Estimating correlation between competing risks

Valerie Gares[1], Malcolm Hudson[1,2], Val Gebski[1]

[1]NHMRC Clinical Trials Centre, Sydney Univ.

[2]Department of Statistics, Macquarie Univ.

December 2015, *Biometrics by the harbour*

Introduction

- In problems with a time to event outcome, **competing risks** may arise when subjects experience events which prevent the outcome of interest in the study being observed
- Methods analogous to the logrank test and proportional hazard models are commonly used to account for the competing risks (Fine and Gray model)

- Correlated competing risk is problematic
  - Poor understanding of their effect in inference
  - Estimation of the correlation is methodologically problematic

Introduction

- In problems with a time to event outcome, **competing risks** may arise when subjects experience events which prevent the outcome of interest in the study being observed
- Methods analogous to the logrank test and proportional hazard models are commonly used to account for the competing risks (Fine and Gray model)

- Correlated competing risk is problematic
  - Poor understanding of their effect in inference
  - Estimation of the correlation is methodologically problematic

Aims

- Knowledge of the level of **correlation** between two different risks
- Assess the **effects of correlation** between two risks on the hazard ratio (HR) estimator for a treatment versus a control
  - assist in the understanding and interpretation of the results of competing risk analysis
- Apply this to an actual study in Head and Neck cancer where competing risks are present

Aims

- Knowledge of the level of **correlation** between two different risks
- Assess the **effects of correlation** between two risks on the hazard ratio (HR) estimator for a treatment versus a control
  - assist in the understanding and interpretation of the results of competing risk analysis
- Apply this to an actual study in Head and Neck cancer where competing risks are present

Aims

- Knowledge of the level of **correlation** between two different risks
- Assess the **effects of correlation** between two risks on the hazard ratio (HR) estimator for a treatment versus a control
  - assist in the understanding and interpretation of the results of competing risk analysis
- Apply this to an actual study in Head and Neck cancer where competing risks are present

The bivariate normal censored model

- We examine the problem of correlation assuming the normal distribution where methods are both developed and understood
- The logarithm of a time to event may satisfy the **normality assumption** (e.g. symmetric distribution)
- We propose a **bivariate censored normal model**
  - We can use **EM algorithm** to estimate the means, variances and **correlation** between two competing events
  - We can also estimate these quantities using EM algorithm for a **prespecified correlation**
- This approach allows us to **impute** the survival time for the censored events where censoring was due to occurrence of a competing event, loss to follow up or end of the study

Bivariate survival data

- Y_1: logarithm of the time to event of the cause of interest *(event 1)*
- Y_2: logarithm of the time to event of another cause *(event 2)*
- Z: covariates vector
- (Y_1, Y_2) follows a **bivariate normal distribution** with, for $i = 1, 2$,
  - $\mathbb{E}(Y_i) = \mu_i = \beta_i Z'$, where β_i is coefficient vector
  - $\text{Var}(Y_i) = \sigma_i^2$
  - $\text{Corr}(Y_1, Y_2) = \rho$
- n: the sample size
- τ: the end of the study

Bivariate survival data

For each individual

- We observe:

$$X = \min(Y_1, Y_2, \tau)$$

- and the indicating variable

$$\delta = \begin{cases} 1: & \text{if the event of interest occurs} \\ 2: & \text{if a competing risk occurs} \\ 0: & \text{if the subject is censored} \end{cases}$$

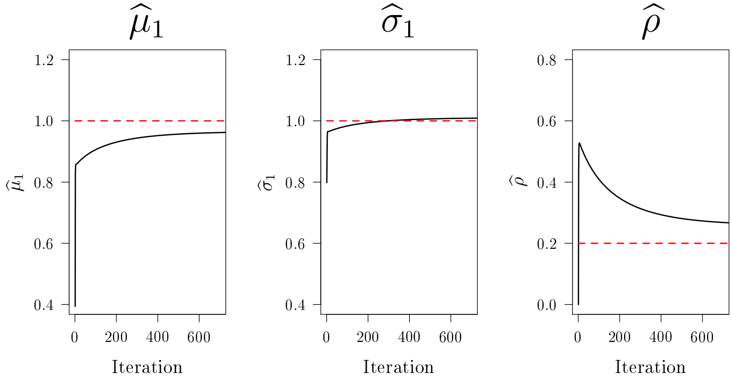
- We assume that 80% of events of interest would occur before the end of follow-up (but may be censored by the other cause)

