Landmarking, immortal time bias and dynamic prediction

Hein Putter

Department of Medical Statistics and Bioinformatics
Leiden University Medical Center

EFSPI Meeting on Survival Analysis, Brussels
November 7, 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

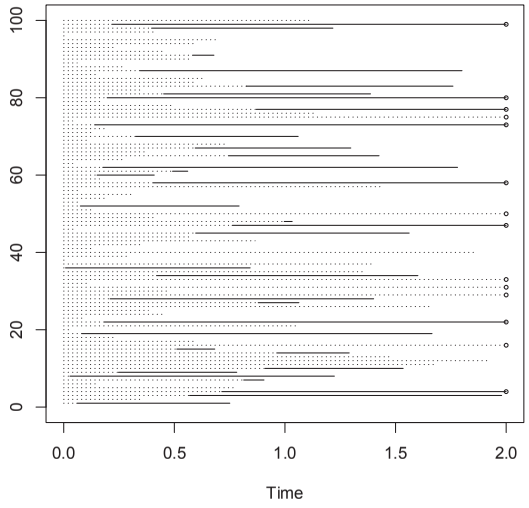
- Crucial issue: "responder" versus "non-responder" is something that is not known at baseline
- When studying survival, it is not allowed to make groups based on something that will happen in the future
- Two alternatives proposed
  - Time-dependent covariate
  - Landmark
    - Consider response at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

Example

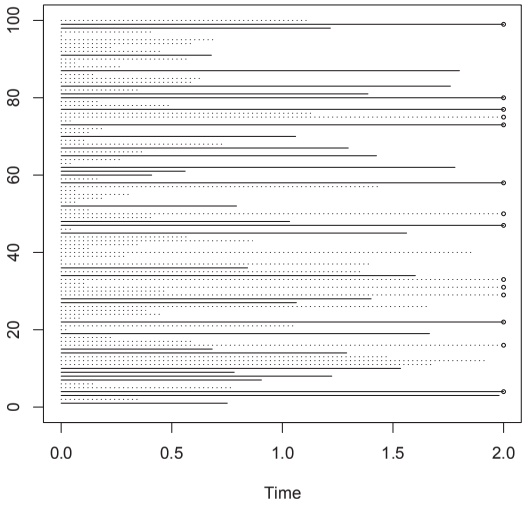
Simulated data loosely based on response to chemotherapy

- $n = 100$
- Time to response T_{resp} uniform on $(0, 1)$ with probability 0.5, no response ($T_{\text{resp}} = \infty$) with probability 0.5
- Time to death T_{death} exponential with mean 1, independent of T_{resp}
  - Could happen before response, in which case response is not observed
- Censoring at 2 (years)

Simulated data



Groups made based on response status



Analyses

Wrong

- Use response status at end of follow-up as if that was known at baseline
- Cox regression gives estimated coefficient of -0.890 with SE of 0.235 (p=0.00015)
- Response to chemotherapy significantly improves survival

