Semiparametric Multi-State Models

Thomas Kneib

Department of Mathematics
Carl von Ossietzky University Oldenburg

joint work with

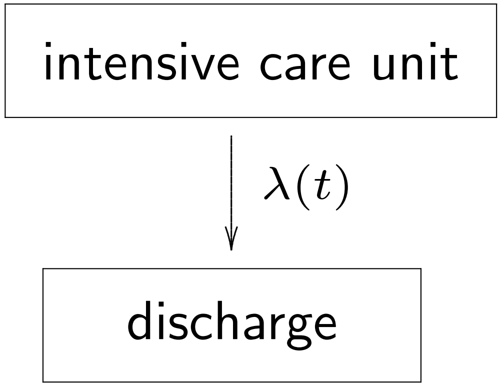
Andrea Hennerfeind, Jan Beyersmann, Martin Schumacher & Petra Gastmeier

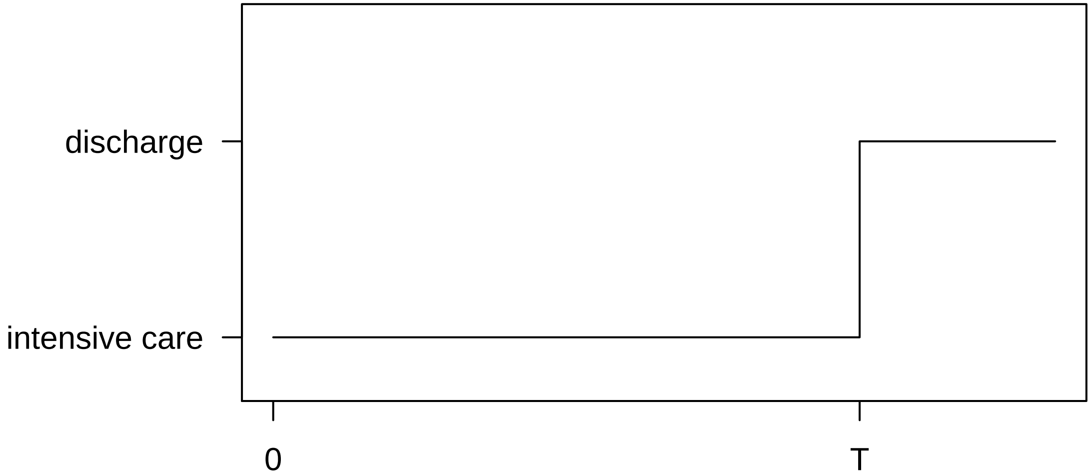
CARL
VON
OSSIETZKY
universität OLDENBURG
FAKULTÄT V
MATHEMATIK UND
NATURWISSENSCHAFTEN

Prague, 25.8.2009

Multi-State Models

- Multi-state models describe the temporal evolution of discrete phenomena in continuous time.
- Simple special case: Survival times or more generally duration times.

