A relative survival model for clustered responses - Comparing SAS PROC NLMIXED and WinBUGS for parameter estimation

Oliver Kuß*, Thomas Blankenburg**, Johannes Haerting*

*Institute of Medical Epidemiology, Biostatistics, and Informatics, University of Halle-Wittenberg, Halle (Saale)

**City Hospital Martha-Maria Dölau, Halle (Saale)

„Markov-Chain-Monte-Carlo - Methoden und Anwendungen“, Workshop der AG „Bayes-Methodik“ der DR der IBS, Mainz, 1.12.2006

Contents

- Relative Survival
- Motivation
- A Relative Survival Model for Clustered Responses
- Computation
- Results
- Conclusion

Relative Survival I: Definition

For a group of patients:

$$\text{Relative Survival} = \frac{\text{Observed Survival}}{\text{Expected Survival}}$$

where expected survival is derived from published age-, sex-, and calendar time-specific mortality rates.

Interpretation: Relative Survival describes survival in a hypothetical population where the disease of interest is the only cause of death (and is therefore the standard method in disease registries).

Relative Survival II: Properties

Advantages:

- Information on cause of death is not needed.
- Cure (in a statistical sense) can be described.

Disadvantages:

- Information on mortality of the general population is needed.
- Patients group must be a sample from the general population.

Relative Survival III: Regression Models

Generalizing the pure description, regression models for relative survival have been proposed to describe influence of prognostic and risk factors (Hakulinen/Tenkanen, 1987; Estève et al., 1990)

Owing to the principle of relative survival these are all *additive* hazard models:

$$\lambda_{obs} = \lambda_{pop} + \lambda_{excess} \tag{1}$$

with λ_{obs} = observed hazard, λ_{pop} = population hazard, $\lambda_{excess} = \exp(X\beta)$: excess hazard, function of the covariates

Compare this to the Cox model: $\lambda_{obs} = \lambda_0 \exp(X\beta)$ (*multiplicative* model)

Relative Survival IV: The Estève model as a GLM I

Dickmann et al., 2004, showed that the Estève model can be written as a GLM with a binary response, a Poisson likelihood, an offset and a specific individualized link function.

Notation: Given are $i = 1, \ldots, N$ patients, each one observed for $j = 1, \ldots, J_i$ annual intervals.

δ_{ij} is the event indicator in the ij-th interval ($\delta_{ij} = 1$ refers to dying, $\delta_{ij} = 0$ to surviving).

r_{ij} denotes the time at risk (in %), and $e^*_{ij} = (\lambda_{pop} * r_{ij})$ the weighted population hazard in the ij-th interval.

Relative Survival V: The Estève model as a GLM II

The model equation is

$$\ln(\mu_{ij} - e^*_{ij}) = \ln(r_{ij}) + x_i\beta. \tag{2}$$

There is no correlation induced by the J_i observations per proband!

Model assumes proportional hazard assumption for the covariates and constant hazard in annual intervals!

