A relative survival model for clustered responses

Oliver Kuss∗1, Thomas Blankenburg2, and Johannes Haerting1

1 Institute of Medical Epidemiology, Biostatistics, and Informatics, University of Halle-Wittenberg, Halle (Saale), Germany
2 City Hospital Martha-Maria Dölau, Halle (Saale), Germany

Received 15 November 2004, revised 30 November 2004, accepted 2 December 2004
Published online 3 December 2004

Summary

Relative Survival is the ratio of the overall survival of a group of patients to the expected survival for a demographically similar group. It is commonly used in disease registries to estimate the effect of a particular disease when the true cause of death is not reliably known. Regression models for relative survival have been described and we extend these models to allow for clustered responses by embedding them into the class of Generalized linear mixed models (GLMM). The method is motivated and demonstrated by a data set from the HALLUCA study, an epidemiological study which investigated provision of medical care of lung cancer patients in the region of Halle in the eastern part of Germany.

Key words: Relative survival, Random effects, Generalized linear mixed models (GLMM), SAS

1 Introduction

1.1 Relative survival

Relative Survival is the ratio of the overall survival of a group of patients to the expected survival for a demographically similar group from a reference population, where this expected survival is derived from published age-, sex-, and calendar-time-specific mortality rates (Buckley, 1984). It is commonly used to estimate the effect of a particular disease when the true cause of death is not reliably known and is therefore the preferred analysis for survival experience in cancer registries thus avoiding the problem of inaccurate or non-available death certificates (Estève et al., 1990). Relative survival can be interpreted as the survival in a hypothetical population where the disease of interest is the only cause of death (Berkson and Gage, 1950) where this interpretation relies (1) on the accurate estimation of the expected survival and (2) on the patient group being a representative sample from the reference population. The relative survival approach has several attractive features, for example, it allows for claiming cure (in a statistical, not a clinical sense) in the case where the observed survival in a group of patients equals the expected survival in the population (Estève, Benhamou, and Raymond, 1994). Moreover, comparisons between international registries are facilitated because survival experience in different countries is adjusted for the respective underlying population (Estève et al., 1994).

1.2 Regression models for relative survival

Generalizing the pure description of relative survival, regression models have been proposed (Estève et al., 1990; Hakulinen and Tenkanen, 1987; Weller et al., 1999) to judge influence of prognostic or risk factors on relative survival. Owing to the principle of relative survival (as a ratio) and the relation between survival distribution and hazard function, all regression models for relative survival are necessarily additive hazard regression models. The observed hazard λ(t, x, x∗) at time t since diagnosis for a patient with covariate

∗ Corresponding author: e-mail: oliver.kuss@medizin.uni-halle.de, Phone: +00 49 345 557-3582, Fax: +00 49 345 557-3580
vector $x$ and population hazard covariates $x^*$ is modelled as a sum of the population hazard $\lambda^*(t, x^*)$ and the excess hazard $\nu(t, x)$ due to the diagnosis of the disease of interest:

$$\lambda(t, x, x^*) = \lambda^*(t, x^*) + \nu(t, x),$$

(1)

where the patient covariates $x$ are assumed to act multiplicatively on the excess hazard: $\nu(t, x) = \exp(x^T \beta)$. The population hazard covariates $x^*$ might be a subgroup of the patient covariates $x$, but this is not a necessary condition. It should be noted that using $\nu(t, x) = \exp(x^T \beta)$ implies two things. First, follow-up time $t$ must be explicitly incorporated into the covariate vector $x$. This is accomplished by splitting follow-up time in pre-specified intervals, where these intervals can be of any length and intervals of different length can be used. For actual analysis, a set of indicator variables is constructed and incorporated into the vector of ‘external’ covariates like age, sex, or stage of disease (Dickman et al., 2004). Second, assuming the effects of the external covariates being independent of time imposes the familiar proportional hazard assumption on these covariates. However, non-proportional hazards might be modelled by including interaction terms of follow-up time indicators and external covariates into the model (Dickman et al., 2004).