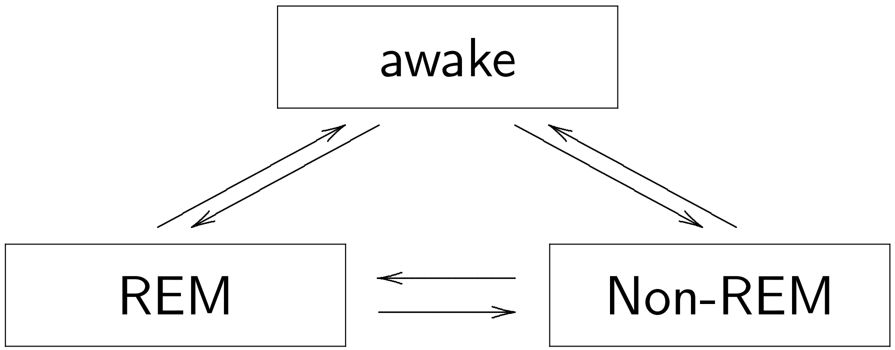




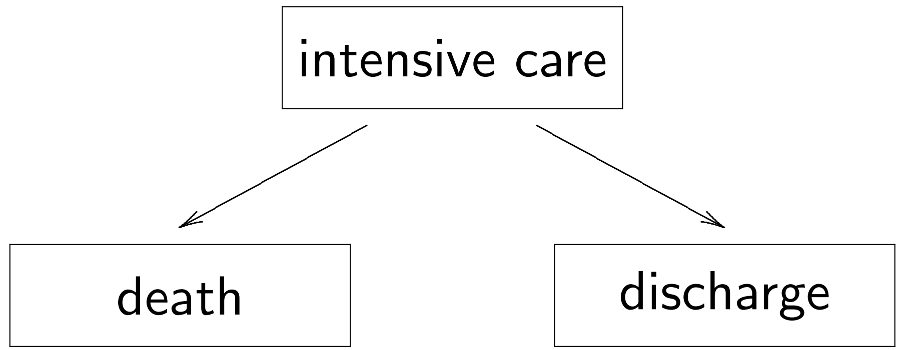
- We are interested in transition intensities at time t, i.e. rates

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}.$$

- Multi-state models describe several transitions between more than two states.
- Examples:



recurrent events

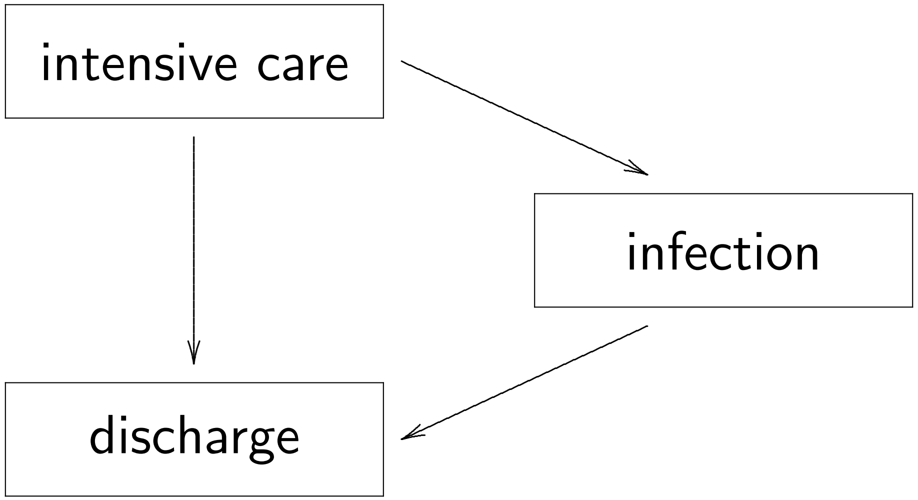


competing risks

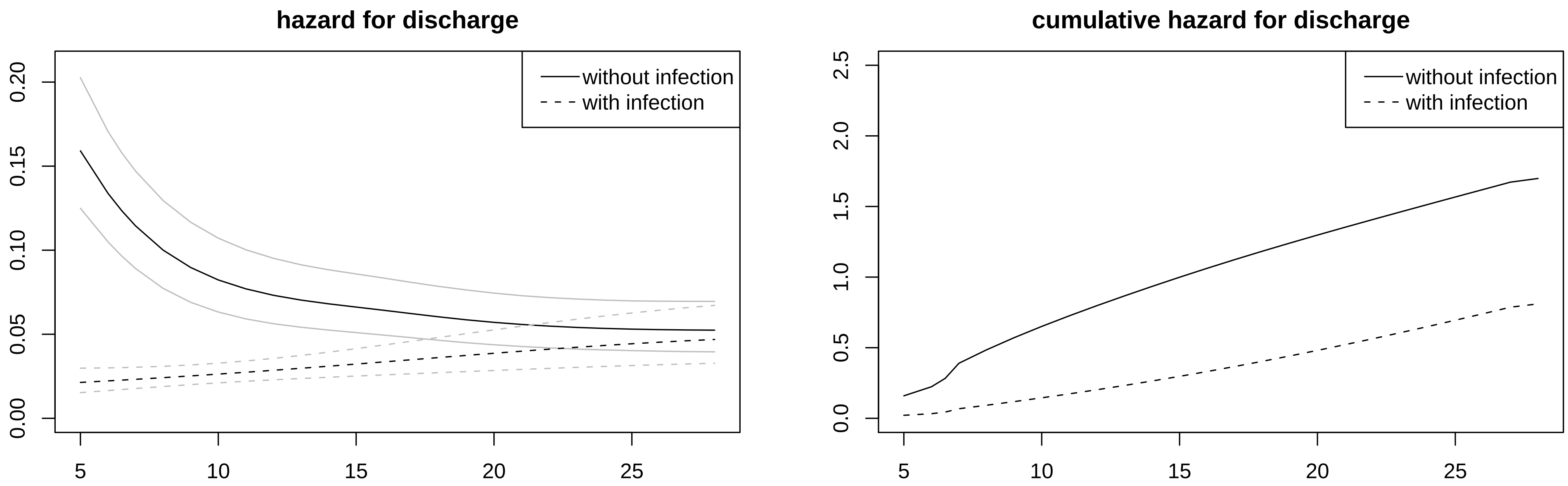
- Each transition is assigned a separate hazard function $\lambda_h(t)$.