Motivation I: The HALLUCA study

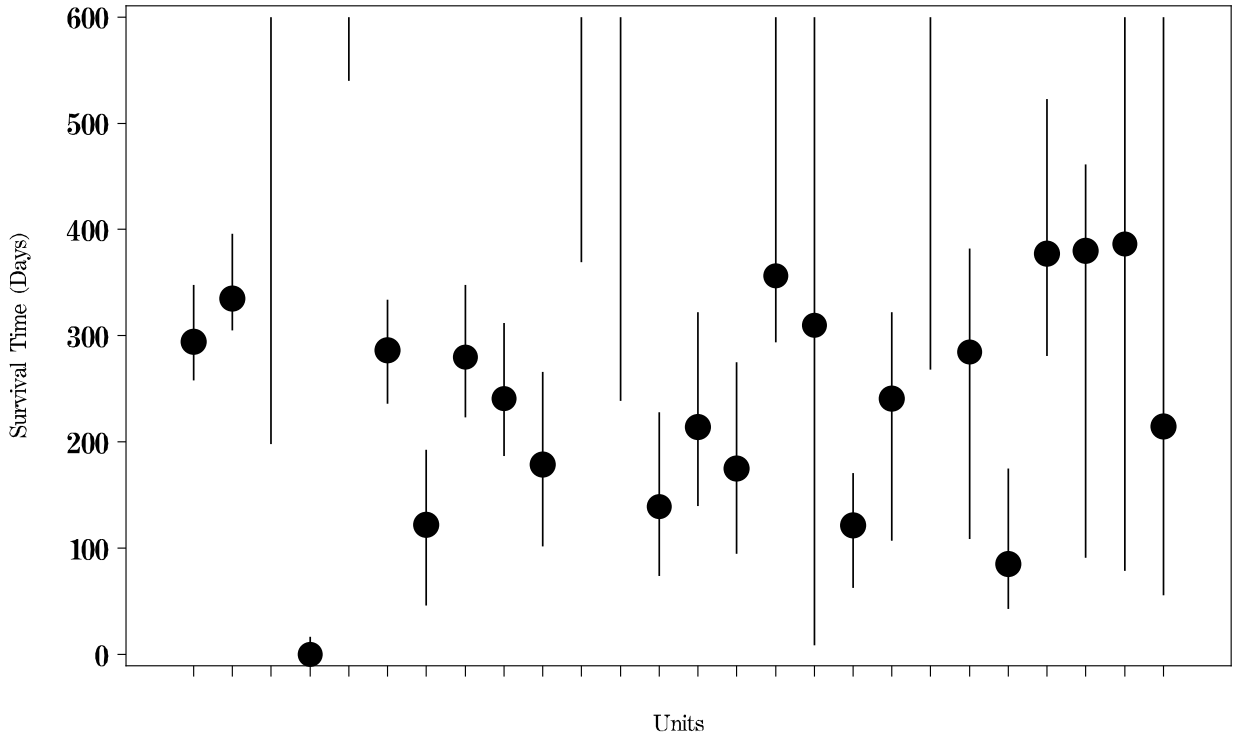
HALLUCA-(= Halle Lung Carcinoma)-study, an epidemiological study which investigated provision of medical care of lung cancer patients in the region of Halle.

Standardized recruiting of all lung cancer patients from 4/1996 to 9/1999, follow-up until 9/2000.

N=1696 lung cancer patients, 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').

Motivation II: Heterogeneous Survival in Diagnostic Units



Observed median survival (with 95% confidence intervals) in the 26 diagnostic units with more than 5 patients.

A Relative Survival Model for Clustered Responses I

Generalize Dickman's model to account for clustered (or, equivalent, correlated within units) responses by adding a random effect for the diagnostic unit in the linear predictor, achieving a generalized linear mixed model (GLMM).

To be concrete, δ_{hij} denotes the event indicator for individual i from cluster h ($h = 1, \ldots, H$), then

$$\ln(\mu_{hij} - e^*_{ij}) = \ln(r_{ij}) + x_i\beta + u_h \tag{3}$$

The random intercept u_h is assumed to be normally distributed with variance σ^2_h, $u_h \sim N(0, \sigma^2_h)$.

A Relative Survival Model for Clustered Responses II

Parameter estimation in this random effects relative survival models, as in all GLMM, is complicated by the fact that the likelihood function consists of H integrals which are not analytically tractable.

We used numerical (SAS PROC NLMIXED) and stochastical integration (WinBUGS) for parameter estimation.

