Landmarking 2.0

Bridging the gap between landmarking and joint models

Hein Putter, Hans C. van Houwelingen

Biomedical Data Sciences
Leiden University Medical Center
The Netherlands

SESSTIM QuanTIM Webinar
April 18, 2025

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

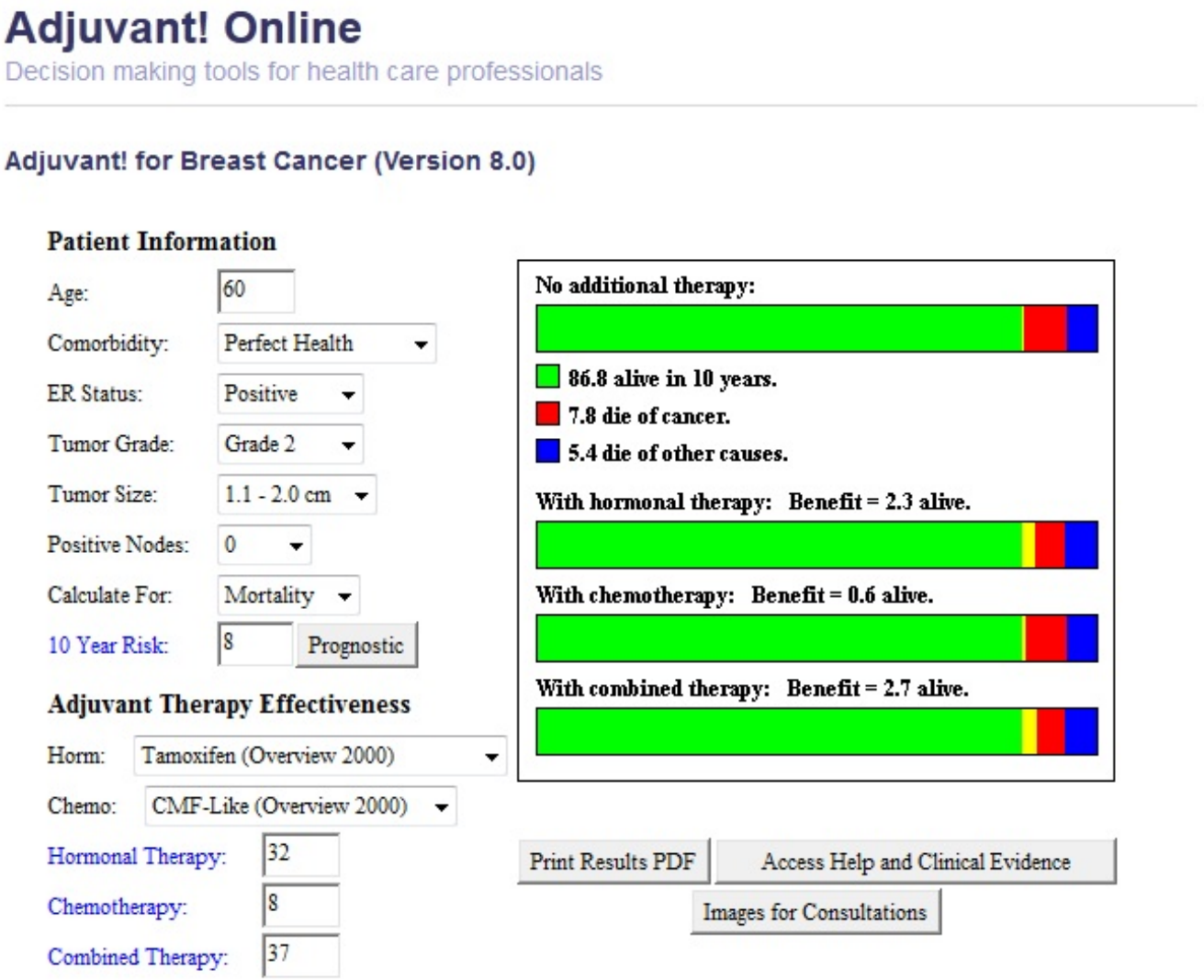
Komt een vrouw bij de dokter ...

- 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

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

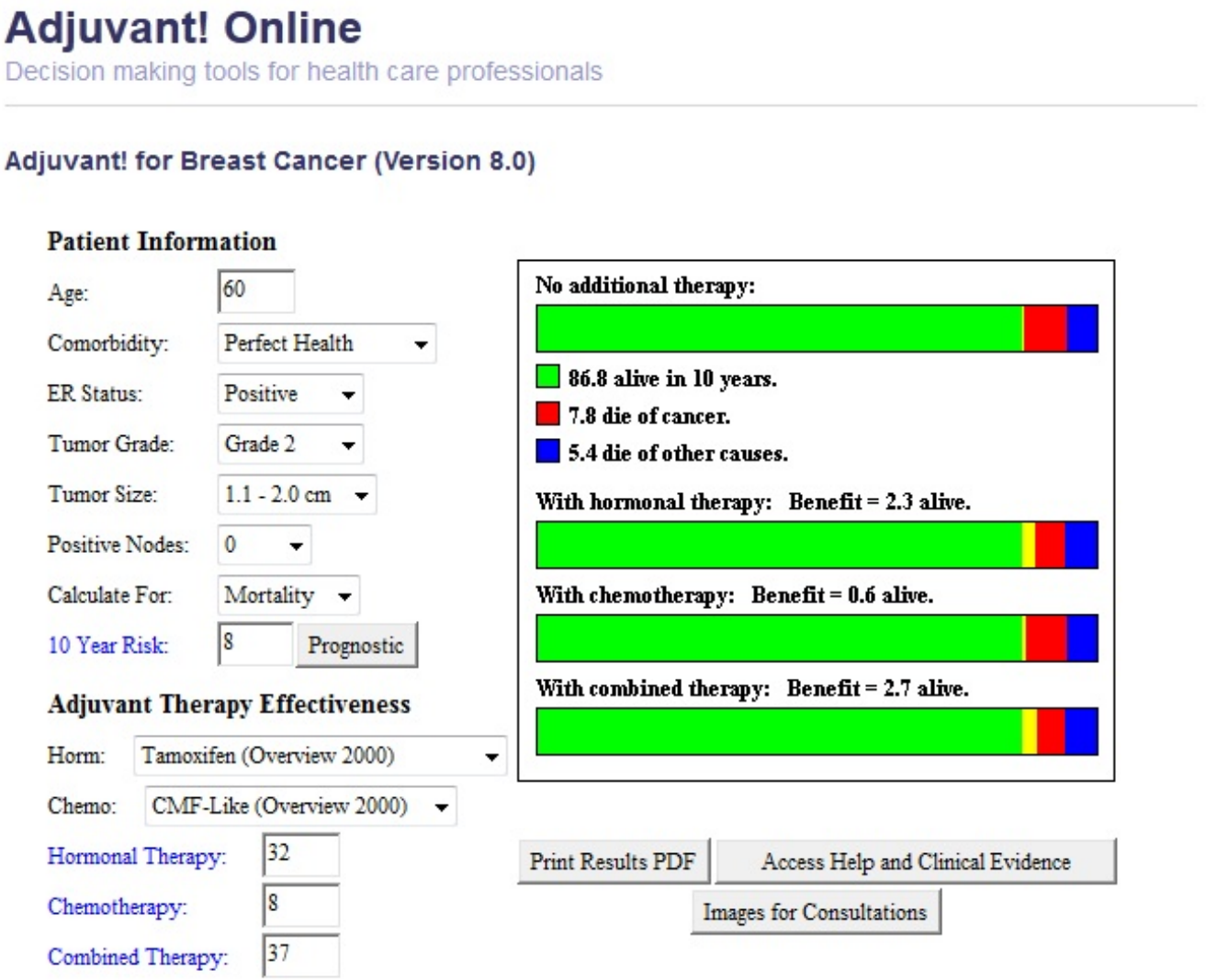
Komt een vrouw bij de dokter ...

- 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)

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 or 10 years from now?

