Bayesian Competing-Risks Analysis of an Epidemiological Breast Cancer Data Set

Author: Lisa Gibbons[1] Supervisor: Matthias Bruekner[2]

[1] PHASTAR, Unit 2, 2A Bollo Lane, London W4 5LE [2] Department of Mathematics and Statistics, Lancaster University, LA1 4YF

Background

The Yang-Prentice model allows for the possibility of crossing hazard functions. This is helpful to determine if a treatment will be useful in the long run even if it causes certain adverse effects early on.

When analysing which factors can influence a breast cancer patient's survival, competing risks must be taken into account (Figure 1). Competing risks often appear in epidemiological studies, however there is no standard method to account for them.

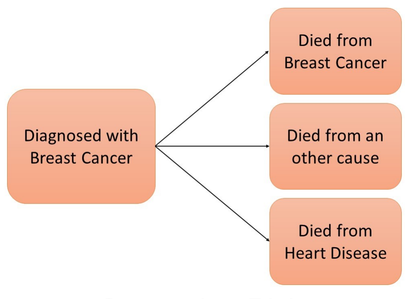

Figure 1: Competing Risks example

Aims

- Implement a parametric Bayesian competing risks survival model using a Yang-Prentice model for the cause-specific hazards
- Validate the implementation of the model using simulated data
- Apply the model to a subset of the Surveillance, Epidemiology, and End Results (SEER) breast cancer data set
- Assess the long and short term effects of the covariates on the mortality of breast cancer patients
- Compare these results with those provided by Hinchliffe and Lambert [2013] who apply a flexible parametric proportional hazards model

Methodology

Data Simulation

- To allow for model comparison and validation
- Simulate competing risks data using Beyersmann [2012]
- Use the Yang-Prentice model for the cause specific hazards with a Weibull baseline survival function

Frequentist Vs. Bayesian

- The Frequentist approach is generally used with Survival analysis
- Implement two different Monte Carlo Markov Chain (MCMC) methods to provide samples from the posterior distribution for the Bayesian approach.
- WinBUGS implements Bayesian inference using Gibbs sampling (BUGS). Whereas RStan implements the No-U-Turn-Sampler (NUTS).

Models

Yang-Prentice Model

Yang and Prentice [2005] have defined a two sample semi-parametric model which can accommodate crossing hazard functions as

$$\lambda_T(t) = \frac{\theta_1\theta_2}{\theta_1 + (\theta_2 - \theta_1)S_C(t)}\lambda_C(t) \quad (t > \tau_0). \tag{1}$$

Where $S_c(t) = exp\left(-\int_0^t \lambda_C(u)\, du\right)$ is known as the baseline survival function and $\lambda_T(t)$ and $\lambda_C(t)$ represent the hazard of the treatment and control group. We can then define θ_1 and θ_2 as the short term and long term hazard ratios respectively:

$$\theta_1 = lim_{t\downarrow 0}\frac{\lambda_T(t)}{\lambda_C(t)}, \quad \theta_2 = lim_{t\uparrow\tau_0}\frac{\lambda_T(t)}{\lambda_C(t)}. \tag{2}$$

Cox Model

A special case of the Yang-Prentice model is that it can be reduced to the Cox model when $\theta_1 = \theta_2$. We can express the hazard of an individual with covariates $\mathbf{x}_i$, where h_0 is some unspecified baseline hazard function, as

$$h(t;\mathbf{x}_i) = \psi(\mathbf{x}_i)h_0(t) \tag{3}$$

Extended Cox Model

To account for non-proportional hazards we can also extend the Cox model using a step function $\beta(t)$ to split time into two intervals.

$$h(t;\mathbf{x}_i) = h_0(t)\exp\{\beta_1 I(t \le t')X + \beta_2 I(t > t')X\} \tag{4}$$

Simulated Study Results: $\theta_1 = \theta_2$

Approach	Parameters	Mean	95% CrI/CI	Lies in CrI/CI
WinBUGS	HR1	0.79	(0.41,1.40)	89%
RStan		0.85	(0.23,1.16)	90%
Frequentist		0.91	(0.55,1.67)	94%
WinBUGS	HR2	0.57	(0.44,1.50)	95%
RStan		0.60	(0.24,1.22)	94.2%
Frequentist		0.49	(0.22,1.07)	97%

Table 1: Cox model estimation

Credible Interval (CrI) Confidence Interval (CI)

The data is simulated with constant hazard ratios. Taking: $HR1 = \theta_1 = \theta_2 = 0.9$ for an event of type 1 and $HR2 = \theta_1 = \theta_2 = 0.5$ for an event of type 2. Table 1 then shows how well the Cox model estimated these hazard ratios.

Simulated Study Results: $\theta_1 \neq \theta_2$

Data is now simulated with $\theta_1 \neq \theta_2$. For an event of type 1: $\theta_1 = 1.2$ and $\theta_2 = 0.8$. For an event of type 2: $\theta_1 = 0.9, \theta_2 = 0.6$.The figures below display the Survival curves using the Yang-Prentice model given our chosen θ values.