Performance of the EM algorithm in bivariate normal censored model

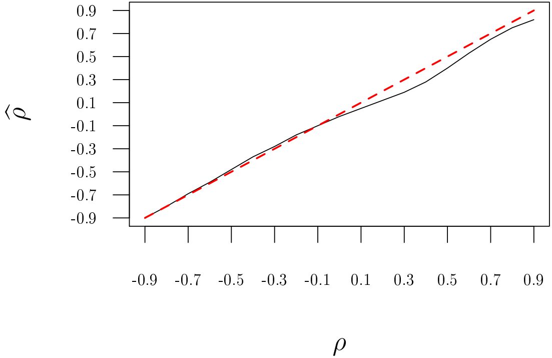
Simulation studies:

- $\mu_1 = \mu_2 = \sigma_1 = \sigma_2 = 1$ and for different values of ρ, generate a dataset of $n = 1000$ pairs (X, δ)
- For this data, we use the EM algorithm to estimate the means, variances and correlation
- Repeat this 1000 times

Plot of the EM algorithm convergence for a typical dataset $\rho = 0.2$



Estimation of the correlation



Effects of correlation on hazard ratio estimation

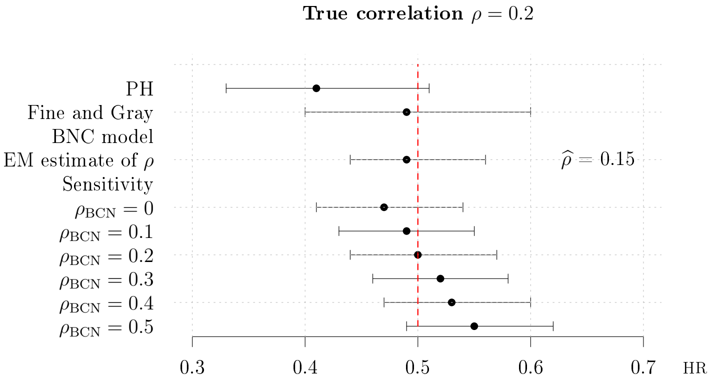
Simulation studies:

- Clinical trial with two groups
  - 500 patients are under treatment (T)
  - 500 patients are under control (C)
  - 80% of events of interest would occur before the end of follow-up
  - *Event 1*: $\mu_1^T = 1$, $HR_1 = 0.5$, $\sigma_1 = 1$
  - *Event 2*: $\mu_2^T = 1$, $HR_2 = 1$, $\sigma_2 = 1$
  - Different correlations $\rho \in \{0, 0.2, 0.4\}$

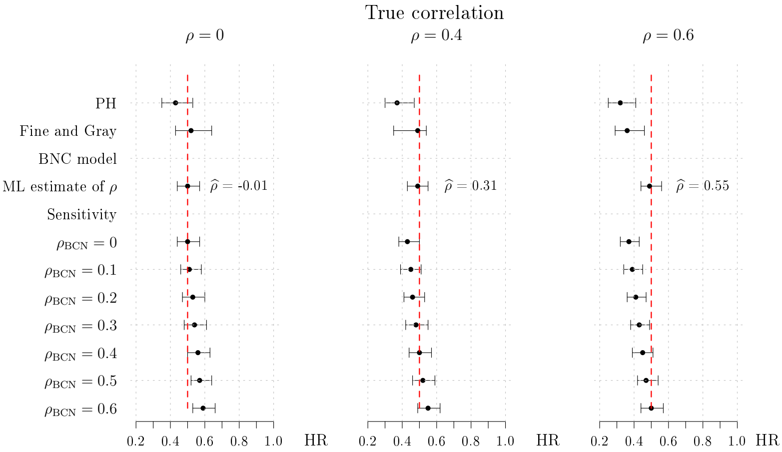
Effects of correlation on hazard ratio estimation

- Methods compared are regression coefficient estimates in:
  - the PH model (subject who have the competing risk are censored)
  - the Fine and Gray model
  - the **bivariate censored normal model** using EM algorithm
    - Imputation in the E-step can use prespecified correlation (ρ_{BCN})
    - Imputation can use EM estimator of correlation

