Section 3.1. The frailty variance reflects now the correlation between the transition

from healthy to death and the transition from illness to death. When the frailties are linked over both the origin and the destination states we only have one individual specific frailty shared over all the four possible transitions. This implies that we have a four dimensional joint unconditional survival function and that the frailty variance is a measure of correlation between these four transitions.

Concerning correlated frailty models we additionally have the choice between three different ways (just as in Section 3.2) of generating correlation between the linked transitions, an additive frailty model, a nested frailty model or a joint frailty model. When the correlation is based on the destination states we only need to model the correlation between v_{13} and v_{23} and we are back at the two-dimensional situation discussed in Section 3.2. With correlation based on origin states we have two sets of mutually independent correlated frailties, the frailties of the healthy state, v_{12} and v_{13} , and the frailties of the illness state, v_{21} and v_{23} . For both sets we can follow the arguments in Section 3.2. When all four frailties are correlated we can use four-dimensional frailty models. Note that the additive gamma frailty model become very complex for four dimensional frailties. For the discrete correlated frailty models a two-factor loading model specification would leave the parameter space manageable without putting too much restriction on the correlations. A model with the frailties shared over the origin states and correlated over the destination states, or vice versa, is not possible.

In the illness-death model only the healthy and illness state can be recurrent. Still, it might be plausible that the transition rate to death from each of these states also changes with re-occurrence. In the first case, only the transition rates from healthy to ill, $\lambda_{12}(t)$ and from ill to healthy, $\lambda_{21}(t)$ depend on how many times the individual has been ill (healthy), while in the latter case all transition rates are recurrent dependent. We focus on the latter case. The transition rates then also depend on the number of times, k , the individual has been ill

$$\lambda_{12k}(t|X_{12k}, v_{12k}) = v_{12k} \lambda_{012k}(t) \exp(\beta_{12k} X_{12k}) \quad (35)$$

$$\lambda_{13k}(t|X_{13k}, v_{13k}) = v_{13k} \lambda_{013k}(t) \exp(\beta_{13k} X_{13k}) \quad (36)$$

$$\lambda_{21k}(t|X_{21k}, v_{21k}) = v_{21k} \lambda_{021k}(t) \exp(\beta_{21k} X_{21k}) \quad (37)$$

$$\lambda_{23k}(t|X_{23k}, v_{23k}) = v_{23k} \lambda_{023k}(t) \exp(\beta_{23k} X_{23k}) \quad (38)$$

The number of possible frailty structures now becomes rather large. We try to capture them all in a very brief manner and focus on the differences with the models mentioned above. When the frailties are shared over the occurrences, i.e. $v_{12k} = v_{12}$ and $v_{21k} = v_{21}$ etc., the possible models are basically the same as mentioned above, with the only difference that the baseline duration and the regression function is stratified by occurrence. With K the maximum number of illness recurrences we have for an illness-death model with recurrent illness $4K$ possible correlated frailties: $\{v_{121}, \dots, v_{12K}, v_{131}, \dots, v_{13K}, v_{211}, \dots, v_{21K}, v_{231}, \dots, v_{23K}\}$. Thus, assuming mutually independent frailties implies that the model reduces to $2K$ independent competing risks models, one for each origin state that are the first time in the healthy state, the first time in the ill state, the second time in the healthy state etc. When the frailties are only independent over the recurrences we have for each recurrence a separate shared or correlated model, with the three possible linkages over the origin states, the destination states or both the origin and destination states.

In principle a very complicated illness-death model with frailties correlated over recurrences, over origin states and over destination states is possible. The analysis of such a model is, however, very complex as it involves the integration of a $4K$ -dimensional frailty distribution that contains many additional parameters to be estimated. The computational burden and the dimension of the parameter space can be reduced by assuming a factor loading model for the frailties. Assuming a one-factor loading model for all the (correlated) frailties would lead to (at least) $4K + 1$ additional frailty parameters. However, the one-factor loading model implies a perfect correlation between the frailties. Assuming a two-factor loading model, which is more flexible, implies (at least) $8K + 2$ additional parameters. Of course, models with frailties that are correlated over recurrences but independent or shared over origin states and/or over destination states are also possible.

4.3 General multistate models with unobserved heterogeneity

For general, beyond the simple illness-death model, multistate models many alternative correlation structures for the frailties are possible. In principle a multistate model has three dimensions; the origin states, the destination states and the recurrent events of a particular state. For each of these dimensions the frailties can either be independent, shared or correlated. Table 1 to Table 3 describe the possible frailty structures. When the frailties are correlated on at least one of the three dimension the tables provide the implies correlation structure. Otherwise just the dimension, i, m for origin, j, r for destination and k, g for recurrent events, are given. The three tables are split according to the three possible correlation for the recurrent events; shared recurrent events, Table 1; independent recurrent events, Table 2; correlated recurrent events, Table 3. In the general multistate model the hazard from state i to state j ($i \neq j$) for the k^{th} time is

$$\lambda_{ijk}(t|X_i, V_{ijk}) = V_{ijk}\lambda_{ijr0}(t) \exp(\beta'_{ijk}X_{ijk}),$$

Of course, it is allowed to put some restrictions on the duration dependence, on the observed characteristics (that are allowed be time-dependent) or on the effect of the observed characteristics on the hazard. For example, the duration dependence might be shared for all exits of one origin state, $\lambda_{ijr0}(t) = \lambda_{ir0}(t)$, the observed characteristics might be shared over all recurrent events, $X_{ijk} = X_{ij}$, and the effect of these factors might only depend on the destination state $\beta_{ijk} = \beta_j$. Here we only focus on the structure of the frailties. The

Table 1: Possible structures of frailty in a multistate model

Shared over recurrent events; $k = g$ and $k \neq g$

Origin i and $m \neq i$	Destination state; $j \neq r$		
	Independent	Shared	Correlated
Independent	$v_{ijk} = v_{ij}$	$v_{ijk} = v_{irg} = v_i$	$\rho(v_{ijk}, v_{irg}) = \rho_{jr}^i$ $\rho(v_{ijk}, v_{mjg}) = 0$ $\rho(v_{ijk}, v_{mrg}) = 0$ $\rho(v_{ijk}, v_{ijg}) = 1$
Shared	$v_{ijk} = v_{mjg} = v_j$	$v_{ijk} = v$	$\rho(v_{ijk}, v_{irg}) = \rho_{jr}$ $\rho(v_{ijk}, v_{mjg}) = 1$ $\rho(v_{ijk}, v_{mrg}) = \rho_{jr}$ $\rho(v_{ijk}, v_{ijg}) = 1$
Correlated	$\rho(v_{ijk}, v_{irg}) = 0$ $\rho(v_{ijk}, v_{mjg}) = \rho_{im}^j$ $\rho(v_{ijk}, v_{mrg}) = 0$ $\rho(v_{ijk}, v_{ijg}) = 1$	$\rho(v_{ijk}, v_{irg}) = 1$ $\rho(v_{ijk}, v_{mjg}) = \rho_{im}$ $\rho(v_{ijk}, v_{mrg}) = \rho_{im}$ $\rho(v_{ijk}, v_{ijg}) = 1$	$\rho(v_{ijk}, v_{irg}) = \rho_{jr}^i$ $\rho(v_{ijk}, v_{mjg}) = \rho_{im}^j$ $\rho(v_{ijk}, v_{mrg}) = \rho_{ijmr}$ $\rho(v_{ijk}, v_{ijg}) = 1$

