Discussion: Modelling baseline excess hazard function in relative survival

Patrick Royston, MRC Clinical Trials Unit, London

Stockholm meeting on relative survival, Sep 6th, 2007

1

Statement of Interest!

- I am biased!
- Nelson's approach is essentially that of Royston & Parmar (2002)
- 5 years later, I still like the approach
- I therefore encourage Nelson and others to research this family of models further

2

Bernard Rachet

- Time and covariates are not (or need not be) categorised
- FPs and/or regression splines as functions of time for modelling baseline excess hazard
- … also for interactions of covariates with time
- Principled approach to model selection using MFP

3

Issues with Rachet's approach

- Modelling the excess hazard with FPs/splines may give artefacts in the curves
- "How flexible?" seems to be a general issue
  - Particularly with large datasets
  - Criteria for model selection?
- Selecting interaction models – may be too flexible, leading to instability/possible interpretation problems
  - Breast cancer example
- May need expert "tuning" to get good results

4

Chris Nelson

- Main difference: modelling *cumulative* excess hazard function
- Splines in log t for this function
- Can also model covariates in similar fashion

5

Some issues with categorisation (Nelson)

- Several methods split the timescale into manageable chunks for estimation purposes
  - Necessary with large datasets
- Collapse on time and grouped covariates
  - Can't handle continuous covariates
- Piecewise models could be considered as biologically implausible, both for time and covariates (also, biased?)

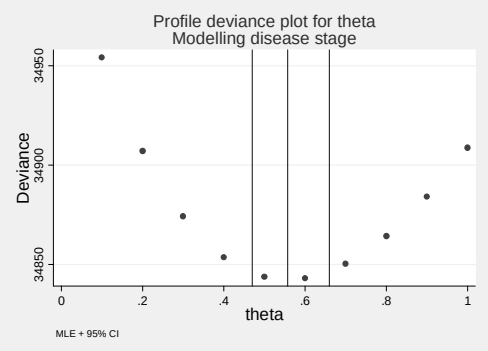
6

- Many parameters for time-dependent effects
  - Time chunks x covariate groups = lots
  - Overfitted, unrealistic
  - Can lead to problems in smaller datasets, e.g. zero cells
- If have many observations, may get computation-time problems on split data
- Methods that do *not* split time require integration to estimate survival function
  - Because estimating excess hazard, not cumulative

Comments

- For me, Nelson's approach is close to the method of choice
- Could substitute fractional polynomials for splines in some applications (Lambert, Rachet)
  - Model baseline cumulative excess hazard
  - Model continuous covariates
- Don't forget the other link functions
  - Θ = 1 (cumulative proportional odds)
  - Θ ≠ 0, 1 (can be estimated from the data)

Colon cancer data – stage known



Regression splines – in favour

- Parametric
  - Concise description of function available
- Flexible
  - Can cope with many different curve shapes
- Can control flexibility via number of knots
  - Not usually too sensitive to exact knot positioning
- Easily incorporated in regression models
  - No special computational methods required

Regression splines – against

- Baseline cumulative excess hazard function not guaranteed monotonic in time
- Model (knot) selection – is AIC satisfactory?
  - No simulation evidence available
- Can get artefacts in fitted function
  - Problems with over-interpretation
  - May not be transportable to other datasets (validation)
- Highly correlated terms
  - Need to orthogonalise

A difficult issue: model selection

- Scale of model
  - Hazard, odds, probit, etc.
- Baseline (spline terms)
  - Done separate or simultaneously with covariates?
- How to do scale, variable and function selection?
  - E.g. when have prognostic factors in cancer
- Many possible covariate x time interactions give serious multiplicity problem
  - Forward selection?
- Assessing model stability (e.g. bootstrap)

Paul Lambert

- Cure models – mixture and non-mixture formulations
- Seems particularly appropriate approach in relative survival context
  - In most primary cancers, some patients will be cured
  - For metastatic (e.g. stage 4) cancer, cure proportion is nearly 0
- Preference for non-mixture model
  - More realistic
  - Is also interpretable as a mixture model

13

Issues

- Are the models identifiable?
  - Model with excess hazards converging towards zero may fit equally well – and could use more flexible distributional specification
- Robustness - estimate of cure fraction (π) depends on choice of parametric survival distribution
- Is it possible to fit Royston-Parmar-Nelson distributions to S(t)?
- How to choose a model – many possibilities?

14

Ula Nur

- Handle missing covariate values by multiple imputation (MI)
- Create *m* complete datasets (e.g. *m* = 10)
- Requires a (complex) imputation model
  - To make MAR assumption more plausible
- Do MI using **`ice`**
- Combine results from models fit by **`strs`**, **`strel`** etc using **`micombine`** or **`mim`**
  - **`mim`** is preferred command now

15

Issues

- Handling of categorical covariates needs care
  - Use **`passive()`** and **`substitute()`** options in **`ice`**
- Not clear how/whether survival time should be transformed in **`ice`** model
  - General issue in survival analysis with missing data
  - Probably depends on substantive model to be fitted
- Since covariate x time interactions are a feature of RS models, should consider including them in **`ice`** model
  - Omission will cause bias towards zero

16

Future

- What are the main aims of relative survival models?
- If *description* then *goodness of fit* and *interpretability* are important
- If *prediction* in independent data then *external validation* is important
- What implications would these aims have for the preferred method of modelling?
- E.g. (when) should cure models be the approach of choice?

17

Further applications

- Meta analysis of relative survival functions across hospitals, regions, countries, etc?
- Role in regular prognostic modelling?
- Role in economic evaluation of cancer therapies using data from clinical trials?
  - Estimate disease-related years of life lost

18

MRC Clinical Trials Unit

Stata

- Standardise popmort file specification for use by strel, strs* and other applications?
- Recommendation of particular approaches/Stata commands to answer specific questions?

19