

CALCULATION OF LTC PREMIUMS BASED ON DIRECT ESTIMATES OF TRANSITION PROBABILITIES

BY

FLORIAN HELMS, CLAUDIA CZADO AND SUSANNE GSCHLÖBL

ABSTRACT

In this paper we model the life-history of LTC-patients using a Markovian multi-state model in order to calculate premiums for a given LTC-plan. Instead of estimating the transition intensities in this model we use the approach suggested by Andersen et al. (2003) for a direct estimation of the transition probabilities. Based on the Aalen-Johansen estimator, an almost unbiased estimator for the transition matrix of a Markovian multi-state model, we calculate so-called pseudo-values, known from Jackknife methods. Further, we assume that the relationship between these pseudo-values and the covariates of our data are given by a GLM with the logit as link-function. Since the GLMs do not allow for correlation between successive observations we use instead the “Generalized Estimating Equations” (GEEs) to estimate the parameters of our regression model. The approach is illustrated using a representative sample from a German LTC portfolio.

KEYWORDS

Markovian Multi-State Model, Transition Probabilities, Aalen-Johansen Estimator, Pseudo-Values, GLM, GEE, LTC, Premium.

1. INTRODUCTION

Increasing life expectancy and the demographic development in most industrialized countries challenge both social welfare and private insurance systems (see CoE (2003)). The numbers from the Federal Statistical Office in Germany from the “Abbreviated Mortality Table for 1999/2001” show that a 60-year old male can still expect to live for another 19.5 years, a female of the same age for another 23.7 years. Due to a different family structure people tend to live more frequently alone and require at older age some kind of external assistance to manage the tasks of daily life. To pay for this external assistance, long term care insurance (LTCI) was established and added to the social welfare system. Insurance companies also have developed full or additional insurance cover for long term care (LTC).

The goal of this paper is to introduce an approach by Andersen et al. (2003) used in Biostatistics to calculate the transition probabilities of a Markovian multi-state model to actuarial scientists.

In contrast to Cox's proportional hazard model (Czado and Rudolph (2002)), where the transition probabilities are calculated from the transition intensities using a relationship given by the set of Kolmogorov forward differential equations (Haberman and Pitacco (1999)), Andersen et al. (2003) developed a method that models the transition probabilities directly. This method calculates pseudo-values based on the Aalen-Johansen estimator, an almost unbiased estimator of the transition matrix of a Markovian multi-state model. These pseudo-values are used in generalized estimating equations (GEEs), see for example Hardin and Hilbe (2003), that take in contrast to Maximum Likelihood estimation correlation between observations into account, to estimate the parameters of the model.

In this paper we provide a complete description of all necessary statistical and actuarial tools and illustrate the application of above mentioned method by Andersen et al. (2003) to a representative random sample from the German private compulsory LTCI-system.

In Germany the LTCI-benefits depend generally on the place of care (at home or in a nursing home) and on the level of care, the severeness of assistance needed: People in "Level 1" are in considerable need of care, in "Level 2" in serious need of care and in "Level 3" in extreme need of care.

This situation of different levels and places of care can be modeled by a three-state Markovian model, with states "Care at home", "Care in a nursing home" and "Death" (see Section 2). To calculate the necessary actuarial values we have to estimate the transition probabilities from our data. Thus we define in Section 3 the Aalen-Johansen estimator. Since the Aalen-Johansen estimator does not allow for covariates such as sex and age of claimants and only generates one outcome for a given set of data, we need pseudo-values to generate the data required for a regression analysis and to construct a relationship between the Aalen-Johansen estimator and covariates associated with the observations. Therefore we introduce pseudo-values in Section 4 and explain their use in generalized linear models (GLMs), which are popular extensions of linear models (McCullagh and Nelder (1989) and Fahrmeir and Tutz (1994)). The pseudo-values are calculated at different points in time involving the same observations. Therefore the assumption of independence required for GLMs does no longer hold and we have to introduce the generalized estimating equations (GEEs) in Section 5 that take correlation between observations into account.

In Section 6 we apply these methods to our data: We calculate the Aalen-Johansen estimator of a three-state model, derive the pseudo-values and thus generate the data for a regression analysis using GEEs, where we specify the logit as link-function in a linear model. With the estimates obtained from this regression analysis we calculate finally the one-year transition probabilities and use them to compute the actuarial values for a given LTC-plan and derive the required premiums. A summary with a comparison of our premiums with premiums offered by a German health insurer completes our analysis.

2. MARKOVIAN MULTI-STATE INSURANCE MODELS

We use a Markovian three-state model with states “Care at home”, “Care in a nursing home” and “Death”. For the calculation of the actuarial values the state “Active” has to be added to this model (Figure 1). We are able to calculate the transition probabilities for the area within the dotted line using our data, whereas for transition from outside this area additional information is necessary, such as incidence rates for LTC and mortality rates for active lives. We use incidence rates from “Custodial Insurance, Japan”, that certainly differ from the rates in Germany due to different LTC-definitions, and mortality rates from the “Bavarian life tables 1986-1988” (Rudolph (2000) Appendix C.1 and C.2).