Adjuvant! Online



Using Adjuvant! Online

- First temptation would be just to use 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

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

Using Adjuvant! Online

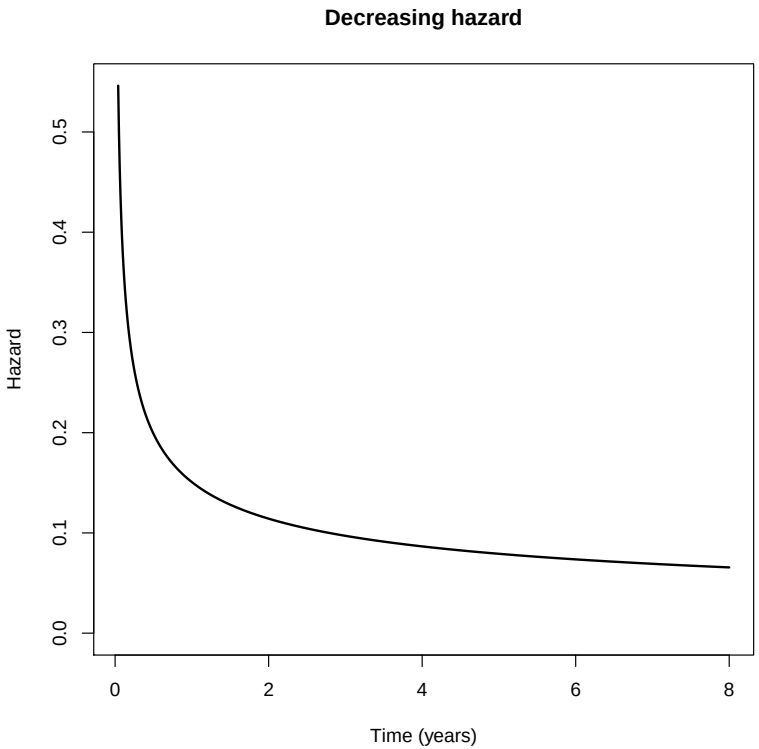
- First temptation would be just to use Adjuvant! Online

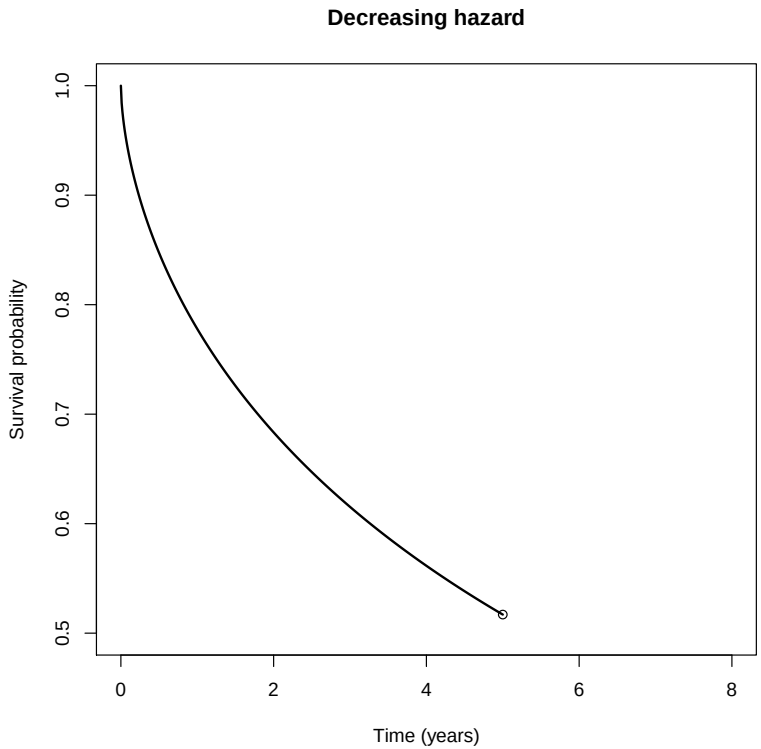
Why this isn't a good idea

- 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”

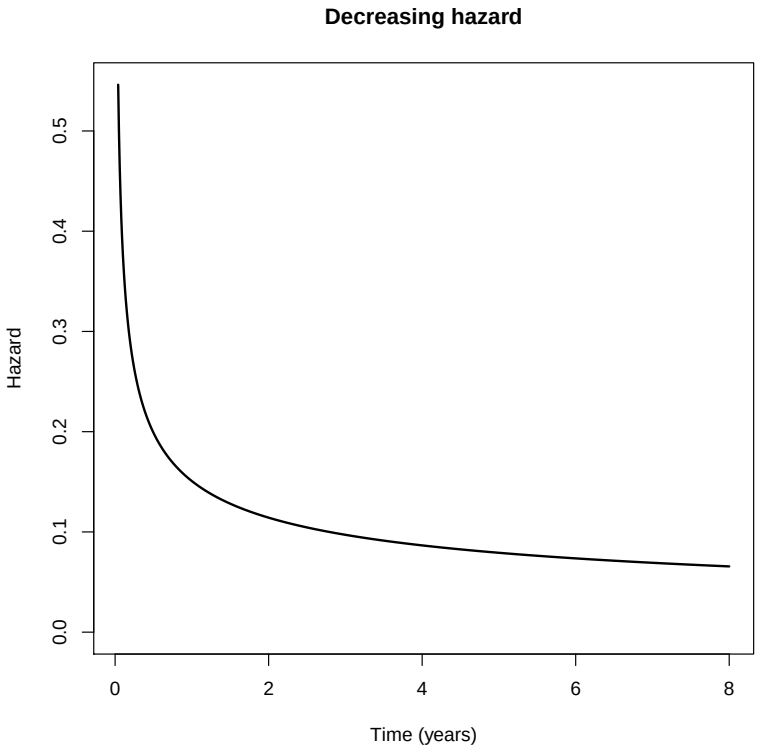
Prognosis may improve

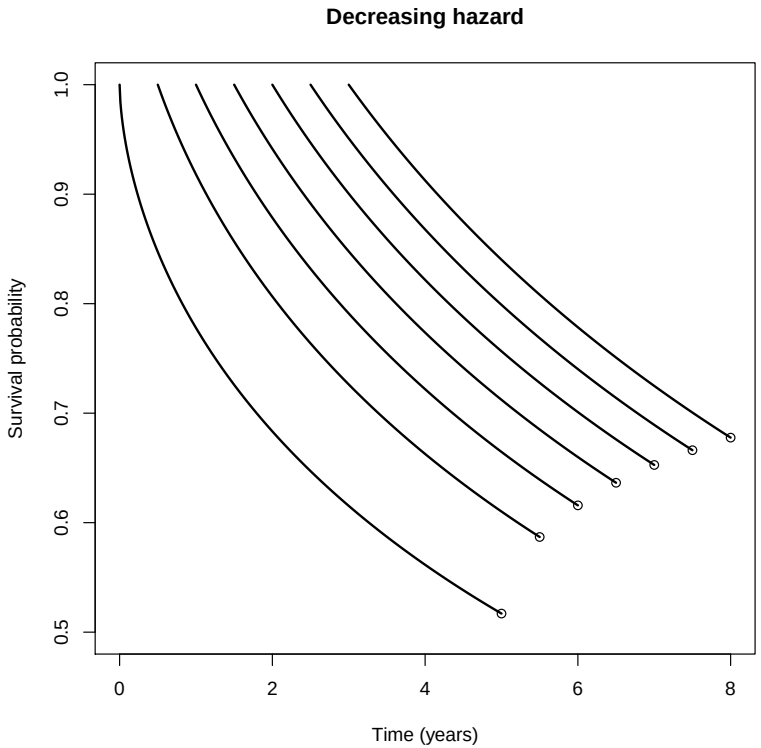




The effect of “being alive”

