Modeling long term survival data: beyond the proportional hazards model

Hans C. van Houwelingen
Department of Medical Statistics and Bioinformatics
Leiden University Medical Center
The Netherlands

jcvanhouwelingen@lumc.nl
Abstract

In modeling long term survival data the assumption of proportional hazards, used in the widely applied Cox regression, will usually not hold.

The talk will discuss different extensions of the Cox model that allow for non-proportional hazards: time-varying effect model, frailty model and cure model. It will be argued that the choice of a model heavily depends on the questions to be answered and that in most cases simple questions can be answered by simple models.
Personal experiences with registry data

• Mainly registries of follow-up data of patients with similar diagnosis,
  o mostly some type of cancer, such as
    ▪ ovarian cancer
    ▪ breast cancer (EORTC, Dutch Comp Cancer Centers, Greek Breast Cancer Registration)
    ▪ leukemia (EBMT, SKION (childhood cancers)
  o transplantation data (Eurotransplant)

• Early phase of follow-up based on clinical trial (hospital) data.Later phase could involve linking to other data bases, like the Dutch mortality registration
Schematic presentation

Intake
Diagnosis
Start Treatment

End
Treatment

Intermediate
events

Death
Loss of follow-up

Information
at baseline

Evaluation
of Treatment

Follow-up information
Questions to be answered from the data

1. What are the (x-year) survival probabilities at the start (t=0)?
2. How do these probabilities change over time when new information might become available?
Primary questions

1. What are the (x-year) survival probabilities for a patient after diagnosis/start of treatment?

2. How do these probabilities depend on
   a. demographic information: age, gender, ...
   b. diagnostic information: disease stage, performance, ...
   c. treatment characteristics
      i. surgery
      ii. radiation
      iii. (adjuvant) chemo
   d. response to treatment (actually, a secondary question)
Model of choice in clinical practice

- Kaplan-Meier survival curves
- Cox Proportional Hazards Regression, followed by Kaplan-Meier curves for subgroups. The use of the Kaplan-Meier curves is caused by lack of understanding and/or mistrust of the Proportional Hazards model
Problem for statistical modelers

The “standard” Proportional Hazards model is inadequate for long term survival data.

Explanations

1. Heterogeneity (omitted covariates, frailty)
2. Measurement error
3. “Aging” of covariates measured at baseline, time-dependent frailty

Remedies

1. Extension of the model
2. “Trimming” of the data, followed by
   i. PH Cox modeling
   ii. something simpler
Example:

**Greek breast cancer data** (Perperoglou et al., SiM, 2007)

2433 women with operable breast cancer

![Survival function](image)

Figure 1. Survival function (Kaplan–Meier) of breast cancer patients. Dotted line presents the censoring distribution.
Cox model

hazard (probability of dying tomorrow, if you are alive today)

\[ h(t | X_1, ..., X_k) = h_0(t) \exp(\beta_1 X_1 + .. + \beta_k X_k) \]

survivor function

\[ S(t | X_1, .., X_k) = \exp(-H_0(t) \exp(\beta_1 X_1 + .. + \beta_k X_k)) \]

with

\[ H_0(t) = \int_0^t h_0(s)ds \] (cumulative base-line hazard)
Cox model on relevant prognostic variables

Table I. Results from Cox regression, age in years, tumour size (in mm), LN+ number of positive lymph nodes, tumour Grade (Bloom-Richardson), chemotherapy (0 no, 1 yes), radiotherapy (0 no, 1 yes) and hormonal treatment (0 no, 1 yes).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE (Coefficient)</th>
<th>z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.017</td>
<td>0.003</td>
<td>4.804</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Tumour</td>
<td>0.011</td>
<td>0.002</td>
<td>4.539</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ln+</td>
<td>0.032</td>
<td>0.003</td>
<td>8.703</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Grade</td>
<td>0.373</td>
<td>0.070</td>
<td>5.299</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>0.006</td>
<td>0.099</td>
<td>0.060</td>
<td>0.951</td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>-0.208</td>
<td>0.095</td>
<td>-2.179</td>
<td>0.029</td>
</tr>
<tr>
<td>Hormonal</td>
<td>-0.623</td>
<td>0.087</td>
<td>-7.113</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Extension of the Cox-model: time-varying effects (B-splines)

\[ h(t \mid X_1, \ldots, X_k) = h_0(t) \exp(\beta_1(t)X_1 + \ldots + \beta_k(t)X_k) \]
Explanations for the time-varying effects

1. Heterogeneity (omitted covariates, *frailty*)
2. Measurement (misclassification) error
3. “Aging” of covariates measured at baseline, time-dependent frailty

Prediction can be built based on

1. Plausible explanation of time-varying effects (as above)
2. Parsimonious extension of the Cox model
General frailty model

\[ h(t \mid X_1, \ldots, X_k, Z) = Z h_0(t) \exp(\beta_1 X_1 + \ldots + \beta_k X_k) \]

with unobservable \( Z \)