Below the previously mentioned regression models (Estève, Hakulinen/Tenkanen, and Weller), the Estève model (Estève et al., 1990) is the only one using explicitly the information of each individual of follow-up time indicators and external covariates into the model (Dickman et al., 2004). More specifically, we assume that $\delta_i$ is the event indicator in the $i$-th interval, respectively, for each of $i = 1, \ldots, N$ individuals. The likelihood for a general survival model, written solely in terms of the hazard function $\lambda$, is (Fahrmeir and Tutz, 2001, p. 392)

$$L(t, \delta) = \prod_{i=1}^{N} \exp\left(-\int_{0}^{t_i} \lambda(s)ds\right) \prod_{i=1}^{N} \lambda(t_i)\delta_i.$$

(2)

Inserting the additive hazard relation (1) in $L(t, \delta)$, taking logarithms and rearranging terms results in the log-likelihood function of the Estève model

$$l(t, \delta, \beta) = -\sum_{i=1}^{N} \int_{0}^{t_i} \lambda^*(s)ds - \sum_{i=1}^{N} \int_{0}^{t_i} \nu(s)ds + \sum_{i=1}^{N} \delta_i \ln[\lambda^*(t_i) + \nu(t_i)].$$

(3)

For parameter estimation, it is convenient to split $l(t, \delta, \beta)$ to obtain a separate observation for each follow-up time interval within individual $i$. Indexing these observations with $j = 1, \ldots, J_i$, and using $\nu(t_i) = \exp(x_i^T \beta)$ the contribution of each $ij$-th interval to the log-likelihood is

$$l_{ij}(\delta_{ij}, \beta) = -r_{ij} \exp(x_i^T \beta) + \delta_{ij} \ln[\lambda_{ij}^*(x_i) + \exp(x_i^T \beta)],$$

(4)

where $r_{ij}$ and $\lambda_{ij}^*$ denote the time at risk and the expected population hazard for individual $i$ in its $j$-th interval, respectively. $\delta_{ij}$ is the event indicator in the $ij$-th interval, defined analogously to $\delta_i$ above. Omitting the first term from $l(t_i, \delta_i, \beta)$ in (3) is only possible as it includes no parameters to estimate, this implying a constant hazard and thus an exponential survival distribution in each $ij$-th interval.

It is well known that the likelihood corresponding to an exponential survival distribution for censored observations is equivalent to the likelihood of a Poisson model with a specific offset (Aitkin and Clayton, 1980). As such, the relative survival model just described can also be interpreted and fitted as a generalized linear model with a binary response, a Poisson likelihood, an offset and a specific link function which is different in each $ij$-th interval (Berry, 1983; Dickman et al., 2004). More specifically, we assume that $\delta_{ij} \sim \text{Poisson}(\mu_{ij})$ with $\mu_{ij} = \lambda_{ij}^* r_{ij}$. Using $\lambda_{ij} = \lambda_{ij}^* + \exp(x_i^T \beta)$ (see (1)) and $e_{ij} = \lambda_{ij}^* r_{ij}$ as the weighted expected population hazard in the $ij$-th interval, the model equation is

$$\ln(\mu_{ij} - e_{ij}^*) = \ln(r_{ij}) + x_i^T \beta.$$

(5)
Up to now and to our knowledge there has been no generalization of the current additive relative survival regression models to allow for clustered responses. In the following we present our data set of lung cancer patients and the need for a correction for clustered responses. In chapter 3 we give an extension of the Estève model to clustered or correlated responses and chapter 4 gives the results for our data set. Chapter 5 discusses some aspects of the model and concludes. In an appendix we give a short description how the model can be estimated with SAS PROC NLMIXED.

