The Landmark Approach: An Introduction and Application to Dynamic Prediction in Competing Risks

Hein Putter

Department of Medical Statistics and Bioinformatics
Leiden University Medical Center

Dynamic prediction workshop, Bordeaux
October 10, 2013
Outline

Landmarking
  Background

Dynamic prediction
  Why dynamic prediction?
  Illustration
  Multi-state approach

Dynamic prediction and landmarking
  Basic idea
  Landmark (super) models
  Landmarking in action

Landmarking and competing risks
  Competing risks
  Dynamic pseudo-observations

Discussion
Landmarking

Origin of landmarking

- Common way of analysis: make two groups, a "responder" group and a "non-responder" group and compare survival between these two groups
- Problem with this approach: a potential responder will only belong to the "responder" group if he/she survives until time of response
- Individuals in the responder group are immortal for some time, this gives them an unfair survival advantage: immortal time bias
Time-dependent covariates

- The problem comes in a number of disguises
  - Effect of recurrence on survival in cancer
  - Effect of transplant failure on survival in transplant studies
  - Effect of compliance on recurrence
  - Effect of drug-specific adverse events on recurrence
  - Effect of winning an Oscar on survival for US actors (*Ann Intern Med*)

- Unfortunately the incorrect approach is still prevalent in medical journals
Correct approaches

- Crucial issue: "responder" versus "non-responder" is something that is not known at baseline
- When studying survival, it is not allowed to make groups based on something that will happen in the future
- Two alternatives proposed
  - Time-dependent covariate
  - Landmark
    - Consider response at fixed point in time (landmark)
    - Remove patients with event (or censored) before landmark from analysis
Dynamic prediction

- Prediction is often well known from start treatment/diagnosis/...
- Depends on patient characteristics known at baseline
- Patient comes back for regular (6 months eg) checks
  - Baseline covariates have not changed
  - But event history (clinical events) may have changed
  - Biomarkers ...
- As a result, prognosis will have changed
  - Also if patient has had no events
- Prediction needs to be updated (dynamic prediction)
US Health and Retirement Study (HRS)

- HRS: number of nationally representative cohorts
- Here only oldest cohort (AHEAD) used, born before 1923, aged 70 and older in 1993
- Outcome of interest: overall survival; time scale is age
- Time-fixed covariates $Z$ at entry: gender, education, BMI, smoking
- Time-dependent covariate

$$Z_{\text{ADL}}(t) = \begin{cases} 
1, & \text{if subject is ADL-disabled at age } t; \\
0, & \text{otherwise.}
\end{cases}$$

- Basic Activities of Daily Living (walking, bathing, dressing, toileting, feeding), disabled if "with difficulty" for $\geq 1$ item
US Health and Retirement Study (HRS)

- HRS: number of nationally representative cohorts
- Here only oldest cohort (AHEAD) used, born before 1923, aged 70 and older in 1993
- Outcome of interest: overall survival; time scale is age
- Time-fixed covariates $Z$ at entry: gender, education, BMI, smoking
- Time-dependent covariate

$$Z_{ADL}(t) = \begin{cases} 
1, & \text{if subject is ADL-disabled at age } t; \\
0, & \text{otherwise.}
\end{cases}$$

- Basic Activities of Daily Living (walking, bathing, dressing, toileting, feeding), disabled if "with difficulty" for $\geq 1$ item
- Objective: dynamic prediction of survival of at least 10 years beyond age $s$, with given ADL-disability status at age $s$ and given covariates
## Covariates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>N</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1564</td>
<td>(39%)</td>
</tr>
<tr>
<td>Female</td>
<td>2468</td>
<td>(61%)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>1736</td>
<td>(43%)</td>
</tr>
<tr>
<td>High school</td>
<td>1212</td>
<td>(30%)</td>
</tr>
<tr>
<td>Some college</td>
<td>1084</td>
<td>(27%)</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\leq$ 25</td>
<td>2244</td>
<td>(56%)</td>
</tr>
<tr>
<td>25 – 30</td>
<td>1388</td>
<td>(34%)</td>
</tr>
<tr>
<td>$&gt; 30$</td>
<td>390</td>
<td>(10%)</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1997</td>
<td>(50%)</td>
</tr>
<tr>
<td>Past</td>
<td>1683</td>
<td>(42%)</td>
</tr>
<tr>
<td>Current</td>
<td>324</td>
<td>(8%)</td>
</tr>
</tbody>
</table>
Multi-state approach

