Survival Analysis I (CHL5209H)

Olli Saarela

Dalla Lana School of Public Health
University of Toronto

olli.saarela@utoronto.ca

March 6, 2019

Time-dependent covariates

Motivation

- The interpretation of the hazard function through the relationship

$$P(\mathrm{d}N_i(t) = 1 \mid \mathcal{F}_{t^-}) = E[\mathrm{d}N_i(t) \mid \mathcal{F}_{t^-}] = Y_i(t)\lambda_i(t)\,\mathrm{d}t$$

  fundamentally implies that survival analysis involves a prediction problem; given all information we have observed just before time t (the history $\mathcal{F}_{t^-}$), we are trying to say something about the probability of the event occurring at time t.
- When predicting, we cannot look into the future.
- Pretending that we know the future is a common source of erroneous results.

A case in point: immortal time bias

- In statistics, 'bias' is understood as the difference between the true parameter value and the expected value of an estimator.
- In epidemiology, 'bias' is more generally a consequence of some particular flaw in the study design or analysis (e.g. selection bias, confounding bias, lead-time bias, length bias, publication bias). The magnitude of this would be measured by the statistical bias, cf. Miettinen (2011, p. 98):

  > *Bias - Consequence of study methodology such that even with infinite study size the result (free of imprecision) would be at variance with the parameter value at issue; also: the extent of this discrepancy.*

- There is a bias that refers to the result of looking into the future in survival analysis, namely *immortal time bias* (e.g. Suissa 2008; Hanley & Foster 2014).

Brøndum-Jacobsen et al. (2013):