2 The data

The data which motivated our work is from the HALLUCA (= Halle Lung Carcinoma)-study, an epidemiological study which investigated provision of medical care to lung cancer patients in the region of Halle in the eastern part of Germany (Bollmann et al., 2004). In close cooperation with the regional clinical tumor registries all lung cancer patients in the study region were recorded in a standardized way from April 1996 to September 1999, follow-up was guaranteed until September 2000. A total of 1696 lung cancer patients was observed, survival was defined as the time from clinical, histological or cytological diagnosis to death or to censoring. 1349 patients (79.5%) died until the end of follow-up, median survival in the study population was 284 days (= 9.3 months). Data on population mortality was achieved from the Statistical Office of the State of Saxony-Anhalt (‘Statistisches Landesamt Sachsen-Anhalt’), from which the study region is a part of, in two consecutive intervals of 1995-1997 and 1998-2000. Both population life tables were stratified by age (in years) and sex.

To judge influence of prognostic and risk factors on relative survival, four fixed effects patient covariates, known to be important predictors for survival, were used for regression modelling, all of them being categorical. To avoid problems with inflated p-values through model selection procedures, all covariates were chosen and categorized a priori of analysis and independently of the response values. Follow-up time was divided in five intervals of length 1 year. Table 1 gives the covariates and relative frequencies across their categories in the study population. We initially decided to model the missing values for the various covariates as separate categories, though we were aware of the dangers of using this procedure (Greenland and Finkle, 1995). However, otherwise we would have been forced to delete these observations or to use more complicated imputation methods. Having certain amounts of missing values is a common problem in cancer registries and seems to be unavoidable. We further explicitly omitted the given primary treatment as a covariate. This was to prevent from unjustified treatment recommendations which should only be derived from randomized trials.

As early as at the stage of pure descriptive analysis we had noticed a very heterogeneous survival experience in the 56 different diagnostic units in our study region. In figure 1 we give the median survival estimates with their respective 95% confidence intervals for 26 of those diagnostic units. Number of patients in our 56 study units is rather different, we have a median number of patients of 3.5 (Q1: 1, Q3: 21), with a maximum of 392 patients, and a minimum of 1 patient. To improve readability of the plot we therefore omitted all units that had less than 5 patients. Moreover, we cut the range of the y-axis (survival time) at 600 days. We notice that the two largest units (Units 1 and 2) are well within the range of the observed overall survival time. However, there are also units (7, 18, and 22) that have significantly lower survival times, and on the other hand, we find units (5, 11, 12, and 20) that have estimated median survival of above 600 days, the median estimates not even appearing on the plot. Unit 4 is a special case (the pathology department of the university hospital) where only 21 DCO (= Death Certificate Only) cases were observed, resulting in a median survival time of 0 days in this unit.

These differences in survival can maybe be explained from a clinical viewpoint. General practitioners noticing symptoms of lung cancer refer patients to those diagnostic units of which they know that the patient will also have adequate treatment after the possible confirmation of diagnosis. To be more specific, a patient in an inoperable stage of the disease will not be sent to a unit where the only possible therapy is surgery.
Table 1 Description of the HALLUCA study population in terms of relevant risk factors which were included as covariates in the various relative survival regression models. Given are the factors and their observed categories with the respective relative frequencies. Histological type abbreviations are ‘SCLC’ for small-cell lung cancer and ‘NSCLC’ for non-small lung cancer.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Category</th>
<th>Relative frequency (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>81.0</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>19.0</td>
</tr>
<tr>
<td>Age</td>
<td>&lt; 57 years</td>
<td>20.1</td>
</tr>
<tr>
<td></td>
<td>57 − &lt; 63 years</td>
<td>21.1</td>
</tr>
<tr>
<td></td>
<td>63 − &lt; 67 years</td>
<td>17.2</td>
</tr>
<tr>
<td></td>
<td>67 − &lt; 73 years</td>
<td>22.7</td>
</tr>
<tr>
<td></td>
<td>≥ 73 years</td>
<td>19.0</td>
</tr>
<tr>
<td>Histological type</td>
<td>SCLC</td>
<td>21.6</td>
</tr>
<tr>
<td></td>
<td>NSCLC</td>
<td>69.7</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>8.7</td>
</tr>
<tr>
<td>Tumor stage</td>
<td>I</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>IIIa</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td>IIIb</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>36.6</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>19.8</td>
</tr>
</tbody>
</table>