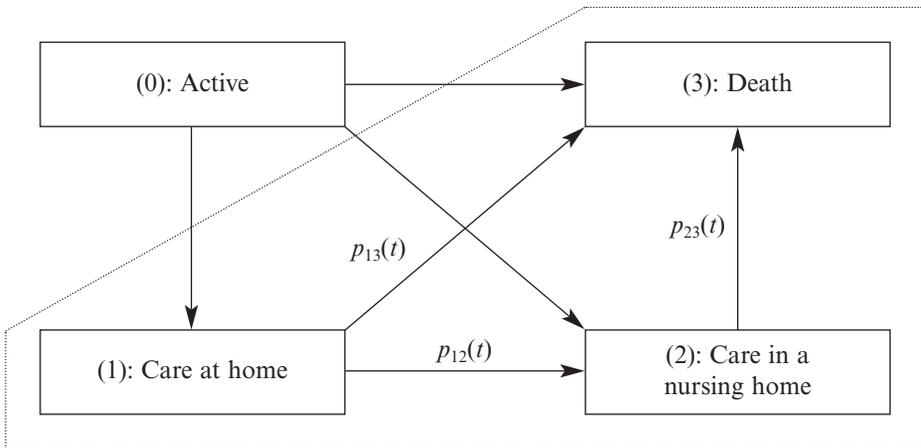


FIGURE 1: Markovian four-state Model for LTCI

The quantity $p_{gh}(t)$ denotes the one-year transition probability for a transition from state g to state h . We exclude transitions from state 2 to 1 since such transitions are observed very rarely, and we consider state 3 as an absorbing state for natural reasons, consequently $p_{21}(t) = p_{31}(t) = p_{32}(t) = 0$ for all t .

The life-history of an individual in above introduced Markovian four-state model can be described by a time-continuous Markov process $S(t)$ with state space $S = 0, 1, 2, 3$. The derivation of actuarial values both in time-continuous and time-discrete contexts, is explained in detail in Haberman and Pitacco (1999).

For our purposes we assume the time-discrete version: An insured pays a premium π in state 0, i.e. “Active”. A lump sum c_{0h} is paid at the end of the year of a transition to state h and an annuity b_h in state h until death, where h might be 1 or 2, i.e. “Care at home” or “Care in a nursing home”. Let $P_{0h}(0, t)$ be the probability of an insured to be in state “h” in year t and $P_{00}(0, t)p_{0h}(t)$ the probability for an insured in state “Active” to transfer to state h in year t . The actuarial value of above payments is the sum over the discounted outcome

multiplied by the corresponding probability of this outcome happening. Thus we obtain the following actuarial values for a z -year old insured, discounted by the factor v :

$$\begin{aligned}
 P_{0,\pi} &:= \sum_{t=0}^{\omega-z-1} P_{00}(0,t) v^t \pi; \\
 B_{0,c_{0h}} &:= \sum_{t=0}^{\omega-z-1} P_{00}(0,t) p_{0h}(t) v^{t+1} c_{0h}; \\
 B_{0,b_h} &:= \sum_{t=0}^{\omega-z-1} P_{0h}(0,t) v^t b_h
 \end{aligned}$$

where ω is the limiting age, i.e. the probability to survive beyond ω is assumed to be zero, and $P_{0h}(0,t)$ can be expressed in terms of the one-year transition probabilities $p_{0h}(t)$, $h \in \mathcal{S}$ (see Haberman and Pitacco (1999)). Outflows occur for transitions to state 1 and 2. Therefore we have to add up the actuarial values of these outflows and obtain, using the principle of equivalence, the following equation:

$$P_{0,\pi} = \sum_{t=0}^{\omega-z-1} P_{00}(0,t) v^t \pi = \sum_{h=1}^2 B_{0,c_{0h}} + \sum_{h=1}^2 B_{0,b_h}$$

Solving for π , the premium to be charged for any values of b_h and c_{0h} is obtained.

3. AALEN-JOHANSEN ESTIMATOR

For the estimation of the transition probabilities of a Markovian multi-state model we define the following quantities:

- $t_1 < t_2 < \dots$ are times, where transitions are observed.
- $d_{gh}^{(j)}$ is the number of lives, that transfer from state g to h at time t_j .
- $r_g^{(j)}$ is the number of lives in state g , alive and uncensored just prior to time t_j .

The Aalen-Johansen estimator, a non-parametric estimator for the transition matrix of a Markovian multi-state model, $P(t,u) = (P_{gh}(t,u))_{g,h \in \mathcal{S}}$, is according to Aalen and Johansen (1978) defined as

$$\hat{P}(t,u) := \prod_{j: t < t_j \leq u} \begin{pmatrix} 1 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & 1 - \frac{d_{gh}^{(j)}}{r_g^{(j)}} & \dots & \frac{d_{gh}^{(j)}}{r_g^{(j)}} & \dots & 0 \\ 0 & \dots & \dots & 1 & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & 1 & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 1 & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & 1 \end{pmatrix} \tag{1}$$

assuming that only one transition is observed at all times t_j and the transition at time t_j occurs from state g to state h . This estimator is a product of matrices at each time a transition is observed. Element (g, g) is equal to $1 - d_{gh}^{(j)}/r_g^{(j)}$ and element (g, h) is equal to $d_{gh}^{(j)}/r_g^{(j)}$, which is the number of lives that transferred from state g to h at time t_j divided by the number of lives in state g just prior to time t_j ; all other diagonal elements are equal to one, whereas other off-diagonal elements equal to zero.

If we take a two-state model with states “Alive” and “Death” the Aalen-Johansen estimator reduces to the well-known Kaplan-Meier estimator (Kaplan and Meier (1958)).

From definition (1) it follows that $\hat{P}(t, u)$ is a stochastic matrix $\forall t, u$ and one can see that the Chapman-Kolmogorov equations hold. So all requirements on the transition probabilities of a Markovian multi-state model are fulfilled. If more than one transition or transitions to different states at the same time occur, the entries of this matrix have to be modified accordingly (Helms (2003)).

