Survival analysis, cont.

Gustaf Edgren, PhD

Outline

- Repetition of some of what Micke just said…
- Regression models for survival and prognosis studies
- Time dependent covariates
- Standardized mortality ratios (SMR:s)
- Relative survival and cure rates

Risk vs. survival models?

- Thematically, standard risk models are based on the two-by-two table (i.e. binary outcomes)

Exposure	Dead: Yes	Dead: No	
+	13	42	55
-	2	53	55

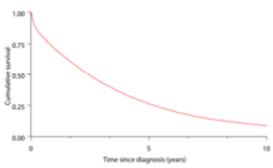
- In survival (which, face it, is 0 in everyone) analysis, <u>time to</u> the binary event is also of interest
- Other models are therefore necessary

Expression of survival

- In clinical settings, patient survival is often expressed in terms of X% 5-year survival or that the average survival was Y years
- This is fine and dandy – and very intuitive – but such expressions really measure different things and are frequently setting-dependent:
  - When did they die?
  - Keep in mind age, sex, SES, etc.
  - Death from other causes?

Time-to-event modeling – rationale

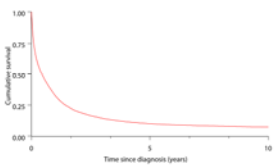
Myeloma survival



Median survival=2.44 years

10-year survival=8.7%

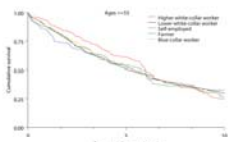
AML survival



Median survival=0.44 years

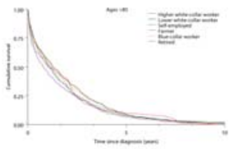
10-year survival=7.3%

Survival regression models – rationale



5.1 year median survival

What is the principal explanation for the difference?



1.1 year median survival

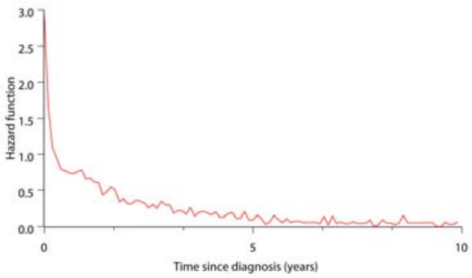
Wish-list for survival regression

- A suitable regression model for survival modeling needs to be able to:
  - Handle censoring
  - Adjust for confounders
  - Handle time-dependence
  - Handle baseline hazard functions with very varying shapes
  - Ideally, should be able to produce estimates of both relative and absolute risk

Regression models for survival

- The most commonly used model for survival regression is probably the Cox model – often referred to the proportional hazards model
- It was proposed by Sir David Cox, and is based on the modeling of (not survival), but rather of hazard functions
- Recognizing the varying shapes of hazard functions, Cox ingeniously introduced an assumption under which the baseline function could be ignored

Hazard function for AML



Hazards

- Theoretically, the hazard function is a theoretical measure of the instantaneous mortality ratio
- It relates to survival as:

S(t) = ∫h(t)	i.e. Survival is 1 minus the sum of all hazard
h(t)=-dS(t)	i.e. The hazard is the negative instantaneous change in survival

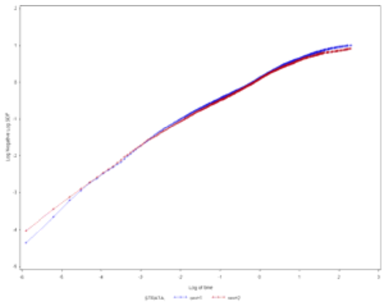
Cox regression basics

- The Cox model models all risks in relation to an baseline hazard (which is not estimated, but factored out)
- The Cox model assumes that all covariates act on this baseline risk multiplicatively
  - I.e. it is relative risk model – cannot calculate absolute risks
- The Cox model assumes that all covariates act equally (proportionally) on the baseline hazard over time
  - I.e. the effects of the covariates do not change with time
- The Cox model handles both right-censured and interval-type data
  - I.e. it is able to handle time-dependent covariates

Cox regression assumptions

- Main assumptions for Cox regression:
  - Proportional hazards*
  - Non-informative censoring
  - Sufficient sample size**
  - *Not too many ties (relative)*

Proportional hazards?



Cox model results

- The Cox model produces relative risks expressed as hazard ratios
- For all practical purposes, hazard ratios can be interpreted as incidence rate ratios
- In theory, the hazard ratio is the ratio of instantaneous mortality intensities

Cox model – pro and con

- Advantages
  - Quick to converge
  - Robust and forgiving model
  - Commonly used (!)
- Disadvantages
  - The proportional hazards assumption
  - No absolute risks can be computed
  - Does not handle truly aggregated data
  - Really requires access to reliable cause-specific death rates

Poisson regression

- Poisson regression is a commonly used alternative to Cox regression
- For all practical purposes, it is equivalent to Cox regression *(who needs a Cox model anyway?)*
- However, there are some situations where Poisson analysis is preferential to Cox:
  - Aggregate data
  - Truly LARGE studies
  - Serious problems with non-proportional hazards*

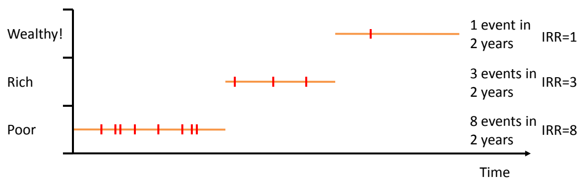
Note: Poisson regression is NOT a survival method, but is used for such purposes

Time-dependent covariates

- In many situations, covariates change in a meaningful way
- Depending on the situation, this can and can not be accounted for in the analysis
- Typical examples of time-dependent covariates are:
  - Age
  - Calendar period
  - Cumulative exposure to (some environmental agent)
  - Income
  - Etc.

