Session 3. Time-dependent covariates

Content

< simulation study

< example CML trial

< irregularly observed time-dependent covariate

< predicting by landmarking

< predicting by retrospective modeling

Simulation study

- Time-dependent covariate $X(t)$

  Wiener process with

  $$E[X(t)] = 0, \; cov(X(t),X(s)) = \sigma^2 e^{-?|t-s|} \text{ with } \sigma = 1 \text{ and } ? = 1$$

  simple stationary autoregressive model

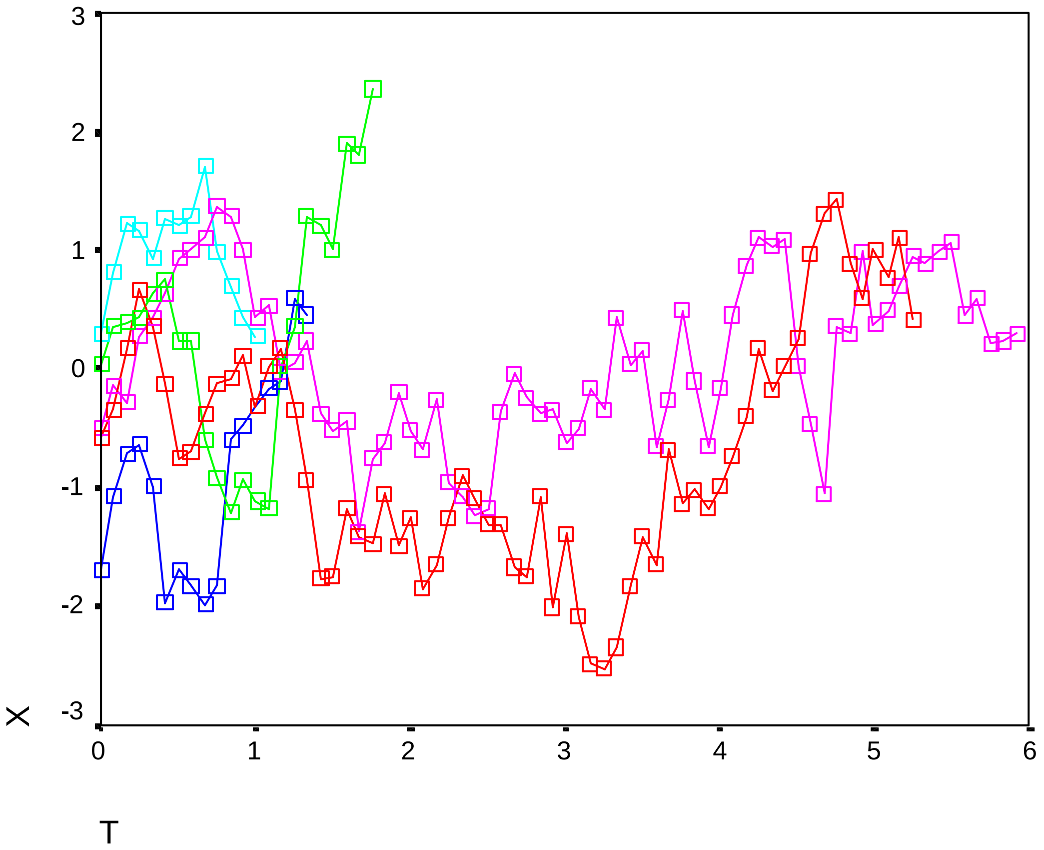
  no measurement error

- hazard function $h(t) = h_0 e^{-\beta X(t)}$ with $h_0 = 0.1$ and $\beta = 2$

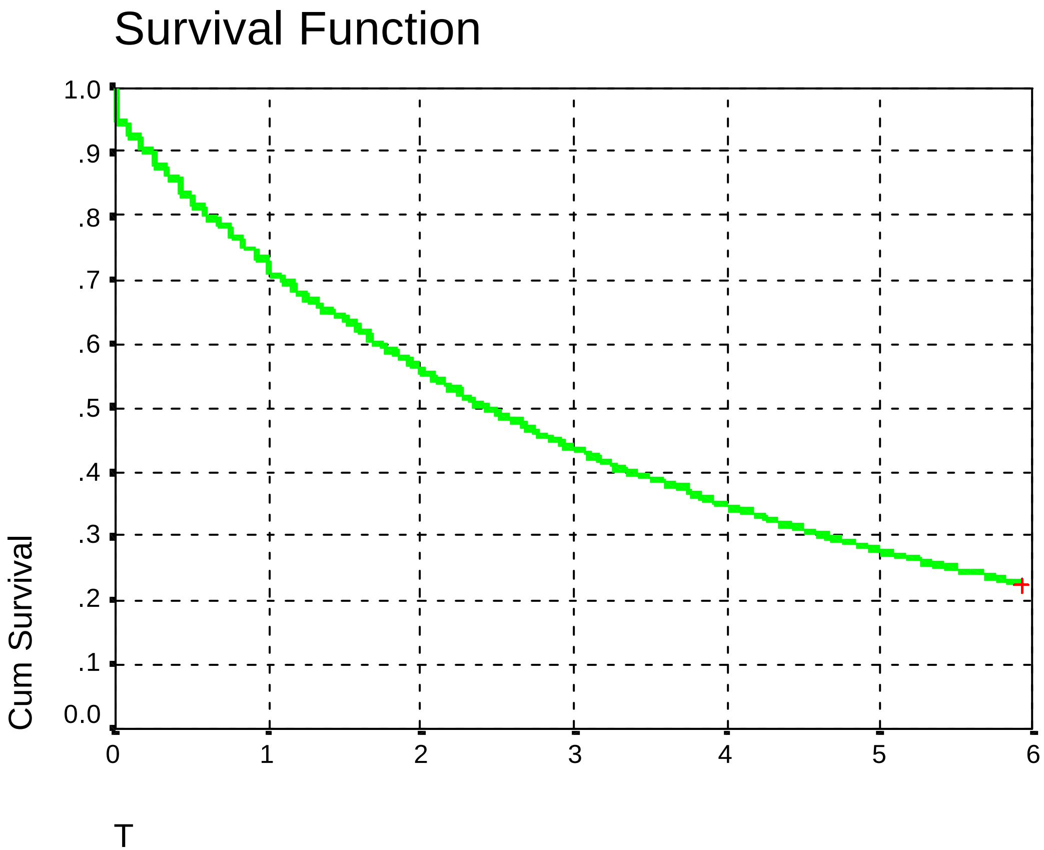
- discretized in staps of $? = 1/12$

- simulated on 2000 patients with follow-up till t=6

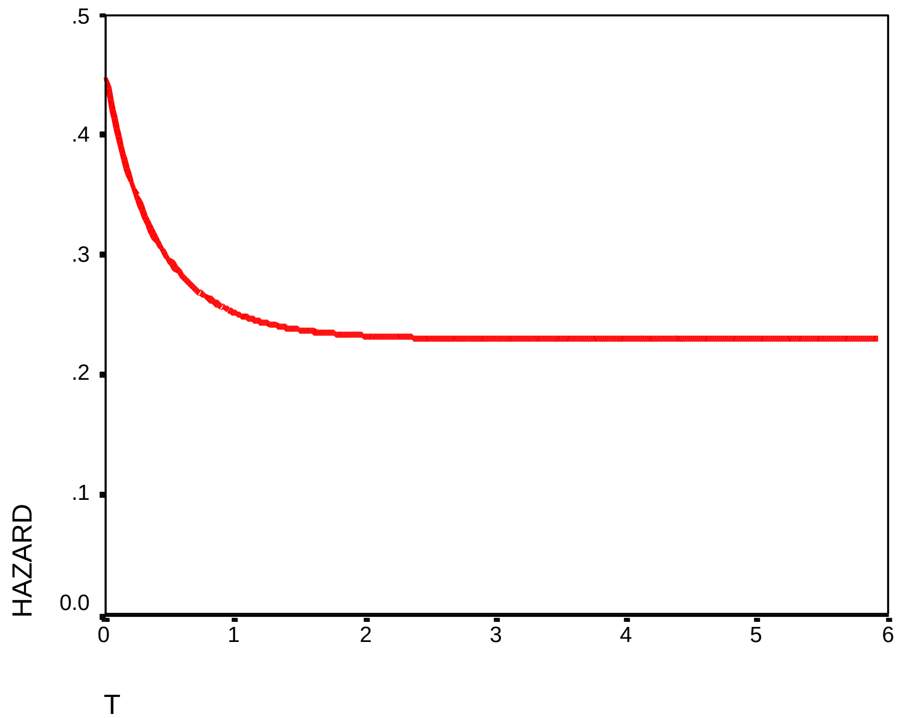
The first 5 simulated patients



Resulting survival function



The corresponding hazard looks like



It is not constant and much larger than 0.1 !!!!

Explanation:

$$h(t) = E[h_0 \exp(\beta X(t) | T \geq t]$$

at t=0

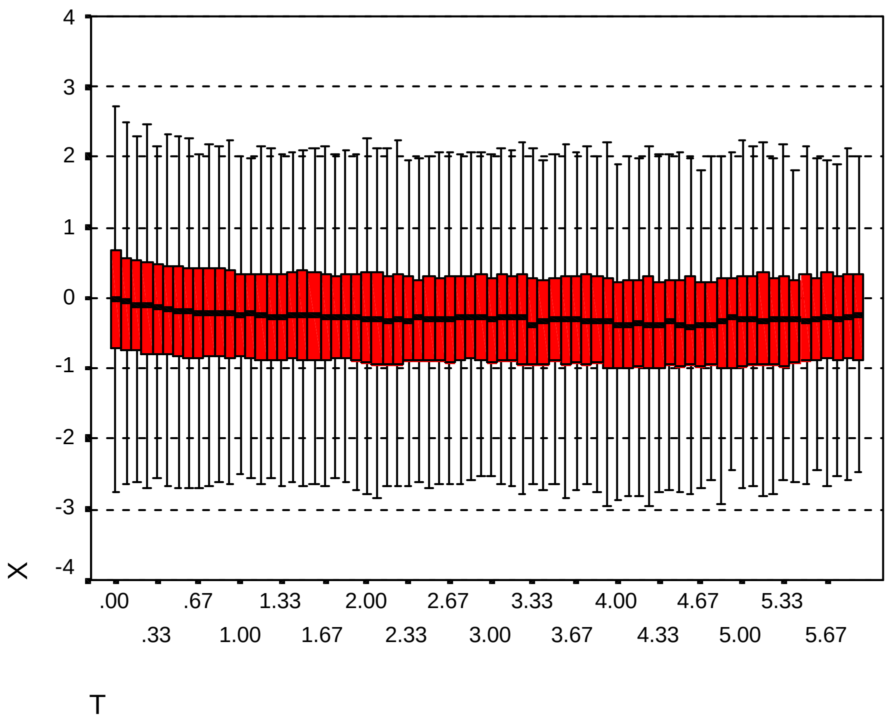
$$h(0) = h_0 \exp(\tfrac{1}{2}\beta^2) = 0.1 \times \exp(2) = 0.73$$

(not quite so in the data, due to discretization)