- Multi-state model; process $X(t)$ in time, taking values
  - 0: alive without ADL disability
  - 1: alive with ADL disability
  - 2: dead

- Reversible illness-death model
Multi-state approach

- Multi-state model; process $X(t)$ in time, taking values
  - 0: alive without ADL disability
  - 1: alive with ADL disability
  - 2: dead

- Reversible illness-death model

- Objective: estimation of $P(X(s + w) < 2|X(s) = 0, Z^*)$ and $P(X(s + w) < 2|X(s) = 1, Z^*)$ for given $Z^*$
Estimated transition hazards

- Healthy→Disabled
- Healthy→Dead
- Disabled→Healthy
- Disabled→Dead
Dynamic prediction for the multi-state model

If the Markov assumption holds

- There are no easy closed-form expressions for the prediction probabilities (because the multi-state model is reversible)
- The Aalen-Johansen estimator can be used to obtain estimates of prediction probabilities but is laborious
- Implemented in \texttt{mstate}

Otherwise

- If the Markov assumption does not hold, we would have to use (micro-)simulation
  - In our data, there is evidence that history of ADL-disability increases disability rate ($\Rightarrow$ violation of Markov assumption)
Dynamic prediction and landmarking

- Idea to use landmarking for dynamic prediction stems from van Houwelingen (2007)
- Suppose we want to estimate the probability, given alive at age 80, of survival until age 90
Dynamic prediction and landmarking

- Idea to use landmarking for dynamic prediction stems from van Houwelingen (2007)
- Suppose we want to estimate the probability, given alive at age 80, of survival until age 90
- The basic idea
  - Suppose that we had an enormous database at our disposal
  - We would select a subset of the data, consisting of everyone alive at age 80
Dynamic prediction and landmarking

▷ Idea to use landmarking for dynamic prediction stems from van Houwelingen (2007)

▷ Suppose we want to estimate the probability, given alive at age 80, of survival until age 90

▷ The basic idea
  ▷ Suppose that we had an enormous database at our disposal
  ▷ We would select a subset of the data, consisting of everyone alive at age 80 (a landmark data set)
Dynamic prediction and landmarking

Idea to use landmarking for dynamic prediction stems from van Houwelingen (2007)

Suppose we want to estimate the probability, given alive at age 80, of survival until age 90

The basic idea

Suppose that we had an enormous database at our disposal

We would select a subset of the data, consisting of everyone alive at age 80 (a landmark data set)

And simply count how many are alive at age 90 and calculate proportion
Dynamic prediction and landmarking

- Idea to use landmarking for dynamic prediction stems from van Houwelingen (2007)
- Suppose we want to estimate the probability, given alive at age 80, of survival until age 90
- The basic idea
  - Suppose that we had an enormous database at our disposal
  - We would select a subset of the data, consisting of everyone alive at age 80 (a landmark data set)
  - And simply count how many are alive at age 90 and calculate proportion
  - If there is censoring, we would estimate the probability using Kaplan-Meier
  - If there are also covariates involved, we could incorporate them in a Cox model
Landmarking in general terms

For each of a set of landmark time points \( s \in [s_0, s_1] \)

- Construct corresponding landmark data set, by selecting all individuals at risk at \( s \)
- Define \( Z(s) \): current vector of predictors, including intermediate events (depends on landmarking time point \( s \))
- Fit simple Cox model

\[
\lambda(t \mid Z(s), s) = \lambda_0(t \mid s) \exp(\beta(s)^\top Z(s))
\]

for \( s \leq t \leq t_{\text{hor}} \), enforcing administrative censoring at \( t_{\text{hor}} \)

- After having obtained estimates \( \hat{\beta}(s) \) and \( \hat{\Lambda}_0(t \mid s) \):
- Estimate of prediction probability \( P(T > t_{\text{hor}} \mid T > s, Z^*(s)) \) is then given by

\[
\exp(-\exp(\hat{\beta}(s)^\top Z^*(s))\hat{\Lambda}_0(t_{\text{hor}} \mid s))
\]
Robustness

- **Note**: for fixed $s$ and $t_{\text{hor}}$, the Cox model

\[
\lambda(t \mid Z(s), s) = \lambda_0(t \mid s) \exp(\beta(s)^\top Z(s))
\]

uses $Z(s)$ as time-fixed covariates and $\beta(s)$ as time-fixed covariate effects.

