Session 2

Non-proportional hazards models and model comparisons

Content

- Survival as Generalized Linear Model

(Compstat, 2000)

- Frailty and Cure models
- Model comparison

(IBC,2000)

Some notation

X: covariates

T: survival time

survivor function

$$S(t|X) = P(T>t|X)$$

density

$$f(t|X) = \frac{d(1-S(t|X))}{dt}$$

hazard (force of mortality)

$$h(t|X)dt = P(T<t+dt|T\geq t,X) = -\frac{d\ln(S(t|X))}{dt}$$

cumulative hazard

$$H(t|X) = \int_0^t h(s|X)ds = -\ln(S(t|X))$$

links:

$$S(t|X) = \exp(-H(t|X))$$

$$f(t|X) = S(t|X)h(t|X)$$

Right censoring

- censoring time C, due to
  - end of follow-up
  - competing risk
  - lost-to-follow-up
- actually observed (and in data base)

  $T \leftarrow \min(T,C)$ (ignore *)

  $d \leftarrow [T \# C]$ (indicator of event)
- likelihood of one individual's observation

  $L \leftarrow f(T|X)^{d} S(T|X)^{1-d} \leftarrow S(T|X) h(t|X)^{d}$

Example. Advanced ovarian cancer (358 patients)

(van Houwelingen et al, 1989)

DIAM (size residual tumour after surgery)

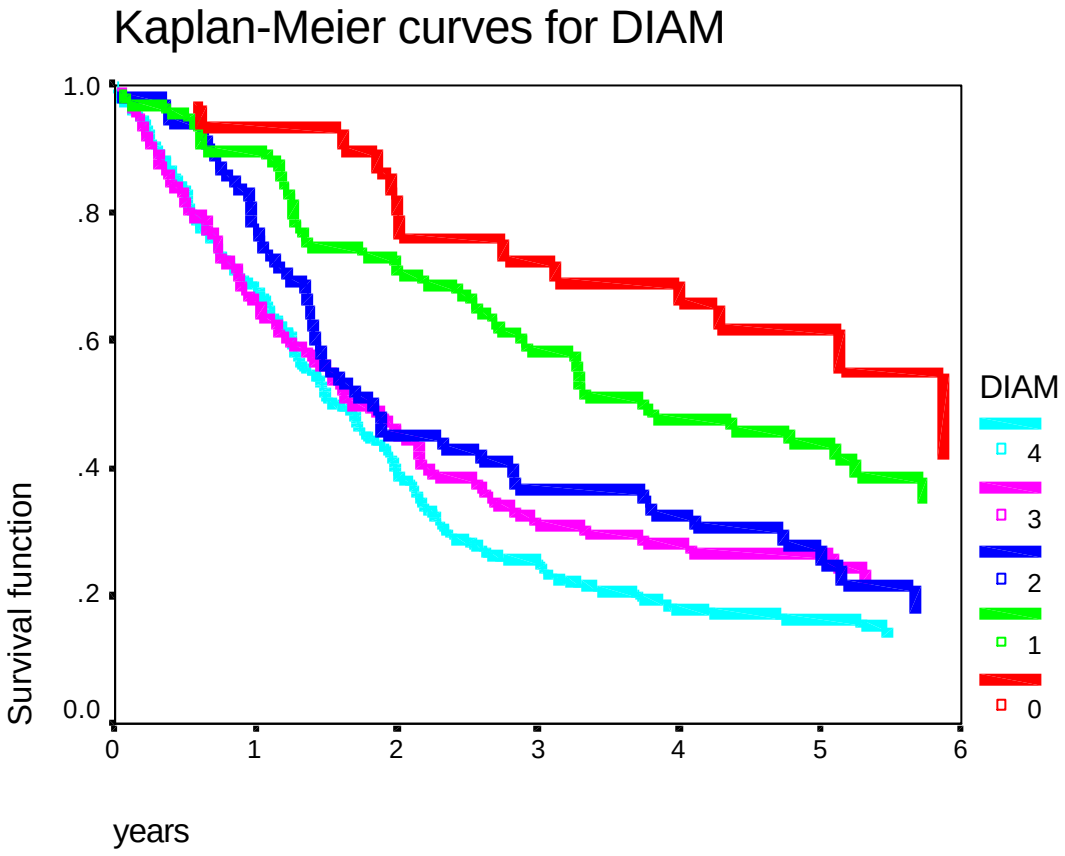
```
0  microscopic   29 ( 8.1%)
1  < 1cm         67 (18.7%)
2  1-2 cm        49 (13.7%)
3  2-5 cm        68 (19.0%)
4  > 5cm        145 (50.5%)
```