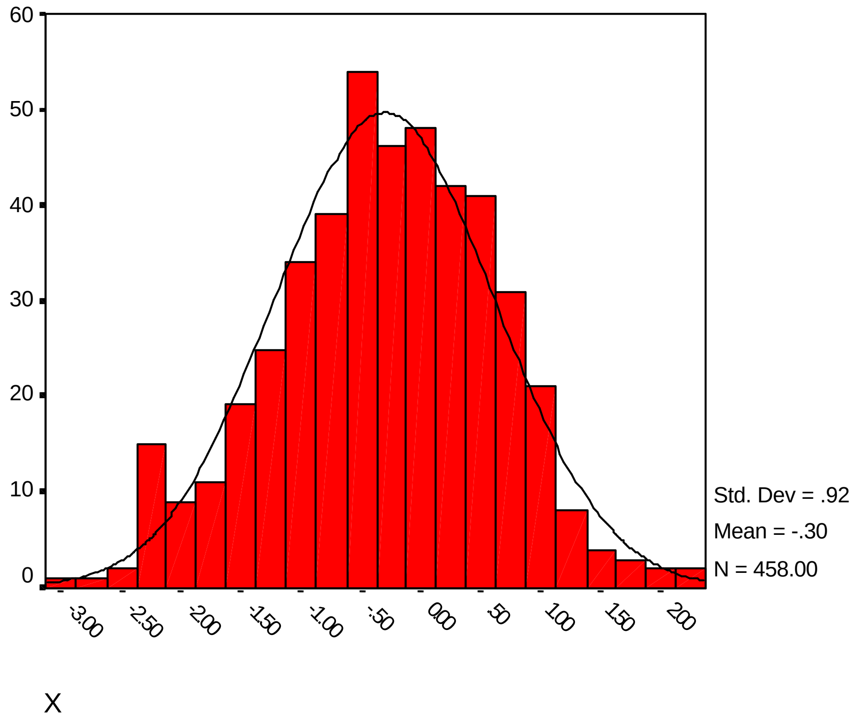
Distribution of $X(t)$ changes over time

< selection on survival „

< regression to the mean *f*



Histogram at t=6



Mean below zero and slightly skewed to the left.

No explicit solution for equilibrium distribution

Since death and censoring are typically Missing at Random the model can completely be recovered from the data.

Model easy to fit because of monotone drop-out

< X-model by regression on the past

< hazard model by logistic regression in risk sets

Big question: how to predict.

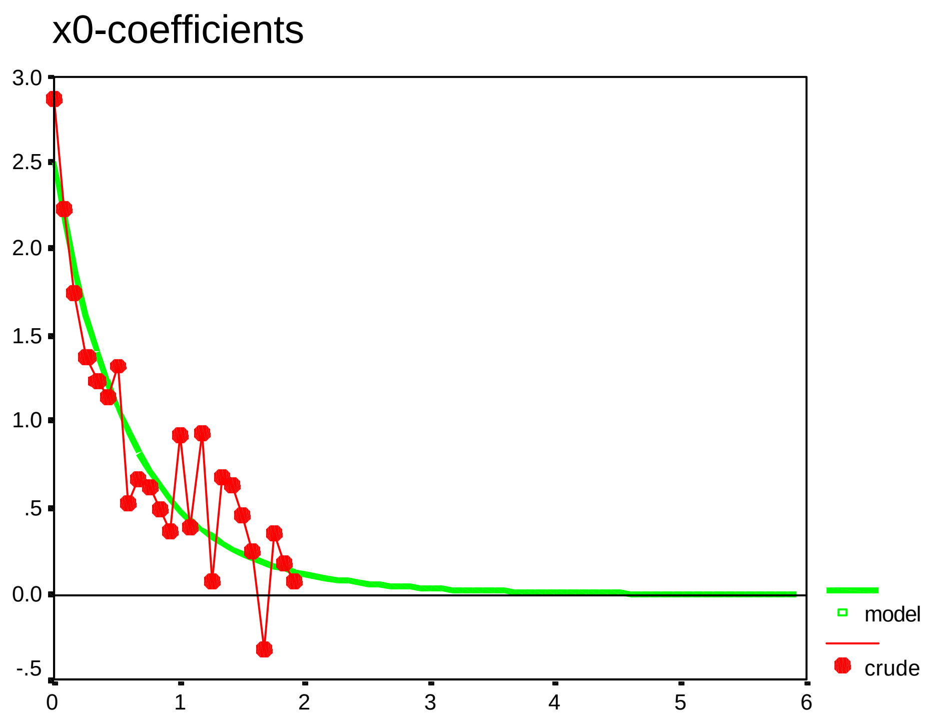
< Need only the last observation.

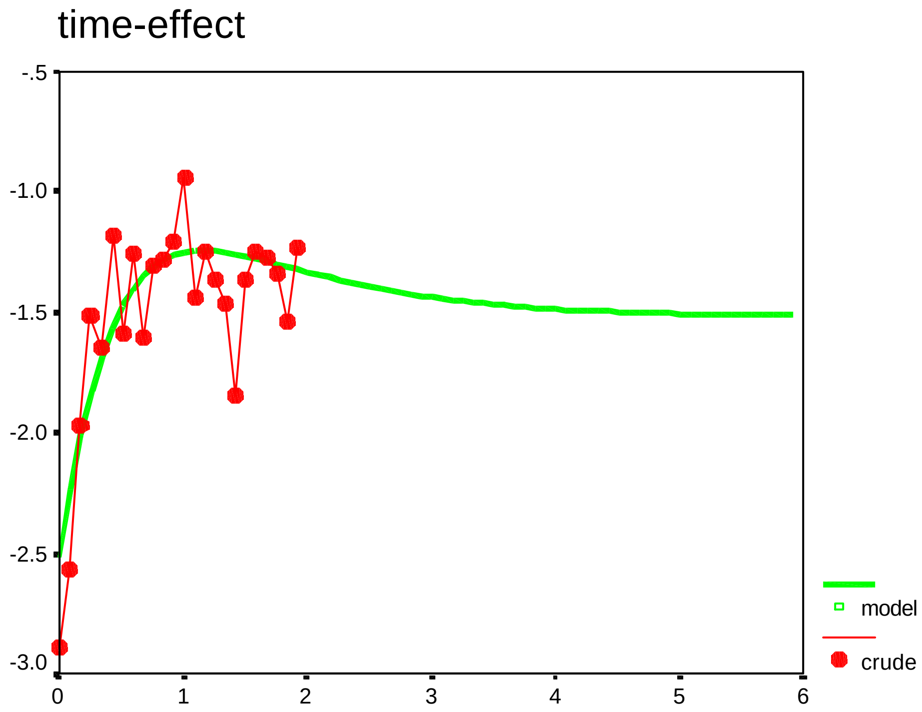
< Consider prediction based on first X-value X_0

- Formally $h(t|X_0) = E[h_0 \exp(\beta X(t)|X(0) = X_0, T \geq t]$
- No explicit solution as in the Manton-Woodbury model
- Approximation based on ignoring $T \geq t$ condition
  - $X(t)|X(0) \sim N(e^{-t}, 1-e^{-2t})$
  - $E[\exp(\beta X(t)\ X_0] = \exp(\beta e^{-t} X_0 + \frac{1}{2}\beta^2(1-e^{-2t}))$
  - $\ln(h(t|X_0)) = \ln(h_0) + e^{-t} X_0 \beta + \frac{1}{2}(1-e^{-2t})\beta^2$

Approximation is not more than approximation.

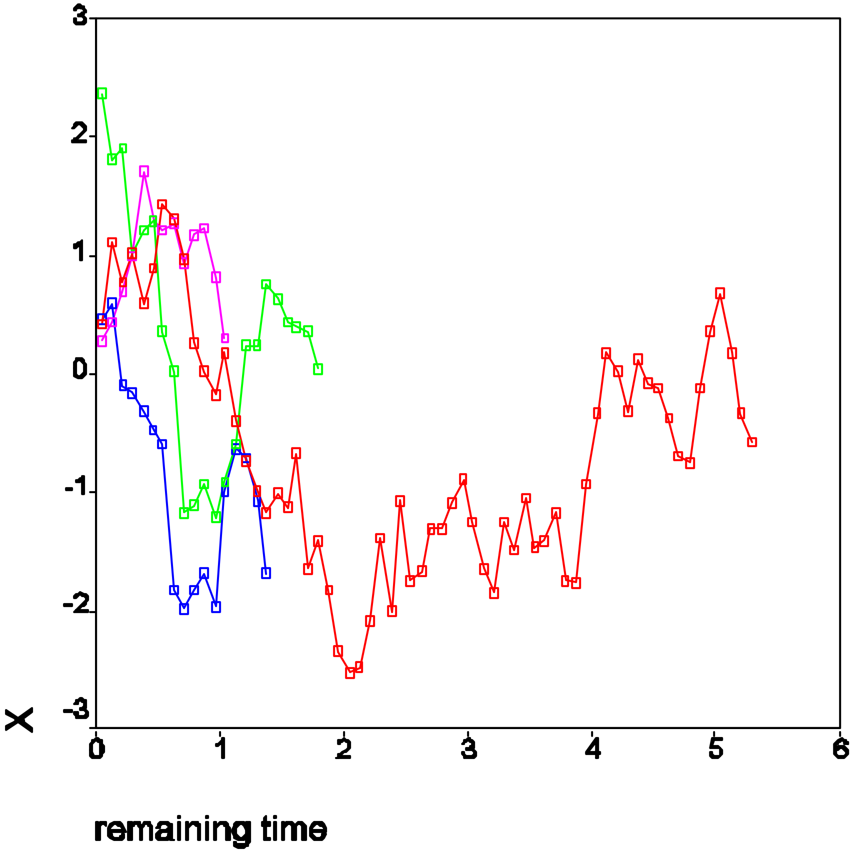
Modelling by exponentials .....



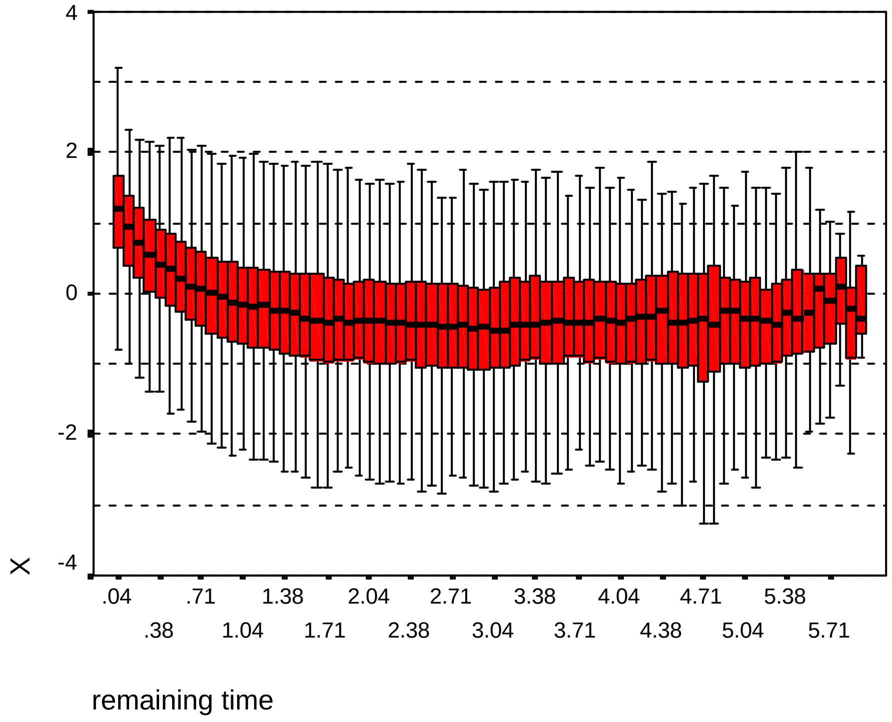


Conclusion: long-range hazard hard to model because of unpredictable behaviour of hazard