Length of Hospital Stays

- Data on 1876 intensive care patients from the Charité University Hospital in Berlin collected from February 2000 until July 2001.
- 158 with nosocomial pneumonia.
- Structure of the associated multi-state model:



- Scientific question: Does an infection prolong the length of admission to intensive care?



- Significantly reduced hazard for the transition to discharge if an infection is acquired.

Semiparametric Hazard Specification

- Regression models for the hazard rates:

$$\lambda_h(t) = \exp(\eta_h(t))$$

  where $\eta_h(t)$ is a semiparametric, time-dependent predictor.

- Different types of models explained for the human sleep data example:
  - Cox-type models:
  $$\eta_h(t) = g_{0,h}(t) + x'\beta_h,$$
  where $g_{0,h}(t)$ is the log-baseline and $x'\beta_h$ comprises parametric effects, e.g. of gender.
  - Special case: Models without any further covariates such as the nosocomial infection model.

– Time-varying effects:

$$\eta_h(t) = \ldots + x \cdot g_h(t) + \ldots$$

where $g_h(t)$ is the smooth, time-varying effect of a covariate x (for example impact of cortisol level on transition to REM sleep).

– Frailties:

$$\eta_h(t) = \ldots + b_c + \ldots$$

where b_c is a cluster-specific frailty term (for example individual-specific effects to account for different sleeping behaviour).

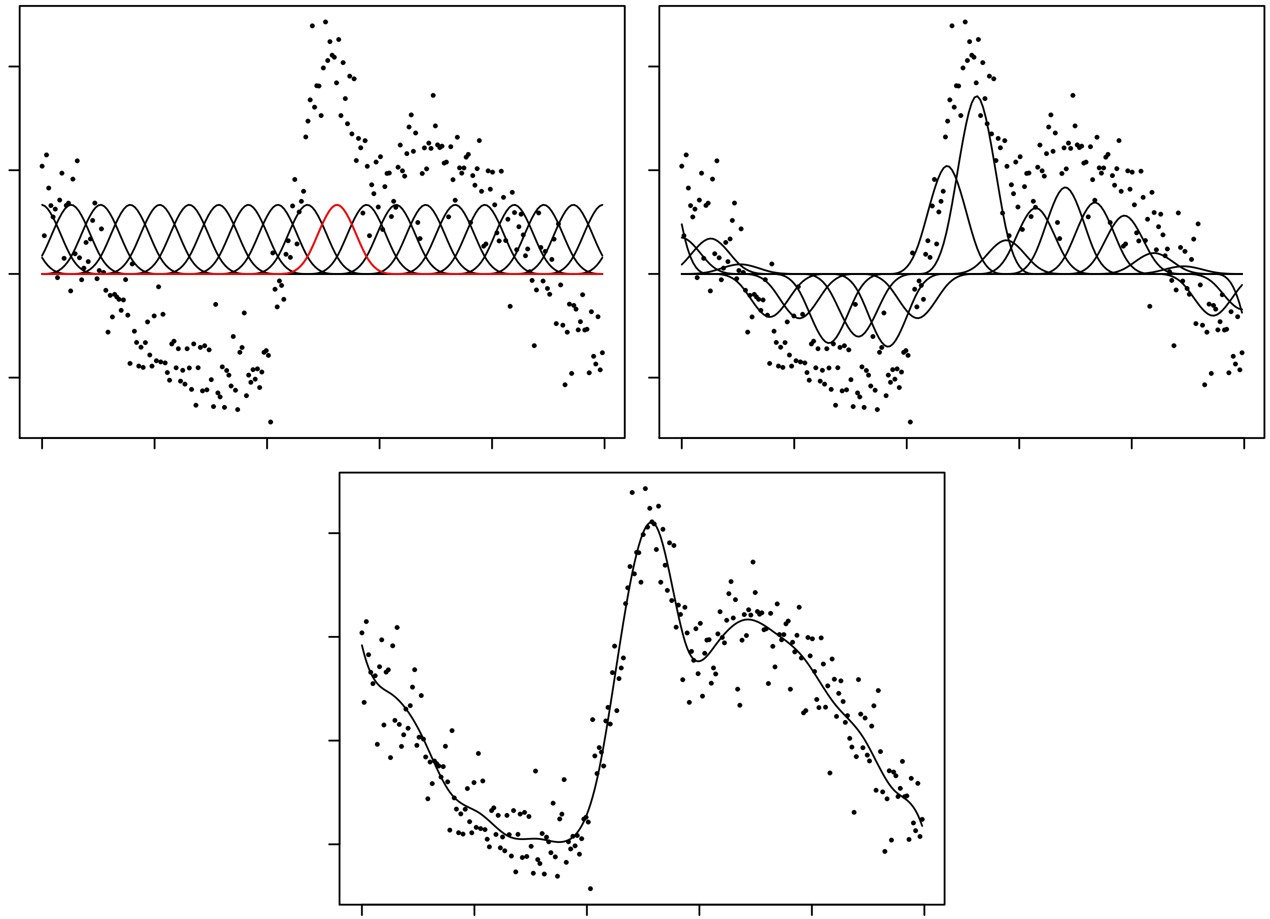
– Nonlinear effects:

$$\eta_h(t) = \ldots + f_h(x) + \ldots$$

where f_h is a smooth, nonlinear function of a continuous covariate x (for example nonlinear effect of cortisol concentration).

- Nonlinear and time-varying effects are modelled based on penalised splines:
  - Approximate unknown functions in terms of flexible basis functions, e.g.

$$g_h(t) = \sum_j \gamma_j B_j(t).$$

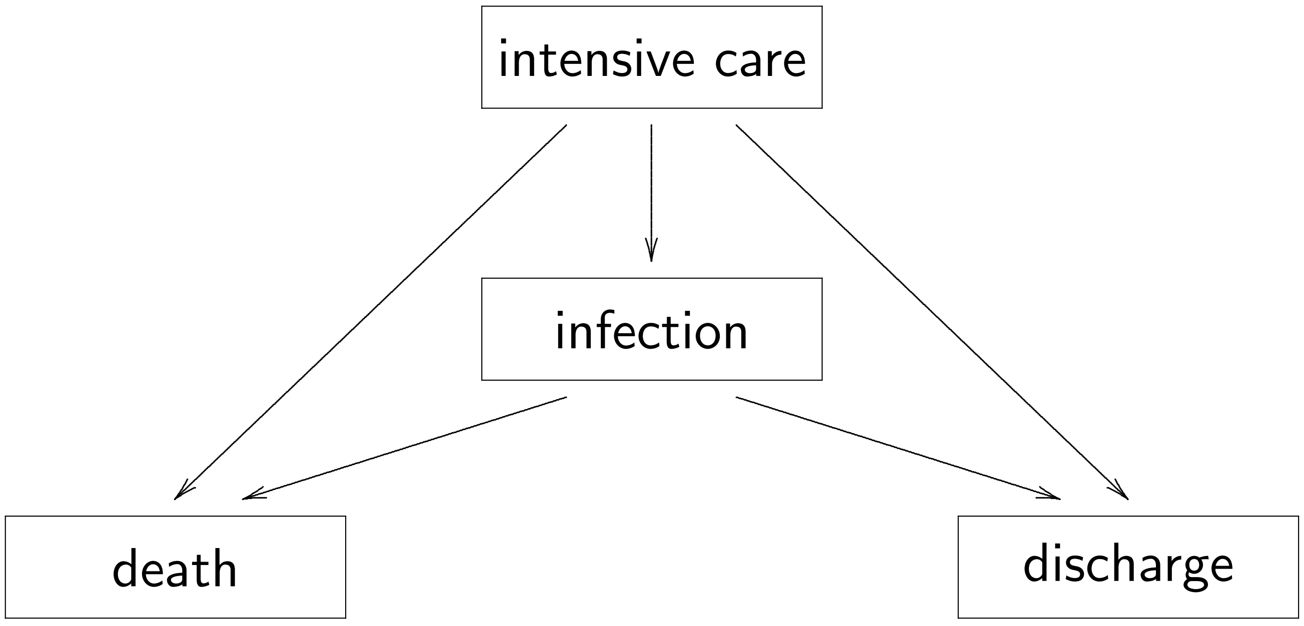


  - Assign an additional penalty to the basis coefficients to enforce smoothness of the functions.

- Inference in semiparametric multi-state models relies on
  - a counting process representation and
  - penalised maximum likelihood estimation.
- The counting process representation also yields residual processes.
- The same methods can also be applied in regression models for survival times.
- Advantages of semiparametric models:
  - Inclusion of flexible covariate effects.
  - Inclusion of frailties for unobserved heterogeneity.