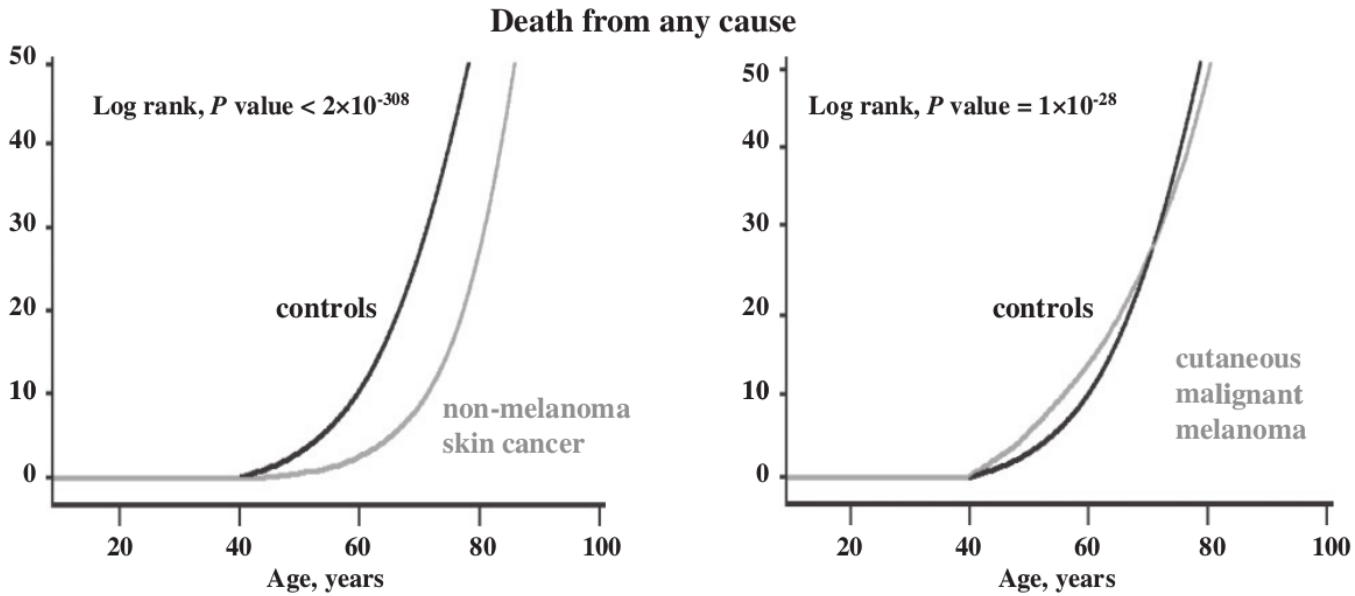

Figure 1 The cumulative incidence of myocardial infarction, hip fracture and death as a function of age in individuals above age 40 years ever diagnosed with non-melanoma skin cancer and cutaneous malignant melanoma. Cumulative incidence curves were generated from Kaplan–Meyer estimates, comparing individuals with non-melanoma skin cancer and cutaneous malignant melanoma vs individuals free of both diseases. *P*-values are for comparison between groups by log rank tests

Remarks

- In this study, the hypothesis was that sun exposure can have beneficial health effects which would manifest as decreased all-cause mortality.
- Skin cancer was used as a surrogate for sun exposure.
- The all-cause mortality of individuals who were ever diagnosed of skin cancer was compared to all-cause mortality of the general population.
- The results, especially when using non-melanoma skin cancer as the surrogate exposure, were striking; the mortality of the skin cancer group was much lower compared to the general population.
- Is this evidence that sun exposure has beneficial effect on all-cause mortality?
- Why were the results less striking when using malignant melanoma as the surrogate exposure?

Source of the bias

- In the study, individuals were classified based on whether they had been ever diagnosed of skin cancer.
- By definition, those classified as having had a skin cancer diagnosis had survived until their skin cancer diagnosis (were 'immortal' during this period).
- Presumably, the risk of skin cancer increases with age; thus, at the time of their diagnosis, the skin cancer patients are older than the general population.
- The reason for difference in mortality between the skin cancer group and the general population should now be obvious.
- Be prepared to explain immortal time bias to a researcher who has never heard of it, and wants you to do a similar analysis.

Recognizing immortal time (Hanley & Foster 2014)

Table 1. Ways to recognize immortal time

Suggestion	Remarks/tests
Distinguish state from trait	A trait (e.g. blood group) is usually forever; people and objects move between states (on/off phone; intoxicated/not; on/off medication; failed allograft in place/removed)
Distinguish dynamic from closed population	Membership in a closed population (cohort) is initiated by an event (transition from a state) and is forever; in a dynamic population, it is for the duration of a state. Dynamic populations are the only option for studying transient exposures with rapid effects (e.g. cellphone/alcohol use vs the rate of motor vehicle accidents)
Focus on person-time in index and reference categories, rather than on people in exposed and unexposed 'groups'	These refer to exposure categories, not to people per se; a person's time may be divided between exposure categories; unless people remain in one category, it is misleading to refer to them as a 'group'
If authors used the term 'group', ask …	When and how did persons enter a 'group'? Does being in or moving to a group have a time-related requirement? Is the classification a fixed one based on the status at time zero, or later? Is it sufficient to classify a person just once, or do we need to classify the 'person-moments,' that is the person at different times?
Sketch individual timelines	If there are two time scales, a Lexis diagram can help; use different notation for the time portion of the timeline where the event-rate of interest might be affected, and the portion where it cannot (see Figures)
Measure the apparent longevity- or time-extending benefits of inert agents/interventions	After the fact, use a lottery to assign virtual (and never actually delivered) interventions, but with same timing as the one under study. Or use actually-received agents with same timing
Imagine this agent/intervention were being tested within a randomized trial	How, and when after entry, would the agent be assigned? Administered? How would event rates be computed? How would Farr have tested his 'early-promotion' suggestion?
Think short intervals and hazard rates, even if the hazard rates do not change abruptly	In addressing the present, conditional on the past, the hazard approachhas already correctly documented the experience in each small past interval; the natural left to right time-ordering of the short intervals allows for correct recognition of transitions between exposure states. By computing a mortality rate over a longer time-span defined after the fact, one may forget that in order to contribute time to the index category, people had to survive the period spent in the (initial) reference category

Lessons to be learned

- The hazard function is defined conditional on the past.
- In survival analysis, never ever classify individuals into groups based on future events.
- Kaplan-Meier curves and related comparisons are not appropriate if the groups to be compared cannot be defined at the start of the follow-up.
- If the exposure is defined through a state, it can change over time, and thus it is impossible to classify individuals into exposure categories (they can be both exposed and unexposed during their follow-up).
- Carrying out the correct survival analysis in this kind of setting is not particularly difficult, requiring only a straightforward generalization of the familiar Cox model, and certain computational tricks.