Shared frailties over recurrent events is the most common choice in multistate modelling and is therefore discussed first. The alternative frailty structures for a multistate model with shared recurrent events frailties are given in Table 1. With both independent frailty by origin and by destination state, we have for each origin-destination pair an independent frailty, v_{ij} . When the frailty is shared over destination states, i.e. all destinations from one origin share the same frailty, then the frailty distribution only depends on the origin the individual is v_i and similarly when the frailty frailty is shared over origin states. When the frailty is shared over origin and destination states we have only one frailty value for each individual. In all three cases the choice of the shared frailty distribution

and its implication for the unconditional survival and hazard functions can be derived by arguments similar to those in Section 3.1. Of course, for a shared frailty model the frailties are correlated in the dimension we shared.

When the frailties are correlated in either dimension Table 1 provides the correlation structure. When the frailties are independent over one dimension and correlated over the other (still assuming shared frailties over recurrent events), the correlation structure is very simple. For example, when the frailties are correlated over origin states the frailties from two different origins, i and m , to the same destination j are correlated and depend on the destination state, i.e. $\rho(v_{ijk}, v_{mjk}) = \rho_{im}^j$. When the frailties are shared over one dimension and correlated over the other, the number of possible non-zero correlations decreases. For example, when the frailties are correlated over destination states the frailties to two different destination, j and r , from any origin are correlated and independent of the origin state i or m , i.e. $\rho(v_{ijk}, v_{mrk}) = \rho_{jr}^k$. Then the correlation of frailties from two different origins, i and m , to the same destination j are perfectly correlated, $\rho(v_{ijk}, v_{mjk}) = 1$. In the case that the frailties are correlated over both origin states and destination states the correlation is defined in all possible origin-destination combinations, e.g. $\rho(v_{ijk}, v_{mrk}) = \rho_{ijmr}^k$. In all the correlated cases we can follow the arguments in Section 3.2 to define the correlation structure. Note that for all cases in Table 1 the frailties in recurrent events are all correlated.

Table 2: Possible structures of frailty in a multistate model
Independent over recurrent events; $k \neq g$

Origin $i \neq m$	Destination state; $j \neq r$		
	Independent	Shared	Correlated
Independent	v_{ijk}	$v_{ijk} = v_{irg} = v_{ik}$	$\rho(v_{ijk}, v_{irk}) = \rho_{jr}^{ik}$ $\rho(v_{ijk}, v_{mjk}) = 0$ $\rho(v_{ijk}, v_{mrk}) = 0$ $\rho(v_{ijk}, v_{mrg}) = 0$
Shared	$v_{ijk} = v_{mjk} = v_{jk}$	$v_{ijk} = v_k$	$\rho(v_{ijk}, v_{irk}) = \rho_{jr}^k$ $\rho(v_{ijk}, v_{mjk}) = 1$ $\rho(v_{ijk}, v_{mrk}) = \rho_{jr}^k$ $\rho(v_{ijk}, v_{mrg}) = 0$
Correlated	$\rho(v_{ijk}, v_{irk}) = 0$ $\rho(v_{ijk}, v_{mjk}) = \rho_{im}^{jk}$ $\rho(v_{ijk}, v_{mrk}) = 0$ $\rho(v_{ijk}, v_{mrg}) = 0$	$\rho(v_{ijk}, v_{irk}) = 1$ $\rho(v_{ijk}, v_{mjk}) = \rho_{im}^k$ $\rho(v_{ijk}, v_{mrk}) = \rho_{im}^k$ $\rho(v_{ijk}, v_{mrg}) = 0$	$\rho(v_{ijk}, v_{irk}) = \rho_{jr}^{ik}$ $\rho(v_{ijk}, v_{mjk}) = \rho_{im}^{jk}$ $\rho(v_{ijk}, v_{mrk}) = \rho_{ijmr}^k$ $\rho(v_{ijk}, v_{mrg}) = 0$

The alternative frailty structures for a multistate model with *independent recurrent events frailties* are presented next, in Table 2. Now the correlation depends on the reoccurrence, with indicator k . For example, with both independent frailty over origin and over destination states, each origin-destination-reoccurrence pair has a independent frailty v_{ijk} and, with independent frailty over destination and correlated frailty over origin, the correlation between the frailties of two origins leading to the same destination (and the same reoccurrence) now also depends on the reoccurrence, $\rho(v_{ijk}, v_{mjk}) = \rho_{im}^{jk}$.

Finally, the alternative frailty structures for a multistate model with *correlated over recurrent events frailties* are presented in Table 3. It shows that the correlation structure can be rather complex. However, when the frailties are independent over origin and over destination states, the correlation structure is only non-zero among recurrent events for the same origin-destination pair. When the frailties are shared over either origin or destination states, the correlation between the frailties. In the

Table 3: Possible structures of frailty in a multistate model, *Correlated over recurrent events*; $k \neq g$