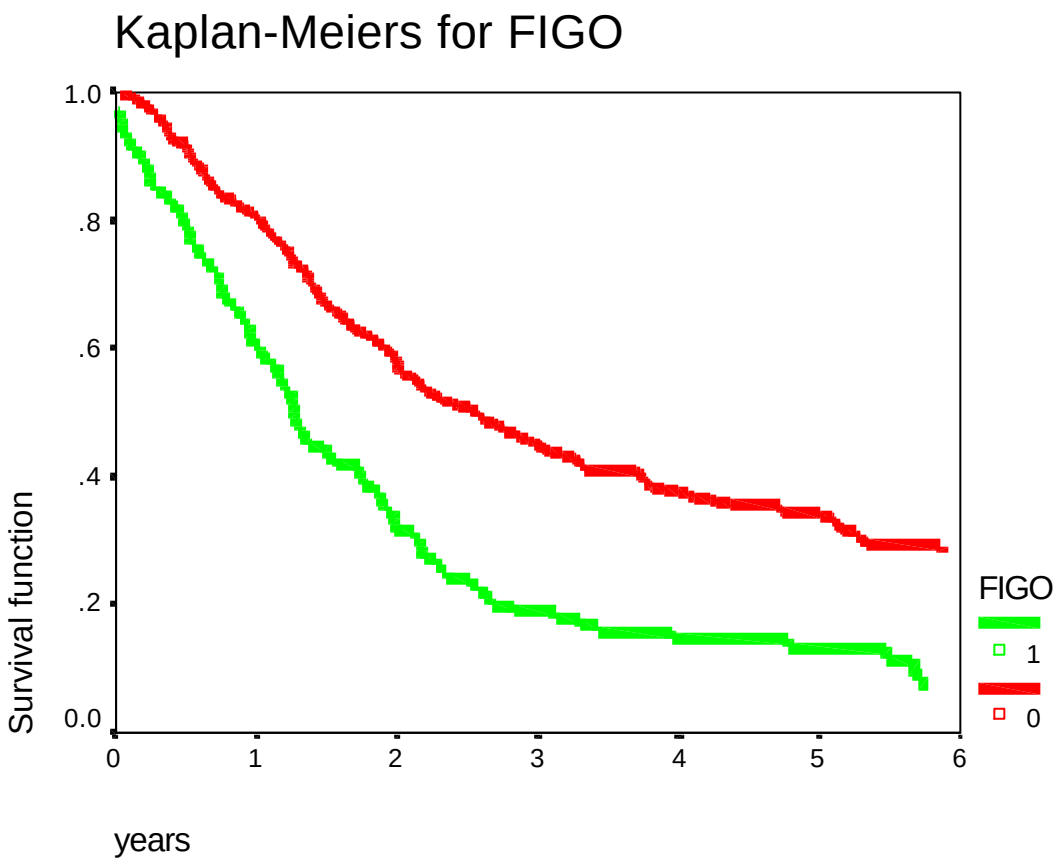
FIGO (stage of cancer)