Retrospective time

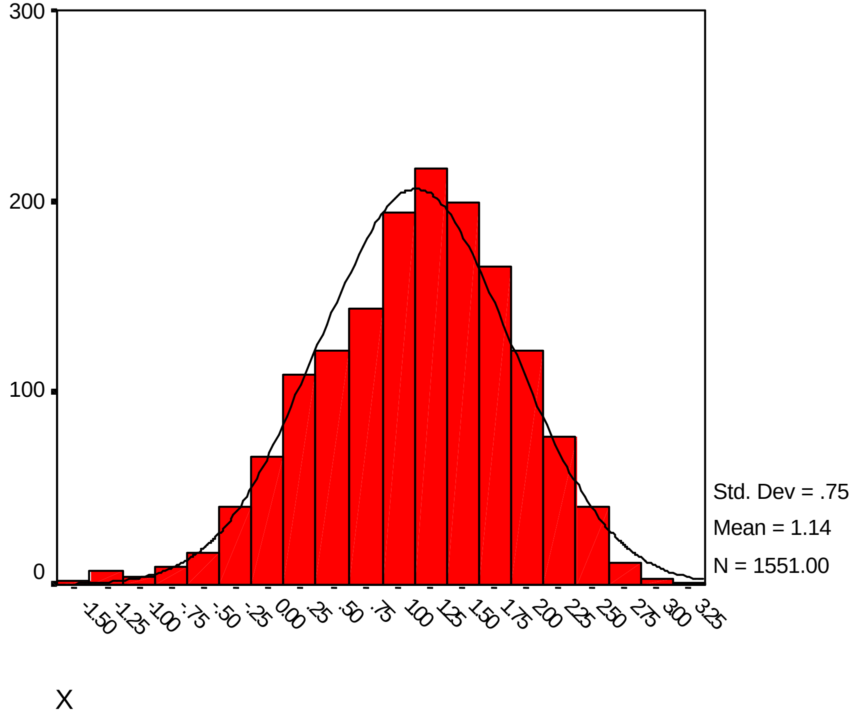


Boxplots

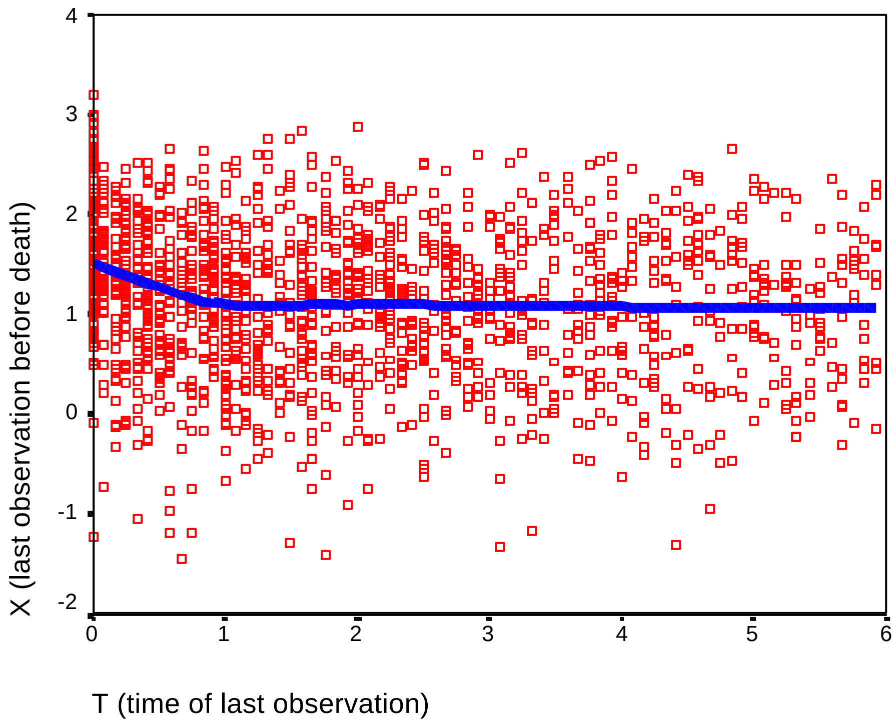


4
2
0
-2
-4
X
.04
.71
1.38
2.04
2.71
3.38
4.04
4.71
5.38
.38
1.04
1.71
2.38
3.04
3.71
4.38
5.04
5.71
remaining time

Histogram of last measurement before death



Some correlation with time of last observation



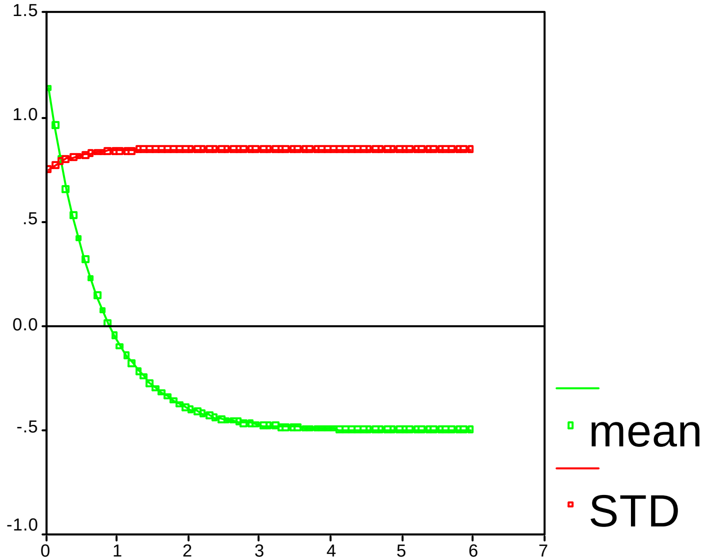
Model: $1.06\%0.56\,(\,\exp(\&4.55\,(\,t_{last})$ $\quad s\ '\ .73$

Retrospective model is purely Markovian, but not stationary.

Autoregressive model per month

X_t ' &0.055%0.879 ($X_{t\&1}$%0.385 (? (limiting s ' 0.84)).

Graph shows mean and st.dev as function of time remaining



Purely Markovian model makes that prediction depends only on last observation.

$$f(t|X(t_{last})) = \frac{f(t)f(X(t_{last})|T=t)}{\int_{t_{last}} f(t)f(X(t_{last})|T=t)dt} \quad \text{with}$$

$$f(X(t_{last})|T=t) \sim N(\mu_{t-t_{last}}, \sigma^2_{t-t_{last}})$$

Not identical with forward prediction, but close (study going on)

CML-trial

- 190 patients
- variables at baseline
  - age
  - sokal (scoring system for overall prognosis)
- WBC ($10^9/l$) measured during follow-up

  transformed to $\,^{10}log(WBC)-0.95$
- endpoint survival

Overall Kaplan-Meier

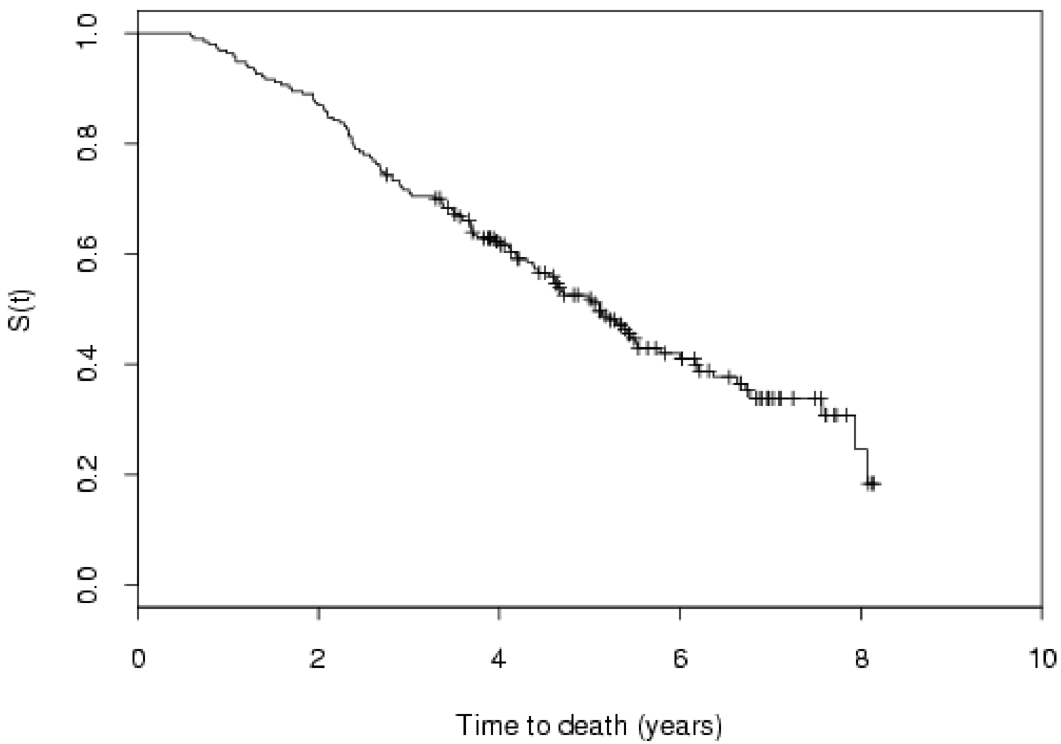


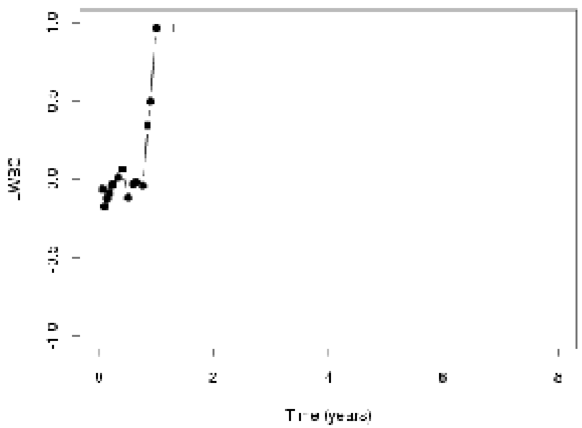
Table 1: Summary of the *WBC* measurements in the data.

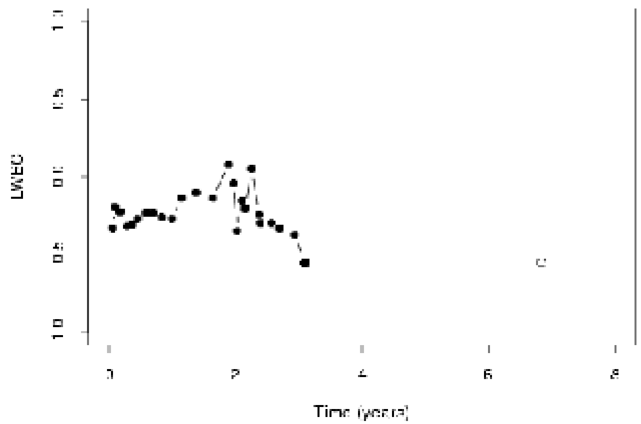
Number of *WBC*-measurements	Number of patients	Percentage (%)
0 – 10	23	12.1
11 – 15	24	12.6
16 – 20	17	8.9
21 – 30	48	25.3
31 – 40	28	14.7
41 – 50	27	14.2
51 – 60	12	6.3
$\geq$ 60	11	5.8
Total	190	100.0

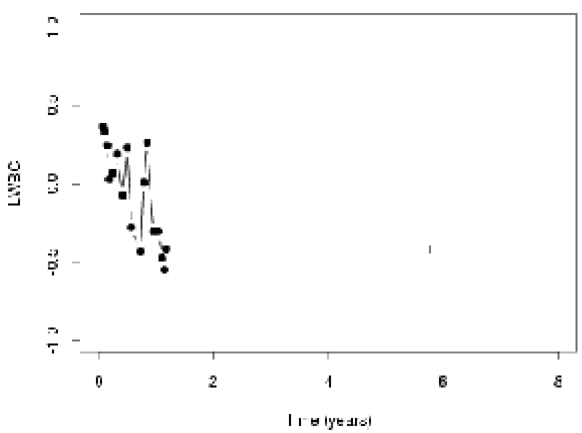
Table 2: Summary of the data.