- Xu & O’Quigley (2000) and van Houwelingen (2007): *even if the effect of $Z(s)$ is time-varying*, the above model give accurate (dynamic) predictions provided
  - Administrative censoring is enforced at $t_{\text{hor}}$ during estimation of the Cox model
  - Prediction is only used at $t_{\text{hor}}$
Combining information

- Estimate parameters by fitting simple Cox model

\[
\lambda(t \mid Z(s), s) = \lambda_0(t \mid s) \exp(\beta(s)^\top Z(s))
\]

for \( s \leq t \leq t_{\text{hor}} \), enforcing administrative censoring at \( t_{\text{hor}} \)

- Can be done for each landmark point separately

- But we would expect the coefficients \( \beta(s) \) to depend on \( s \) in a smooth way

- Can use splines or parametric model, eg

\[
\beta(s) = \beta_0 + \beta_1 s
\]
How to implement it

- Fitting this combined model can be done using standard software
  - Stack the landmark data sets
  - Stratify by landmark
- Estimated coefficients are correct, but for standard errors we need correction for the fact that data of the same patient are used repeatedly
  - Sandwich estimators (Lin & Wei, 1989)
- Baseline hazard estimated by Breslow estimator
- Depends on \( s \) unless both \( Z(s) \) and \( \beta(s) \) are constant
Baseline hazards

- Baseline hazards for different landmark time points $s$ may be combined
- To add more structure and to make it easier to interpret the models
- We may assume a model
  \[
  \lambda_0(t \mid s) = \lambda_0(t) \exp(\theta(s))
  \]
  with $\theta(s_0) = 0$ for identifiability
- In our application we take
  \[
  \theta(s) = \theta_1 s + \theta_2 s^2
  \]
- Model can be fitted directly by applying a simple Cox model to the stacked data set
- Landmark time $s$ not used as stratifying variable but as covariate
Landmarking in action

Set-up

- Endpoint is survival in a window of fixed width $w = 10$ years from the moment of prediction.
- Landmark time points used: 16 points, equally spaced, from age 75 to age 90.
- For each landmark (prediction) time point, construct landmark data set, containing all relevant information needed for the prediction.
- In all data sets we take all patients still at risk (alive), compute the current value of ADL-disability, and set the horizon at $t_{\text{hor}} = t_{\text{LM}} + 10$ years.
- At each landmark point we fit a simple Cox model on $(t_{\text{LM}}, t_{\text{hor}})$ and use that to obtain a prediction of survival at $t_{\text{hor}} + 10$. 
Landmarking in action

The landmark data sets

The diagram shows the landmark data sets for ADL disabled and Healthy individuals over the years 75 to 90. The data is represented in bars, with the height indicating the number of observations for each year and the color indicating whether the individual is ADL disabled or Healthy.
Regression coefficients

Regression coefficients with 95% confidence intervals

- SMOpast
- SMOcurrent
- BMI325–30
- BMI3>30
- ADL
- disabled
- educ3HighSchool
- educ3SomeCollege
- sexfemale

