Model selection and explained variation of survival from cancer

Camille Maringe, Laura Woods, Janez Stare, Bernard Rachet*
CRUK Cancer Survival Group
London School of Hygiene and Tropical Medicine

Background and aims

• Model selection for estimation of cancer survival
  • Akaike Information Criterion (AIC)

• Model selection for prediction of cancer survival
  ➢ Ideas:
  • AIC-based measures of goodness of fit
  • Measures of explained variation
AIC-based measures of Goodness of Fit

- To help detect likely form of the predictors
- To investigate model-averaging

- AIC & AIC-derived measures: overall and by age group
  - AIC-differences: \( d_i = \text{AIC}_i - \min_{j=1:N}(\text{AIC}_j) \)
  - Model likelihood: \( l_i = \exp(-d_i/2) \)
  - AIC weights: \( w_i = l_i/\sum_{j=1:N}(l_j) \)

  \( i \) – model in consideration

  \( N \) – total number of models considered

A measure of explained variation: \( R_E \)

- \( R_E \)
  - Stands for “Ranks Explained”
  - Non-parametric
  - Model-free interpretation
  - Well-understood scale: \([-1; +1]\)
  - Consistent with independent censoring mechanisms


  ➢ To be adapted to the net survival context
RE in the overall survival context

\[ R_E = \frac{\sum_i (r_{i,\text{null}} - r_{i,\text{model}})}{\sum_i (r_{i,\text{null}} - r_{i,\text{perfect}})} \]

Summation over all observations that fail

i: time at which the \(i^{th}\) observation fails

\(r_{i,\text{null}}\): rank of observation \(i\) under the null model

\(\rightarrow\) all observations in risk set have equal chances to fail at time \(t_i\)

\(r_{i,\text{perfect}}\): rank of observation \(i\) under the perfect model

\(\rightarrow\) observation \(i\) is given rank 1

RE in the net survival context

\[ R_E = \frac{\sum_i (r_{i,\text{LT}} - r_{i,\text{model}})}{\sum_i (r_{i,\text{LT}} - r_{i,\text{perfect}})} \]

Summation over all observations that fail

i: time at which the \(i^{th}\) observation fails

\(r_{i,\text{LT}}\): rank of observation \(i\) by the life tables

\(\rightarrow\) all observations in risk set have unequal chances to fail at time \(i\)

\(r_{i,\text{perfect}}\): rank of observation \(i\) under the perfect model

\(\rightarrow\) observation \(i\) is given rank 1
Test of $R_E$ in the net survival context

- Simulated datasets
  - 2000 colon cancer patients
  - Age at diagnosis 15-99 years (mean age 70)
  - Distribution of stage at diagnosis
    - 13% - stage I
    - 40% - stage II
    - 27% - stage III
    - 20% - stage IV

- Scenarios: varying effects of age and stage at diagnosis on cancer-survival
  - Time-fixed vs. time-varying
  - Linear vs. non-linear

- 100 datasets per scenario

Results: $R_E$ in the net survival context

*7 years of follow-up*

![Simulation: LaNLS, 7 years of F-up. Re measured at 5 years](#)
Results: $R_E$ in the net survival context

15 years of follow-up

Censoring

Simulation: LaNLs, 15 years of F-up
Re measured at 10 years

Simulation: LaNLs, 7 years of F-up
Re measured at 5 years
Results: $R_E$ in the net survival context

*Time-varying effect*

Simulation: LaNLs, 7 years of Fup

Discussion

- $R_E$ not sensitive to models which do not affect the rank order
- Informative null model
  - Pros: reflect the nature of net survival model
  - Cons:
    - values never high (survival issue)
    - values for $R_E$ can go beyond -1
- Local $R_E$: tool for understanding time-varying impact of risk factors on cancer survival
- Additional potential predictors
  - Future
    - Model-averaging
    - Frailty model