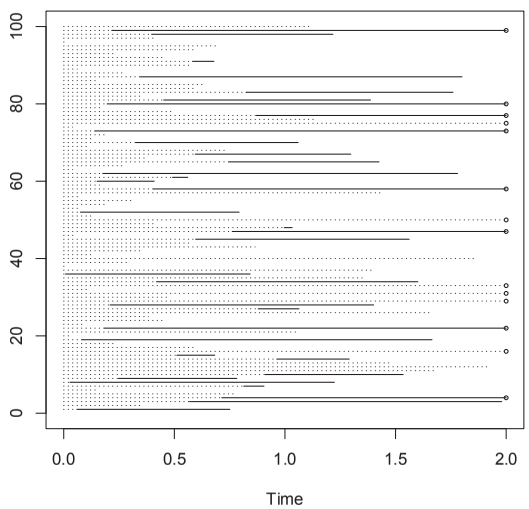
Correct I

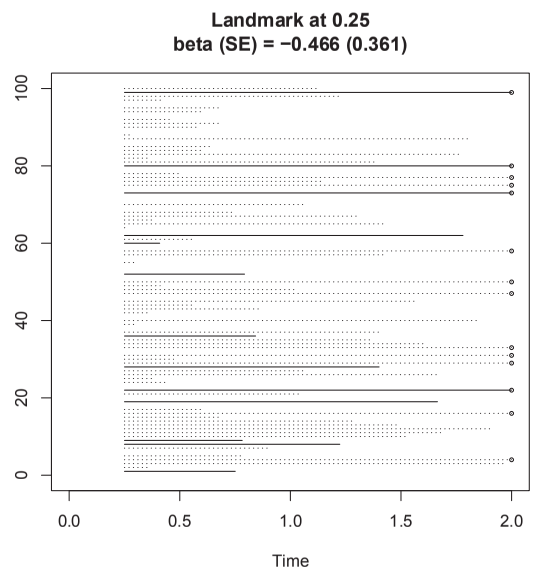
- Use response status as time-dependent covariate
- Cox regression gives estimated coefficient of -0.176 with SE of 0.258 (p=0.50)
- Response to chemotherapy does not affect survival

Analyses

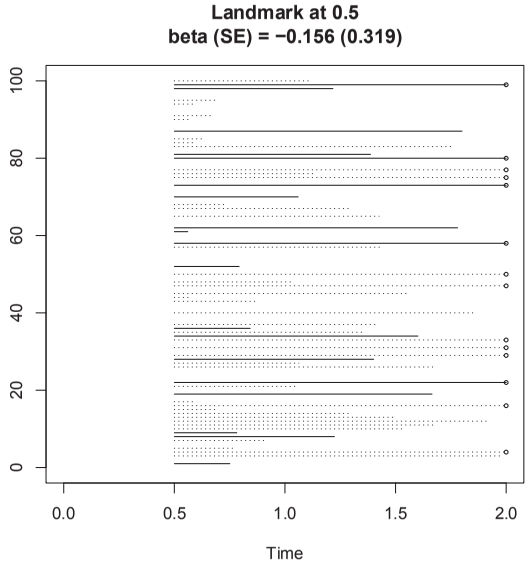
Correct II

- Fix landmark time point t_{LM}
- Create a "landmark data set" by
  - Removing everyone with event or censored before t_{LM}
  - Creating response groups based on response status at t_{LM}
- Perform Cox regression with these response groups as time-fixed covariate
- Illustrated for series of landmark time points $t_{\mathrm{LM}} = 0.25, 0.5, \ldots, 1.5, 1.75$