Further one can show, using the theory of counting processes, more precisely Duhamel’s equation, that the Aalen-Johansen estimator is almost unbiased. It is even an unbiased estimator if the probability that $r_g^{(j)} = 0$ is equal to zero for all times t_j and states g . Given a large sample of observations this is usually the case and we refer in the following to the Aalen-Johansen estimator as an unbiased estimator. The necessary theory and a proof of this result can be found in Andersen et al. (1993).

The Aalen-Johansen estimator is calculated using the life-history of all lives, but does not take their associated covariates into account. Since one receives only a single estimate for the transition matrix, Andersen et al. (2003) suggested to calculate pseudo-values in order to generate the data required for a regression analysis and construct a relationship between the Aalen-Johansen estimator and the covariates of each live under observation.

4. PSEUDO-VALUES

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a sample of n observations, e.g. the claim-history of LTC-patients, and $\hat{P}(t, u) = s(\mathbf{x})$ the Aalen-Johansen estimator. Then the i^{th} -jackknife replication of $\hat{P}(t, u)$, the so-called “leave-one-out” estimator, is defined using the i^{th} -jackknife sample of \mathbf{x} , introduced by Efron and Tibshirani (1993) as $\mathbf{x}_{-i} := (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, as

$$\hat{P}_{-i}(t, u) := s(\mathbf{x}_{-i}).$$

This is the estimator $\hat{P}(t, u)$ without the i^{th} -observation. The i^{th} -pseudo-value of $\hat{P}(t, u)$ is defined as

$$\begin{aligned} \tilde{P}_{-i}(t, u) &:= \hat{P}(t, u) + (n - 1) \cdot (\hat{P}(t, u) - \hat{P}_{-i}(t, u)) \\ &= n \cdot \hat{P}(t, u) - (n - 1) \cdot \hat{P}_{-i}(t, u) \end{aligned}$$

Jackknife is usually used to detect outliers and to check the bias and precision of an estimator: For a good estimator $\hat{P}(t, u)$ one would assume the i^{th} -jackknife

replication of $\hat{\mathbf{P}}(t, u)$ to be close to $\hat{\mathbf{P}}(t, u)$ and consequently equal to the i^{th} -pseudo-value. Andersen et al. (2003) suggested to go a step further and perform a regression analysis using these pseudo-values $\tilde{\mathbf{P}}_i(t, u)$:

For n observations n pseudo-values can be calculated and thus the data required for a regression analysis can be generated. The missing link between the i^{th} -pseudo-value and the covariates was provided by Andersen et al. (2003). While Andersen et al. (2003) referred in their proofs to both the state and the transition matrices, they made only use of the former. Since we require the transition matrix we focus on the Aalen-Johansen estimator right from the beginning.

Let x_i be the realization of an independent and identically distributed random variable X_i , e.g. the claim-history of the LTC-patients, with expectation $\mathbf{P}(t, u)$. Thus, an unbiased estimator, i.e. the Aalen-Johansen estimator $\hat{\mathbf{P}}(t, u)$ is available for $\mathbf{P}(t, u)$.

Given i.i.d. covariates $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})^T$, with distribution function C we can write:

$$\mathbf{P}(t, u) = E[X_i] = E[E[X_i|\mathbf{Z}_i]] = \int_0^\infty E[X_i|\mathbf{Z}_i = \mathbf{z}] dC(\mathbf{z})$$

Andersen et al. (2003) then interpreted the transition matrix $\mathbf{P}(t, u)$ as the simple average of the $\mathbf{P}_i(t, u)$'s, defining $\mathbf{P}_i(t, u) := E[X_i|\mathbf{Z}_i = \mathbf{z}_i]$, where $\mathbf{z}_i, i = 1, \dots, n$ are the observed covariate values, and estimating C , the distribution of the i.i.d. covariates \mathbf{Z}_i , by its empirical distribution \hat{C} , since

$$\mathbf{P}(t, u) \approx \int_0^\infty E[X_i|\mathbf{Z}_i = \mathbf{z}] d\hat{C}(\mathbf{z}) = \frac{1}{n} \sum_{i=1}^n \mathbf{P}_i(t, u).$$

Using the fact that the Aalen-Johansen estimator is unbiased it follows that $E[\hat{\mathbf{P}}(t, u)] \approx \frac{1}{n} \sum_{i=1}^n \mathbf{P}_i(t, u)$. The same holds true if we remove one observation since the data are assumed i.i.d. and we can go through the above steps with the i^{th} -jackknife sample and obtain

$$E[\hat{\mathbf{P}}_{-i}(t, u)] \approx \frac{1}{n-1} \sum_{j \neq i} \mathbf{P}_j(t, u).$$

This provides a link between the estimator $\hat{\mathbf{P}}(t, u)$ and the quantity $\mathbf{P}_i(t, u) = E[X_i|\mathbf{Z}_i = \mathbf{z}_i]$, since using pseudo-values, Andersen et al. (2003) concluded with the above results that the i -th pseudo-value $\tilde{\mathbf{P}}_i(t, u)$ is approximately equal to the quantity, $\mathbf{P}_i(t, u)$, that takes in contrast to $\mathbf{P}(t, u)$ the covariates $\mathbf{Z}_i = \mathbf{z}_i$ into account:

$$\begin{aligned} E[\tilde{\mathbf{P}}_i(t, u)] &= E[n \cdot \hat{\mathbf{P}}(t, u) - (n-1) \cdot \hat{\mathbf{P}}_{-i}(t, u)] \\ &\approx \sum_j \mathbf{P}_j(t, u) - \sum_{j \neq i} \mathbf{P}_j(t, u) = \mathbf{P}_i(t, u) \end{aligned}$$

In other words the covariates of the i^{th} -observation are matched with the i^{th} -pseudo-value. This establishes a relationship between the covariates and the pseudo-values.