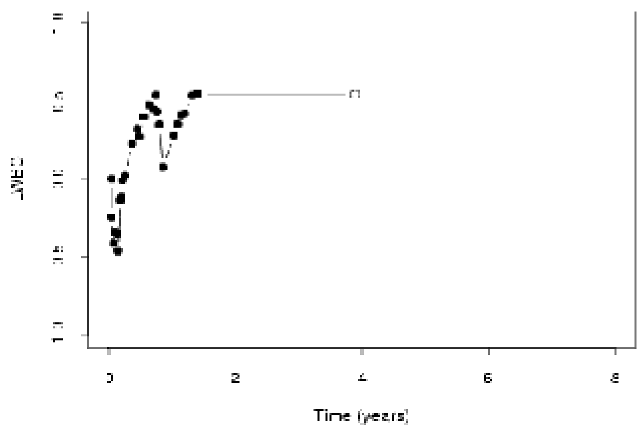
Time (years)	Total patients at start of the year	Mean number of patients in that year	Number of WBC observations per year	Events	Observation rate	Event rate
0-1	190	188.5	2384	7	12.6	0.04
1-2	183	174.2	1179	18	6.8	0.10
2-3	165	149.5	856	30	5.7	0.20
3-4	134	122.5	548	17	4.5	0.14
4-5	102	85.5	336	15	3.9	0.18
5-6	72	53.0	186	13	3.5	0.25
6-7	42	31.4	105	6	3.4	0.19
7-end	17	9.6	40	3	4.2	0.31

Some typical LWBC   patterns









Two time-dependent covariates

- $X_t^{(}$ last observed value of LWBC
- TEL_t time elapsed since last observation

Model , inspired by the simulation study

$$\ln(h(T|X_t^{(},TEL_t,Z)\ '\ \ln(h_0(t))\%Z\text{ß}\%X_t^{(}?(TEL_t)\%\mathrm{d}(TEL_t)$$

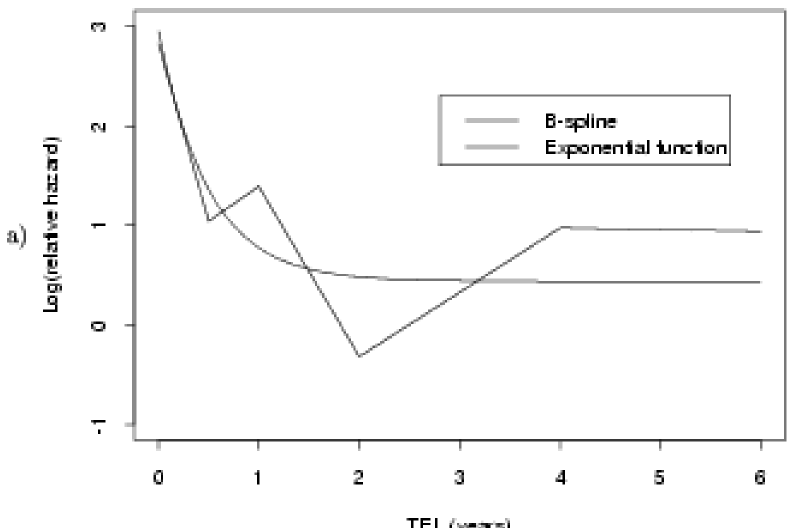
$$?(TEL)\ '\ ?_0\%?_1\exp(\&\mathrm{a}_1 TEL)$$

$$\mathrm{d}(TEL)\ '\ \mathrm{d}\exp(\&\mathrm{a}_2(TEL)$$

Result

Nonparametric vs parametric

? d



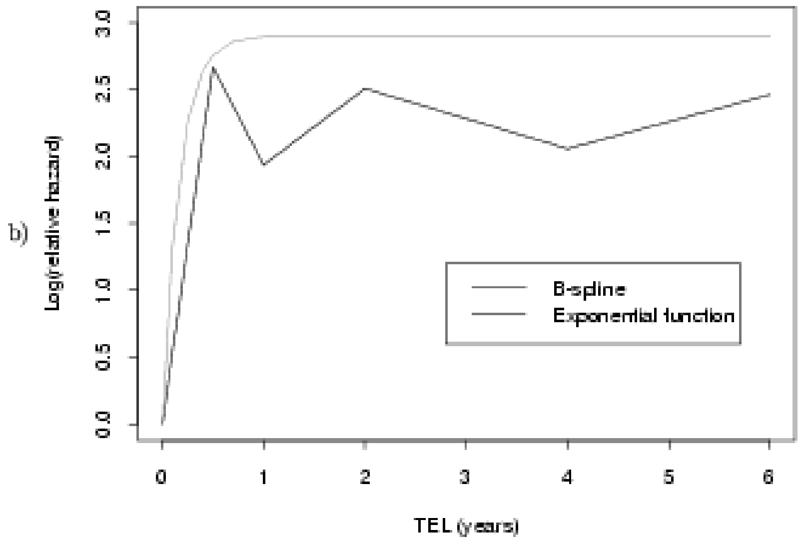


Table 3: Model with TEL and last observed $LWBC$.

	Simple model		Extended model	
Covariate	$\hat{\beta}$	s.e.$(\hat{\beta})$	$\hat{\beta}$	s.e.$(\hat{\beta})$
age	0.018	0.008	0.023	0.008
Sokal score	0.453	0.180	0.345	0.192
$LWBC$	1.534	0.214	0.438	0.325
$LWBCe^{-2TEL}$			2.510	0.636
1-e^{-6TEL}			2.901	0.429
deviance	61.1		123.0	

Refinement

Exponentially weighted linear extrapolation of X as estimate of the current value of X(t), truncated at -1 or +1

b)

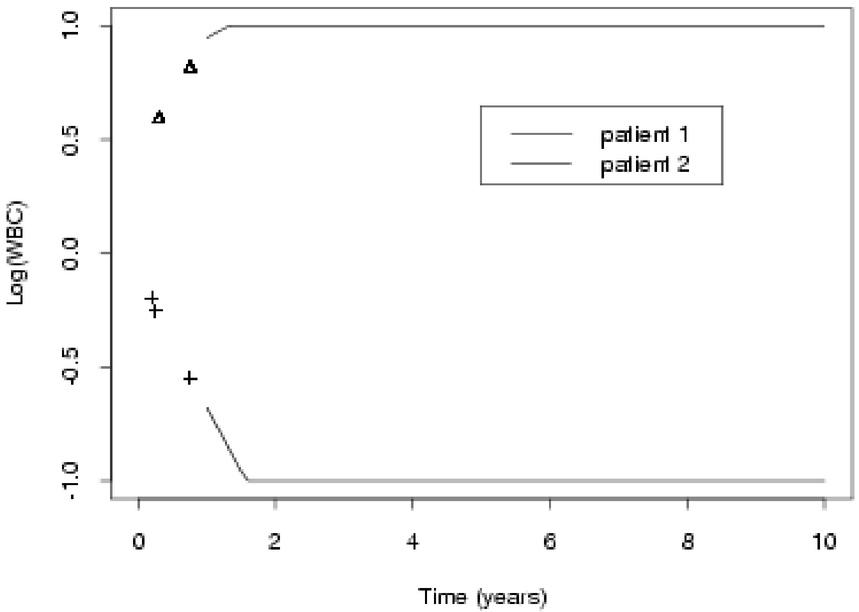


Table 4: Model with Kernel smoother with trend for $LWBC$, with $h = 0.067$.

Covariate	Simple model $\hat{\beta}$	s.e.$(\hat{\beta})$	Extended model $\hat{\beta}$	s.e.$(\hat{\beta})$
age	0.020	0.008	0.023	0.008
Sokal score	0.444	0.185	0.415	0.190
$LWBC_{wls}$	1.059	0.173	0.021	0.179
$LWBC_{wls}e^{-2TEL}$			3.204	0.505
1-e^{-6TEL}			2.988	0.426
deviance	55.3		134.6	

Predicting from the last model, starting at t_0

$$\ln(?(t|Z,X_t^{(},TEL_0)\ '\ \ln(?_0(t))\%Z\text{ß}\%X_t^{(}?(TEL_0\%t\&t_0)\%\mathrm{d}(TEL_0)$$

$X_t^{(}$ prediction of X

TEL_0 age of observation at time of prediction

Choice between $\mathrm{d}(TEL_0)$ and $\mathrm{d}(TEL_0\%t\&t_0)$ depending on interpretation of TEL-effect

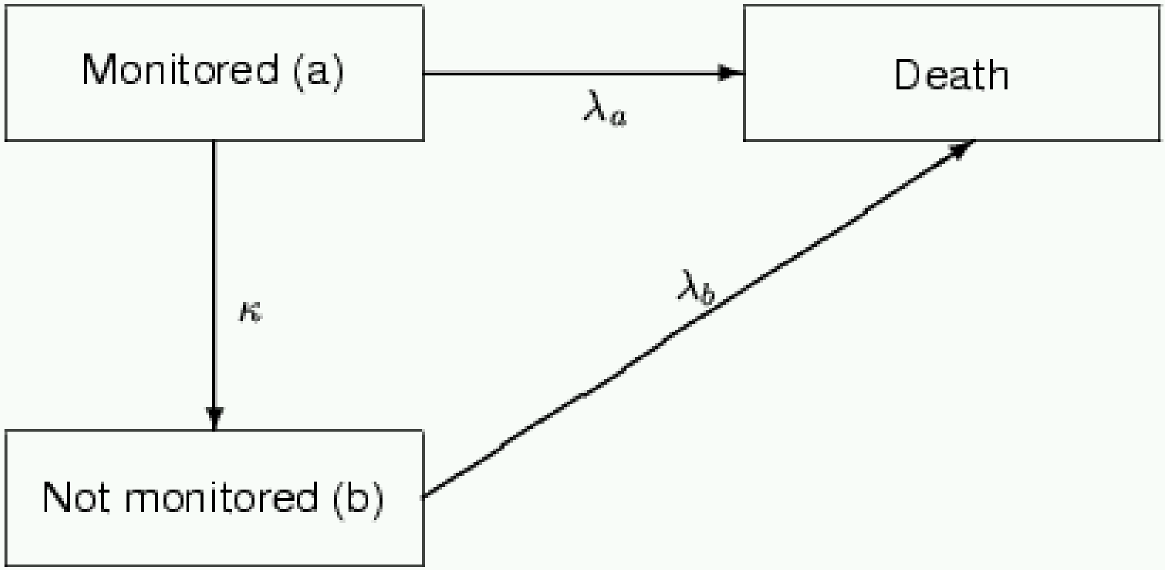


Figure 4: The latent multi-state model with two different states (monitored and not-monitored)with one-way transition intensity κ and corresponding hazard rates λ_a and λ_b.