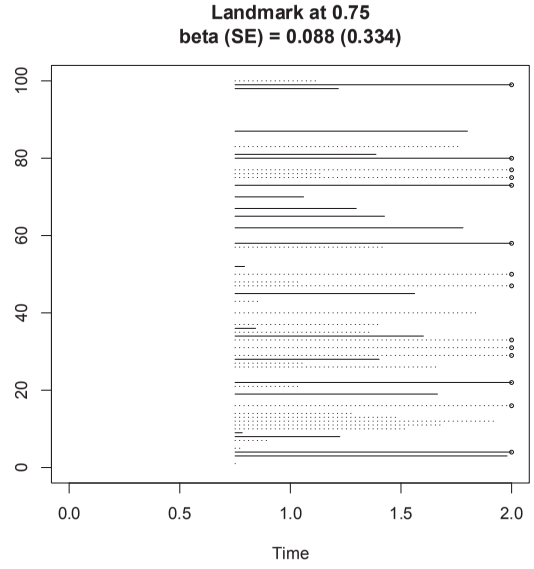




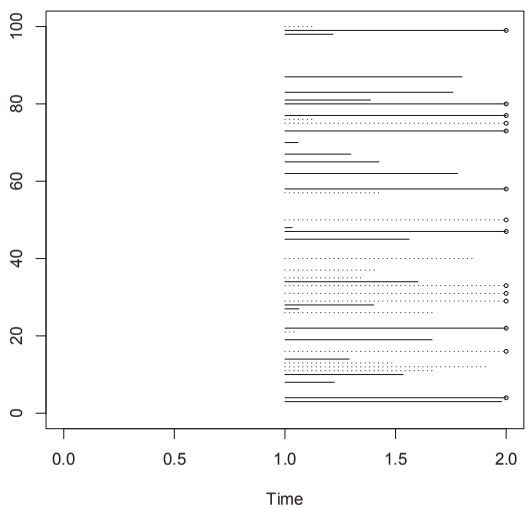
... in action ...



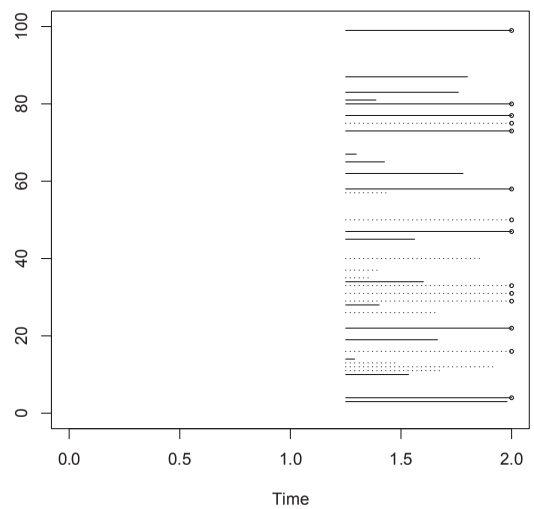
... in action ...

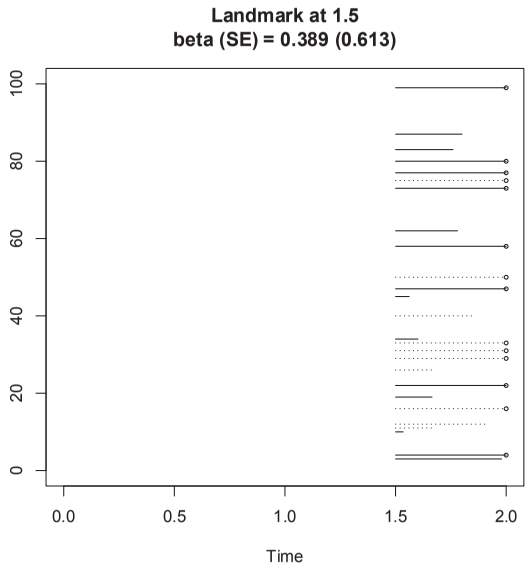


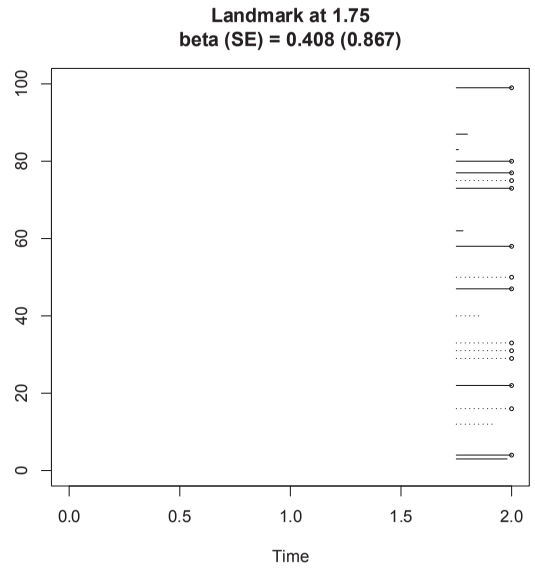
... in action ...