For our purpose we calculate the Aalen-Johansen estimator as defined in (1) based on the entire sample and the “leave-one-out estimator” for all observations $i = 1, \dots, n$ and time-points $t = 1, \dots, T$ giving us nT pseudo-value matrices, that is

$$\tilde{\mathbf{P}}_i(t-1, t) := \tilde{\mathbf{P}}_i(t) := (\tilde{P}_{i,gh}(t), g, h \in S) \tag{2}$$

We assume that the relationship between the pseudo-values $\tilde{\mathbf{P}}_i(t)$ and the covariates is element-wise given by a GLM (McCullagh and Nelder (1989)), where we choose the logit as link-function, the variance-function as constant and use normal errors, i.e.

$$\tilde{P}_{i,gh}(t) = \frac{\exp\{\alpha_i^{(gh)} + \mathbf{Z}_i^T \beta^{(gh)}\}}{1 + \exp\{\alpha_i^{(gh)} + \mathbf{Z}_i^T \beta^{(gh)}\}} + \varepsilon_{i,gh}(t) \tag{3}$$

where $i = 1, \dots, n, t = 1, \dots, T$ and $\varepsilon_{i,gh} \sim N(0, \sigma_{gh}^2)$. As covariates $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{i5})^T$ we include an intercept term ($\mathbf{Z}_{i1} = 1$), the “Age” (\mathbf{Z}_{i2}) at transition time t , “Sex” (\mathbf{Z}_{i3}) and the severeness of assistance needed according to the German LTC-definition denoted by “Level of care 2” (\mathbf{Z}_{i4}) and “Level of care 3” (\mathbf{Z}_{i5}). The unknown parameters are denoted by $\beta^{(gh)} = (\beta_1^{(gh)}, \dots, \beta_5^{(gh)})^T$. The parameter $\alpha_i^{(gh)}$ denotes the effect of the “Duration of Care”, when the duration is in the interval $(t-1, t]$ for the transition from state g to state h . Also other link-functions, for example the probit or the complementary loglog might be used. In this case we focus on the widely used logit which is easier to interpret, since our aim is to fit probabilities and the probit produces usually similar results (see Collett (2002)).

The problem at hand now is a problem of longitudinal data analysis: Since the same lives are observed at different time-points, correlation occurs in our model. The pseudo-values of different individuals can be assumed to be independent since the pseudo-value $\tilde{\mathbf{P}}_i(t)$ approximates the quantity $\mathbf{P}_i(t) := \mathbf{P}_i(t-1, t) = E[X_i | \mathbf{Z}_i = \mathbf{z}_i]$, which are independent random variables for all i . So for large n it seems reasonable to assume that $\tilde{\mathbf{P}}_i(t)$ are independent of $\tilde{\mathbf{P}}_j(t')$, $\forall t, t', \forall i \neq j$. Realizations from the same observation at different times are obviously dependent. Therefore we use generalized estimating equations (GEEs) instead of standard Maximum Likelihood estimation in GLMs to account for this dependency. We will give a short introduction to the GEE approach in the next section.

Finally, we would like to remark, that even tough $\hat{P}_{gh}(t)$ and $\hat{P}_{-i,gh}(t)$ are estimated transition probabilities, i.e. restricted to the interval $(0,1)$, the corresponding pseudo-value $\tilde{P}_{i,gh}(t) = n \cdot \hat{P}_{gh}(t) - (n-1) \cdot \hat{P}_{-i,gh}(t)$ does not need to satisfy this restriction for finite n . Technical difficulties might arise here, if the software package used does not set up the corresponding generalized estimation equations which will be given in Section 5 and uses for example the iterative weighted least-squares algorithm for GLMs.

5. PARAMETER ESTIMATION USING A GEE APPROACH

In the previous section we showed how to generate the data for a regression analysis using pseudo-values and justified matching the i^{th} -pseudo-value with

the covariates of the i^{th} -observation. In particular we propose to use the GLM specification (3). Since the pseudo-values of an individual at different time-points are correlated we have to allow for this correlation when parameters are estimated.

The GEE approach, introduced by Liang and Zeger (1986) and Zeger and Liang (1986) takes this into account and can be seen as the extension of Quasi-Likelihood (see Wedderburn (1974)) to longitudinal data.

For each subject i we observe the response y_{it} at times $t = 1, \dots, T$ with the corresponding covariate vector $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})$. Further we define $\mu_{it} = E[y_{it}]$ and $v_{it} = Var[y_{it}]$ assuming that they exist. In vector notation we have the subject specific response vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})^T$, $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{iT})^T$ and the subject specific design matrix

$$X_i = \begin{pmatrix} x_{i1} \\ \dots \\ x_{iT} \end{pmatrix} = \begin{pmatrix} x_{i11} & \dots & x_{i1p} \\ \dots & \dots & \dots \\ x_{iT1} & \dots & x_{iTp} \end{pmatrix} \in \mathbb{R}^{T \times p}.$$

For $T = 1$ we have the usual situation of Quasi-Likelihood. Additionally, if we specify the distribution of y_{it} as a distribution from the exponential family we recover the situation of the classical likelihood.