Predictions using fixed (genuine) TEL-effect lead to

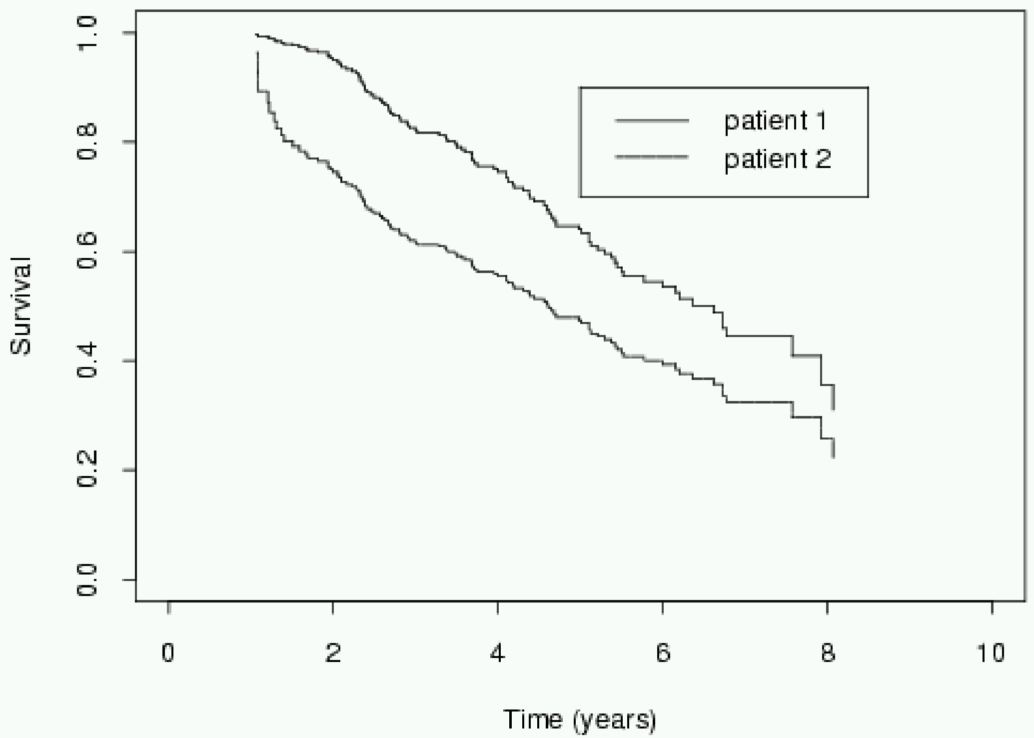


Figure 5: Predictive use of the model, (a) Cumulative baseline hazard; (b) *LWBC* observations and projections for two patients; (c) Predicted survival curves for the same patients.

Predicting by landmarking

Landmark idea:

If you want to predict at some t_0*, use all data available at that landmark point, throw away all time-dependent covariate information after* t_0 *and make a predictive model*.

Different situations

- continuous observations
- irregular observations

Continuous observations

- t_0 landmark point
- $X(t_0)$ last observations
- Z fixed covariate
- s time since t_0 $\quad (t = t_0 + s)$

Model

$$h(t_0 + s | X(T_0), Z) = h_0(s) e^{\beta(s) X(t_0) + \gamma Z}$$

$$\beta(s) = \beta_0 + \beta_1 e^{-\alpha s}$$

(everything may depend on t_0)

Application to CML-trial data with $X(t_0)$ last observation before t_0. Results for t_0=1,2,3 and 4 years

Table 1: Parameter estimates and standard deviations using the separate landmark analyses for the landmarks $t_0 = 1, 2, 3, 4$ year.

	Landmark			
Covariate	1	2	3	4
age	0.17 (0.08)	0.09 (0.09)	0.11 (0.11)	0.08 (0.14)
$Sokal\ score$	0.53 (0.18)	0.68 (0.22)	0.21 (0.37)	0.36 (0.43)
$LWBC(t_0)$	0.73 (0.36)	0.75 (0.43)	0.18 (0.58)	0.42 (0.38)
$LWBC(t_0)e^{-\alpha s}$	3.45 (1.42)	1.97 (1.70)	3.72 (1.50)	
α	2.5	3	1	0

Wanted: model for all t_0.

Simple procedure :

- create prediction data-sets for k landmarks on the interval of interest (the interval [1,4] in the example
- stack the data-sets
- fit the model

$$h(t_0+s|X(t_0),Z,t_0) = h_0(s)e^{\beta(s)X(t_0)+\gamma Z+\delta t_0}$$

(or more complicated if you think this is too simple

Table 2: Parameter estimates and bootstrapped s.e.'s of the multiple landmark model for a different number k of equidistant chosen landmarks at [1,4] year. The sum of the bootstrapped variances SBV is the criterion to decide on the number of landmarks.

k	age	$Sokal\ score$	$LWBC(t_0)$	$LWBC(t_0)e^{-2s}$	t_0	SBV
2	0.15 (0.09)	0.51 (0.16)	0.57 (0.31)	1.39 (1.06)	0.12 (0.06)	1.26
3	0.12 (0.09)	0.52 (0.19)	0.72 (0.32)	0.42 (0.80)	0.11 (0.06)	0.79
4	0.12 (0.09)	0.53 (0.18)	0.62 (0.34)	1.73 (0.72)	0.10 (0.06)	0.67
5	0.12 (0.09)	0.54 (0.19)	0.70 (0.31)	1.65 (0.65)	0.10 (0.06)	0.56
6	0.11 (0.08)	0.59 (0.19)	0.78 (0.32)	1.32 (0.62)	0.11 (0.06)	0.53
7	0.12 (0.09)	0.56 (0.20)	0.64 (0.35)	1.31 (0.62)	0.10 (0.06)	0.56

Questions that arise:

- What exactly are you doing ?
- What are the standard errors of your coefficients?
- How many landmarks should you take.

Approaches

- mathematical via
  - pseudo-likelihood
  - sandwich-estimators
- bootstrapping

Bootstrapping can produce

- covariance matrix of estimated coefficients (st. errors)
- cov. matrix of baseline hazard (not done)
- SBV=trace(cov. matrix) as overall criterion for precision

  Sum of Bootstrapped Variances

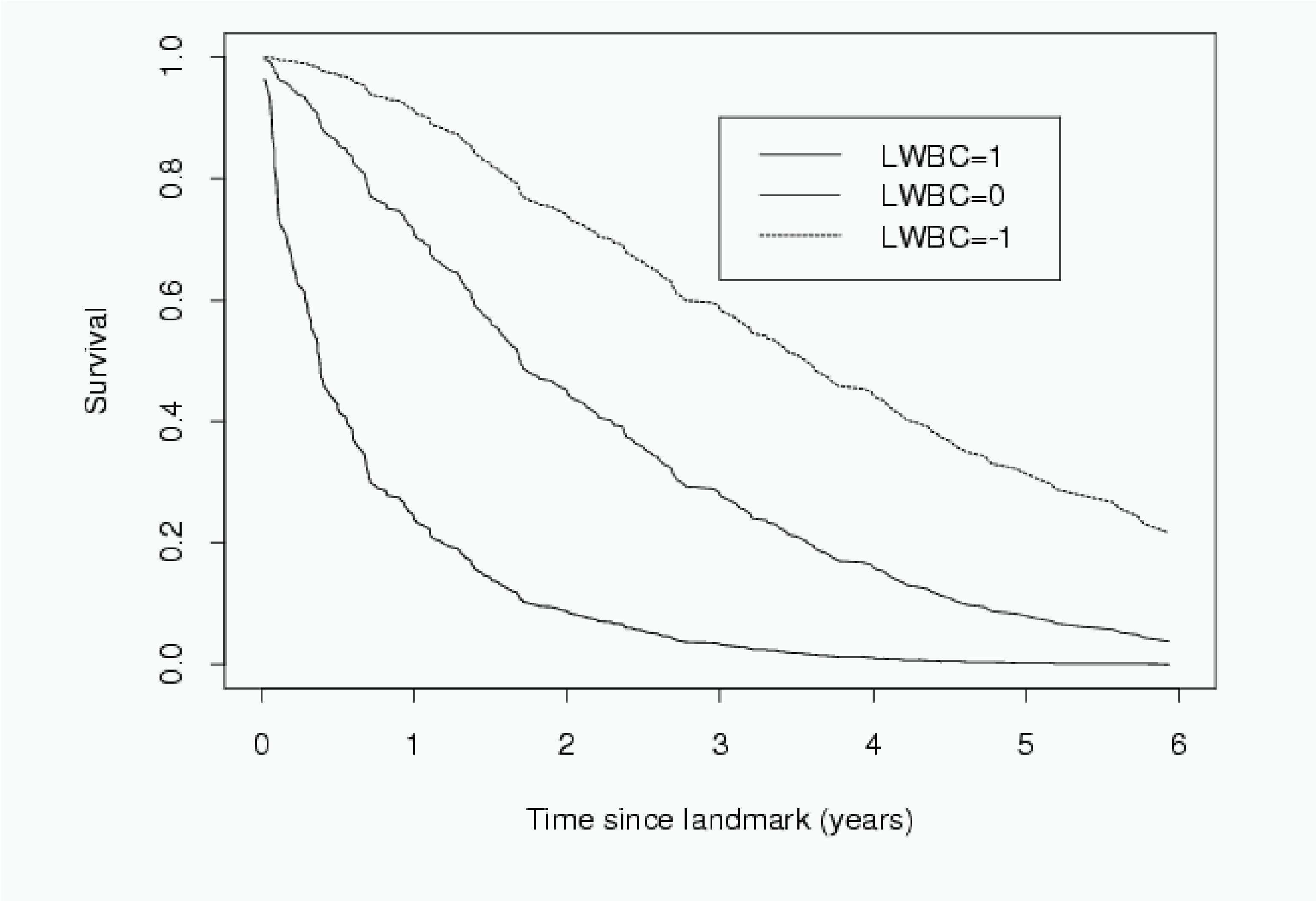
General impression:

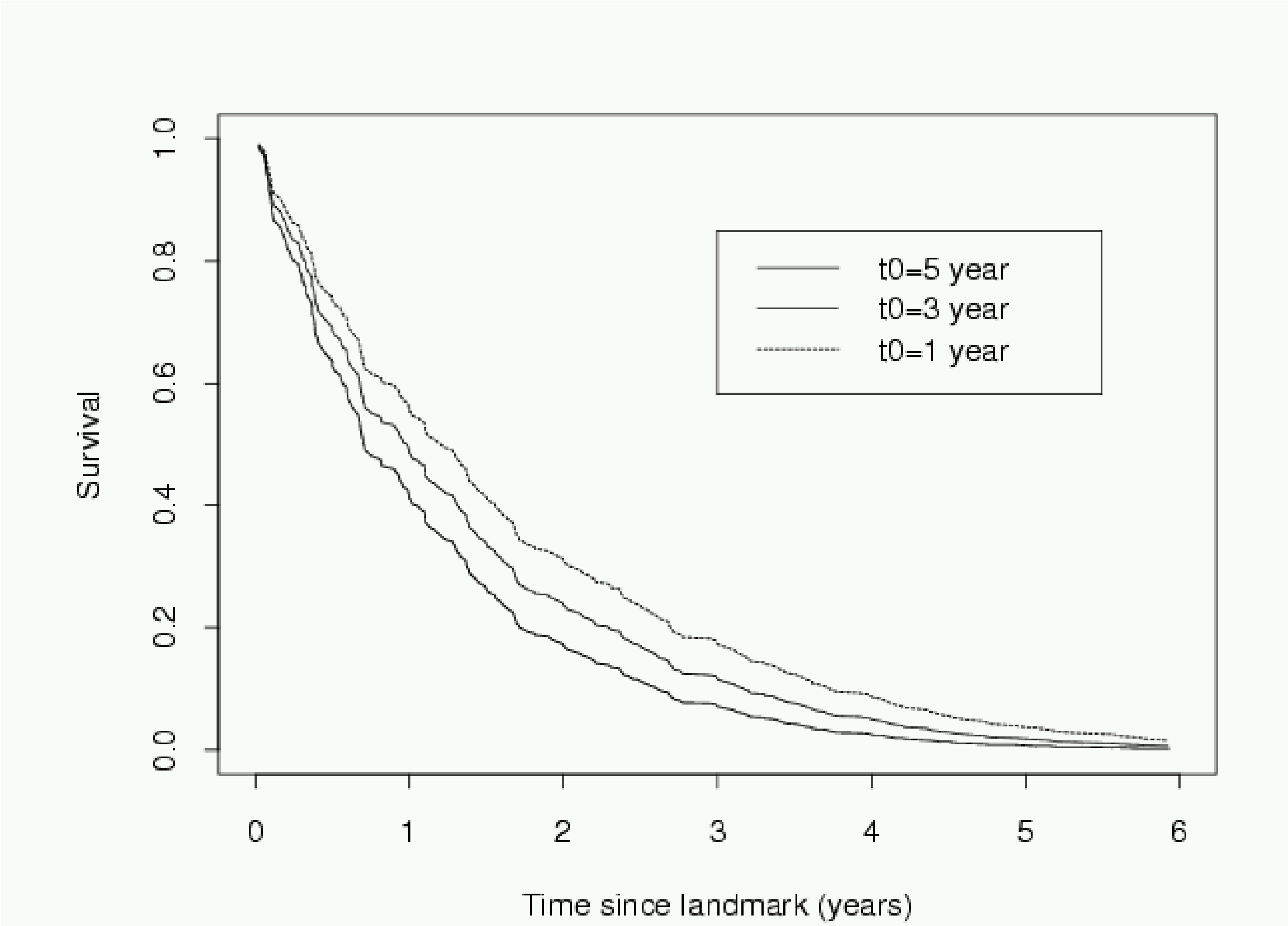
- about five landmarks suffice

Prediction is straightforward from the model, provided we have an estimate of $h_0(s)$. Rather straight forward in stacked data-set.

Consider patient(s) with

- age=60
- sokal=0.84
- t_0=2.5 and $LWBC(t_0)$=-1,0,1
- t_0=1,3,5 and $LWBC(t_0)$=0.5


1.0
0.8
0.6
0.4
0.2
0.0
Survival
LWBC=1
LWBC=0
LWBC=-1
0
1
2
3
4
5
6
Time since landmark (years)


Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
Time since landmark (years)
t0=5 year
t0=3 year
t0=1 year

Irregular observations

Last observation $X(t_0)$ at $t = t_0 - TEL$.

Prediction model

$$h(t_0+s|X(t_0),Z,t_0,TEL) = h_0(s)e^{\beta(TEL,s)X(t_0)+\gamma Z+\delta t_0+\gamma g(TEL,s)}$$

$$\beta(TEL,s) = \beta(TEL+s) = \beta_0+\beta_1 e^{-\alpha(TEL+s)}$$

$$g_1(TEL,s) = (1-e^{-\varphi\, TEL}) \quad \text{or} \quad g_2(TEL,s) = (1-e^{-\tau(s+TEL)})$$

Table 3: Results for the SBV criterion and the bootstrapped deviance. These results are given for the extended prediction model with both a fixed and time-dependent $g(TEL, s)$.

k	SBV		$BDEV$	
	TEL fixed	TEL time-dependent	TEL fixed	TEL time-dependent
2	3.66	5.13	98.0	91.4
3	2.59	3.41	79.8	81.0
4	1.52	1.98	122.8	118.6
5	1.58	2.11	99.0	97.4
6	1.47	1.87	156.6	139.0
7	1.44	1.80	106	103.4

Models compared on the basis of

bootstrapped deviance $BDEV = \hat{\beta}\Sigma_{Boot}(\hat{\beta})^{-1}\hat{\beta}$

Fixed model g_1 seems to fit slightly better.

Final model

age	$Sokal\ score$	$LWBC(t_0)$	$LWBC(t_0)$ $e^{-2(s+TEL)}$	$1-e^{-TEL}$	t_0
0.14 (0.09)	0.52 (0.20)	0.58 (0.34)	3.74 (1.10)	1.11 (0.39)	0.01 (0.07)

Making predictive model for an irregularly observed time-dependent covariate is much harder than for a continuous one.

Interesting question: **coherence?**

- last observation at t_{obs}
- prediction wanted at t_{pred}
- could use any landmark $t_{obs}\#t_0\#t_{pred}$

Would that matter? Yes, but not very much.

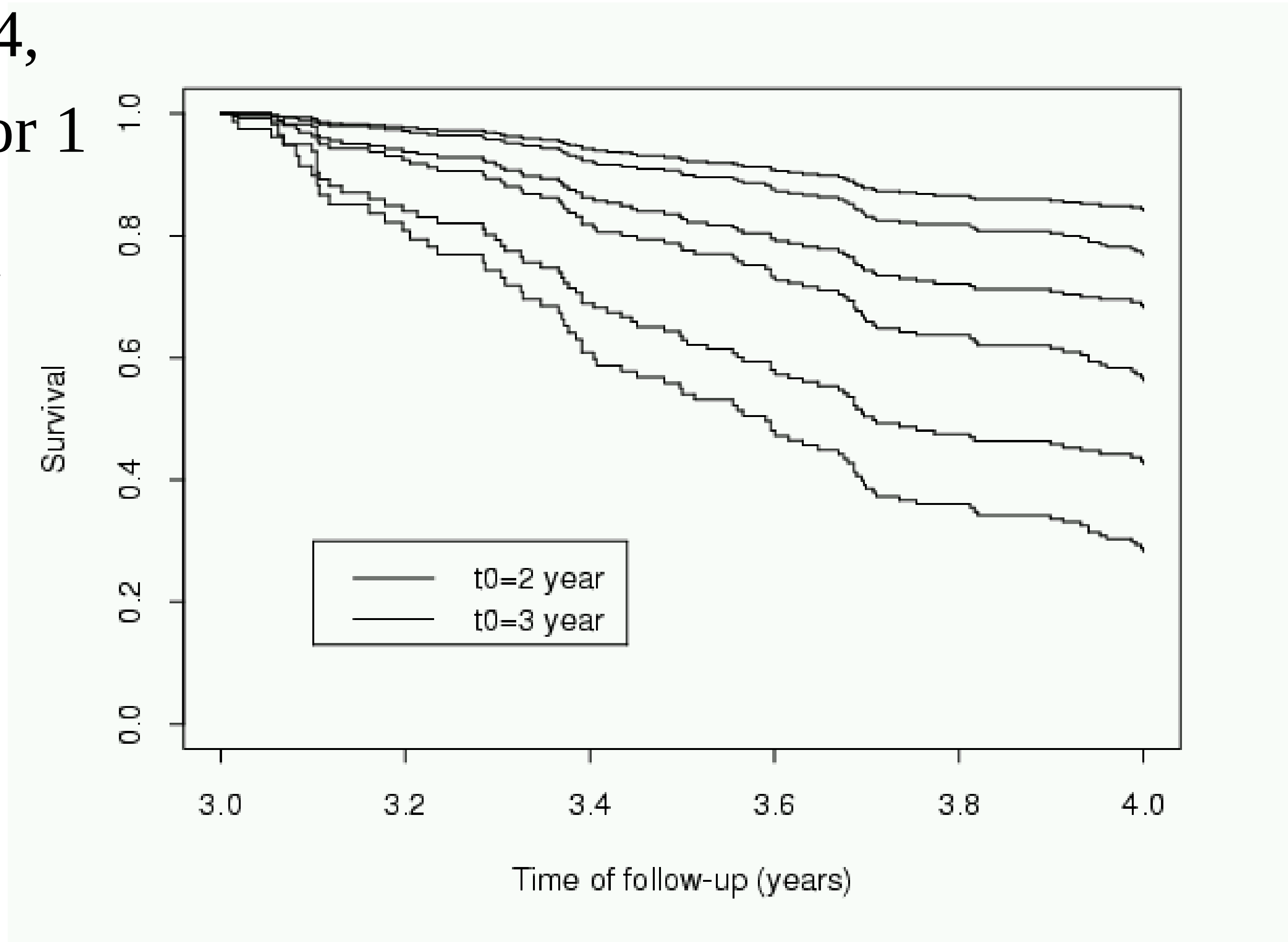
Example:

age=60 Sokal=0.84,

$LWBC(t_{obs})$=-1, 0 or 1

t_{obs} ' 2 , t_{pred} ' 3 and

t_0=2 or 3



Retrospective modeling

- T survival time
- X time-dependent covariate (process)
- Z fixed covariate

Prospective model

$$f(X,T|Z) = f(X|Z)f(T|X,Z)$$

Retrospective modeling (pattern-mixture à la Hogan & Laird)

$$f(X,T|Z) = f(T|Z)f(X|T,Z)$$

Type of models

- $f(T|Z)$ Cox-model or parametric survival model
- $f(X|T,Z)$ Generalized Linear Mixed Model

Censored observations handled by either

- using observed T's only in modeling $f(X|T,Z)$ (HL)
- multiple imputation (discussed here) (PMDA)
- *full ML (combined with EM) (not discussed)*

HL Hogan & Laird

PMDA Poor man's Data Augmentation

Prediction of survival after t_0 using **all relevant information** before t_0 by means of Bayes Theorem