3 A relative survival model for clustered responses

To adequately adjust for the observed heterogeneous survival experience in the different diagnostic units we derived a generalization of the Estève relative survival model. In section 2 we showed how the original model can be interpreted as a generalized linear model with a binary response $\delta_{ij}$, a Poisson likelihood, an offset $\ln(r_{ij})$ and the link function $\ln(\mu_{ij} - e^{*}_{ij})$. A generalization of the model to account for clustered (or, equivalent, correlated within units) responses is now straightforward by adding a random effect for the clustering factor in the linear predictor, achieving a generalized linear mixed model (GLMM). This latter class seems to be today’s standard method for generalized linear models with clustered or correlated responses, a comprehensive overview with an emphasis on application in the medical sciences is given by Brown and Prescott (2006).

To be concrete, we assume that $\delta_{hij}$ denotes if individual $i$ from cluster $h$ ($h = 1, \ldots, H$) survives the $j$-th follow-up time interval ($\delta_{hij} = 0$) or not ($\delta_{hij} = 1$). Relying on the same assumptions as above, our relative survival model for clustered responses written as a GLMM Poisson model becomes

$$\ln(\mu_{hij} - e^{*}_{ij}) = \ln(r_{ij}) + x_i \beta + u_h,$$

with $\delta_{hij} \sim \text{Poisson}(\mu_{hij})$. The random intercept $u_h$ is assumed to be normally distributed with variance $\sigma^2_h$, $u_h \sim N(0, \sigma^2_h)$. It should be noted that $u_h$ is included in the linear predictor and we might think of the random intercept as a sum of latent covariates differing in value between the diagnostic units. Proceeding this way, the random intercept is independent of all the other terms in the likelihood function (that is, $r_{ij}$, $\beta$, $e^{*}_{ij}$) and the equivalence of the Estève model and the Poisson model carries over to the random effect case.

Copyright line will be provided by the publisher
Fig. 1  Observed median survival (with corresponding 95% confidence intervals) in the 26 diagnostic units with more than 5 patients.

The complete likelihood function is now (using the Poisson likelihood)

\[ L_{hij}(\delta_{hij}, \beta, \sigma^2_h) = \prod_{h=1}^{H} \int_{-\infty}^{\infty} \left( \prod_{j=1}^{J_h} \prod_{i=1}^{I} \exp(\delta_{hij} \ln(\mu_{hij}) - \mu_{hij}) \right) f_u(0, \sigma^2_h) \, du_h \]  

(7)

where \( f_u(0, \sigma^2_h) \) is the density of \( N(0, \sigma^2_h) \) and \( \mu_{hij} = \exp(\ln(r_{ij}) + x_i \beta + u_h) + e^*_{ij} \).

The likelihood function consists of a product of \( H \) integrals which are not analytically tractable as \( \mu_{hij} \) is a nonlinear function of the \( u_h \). As such, numerical or stochastical integration (via MCMC) are viable estimation alternatives, moreover a wealth of approximate methods has also been proposed (Agresti et al., 2002). We used adaptive Gaussian quadrature for parameter estimation, mainly because this can conveniently be programmed in SAS PROC NLMIXED, the respective code is given in the appendix.