Origin $i \neq m$	Destination state; $j \neq r$		
	Independent	Shared	Correlated
Independent	$\rho(v_{ijk}, v_{irk}) = 0$ $\rho(v_{ijk}, v_{mjk}) = 0$ $\rho(v_{ijk}, v_{mrk}) = 0$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}^{ij}$ $\rho(v_{ijk}, v_{irg}) = 0$ $\rho(v_{ijk}, v_{mjg}) = 0$ $\rho(v_{ijk}, v_{mrg}) = 0$	$\rho(v_{ijk}, v_{irk}) = 1$ $\rho(v_{ijk}, v_{mjk}) = 0$ $\rho(v_{ijk}, v_{mrk}) = 0$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}^i$ $\rho(v_{ijk}, v_{irg}) = \rho_{kg}^i$ $\rho(v_{ijk}, v_{mjg}) = 0$ $\rho(v_{ijk}, v_{mrg}) = 0$	$\rho(v_{ijk}, v_{irk}) = \rho_{jr}^{ik}$ $\rho(v_{ijk}, v_{mjk}) = 0$ $\rho(v_{ijk}, v_{mrk}) = 0$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}^{ij}$ $\rho(v_{ijk}, v_{irg}) = \rho_{jrk}^i$ $\rho(v_{ijk}, v_{mjg}) = 0$ $\rho(v_{ijk}, v_{mrg}) = 0$
Shared	$\rho(v_{ijk}, v_{irk}) = 0$ $\rho(v_{ijk}, v_{mjk}) = 1$ $\rho(v_{ijk}, v_{mrk}) = 0$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}^j$ $\rho(v_{ijk}, v_{irg}) = 0$ $\rho(v_{ijk}, v_{mjg}) = \rho_{kg}^j$ $\rho(v_{ijk}, v_{mrg}) = 0$	$\rho(v_{ijk}, v_{irk}) = 1$ $\rho(v_{ijk}, v_{mjk}) = 1$ $\rho(v_{ijk}, v_{mrk}) = 1$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}$ $\rho(v_{ijk}, v_{irg}) = \rho_{kg}$ $\rho(v_{ijk}, v_{mjg}) = \rho_{kg}$ $\rho(v_{ijk}, v_{mrg}) = \rho_{kg}$	$\rho(v_{ijk}, v_{irk}) = \rho_{jr}^k$ $\rho(v_{ijk}, v_{mjk}) = 1$ $\rho(v_{ijk}, v_{mrk}) = \rho_{jr}^k$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}^j$ $\rho(v_{ijk}, v_{irg}) = \rho_{jrk}^g$ $\rho(v_{ijk}, v_{mjg}) = \rho_{kg}^j$ $\rho(v_{ijk}, v_{mrg}) = \rho_{jrk}^g$
Correlated	$\rho(v_{ijk}, v_{irk}) = \rho_{jr}^k$ $\rho(v_{ijk}, v_{mjk}) = \rho_{im}^{jk}$ $\rho(v_{ijk}, v_{mrk}) = 0$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}^j$ $\rho(v_{ijk}, v_{irg}) = 0$ $\rho(v_{ijk}, v_{mjg}) = \rho_{imkg}^j$ $\rho(v_{ijk}, v_{mrg}) = 0$	$\rho(v_{ijk}, v_{irk}) = 1$ $\rho(v_{ijk}, v_{mjk}) = \rho_{im}^k$ $\rho(v_{ijk}, v_{mrk}) = \rho_{im}^k$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}^i$ $\rho(v_{ijk}, v_{irg}) = \rho_{kg}^i$ $\rho(v_{ijk}, v_{mjg}) = \rho_{imkg}$ $\rho(v_{ijk}, v_{mrg}) = \rho_{imkg}$	$\rho(v_{ijk}, v_{irk}) = \rho_{jr}^{ik}$ $\rho(v_{ijk}, v_{mjk}) = \rho_{im}^{jk}$ $\rho(v_{ijk}, v_{mrk}) = \rho_{ijmr}^k$ $\rho(v_{ijk}, v_{ijg}) = \rho_{kg}^{ij}$ $\rho(v_{ijk}, v_{irg}) = \rho_{jrk}^i$ $\rho(v_{ijk}, v_{mjg}) = \rho_{imkg}^j$ $\rho(v_{ijk}, v_{mrg}) = \rho_{ijmrkg}$

case that the frailties are correlated over both origin states and destination states the correlation is defined in all possible origin-destination-reoccurrence combinations, e.g. $\rho(v_{ijk}, v_{mrg}) = \rho_{ijmrkg}$. In all the correlated cases we can follow the arguments in Section 3.2 to define the correlation structure.

Multistate models with correlated frailty can become very complex. Of course, the choice of the correlation structure in a multistate model depends on data availability and on the questions one wants to answer. For high dimensional multistate models, including correlated frailties would extend the parameters space too. Dimensionality problems are common to multistate model, not only to multistate models with frailty. One might consider reducing the origin, destination or reoccurrence state space. However, by using factor loading models it is possible to reduce the dimension of correlated frailty distribution a lot, without losing much flexibility.

5 Empirical illustration

Labour market transitions and return migration of immigrants are intertwined (especially for labour migrants) and should, therefore, be analysed in conjunction. We use a multistate duration model to analyse the labour market dynamics together with the migration dynamics of recent labour migrants to The Netherlands.

All legal immigration by non-Dutch citizens to the Netherlands is registered in the Central Register Foreigners (Centraal Register Vreemdelingen, CRV), using information from the Immigration Police (Vreemdelingen Politie) and the Immigration and Naturalization Service (Immigratie en Naturalisatie Dienst, IND). It is mandatory for every immigrant to notify the local population register immediately after the arrival in the Netherlands if he intends to stay in for at least two thirds of the forthcoming six months. Our data comprise the entire *population* of immigrants who entered during our observation window of 1999-2005, and after merging in other administrative registers we obtain a panel.

In addition to the date of entry and exit, the administration also records the migration motive of the individual. Either the motive is coded according to the visa status of the immigrant, or the immigrant reports the motive upon registration in the population register. See Bijwaard (2010) for an extensive descriptive analysis of the various migration motives. Here we focus exclusively on labour migrants, which comprise about 23% of all non-Dutch immigrants in the age group 18-64 years.

This immigration register is linked by Statistics Netherlands to the Municipal Register of Population (Gemeentelijke Basisadministratie, GBA) and to their Social Statistical Database (SSD). The GBA contains basic demographic characteristics of the migrants, such as age, gender, marital status and country of origin. From the SSD we have information (on a monthly basis) on the labour market position, income, industry sector, housing and household situation. Based on the main source of income we distinguish three labour market states: employed (including self-employment), unemployed receiving benefits, and non-participation (no income).

Table 4: Spell dynamics of the labour migrants (# 45,987)

	# of spells	Percentage ending in			
		employed	UI	NP	Abroad
Employed	73375		6%	39%	12%
Unemployed (UI)	8735	46%		28%	4%
Non-participation (NP)	31873	44%	12%		25%
Abroad	22153	10%	1%	4%	

Source: Statistics Netherlands, based on own calculations.

The migrants in our sample show substantial dynamic behavior. Of all the migrants that enter in 1999-2003, including those that arrive in December 2003, 48% leaves the country at least once, 24% has more than one employment spell, 11% has at least one unemployment spell and 40% has at least one non-participation (no-income in host) spell. Table 4 report the observed transitions among the four different states. The majority of employment spells end in non-participation, while the majority of non-participation spells end abroad. The majority of the spells abroad are censored, the migrants are still abroad at the end of the observation period. About half of the, relatively small, number of unemployment spells end in employment. But a third of the unemployed receiving benefits leave the labour market. Very few migrants leave the country from unemployment.

