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

Origin of landmarking

- Origin: debate on the effect of response to chemotherapy on survival (Anderson JR, Cain KC, Gelber RD, 1983, *J Clin Oncol* **1**, 710-719)
- 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

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

  - Basic Activities of Daily Living (walking, bathing, dressing, toileting, feeding), disabled if "with difficulty" for ≥ 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

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

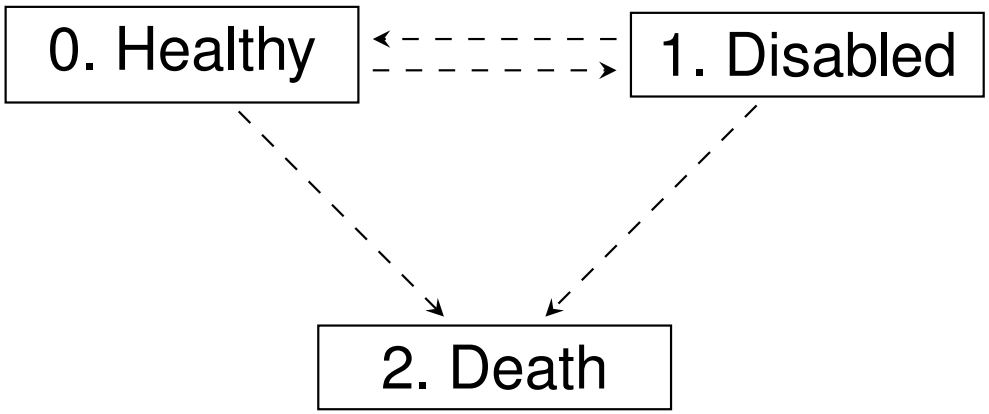
  - Basic Activities of Daily Living (walking, bathing, dressing, toileting, feeding), disabled if "with difficulty" for ≥ 1 item
- Objective: dynamic prediction of survival of at least 10 years beyond age s, with given ADL-disability status at age s and given covariates

Covariates

Covariate	N	(%)
Gender		
Male	1564	(39%)
Female	2468	(61%)
Education		
Less than high school	1736	(43%)
High school	1212	(30%)
Some college	1084	(27%)
BMI		
≤ 25	2244	(56%)
25 – 30	1388	(34%)
> 30	390	(10%)
Smoking		
Never	1997	(50%)
Past	1683	(42%)
Current	324	(8%)

Multi-state approach

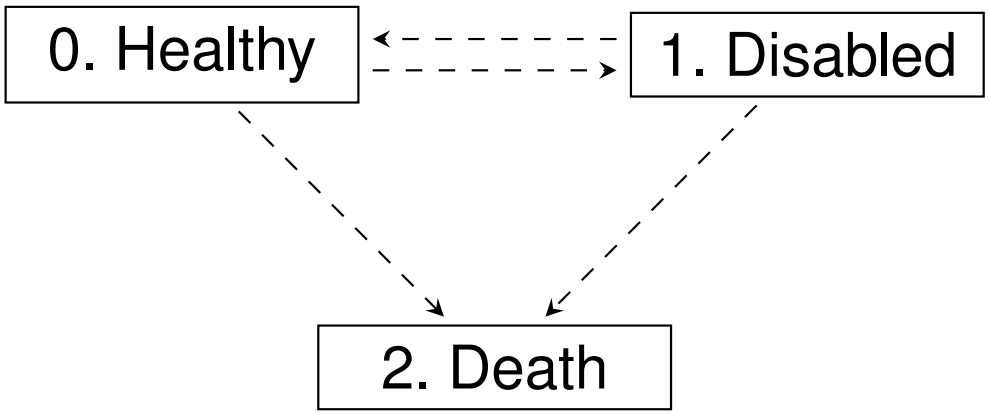
- Multi-state model; process $X(t)$ in time, taking values
  - 0: alive without ADL disability
  - 1: alive with ADL disability
  - 2: dead



- Reversible illness-death model

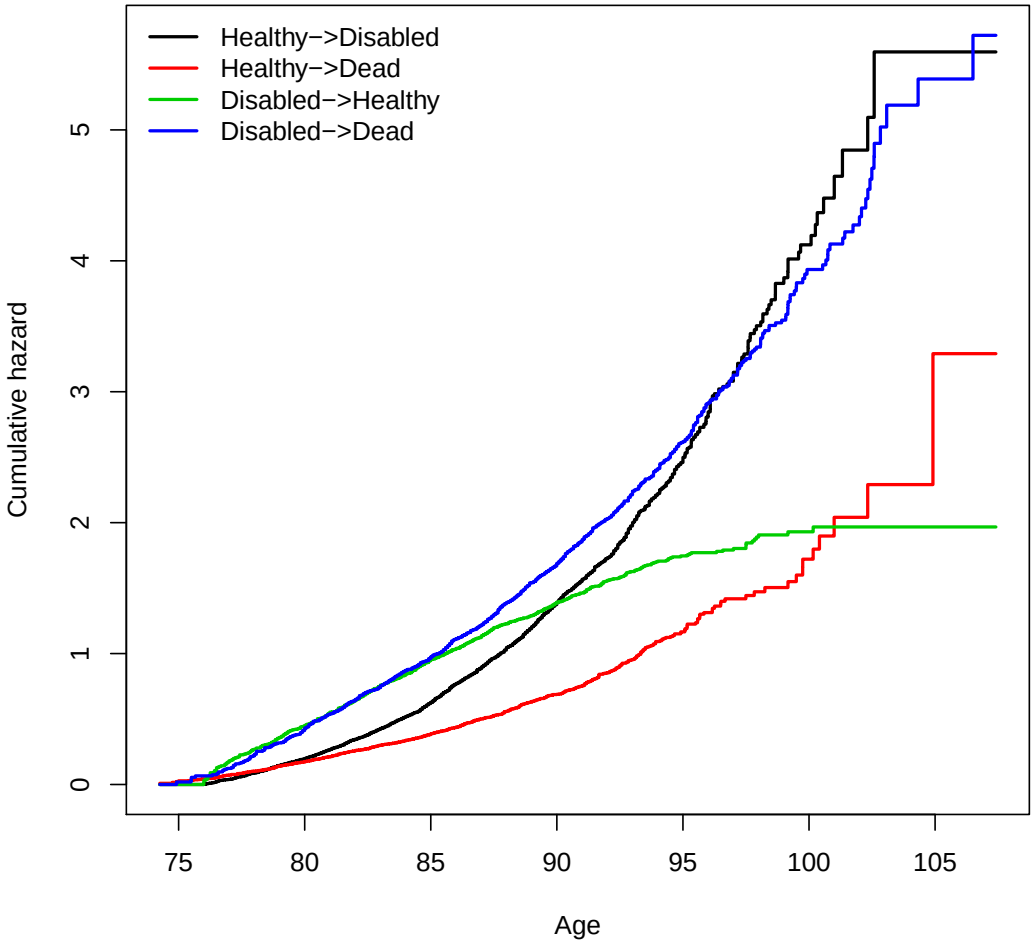
Multi-state approach

- 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



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 **`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_{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_{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_{hor} during estimation of the Cox model
  - Prediction is only used at t_{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_{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

