Survival analysis
Course: Statistical Evaluation of Diagnostic and Predictive Models

Thomas Alexander Gerds (University of Copenhagen)

Summer School, Barcelona, July 2, 2015
Risk dynamics

Dynamic = changing, able to change and to adapt

For the patient
- Environment
- Treatment
- Disease

For the modeller
- Prediction time-point
- Prediction horizon
- Event status, measurements of biomarkers, treatment, questionnaire results, etc.
The role of time

Prediction model timeline

Time point at which patient is provided with prediction

Time point attached to the prediction

baseline

Origin (time 0)

Horizon (time t)

followup

Lost to followup, or (right) censored, means that patient was not followed until horizon time t.

Until time t, three things can happen:

- patient is event-free
- the event of interest has occurred
- a competing event has occurred
Setup

- time origin = 0
- Survival time: \( T \) (time between 0 and an event)
- Covariate vector \( X \in \mathbb{R}^p \) (age, sex, diabetes, stroke score, \ldots)
- \( X \) must be measurable at time 0

Outcome at time \( t \)

\[
Y(t) \in \{0, 1\} \quad 0 = \text{survived at } t, \ 1 = \text{dead between } 0 \text{ and } t
\]

Aim

Predict survival status at time \( t \) with risk prediction model \( \hat{R}_n \)

\[
(X, t) \mapsto [0, 1] \quad \hat{R}_n(t|X) \approx P(Y(t) = 1|X)
\]
Predicting survival using a Cox regression model

From a fitted Cox regression model predictions of the survival probability at time $t$ are obtained with the following formula:

$$
\hat{S}_{Cox}(t|X) = \exp \left( - \int_0^t \hat{\alpha}_0(s) \exp(\hat{\beta}_1 X_1 + \cdots + \hat{\beta}_K X_p) \right)
$$

- $\hat{\beta}_k$ is the partial likelihood estimate of the log-hazard ratio
- $\hat{\alpha}_0$ is the Efron or Breslow estimate of the baseline hazard rate
- Predictions are time-dependent
Preding survival using other tools

- parametric Cox regression (survival, rms)
- Aalen’s non-parametric additive hazard model (timereg)
- penalized Cox regression (penalized)
- random survival forest (randomForestSRC)
- boosting (CoxBoost)
- Binomial regression (timereg, riskRegression)
Performance measures are time-dependent

Brier score

\[
\text{Brier}(t) = E_{Y_i,X_i}\{Y_i(t) - \hat{R}_n(t|X_i)\}^2
\]

- Subject i was not used to train \( \hat{R}_n \).
- The “null” model is the Kaplan-Meier estimate

Area under the time-dependent ROC curve

\[
AUC(t) = E_iE_j(\mathcal{I}\{\hat{R}_n(t,X_i) > \hat{R}_n(t,X_j)\}|T_i \leq t, T_j > t).
\]

- Subjects i,j were not used to train \( \hat{R}_n \).
- Retrospective interpretation
Uncensored observations

![Graph showing uncensored observations with time and status axes. Status transitions at time $T_i$.]
Right censored observations

Time

Status

0 1

0 1

0 1

0 1

Time

C_i
Censored data problem

All performance parameters (expected Brier scores, concordance probabilities, etc.) depend on the unknown distribution of the future patients.

This data generating mechanism is best estimated by the empirical distribution of the validation data set.

However, when some patients are lost to follow-up (event-free) before the time horizon \( t \), then their event status is unknown. Also the order of some pairs of patients is unknown (unusable).

In this case the performance parameters have to be estimated using some technique for right censored data to avoid bias.
Observations

Random variables

\[ T \quad \text{event time} \]
\[ C \quad \text{censoring time} \]
\[ \tilde{T} = \min(T, C) \]
\[ \Delta = 1 \{T \leq C\} \]
\[ X \quad \text{predictors} \]

Assume (at least) that \( C \) is conditionally independent of \( T \) given \( X \).

Otherwise the joint distribution of \((T, X)\) is not identifiable from the observations.
To deal with censoring, we need a second model (a nuisance) to estimate the distribution $E_i$ of the validation data set. It is possible (but not recommended) to use the model whose performance is studied.