By definition all labour migrants start in the employed state at entry. Soon after arrival some migrants move to the other states. Some may return and some may move on to another state. But the migrant is always in one of the four states. In Figure 2 we depict the development of the distribution over the four states for the 1999-entry cohort. After six years less than 40% of these labour migrants is still employed. The proportion of migrants abroad continuously increases. Six years after arrival more than 50% of the labour migrants have left the country. The remaining migrants who are not

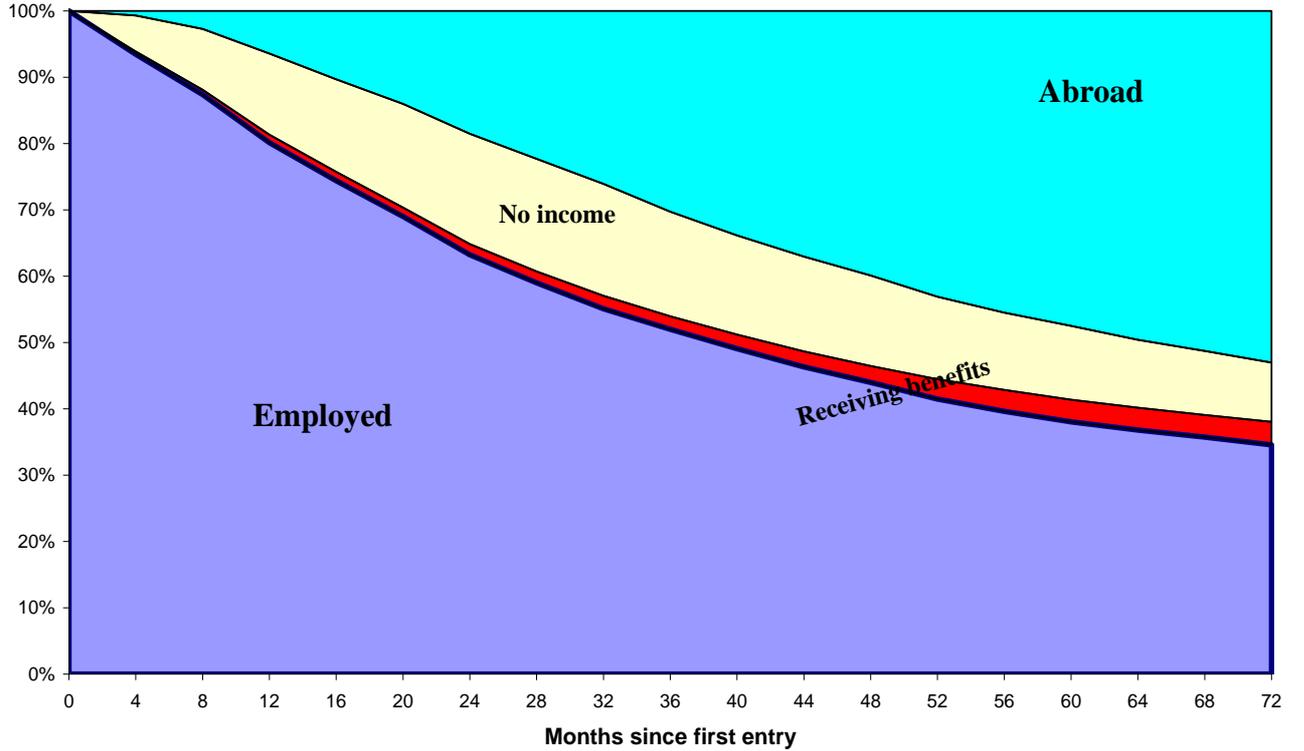


Figure 2: Development of SES of labour immigrants arriving in 1999

employed are mainly in the country without income. Only a few migrants get unemployment benefits, possibly because they do not have gained any benefit rights in the Netherlands. However, this number slowly increasing. For more information on these data see Bijwaard (2009).

5.1 Multistate model

We view the migrant behavior as a semi-Markov process with individuals moving between the four states: (1) Employed in the Netherlands; (2) Unemployed and receiving benefits in the Netherlands; (3) Out of the labour market (and not receiving benefits= non-participating) in the Netherlands (NP); (4) Living abroad. These states are mutually exclusive and exhaust all possible destinations.¹ A migrant may leave a state $j = \{e, u, n, a\}$ for any of the other destination states. The 4-state multistate model is depicted in Figure 3.

We use a competing risks model hazard model for each origin-destination pair. Define the random variables T_{jk} that describe the time since entry in j for a transition from j to k . We assume a (mixed) proportional hazard model for which the intensity for the transition from j to k is:

$$\lambda_{jk}(t|\bar{X}_{jk}(t), v_{jk}) = v_{jk}\lambda_{0jk}(t) \exp(\beta'_{jk}X_{jk}(t)) \quad (39)$$

where $\bar{X}_{jk}(t) = \{X_{jk}(s)|0 \leq s \leq t\}$ is the sample path of the observed characteristics up to time t , which is, without loss of generality, assumed to be left continuous. We assume that the frailty is shared over possible recurrent events. For the baseline duration $\lambda_{0jk}(t)$ we assume that it is piecewise constant on eleven intervals (every six months and beyond five years).

We use three different frailty models: (1) a PH model, thus a model without unobserved heterogeneity (PH); (2) uncorrelated MPH model with a two-point discrete unobserved heterogeneity (MPH); and (3) a two-factor loading frailty correlated over the origin state (correlated). We estimate

¹The death rate for the age range 18-64 is small enough to ignore deaths.