The relationship between the linear predictor $\eta_{it} := \mathbf{x}_{it}\boldsymbol{\beta} = x_{it1}\beta_1 + \dots + x_{itp}\beta_p$ and μ_{it} is given by the link-function $g(\cdot)$:

$$g(\mu_{it}) = \eta_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \quad \text{or} \quad \mu_{it} = g^{-1}(\eta_{it})$$

The variance v_{it} is defined as the variance-function $k(\cdot)$ evaluated at μ_{it} and divided by ϕ , the so-called scale parameter, that is $v_{it} := k(\mu_{it}) / \phi$. This gives us for each observation the following derivative of the Quasi-Likelihood function $K(y_{it}, \mu_{it})$ with respect to the mean function μ_{it} :

$$\frac{\partial K(y_{it}, \mu_{it})}{\partial \mu_{it}} = \frac{y_{it} - \mu_{it}}{v_{it}}$$

For each \mathbf{y}_i we define the same $T \times T$ correlation matrix $R(\alpha)$, which is fully parametrized by the $s \times 1$ correlation parameter vector α . We only require $R(\alpha)$ to be a correlation matrix and call it the “working” correlation matrix, since we do not expect it to be correctly specified, though we want consistent estimates and consistent variances of these estimates. We define V_i using the $T \times T$ diagonal matrix $A_i := \text{diag}(\mathbf{v}_i \phi)$ with $\mathbf{v}_i = (v_{i1}, \dots, v_{iT})^T$ as

$$V_i := A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}} / \phi. \tag{4}$$

If $R(\alpha)$ is the true correlation matrix for \mathbf{y}_i , the matrix V_i is equal to the true covariance matrix for \mathbf{y}_i , that is $V_i = Cov[\mathbf{y}_i]$, since

$$Cov[\mathbf{y}_i] = Var[\mathbf{y}_i]^{\frac{1}{2}} Corr[\mathbf{y}_i] Var[\mathbf{y}_i]^{\frac{1}{2}} = \text{diag}(\mathbf{v}_i \phi)^{\frac{1}{2}} R(\alpha) \text{diag}(\mathbf{v}_i \phi)^{\frac{1}{2}} / \phi = V_i.$$

In addition to β and ϕ we now also have to estimate α . To obtain the GEE estimates of β we solve the score equations

$$U_G(\beta) := \sum_{i=1}^n \frac{\partial l(\mathbf{y}_i; \beta)}{\partial \beta} = \sum_{i=1}^n \frac{\partial l}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta} = \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^T V_i^{-1} (\mathbf{y}_i - \mu_i) = \mathbf{0} \quad (5)$$

With $D_{it} := \partial \mu_{it} / \partial \beta$ and $D_i := (D_{i1}, \dots, D_{iT})^T = \partial \mu_i / \partial \beta \in \mathbb{R}^{T \times p}$ the score equations can be rewritten as

$$U_G(\beta) = \sum_{i=1}^n D_i^T V_i^{-1} S_i = \sum_{i=1}^n U_i(\beta, \alpha) = \mathbf{0}, \quad (6)$$

where $U_i(\beta, \alpha) := D_i^T V_i^{-1} S_i$ and $S_i = \mathbf{y}_i - \mu_i$. If only one observation is available for each subject, that is $T = 1$, this equation becomes identical to the score-like equation from Quasi-Likelihood.

In contrast to the Quasi-Likelihood approach from Wedderburn (1974), the matrix V_i in the function $U_i(\beta, \alpha)$ depends for each i not only on the parameter β but also on the parameter α . A next step is now to replace α and ϕ by any $n^{\frac{1}{2}}$ -consistent estimators. Recall that an estimator $\hat{\theta}_n$ for θ is $n^{\frac{1}{2}}$ -consistent if $\forall \eta > 0 \exists$ a constant $k(\eta)$ and an integer $n(\eta)$ such that $P(|n^{-\frac{1}{2}}(\hat{\theta} - \theta)| \leq k(\eta)) \geq 1 - \eta \forall n > n(\eta)$. Assuming that β and ϕ are known we denote the estimator for α by $\hat{\alpha}(\beta, \phi) := \hat{\alpha}(Y, \beta, \phi)$. Given β we take $\hat{\phi}(\beta) := \hat{\phi}(Y, \beta)$ as estimator for ϕ . We insert these $n^{\frac{1}{2}}$ -consistent estimators in (6) and obtain

$$U_G(\beta) \approx \sum_{i=1}^n U_i(\beta, \hat{\alpha}(\beta, \hat{\phi}(\beta))) = \mathbf{0}.$$

We define the GEE estimator of β , denoted by $\hat{\beta}_G$, as the solution of this equation. Under mild regularity conditions, it holds that

$$n^{\frac{1}{2}}(\hat{\beta}_G - \beta) \xrightarrow{\mathcal{L}} \mathbf{Z} \quad \text{as } n \rightarrow \infty$$

where \mathbf{Z} has a $N_p(\mathbf{0}, V_G)$ distribution. Further

$$V_G = \lim_{n \rightarrow \infty} n \left(\sum_{i=1}^n D_i^T V_i^{-1} D_i \right)^{-1} \left(\sum_{i=1}^n D_i^T V_i^{-1} \text{Cov}[\mathbf{y}_i] V_i^{-1} D_i \right) \left(\sum_{i=1}^n D_i^T V_i^{-1} D_i \right)^{-1}. \quad (7)$$

Here $N_p(\mu, \Sigma)$ denotes the multivariate normal distribution with mean vector μ and covariance matrix Σ . A sketch of the proof of this result can be found in Liang and Zeger (1986) while Helms (2003) provides further details. The GEE estimates of β , α and ϕ are determined by iterating between a iterative weighted least-squares algorithm based on Fisher-Scoring for β and moment estimation for α and ϕ .