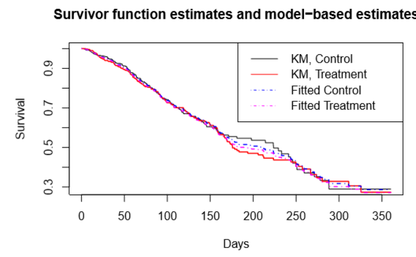

Figure 2: Type 1 Event

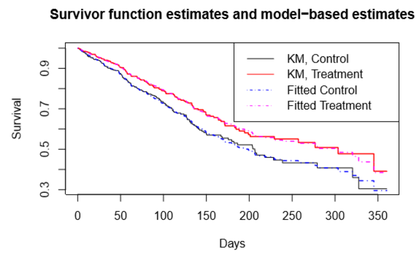

Figure 3: Type 2 Event

Table 2 displays the results for the Extended Cox model and the Yang-Prentice model. For the Extended Cox model, using a time split of 0.5 we make the assumption that the hazard ratio for ≤ 0.5 can be interpreted as θ_1 and the hazard ratio for > 0.5 as θ_2.

			Extended Cox model			Yang-Prentice model		
Approach	Event	Parameters	Mean	95% CrI/CI	Lies in CrI/CI	Mean	95% CrI/CI	Lies in CrI/CI
RStan	Type 1	θ_1	1.11	(0.69,1.67)	94%	1.15	(0.80,1.66)	96.1%
		θ_2	1.43	(0.31,4.90)	91%	0.87	(0.56,1.42)	97.5%
	Type 2	θ_1	0.96	(0.59,1.47)	97%	0.89	(0.61,1.32)	97.1%
		θ_2	1.23	(0.19,4.16)	92%	0.72	(0.42,1.32)	95.9%
Frequentist	Type 1	θ_1	1.14	(0.74,1.75)	93.2%	1.20	(0.04,35.93)	100%
		θ_2	115260.3	$(0.34,\infty)$	95.2%	1.00	(0.44,2.30)	84.6%
	Type 2	θ_1	0.85	(0.54,1.33)	93.1%	0.96	(0.01,67.81)	99.8%
		θ_2	91609.31	$(0.21,\infty)$	97.4%	0.96	(0.44,2.12)	74%

Table 2: Non-proportional hazards estimation

Neither the Frequentist nor Bayesian method estimated close to the true values of θ_2 for the Extended Cox model. This could be due to outlying data, the choice of time split or the fact that it violates the standard proportional hazard assumption that the Cox model is known for. For the Yang-Prentice model as the Bayesian estimates lie within the credible interval 95% of the time the implementation of the model is validated.

SEER Data Results

The model is applied to a subset (30,000 patients) from the SEER breast cancer data. Four possible competing outcomes (dying from: breast cancer, other cancer, heart disease and other cause of death) were applied, Table 3 displays two of these.

	Breast Cancer		Heart Disease	
Parameters	$log\theta_1$ (95% CrI)	$log\theta_2$ (95% CrI)	$log\theta_1$ (95% CrI)	$log\theta_2$ (95% CrI)
Surgery	-2.80 (-2.88, -2.73)	-0.23 (-0.36, -0.09)	-0.31 (-0.41, -0.21)	0.20 (0.10, 0.30)
Ages 60-69	0.13 (0.06, 0.20)	-0.17 (-0.27, -0.07)	0.85 (0.77, 0.93)	0.02 (-0.08, 0.13)
Ages 70-79	0.28 (0.21, 0.35)	-0.17 (-0.27, -0.07)	1.73 (1.65, 1.82)	0.14 (0.04, 0.24)
Ages 80+	0.52 (0.44, 0.59)	-0.28 (-0.38, -0.17)	2.53 (2.45, 2.61)	0.21 (0.08, 0.31)
Regional	1.21 (1.14, 1.27)	0.26 (0.16, 0.36)	0.02 (-0.06, 0.11)	-0.14 (0.25, -0.03)
Distant	3.55 (3.48, 3.62)	2.12 (2.03, 2.21)	0.16 (0.05, 0.27)	-0.19 (-0.31, -0.07)

Table 3: SEER Results

No surgery, Ages 18-59 and Localised are the reference categories for surgery, age and stage of breast cancer respectively

Interpretations

The covariate results for the short term hazard are similar to those provided in Hinchliffe and Lambert [2013]. Further validating the model we can see that as age at diagnosis increases then the rate of death also increases, which is expected. In order to fully assess how these covariates affect a patients survival in relation to the competing risks, the covariate results would be used to calculate the cause-specific hazards. This would allow us to confirm inferences such as

- As age increases, the long term hazard indicates that you are more likely to die from something other than breast cancer
- θ_2 increasing as stage of breast cancer increases indicates that a subject with distant breast cancer is more likely to die from that than any other competing cause

Conclusion

Through data simulation it has been shown how the choice of model can affect results depending on the type of data available. Furthermore it shows how Bayesian implementations of models can be of use as opposed to the standard Frequentist method. It has also been shown how a Bayesian estimation of the Yang-Prentice model can be adapted to account for competing risks and multiple covariates in an Epidemiological setting.

References

Jan Beyersmann. *Competing risks and multistate models with R [electronic resource]*. Use R! Springer, New York, 2012. ISBN 9781461420347.

Sally R Hinchliffe and Paul C Lambert. Flexible parametric modelling of cause-specific hazards to estimate cumulative incidence functions. *BMC medical research methodology*, 13(1), February 2013. ISSN 1471-2288.

Song Yang and Ross Prentice. Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika*, 92(1):1–17, March 2005. ISSN 0006-3444.

Contact:lisa.gibbons@phastar.com