Landmark age
- 75
- 80
- 85
- 90

Regression coefficient
- −1.0
- −0.5
- 0.0
- 0.5
- 1.0
- 1.5

Landmarking in action

Dynamic prediction and landmarking

Landmarking and competing risks

Discussion
## Landmark super model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Category</th>
<th>B</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female</td>
<td>-0.465</td>
<td>0.063</td>
</tr>
<tr>
<td>Education</td>
<td>High school</td>
<td>-0.111</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>College</td>
<td>-0.234</td>
<td>0.072</td>
</tr>
<tr>
<td>BMI</td>
<td>25–30</td>
<td>-0.344</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>&gt; 30</td>
<td>-0.135</td>
<td>0.098</td>
</tr>
<tr>
<td>ADL</td>
<td>ADL disabled</td>
<td>0.636</td>
<td>0.050</td>
</tr>
<tr>
<td>Smoking</td>
<td>Past smoker</td>
<td>0.389</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>$\times \bar{s}$</td>
<td>-1.020</td>
<td>0.586</td>
</tr>
<tr>
<td></td>
<td>$\times \bar{s}^2$</td>
<td>0.739</td>
<td>0.506</td>
</tr>
<tr>
<td></td>
<td>Current smoker</td>
<td>1.024</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>$\times \bar{s}$</td>
<td>-1.460</td>
<td>0.810</td>
</tr>
<tr>
<td></td>
<td>$\times \bar{s}^2$</td>
<td>0.538</td>
<td>0.794</td>
</tr>
<tr>
<td>$\theta(s)$</td>
<td>$\bar{s}$</td>
<td>0.971</td>
<td>0.424</td>
</tr>
<tr>
<td></td>
<td>$\bar{s}^2$</td>
<td>0.021</td>
<td>0.356</td>
</tr>
</tbody>
</table>

$$\bar{s} = (s - 75)/15$$
Landmarking in action

With corresponding baselines

![Graphs showing cumulative hazard and exponential of theta(s) vs age.](image-url)

Dynamic prediction Hein Putter
**Regression coefficients**

Blue lines are the (landmark-varying) supermodel effect estimates

![Bar chart showing regression coefficients for different factors such as SMOpast, SMOcurrent, BMI3-30, BMI3>30, ADLADL disabled, educ3HighSchool, educ3SomeCollege, and sexfemale.](chart.png)
Dynamic predictions from the landmark super model

- Male – Never smoker
- Male – Past smoker
- Male – Current smoker
- Female – Never smoker
- Female – Past smoker
- Female – Current smoker

Age at prediction
Probability of surviving next 10 years
0.2
0.4
0.6
0.8
75 80 85 90
Healthy
ADL disabled
Software

**dynpred**

- It is not so difficult to write your own code in the statistical package of your choice
- In R, package *dynpred* is available on CRAN ([cran.r-project.org](http://cran.r-project.org))
  - The companion package of the book "Dynamic Prediction in Clinical Survival Analysis" by Hans van Houwelingen and myself (Chapman & Hall)
  - Functions available to create landmark data sets, applying administrative censoring at horizon (*cutLM*), and to calculate dynamic "death within window" curves (*Fwindow*)
- On the book website [http://www.msbi.nl/DynamicPrediction](http://www.msbi.nl/DynamicPrediction), R code (using the *dynpred* package) of all the analyses in the book is available for download
**Data (EBMT)**

- 5582 CML patients with transplantation (SCT) between 1997 and 2003
- Two competing risks: "relapse" (Rel) and "non-relapse mortality" (NRM)
Events of interest

- Objective:
  - To give prognosis of the disease/recovery process after SCT for a patient with a given post-transplant history

- Prediction to be based on covariates
  - Baseline: year of SCT, risk score (low, medium, high)
  - Time-dependent: Acute Graft-versus-Host-Disease (aGvHD)
    - Low or high grade

![Cumulative incidences of aGvHD](image)
Time-dependent covariates

- Define

\[ Z_{\text{low}}(t) = \begin{cases} 1 & \text{occurrence of low grade aGvHD before time } t \\ 0 & \text{otherwise} \end{cases} \]
\[ Z_{\text{high}}(t) = \begin{cases} 1 & \text{occurrence of high grade aGvHD before time } t \\ 0 & \text{otherwise} \end{cases} \]

Objective

- For a patient, alive without relapse at time \( s \) after SCT, with given covariates \( Z^*(s) \), what is the probability that he/she will
  - have had relapse before time \( s + w \)
  - have died without relapse before time \( s + w \)
- In our application: \( w \) is five years
Competing risks: notation