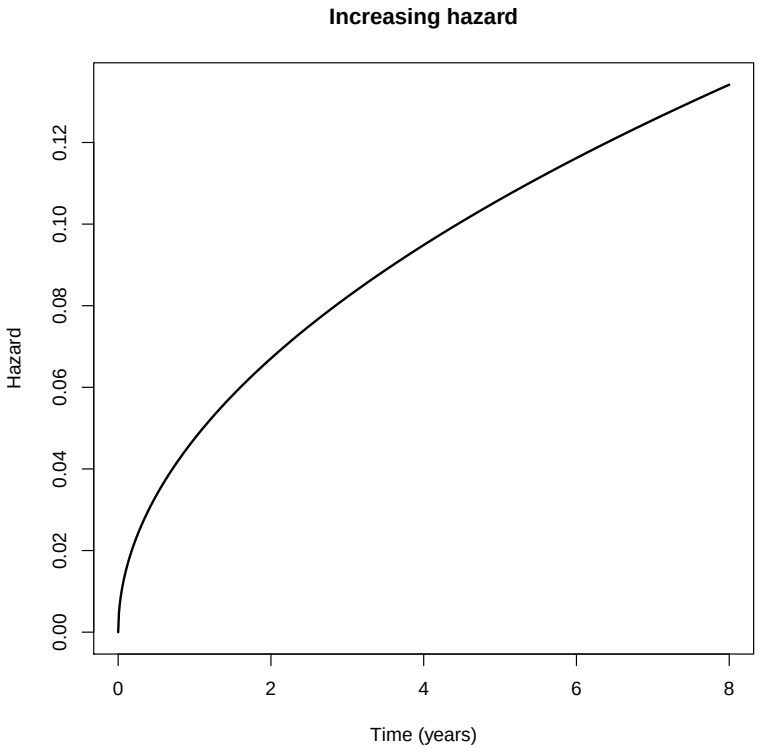
Prognosis may improve

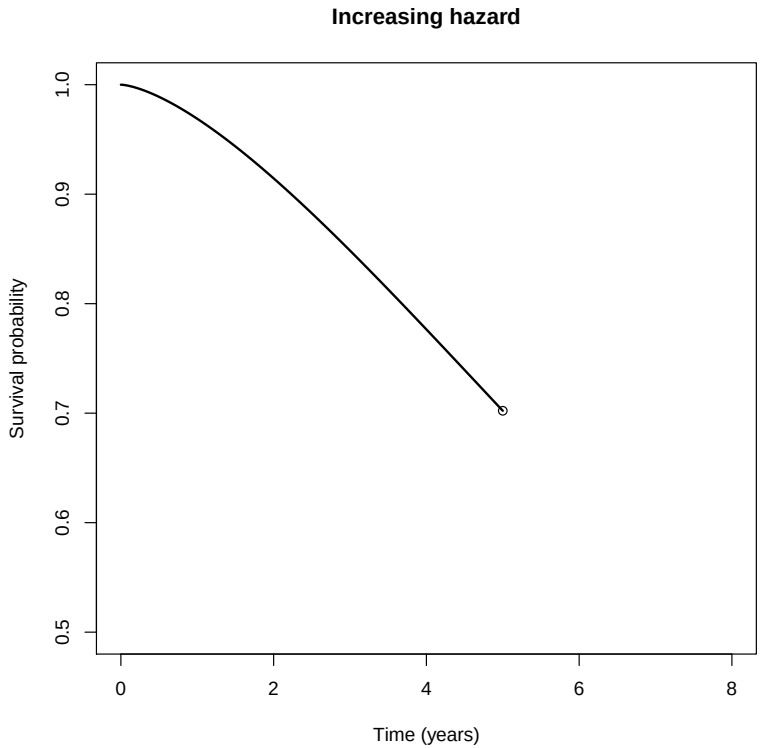




The effect of “being alive”

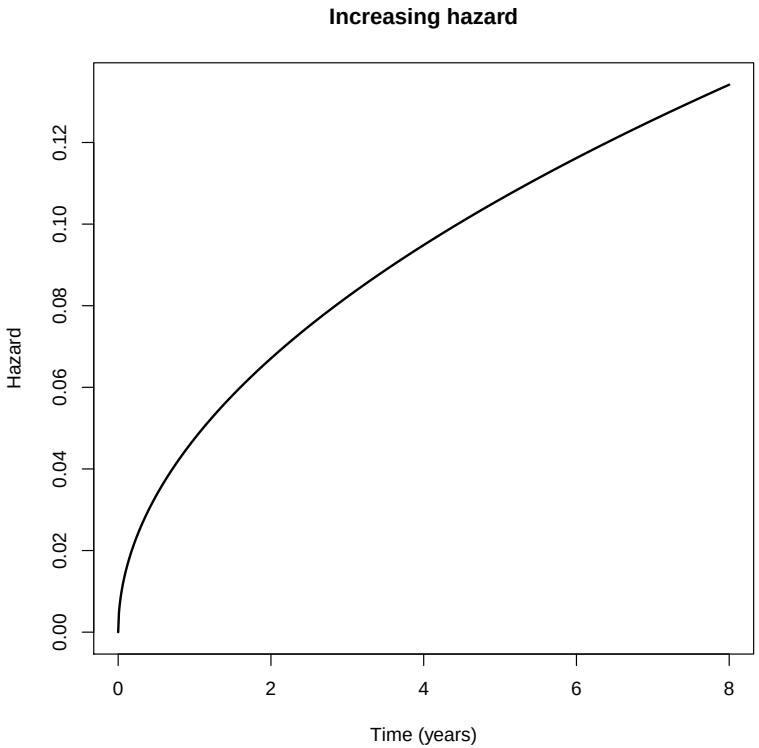
Prognosis may become worse

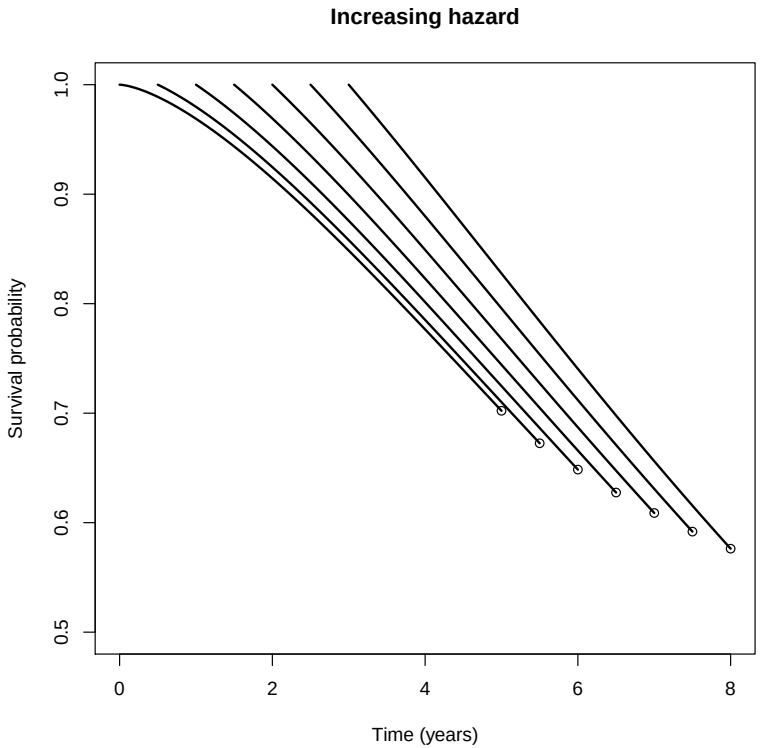




The effect of “being alive”

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 with biomarkers

Two "schools"

- Joint models
  - Mixed model for biomarkers linked to survival outcome
  - Advantage: efficient when well specified, software available
  - Disadvantage: computationally heavy
- Landmarking
  - Pragmatic approach, no model for longitudinal markers
  - Advantage: easy to implement, no specialised software needed
  - Disadvantage: less efficient
- Objective: to bring the two together

Dynamic prediction with biomarkers

Context / notation

- Patients ($i = 1, \ldots, n$) are followed from time $t = 0$ until an event (called death) occurs; T_i is the time of death
- Continuous biomarker process $X_i(t)$, defined as long as individual i is alive
- Process is observed at observation times t_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$
- Observations have measurement error / day-to-day variation (white noise)
- Actual observations denoted as X_{ij}
- Other covariates might be present, but will be ignored for the sake of presentation

Objective

- Objective is dynamic prediction: we want to estimate the conditional survival probability

$$\pi_i(s+w \mid s) = P(T_i > s+w \mid T_i \geq s, \mathcal{H}_i(s)),$$

with $\mathcal{H}_i(s)$= history (including biomarkers) up to s

- A Cox model with a time-dependent covariate $X(t)$

$$\lambda(t \mid X(t)) = \lambda_0(t) \exp(\beta X(t))$$

is helpful in understanding biology

Objective

- Objective is dynamic prediction: we want to estimate the conditional survival probability

$$\pi_i(s+w \mid s) = P(T_i > s+w \mid T_i \geq s, \mathcal{H}_i(s)),$$

with $\mathcal{H}_i(s)$= history (including biomarkers) up to s

- A Cox model with a time-dependent covariate $X(t)$

$$\lambda(t \mid X(t)) = \lambda_0(t) \exp(\beta X(t))$$

is helpful in understanding biology, but useless in predicting the future

  - Reason, we need $X_i(t)$ between s and $s+w$ to evaluate $\exp\left(-\int_s^{s+w} \lambda_0(t) \exp(\beta X_i(t)) dt\right)$

Objective

- Objective is dynamic prediction: we want to estimate the conditional survival probability

$$\pi_i(s+w \mid s) = P(T_i > s+w \mid T_i \geq s, \mathcal{H}_i(s)),$$

with $\mathcal{H}_i(s)$= history (including biomarkers) up to s

- A Cox model with a time-dependent covariate $X(t)$

$$\lambda(t \mid X(t)) = \lambda_0(t) \exp(\beta X(t))$$

is helpful in understanding biology, but useless in predicting the future

  - Reason, we need $X_i(t)$ between s and $s+w$ to evaluate $\exp\left(-\int_s^{s+w} \lambda_0(t) \exp(\beta X_i(t)) dt\right)$

- Two common approaches
  - Joint model
  - Landmarking

Joint model

- $X_i(t)$ follows a Gaussian process with mean $\mu_i(t)$ and covariance function $C(t_1, t_2) = \text{cov}(X_i(t_1), X_i(t_2))$
- Popular choice is a linear mixed model like $\mu_i(t) = \beta_0 + b_{i0} + (\beta_1 + b_{i1})t$, with random effects (b_{i0}, b_{i1}) assumed to be bivariate normal with mean zero
- $X_i(t)$ observed at t_{ij} with independent errors e_{ij}
- The hazard of dying at time t depends on the current value of the biomarker, for instance given by the PH-model

$$\lambda(t \mid X_i(t)) = \lambda_0(t) \exp(\beta \mu_i(t))$$

- More refined options (slope, AUC) possible
- Important to correctly specify, especially mean model (Ferrer et al. 2019)