  - Methods imbedded in regression models $\Rightarrow$ tests, confidence intervals, etc.
  - Direct estimation of the hazard instead of the cumulative hazard.
  - Smooth time-varying effects instead of step functions.

Advantages of Multi-State Models

- Simultaneous analyses of several transitions allows for a relative interpretation of hazards.
- Example: Analysis on survival times while accounting for nosocomial infections:



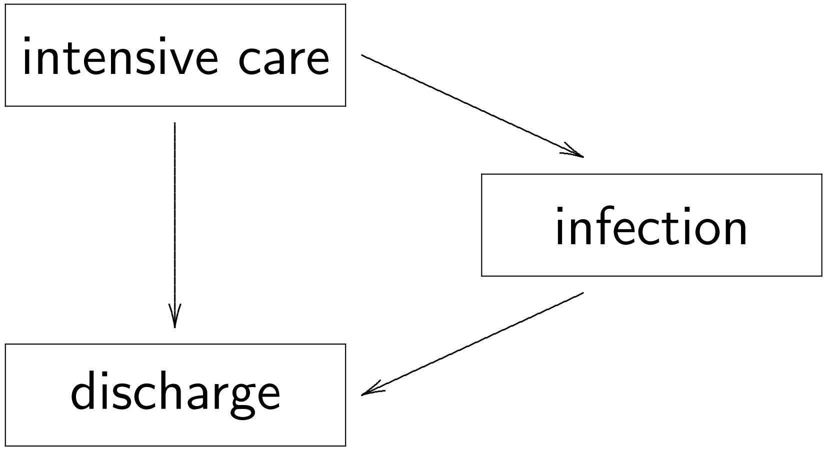
- Hypothetical result:
  - No differences in death rates between patients with and without infection
  - Empirical observation: More deaths following infection.
- Explanation obtained from the multi-state model:
  - Lower discharge rate for patients with infection results in longer hospital stays.
  - Patients with infection are under risk for death for a longer time.

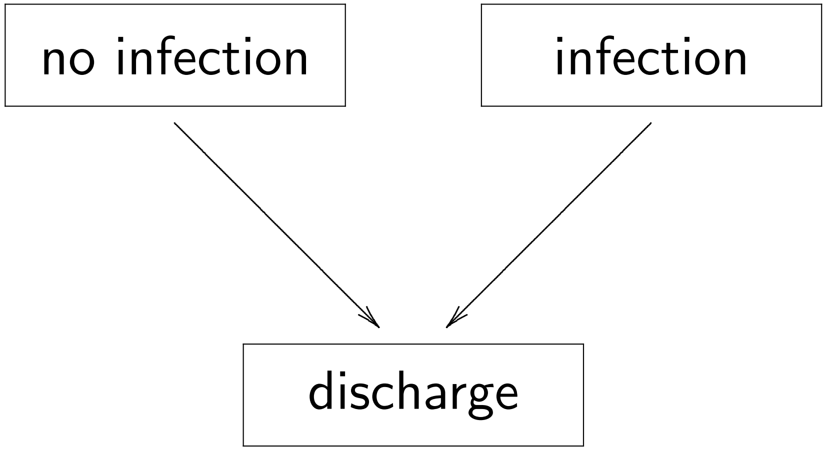
$\Rightarrow$ Take care when interpreting separate hazard functions.