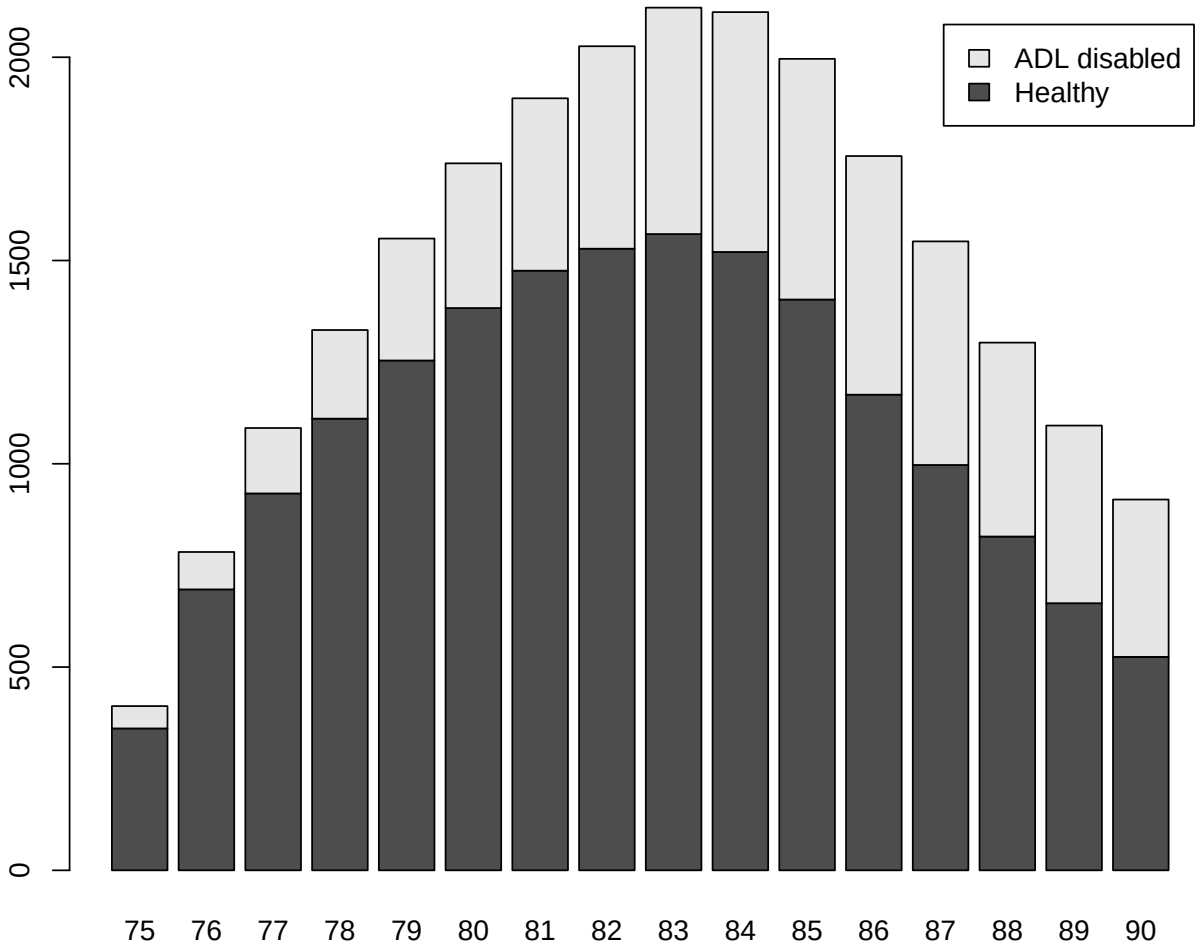
$$\theta(s) = \theta_1 s + \theta_2 s^2$$

- 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

Set-up

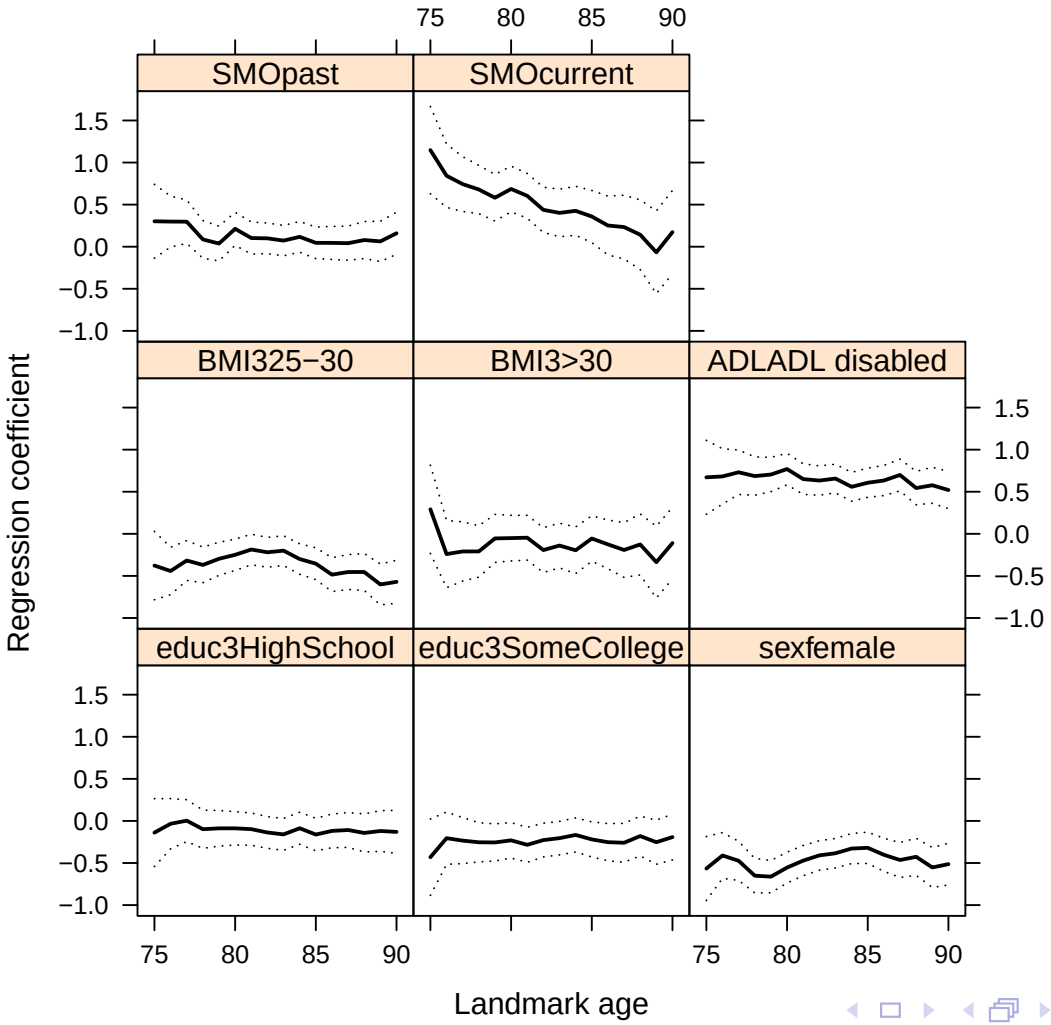
- Endpoint is survival in a window of fixed width $w = 10$ years from the moment of prediction
- Landmark time points used: 16 points, equally spaced, from age 75 to age 90
- For each landmark (prediction) time point, construct landmark data set, containing all relevant information needed for the prediction
- 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 $t_{\text{hor}} + 10$