Landmarking 1.0 van Houwelingen, SJoS, 2007

- Objective: the conditional survival
  $\pi_i(s+w \mid s) = P(T_i > s+w \mid T_i \geq s, \mathcal{H}_i(s))$
- Approach:
  - Use all data at time s
  - Summarize the history up to s
  - Apply administrative censoring at $t = s+w$, use the simple Cox model ("stopped Cox")
    - Remaining concern: For long term prediction (large w) time-varying effects might lead to bias
- Summaries of the history:
  - Simple summary: LOCF (last observation carried forward)
  - Extension: use the "age" of the last observation (difference between s and last observed time before s) as additional covariate
  - Concern: the staleness ("aging") of the predictor based on $X(t)$ calls for time-varying effects, but no simple model; threat of overfitting

Landmarking 1.5: two-stage approach

- For instance Sweeting et al. (2017)
- Use the data of the time-dependent covariate(s) before the landmark prediction time-point $t_{\mathrm{LM}} = s$
- Fit a mixed model to those data (or all data)
- Use the Empirical Bayes BLUP as a predictor at $t_{\mathrm{LM}} = s$
- It is called "error free" but that could be too optimistic
- It partly solves the staleness problem of the predictor at t_{LM}

Landmarking 2.0: getting closer to the joint model

Using the joint model for prediction

- The joint model approach leads to the following model for the conditional survival (see Putter & van Houwelingen, Stat Biosc 2017)

$$\pi_i(s+w \mid s) = E\left[\exp\left(-\int_s^{s+w} \lambda_0(t)\exp(\beta X_i(t))dt\right) \mid \mathcal{H}_i(s)\right]$$

- Following Tsiatis et al. (1995) in their treatment of measurement errors in survival analysis, the conditional survival can be approximated by

$$\pi_i(s+w \mid s) = \exp\left(-\int_s^{s+w} \tilde{\lambda}_0(t)\exp\left(\tilde{\beta}E[X_i(t) \mid \mathcal{H}_i(s)]\right)dt\right)$$

How to fit a landmarking 2.0 model

- Define and fit a working Gaussian process with trend $\mu(t)$ and covariance matrix $C(t_1, t_2)$ to the observed X_{ij}
- Use that to estimate $E[X_i(t) \mid \mathcal{H}_i(s)]$ for $t \geq s$ by least squares yielding the predictable time-dependent covariate $\hat{X}_i(t \mid s)$
- Fit a landmark Cox model with a fixed effect of the predictable time-dependent covariate $\hat{X}_i(t \mid s)$, yielding estimates $\hat{\beta}$ and $\hat{\lambda}_0(t)$

How to use a landmarking 2.0 model for prediction

- Use the Gaussian process again to estimate $E\{X^*(t) \mid T \geq s, \mathcal{H}^*(s)\}$ for $t \geq s$ by least squares yielding the predictable time-dependent covariate $\hat{X}^*(t \mid s)$
- Calculate the predicted hazard increments $\hat{\lambda}_0(u) \exp\{\hat{\beta}\hat{X}^*(u \mid s)\}$ for each event time point u between s and $s + w$ in the data
- Estimate

$$\hat{\pi}(s + w|s) = \exp\Big[- \sum_{s<u\leq s+w} \hat{\lambda}_0(u) \exp\{\hat{\beta}\hat{X}^*(u \mid s)\}\Big].$$

Properties

- The approach avoids latent variables and integration over random effects
- It gives a robust estimate of the survival given the predictable $\hat{X}(t \mid s)$
- It might be less efficient than the joint model, but it allows closer inspection and direct modeling of the trajectories of the survivors before estimating the regression parameters of the survival model

Working longitudinal model

- Variance components approach related to autoregressive model (as in Dempsey and McCullagh 2018)
- Separate models for trend $\mu(t)$ and temporal covariance $C(t_1, t_2)$
- Independent variance components

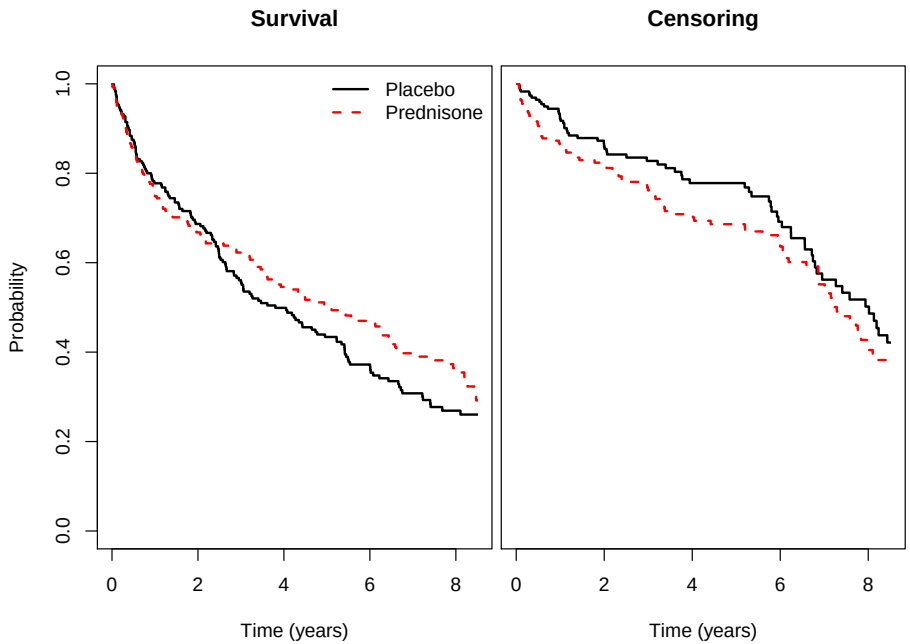
Component	Variance	Temporal correlation
Between individuals	σ_1^2	1
Within individuals	σ_2^2	$\exp(-\lambda\|t_1 - t_2\|)$
White noise	σ_3^2	0

leading to

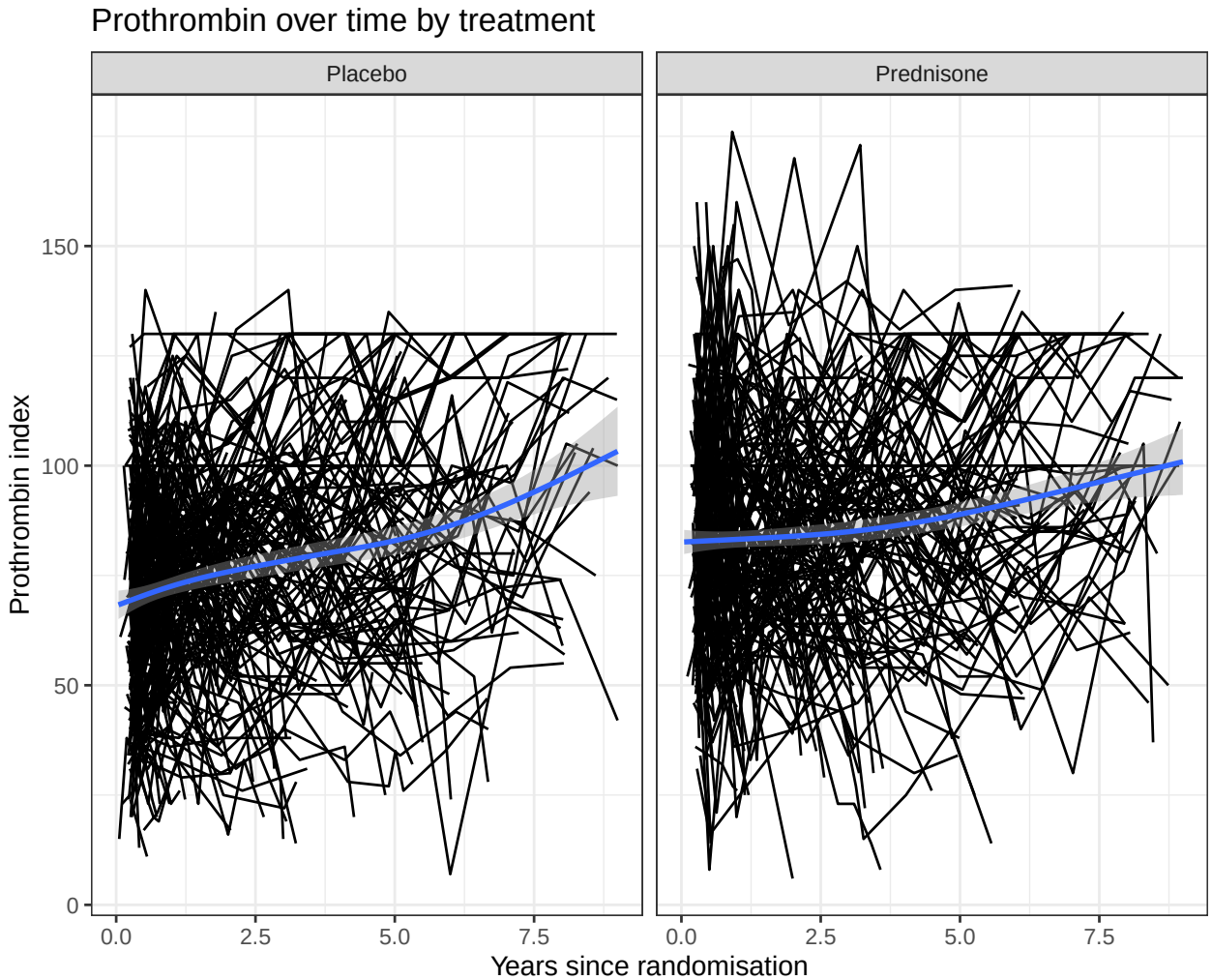
$$C(t_1, t_2) = \sigma_1^2 + \sigma_2^2 \exp(-\lambda|t_1 - t_2|) + \sigma_3^2 \mathbf{1}\{t_1 = t_2\}$$