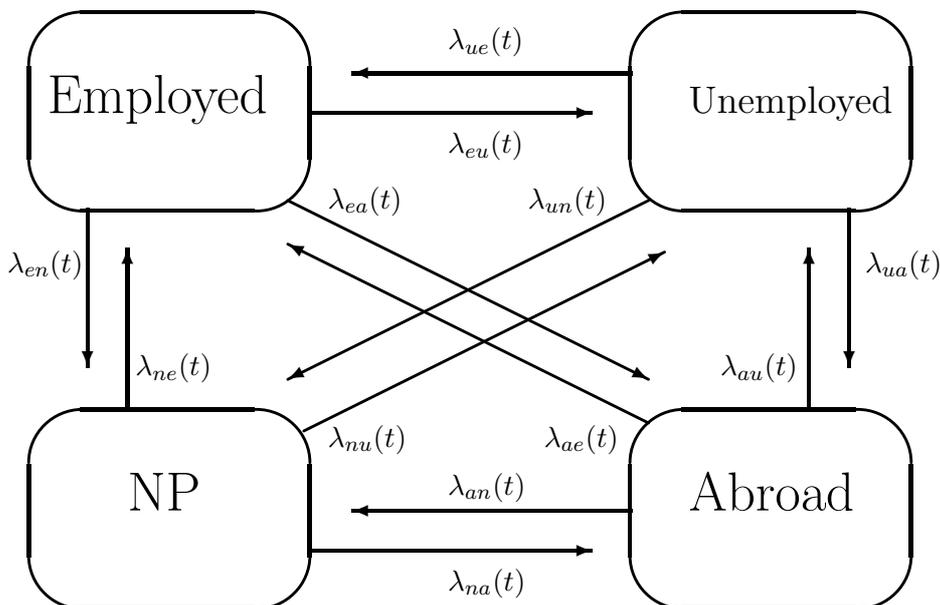


Figure 3: Multistate model for labour-migration dynamics

all models using a maximum likelihood procedure. The covariates included in the model refer to demographic (gender, age, marital status and age of children), country of origin, individual labour market characteristics (monthly income, industry sector), labour market history and migration history. We control for business cycle conditions by including the national unemployment rate, both at the moment of first entry to the country and the time-varying monthly rate. The unemployment rate at entry captures the ‘scarring effect’ of migrants, while the running unemployment rate captures the impact of the business cycle on the transition intensities. With the abundant information on the migrants the model contains many parameters. Here we only discuss the parameter estimates for the transition from employed to abroad, $\lambda_{ea}(t)$ and, focus on the differences induced by the alternative frailty assumptions.²

The estimated duration dependence and covariate effects of $\lambda_{ea}(t)$ are reported in Table 5. As expected ignoring unobserved heterogeneity leads to large negative duration dependence in the hazard of leaving the country, although departure still exhibits a peak after 2 to 3 years in the Netherlands. Allowing for correlation among all the three competing frailties starting in the employed state; employed to abroad, v_{ea} , employed to unemployed, v_{eu} , and employed to non-participation, v_{en} , reduces the positive duration dependence implied by the uncorrelated frailty model. Although we do not encounter changes in sign including unobserved heterogeneity can have rather large effect on the size of the effect on the departure hazard. Most effects get more pronounced after allowing for unobserved heterogeneity, with the notable exception that the effect of repeated entry declines after allowing for unobserved heterogeneity. In general allowing for correlation between the competing frailties affects the parameter estimates less than the introducing unobserved heterogeneity. However, the effect of repeated employment on departure changes from increasing the hazard by a factor 2.9 in an uncorrelated frailty model (and from 2.2 in the model without unobserved heterogeneity) to a factor 4 in the correlated frailty model.

Table 6 reports the implied frailty variance and frailty correlation. For both the uncorrelated and the correlated frailty model we find that the frailty variance is significant. For the correlated frailty model the frailty variance is much larger. In the correlated model we also find significant correla-

²All estimation results are available from the author. In Bijwaard (2009) the marginal effects of the MPH with uncorrelated frailties is discussed.

Table 5: Parameter estimates transition from employed to abroad, $\lambda_{ea}(t)$

	<i>Independent</i>		<i>Correlated</i>
	PH	MPH	MPH
female	-0.137** (0.028)	-0.155** (0.032)	-0.146** (0.032)
self-employed	-1.251** (0.165)	-1.412** (0.173)	-1.396** (0.171)
income < 1000	-0.442** (0.064)	-0.317** (0.052)	-0.328** (0.052)
income 1000-2000	-0.097* (0.039)	-0.095* (0.041)	-0.133** (0.041)
income 3000-4000	0.462** (0.043)	0.492** (0.046)	0.495** (0.046)
income 4000-5000	0.733** (0.053)	0.820** (0.057)	0.816** (0.057)
income > 5000	1.197** (0.040)	1.348** (0.044)	1.313** (0.043)
married	-0.506** (0.030)	-0.645** (0.038)	-0.600** (0.037)
divorced	-0.817** (0.103)	-0.956** (0.114)	-0.925** (0.111)
repeated entry	0.968** (0.051)	0.684** (0.049)	0.663** (0.052)
repeated employment	0.773** (0.026)	1.060** (0.041)	1.392** (0.054)
Unemployment rate at entry	0.073** (0.020)	0.121** (0.023)	0.144** (0.023)
Unemployment rate	0.280** (0.016)	0.270** (0.016)	0.254** (0.016)
α_2 (7-12 months)	0.128** (0.038)	0.293** (0.039)	0.205** (0.039)
α_3 (12-18 months)	0.146** (0.040)	0.421** (0.044)	0.289** (0.044)
α_4 (18-24 months)	0.353** (0.040)	0.705** (0.046)	0.531** (0.046)
α_5 (24-30 months)	0.210** (0.045)	0.640** (0.053)	0.430** (0.052)
α_6 (30-36 months)	0.265** (0.047)	0.749** (0.057)	0.513** (0.056)
α_7 (36-42 months)	0.186** (0.054)	0.729** (0.064)	0.467** (0.063)
α_8 (42-48 months)	0.134* (0.058)	0.719** (0.070)	0.436** (0.070)
α_9 (48-54 months)	-0.038 (0.069)	0.584** (0.080)	0.286** (0.080)
α_{10} (54-60 months)	-0.106 (0.078)	0.543** (0.090)	0.228* (0.090)
α_{11} (> 60 months)	-0.157* (0.070)	0.533** (0.084)	0.204* (0.086)

Country of birth dummies, sector dummies and age at entry dummies are also included. * $p < 0.05$ and ** $p < 0.01$

Table 6: Estimated frailty variance and correlation transition from employed to abroad

	Uncorrelated	Correlated
Frailty variance	1.036** (0.097)	6.554** (1.974)
Correlation		
$\rho(v_{eu}, v_{en})$	-	-0.141* (0.067)
$\rho(v_{eu}, v_{ea})$	-	-0.199** (0.025)
$\rho(v_{en}, v_{ea})$	-	-0.926** (0.021)

* $p < 0.05$ and ** $p < 0.01$

tion between the three competing frailties from the employed state. The large negative correlation between the frailty of the transition from employment to non-participation and the frailty of the transition from employment to abroad implies that employed migrants who are more prone to stay in the country without income are less prone to leave. The frailty for the transition from employment to unemployment and the frailty for the transition from employment to abroad are also negatively correlated.