With Poisson response data, we necessarily assume the mean of the response to equal its variance. Whenever the variance exceeds the mean the data are said to show ’overdispersion’ (Cox, 1983). Overdispersion is a problem frequently encountered in real data sets, the most prominent remedy for modelling overdispersion is to assume the response \( \delta \) to have a negative binomial distribution \( \delta \sim nb(k, \mu) \) yielding moments \( E(\delta) = \mu \) and \( \text{Var}(\delta) = \mu + k\mu^2 \) (McCullagh and Nelder, 1983). Hence, \( k \) quantifies the amount of overdispersion, with \( k = 0 \) corresponding to no overdispersion. Booth et al. (2003) gave an extension of the negative binomial model for correlated responses by also adding random effects to the linear predictor. We followed their idea and extended our random effect relative survival model by an additional overdispersion factor.
4 Results

In table 2 the results from the three different models previously described for the HALLUCA data set are given. Parameters are displayed as $\exp(\beta)$ and can be interpreted as relative excess hazards. It can easily be seen that the parameter estimates are only quantitatively, but not qualitatively different between the three models. We notice, as expected from common knowledge of lung cancer, a certain sex effect, where females have lower excess hazards, and an age effect, where excess hazards grow with increasing age. This emphasizes again the fact (Estève et al., 1990) that estimates from relative survival models are not completely adjusted for age and sex. The estimates for the other covariates also give no surprising results. As expected, we find a very strong, strictly increasing excess hazard through the various tumor stages. Concerning follow-up time the excess hazards fall with increasing time. That is, the conditional probability of survival (conditional on surviving up to the present year) rises with each year survived.

More interesting in our context, however, is to compare the different models with respect to the cluster effect and the additional overdispersion and we chose the BIC criterion (Schwarz, 1978) to compare the fit between the different models where smaller values of the BIC correspond to a better fitting model. Comparing the standard Estève model to the Poisson random effect model, we see a clear fall in the BIC values and thus the Poisson random effect model is preferred. The estimated value of the cluster variance $\hat{\sigma}^2 = 0.051$ can be deemed statistically significant ($p = 0.028$) if we compare the difference in the log-likelihoods where this $p$-value has already been halved because the null value of the parameter lies on the boundary of the parameter space (Stram and Lee, 1994). Comparing the Poisson random effect model to the negative binomial random effect model, the simple Poisson random effect model results in a better model in terms of the BIC. Relying on a LR test for the overdispersion parameter $k$, we find this difference also to be non-significant ($p = 0.264$).

To give some insight into interpretation and relevance of the estimated value of the random effect variance $\sigma^2_k$ we refer to figure 2. There we give unit-specific estimated baseline hazards for a reference patient (with all values of the covariates set to the reference level) for the 26 units from figure 1. In addition, the overall estimated baseline hazard is also given. Note that this graph does not give Kaplan-Meier estimates, that is, increasing values might be possible but were not found in our data set. As already expected from the heterogeneous values of median survival (figure 1) we see considerable spread in the units’ baseline hazards.
5 Discussion

We have shown how the relative survival regression model of Estève can be extended to account for clustered responses by embedding the model into the class of generalized linear mixed models. We learned on our own HALLUCA data that adjusting for diagnostic units indeed leads to an improved model fit in terms of the BIC. Parameter estimation in generalized linear mixed models is meanwhile straightforward and we used SAS PROC NLMIXED for this task.

Concerning the parameter estimates for the fixed effects covariates we found only small differences between the competing models, at least in clinical terms. We feel that this is due to the cancer type involved in our study. Lung cancer is a very aggressive cancer (median survival for the HALLUCA patients being 9.3 months) and survival of patients is maybe too short for differences in prognostic factors between models to emerge. We expect larger differences between fixed effect parameter estimates with less aggressive cancer types.