However, when the aim is to compare two models, the nuisance model should definitely be independent of these two models.

Possible solutions:

- Inverse of probability of censoring weights (IPCW)
- Jackknife pseudo-values

Both depend on a model for the censoring distribution!
IPCW estimate of the expected Brier score

We assume conditional independent censoring and a working model for the conditional censoring distribution $G(t|X) = P(C > t|X)$. Then weights are constructed based on an estimate $\hat{G}$ of $G$:

$$\omega_i(t) = \left\{ \frac{\mathcal{I}\{T_i \leq t, \Delta_i = 1\}}{\hat{G}(T_i - |X_i)} + \frac{\mathcal{I}\{T_i > t, C_i > t\}}{\hat{G}(t|X_i)} \right\}$$

The estimate of the expected Brier score in validation data $^1$

$$\frac{1}{n} \sum_{i \in \text{validation data}} \omega_i(t) \left\{ Y_i(t) - \hat{R}_n(t|X_i) \right\}^2.$$ 

If the working model is correctly specified the estimate is consistent.

$^1$hidden in training of $\hat{R}_n$
IPCW estimate of the time-dependent Brier score

```r
library(pec)
library(survival)
data(pbc)
pbc$logbili <- log(pbc$bili)
fit0 <- coxph(Surv(time,status!=0)~age+sex,data=pbc)
fit1 <- coxph(Surv(time,status!=0)~age+sex+logbili,data=pbc)
plot(pec(list("Cox (age,sex)"=fit0,"Cox (+bili)"=fit1),data=pbc),cens.
    model="cox")
```

![Graph showing prediction error over time for Cox (age, sex) and Cox (+bili) models.](image-url)
Pseudo-value approach to calibration plot

- Fix a prediction horizon \( t \).
- Compute the Kaplan-Meier estimate \( \hat{S} \) in all data.
- Compute the Kaplan-Meier estimate \( \hat{S}^{(i)} \) leaving out the data of subject \( i \).
- Replace the possibly unknown status \( Y_i(t) \) of subject \( i \) by the pseudo-value:

\[
\tilde{Y}_i(t) = n\hat{S}(t) - (n - 1)\hat{S}^{(i)}(t)
\]

Then, if censoring is independent of the predictors, we can estimate the calibration curve by:

\[
p \mapsto \frac{1}{|V|} \sum_{i \in \text{validation set}} \left( \tilde{Y}_i(t) \mathbb{I}\{\hat{R}_n(t|X_i) \in [p - \epsilon; p + \epsilon]\} \right)
\]
Calibration in the large
bandwidth=1

Localized calibration
bandwidth=0

Kernel smoother
automatically selected
bandwidth= 0.194

Kernel smoother
bandwidth=0.1
Exercise 4.1

In the cost data (library(pec); data(cost)) consider the three Cox regression models (same as in exercise 3.3)

- a Cox regression model which includes the 13 predictor variables

- a Cox regression model which uses a restricted cubic spline for the 3 continuous predictor variables.
  \[
  \text{rms::cph(Surv(time,status)~alcohol+rcs(age)+rcs(cholest)+hemor+...)}
  \]

- a Cox regression model which uses backward elimination on the 13 variables
  \[
  \text{ModelGood:::selectCox(Surv(time,status)~alcohol+age+cholest+...)}
  \]

- Compute the cross-validated Brier score with the function pec based on 100 bootstrap samples of size 450 drawn without replacement from the cost data.

- Plot the results
It is sometimes not possible to observe very large or very small values due to a detection limit.

In survival analysis the outcome is time-to-event and large values are not observed when the patient was lost-to-follow-up before the event occurred.

Data are called right-censored when the event for a patient is unknown, but it is known that the event time exceeds a certain value.

A competing risk is an event after which it is clear that the patient will never experience the event of interest.
Competing risk

Speed = 0, arrives never!

Censored

Speed = ? arrival time?
Decision making in the presence of competing risks

Suppose a 40 year old and a 80 year old person need to decide for or against prophylactic coagulation therapy.