- $J$ types of failure
- $\tilde{T}$ time of failure, $C$ censoring time, $T = \min(\tilde{T}, C)$
- $D$: type of failure, $\Delta = D$ if failure occurred, 0 otherwise
- $Z(t)$ vector of covariates (possibly time-dependent)
- We observe $(T_i, \Delta_i, Z_i(\cdot))$ for individual $i$
- Assume censoring is independent of $(\tilde{T}, D)$ given covariates
- Cumulative incidences

$$F_j(t|s) = P(T \leq t, D = j| T > s)$$
Landmarking and competing risks

Approach based on cause-specific hazards

- Select a number of landmark time points
- For landmark time point \( s \)
  - Construct landmark data set, by selecting subjects at risk at \( s \); denote by \( \mathcal{D}_s \)
  - Define \( Z(s) \): current (at \( s \)) vector of predictors
  - Fit Cox models on the cause-specific hazards

\[
\lambda_j(t \mid Z(s), s) = \lambda_{j0}(t \mid s) \exp(\beta_j(s)^\top Z(s))
\]

- Obtain estimates \( \hat{\beta}_j(s) \) and \( \hat{\lambda}_{j0}(t \mid s) \)
- For given (new) patient calculate patient-specific
  \[
  \hat{\lambda}_j(t \mid Z^*(s), s) = \hat{\lambda}_{j0}(t \mid s) \exp(\hat{\beta}_j(s)^\top Z^*(s))
  \]
- Calculate
  \[
  \hat{F}_j(t \mid Z^*(s), s) = \int_s^t \hat{\lambda}_j(u \mid Z^*(s), s) \hat{S}(u - \mid Z^*(s), s) du
  \]
Cause-specific hazards approach

- Cortese & Andersen (2010) looked at fixed set of landmark time points and over the whole future
  - No administrative censoring applied
- Nicolaie et al. (2012): interested in fixed width predictions
  - To be obtained at a continuum of prediction time points
  - Administrative censoring applied (robust against violations of proportional hazards)
  - Super models, combining different landmark models in one, were used
- For both approaches
  - Covariate effects are expressed in terms of the rates (cause-specific hazards), not directly on the risks (cumulative incidences)
- Fine-Gray type approach combined with landmarking
  - Done in Cortese & Andersen (2010) approach (straightforward, because no super models)
  - Not directly for Nicolaie et al. (2012)
Modeling: intuition from complete data

Fix cause of interest \( j \) and fix \( s \)

- For individual \( i \) within \( D_s \), the survivors at \( s \), define
  \[
  Y_i = 1\{ T_i \leq s + w, D_i = j \}
  \]

- Expectation of \( Y_i \) in \( D_s \) is \( P(T_i \leq s + w, D_i = j \mid T_i > s) \)

- General idea: specify a model for \( \mu_i(\beta) = E(Y_i \mid T_i > s, Z_i) \)

- For some link function \( g \), postulate
  \[
  g(\mu_i(\beta)) = \beta_0(s) + \beta^\top(s)Z_i(s)
  \]

- Score equation from binomial likelihood
  \[
  \sum_{i=1}^{n_s} \frac{\partial \mu_i(\beta)}{\partial \beta} \cdot \frac{1}{\mu_i(1 - \mu_i)} \cdot [y_i - \mu_i] = 0
  \]
Obtaining dynamic predictions

- For a new patient with covariate vector $Z^*$
- Estimate of $P(T \leq s + w, D = j | T > s, Z^*(s))$ is

$$
\hat{F}_j(s + w|Z^*(s), s) = g^{-1}(\hat{\beta}_0(s) + \hat{\beta}^\top(s)Z^*(s)),
$$

- Its variance is estimated consistently by

$$
\left( \frac{dg^{-1}(x)}{dx} \right)^2 \bigg|_{x=\hat{\beta}^\top Z^*} \cdot (Z^*)^\top \cdot \text{var}(\hat{\beta}) \cdot Z^*
$$

(delta-method)
Dynamic pseudo-observations

- Recall

\[ Y_i = 1 \{ T_i \leq s + w, D_i = j \} \]

- Unfortunately, not available for censored patients!