To estimate the values of α and ϕ , Liang and Zeger (1986) use Pearson residuals, that can be calculated in each step of the iteration given the current value $\hat{\beta}_C$ for $\hat{\beta}_G$ as

$$\hat{r}_{it} := \frac{y_{it} - \hat{\mu}_{it}}{k(\hat{\mu}_{it})^{\frac{1}{2}}}$$

where we calculate $\hat{\mu}_{it} := g^{-1}(\mathbf{x}_{it}\hat{\beta}_C)$. Let N be again the number of all observations, that is $N = nT$. We can get a new estimate of ϕ by

$$\hat{\phi}^{-1} = \sum_{i=1}^n \sum_{t=1}^T \frac{\hat{r}_{it}^2}{N - p},$$

which is the longitudinal analogue of the familiar Pearson statistic (Zeger and Liang (1986)).

The estimation of the parameter α depends on the correlation structure selected for the working correlation matrix. Liang and Zeger (1986) present five different types: An “independent”, “exchangeable”, “unstructured”, “autoregressive (AR-I)” and “one-dependent” working correlation.

Since $\hat{\beta}_G$ and V_G are robust to the choice of the correlation structure (Liang and Zeger (1986)) we obtain, using the asymptotic normality, unbiased estimates even if the correlation structure is misspecified. Clearly, if we choose the working correlation matrix close to the true one, the estimates will be more efficient. For details on simulation studies using different correlation structures and misspecified correlation structures see Liang and Zeger (1986).

GEE estimation can be facilitated in the program by Mark X. Norleans designed for the statistical software program Splus that can be obtained from <http://lib.stat.cmu.edu/> and the program called “Oswald” developed by the Statistics Group at the University of Lancaster that can be obtained from <http://www.maths.lancs.ac.uk/Software/Oswald/> as a Splus library.

6. APPLICATION TO LTC-DATA

Now all necessary tools are given to estimate transition probabilities from a set of data containing the claim-history of LTC-patients, directly. The sample includes 5593 individuals, 3505 females and 2088 males, with ages ranging from 0 to 100 years. The median is about 83 years. Their claim-history is given with corresponding levels and places of care at the times a transition occurred. Out of the 5593 individuals, 3264 were censored during the survey due to various reasons. In total 7348 transitions could be observed. In summary we proceed as follows:

First we calculate the non-parametric Aalen-Johansen estimates $\hat{\mathbf{P}}(t-1, t)$ defined in (1) for $t = 1, \dots, T$ of the transition matrices, ignoring the covariate information. In the next step we generate the i^{th} pseudo-value matrices $\hat{\mathbf{P}}_i(t)$ defined in (2) for $i = 1, \dots, n$. Assume that the GLM given in (3) holds for $\hat{\mathbf{P}}_i(t)$ element by element. To adjust for the correlation present among observations $\hat{\mathbf{P}}_{i,gh}(t)$, $t = 1, \dots, T$, $i = 1, \dots, n$, $g, h \in S$ we apply now the GEE estimation approach described in Section 5.

For a specified working correlation matrix $R(\alpha)$ and $\phi = \sigma^2$ we estimate β , α and σ^2 by GEE. For choosing $R(\alpha)$ we first fitted an “unstructured” working

correlation. Based on the correlation matrix estimated in this approach we decided then which correlation structure would represent this estimated correlation matrix best and performed another regression using this new correlation structure as working correlation matrix.

Since for a transition from state 1 to 2, that is from “Care at home” to “Care in a nursing home”, the values in the “unstructured” correlation matrix were quite small we chose the “independence” working correlation matrix, which is fixed to the identity matrix and does not need to be estimated. In contrast, for a transition from state 1 to 3 we obtained larger estimated correlations using the “unstructured” working correlation matrix. Since the correlations tend to decrease when the distance in time increases we considered an “autoregressive (AR-I)” correlation structure to be sufficient for this transition probability. In the case of a transition from state 2 to 3 we also decided to use the “autoregressive (AR-I)” working correlation matrix, since a similar behavior in the correlation estimate of the “unstructured” working correlation matrix can be observed.

In Table 6 we summarized the GEE estimates of all possible transitions of the model shown in Figure 1, using above mentioned working correlations. The corresponding p -values are based on the asymptotic normality result presented in Section 5. In particular standard error estimates of the regression parameter estimates are based on an estimate of V_G given in (7). The validity of (7) as variance estimator relies on the independence of $\tilde{P}_i(t)$ across subjects. A justification for this assumption has been given in Section 4.

Although some p -values for the Duration of care are highly insignificant, we include all durations in the model since they are required for a consistent calculation of the actuarial values in our model.

Generally the probability for a transition from state 1 to 2 or 3 increases with age, whereas it decreases for a transition from state 2 to 3, but very slightly. The values of all three estimated transition probabilities are higher for males than the values for females as indicated by a significant positive value of the estimated regression coefficient corresponding to the covariate “Sex”.

Transitions out of state “Care at home” are more likely to happen to individuals in “Level 1”, whereas the values for individuals in “Level 2” and “Level 3” are nearly the same. This can be understood by looking at the regression coefficient estimates for “Level of care 2” and “Level of care 3” which are nearly the same for these transition probabilities. The probability of dying in state “Care at home” decreases from “Level 1” to “Level 2” before it increases to “Level 3” but does not reach the value from “Level 1” again. In contrast, for a transition from state 2 to 3 one observes decreasing transition probabilities for an increase in the severeness of care needed.

The “Duration of Care” causes rising transition probabilities for transitions out of state 1 until a duration of four years, then a decrease can be observed for three years until they increase again. For the transition from state “Care in a nursing home” to “Death” the values are very close together for all durations and only small changes occur.