Illustration: CSL-1 trial

- Randomized clinical trial, to evaluate the efficacy of prednisone ($n = 251$) vs placebo ($n = 237$) in patients with histologically verified liver cirrhosis
- Biomarker: prothrombin index, indicator of liver functioning



Marker values over time



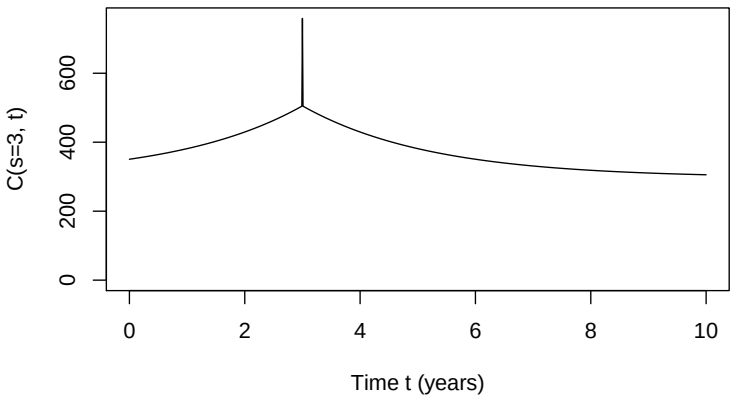
Longitudinal model

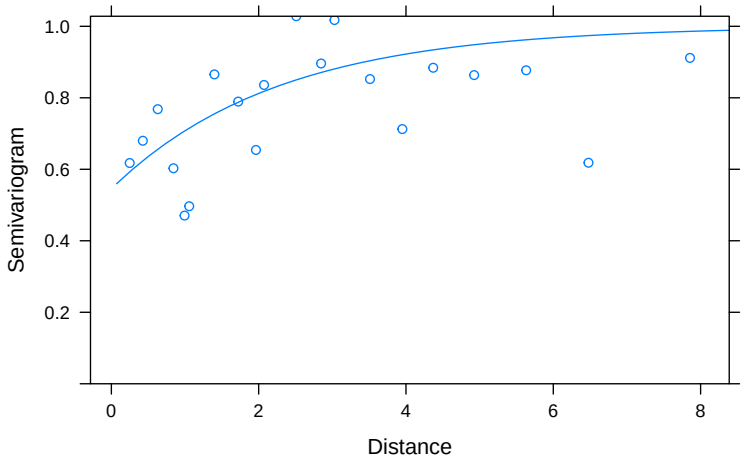
- We fitted the model mentioned earlier on both prednisone and placebo patients
  - Different trends
  - Same covariance parameters
- Trend: $69.03 + 2.19t$ for the placebo arm and $80.57 + 1.03t$ for prednisone arm
- Covariance parameters

Component	Variance	Estimate
Between individuals	σ_1^2	308.4
Within individuals	σ_2^2	240.8
Temporal decay parameter	λ	0.52
White noise	σ_3^2	184.3

Covariance parameters visualized

- Plot below shows $C(t_1, t_2) = \mathrm{cov}(X(t_1), X(t_2))$, for $t_1 = 3$, as function of t_2, and semi-variogram



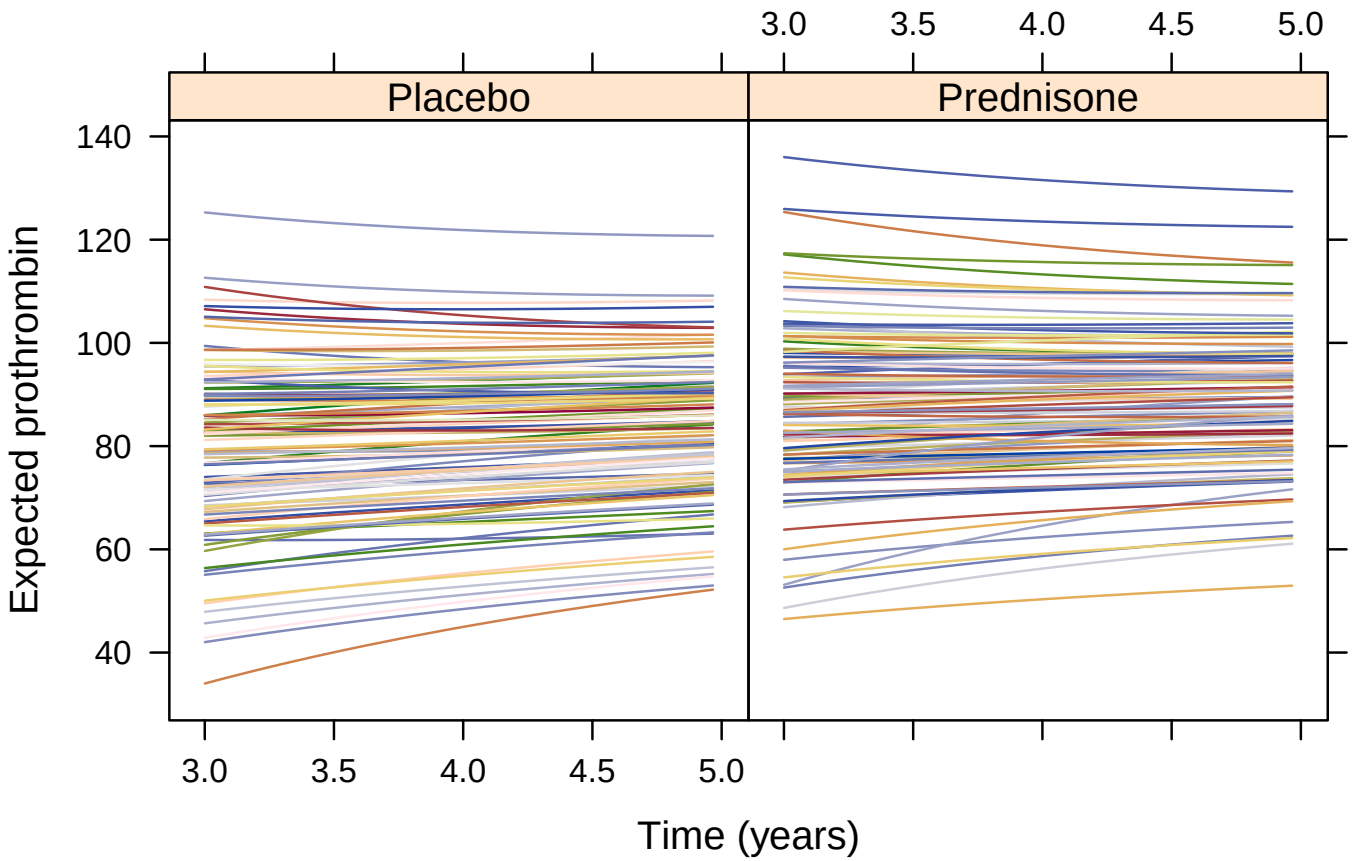


Landmark models

- Aim is dynamic prediction of survival at $s + w$ years for those alive at $s = 3$ years
- We take $w = 2$
- Using marker values up to s
- Four Cox models fitted
  - Based on last observed measurement (LOCF)
  - Based on the BLUP $\hat{X}(s \mid s)$
  - Based on $\hat{X}(t \mid s)$ as predictable time-dependent covariate from Gaussian process
  - Based on $\hat{X}(t \mid s)$ as predictable time-dependent covariate from revival model (in a minute)

$\hat{X}(t \mid s)$

- Plot of the individual trajectories of $\hat{X}(t \mid s)$ over time
- For t between s and $s + w$