- Multi-state models also allow to study the temporal evolution of the phenomena of interest.

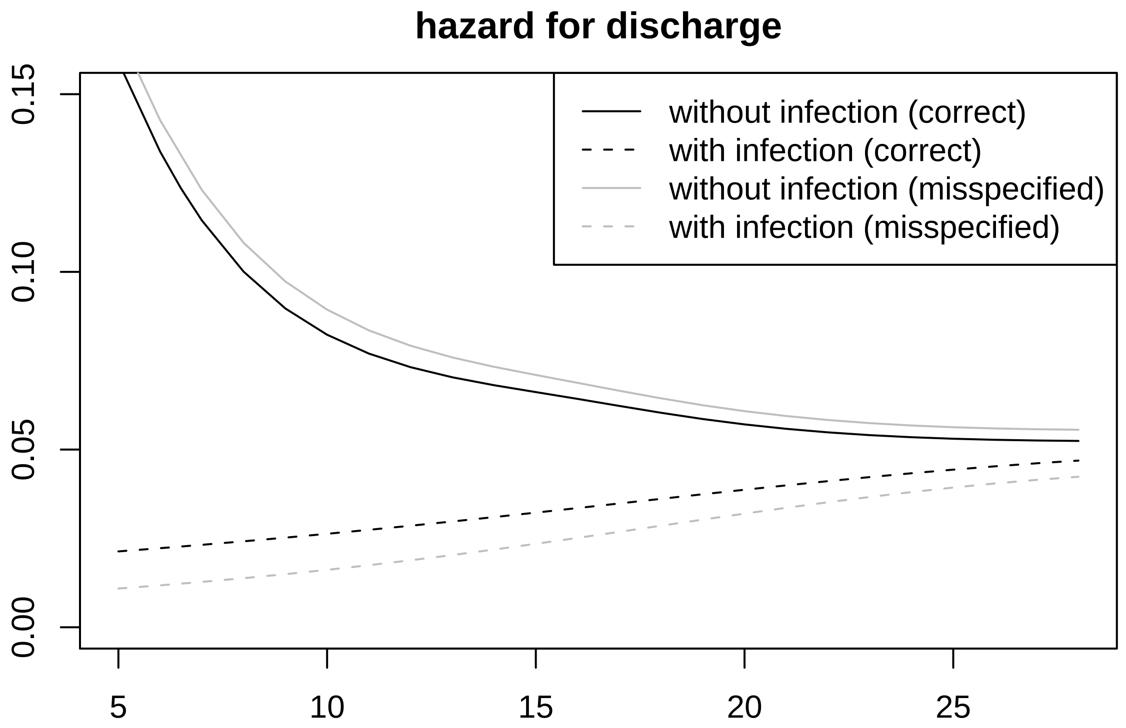
- Earlier studies on nosocomial infections often treated the infection status as a time-constant risk factor.

- Misspecification of the multi-state model:





- Impact on the estimation results:



⇒ The impact of infection is over-estimated in the misspecified model.

Summary & Outlook

- Multi-state models are a useful statistical tool to study temporal dynamics of health-related processes.
- The semiparametric model specification allows to include flexible covariate structures for example in terms of frailties, time-varying effects or nonlinear effects.
- Implemented in the free software package BayesX:

  `http://www.stat.uni-muenchen.de/~bayesx`

- Survival time regression is included as a special case.
- The counting process representation enables the construction of martingale residuals.

- References:

  - Kneib, T., & Hennerfeind, A. (2008): Bayesian Semiparametric Multi-State Models. Statistical Modelling, 8, 169-198.
  - Beyersmann, J., Kneib, T., Schumacher, M. & Gastmeier, P. (2009): Nosocomial infection, length of stay and time-dependent bias. Infection Control & Hospital Epidemiology, 30, 273-276.

- Acknowledgements:

  - Andrea Hennerfeind (Ludwig-Maximilians-University Munich).
  - Jan Beyersmann & Martin Schumacher (Albert-Ludwigs-University Freiburg).
  - Spread of Nosocomial Infections and Resistant Pathogens (SIR) Network.

- A place called home:

  `http://www.staff.uni-oldenburg.de/thomas.kneib`