Time-dependent covariates, cont.

- The principal fashion for managing time-dependent covariates is to “split” follow-up time and events according to which exposure stratum they contributed:



Time-dependent covariates, cont.

- While conceptually simple:
  - Time-dependent covariates is difficult to execute,
  - May give you results that are difficult to interpret
  - Requires CAREFUL thought
- Some general rules:
  - Never know what you don't know (at that time)
  - Never condition on the future

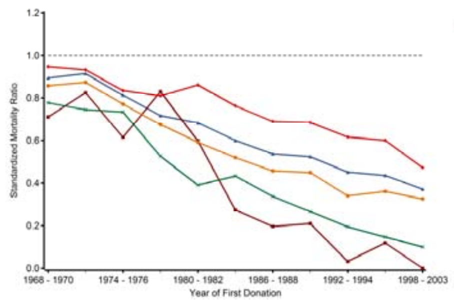
Time-dependent covariates, cont.

- For time-dependent covariates with a fixed "origin," there are very standardized solutions:
  - Lexis (implemented in Stata, macro for SAS)
  - Fstpyrs macro for SAS (http://sourceforge.net/projects/pyrsstep)
- These can chop time up by age and calendar period relatively easily

Standardized mortality ratios

- We've talked about SMR:s previously
- Recap:
  - External comparison to the general population
  - Handles confounding by age, sex and calendar period with standardization
  - Interpreted just like a relative risk (observed number of deaths/expected number of cases)

Standardized mortality ratios, ex.



Relative survival

- Background:
  - While SMR:s give us the opportunity to study relative mortality, it says little about survival
  - Likewise, Kaplan-Meier methods fail to consider the expected mortality (due to age, etc.)
- Thus, we want to be able to assess survival relative to the general population

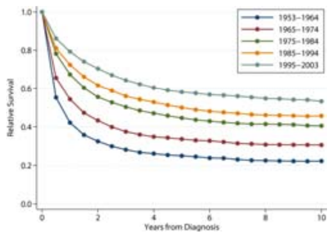
Relative survival, cont.

$$\text{Relative survival} = \frac{\text{Observed survival}}{\text{Expected survival}}$$

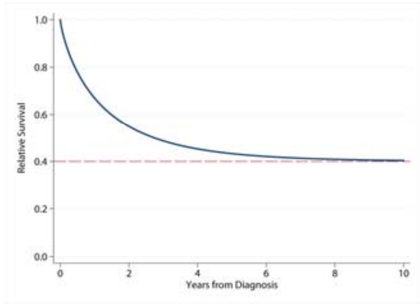
- Expected survival obtained from national population life tables stratified by age, sex and other covariates
- Estimate of mortality associated with a disease without requiring information on cause of death. Can also be expressed on hazard scale:

Excess mortality=observed mortality - expected mortality

Relative survival, colon cancer



Cure?



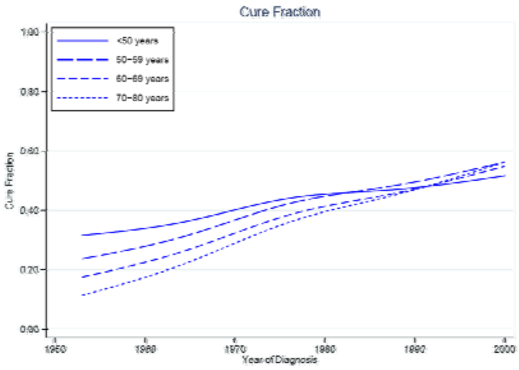
Statistical cure?

- From a relative survival model, it is possible to estimate the point at which the survival of your population returns to that of the background population
- This point is generally referred to as the cure point, and the relative survival at that point is referred to as cure fraction

Cure analysis

- Cure rate analysis is a relatively new field
- It assesses the occurrence of statistical cure
  - I.e. the population cure rate
- It does NOT say anything about individual cure
- With this comes some advantages:
  - The cure fraction is not affected by things like lead time
  - It allows the reliable comparison of calendar period effects with

Cure rate



Summary

- Adequate analysis of survival requires consideration to time-to-event
- Several models for survival* analysis exist:
  - Kaplan-Meier analysis
  - Cox regression
  - Poisson regression
  - SMR
  - Relative survival

Summary, cont.

- Cox and Poisson regression are common models for analysis of predictors of survival
  - Both produce relative risk estimates (HR, IRR)
- Alternatives to Cox, Poisson and KM are SMR:s and relative survival models which account for the background population mortality