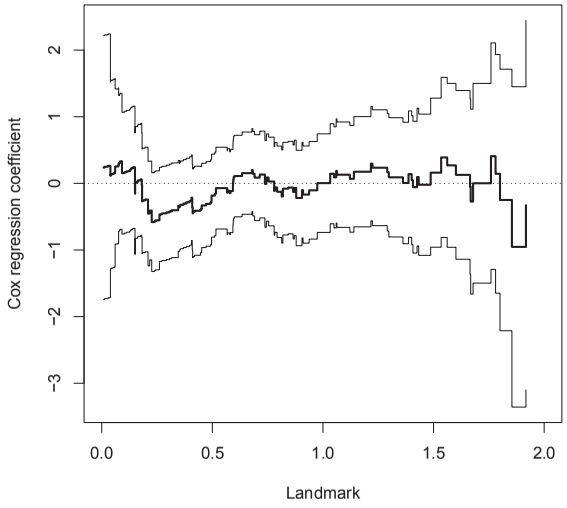
... in action ...







For all possible landmark points



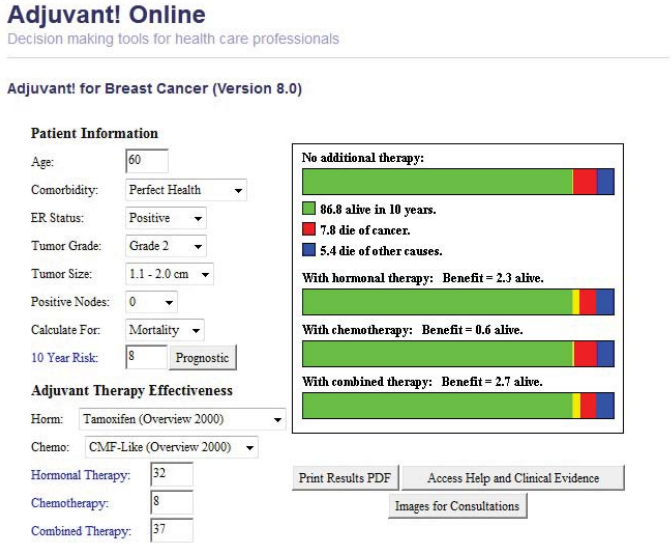
Prediction models

- Prediction models used in wide variety of diseases
- They are important, used to guide therapy choices, to inform patients
- Famous examples: Apgar score, Framingham risk score, the Gail model, Adjuvant! Online

Komt een vrouw bij de dokter ...

- Woman, 60 years, diagnosed with breast cancer
- ER+, Grade II, no additional health problems
- Tumor to be removed with mastectomy plus radiotherapy
- Tumor size 1.5 cm, no lymph nodes involved
- What is the probability that she will be alive 5 years from now?
  - With hormonal therapy
  - With chemotherapy

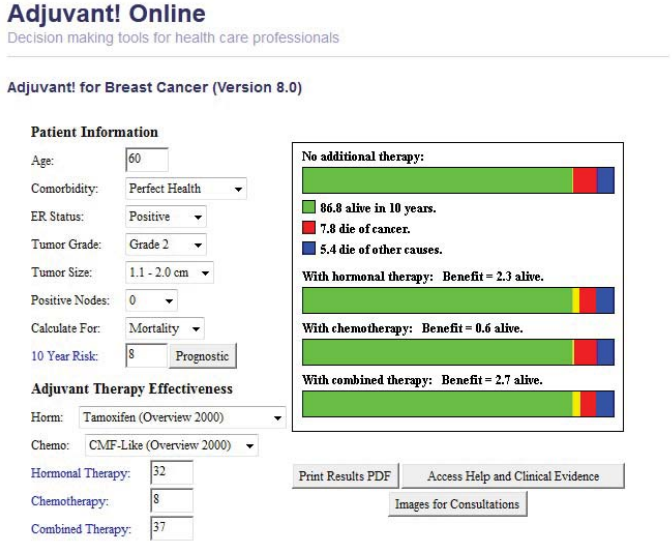
Adjuvant! Online (10 years)



Komt een vrouw bij de dokter ...