Suppose further the predicted risk of dying from cardiovascular disease within the next 10 years for both persons is 12%. How could this happen?

One plausible explanation is that the 40 year old person has other risk factors that the 80 year old person does not have.

Another plausible explanation is that the 80 year old person has a much higher risk to die due to non-cardiovascular disease within the next 10 years than the 40 year old person.
What is the difference between censoring and (non-fatal) competing risks?

- Non-informative censoring does not change the event rates.
- A fatal competing risk changes the event rates from positive to zero.
- Generally the occurrence of a competing risks changes the event rate or changes the interest in the event (example: failure of a filling in a primary tooth).
- In a competing risk model interest is in time to first event. What happens after the first event is not analysed.
- The effect of non-fatal events can also be studied in more complex multi-state models.
Competing risks

Any other event which changes the risk of the event being predicted may be considered a separate state in a competing risk model.

Most commonly this would be dying from other causes.

Competing risks affect all stages of the process from the discovery of markers over modelling and assessment of risk predictions to medical decision making.
Competing risks model

Event-free

Event 1
e.g. cancer

Event 2
e.g. death
without cancer

$\lambda_1(t)$

$\lambda_2(t)$
Setup

- time origin = 0
- Event time: $T$ (time between 0 and an event)
- Observed event time: $\min(T, C)$
- Cause of the event: $D \in \{1, \ldots, K\}$
- Covariate vector $X = X_1, \ldots, X_p$
- $X$ must be measurable at time 0

Event status

$N_k(t) = 1\{T \leq t, D = k\}$

Assess performance of personalized risk prediction

$\hat{R}_n(t|X_i) \approx N_{ik}(t) = 1\{T_i \leq t, D_i = k\}$
Parameters in the presence of competing risks

- The cause-k specific hazard function for a subject characterized by covariate vector $X$

$$\lambda_k(t|X) \approx \text{the probability of an event of type } j \text{ tomorrow [s]}$$
$$\text{given no event until today [s-]}.$$

- The cause-k specific cumulative incidence function

$$F_k(t|X) = P\{ T \leq t, D = k | X \}.$$  
$$= \text{absolute risk of event j before time t}$$
Prediction in the presence of competing risks

First pick a time origin at which it is of interest to predict the future status of a subject.

Until time $t$ after the time origin three things can happen:

1. the event has occurred (e.g. cancer diagnosis)
2. a competing event has occurred (e.g. death without cancer)
3. the patient is alive and event-free.

When the individual observation period of a subject is shorter than time $t$, the subject’s event time is right censored.
Relation between hazards and absolute risks

\[ F_k(t|X) = \]

\[ \int_0^t \exp \left( - \int_0^s \{ \lambda_1(u|X) + \cdots + \lambda_k(u|X) \} \, du \right) \underbrace{\lambda_k(s|X)}_{\text{Event type } j \text{ at } s} \, ds. \]

- a covariate that reduces the cause-specific hazard of a competing risk indirectly increases the cumulative incidence of event \( j \).

- covariates found to change \( F_k \) are those that change any of the cause-specific hazard functions.
Different tasks require different methods

1. We focus on the cause-k specific hazard to identify variables that affect the biology of cause-k events:
   - cause-specific log-rank test:
     \[ H_0 : \lambda_k(t|A) = \lambda_k(t|B) = \lambda_k(t|C) \]
   - cause-specific Cox regression
   - technique: treat competing events as if they were right censored.