Observable hazard

\[ h^* (t \mid X_1, \ldots, X_k) = h_0(t) \exp(\beta_1 X_1 + \ldots + \beta_k X_k) E[Z \mid T \geq t, X] \]

The selection factor \( E[Z \mid T \geq t, X] \) causes the violation of the PH-model.
Popular frailty models

Gamma-frailty model

\[ Z \sim \text{Gamma}(1 / \xi, 1 / \xi), \quad E[Z] = 1, \quad \text{var}(Z) = \xi \]

leading to the Burr model

\[
h^*(t \mid X_1, \ldots, X_k) = \frac{h_0(t) \exp(\beta_1 X_1 + \ldots + \beta_k X_k)}{1 + \xi H_0(t) \exp(\beta_1 X_1 + \ldots + \beta_k X_k)}
\]

(Highly significant in the Greek data)
Cure model

\[ Z = 0 \text{ (cured)} \text{ or } Z = 1 \text{ (not-cured)} , \ P(Z = 1) = \pi \]

Selection factor

\[
E[Z \mid T \geq t, X] = \frac{(1 - \pi)\exp(-H_0(t)\exp(X_1\beta_1 + \cdots + X_k\beta_k))}{(1 - \pi)\exp(-H_0(t)\exp(X_1\beta_1 + \cdots + X_k\beta_k)) + \pi}
\]

Survival model

\[ S(t \mid X) = \pi + (1 - \pi)\exp(-H_0(t)\exp(X_1\beta_1 + \cdots + X_k\beta_k)) \]

You might also let \( \pi \) depend on the covariates through a logistic regression model. **Beware of over-interpreting** \( \pi \)
Application to the Greek data

Table V. Estimated coefficients, standard errors and Wald statistics (W) from a semiparametric cure model.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>SE ($\beta$)</th>
<th>W</th>
<th>$\beta^*$</th>
<th>SE ($\beta^*$)</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.017</td>
<td>0.003</td>
<td>4.746</td>
<td>-0.032</td>
<td>0.021</td>
<td>-1.549</td>
</tr>
<tr>
<td>Age</td>
<td>0.013</td>
<td>0.002</td>
<td>5.234</td>
<td>-0.036</td>
<td>0.013</td>
<td>-2.777</td>
</tr>
<tr>
<td>Tumour</td>
<td>0.051</td>
<td>0.004</td>
<td>12.612</td>
<td>-0.173</td>
<td>0.021</td>
<td>-8.318</td>
</tr>
<tr>
<td>Ln+</td>
<td>0.292</td>
<td>0.072</td>
<td>4.068</td>
<td>0.564</td>
<td>0.389</td>
<td>1.452</td>
</tr>
<tr>
<td>Grade</td>
<td>-0.063</td>
<td>0.100</td>
<td>-0.631</td>
<td>-0.633</td>
<td>0.986</td>
<td>-0.643</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>-0.289</td>
<td>0.097</td>
<td>-2.975</td>
<td>0.821</td>
<td>0.622</td>
<td>1.320</td>
</tr>
<tr>
<td>Hormtherapy</td>
<td>-0.7504</td>
<td>0.087</td>
<td>-8.628</td>
<td>3.786</td>
<td>0.712</td>
<td>5.312</td>
</tr>
</tbody>
</table>

effect on hazard  
effect on cure rate
Alternative: parsimonious extension of the Cox model via reduced rank. (Main focus of Perperoglou’s research)

Simplest model: rank=1 model

\[ h(t \mid X_1, \ldots, X_k, Z) = h_0(t) \exp((\beta_1 X_1 + \ldots + \beta_k X_k) f(t \mid \gamma)) \]

\[ \beta_i(t) = \beta_i f(t \mid \gamma) \]
It is very hard to tell the difference between the frailty models and related models as described in Perperoglou’s paper(s).

All models give about the same prediction and perform equally well.
**Remedies**

1. **Extension of the model** *(already discussed)*

2. “Trimming” of the data, followed by
   i. PH Cox modeling
   ii. something simpler

Instead of adapting the model, it might me easier to adapt the question.
Recall (previous slides)

Model of choice in clinical practice

- Kaplan-Meier survival curves
- Cox Proportional Hazards Regression, followed by Kaplan-Meier curves for subgroups. The use of the Kaplan-Meier curves is caused by lack of understanding and/or mistrust of the Proportional Hazards model

Primary questions

1. What are the survival probabilities for a patient after diagnosis/start of treatment?
2. How does that probability depend on
   a. demographic information: age, gender, ...
   b. diagnostic information: disease stage, performance, ...
   c. treatment characteristics
Instead of all the complex modeling, we might as well answer our primary questions directly by modeling the survival function \( S(t \mid X) = P(T \geq t \mid X) \) for a range of \( t \) values, for example \( t=1,2,3,... \) years

**Complication:**
Censoring

**Solution(s)**
Jackknife (Andersen, Klein) + logistic regression
(In Perperoglou’s paper)
Trimming + Cox regression
(Like in my Landmarking paper (SJoS, 2007))
Creating pseudo-observations by Jackknifing

\[ PsObs_i(t) = nKM(t) - (n - 1)KM^{(-i)}(t) \]

is an unbiased estimated of \([T_i \geq t]\).