- Woman, 60 years, diagnosed with breast cancer
- ER+, Grade II, no additional health problems
- Tumor to be removed with mastectomy plus radiotherapy
- Tumor size 1.5 cm, no lymph nodes involved
- Surgery was three years ago, after consulting Adjuvant! Online, it was decided to add hormonal therapy and chemotherapy
- Today woman comes for regular visit, she is doing fine
- Three years without evidence of disease (no local recurrence or distant metastasis)
- Does she need to worry that disease comes back?
- What is the probability that she will be alive and disease-free in 5 years from now?

Adjuvant! Online



Using Adjuvant! Online

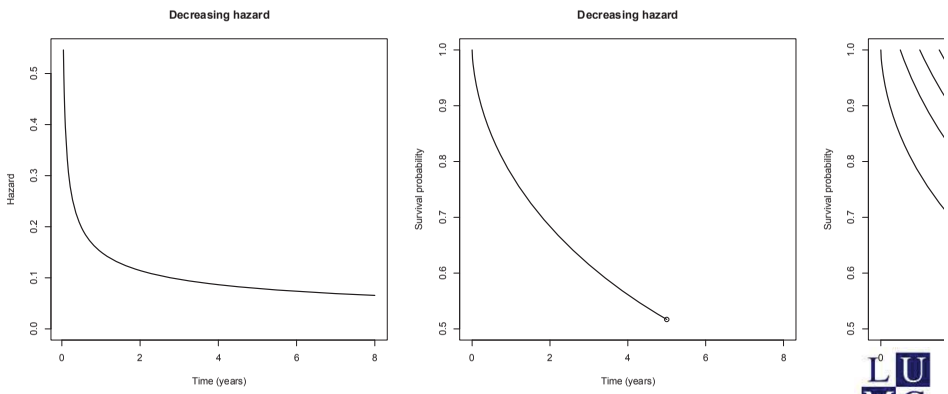
- First temptation would be just to use Adjuvant! Online

Why this isn't a good idea

- Not using information that has become available
- Some covariates may have time-varying effects, typically strong in the beginning, less important later in follow-up
- The very fact of being alive changes prognosis

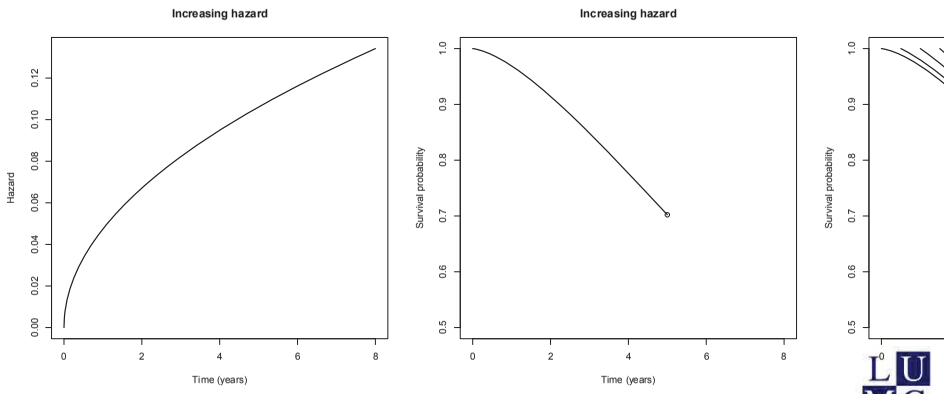
The effect of "being alive"

Prognosis may improve



The effect of "being alive"

Prognosis may become worse



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)

