A relative survival model for clustered responses

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)

Herbstworkshop „Statistische Methoden für korrelierte Daten“,
Bochum, 23.11.2006

Contents

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

Relative Survival I: Definition

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

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).
- Measure for excess mortality in the patient group.

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, baseline hazard λ_0 is not estimated)

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 binomial 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

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

The response δ_{ij} can only take values 0 or 1 and the sum over all j for a specific patient i can be at most equal to 1 ($\sum_{j=1}^{J_i} \delta_{ij} = 0$ in the case of censoring or $\sum_{j=1}^{J_i} \delta_{ij} = 1$ in the case of death).

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

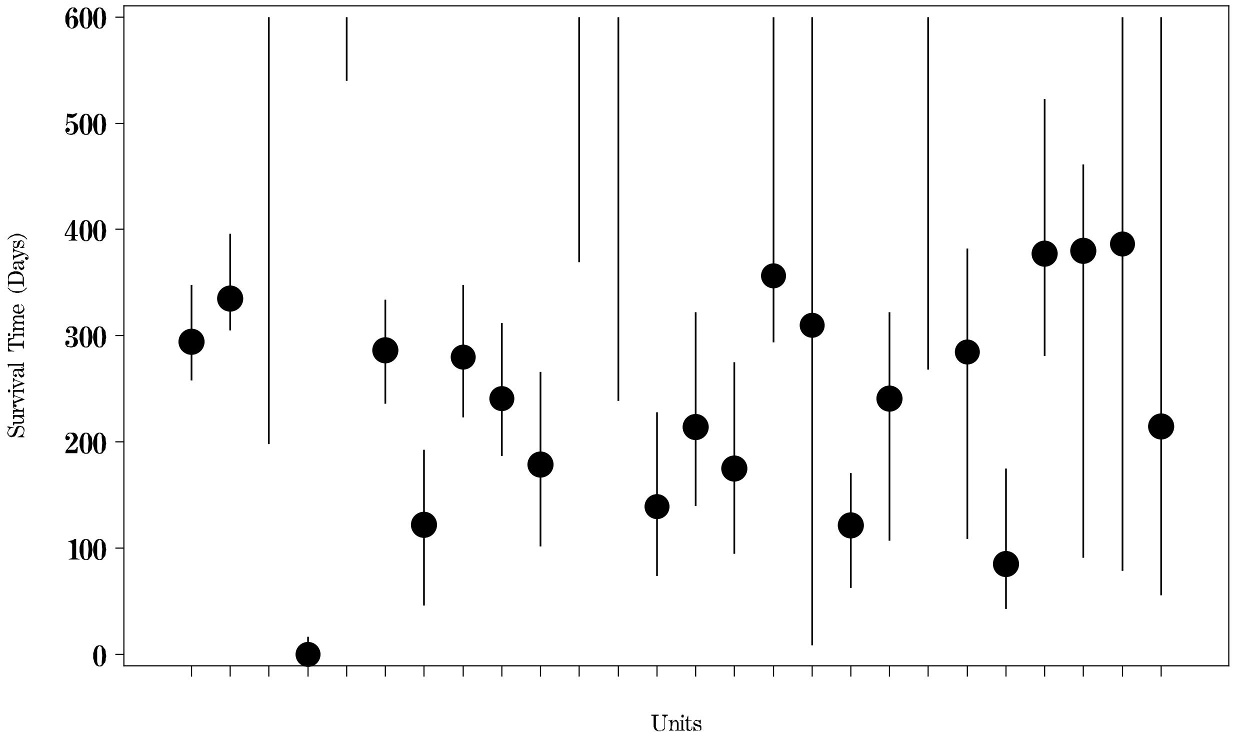
HALLUCA-(= Halle Lung Carcinoma)-study, an epidemiological study which investigated provision of medical care of lung cancer patients in the region of Halle.

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 if individual i from cluster h $(h = 1, \ldots, H)$

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

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

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

Results I: Fixed effects

Covariate	Category	Standard Estève β (SE)	Clustered Responses β (SE)
Gender	Female	-0.158 (0.074)	-0.161 (0.076)
Age	>= 65 years	0.120 (0.058)	0.118 (0.060)
Histological type	SCLC	0.119 (0.069)	0.120 (0.071)
	Missing	-0.142 (0.116)	-0.143 (0.120)
Performance status (ECOG)	3-4	0.692 (0.108)	0.714 (0.114)
	Missing	0.136 (0.062)	0.145 (0.065)
Tumor stage	II	0.519 (0.183)	0.528 (0.187)
	IIIa	0.672 (0.143)	0.683 (0.146)
	IIIb	1.051 (0.132)	1.074 (0.136)
	IV	1.452 (0.122)	1.485 (0.126)
	Missing	0.553 (0.134)	0.575 (0.138)
Year of Follow-up	2	-0.202 (0.073)	-0.160 (0.078)
	3	-0.732 (0.145)	-0.670 (0.151)
	4	-0.906 (0.265)	-0.816 (0.272)
	5	-1.392 (0.896)	-1.290 (0.899)

Results II: Random effects, Model fit

Parameter	Standard Estève	Clustered Responses
σ_h^2	–	0.053 (0.037)
-2LogL	6693.0	6690.1
BIC	6820.2	6807.5

p(LR-test, $\sigma_h^2 = 0$, $\chi^2_{obs} = 2.888$): 0.045

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;

     Mu    = exp(Xbeta+log_risk) + e;

     loglike = survival_status*log(Mu) - Mu;
     model survival_status ~ general(loglike);
     random u ~ normal(0,sd2) subject=DiagnosticUnit;
run;
```

WinBUGS

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

        logrisk[i]<- log(risk[i]);
        log(mue[i]) <- logrisk[i] + Xbeta[i]+ exp(e[i]);
        survival_status[i] ~ dpois(mue[i]);
                   }
    for (h in 1:H){
         u_zentrum[h]~ dnorm( 0.0000, tau_zentrum);
                  }
    tau_zentrum ~ dgamma(0.001,0.001);
    var_zentrum <- 1 / tau_zentrum;

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

Conclusion

- 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.
- For our data set we achieved a better fit with the new model but clinical conclusions remained essentially unchanged.
- Several extensions are possible: Using a negative binomial likelihood, more hierarchical levels, random covariates, other estimation methods

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.