Additional complication: individualized link functions

Computation I: SAS PROC NLMIXED

```
proc nlmixed data=... ;
     parms int=-1 b_stage2=0.5 b_stage3=0.7 ...  sd2=1;

     Xbeta = int + b_stage2*stage2 + b_stage3*stage3 + ... + u_h;

     Mu    = exp(Xbeta+log_r_ij) + e_ij;

     loglike = delta_ij*log(Mu) - Mu;
     model delta_ij ~ general(loglike);
     random u ~ normal(0,sd2_h) subject=DiagnosticUnit;
run;
```

Computation II: WinBUGS

```
model; {
    for (i in 1:N){
        Xbeta[i] <- int + b_stage2*stage2[i] + b_stage3*stage3[i] + ...
                        + u_DiagnosticUnit[DiagnosticUnit[i]];

        log(mu[i]) <- log(r_ij[i]) + Xbeta[i]+ exp(e_ij[i]);
        delta_ij[i] ~ dpois(mu[i]);
                    }
    for (h in 1:H){
         u_DiagnosticUnit[h]~ dnorm(0.0000, tau_DiagnosticUnit);
                   }

    tau_DiagnosticUnit ~ dgamma(0.001,0.001);
    var_DiagnosticUnit <- 1 / tau_DiagnosticUnit;

   # priors
     int~ dnorm(0.0,1.0E-6) b_stage2~ dnorm(0.0,1.0E-6) ...
         }
```

Results I: Fixed effects (selected)

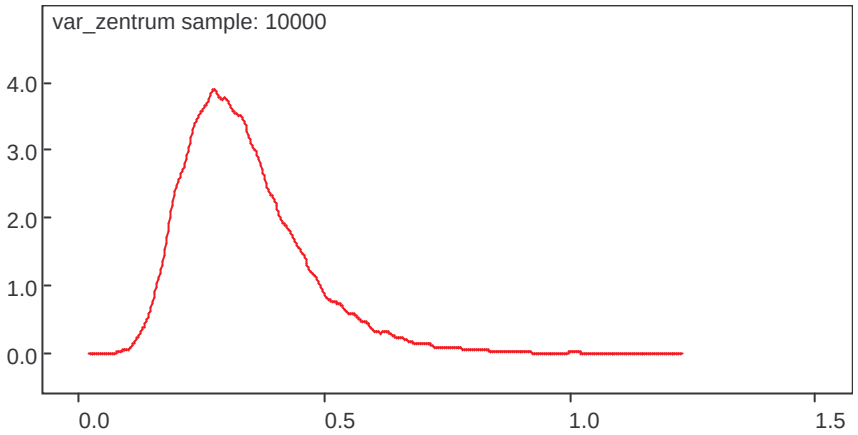
Covariate	Category	PROC NLMIXED β (SE)	WinBUGS * β (SE) **
Gender	Female	-0.161 (0.076)	-0.152 (0.073)
Age	>= 65 years	0.118 (0.060)	0.131 (0.057)
Histological type	SCLC	0.120 (0.071)	0.091 (0.068)
	Missing	-0.143 (0.120)	-0.140 (0.115)
Performance status (ECOG)	3-4	0.714 (0.114)	0.652 (0.110)
	Missing	0.145 (0.065)	0.158 (0.065)

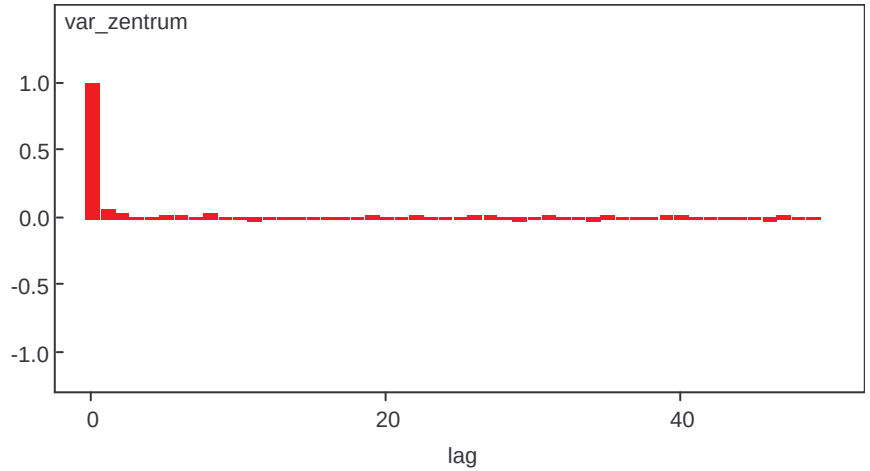
* 10.000 runs burn-in, 100.000 runs, thinning 1:10, non-informative priors