- Define the dynamic pseudo-observation

\[ \hat{\theta}_{is} = n_s \hat{F}_j(s + w|s) - (n_s - 1) \hat{F}_j^{(-i)}(s + w|s), \]

where

\[ \hat{F}_j(s + w|s) = \sum_{s < t_i \leq s + w} \frac{d_{ij}}{n_i} \cdot \hat{S}(t_i - |s), \]

with

- \( d_{ij} = \) no of cause \( j \) events at \( t_i \); \( n_i = \) no at risk at \( t_i \)
Properties of dynamic pseudo-observations

No censoring

\[ \hat{\theta}_{is} = 1 \{ T_i \leq s + w, D_i = j \} \]
Properties of dynamic pseudo-observations

No censoring

\[ \hat{\theta}_{is} = 1\{T_i \leq s + w, D_i = j\} \]

With censoring

- **(P1)** \( \hat{\theta}_{is} \) is asymptotically independent of \( \hat{\theta}_{ls} \) for individuals 
  \( i \neq l \) as \( n_s \to \infty \)

- **(P2)** \( \hat{\theta}_{is} \) is asymptotically independent of \( \hat{\theta}_{ls'} \) for individuals 
  \( i \neq l \) and landmark time points \( s \neq s' \) as \( n_s, n_{s'} \to \infty \)

- **(P3)** \( E[\hat{\theta}_{is} | T_i > s, Z_i] \) equals asymptotically its theoretical 
  counterpart \( E[1\{T_i \leq s + w, D_i = j\} | T_i > s, Z_i] \) as \( n_s \to \infty \)

(Follow more or less directly from results in Graw et al. (2009).)
Censored data

Fixed $s$

- $Y_i = 1 \{ T_i \leq s + w, D_i = j \}, \ i \in D_s$ not observed
- Retain score equations and replace $y_i$ by pseudo-observations $\hat{\theta}_{is}$
- The quasi-score equation is

$$\sum_{i=1}^{n_s} \frac{\partial \mu_i(\beta)}{\partial \beta} \cdot \frac{1}{\mu_i(1 - \mu_i)} \cdot [\hat{\theta}_{is} - \mu_i] = 0$$

- Estimating equations are asymptotically unbiased ($\Rightarrow$ asymptotic normality of $\hat{\beta}$)
- Note: we calculate only one pseudo-observation per individual
Obtaining dynamic predictions

- For a new patient with covariate vector $Z^*$
- Estimate of $P(T \leq s + w, D = j|T > s, Z^*(s))$ is

$$
\hat{F}_j(s + w|Z^*(s), s) = g^{-1}(\hat{\beta}_0(s) + \hat{\beta}^\top(s)Z^*(s)),
$$

- Its variance is estimated consistently by

$$
\left(\frac{dg^{-1}(x)}{dx}\right)^2 \bigg|_{x=\hat{\beta}^\top Z^*} \cdot (Z^*)^\top \cdot \text{var}(\hat{\beta}) \cdot Z^*
$$

(delta-method)
Super models: setup

- Define a set of landmark time points $0 \leq s_1 < \ldots < s_K \leq \tau$
- Construct the corresponding landmark data sets $\mathcal{D}_k := \mathcal{D}_{s_k}$
- Define the dynamic pseudo-observation $\hat{\theta}_{ik}$ of individual $i$ at time $s_k + w$
- Note again: only one per subject per $s_k$ (namely at $s_k + w$)
- Define $\hat{\theta}_i = (\hat{\theta}_{ik}, \ k \in \mathcal{L}_i), \ \mathcal{L}_i \subset \{1, \ldots, K\}$
- Longitudinal vector $\hat{\theta}_i$ truncated by death (Kurland & Heagerty 2005)
Super models

- Define $\mu_{ik} = \mu_i(s_k) = P(T_i \leq s_k + w, D_i = j \mid T_i > s_k)$
- GLM

\[ g(\mu_{ik} \mid Z_i(s_k)) = \beta_0(s_k) + \beta^\top(s_k)Z_i(s_k), \]

- Choose smooth model for $l^{th}$ component of $\beta(s)$

\[ \beta_l(s) = \beta_l^\top h_l(s) \quad \Rightarrow \quad \beta(s) = H(s)\beta \]

- The quasi-score equation for regression parameter $\beta$ is

\[ U(\beta) = \sum_{i=1}^{n} U_i(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i(\beta)}{\partial \beta} \cdot V_i^{-1}(\hat{\theta}_i - \mu_i) = 0 \]

- Asymptotic unbiasedness of these estimation equations follows (only for independence working correlation!)
Dynamic pseudo-observations