Revival

- Different approach of obtaining $\hat{X}(t \mid s)$
- Interesting paper by Dempsey and MacCullagh (2018)
- Fit model to biomarker data in reverse time (backward from death)
- Problem: how to deal with censoring
- Define late horizon τ, and
  - Subject i dead at $t_i < \tau$: $Z_i(u) = X_i(t_i - u)$ (back from death)
  - Subject i alive at τ or censored before τ: $Z_i(u) = X_i(\tau - u)$ (back from horizon)

Revival and dynamic prediction

Direct revival

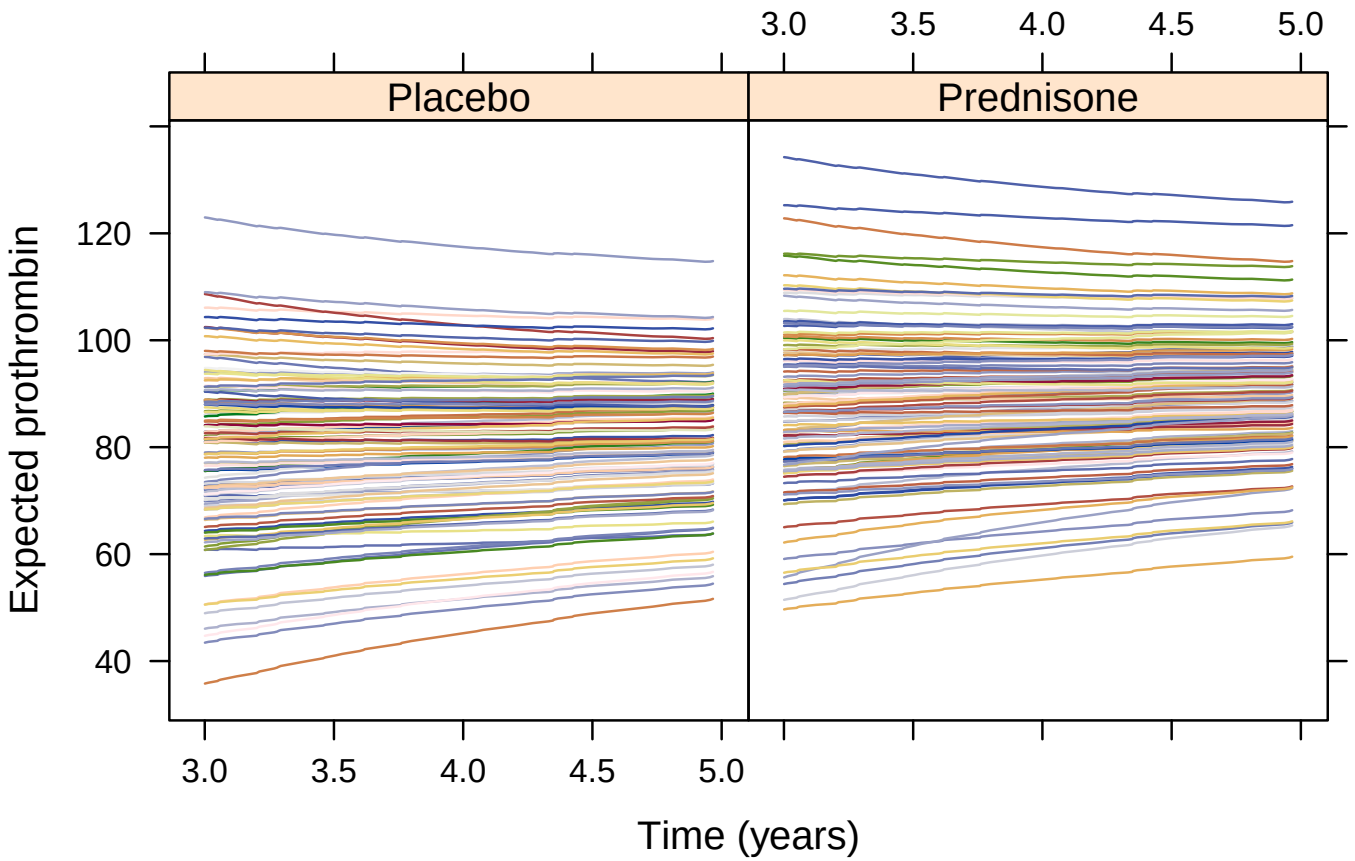
- Fit longitudinal models to observed $Z_i(u)$, separately for "dead" and "survivors"
- Use marginal model for time to event, fitted longitudinal models and Bayes' rule to obtain (t all event time points before τ (dead) and τ (survivors))

$$P(T = t \mid T > s, \overline{X}(s)) \propto P(\overline{X}(s) \mid T = t, T > s) \cdot P(T = t \mid T > s)$$

Revival and landmarking 2.0

- Going one step further (formulas ugly and not shown), one can also derive $\hat{X}_i(t \mid s) = E[X_i(t) \mid \mathcal{H}_i(s)]$ for $t \geq s$ from these same models and use these in landmarking 2.0

The predicted trajectories based on revival



Results of Cox models

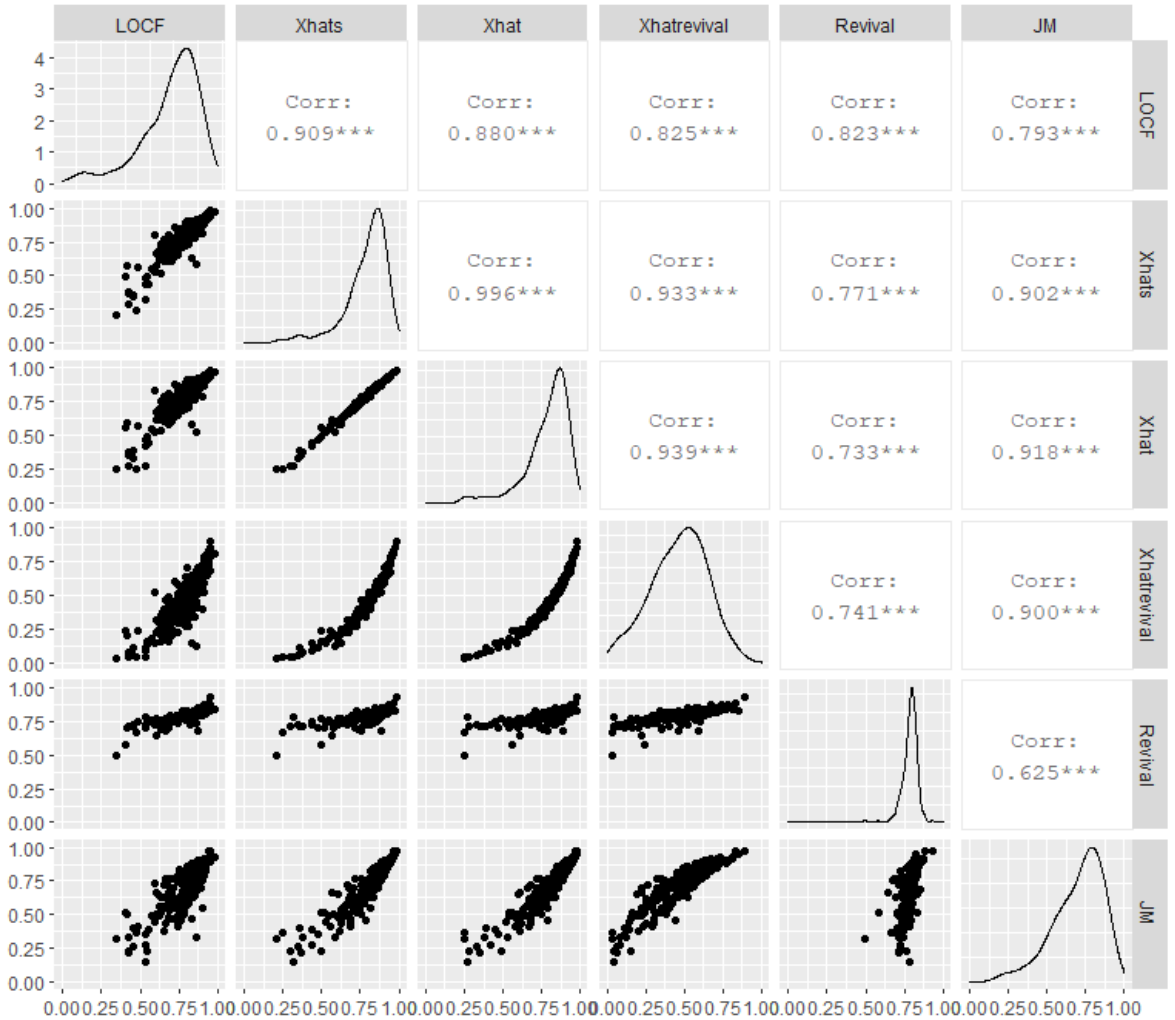
Model	Beta	SE	χ^2
Last observation	-0.0260	0.0063	17.016
$\hat{X}(s \mid s)$	-0.0446	0.0091	24.246
$\hat{X}(t \mid s)$	-0.0506	0.0010	25.614
$\hat{X}(t \mid s)$ revival	-0.0576	0.0106	29.485