The landmark data sets



Regression coefficients

Regression coefficients with 95% confidence intervals

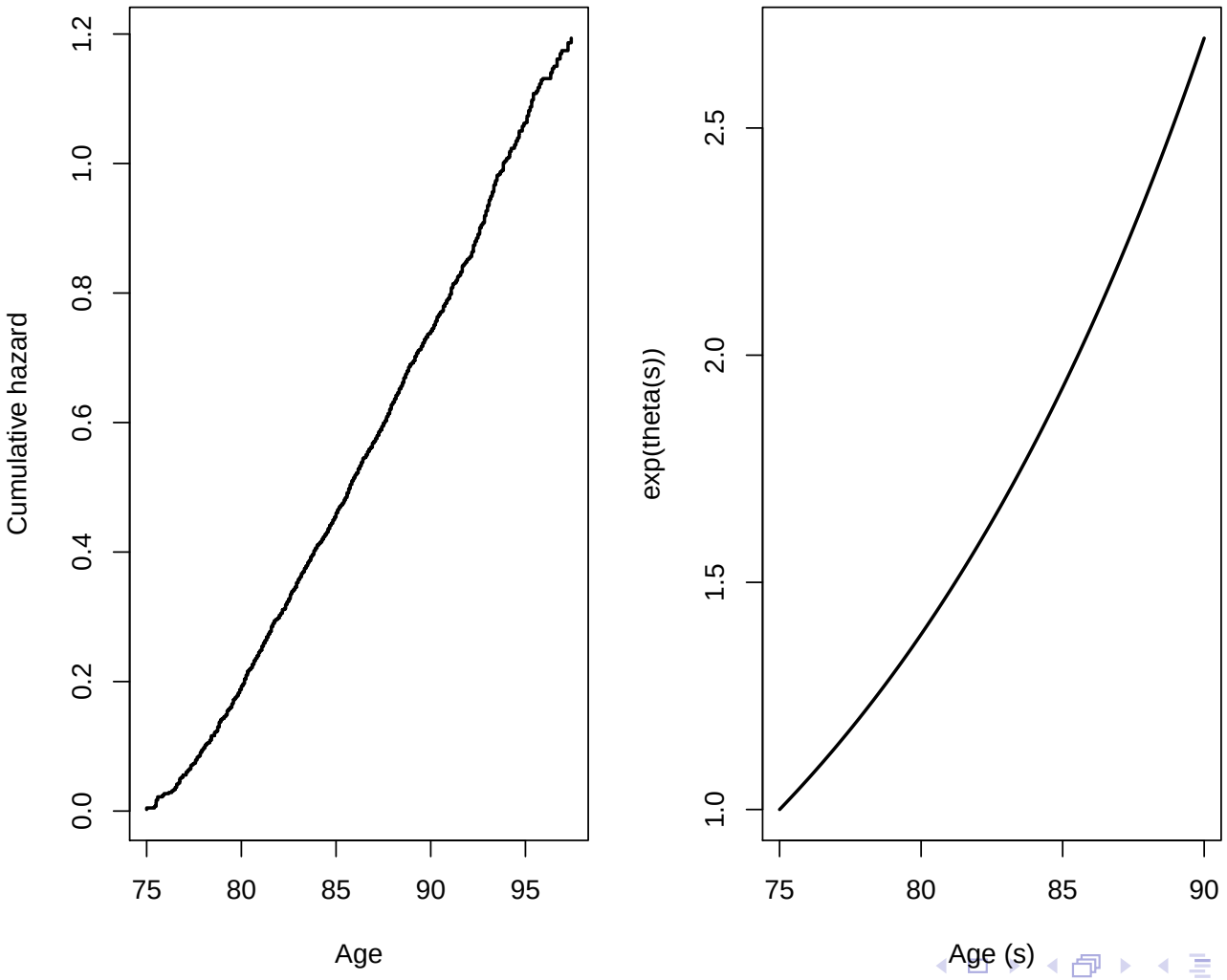


Landmark super model

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

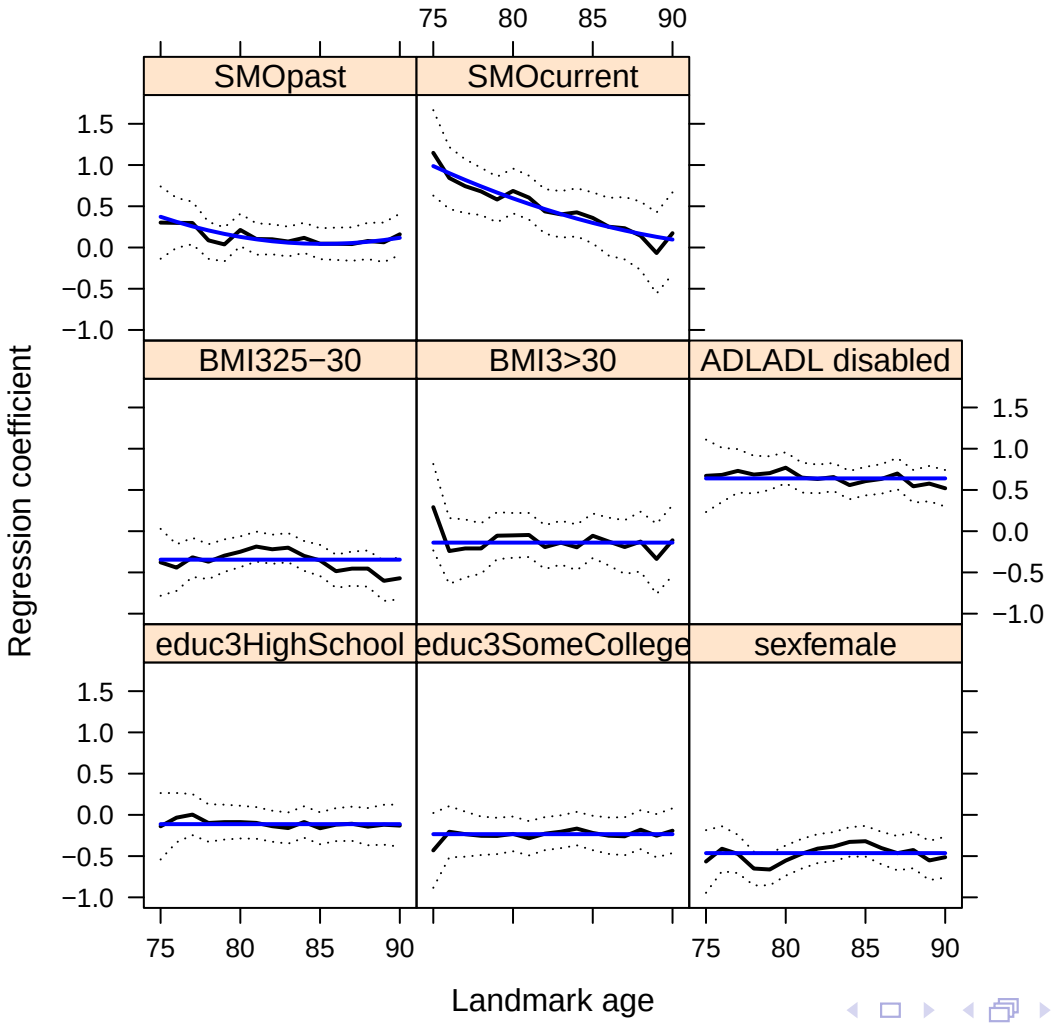
$\bar{s} = (s - 75)/15$

With corresponding baselines

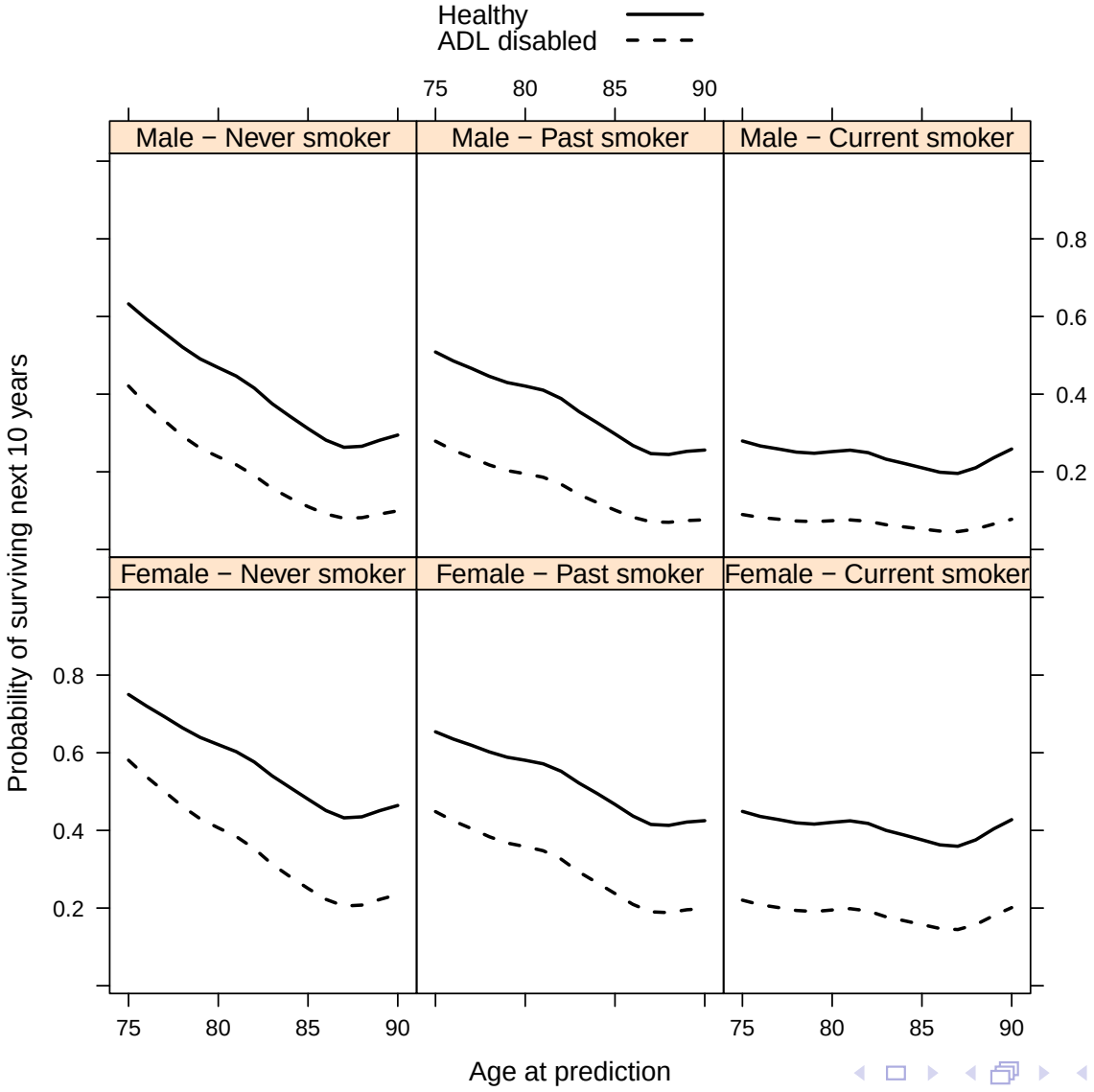


Regression coefficients

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



Dynamic predictions from the landmark super model



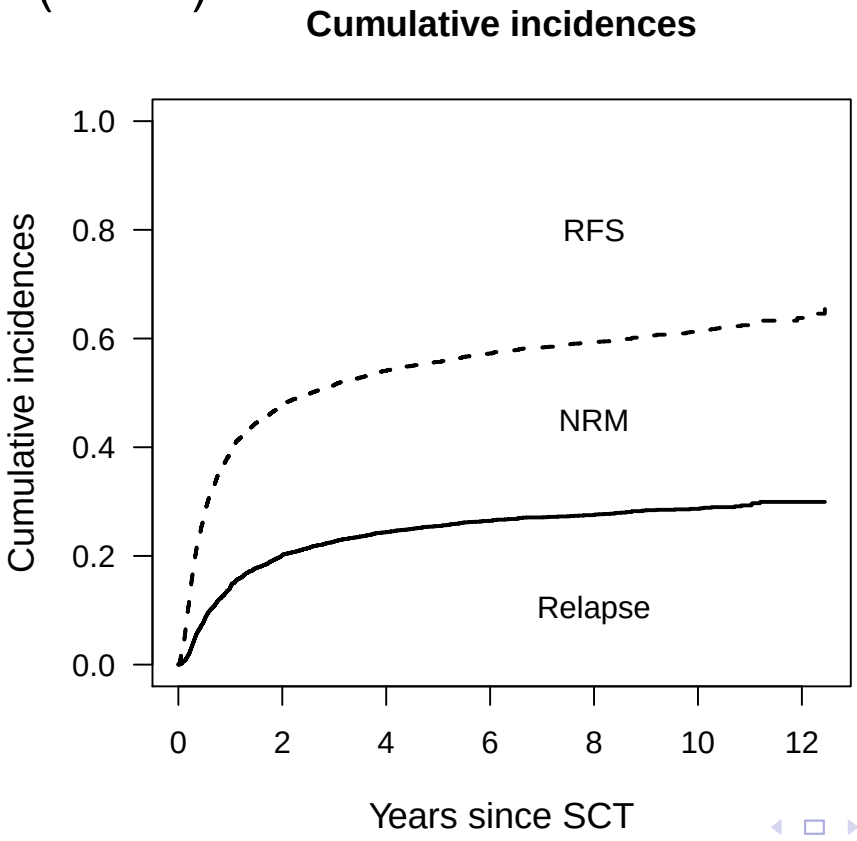
Software

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`)
  - 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`, R code (using the **dynpred** package) of all the analyses in the book is available for download

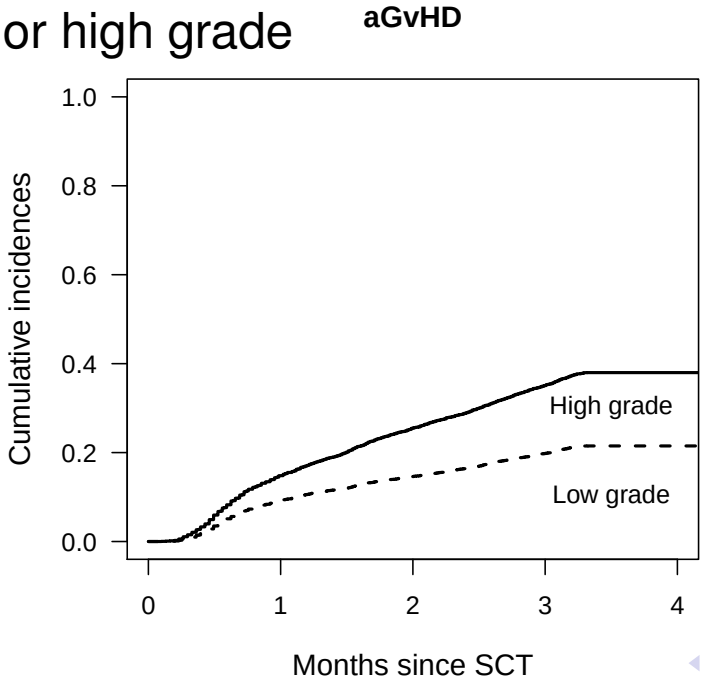
Data (EBMT)