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 three years after surgery, to live another 5 years
- The basic idea
  - Suppose that we had an enormous database of breast cancer patients at our disposal
  - We would select a subset of the data, consisting of everyone alive 3 years after surgery (a landmark data set)
  - And simply count how many are alive 5 years later 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

$$h(t \mid Z(s), s) = h_0(t \mid s) \exp(\beta(s)^\top Z(s))$$

  for $s \leq t \leq t_{\text{hor}}$, enforcing administrative censoring at t_{hor}
- After having obtained estimates $\hat{\beta}(s)$ and $\hat{h}_0(t \mid s)$:
- Estimate of prediction probability $P(T > t_{\text{hor}} \mid T > s, Z^*(s))$ is then given by $\exp(-\exp(\hat{\beta}(s)^\top Z^*(s))\hat{H}_0(t_{\text{hor}} \mid s))$

Robustness

- Note: for fixed s and t_{hor}, the Cox model

$$h(t \mid Z(s), s) = h_0(t \mid s) \exp(\beta(s)^\top Z(s))$$

  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

$$h(t \mid Z(s), s) = h_0(t \mid s) \exp(\beta(s)^\top Z(s))$$

  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

$$h_0(t \mid s) = h_0(t) \exp(\theta(s))$$

  with $\theta(s_0) = 0$ for identifiability
- In our application we take

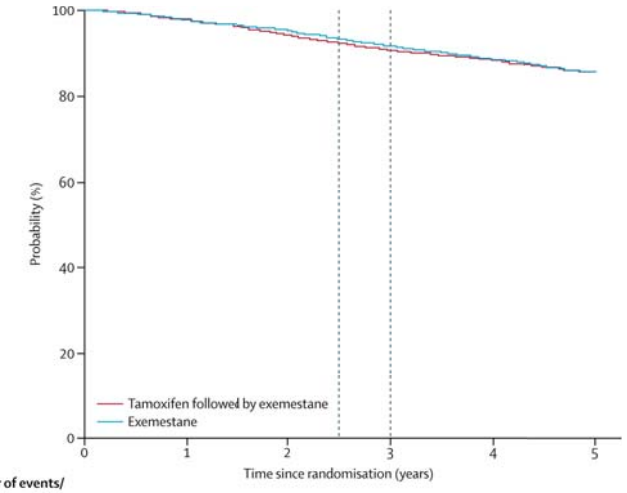
$$\theta(s) = \theta_1 s + \theta_2 s^2$$

- 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

TEAM study

- Multinational open-label phase III randomized clinical trial in postmenopausal hormone-sensitive breast cancer patients
- Randomized to receive
  - Exemestane (25mg once-daily) for 5 years, or
  - Tamoxifen (25mg once-daily) for 2.5-3 years, followed by exemestane (25mg once-daily) for 2-2.5 years, for a total of 5 years
- Participants enrolled in nine countries worldwide
- Current analysis based on the Dutch TEAM patients
- Primary endpoint: disease-free survival
- Primary endpoint not significant (HR=0.97; 95% CI 0.88-1.08) (van de Velde et al. Lancet 2011)

TEAM study



Number of events/ number at risk						
Tamoxifen followed by exemestane	4868	111/4660	160/4436	155/4140	108/3377	100/2529
Exemestane	4898	109/4716	117/4533	166/4272	133/3575	107/2564

