UIC INSTITUTE FOR HEALTH RESEARCH AND POLICY

Time-dependent covariates in the Cox proportional hazard regression model

Oksana Pugach, PhD

Institute for Health Research and Policy

University of Illinois at Chicago

October, 2014

Overview

- Introduction to survival analysis
- Specifics of survival data
- Cox proportional model
- Time dependent covariates
- Time-varying covariates
- Example

Introduction to Survival Analysis

The prototypical event is death, hence the name

Survival analysis is also known as

- Event history analysis (sociology)
- Duration models (political science, economics)
- Hazard models (epidemiology, biostatistics)
- Failure-time models (engineering)

Examples not from clinical/medical research:

political careers, length of civil wars, coalition durability, strike duration, work career, criminal careers, marriage.

Specifics

Models for time-to-event data

- Survival times are non-negative and usually positively skewed
- Presence of censored observations
  - Right censored
  - Left censored
  - Doubly censored
  - Interval censored
- Truncation – individuals never considered for inclusion into the study (Example: survival study of residents of a retirement center (left truncation))

Cox Proportional Hazard Model

- Popular method of analyzing survival data

$$h_i\left(t \mid X\right) = h_0(t)\exp\left(\sum_{k=1}^{p} b_k x_{ik}\right)$$

- Assumptions
  - Effect of covariate X_k on hazard rate is constant over time
  - Survival time is independent of censored time

What if hazards are non-proportional

1. Stratify.

   The covariates with non-proportional effect may be incorporated into the model as stratification factor. Effect of stratification variable on survival cannot be estimated. Works naturally for categorical variables. Analysis less efficient.

2. Partition time axis.
3. Use time-dependent covariates.
4. Use different model such as accelerated failure time or additive hazards model

Time-Dependent Covariates

- Values of some of explanatory variables in a study change over time. In this case the most recent value is used at each specific time in the modeling process.
- Two types of variables that change over time:
  - Internal variables: relate to a particular individual; reflect condition of a patient and their values may well be associated with the survival time of the patient
  - External variables: exists independently of survival experience of a subject (level of atmospheric sulphur dioxide); variable changes in a way that its value will be known in advance at any future time (patient age)

- The Cox regression model takes the form

$$h_i\left(t \mid X\right) = h_0(t)\exp\left(\sum_{k=1}^{p} b_k x_{ik}\left(t\right)\right)$$

- Both the data management and conceptual treatment of time-dependent covariates are facilitated by the 'counting process' approach

Interpretation of parameters

- Baseline hazard $h_0(t)$
  - Hazard of death at time *t* for individual with all covariates at 0 at the time origin and remain at this same value through time

- β_k - log-hazard ratio for two individuals whose value of the *k*-th explanatory variable at any time *t* differs by one unit

Note: since covariates depend on time, the relative hazard is also time-dependent. The model is no longer a proportional hazard model

The choice of time-dependent covariate involves the choice of a functional form for the covariate. This choice is usually not self-evident and may be suggested by biological underlying mechanisms. (Fisher and Lin, 1999)

Time-Varying Effects

- Coefficient of a time-constant explanatory variable is a function of time: the log-hazard ratio is time-dependent and Cox model is no longer proportional hazard model

$$h_i\left(t \mid X\right) = h_0(t)\exp\left(\beta\left(t\right)X\right)$$

$\beta\left(t\right) = \beta_0 + \beta_1 t$ - for example, then the model is the same as with time-dependent (external) covariate and constant coefficient

- If $\beta\left(t\right)$ is non-linear function, it can be estimated by smooth functions
- Also used to test proportional hazard assumption

Example: Stanford Heart Transplant

Problem: Evaluate whether patients receiving heart transplant will benefit with increased survival