- 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



Time-dependent covariates

- Define

$$Z_{\text{low}}(t) = \mathbf{1}\{\text{occurrence of low grade aGvHD before time } t\}$$
$$Z_{\text{high}}(t) = \mathbf{1}\{\text{occurrence of high grade aGvHD before time } t\}$$

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 $\mathcal{D}_s$, the *survivors* at s, define

$$Y_i = \mathbf{1}\{T_i \leq s + w, D_i = j\}$$

- Expectation of Y_i in $\mathcal{D}_s$ is $P(T_i \leq s + w, D_i = j \mid T_i > s)$
- General idea: specify a model for $\mu_i(\beta) = E(Y_i | T_i > s, Z_i)$
- For some link function g, postulate

$$g(\mu_i(\beta)) = \beta_0(s) + \beta^\top(s)\mathbf{Z}_i(s)$$

- 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 $\mathbf{Z}^*$
- Estimate of $P(T \leq s + w, D = j | T > s, \mathbf{Z}^*(s))$ is

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

- Its variance is estimated consistently by

$$\left(\frac{\mathrm{d}g^{-1}(x)}{\mathrm{d}x}\right)^2_{|x=\widehat{\beta}^\top \mathbf{Z}^*} \cdot (\mathbf{Z}^*)^\top \cdot \widehat{\mathrm{var}}(\widehat{\beta}) \cdot \mathbf{Z}^*$$

(delta-method)

Dynamic pseudo-observations

- Recall
$$Y_i = \mathbf{1}\{T_i \leq s + w, D_i = j\}$$
- 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

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

Properties of dynamic pseudo-observations

No censoring

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

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)** $E\left[\hat{\theta}_{is} | T_i > s, \mathbf{Z}_i\right]$ equals asymptotically its theoretical counterpart $\mathbf{E}\left[\mathbf{1}\{T_i \leq s + w, D_i = j\} | T_i > s, , \mathbf{Z}_i\right]$ as $n_s \to \infty$