These results show that in this multistate model for the labour–migration dynamics of recent labour migrants to the Netherlands it is important to include frailties and to allow these frailties to be correlated within each origin state of the multistate model.

6 Other issues related to multistate models with unobserved heterogeneity

In this article we have addressed how to incorporate frailty in multistate duration models of the mixed proportional hazard type in which, conditionally on the value of the frailty, the semi-Markov property holds. Many relevant aspects of multistate frailty modelling we did not cover. In this section we briefly mention some other important issues. Additional references are given for further reading. To limit the size of the article many other issues related to multistate frailty models are not covered.

6.1 Estimation procedures

There are four main estimation methods for duration models with unobserved heterogeneity: maximum likelihood methods, the Expectation-Maximization (EM) algorithm, penalized partial likelihood, and Bayesian Markov Chain Monte Carlo methods. When the conditional survival and conditional hazard functions have an analytical solution it is possible to use maximum likelihood methods. Even without analytical solutions the maximum likelihood estimation can be used when good numerical approximation procedures for the integrals involved are used. For instance, for the integrals implied by the (multivariate) log-normal frailty distribution good approximation methods exist. Just as for event history models without frailty the likelihood is adjusted for (right)censoring and truncation. Details on how to handle censoring and truncation in the likelihood are presented in most books on event history analysis, e.g. Andersen et al. (1993), Klein and Moeschberger (2003) or Aalen et al. (2008).

The second important estimation approach to frailty models is the EM-algorithm. The EM-algorithm allows parameter estimation in semi-parametric multistate gamma and discrete frailty models. In semi-parametric multistate models the baseline duration is treated as a nuisance parameter. This algorithm was suggested by Dempster et al. (1977) and is often used in the presence of unob-

served data. It was adopted for parameter estimation in frailty models by Nielsen et al. (1992) and Klein (1992). The EM algorithm iterates between two steps. In the first step the expectation of the unobserved frailties based on observed data is estimated. These estimates are then used in the second, maximization, step to obtain new parameter estimates given the estimated frailties. See Klein and Moeschberger (2003) and Bijwaard et al. (2006) for a detailed description of using the EM-algorithm to estimate the parameters of a multivariate shared gamma-frailty model.

The third important estimation approach to frailty models is the penalized partial likelihood method, introduced by McGilchrist and Aisbett (1991). It has been applied to the log-normal frailty, Therneau et al. (2003), and the gamma-frailty, Rondeau et al. (2003). It leads to the same estimates as the EM algorithm in the gamma frailty model (Therneau and Grambsch (2000), Duchateau and Janssen (2008)), but not for other frailty distributions. The penalized likelihood penalizes large deviations of the random frailties from its expectation. An advantage of the penalized partial likelihood method is that it can be used to fit log-normal frailty models. This method is also much faster than the EM algorithm. A major disadvantage is that it is very complicated to obtain a valid estimate of the standard error of the frailty variance.

The Markov Chain Monte Carlo (MCMC) method is particularly useful for fitting (correlated) log-normal frailty distributions. For log-normal frailty distributions it is impossible to derive an analytical solution of the conditional survival and conditional hazard functions. Bayesian MCMC methods have been developed by Clayton (1991) and further extended to shared frailty models by Sinha and Dey (1997) and to correlated frailty models by Xue and Ding (1999). In the Bayesian context the frailty distribution represents a prior of the model and its parameters (hyperparameters) are considered as random variables. The MCMC method consists of generating a set of Markov chains whose stationary distribution corresponds to the joint posterior of the model.

6.2 Copula models

The idea behind copulas is to study the dependence, when the influence of the marginal distributions is removed. Often more information about marginal distributions of related variables is available than their joint distribution. The copula approach is a useful method for deriving joint distributions given the marginal distributions, especially when the variables are nonnormal. Another reason to use copulas is that, in a bivariate context, copulas can be used to define nonparametric measures of dependence for pairs of random variables. When fairly general and/or asymmetric modes of dependence are relevant, such as those that go beyond correlation or linear association, then copulas play a special role in developing additional concepts and measures. Finally, copulas are useful extensions and generalizations of approaches for modeling joint distributions and dependence that have appeared in the literature.

Copula models are closely related to shared frailty models, see Goethals et al. (2008). In fact many popular duration models are special cases of the copula model. As there are many different families of copulas, Nelsen (2006), the model allows for flexible specification of the dependence structure between event history times. The significance of copulas lies in the fact that by way of transformation, any joint distribution function can be expressed as a copula applied to the marginal distributions, see Sklar (1959).

A copula is a multivariate joint distribution function defined on the n -dimensional unit cube $[0, 1]$ such that every marginal distribution is uniform on the interval $[0, 1]$. For example the Clayton (1978) copula is

$$S(t_1, \dots, t_n) = C(S_1(t_1), \dots, S_n(t_n)) = \left(\sum_{j=1}^n S_j(t_j)^{-\theta} - (n-1) \right)^{-1/\theta} \quad (40)$$

This looks very similar to the unconditional joint survival function of the gamma-shared frailty model in equation (15). There is, however, a substantial difference that the marginal unconditional survival functions from the copula and the frailty model are not the same. Copula models also covers the extension of the gamma frailty model with negative dependence.

6.3 Dependence measures

It is not straightforward how to assess or quantify dependence in a multivariate duration model, see Hougaard (2000). In traditional multivariate analysis, when a multivariate normal distribution is assumed, the ordinary product moment correlation (Pearson correlation) is used for measuring the dependence between outcomes. However, Pearson correlation only measures the linear dependence. For duration data the marginal distributions are not normal and the dependence structure is often nonlinear. There are at least six dependence measures.

A first dependence measure is the standard Pearson correlation coefficient, which is in the bivariate case $\text{Cov}(T_1, T_2) / \sqrt{\text{Var}(T_1)\text{Var}(T_2)}$. Although it is a commonly used measure that is readily understood it is only useful for linear dependence and multivariate normal distribution. Van den Berg (1997) shows that for a bivariate MPH model without duration dependence the Pearson correlation coefficient is bounded by $-\frac{1}{3}$ from below and $\frac{1}{2}$ from above, regardless of the shape of the joint frailty distribution. In general the correlation coefficient can be low, in absolute value, if the dependence is nonlinear, even with a high degree of dependence.