Effects of correlation on hazard ratio estimation



Effects of correlation on hazard ratio estimation



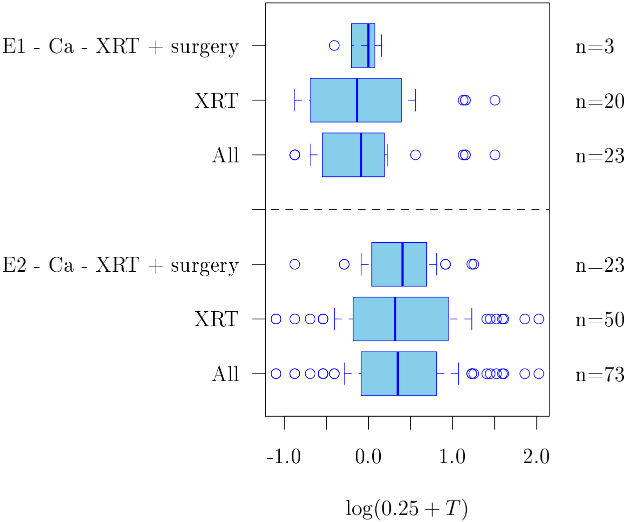
Example

- The study compared patients with Head and Neck cancer:
  - receiving radiotherapy (XRT)
  - receiving surgery with adjuvant radiotherapy
- The events are recorded as:
  - local regional relapse only (Event 1 - Ca)
  - other causes: distant relapse, intercurrent deaths (Event 2 - OC)
- Patients are censored due to loss to follow up or the end of study

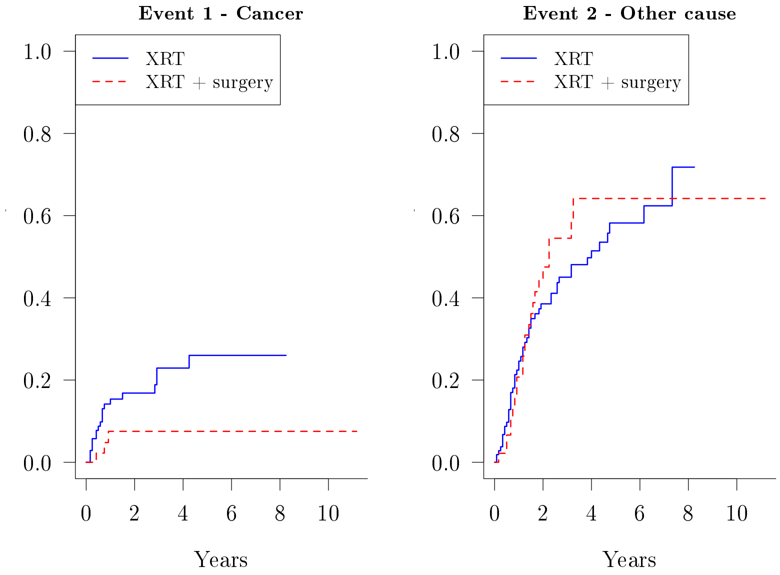
Example

- The study compared patients with Head and Neck cancer:
  - receiving radiotherapy (XRT)
  - receiving surgery with adjuvant radiotherapy
- The events are recorded as:
  - local regional relapse only (Event 1 - Ca)
  - other causes: distant relapse, intercurrent deaths (Event 2 - OC)
- Patients are censored due to loss to follow up or the end of study

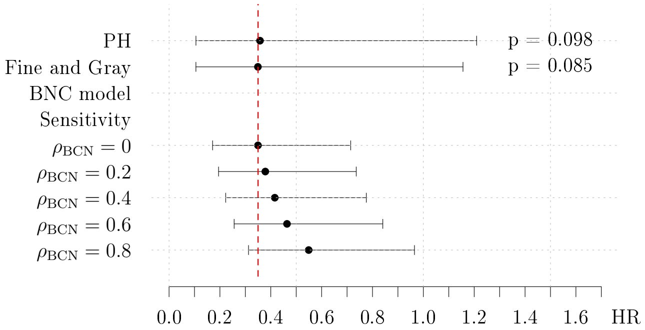
Distribution



Cumulative incidence functions



Hazard ratio estimation



Conclusion

- Slight underestimation of correlation with bivariate censored normal model
  - We need to do a sensitivity analysis for prespecified ρ
- The correlation level has an impact on the hazard ratio estimation
  - Fine and Gray appears robust when the risks are moderately correlated
  - The BNC model is an alternative in presence of high correlation
- Sensitivity analysis can help to determine the upper bound of potential correlation
  - provides a method to assist estimation of correlation between risks
  - may be useful in the design of later studies.

Thanks

Thank you for your attention

- Buckley, J. and James, I. Linear regression with censored data. Biometrika, 66(3):429.436, 1979.
- Jin Z., Lin D.Y. and Ying Z. On least-squares regression with censored data. Biometrika, 93(1):147.161, 2006.
- Crowder M. On the identialibility crisis in competing risks analysis. Scandinavian Journal of Statistics, 18(3):22.233, 1991.
- Fine J. P. and Gray R. J. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association, 94:496. 509, 1999.