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

Censored data

Fixed s

- $Y_i = \mathbf{1}\{T_i \leq s + w, D_i = j\},\ i \in \mathcal{D}_s$ 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 $\mathbf{Z}^*$
- Estimate of $P(T \leq s + w, D = j | T > s, \mathbf{Z}^*(s))$ is

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

- Its variance is estimated consistently by

$$\left(\frac{\mathrm{d}g^{-1}(x)}{\mathrm{d}x}\right)^2_{|x=\widehat{\beta}^\top \mathbf{Z}^*} \cdot (\mathbf{Z}^*)^\top \cdot \widehat{\mathrm{var}}(\widehat{\beta}) \cdot \mathbf{Z}^*$$

(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 $\widehat{\theta}_{ik}$ 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
$$g(\mu_{ik}|Z_i(s_k)) = \beta_0(s_k) + \beta^\top(s_k)\mathbf{Z}_i(s_k),$$
- Choose smooth model for l^{th} component of $\beta(s)$
$$\beta_l(s) = \beta_l^\top h_l(s) \qquad \Rightarrow \qquad \beta(s) = H(s)\beta$$
- The quasi-score equation for regression parameter β is
$$U(\beta) = \sum_{i=1}^{n} U_i(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i(\beta)}{\partial \beta} \cdot V_i^{-1}(\widehat{\theta}_i - \mu_i) = 0$$
- Asymptotic unbiasedness of these estimation equations follows (only for independence working correlation!)

Dynamic prediction in super models

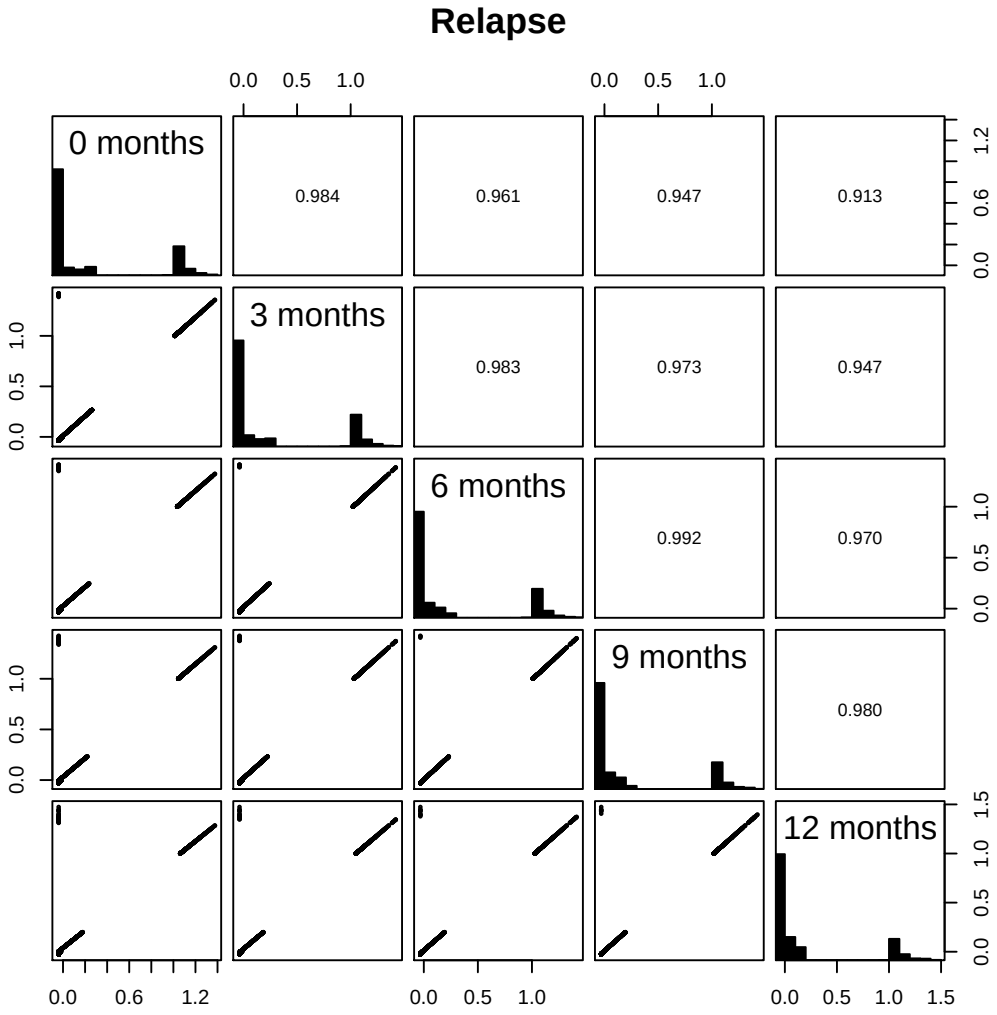
- For a new patient with covariate vector $\mathbf{Z}^*$
- The estimate of $P(T \leq s + w, D = j | T > s, \mathbf{Z}^*(s))$ is

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

  with $\widehat{\beta}(s) = H(s)\hat{\beta}$
- Its variance is estimated consistently by

$$\left(\frac{\mathrm{d}g^{-1}(x)}{\mathrm{d}x}\right)^2_{|x=(\widehat{\beta})^\top \mathbf{Z}^*} \cdot (\mathbf{Z}^*)^\top \cdot H(s) \cdot \widehat{\mathrm{var}}(\widehat{\beta}) \cdot H(s)^\top \cdot \mathbf{Z}^*$$