Time-dependent covariates

- Introduce notation for the covariate history of individual i until time t:

$$Z_i(t) \equiv \{z_i(u) : 0 \leq u \leq t\},$$

where $z_i(t)$ is the covariate vector at time t.

- The hazard function can now be redefined through

$$\lambda_i(t)\,\mathrm{d}t = P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid Z_i(t), \tilde{T}_i \geq t).$$

- A simplification is obtained if we assume that the hazard function depends only on the current values of the covariates, rather than the cumulative history of these:

$$\lambda_i(t)\,\mathrm{d}t = P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid z_i(t), \tilde{T}_i \geq t).$$

Cox model with time-dependent covariates

- The corresponding proportional hazards model can be specified as

$$\lambda_i(t) = \lambda_0(t) \exp\{\beta' z_i(t)\}.$$

- Usually, some of the covariates are constant over time and some are time-dependent.
- Note that this model still assumes the covariate effects to be constant over time, even though the covariates may change.
- The corresponding Cox partial likelihood is simply

$$\prod_{i=1}^{n} \left(\frac{\exp(\beta' z_i(t_i))}{\sum_{l=1}^{n} Y_l(t_i) \exp(\beta' z_l(t_i))} \right)^{e_i}.$$

- Note that the covariate values of the riskset members are evaluated at the event time of the individual with the outcome event.

Interpretation

- Suppose that we have only a single binary time dependent covariate $z_i(t) = \mathbf{1}_{\{v_i<t\}}$, where v_i is the time of an event that determines the subsequent exposure status (say, a treatment procedure or start of prescription medication), the model is

$$\lambda_i(t) = \lambda_0(t) \exp\left\{\beta \mathbf{1}_{\{v_i<t\}}\right\}.$$

- For two individuals i and j with $v_i < t$ and $v_j > t$ we then get $\lambda_i(t)/\lambda_j(t) = \exp\{\beta\}$.
- While the hazard ratios are interpretable, calculating predictive probabilities (risk or survival) based on such a model is more complicated, because we don't know the future exposure status. For this, we would have to predict the covariate process as well.

Creating time-dependent covariates in R

- In the `coxph` function, certain kinds of time-dependent covariates can be created through a time-transform function, specified through the `tt` argument, which calculates and returns the covariate value at a given point in time.
- Other option is the create a long format dataset, where the individual contributes one row at each constant value of the time-dependent covariate. Here variables need to be specified for the starting and ending times of the interval, the outcome indicator, and the covariate value.
- Examples to follow.

The effect of transplantation on survival

- Typical examples of a time-dependent exposure are studies of survival of patients on a waiting list for an organ transplant.
- We want to know whether the transplantation improves survival, but only part of the individuals on the waiting list received the transplant during the follow-up period.
- Incorrect analysis of such a setting by classifying the patients into 'transplanted' and 'not transplanted' groups would easily lead to immortal time bias. (Why?)
- Demonstrate the incorrect analysis and immortal time bias, as well as the correct analysis, with the Stanford Heart Transplant dataset...

References

- Brøndum-Jacobsen P, Nordestgaard BG, Nielsen SF, Benn M (2013). Skin cancer as a marker of sun exposure associates with myocardial infarction, hip fracture and death from any cause. *International Journal of Epidemiology* 42:1486–1496.
- Miettinen OS (2011). Epidemiological Research: Terms and Concepts. Springer, Dordrecht.
- Hanley JA, Foster BJ (2014). Avoiding blunders involving 'immortal time'. *International Journal of Epidemiology* 43:949–961.
- Suissa S (2007). Immortal Time Bias in Pharmacoepidemiology. *American Journal of Epidemiology* 167:492–499.