2. We focus on the cumulative incidence(s) to predict the risk(s) for patient counseling:
   - Gray test:
     \[ H_0 : F_k(t|A) = F_k(t|B) = F_k(t|C) \]
   - Combine cause-specific Cox regression models
   - Absolute risk regression model
   - Fine-Gray regression model
Predicting event risk in presence of competing risks

The aim is to predict the absolute risk of an event of type 1, say, \( t \)-years after the time origin (or landmark):

\[
F_1(t|X) = \text{cumulative incidence function}
\]

Common regression models for \( F_1 \) are either based on:

\[
F_1(t|X) = \int_0^t \exp \{- \Lambda_1(s|X) - \Lambda_2(s|X)\} \Lambda_1(ds|X)
\]

or

\[
F_1(t|X) = \int_0^t P(\min(T, C) > s|X) \frac{\Lambda_1(ds|X)}{G(s - |X)}
\]
Prediction of absolute risk (formula 1)

\[ F_1(t|X) = \text{Cumulative incidence of event 1} \]

\[
\int_0^t \exp \left( - \int_0^s \{ \hat{\lambda}_1(u|X) + \hat{\lambda}_2(u|X) \} \, du \right) \left( \hat{\lambda}_1(s|X) \right) \, ds.
\]

- No event of any cause until \( s \)
- Event type 1 at \( s \)

- **Cox regression for events of type 1, e.g., stroke hazard:**
  \[
  \hat{\lambda}_1(u|X) = \hat{\lambda}_{01}(u) \exp(\hat{\beta}X)
  \]

- **Cox regression for competing events, e.g., hazard of death other causes:**
  \[
  \hat{\lambda}_2(u|X) = \hat{\lambda}_{02}(u) \exp(\hat{\gamma}X)
  \]
Combination of cause-specific Cox regression

Remarks:

- A covariate that reduces the cause-specific hazard of a competing risk indirectly increases the cumulative incidence of event $j$.

- Covariates found to change the absolute risk of stroke either change the cause-specific hazard of stroke or change the cause-specific hazard of death other causes or change both.

Disadvantages:

- We need to model all competing risks or the overall survival hazard function.

- Changes of the absolute risk depend on changes of a single covariate in a complicated way.
Direct transformation models (formula II)

\[ h\{F_1(t|X)\} = F_{01}(t) + \eta_1^T X \]

- \( h(x) = \log(-\log(x)) \) (Fine-Gray model)
- \( h(x) = \log(x/(1-x)) \) (Logistic model)
- \( h(x) = \log(x) \) (Log-binomial model)

Requires

\[ G(t|X) = \text{regression model for the censoring times} \]

\(^2\text{Fine-Gray model: } \log(-\log(-F_1(t|X))) = \eta X; \text{"sub-hazard ratio" } \eta \text{ has no clear interpretation.}\)
Prediction of absolute risk (formula II)

It would be desirable to have a model in which the regression coefficients $\beta$ have the following interpretation:

*The absolute risk of a stroke during the next ’t’ years changes with a factor $\exp(\beta)$ for a one unit change of age, given fixed values for the other predictor variables.*

Absolute risk regression:

$$F_1(t|X) = F_{01}(t) \exp(\beta X) \text{ (Log-binomial model)}$$

Disadvantages:

- estimation of $\beta$ requires a (regression) model for the probability of not being lost to followup, i.e., a regression model for the censoring times.
- model has numerical problems with effects of continuous covariates

---

Right censored observations

Random variables

\[
\begin{align*}
T & \quad \text{event time} \\
D & \quad \text{type of event} \\
C & \quad \text{censoring time} \\
\tilde{T} & = \min(T, C) \\
\Delta & = 1\{T \leq C\} \\
\tilde{D} & = D \times \Delta \\
X & \quad \text{predictors}
\end{align*}
\]

Assume (at least) that \( C \) is conditionally independent of \( T \) given \( X \).

Otherwise the joint distribution of \((T, X)\) is not identifiable from the observations.
Brier score in the presence of competing risks

\[
\text{Brier}(t, k) = \mathbb{E}_{T, D, X} \left\{ N_k(t) - \hat{R}_n(t|X_i) \right\}^2
\]

- Null model is Aalen-Johansen estimate

IPCW estimate of the expected Brier score (cause k)

We need a "working model" and a corresponding estimate \( \hat{G} \) for the conditional censoring distribution \( G(t|X) = P(C > t|X) \)

\[
\frac{1}{n} \sum_{i \in \text{validation data}} \left\{ \frac{\mathcal{I}\{T_i \leq t, \Delta_i = 1\}}{\hat{G}(T_i - |X_i)} + \frac{\mathcal{I}\{T_i > t, C_i > t\}}{\hat{G}(t|X_i)} \right\} \left\{ N_{ik}(t) - \hat{R}_n(t|X_i) \right\}^2.
\]