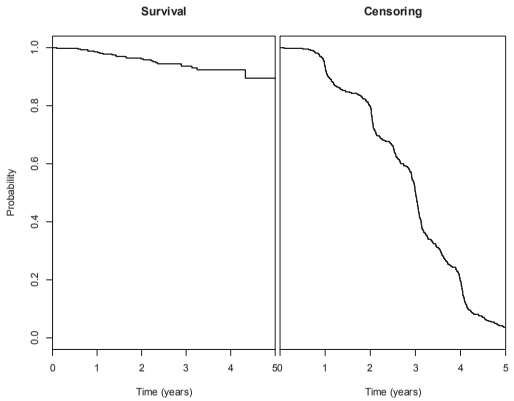
Characteristic		*n*	(%)
Age	< 65	1447	(56%)
	65-74	721	(28%)
	≥ 75	429	(17%)
Tumor stage	T0/T1	1132	(44%)
	T2	1275	(49%)
	T3/T4	190	(7%)
Nodal stage	N0	820	(32%)
	N1	1342	(52%)
	N2/N3	435	(17%)
Histological grade	BR I	382	(15%)
	BR II	1198	(46%)
	BR III	1017	(39%)
Estrogen receptor status	Negative	57	(2%)
	Positive	2540	(98%)
Progestrogene receptor status	Negative	578	(22%)
	Positive	2019	(78%)
Most extensive surgery	Mastectomy	1417	(55%)
	Wide local excision	1180	(45%)
Radiotherapy	Yes	1716	(66%)
	No	881	(34%)
Chemotherapy	Yes	840	(32%)
	No	1757	(68%)

Set-up

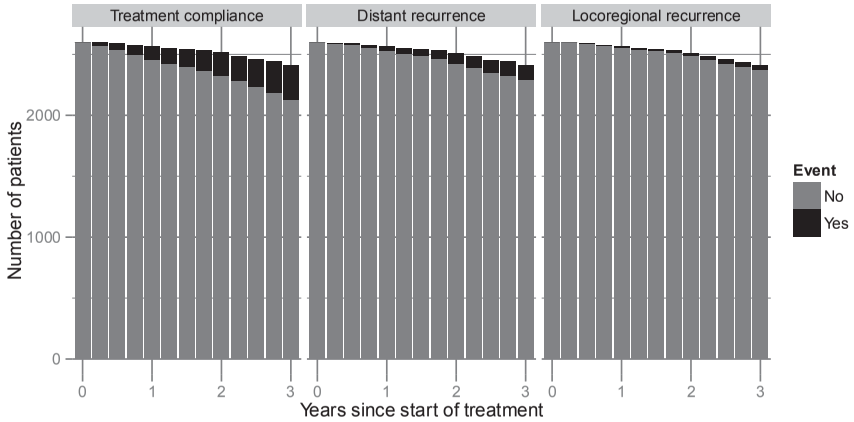
- Endpoint is survival in a window of fixed width $w = 5$ years from the moment of prediction
- Landmark time points used: equally spaced 3 months apart, from $s = 0$ to $s = 3$ years
- For each landmark (prediction) time point, construct landmark data set, containing all relevant information needed for the prediction
- In all data sets we take all patients still at risk (alive), compute the current value of LR, DM and compliance, and set the horizon at $t_{\text{hor}} = t_{\text{LM}} + 5$ 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 $t_{\text{hor}} + 5$

TEAM NL

- Based on patients with complete covariate information (2792/3157)
- Events: 90 local recurrences, 410 distant recurrences, 561 deaths



The landmark data sets



Landmark super model

Time-constant effects

Covariate	Category	B	SE
Age	< 65		
	65-74	0.277	0.126
	≥ 75	1.084	0.134
Tumor stage	T0/T1		
	T2	0.259	0.104
	T3/T4	0.333	0.175
Histological grade	BR I		
	BR II	0.000	0.153
	BR III	0.353	0.157
Estrogen receptor status	Positive		
	Negative	0.569	0.317
Progestrogene receptor status	Positive		
	Negative	0.443	0.097
Most extensive surgery	Mastectomy		
	Wide local excision	0.061	0.132
Radiotherapy	Yes		
	No	0.267	0.133
Chemotherapy	Yes		
	No	0.193	0.135