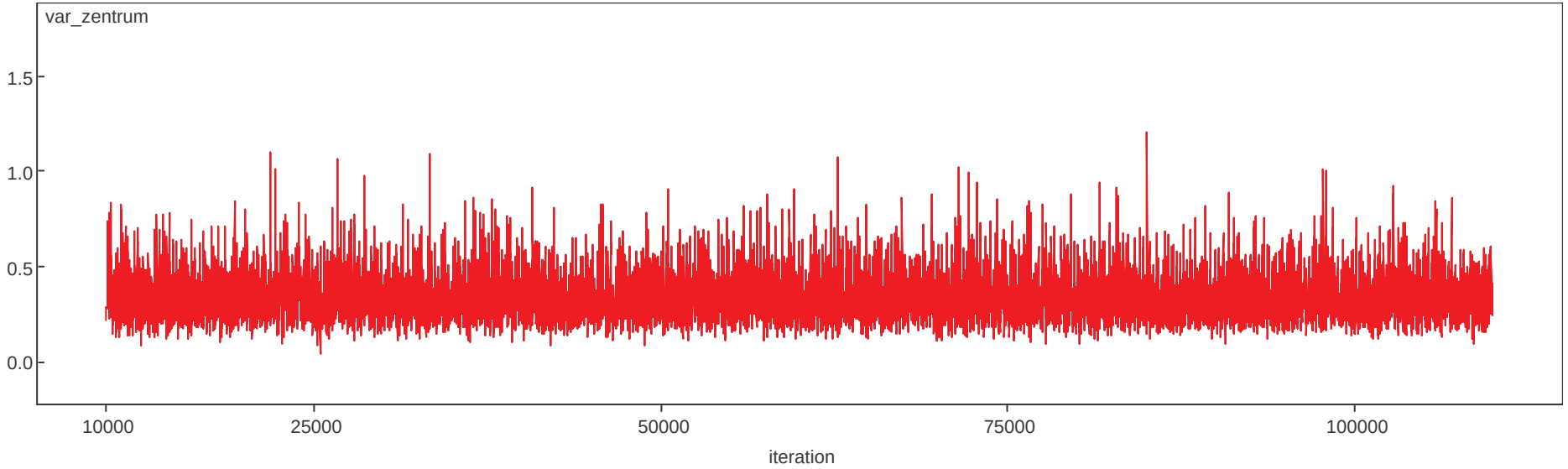
** Posterior mean

Results II: Random effects

Parameter	PROC NLMIXED	WinBUGS
σ_h^2	0.053 (0.037)	0.338 (0.125)


var_zentrum sample: 10000
4.0
3.0
2.0
1.0
0.0
0.0
0.5
1.0
1.5


var_zentrum
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


var_zentrum
1.5
1.0
0.5
0.0
10000
25000
50000
75000
100000
iteration

Conclusion I

- A relative survival model for clustered responses can be easily defined by embedding Dickman's version of the Estève version into the class of generalized linear mixed models.

- Parameter estimation is straightforward, SAS PROC NLMIXED and WinBUGS can be used (besides others).

- For our data set fixed effects estimates in NLMIXED and WinBUGS did not differ, but random effects estimates did. This is compatible with our experience on other data sets.

Conclusion II

- Coding complicated models in different software packages is a good idea and gives impression of robustness of results.

- Advantages PROC NLMIXED: ease of data handling, computation time

- Advantages WinBUGS: allows generalization to more random effects.

References

1. Estève J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: Elements for further discussion. Stat Med 1990; 9:529-538.

2. Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. Appl Stat 1987: 36:309-317.

3. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression Models for Relative Survival. Stat Med 2004; 23:51-64.