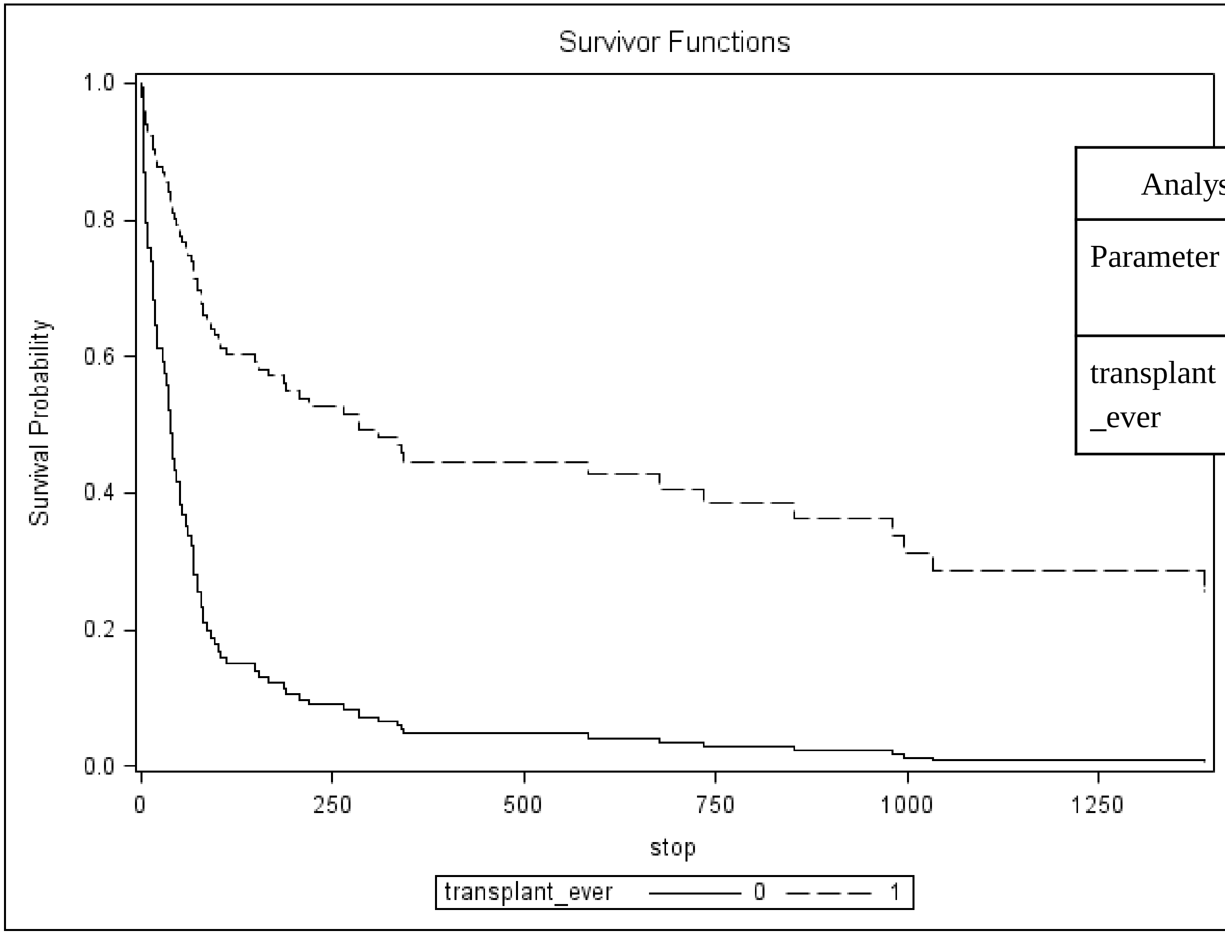
Some approach:

Identify patients that received heart transplant and those that did not. Estimate survival time from the time they entered the study for two groups and compare the survival times using log rank test.

Patients that died earlier did not have a chance to receive heart transplant. Two groups are selectively biased favoring the heart transplant patients.

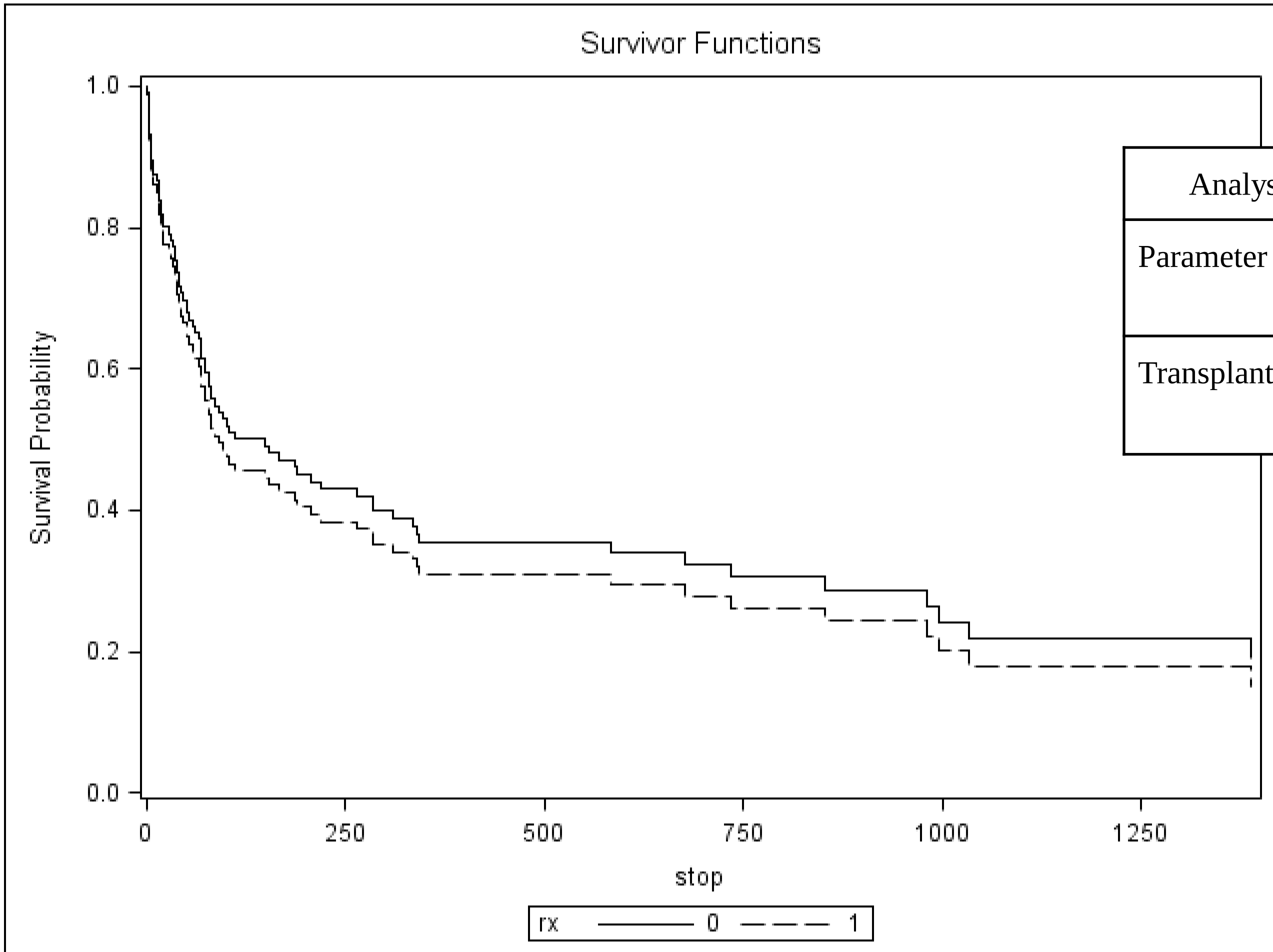
Source: J Crowley and M Hu (1977), Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, **72**, 27–36.