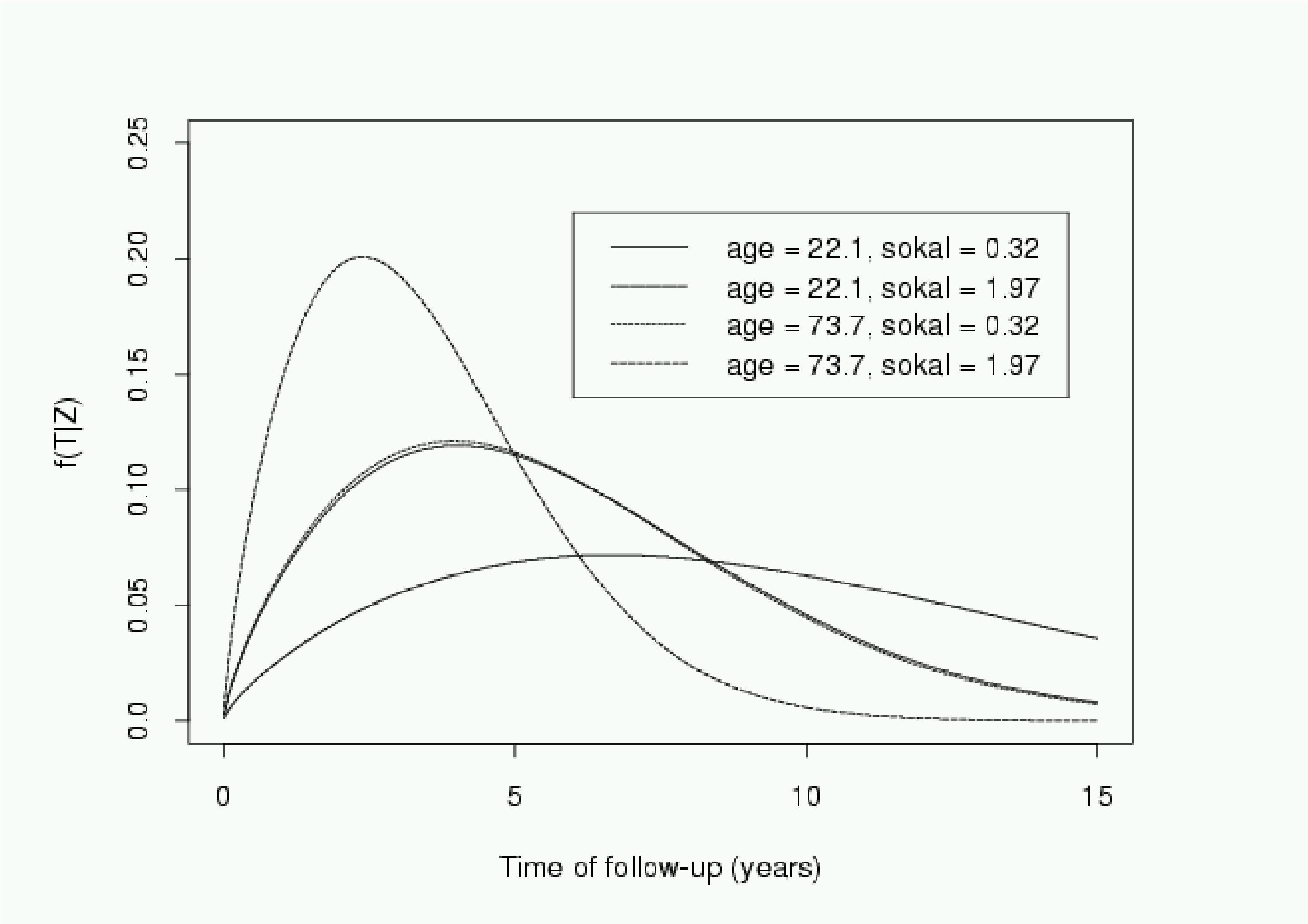
$$f(T|X(..t_0),Z,T \geq t_0) = \frac{f(X(..t_0)|T,Z)f(T)}{\int_{t_0} f(X(..t_0)|T,Z)f(T)dT}$$

$X(..t_0)$ up to t_0

Application to example

Table 1: Loglikelihood of several parametric survival functions for the initial estimation of $f(T|Z)$, including the fixed covariates age and $Sokal\ score$.

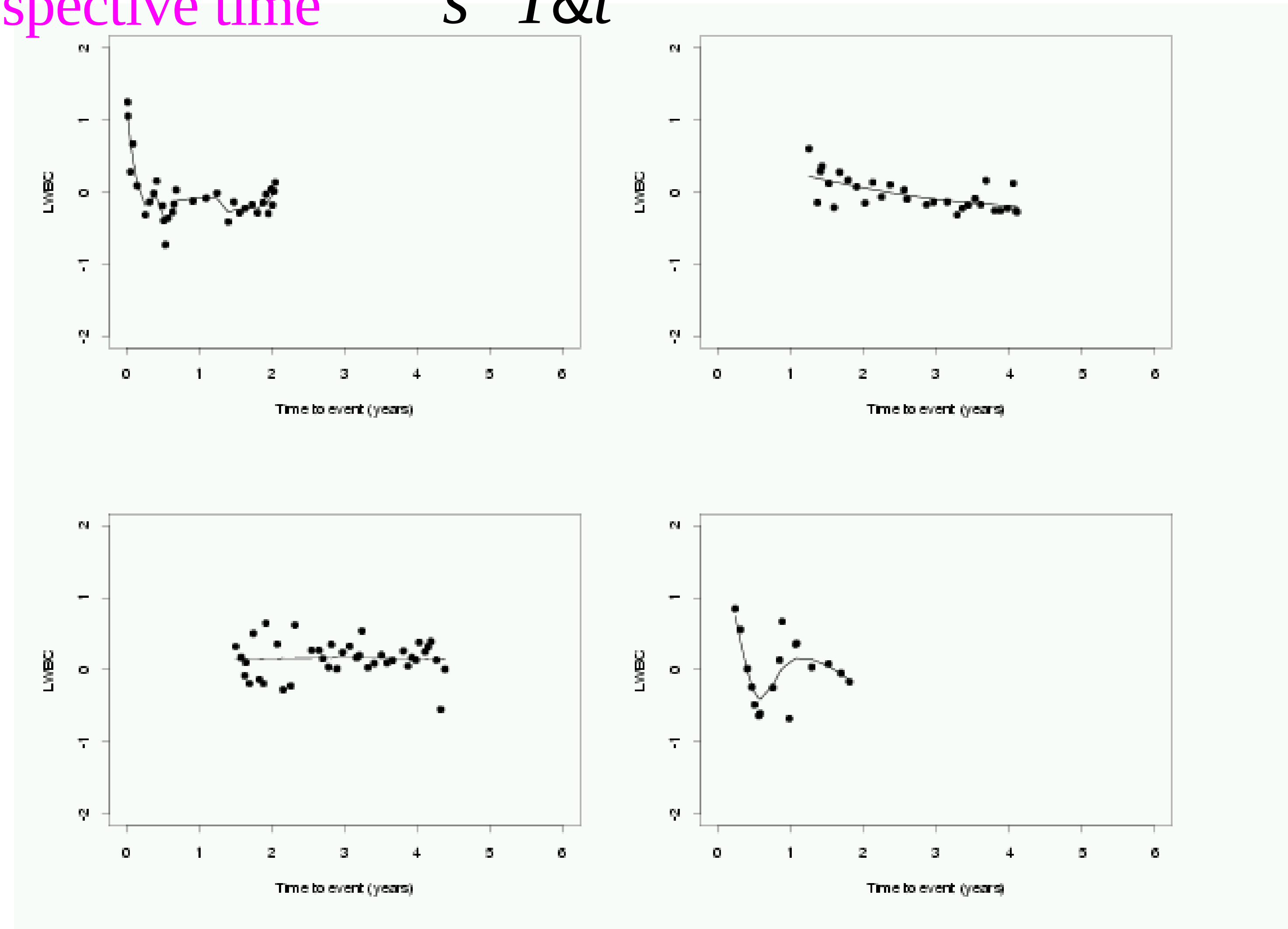
Parametric survival function	loglikelihood
Exponential	-280.0
Weibull	-276.3
Gaussian	-296.8
Logistic	-300.8
Log logistic	-278.4
Log normal	-281.9



0.25
0.20
0.15
0.10
0.05
0.0
f(T|Z)
age = 22.1, sokal = 0.32
age = 22.1, sokal = 1.97
age = 73.7, sokal = 0.32
age = 73.7, sokal = 1.97
0
5
10
15
Time of follow-up (years)

Model for covariate process given survival time formulated in

retrospective time *s ' T&t*



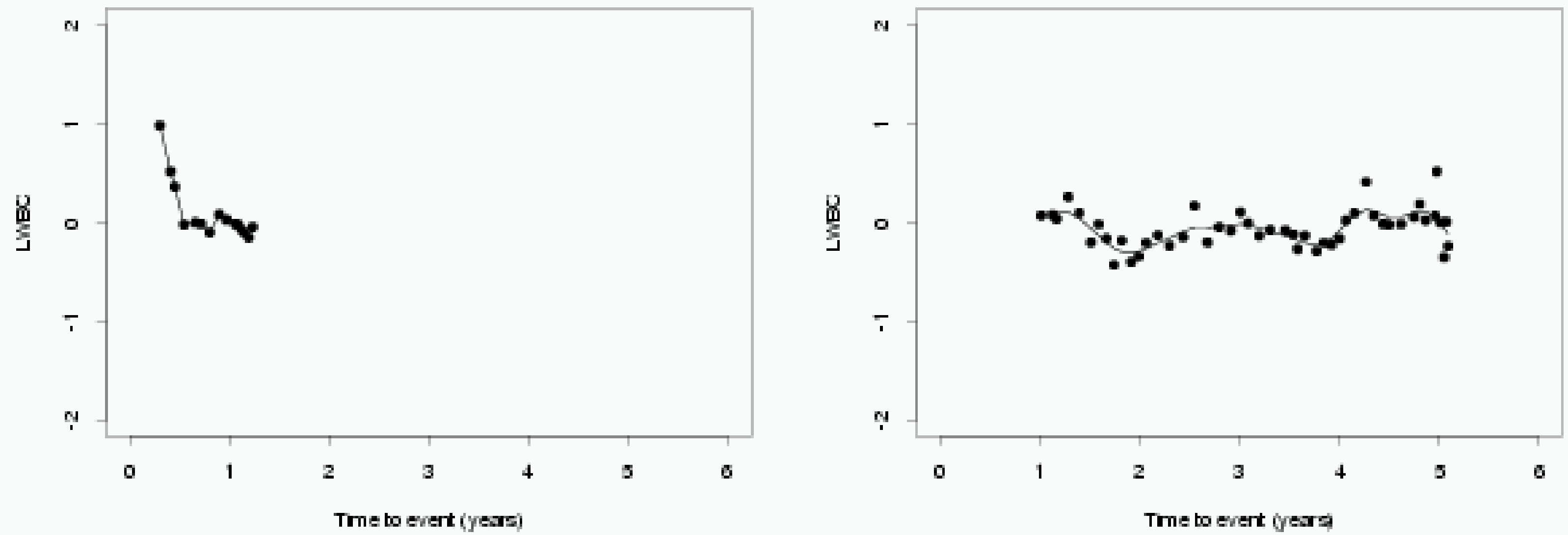


Figure 2: Smoothing splines for the $LWBC$ courses in retrospective time for 6 randomly drawn patients with an event.