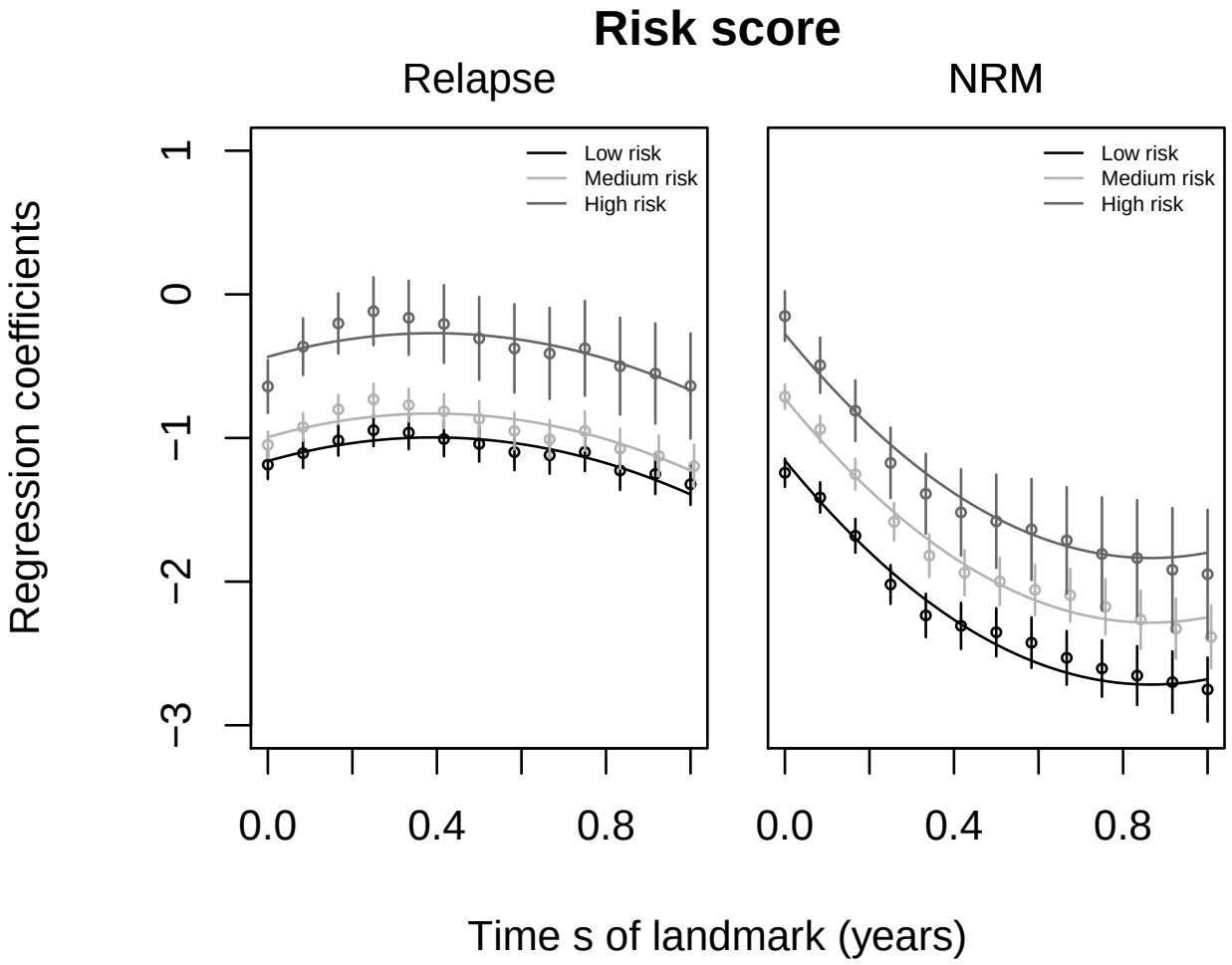
Scatter-plot



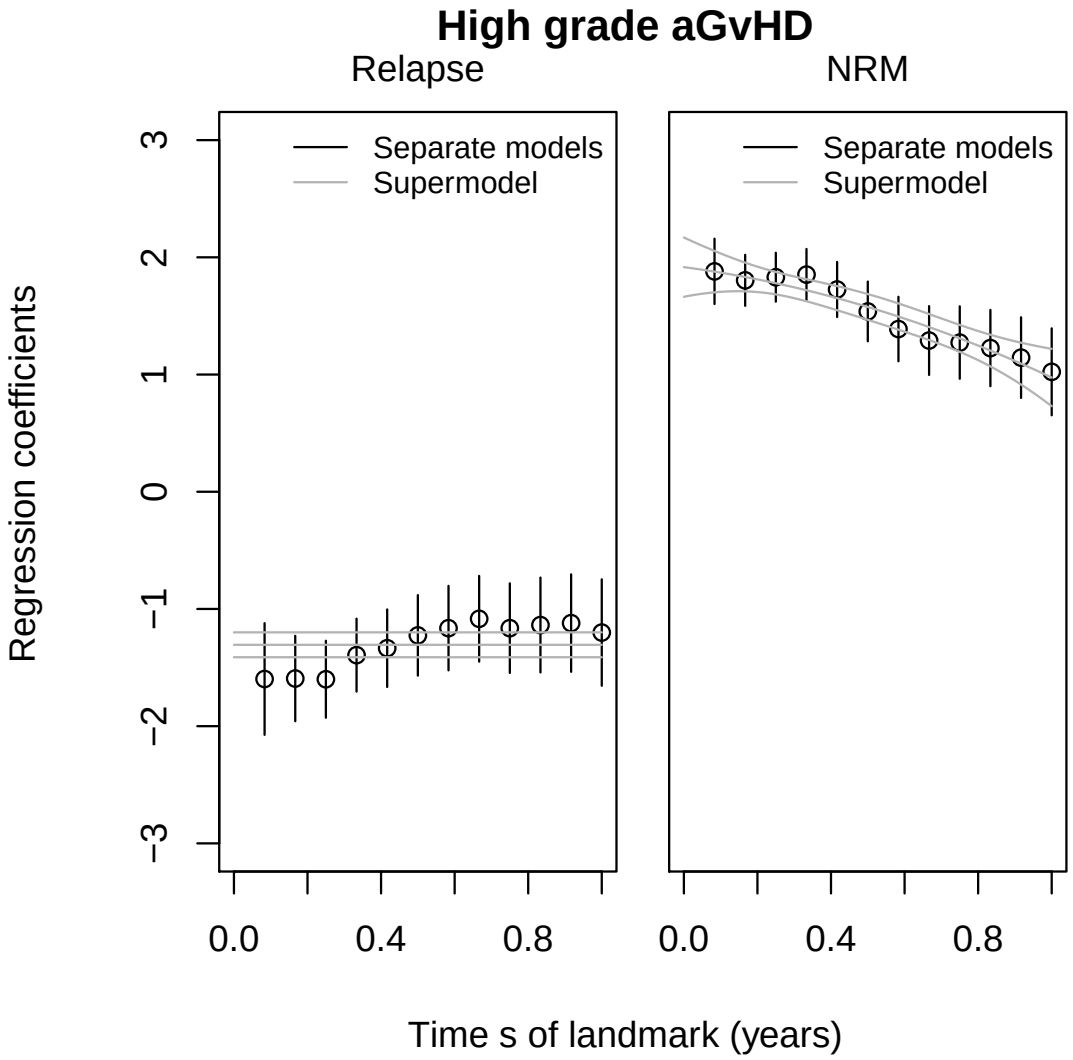
Model

Covariate	Relapse		NRM	
	$\widehat{\beta}$	SE($\widehat{\beta}$)	$\widehat{\beta}$	SE($\widehat{\beta}$)
Intercept				
Constant	-1.160	0.027	-1.156	0.029
s	0.839	0.126	-3.603	0.165
s^2	-1.072	0.129	2.080	0.183
Year of transplantation				
Constant	0.530	0.126	-0.591	0.125
s	-1.165	0.657	1.604	0.718
s^2	0.553	0.678	-1.079	0.766
Risk score				
Low risk				
Medium risk	0.166	0.022	0.431	0.025
High risk	0.725	0.039	0.880	0.042
Low grade aGvHD				
Constant	-0.490	0.030	0.168	0.102
s			2.032	0.461
s^2			-1.738	0.436
High grade aGvHD				
Constant	-1.305	0.054	1.916	0.129
s			-0.416	0.579
s^2			-0.524	0.545