We deliberately excluded the given primary therapy from our list of a priori chosen covariates. This was to prevent from unjustified treatment recommendations in our observational setting. We initially speculated that the observed cluster effect can be mainly explained by the specific referral behavior of the general practitioners which send patients to those diagnostic units which have also adequate treatment options at their sites. As such, we expect the estimated cluster effects to be lower if we adjust for treatment in the respective models. However, this was not the case, in the Poisson random effect model $\hat{\sigma}_h^2$ raised from 0.051 to 0.070 (0.040). In table 3 we report on the BIC values for the four different models with the additional covariate of treatment. We see that treatment is considered very influential, we find considerably lower BIC values as compared to those from table 2. We also note that the ordering of models with the Poisson model ranking the highest and the negative binomial model second best is not preserved. The
Table 2 Results from fitting the HALLUCA data set via the different methods. All models were fit with SAS PROC NLMIXED. In the case of the random effect models adaptive Gaussian quadrature with 50 quadrature points was used. Given are the estimated parameters with standard errors in parentheses after exponentiation.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Category</th>
<th>Standard Poisson</th>
<th>Negative binomial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estève Random effect</td>
<td>Random effect</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>-1.212 (0.130)</td>
<td>-1.245 (0.134)</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>0.870 (0.064)</td>
<td>0.865 (0.066)</td>
</tr>
<tr>
<td>Age</td>
<td>57− &lt; 63 years</td>
<td>1.029 (0.091)</td>
<td>1.026 (0.093)</td>
</tr>
<tr>
<td></td>
<td>63− &lt; 67 years</td>
<td>1.131 (0.104)</td>
<td>1.130 (0.107)</td>
</tr>
<tr>
<td></td>
<td>67− &lt; 73 years</td>
<td>1.186 (0.104)</td>
<td>1.175 (0.106)</td>
</tr>
<tr>
<td></td>
<td>&gt;= 73 years</td>
<td>1.308 (0.122)</td>
<td>1.322 (0.127)</td>
</tr>
<tr>
<td>Histological type</td>
<td>SCLC</td>
<td>1.136 (0.078)</td>
<td>1.131 (0.081)</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>0.882 (0.102)</td>
<td>0.882 (0.105)</td>
</tr>
<tr>
<td>Tumor stage</td>
<td>II</td>
<td>1.668 (0.305)</td>
<td>1.688 (0.315)</td>
</tr>
<tr>
<td></td>
<td>IIIa</td>
<td>1.929 (0.276)</td>
<td>1.964 (0.286)</td>
</tr>
<tr>
<td></td>
<td>IIIb</td>
<td>2.852 (0.375)</td>
<td>2.927 (0.395)</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>4.376 (0.530)</td>
<td>4.531 (0.567)</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>1.794 (0.236)</td>
<td>1.836 (0.248)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Category</th>
<th>Standard Poisson</th>
<th>Negative binomial</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estève Random effect</td>
<td>Random effect</td>
<td></td>
</tr>
<tr>
<td>Year of Follow-up</td>
<td>2</td>
<td>0.816 (0.059)</td>
<td>0.851 (0.065)</td>
<td>0.841 (0.067)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.484 (0.070)</td>
<td>0.851 (0.065)</td>
<td>0.503 (0.077)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.407 (0.108)</td>
<td>0.447 (0.121)</td>
<td>0.433 (0.119)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.247 (0.222)</td>
<td>0.274 (0.247)</td>
<td>0.253 (0.234)</td>
</tr>
</tbody>
</table>

\( \hat{\sigma}_h^2 \) \( \hat{k} \)

| -2LogL | 6725.4 | 6721.7 | 6720.5 |
| BIC    | 6860.4 | 6846.0 | 6851.6 |

negative binomial model is now considered the least fitting model in terms of the BIC. Interestingly, in the negative binomial random effect model we found \( \hat{\sigma}_h^2 = 0 \) and \( \hat{k} = 0.228 (0.081) \), that is, the whole cluster effect is absorbed by the overdispersion factor \( k \). We finally judge that the referral behavior of the general practitioners is not responsible for the clustering of survival times.