- Notable difference between LOCF and $\hat{X}(s \mid s)$, very small difference between $\hat{X}(s \mid s)$ and $\hat{X}(t \mid s)$; $\hat{X}(t \mid s)$ based on revival wins

Comparison of dynamic predictions

- Joint model thrown into the mix
  - Standard settings, comparable to landmark models
  - Random intercept and slope, hazard depending on the value of the marker (proportional hazards)
- For proper comparison leave-one-out cross-validation used
  - One by one, leave one observation out, fit models and predict the conditional survival probability on the left out individual
- The revival models not well calibrated
  - Already noticed by Dempsey & McCullagh and discussants
- For revival results showing calibrated dynamic prediction probabilities

Dynamic predictions



Comparison of dynamic predictions

Model	Prediction error	
	Brier	KL
Null model	0.1683	0.5206
Joint model	0.1649 (2.1%)	0.5048 (3.0%)
Direct revival	0.1565 (7.0%)	0.4858 (6.7%)
Last observation	0.1585 (5.8%)	0.4932 (5.3%)
$\hat{x}(s \mid s)$	0.1549 (8.0%)	0.4797 (7.9%)
$\hat{x}(t \mid s)$	0.1549 (8.0%)	0.4791 (8.0%)
$\hat{x}(t \mid s)$ revival	0.1536 (8.7%)	0.4751 (8.7%)

Simulation study

Aim

- Compare predictive accuracy of predicted probabilities obtained using LM1.0, LM1.5, LM2.0 and JM

Simulation study

Data generation: biomarker

- Ornstein-Uhlenbeck (OU) process $X^*(t)$, starting at $X^*(0) = 0$, further defined by

$$\mathrm{d}X^*(t) = -\theta X^*(t)\mathrm{d}t + \sigma\ \mathrm{d}W(t),$$

  with

  - $W(t)$: Wiener process
  - θ: degree of mean reversal (to zero)
  - σ: influence of the random fluctuations of the Wiener process

- Individual biomarker process given by

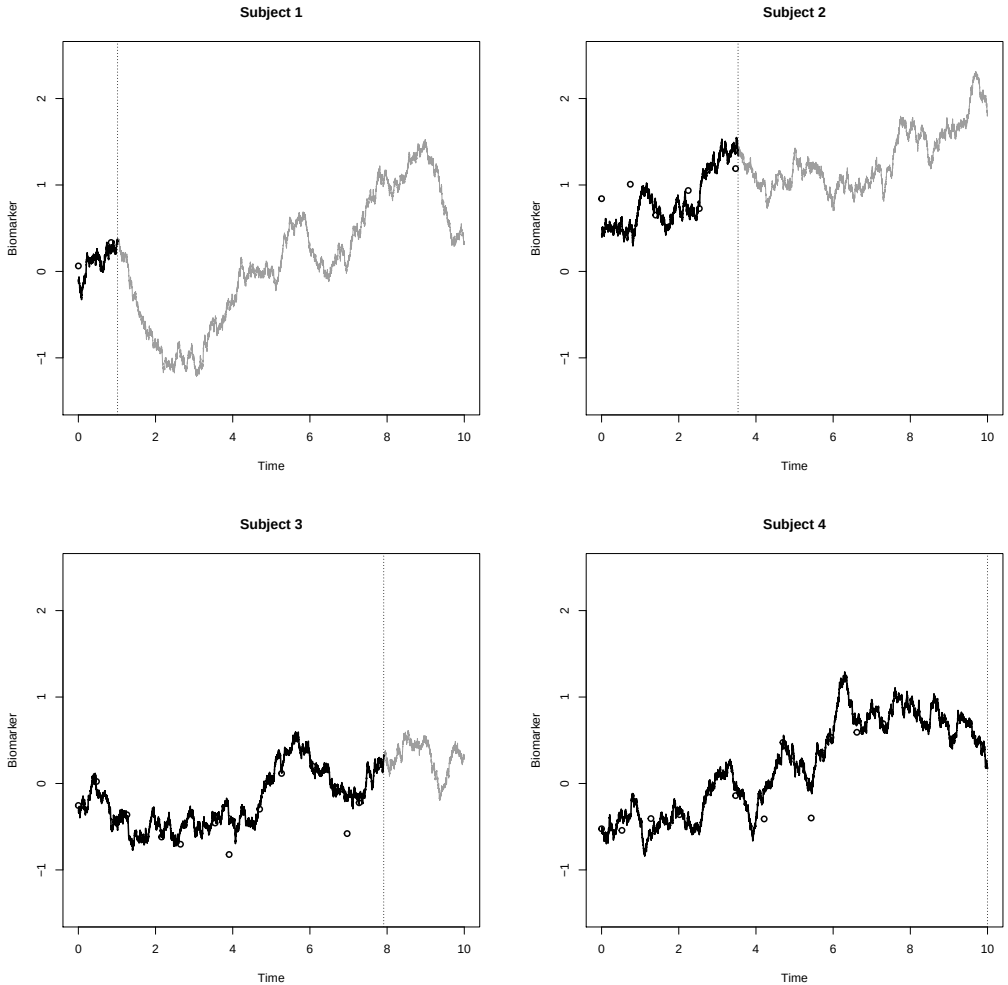
$$X_i(t) = \mu(t) + b_i + X_i^*(t)$$

- Result is Wiener process (Putter & van Houwelingen, 2017) with

$$\mathrm{cov}(X_i(s), X_i(t)) = \omega^2 + \frac{\sigma^2}{2\theta}\exp(-\theta|t-s|) = \sigma^2_{\mathrm{tot}}(\rho + (1-\rho)\exp(-\theta|t-s|))$$