Example: Stanford Heart Transplant – time-constant covariate



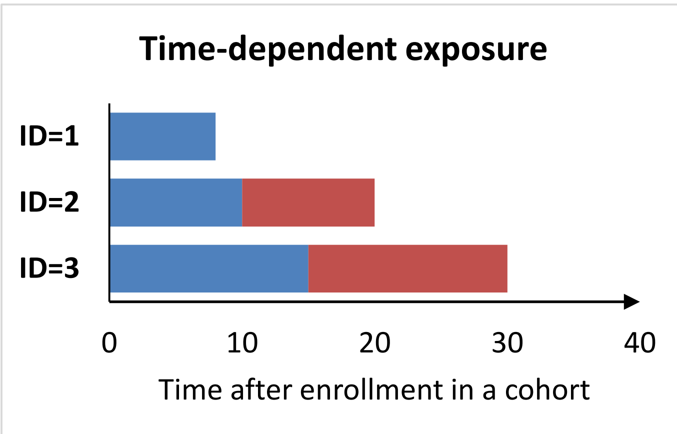
Analysis of Maximum Likelihood Estimates				
Parameter	Parameter Estimate	SE	P-value	HR
transplant_ever	-1.32	0.24	<.0001	0.27

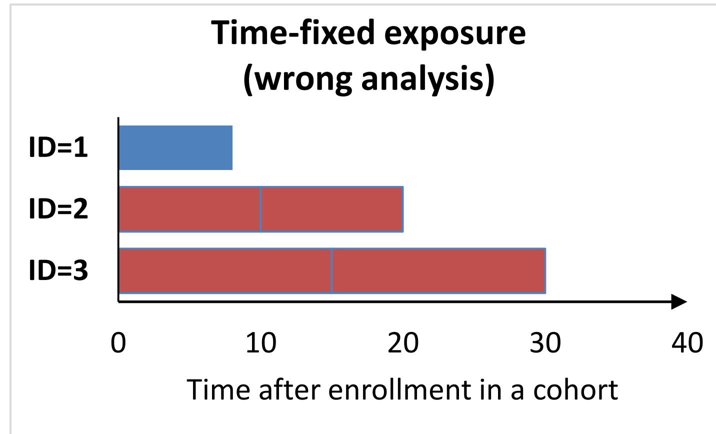
Example: Stanford Heart Transplant – time-dependent covariate



Analysis of Maximum Likelihood Estimates				
Parameter	Parameter Estimate	SE	P-value	HR
Transplant	0.13	0.30	0.67	1.13

Time immortal bias (Survival Bias)





Bias can be introduces if analyzed as ever/never transplant vs non-transplant

With Time-fixed Exposure:

Rate of death in Transplant group = 2/(20+30) = 0.04 per person-month

Rate of death in Non-Transplant group = 1/8 = 0.125 per person-month

Rate Ratio = 0.32

With Time-Dependent Exposure:

Rate of death in Transplant group = 2/(10+15) = 0.08 per person-month

Rate of death in Non-Transplant group = 1/(8+10+15) = 0.03 per person-month

Rate Ratio = 2.67

Attribution to Ayumi Shintani 'Cox Proportional hazard regression with time varying covariate in SPSS' Youtube video

Thank you!