A second measure of dependence is Kendall's $\tau = \text{E}[\text{sign}\{(T_{11} - T_{21})(T_{12} - T_{22})\}]$ (or Kendall's coefficient of concordance). This is the most popular global ordinal measure of association in the literature on multivariate durations, see e.g. Oakes (1989). This measure of dependence is designed as a rank-based correlation type measure. It seeks to compare the orders of event times in the same group. An advantage of this measure of dependence is that it is invariant with both linear and nonlinear monotone transformations. As a result for the MPH multistate duration model it does not depend on the duration dependence or on value of x . Thus, it only depends on the joint frailty distribution. Van den Berg (1997) shows that for a bivariate MPH model with a joint discrete frailty distribution with n point of support that Kendall's τ is bounded by $-1 + \frac{1}{n}$ from below and by $1 - \frac{1}{n}$ from above. For the shared gamma frailty Kendall's τ is equal to $\sigma^2 / (\sigma^2 + 2)$.

A third measure of dependence is Spearman's correlation coefficient. This measure is based on the marginal ranks of survival times and is, therefore, independent of marginal transformations. A disadvantage of this measure is that its value, which is $12 \int_0^1 \int_0^1 S(u, v) du dv - 3$, is in general difficult to calculate.

A fourth measure of dependence is the median concordance that avoids the conceptual difficulties with the Kendall's τ , which requires two pairs to interpret. Instead of comparing with a second pair, one evaluates the concordance of a single observation in relation to a fixed bivariate point, the median duration. Although this measure of dependence might be difficult to interpret it satisfies the same simple properties of the previous two measures. It ranges from -1 to 1 and is zero under independence. However, similar to those measures, a zero value does not imply independence.

In many applications it may also be of interest to examine the dependence of the residual event times if one conditions on survival up to a certain duration. The next two dependence measure are both such local dependence measures. The fifth dependence measure is the cross-ratio function, Clayton (1978),

$$\Theta_{cr}(t_1, t_2) = \frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 > t_2)} = \frac{S(t_1, t_2)S_{t_1 t_2}(t_1, t_2)}{S_{t_1}(t_1, t_2)S_{t_2}(t_1, t_2)} \quad (41)$$

which is also called the odd-ratio function, captures to what extend the hazard rate of T_1 at t_1 depends on knowledge that T_2 is realized at a certain point of time t_2 , relative to when T_2 is realized after t_2 . The sixth dependence measure is the current-versus-alive function, Oakes (1989),

$$\Theta_{cva}(t_1, t_2) = \frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 > t_1)} \quad (42)$$

captures to what extend the hazard rate of T_1 at t_1 depends on knowledge that T_2 is realized at a certain point of time t_2 , relative to when T_2 is not yet realized. These functions are informative about the way in which the dependence changes over time. More details on both local dependence measures can be found in Anderson et al. (1992) and Van den Berg (1997).

6.4 Identification issues

Associated with frailty models is a general identification problem. This issue concerns the logical possibility of decomposing the individual contributions to the average survival probability of the baseline duration dependence, the unobserved frailty and the observed characteristics, given the observed data. More specifically, if the Proportional hazard model were not identified, then it would be logically impossible to separate the individual contributions of duration dependence and frailty. In the econometric literature the case of the univariate MPH model has been investigated in detail. Elbers and Ridder (1982) and Heckman and Singer (1984b) have established the identification of the MPH model under certain conditions. For an overview, see Van den Berg (2001). The most important assumption here is that the frailty has finite mean and some exogenous variation in the observed characteristics. Ridder and Woutersen (2003) show that bounding the duration dependence hazard away from 0 and ∞ at the start is also sufficient for nonparametric identification of the MPH model and with it the finite mean assumption can be discarded.

Honoré (1993) shows that both the frailty distribution and the duration dependence are identified with multivariate event history data under much weaker assumptions. All shared frailty models are identified without additional information such as observed covariates or parametric assumptions about the duration dependence. Furthermore, the duration dependence may depend on observed covariates in a unspecified way, and the frailty and the observed covariates may be dependent. This identifiability property holds for a broader class of frailty models, including correlated frailty models.

A caveat of multistate data is that such data is more sensitive to censoring. With univariate event history data, many types of censoring can be captured by standard adjustments to the likelihood function, see Andersen et al. (1993) and Klein and Moeschberger (2003). With sequential events, either recurrent or from different types, one has to be more careful. Consider two consecutive events with time t_1 and t_2 , and where the data are subject to right-censoring at a fixed time after the starting point or the first event. Then the moment at which t_2 is right-censored is not independent from t_2 itself. For example, individuals with a small value of the frailty will, on average, have a short time till the first event. As a result the time till the second event will start relatively early. This implies that the time till the second event will often be censored after a relatively longer period (or not censored at all). Thus, t_2 and its censoring probability are both affected by the frailty. It may also happen that the process or some of the processes are not observed from the origin. With left-censoring, not to be confused with left-truncation, the analysis is more complicated, see Heckman and Singer (1984a) and Commenges (2002).

7 Summary and concluding remarks

This article has provided an overview of multiple spell multiple states duration (multistate) duration models with unobserved heterogeneity, with an emphasis on semi-Markov multistate models with a mixed proportional hazard structure. The literature on this subject is continuing and growing and with the increased computer power the complexity of the models will not refrain researchers from using them. We have seen that ignoring frailty can have a large impact on the parameters of interest of the transition hazards, the duration dependence and the effect of observed covariates on the hazard. We have shown how different correlation structures of the frailties in a multistate model can be achieved.

Obviously, I did not intend to cover exhaustively all aspects of multistate frailty models. Many issues we did not address receive ample attention in the literature. An important observation is that the literature is highly segmented into mathematical research, biostatistical research, econometric research and demographic research. Although different terms are used, the problems addressed are similar, and the solutions are often very similar too. I advocate to look beyond the borders of your own research discipline to grasp the knowledge of the other fields.