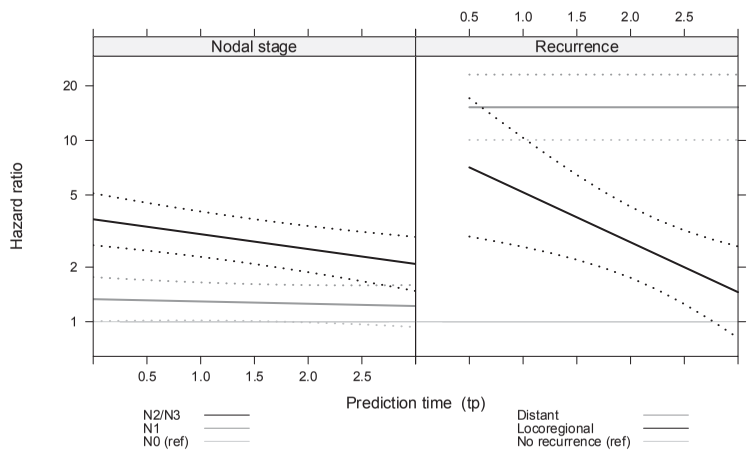
Landmark super model

Time-varying covariates and effects

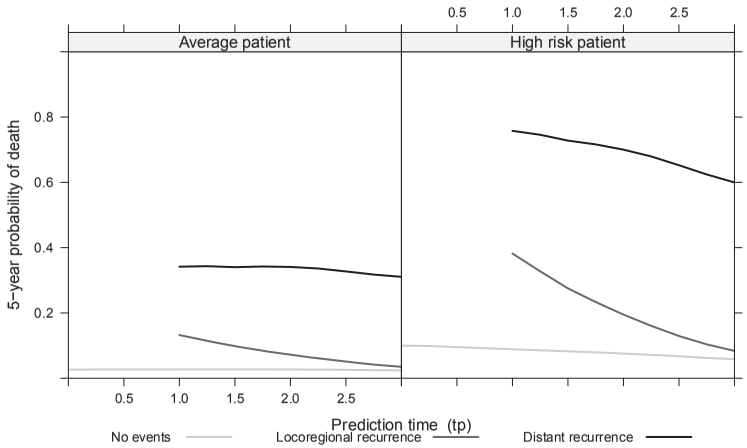
Time-dependent covariate	Category	B	SE
Treatment status	On treatment		
	Off treatment	0.240	0.198
Distant recurrence	No		
	Yes	2.723	0.212
Covariates with time-varying effects			
Prediction time	s	-0.023	0.050
	s^2	-0.028	0.010
Nodal stage	N0		
Constant			
	N1	0.286	0.143
	N2/N3	1.301	0.168
Prediction time			
	N1 * s	-0.029	0.048
	N2/N3 * s	-0.189	0.061
Locoregional recurrence	No		
Constant			
	Yes	2.277	0.551
Prediction time			
	Yes * s	-0.634	0.231

Time-varying effects

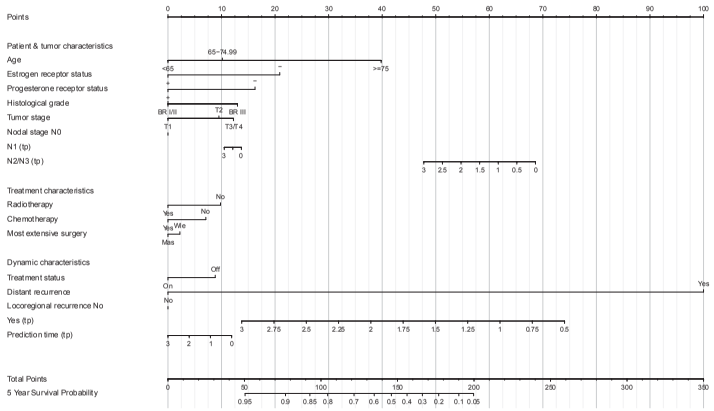
Time-varying hazard ratios for nodal stage and local recurrence



Dynamic predictions from the landmark super model



Dynamic nomogram



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

Discussion

- There may well be way too many prediction models in the medical literature
- But certainly not too many (if any?) dynamic prediction models
- Statistical tools are there
- They are not even difficult to implement
- We just have to use them!

References

van Houwelingen, H. C. (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