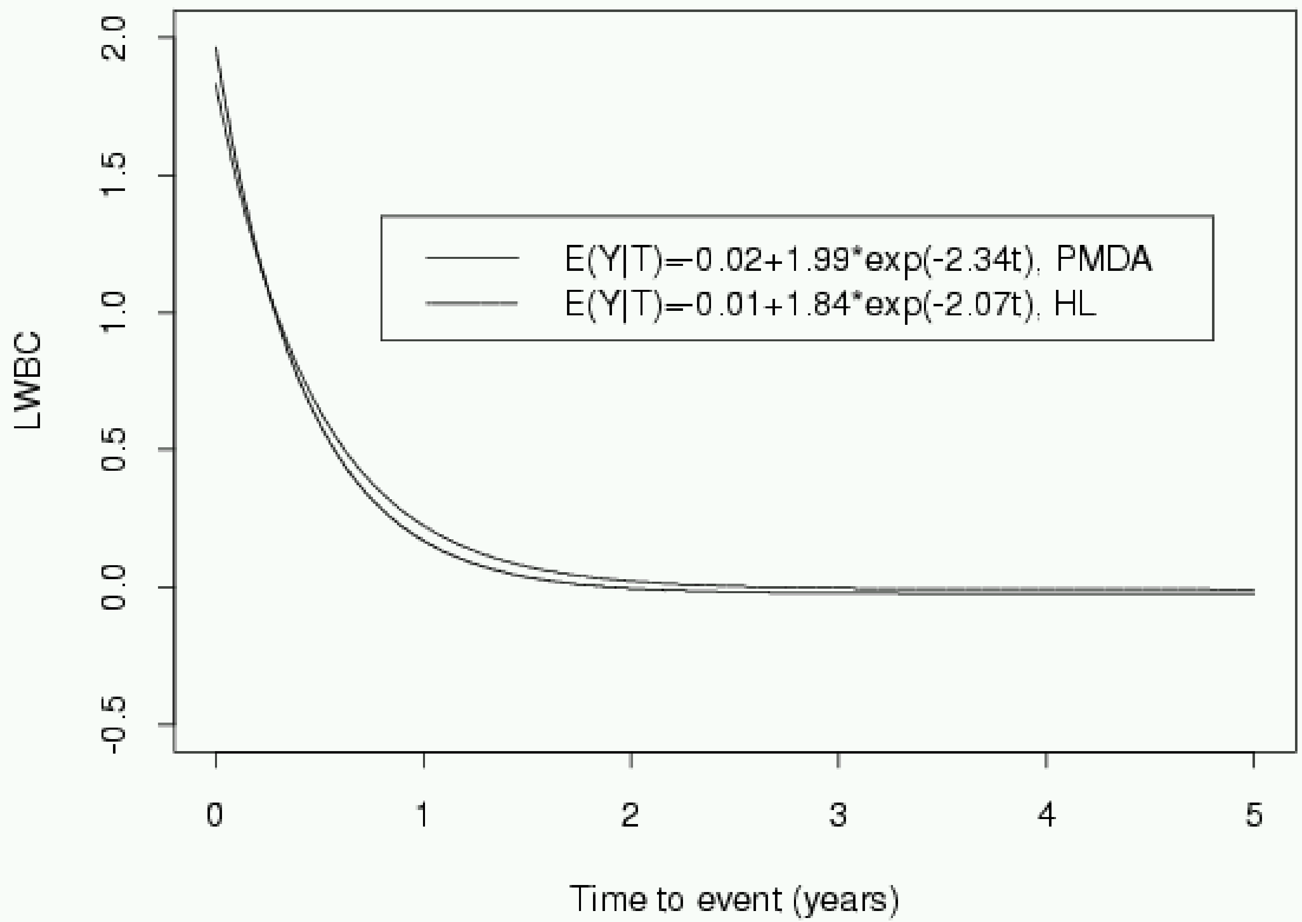
Favorite model

$$X_i(s)\ \text{'}\ ß_0 \% b_{0i} \% (ß_1 \% b_{1i}) e^{\&?(s} \% e_i(s)$$

- fixed parameters $ß_0$, $ß_1$ and ?
- $(b_0, b_1)^{)} \sim N(0, S_b)$
- $e(t)$ autoregressive , mean 0, variance s^2, autocorrelation ?^{t}

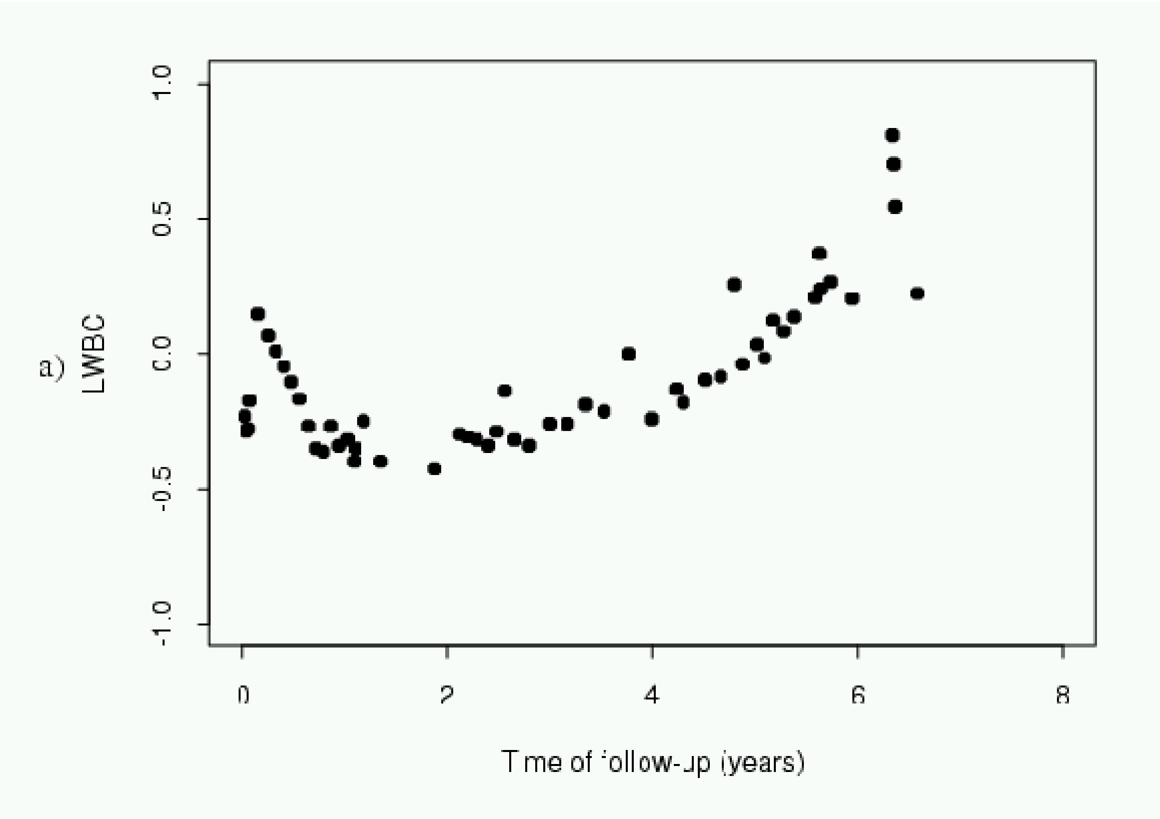
Table 2: Results of the parameter estimates for $f(Y|T)$ and $f(T|Z)$. The s.e.'s are given between brackets. In the third column, the results using the procedure of Hogan and Laird (HL) are given, whereas in the fourth column, the results using the Poor Man's Data Augmentation algorithm (PMDA) are given.

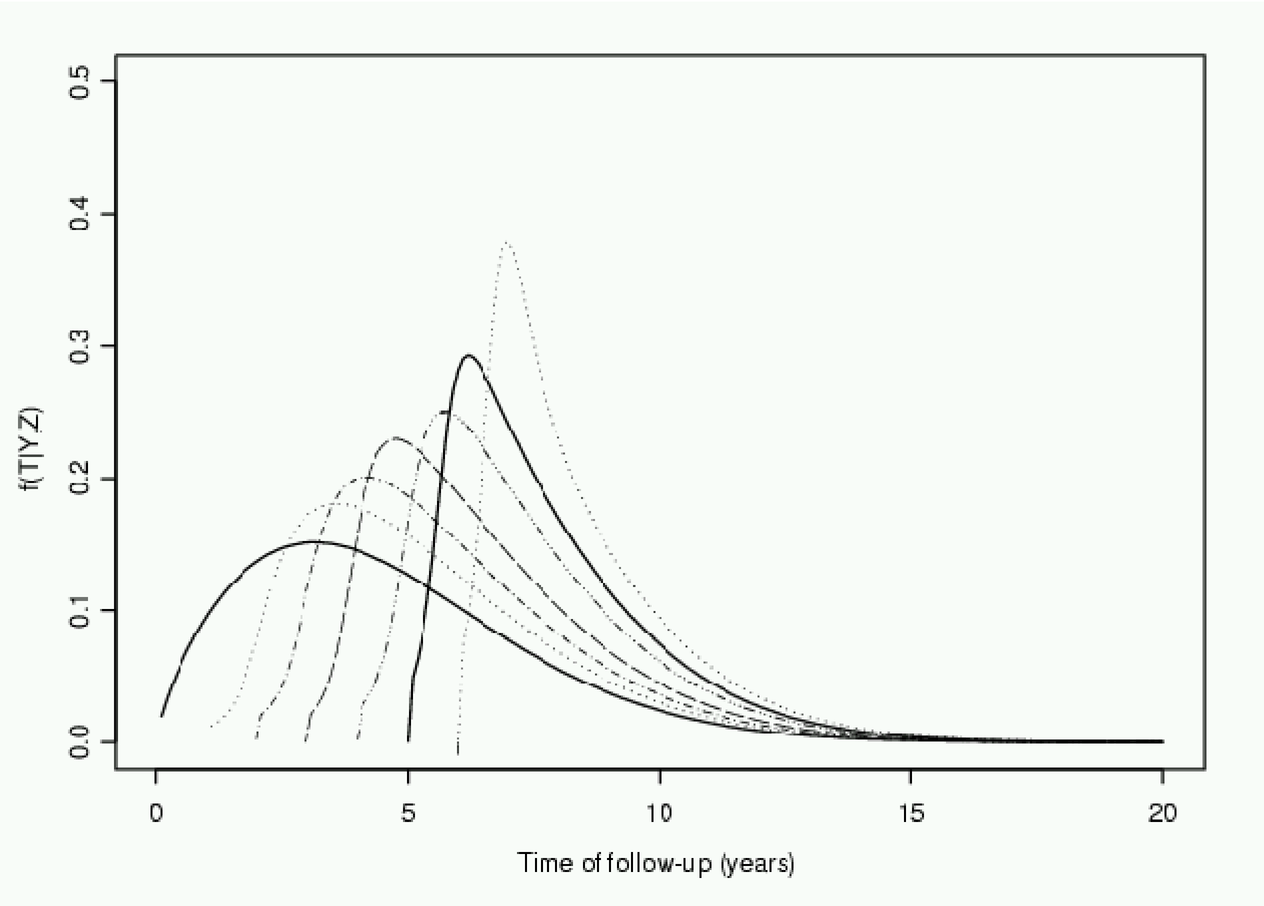
density	parameter	HL	PMDA
$f(Y\|T)$	α_0	-0.007 (0.020)	-0.023 (0.017)
	α_1	1.837 (0.227)	1.986 (0.303)
	γ	2.073 (0.146)	2.335 (0.356)
	ρ	0.316	0.345
	σ^2	0.091	0.080
	Σ_{11}	0.027	0.030
	Σ_{12}	-0.031	-0.023
	Σ_{22}	1.384	1.448
$f(T\|Z)$	*intercept*	2.54 (0.20)	2.50 (0.20)
	age	-0.01 (0.01)	-0.01 (0.01)
	Sokal score	-0.31 (0.10)	-0.33 (0.09)
	scale	0.58 (0.08)	0.57 (0.07)



LWBC
2.0
1.5
1.0
0.5
0.0
-0.5
E(Y|T)=-0.02+1.99*exp(-2.34t), PMDA
E(Y|T)=-0.01+1.84*exp(-2.07t), HL
0
1
2
3
4
5
Time to event (years)

Prediction example: data of some patient that dies after nearly 7 years




0.5
0.4
0.3
0.2
0.1
0.0
f(T|Y,Z)
0
5
10
15
20
Time of follow-up (years)