Dynamic prediction in super models

- For a new patient with covariate vector $Z^*$
- The estimate of $P(T \leq s + w, D = j \mid T > s, Z^*(s))$ is

$$
\hat{F}_j(s + w \mid Z^*, s) = g^{-1}(\hat{\beta}_0(s) + \hat{\beta}^\top(s)Z^*(s)),
$$

with $\hat{\beta}(s) = H(s)\hat{\beta}$

- Its variance is estimated consistently by

$$
\left( \frac{dg^{-1}(x)}{dx} \right)^2_{x=(\hat{\beta})^\top Z^*} \cdot (Z^*)^\top \cdot H(s) \cdot \text{var}(\hat{\beta}) \cdot H(s)^\top \cdot Z^*
$$
Scatter-plot

Relapse

0.0 0.5 1.0

0.984 0.961 0.947 0.913

0.0 0.5 1.0

0.983 0.973 0.947

0.0 0.5 1.0

0.992 0.970

0.0 0.5 1.0

0.980

0.0 0.5 1.0 1.5

0.0 0.6 1.2

0.0 0.5 1.0

0.0 0.5 1.0

0.0 0.5 1.0

0.0 0.5 1.0 1.5

0.0 0.6 1.2
## Model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Relapse</th>
<th>NRM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$</td>
<td>SE($\hat{\beta}$)</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.160</td>
<td>0.027</td>
</tr>
<tr>
<td>$s$</td>
<td>0.839</td>
<td>0.126</td>
</tr>
<tr>
<td>$s^2$</td>
<td>-1.072</td>
<td>0.129</td>
</tr>
<tr>
<td><strong>Year of transplantation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.530</td>
<td>0.126</td>
</tr>
<tr>
<td>$s$</td>
<td>-1.165</td>
<td>0.657</td>
</tr>
<tr>
<td>$s^2$</td>
<td>0.553</td>
<td>0.678</td>
</tr>
<tr>
<td><strong>Risk score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium risk</td>
<td>0.166</td>
<td>0.022</td>
</tr>
<tr>
<td>High risk</td>
<td>0.725</td>
<td>0.039</td>
</tr>
<tr>
<td>Low grade aGvHD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.490</td>
<td>0.030</td>
</tr>
<tr>
<td>$s$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High grade aGvHD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.305</td>
<td>0.054</td>
</tr>
<tr>
<td>$s$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s^2$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Regression coefficients of EBMT risk score

![Graph showing regression coefficients for EBMT risk score over time](image)
Regression coefficients of aGvHD

High grade aGvHD

Relapse

NRM

Separate models
Supermodel

Regression coefficients

Time s of landmark (years)
5-yr prediction probabilities

- Low risk
- Medium risk
- High risk

Predicted probabilities at $s + 5$ years

- No aGvHD
- Low grade
- High grade

Time of landmark (years)

- Low grade
- High grade

Predicted probabilities at $s + 5$ years

- NRM

Time of landmark (years)
Discussion

- For each landmark time point $s_k$ we only use (need) dynamic pseudo-observations at one fixed horizon ($s_k + w$)
- No proportional hazards assumptions needed
- Gain in robustness
- But possible loss of efficiency
- Method is direct and straightforward to implement in GEE software (especially using the pseudo package)
- Correlation structure of dynamic pseudo-observations is ignored in the estimating equations of the super models
Discussion (general)

Advantages of standard approach

- It is standard
- There is software
- You gain biological understanding (hopefully) by modeling the effects of covariates on transitions

Advantages of landmarking

- It is more direct; no need for complicated formulas for prediction
- Predictions obtained from multi-state model may be off the mark if assumptions are violated or if model fit is not good
- Sparse model (considerably fewer parameters than multi-state model)
References


References

Dynamic prediction by landmarking in event history analysis.

Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphois leukemia data.

*Dynamic Predicting in Clinical Survival Analysis.*
Chapman & Hall.

The robust inference for the Cox proportional hazards model.
*JASA* 84: 1074–1078.

Estimating average regression effects under non-proportional hazards.
*Biostatistics* 1: 423–439.

Partly conditional survival models for longitudinal data.
Dynamic Prediction in Clinical Survival Analysis

Monographs on Statistics and Applied Probability 123

Dynamic Prediction in Clinical Survival Analysis

Hans C. van Houwelingen
Hein Putter

CRC Press
Taylor & Francis Group
A Chapman & Hall Book