With these estimates we are finally able to calculate the premiums for the LTC-plan “PET” sold by a German insurer. According to this LTC-plan the

TABLE 1

PARAMETER ESTIMATES AND THEIR *P*-VALUES FOR ALL TRANSITIONS OF THE MARKOVIAN THREE-STATE MODEL FOR LTC IN FIGURE 1 USING THE GEE APPROACH

working correlation	Independence		AR-I		AR-I	
transition from state g to h	$g = 1, h = 2$		$g = 1, h = 3$		$g = 2, h = 3$	
$\hat{\beta}_j^{(gh)}$	Values	Pr($ t >$)	Values	Pr($ t >$)	Values	Pr($ t >$)
Intercept	-5.31	0.00	-2.16	0.00	7.40	0.00
Age	0.02	0.05	0.02	0.02	-0.06	0.00
Sex = male	0.66	0.00	0.36	0.07	1.51	0.00
Level of care = 2	-1.37	0.00	-0.69	0.00	-1.96	0.00
Level of care = 3	-1.38	0.00	-0.53	0.02	-2.27	0.00
$\hat{\alpha}_j^{(gh)}$						
0 < Duration of care <= 1	0.35	0.30	0.02	0.71	0.34	0.19
1 < Duration of care <= 2	0.40	0.25	0.18	0.01	-1.08	0.00
2 < Duration of care <= 3	0.90	0.01	0.88	0.00	-0.37	0.19
3 < Duration of care <= 4	1.24	0.00	0.99	0.00	-1.68	0.00
4 < Duration of care <= 5	0.10	0.85	0.74	0.00	-1.96	0.00
5 < Duration of care <= 6	1.53	0.00	0.41	0.01	-1.06	0.01
6 < Duration of care <= 7	-0.11	0.88	0.24	0.18	-0.65	0.16
7 < Duration of care <= 8	0.25	0.60	0.82	0.00	-0.09	0.88
8 < Duration of care <= 9	0.80	0.11	1.19	0.00	-2.28	0.00
9 < Duration of care <= 10	1.26	0.00	-0.48	0.12	-0.03	0.97

TABLE 2

COMPARISON OF PREMIUMS FOR A 10 EUR DAILY ALLOWANCE

Age	Premium based on GEEs		Premium offered by German health insurer	
	Female	Male	Female	Male
20	4.94	3.85	2.12	1.70
25	6.14	4.80	2.92	2.33
30	7.70	6.05	3.90	3.10
35	9.76	7.72	5.05	4.01
40	12.51	9.97	6.44	5.13
45	16.14	12.95	8.16	6.52
50	21.01	16.95	10.39	8.36
55	27.58	22.40	13.32	10.86
60	36.48	29.88	17.31	14.40
65	48.53	40.06	22.01	18.84
70	64.55	53.62	29.04	25.71

insured receives a certain allowance depending on the level of care needed. This is for “Care at home” 25% in “Level 1”, 50% in “Level 2” and 75% in “Level 3” and for “Care in a nursing home” 100% of the allowance. Thus, the c_{0h} 's are zero, in the case of “Care at home” $b_h = 1 - 0.25 * (4 - h)$ and in the case of “Care in a nursing home” $b_h = 1, h \in \{1, 2, 3\}$, for an unit allowance.

For the calculation of the premiums we use a modified version of a C-program, which needs the benefits, interest rate and transition probabilities as input. For details see Rudolph (2000). The results for the LTC-plan “PET” obtained for a 10 EUR daily allowance and an interest rate of 3.5% can be found in the left columns of Table 2. The right columns show premiums offered by a German health insurer.

We observe higher premiums when the GEE estimation approach is used compared to the commercial premiums, but we can see that the behavior with respect to age is similar as well as the proportion between males and females, see Figure 2. It also should be noted that the incidence rates and mortality rates for “Active” individuals include administrative costs, whereas the transition probabilities do not. Therefore a direct comparison between the calculated premiums using GEEs and the commercial premiums might not be sensible. Including administrative costs in the transition probabilities would result in an even bigger gap between the rates. Further, the LTC-definition in different countries, such as Japan and Germany, varies and therefore country-specific incidence rates might be necessary.

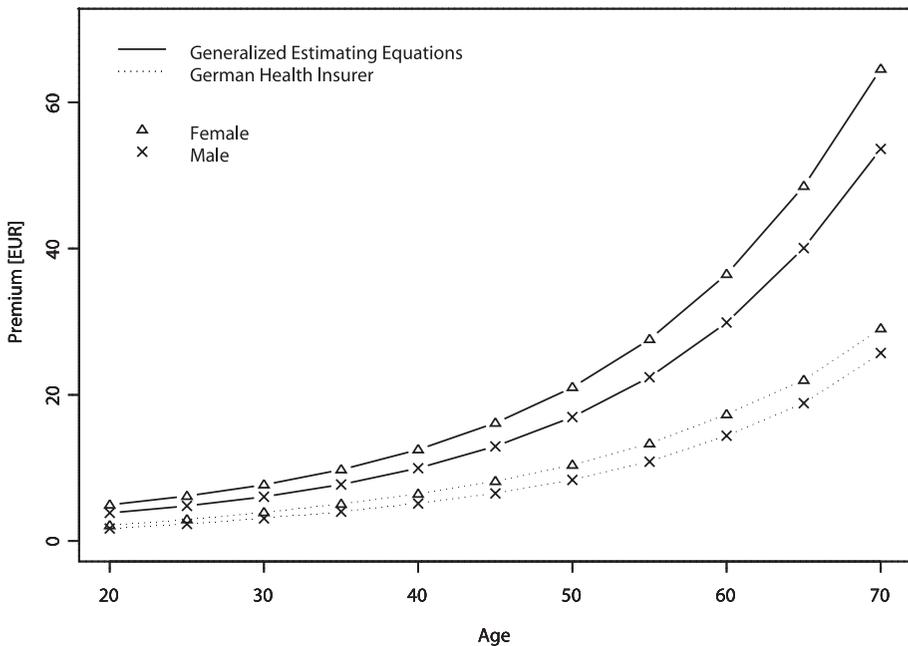


FIGURE 2: Comparison of Premiums.