Simulation study

Four biomarker sample paths



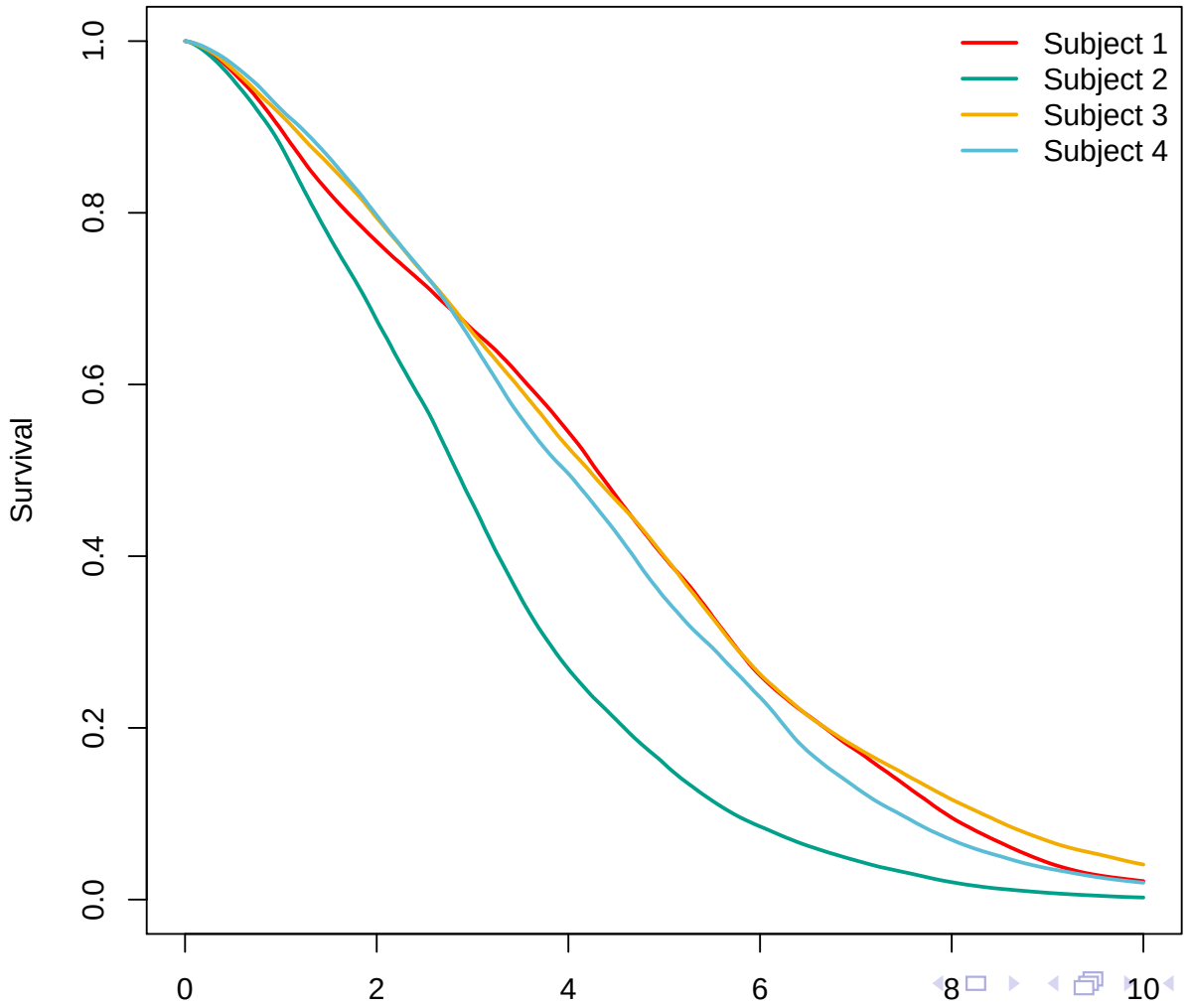
Simulation study

Data generation: the rest

- Baseline Weibull hazard $h_0(t; a, b) = abt^{b-1}$, $a = 0.1$, $b = 1.5$
- Hazard, given biomarker, given by $h(t \mid X(t)) = h_0(t; a, b) \exp(\beta X(t))$, with $\beta = 0.5$
- For given scenario, we first generated a single pool of validation data for a large number ($N = 6250$) of individuals
- Full data for each individual consists of $X_i(t)$ at a fine grid ($\Delta t = 0.01$) from $t = 0$ until $t = \tau = 10$
- True conditional survival probabilities can be calculated through $\pi_i(s, t) = \exp(-\int_s^t h(u \mid X_i(u))\mathrm{d}u)$
- Observed time to event is generated using $T_i = S_i^{-1}(U)$, with $S_i(t) = \exp(-\int_0^t h(u \mid X_i(u))\mathrm{d}u)$
- Observed biomarker data given by $X_i(t)$ observed at more or less regular intervals plus measurement error (SD=0.2)

Simulation study

Four survival curves



Simulation study

Data generation: training data

- We generated $M = 250$ sets of training data, consisting of $n = 500$ individuals each
- Data generated in the same way as the observed validation data

Models

- Four models considered
  - LM1.0: using the last observed biomarker value
  - LM1.5: based on the BLUP approach
  - LM2.0: newly proposed landmark approach
  - JM: joint model with linear trend, random intercept and slope for the longitudinal part, hazard based on the current value of the biomarker (JM), using the JM package
  - NULL: based on Kaplan-Meier estimates

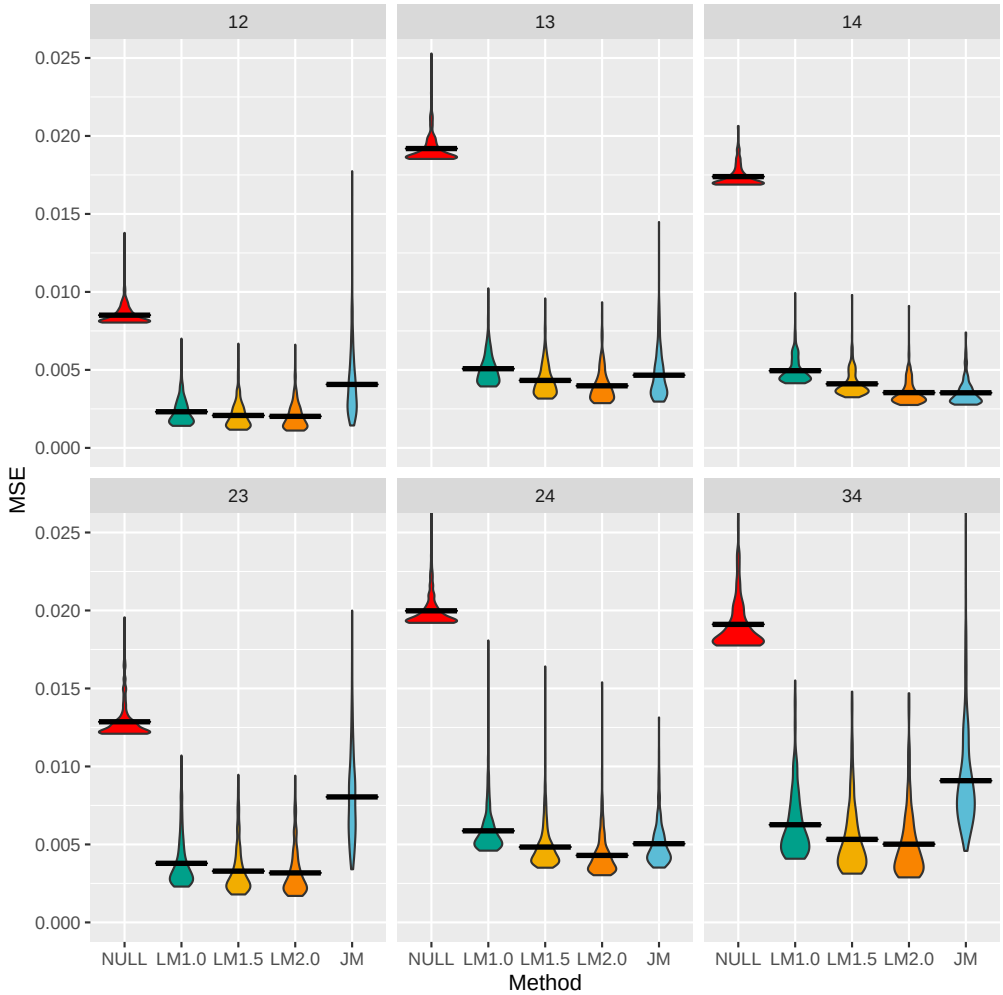
Simulation study

Estimands and performance

- Each model used to estimate $\pi_i(s,t) = P(T_i > t \mid T_i \geq s, \mathcal{H}_i(s))$ for $i = 1, \ldots, N$ in the large pool of validation data, where s is the landmark time point, t is the prediction horizon, and $\mathcal{H}_i(s)$ is the set of observed data of subject i before time s
- For s and t we used the quintiles of the marginal distribution of the generated event times in the validation set (which could differ from scenario to scenario)
- In output, s and t denoted by the numbers of these quintiles
  - For instance, "24" stands for s being the **second** (40%) quintile, and t being the **fourth** (80%) quintile of the marginal time to event distribution
- Estimated $\hat{\pi}_i(s,t)$ compared with the true $\pi_i(s,t)$ values
  - Bias: $N^{-1} \sum_{i=1}^{N} [\hat{\pi}_i(s,t) - \pi_i(s,t)]$
  - Mean squared error (MSE) $N^{-1} \sum_{i=1}^{N} [\hat{\pi}_i(s,t) - \pi_i(s,t)]^2$

Simulation study: results

Mean squared errors



Summary and future directions

- Uncertain in how far results on this one data set (and one choice of s and w) generalize
- The landmarking principle that "prediction should depend only on the past and nothing but the past in a transparent way" firmly stands
- Nevertheless there is a lot to learn from the "future of the past"
- Joint models and "revival" models can be helpful in building models beyond landmarking 1.0
- Including more-dimensional bio-makers needs more thinking how to handle the correlation between the biomarker components
- Competing risks relatively straightforward, also for revival

Material for this study

- Preprint on Stat Med website: `doi:10.1002/sim.9336`
- Full code on all analyses:
  `https://github.com/survival-lumc/Landmarking2.0`

References landmarking


Anderson, JR, Cain, KC and Gelber, RD.
Analysis of survival by tumor response.
J Clin Oncol 1983; **1**: 710–71.


van Houwelingen HC.
Dynamic prediction by landmarking in event history analysis.
Scan J Statist 2007; **34**: 70–85.


van Houwelingen HC and Putter H.
Dynamic Prediction in Clinical Survival Analysis.
Chapman & Hall/CRC, Boca Raton.


Sweeting MJ, Barrett JK, Thompson SG and Wood AM.
The use of repeated blood pressure measures for cardiovascular risk prediction: a comparison of statistical models in the ARIC study.
Stat Med 2017; **36**: 4514–4528.


Ferrer L, Putter H and Proust-Lima C.
Individual dynamic predictions using landmarking and joint modelling: validation of estimators and robustness assessment.
Stat Meth Med Res 2019; **28**: 3649–3666.

References other


Tsiatis A, DeGruttola V and Wulfsohn MS.
Modeling the relationship of survival to longitudinal data measured with error. applications to survival and CD4 counts in patients with AIDS.
JASA 1995; **90**: 27–37.


Putter H and van Houwelingen HC.
Understanding landmarking and its relation with time-dependent Cox regression.
Stat Biosc 2017; **9**: 489–503.


Dempsey W and McCullagh P.
Survival models and health sequences.
Lifetime Data Anal 2018; **24**: 550–584.