References

- Aalen, O. O., Ø. Borgan, and H. K. Gjessing (2008). *Survival and Event History Analysis*. New York: Springer–Verlag.
- Abbring, J. H. and G. J. van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* 94, 87–99.
- Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer–Verlag.
- Andersen, P. K. and R. D. Gill (1982). Cox’s regression model for counting processes: A large sample study. *Annals of Statistics* 10, 1100–1120.
- Anderson, J. E., T. A. Louis, N. V. Holm, and B. Harvald (1992). Time-dependent association measures for bivariate survival distributions. *Journal of the American Statistical Association* 87, 641–650.
- Bijwaard, G. E. (2009). Labour market status and migration dynamics. Discussion Paper No. 4530, IZA.
- Bijwaard, G. E. (2010). Immigrant migration dynamics model for The Netherlands. *Journal of Population Economics* 23, 1213–1247.
- Bijwaard, G. E., P. H. Franses, and R. Paap (2006). Modeling purchases as repeated events. *Journal of Business & Economic Statistics* 24, 487–502.
- Boag, J. W. (1949). Maximum likelihood estimation of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society: Series B* 11, 15–44.
- Bonnal, L., D. Fougère, and A. Sérandon (1997). Evaluating the impact of French employment policies on individual labour market histories. *Review of Economic Studies* 64, 683–713.
- Clayton, D. (1978). A model for the association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65, 141–151.
- Clayton, D. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* 47, 467–485.
- Commenges, D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis* 5, 315–327.
- Commenges, D. (2002). Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research* 11, 167–182.
- Cook, R. J. and J. F. Lawless (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer–Verlag.
- Courgeau, D. and E. Lelièvre (1992). *Event History Analysis in Demography*. Oxford: Clarendon Press.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society: Series B* 39, 1–38.
- Duchateau, L. and P. Janssen (2008). *The Frailty Model*. New York: Springer–Verlag.
- Duchateau, L., P. Janssen, I. Kezic, and C. Fortpiet (2003). Evolution of recurrent asthma event rate over time in frailty models. *Applied Statistics* 52, 355–363.
- Elbers, C. and G. Ridder (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies* 49, 403–409.
- Flinn, C. J. and J. J. Heckman (1983). Are unemployment and out of the labor force behaviorally distinct labor force states? *Journal of Labor Economics* 1, 28–42.

- Fougère, D. and T. Kamionka (2008). Econometrics of individual labor market transitions. In L. Mátyás and P. Sevestre (Eds.), *The econometrics of panel data, Fundamentals and recent developments in theory and practice*, pp. 865–905. Princeton: Princeton University Press.
- Goethals, K., P. Janssen, and L. Dutchateau (2008). Frailty models and copulas: Similarities and differences. *Journal of Applied Statistics* 35, 1071–1079.
- Heckman, J. J. and B. Singer (1984a). Econometric duration analysis. *Journal of Econometrics* 24, 63–132.
- Heckman, J. J. and B. Singer (1984b). The identifiability of the proportional hazard model. *Review of Economic Studies* 51, 231–241.
- Heckman, J. J. and B. Singer (1984c). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52, 271–320.
- Honoré, B. E. (1993). Identification results for duration models with multiple spells. *Review of Economic Studies* 61, 241–246.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer–Verlag.
- Kelly, P. J. and L. Y. Lim (2000). Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine* 19, 13–33.
- Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* 48, 795–806.
- Klein, J. P. and M. L. Moeschberger (2003). *Survival Analysis: Techniques for Censored and Truncated Data (2nd edition)*. New York: Springer–Verlag.
- Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* 47, 939–956.
- Maller, R. A. and X. Zhou (1996). *Survival Analysis with Long-term Survivors*. Chichester: John Wiley.
- Manda, S. O. M. (2001). A comparison of methods for analysing a nested frailty model to child survival in Malawi. *Australian & New Zealand Journal of Statistics* 43, 7–16.
- McGilchrist, C. A. and C. W. Aisbett (1991). Regression with frailty in survival analysis. *Biometrics* 47, 461–466.
- Nelsen, R. B. (2006). *An Introduction to Copulas (2nd edition)*. New York: Springer-Verlag.
- Nielsen, G. G., R. D. Gill, P. K. Andersen, and T. A. I. Sørensen (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* 19, 25–43.
- Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 84, 487–493.
- Oakes, D. A. (1992). Frailty models for multiple event times. In J. P. Klein and P. K. Goel (Eds.), *Survival Analysis: State of the Art*, pp. 371–379. Dordrecht: Kluwer.
- Prentice, R. L., B. J. Williams, and A. V. Peterson (1981). On the regression analysis of multivariate failure time data. *Biometrika* 68, 373–379.
- Putter, H., M. Fiocco, and R. B. Geskus (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine* 26, 2389–2430.
- Ridder, G. and T. Woutersen (2003). The singularity of the efficiency bound of the mixed proportional hazard model. *Econometrica* 71, 1579–1589.
- Rogers, A. (1975). *Introduction to Multiregional Mathematical Demography*. New York: Wiley.

- Rogers, A. (1995). *Multiregional Demography: Principles, Methods and Extensions*. New York: Wiley.
- Rondeau, V., D. Commenges, and P. Jolly (2003). Maximum penalized likelihood estimation in a gamma-frailty model. *Lifetime Data Analysis* 9, 139–153.
- Sastry, N. (1997a). Family-level clustering of childhood mortality risk in northeast Brazil. *Population Studies* 51, 245–261.
- Sastry, N. (1997b). 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.
- Schmidt, P. and A. D. Witte (1989). Predicting criminal recidivism using ‘split population’ survival time models. *Journal of Econometrics* 40, 141–159.
- Schumacher, M., M. Olschewski, and C. Schmoor (1987). The impact of heterogeneity on the comparison of survival times. *Statistics in Medicine* 6, 773–784.
- Sinha, D. and K. Dey (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association* 92, 1195–1212.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l’Institut de Statistique de L’Université de Paris* 8, 229–231.
- Therneau, T. and P. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag.
- Therneau, T. M., P. M. Grambsch, and V. S. Pankratz (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics* 12, 156–175.
- Van den Berg, G. J. (1997). Association measures for durations in bivariate hazard models. *Journal of Econometrics* 79, 221–245.
- Van den Berg, G. J. (2001). Duration models: Specification, identification, and multiple duration. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume V*, Chapter 55, pp. 3381–3460. Amsterdam: North-Holland.
- Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16, 439–454.
- Vaupel, J. W. and A. I. Yashin (1985). Heterogeneity’s ruses: Some surprising effects of selection on population dynamics. *The American Statistician* 39, 176–185.
- Wei, L. J., D. Y. Lin, and L. Weissfeld (1989). Regression analysis of multivariate failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 84, 1065–1073.
- Wienke, A. (2011). *Frailty Models in Survival Analysis*. Boca Raton: Chapman & Hall/CRC.
- Willekens, F. J. (1999). Life course: Models and analysis. In P. A. Dykstra and L. J. G. van Wissen (Eds.), *Population Issues: An Interdisciplinary Focus*, pp. 23–51. New York: Plenum Press.
- Xue, X. and R. Brookmeyer (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis* 2, 277–290.
- Xue, X. and Y. Ding (1999). Assessing heterogeneity and correlation of paired failure times with the bivariate frailty model. *Statistics in Medicine* 18, 907–918.
- Yashin, A. I., J. W. Vaupel, and I. A. Iachine (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies* 5, 145–159.
- Yau, K. K. W. and C. A. McGilchrist (1998). ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine* 17, 1201–1213.