Focussing on methodical issues, several extensions of the proposed model are possible and might be good starting points for future work. In principle, all extensions of the standard relative survival models (non-proportional hazards (Bolard et al., 2001), more flexible survival distributions in the annual intervals (Weller et al., 1999), models with additional cure fraction (De Angelis et al., 1999) or additive covariate
Table 3  Results from fitting the HALLUCA data with the additional covariate of treatment. Given are the BIC values.

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>Poisson</th>
<th>Negative binomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estève effect Random</td>
<td>6656.5</td>
<td>6636.9</td>
<td>6680.0</td>
</tr>
<tr>
<td>effect Random effect</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

effects (Lambert et al., 2005)) could also be used for the random effect models. Extending the model by non-proportional hazards might be the most important extension of the model (Dickman et al., 2004). This can easily be done by inclusion of interaction terms of covariates with follow-up time into the model and would allow the effect of covariates to vary over time.

There are also several possibilities for extension of the random effects part. By using only a random intercept which is independent of follow-up time, we implicitly assume that the cluster effect is constant over time. However, we could also, similar to the fixed effects covariates, include interaction terms of the random intercept effect with follow-up time to allow for the differences between clusters to vary over time. Extension to spatial effects also is straightforward if we seek to model geographical variation. Only recently, Yu et al. (2004) proposed a model with a random cluster effect to estimate region-specific shrunken relative excess risks. The difference to our model is that they do not allow for covariates and that they avoid computational complexity by specifying a gamma distribution for the region-specific excess risks. In our data set we were not able to model the spatial distribution of our diagnostic units, due to confidentiality issues. For the same reason we were also not allowed to model cluster specific covariates, however, the model would easily allow for that. Further, we could generalize the assumption of the random effects distribution from normal to more flexible distributions. A possible option might be a non-parametric maximum likelihood approach where the distribution of the random cluster effect is estimated in parallel to the fixed effects (Aitkin, 1999). Finally, it would also be straightforward to include random covariate effects, maybe to model additional measurement error.

Computing time is always an issue with random effect models, and especially if parameters are estimated by numerical quadrature. The NLMIXED code for the Poisson RE model in the appendix required 45 seconds to converge on an IBM desktop PC (Pentium 4, 3 GHz, 2 GB RAM) for the HALLUCA data. However, computing time for the negative binomial RE model was 15 minutes, 45 seconds, and we are well aware of the fact that, in general, data sets for relative survival analysis are much larger than our HALLUCA data set. We can think of two possible remedies for saving computing time. First, we might return to an approximate ML method (for example, SAS PROC GLIMMIX offers PQL estimation, SAS code is available from the author). Second, it would also be straightforward (Dickman et al., 2004, Lambert et al., 2005) to collapse data within follow-up time intervals to give one single observation for each covariate pattern. That is, \( \delta_i, e_i^* \) and \( r_i \) are summed for each follow-up time interval \( j \) and covariate pattern separately. However, we emphasize that for our random effect extension the different clusters also define different covariate patterns and so there might be only a minor reduction in observation numbers. In our HALLUCA data set collapsing across covariates and clusters would result in a reduction from 1696 patients to 770 covariate patterns. We also like to mention that collapsing data would require a different likelihood function and equivalence with the Estève model would be lost.

To judge the robustness of our results we finally performed some sensitivity analyses. First, we estimated the Poisson RE model by a different estimation method, using MCMC via WinBUGS. Second, we tried an additional partition of follow-up time with an increased number of follow-up time intervals. In this new partition we divided the first follow-up year in four equidistant intervals, the second follow-up year in two equidistant intervals, and used annual intervals thereafter. In all cases the results for the external covariates, the random effect and (in the second case) also for the estimated hazards over follow-up time were remarkably comparable.
To conclude, we recommend routine application of this model (at least of the Poisson random effect model) whenever there is suspected heterogeneity in survival across clusters in the data set at hand. This has two advantages. First, the normal distribution of the random cluster effect then accounts elegantly for potential latent factors in the specific cluster as their sum can be expected to be normally distributed due to the central limit theorem. Second, we get more realistic (and, in general, larger) standard errors for the parameter estimates.