7. SUMMARY AND DISCUSSION

This paper gives a detailed introduction to a method proposed by Andersen et al. (2003) for estimating transition probabilities directly and its application to a large German LTC portfolio. In particular the necessary actuarial setup for calculating premiums based on a Markovian multi-state model is provided. For the required transition probabilities pseudo-values of the Aalen-Johansen estimator which is specific to the claim-history of a LTC patient are generated. These are linked to patient specific covariates in a longitudinal GLM with normal errors and a logistic link-function. The parameters of this longitudinal GLM are estimated using a GEE approach accounting for correlation within the claim-history of a patient. These finally provide the required transition probability estimates. Diverse statistical tools ranging from survival analysis, jack-knifing methods, GLMs to GEE estimation for longitudinal regression models are introduced to give a basis for understanding how this method for estimating transition probabilities directly is working.

Even though Andersen et al. (2003) have investigated the validity of their method through simulation some further points are worth while to be addressed in further research. More precisely, in our case there are no methods available at the moment to examine the goodness-of-fit or confirm the choice of link-function in the longitudinal GLM used. The choice of time-points might also influence the results, but in our case the time-points are given, since we need one-year transition probabilities for the calculation of actuarial values. Further, more precise estimates might be obtained if the correlation matrix is chosen close to the true one.

An alternative to GEE estimation in longitudinal GLMs is to use a Bayesian approach for estimation. This will require the use of Markov Chain Monte Carlo (MCMC) methods (see for example Gilks et al. (1996) and Gamerman (1990)). Bayesian model selection criteria such as Bayes factors (see Kaas and Raftery (1995)) and the deviance information criteria (DIC) by Spiegelhalter et al. (2002) can then be applied to assess the goodness of fit and the choice of the link-function. This approach will be pursued in future research.

ACKNOWLEDGEMENTS

The second author was supported by Sonderforschungsbereich 386 *Statistische Analyse Diskreter Strukturen*, while the third author was supported by a doctoral fellowship within the Graduiertenkolleg *Angewandte Algorithmische Mathematik*, both sponsored by the *Deutsche Forschungsgemeinschaft*.

REFERENCES

- AALLEN, O.O. and JOHANSEN, S. (1978) An Empirical Transition Matrix for Non-homogeneous Markov Chains based on Censored Observations. *Scand J Statist.*, **5**, 141-150.
- ANDERSEN, P.K., BORGAN, Ø., GILL, R.D. and KEIDING, N. (1993) *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

- ANDERSEN, P.K., KLEIN, J.P. and ROSTHØJ, S. (2003) Generalised Linear Models for correlated Pseudo-Observations, with Applications to Multi-State Models. *Biometrika* **90**, **1**, 15-27.
- COE (2003) Recent demographic developments in europe 2002. www.coe.int.
- COLLETT, D. (2002) *Modelling Binary Data 2nd ed.* London: Chapman & Hall.
- CZADO, C. and RUDOLPH, F. (2002) Application of Survival Analysis Methods to Long Term Care Insurance. *Insurance: Mathematics and Economics*, **31**, 395-413.
- EFRON, B. and TIBSHIRANI, R.J. (1993) *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- FAHRMEIR, L. and TUTZ, G. (1994) *Multivariate Statistical Modelling Based On Generalized Linear Models*. New York: Springer.
- GAMERMAN, D. (1990) *Stochastic simulation for Bayesian inference*. London: Champman & Hall.
- GILKS, W.R., RICHARDSON, W.S., and SPIEGELHALTER, D. (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.
- HABERMAN, S. and PITACCO, E. (1999) *Actuarial Models for Disability Insurance*. London: Chapman & Hall.
- HARDIN, J.W. and HILBE, J.M. (2003) *Generalized Estimating Equations*, Chapman & Hall / CRC, Boca Raton, FL.
- HELMS, F. (2003) Estimating LTC-Premiums using GEEs for Peudo-Values. *Technische Universität München Diplomarbeit*.
- KAAS, R.E. and RAFTERY, A.E. (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.
- KAPLAN, E.L. and MEIER, P. (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**(282), 457-481.
- LIANG, K.-Y. and ZEGER, S.L. (1986) Longitudinal Data Analysis using Generalized Linear Models. *Biometrika*, **73**(1), 13-221.
- McCULLAGH, P. and NELDER, J.A. (1989) *Generalized Linear Models 2nd ed.* London: Chapman and Hall.
- RUDOLPH, F. (2000) Anwendungen der Überlebenszeitanalyse in der Pflegeversicherung. *Technische Universität München Diplomarbeit*.
- SPIEGELHALTER, D., BEST, N., CARLIN, B. and VAN DER LINDE, A. (2002) Bayesian measures of model complexity and fit. *J. Roy. Stat. Soc. B.*, **64**, 583-639.
- WEDDERBURN, R.W.M. (1974) Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika* **61**(3), 439-447.
- ZEGER, S.L. and LIANG, K.-Y. (1986) Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, **42**(1), 121-130.

FLORIAN HELMS

*Technische Universität München
Zentrum Mathematik
Email: florian.helms@web.de*

CLAUDIA CZADO

*Technische Universität München
Zentrum Mathematik
Email: cczado@ma.tum.de*

SUSANNE GSCHLÖBL

*Technische Universität München
Zentrum Mathematik
Email: susanne@ma.tum.de*