Can be used in logistic regression if the software like GEE allows an outcome that is not always equal to 0 or 1 (and uses a robust estimate of the standard error)
**Trimming + Cox Regression**

To obtain an estimate for the survival at some $t_{\text{horizon}}$

1. Censor all data with $t > t_{\text{horizon}}$
2. Fit a simple Cox model to the data
3. Use that **only** to estimate $P(T \geq t_{\text{horizon}} \mid X)$

This will produce a robust estimate, if there is not too much censoring.

(If there is heavy censoring, some weighting might be needed.

Van Houwelingen et. al, SiM, 2005)
As a by-product you get an “new” view on the time varying effects

<table>
<thead>
<tr>
<th>horizon</th>
<th>events</th>
<th>age</th>
<th>tumour</th>
<th>lnpos</th>
<th>grade</th>
<th>chemo</th>
<th>radio</th>
<th>hormonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>138</td>
<td>.013</td>
<td>.015</td>
<td>.023</td>
<td>.693</td>
<td>.206</td>
<td>-.095</td>
<td>-1.160</td>
</tr>
<tr>
<td>3</td>
<td>239</td>
<td>.012</td>
<td>.014</td>
<td>.028</td>
<td>.730</td>
<td>.117</td>
<td>-.060</td>
<td>-.918</td>
</tr>
<tr>
<td>4</td>
<td>305</td>
<td>.009</td>
<td>.014</td>
<td>.028</td>
<td>.648</td>
<td>.028</td>
<td>-.091</td>
<td>-.769</td>
</tr>
<tr>
<td>5</td>
<td>356</td>
<td>.011</td>
<td>.014</td>
<td>.030</td>
<td>.596</td>
<td>-.038</td>
<td>-.046</td>
<td>-.778</td>
</tr>
<tr>
<td>6</td>
<td>404</td>
<td>.011</td>
<td>.014</td>
<td>.030</td>
<td>.510</td>
<td>-.029</td>
<td>-.075</td>
<td>-.762</td>
</tr>
<tr>
<td>7</td>
<td>452</td>
<td>.013</td>
<td>.014</td>
<td>.030</td>
<td>.462</td>
<td>.004</td>
<td>-.170</td>
<td>-.736</td>
</tr>
<tr>
<td>8</td>
<td>480</td>
<td>.014</td>
<td>.013</td>
<td>.028</td>
<td>.478</td>
<td>.006</td>
<td>-.153</td>
<td>-.704</td>
</tr>
<tr>
<td>9</td>
<td>504</td>
<td>.015</td>
<td>.013</td>
<td>.030</td>
<td>.462</td>
<td>.010</td>
<td>-.141</td>
<td>-.675</td>
</tr>
<tr>
<td>10</td>
<td>523</td>
<td>.015</td>
<td>.014</td>
<td>.030</td>
<td>.446</td>
<td>.012</td>
<td>-.115</td>
<td>-.637</td>
</tr>
<tr>
<td>11</td>
<td>542</td>
<td>.015</td>
<td>.013</td>
<td>.030</td>
<td>.425</td>
<td>.011</td>
<td>-.121</td>
<td>-.621</td>
</tr>
<tr>
<td>12</td>
<td>560</td>
<td>.015</td>
<td>.013</td>
<td>.031</td>
<td>.422</td>
<td>-.035</td>
<td>-.145</td>
<td>-.648</td>
</tr>
</tbody>
</table>
Some graphics for

**Grade**

Weaker effect over time

**Radiotherapy**

Stronger effect over time
Secondary question

Dynamic adjustment of the prognosis

1. Only based on initial information

2. Taking into account
   a. Follow-up information (time-dependent covariates)
   b. Intermediate events (multi-state)

Solutions

1. Joint model for all possible data

2. “Landmarking”
No time for details.

Just an illustration based on Van Houwelingen & Putter, LIDA, 2008, modeling data from the EBMT registry.
Landmark models can be used directly for dynamic prediction.

Main predictors

1. Age in three groups
2. Occurrence of Acute Graft versus Host Disease (AGvHD)
3. Recent experience (< 3 months) of Platelet Recovery
Fig. 3  Estimated 5-year failure-free survival probabilities for subject with no TCD and transplanted after 1989 based on the landmark model. In light grey age <20, in grey age 20–40, in black age ≥40
Conclusion

Be robust, keep it simple!
References

Long term survival and time-varying effects

Landmarking


Pseudo-observations

- Klein, JP; Gerster, M; Andersen, PK; Tarima, S; Perme, MP. 2008. SAS and R functions to compute pseudo-values for censored data regression. *COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE* 89 (3): 289-300.