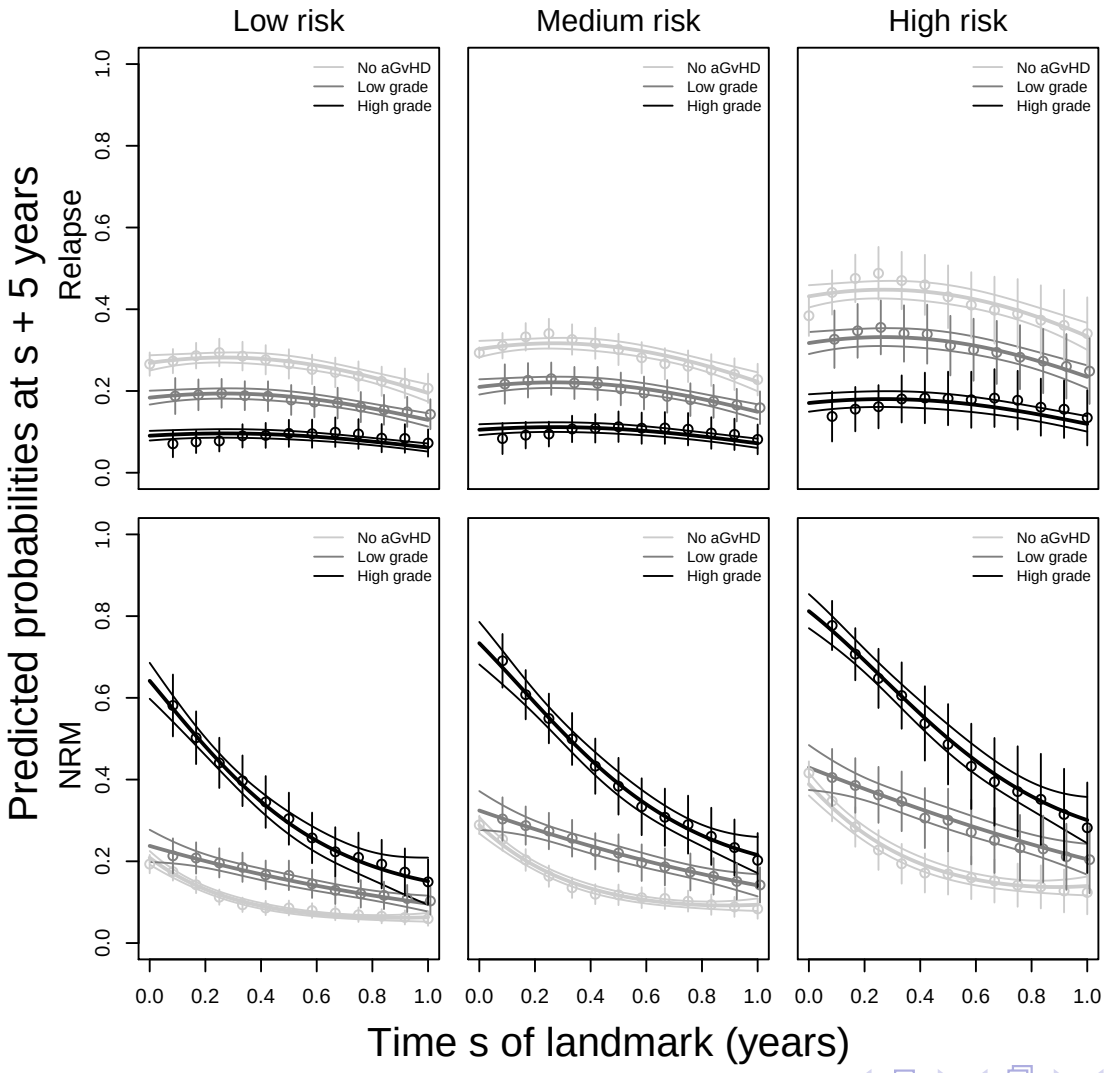
Regression coefficients of EBMT risk score



Regression coefficients of aGvHD



5-yr prediction probabilities



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

Cortese, G. and Andersen, P. K. (2009).
Competing risks and time-dependent covariates.
Biom J **32**: 138–158.

Graw, F. and Gerds, T. A. and Schumacher, M. (2009).
On pseudo-values for regression analysis in multi-state models.
Lifetime Data Anal **15**: 241–255.

van Houwelingen, H. C. (2007).
Dynamic prediction by landmarking in event history analysis.
Scand J Stat **34**: 70–85.

Kurland, B. F. and Heagerty, P. J. (2005).
Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths.
Biostat **6**: 241–258.

Nicolaie, M. A. and van Houwelingen, H. C. and de Witte, T. M. and Putter, H. (2012).
Dynamic prediction in competing risks by landmarking.
Stat Med, in press.

References

H. C. van Houwelingen (2007).
Dynamic prediction by landmarking in event history analysis.
Scand. J. Stat. **34**: 70–85.

H. C. van Houwelingen and H. Putter (2008).
Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphois leukemia data.
Lifetime Data Anal. **14**: 447–463.

H. C. van Houwelingen and H. Putter (2012).
Dynamic Predicting in Clinical Survival Analysis.
Chapman & Hall.

D. Y. Lin and L. J. Wei (1989).
The robust inference for the Cox proportional hazards model.
JASA **84**: 1074–1078.

R. Xu and J. O'Quigley (2000).
Estimating average regression effects under non-proportional hazards.
Biostatistics **1**: 423–439.

Y. Y. Zheng and P. J. Heagerty (2005).
Partly conditional survival models for longitudinal data.
Biometrics **61**: 379–391.

Dynamic Prediction in Clinical Survival Analysis