If the working model is correctly specified the estimate is consistent.
AUC in the presence of competing risks

\[ AUC_k(t) = \]
\[ \mathbb{E}_i \mathbb{E}_j (I\{\hat{R}_n(t, X_i) > \hat{R}_n(t, X_j)\} | D_i = k, T_i \leq t, (T_j > t \text{ or } D_j \neq k)) \]

AUC$_1(t)$ is the probability that a random subject who experienced an event of type D=1 has received a higher predicted risk at baseline than another random subject who did not experience event of type D=1 within $t$ years, i.e., is either alive and event free or died before time $t$ without event type D=1.
IPCW estimate of AUC (cause $k$)

Weights:

\[
\hat{W}_{ij,1} = \frac{\mathcal{I}\{T_i < T_j, T_i < C_i\}}{\hat{G}(T_i - |X_i|) \hat{G}(T_i|X_j)}
\]

\[
\hat{W}_{ij,2} = \frac{\mathcal{I}\{T_i \geq T_j, D_j \neq k, T_j \leq C_j\}}{\hat{G}(T_i - |X_i|) \hat{G}(T_j - |X_j|)}
\]

IPCW estimate:

\[
\sum_{i,j \in \text{validation data}} \left( \hat{W}_{ij,1} + \hat{W}_{ij,2} \right) \mathcal{I}\{\hat{R}_n(t,X_i) > \hat{R}_n(t,X_j)\} \frac{N_{ik}(t)}{\sum_{i,j \in \text{validation data}} \left( \hat{W}_{ij,1} + \hat{W}_{ij,2} \right) N_{ik}(t)}
\]

Note: \( \mathcal{I}\{\hat{R}_n(t,X_i) > \hat{R}_n(t,X_j)\} \) may change over time.
Cause-specific calibration curves

A calibration plot (is still) a visualizing the distance between predicted and expected event probabilities.

In the presence of competing risks:

- Fix a prediction horizon $t$.
- Obtain pseudo values for the possibly unknown status $Y_{ik}(t)$ with the Aalen-Johansen estimate:

$$\tilde{Y}_{ik}(t) = n\hat{F}_k(t) - (n - 1)\hat{F}_k^{(i)}(t)$$

Then if censoring is independent of $X$ we can estimate the calibration curve by:

$$p \mapsto \frac{1}{|V|} \sum_{i \in \text{validation set}} \left( \tilde{Y}_{ik}(t) \mathbb{I}\{\hat{F}_k(t|X_i) \in [p - \epsilon; p + \epsilon]\} \right)$$
Example

Risk reclassification

Multiple regression with ERG status

2-yr prediction horizon

ERG status

○ Negative

● Positive

Multiple regression without ERG status

Calibration

Observed probabilities

2-yr prediction horizon

Prediction error

Reference (no predictor variables)

Multiple regression without ERG status

Multiple regression with ERG status

Discrimination ability

Reference (no predictor variables)

Multiple regression without ERG status

Multiple regression with ERG status

Years on AS

Years on AS
Summary and conclusions

In survival analysis predictions and prediction performance are time-dependent. Performance is a parameter (of the distribution of the validation data) which needs to be estimated usually in the presence of right censored data.

With competing risks the way we do things need some slight adaptation, there are pitfalls, and some formula get more complex, but generally everything seems to be under control.

To interpret a prediction in most applications with competing risks we have to build several models, one for each competing risk.
Exercise 4.2

- Work through the appendix of Absolute risk regression (see course homepage).
- In the pbc data library(survival); data(pbc)
- Compute a cause-specific Cox regression model with the function CSC
- Compute a Fine-Gray regression model for cause 2 with the function FGR
- Predict the event risk of cause 2 with both methods, assess and compare their prediction performance. Hint: use library(pec) for Brier score and library(timeROC) for ROC and AUC.