**Acknowledgements** We are grateful to the HALLUCA study group for helpful discussion during the project. This work profited a lot from SAS code and tutorial slides that Paul Dickman made available on his homepage (www.pauldickman.com).

### 6 Appendix

In this appendix we give a short description how the Poisson random effect model for relative survival can be estimated with SAS PROC NLMIXED.

The following data step creates the data set halluca. To save space only the most relevant variables are given, especially we refer to only one fixed effect (sex).

```sas
DATA halluca;
  INPUT i j r_ij e_star_ij delta_ij DUnit fu2 fu3 fu4 fu5 sex ...;
  CARDS;
  1 1 1.0000 0.0362 0 1 0 0 0 0 1 ... 
  1 2 1.0000 0.0370 0 1 1 0 0 1 ... 
  1 3 1.0000 0.0388 0 1 0 1 0 0 1 ... 
  1 4 0.5890 0.0256 0 1 0 0 1 0 1 ... 
  2 1 0.3480 0.0018 1 1 0 0 0 0 0 ... 
  3 1 1.0000 0.0118 0 2 0 0 0 0 0 ... 
  3 2 1.0000 0.0133 0 2 0 1 0 0 0 ... 
  3 3 0.3178 0.0048 1 2 0 0 1 0 0 ... 
  ...;
RUN;
```

The data set is organized to have a single line for each $ij$-th interval. For example, the first patient ($i=1$) from diagnostic unit 1 ($DUnit=1$) contributes 4 observations, the annual follow-up interval being indicated by the dummy variables $fu2$ to $fu5$. That is, she ($sex=1$) was observed for more then 3 years where observation time in her fourth year was only 58.9% ($r_{ij}=0.589$) and then she was censored ($delta_{ij}=0$). The second patient ($i=2$), also from diagnostic unit 1, was male, was observed for only 34.8% of his first year ($r_{ij}=0.348$) and then died ($delta_{ij}=1$). Patient 3 from diagnostic unit 2 died after having survived 31.8% of his third year.

Parameter estimates can be calculated with the NLMIXED code below. The NLMIXED procedure has been part of the SAS System since Version 7. It attempts to maximize the likelihood function of nonlinear mixed effects models directly by numerical integration methods, more precisely by adaptive Gaussian quadrature. The PARMS statement is used to initialize values of the parameters to be estimated. The MODEL statement declares the response variable and the likelihood structure which will be maximized. The RANDOM statement identifies the random effect $u_h$ as being normally distributed and determines (via the SUBJECT statement) the variable containing the different diagnostic units. The code between the PARMS statement and the MODEL statement constructs the mean of the Poisson distribution.

```sas
PROC NLMIXED DATA=halluca;
  * Initialize the parameters;
  PARMS int=-1.21 b_fu2=-0.20 b_fu3=-0.73 b_fu4=-0.90 b_fu5=-1.40
    b_sex=-0.14 ... variance_u_h=1;
```

Copyright line will be provided by the publisher
Define the linear predictor with the fixed effects and the random effect $u_h$:

$$x_{ibeta} = \text{int} + b_{fu2} \cdot fu2 + b_{fu3} \cdot fu3 + b_{fu4} \cdot fu4 + b_{fu5} \cdot fu5 + b_{sex} \cdot sex + \ldots + u_h;$$

Transform the linear predictor to incorporate the $r_{ij}$ and the $e_{star_{ij}}$:

$$mu_{ijh} = \exp(x_{ibeta} + \log(r_{ij})) + e_{star_{ij}};$$

Specify the Poisson likelihood:

MODEL delta_{ij} \sim POISSON(mu_{ijh});

Declare the random effect:

RANDOM $u_h$ \sim NORMAL(0, variance_u_h) SUBJECT=DUnit;

RUN;

References