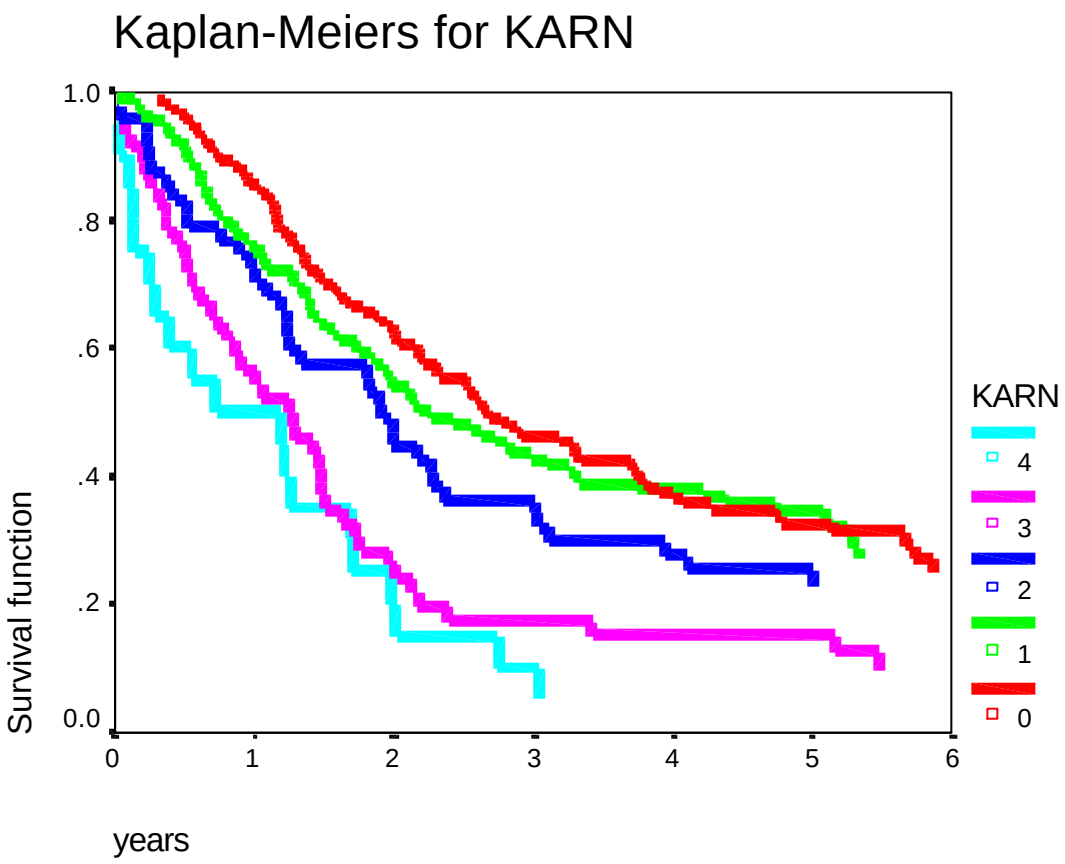
```
0  FIGO=III     262 (73.2%)
1  FIGO=IV       96 (26.8%)
```

Karnofsky (Performance status at t=0)

```
0  100% funct.  137 (38.3%)
1   90%         108 (30.2%)
2   80%          47 (13.1%)
3   70%          46 (12.8%)
4  #60%          20 ( 5.6%)
```


Kaplan-Meier curves for DIAM
Survival function
1.0
.8
.6
.4
.2
0.0
0
1
2
3
4
5
6
years
DIAM
4
3
2
1
0


Kaplan-Meiers for FIGO
Survival function
1.0
.8
.6
.4
.2
0.0
0
1
2
3
4
5
6
years
FIGO
1
0


Kaplan-Meiers for KARN
Survival function
1.0
.8
.6
.4
.2
0.0
0
1
2
3
4
5
6
years
KARN
4
3
2
1
0

Accelerated Failure Time Model

Additive model on ln-scale

< $\ln(T) = X\beta + \sigma e$

< e has cdf F

< $S(t|X) = 1 - F((\ln(t) - X\beta)/\sigma)$

Simplest model: normal

cdf $F(e) = \Phi(e)$

Most popular model: Weibull model

cdf $F(e) = 1 - \exp(-\exp(e))$

Weibull model for the example

ß's	constant	2.1733
	DIAM	-0.2123
	FIGO	-0.5771
	KARN	-0.1743
s		1.0045

$\exp(X\beta)$ models the moment where $S(t) = \exp(-1) = 0.37$

Cox Proportional Hazards model

$h(t|X) = h_0(t)\exp(X\beta)$ (no constant)

Let $H_0(t) = \int_0^t h_0(s)ds$, then

$S(t|X) = S_0(t)^{RR(x)}$ with

$S_0(t) = \exp(-H_0(t))$

$RR(x) = \exp(X\beta)$

Fitting goes in two-stages

- Estimating ß by Maximizing

  Partial Likelihood PL, defined as

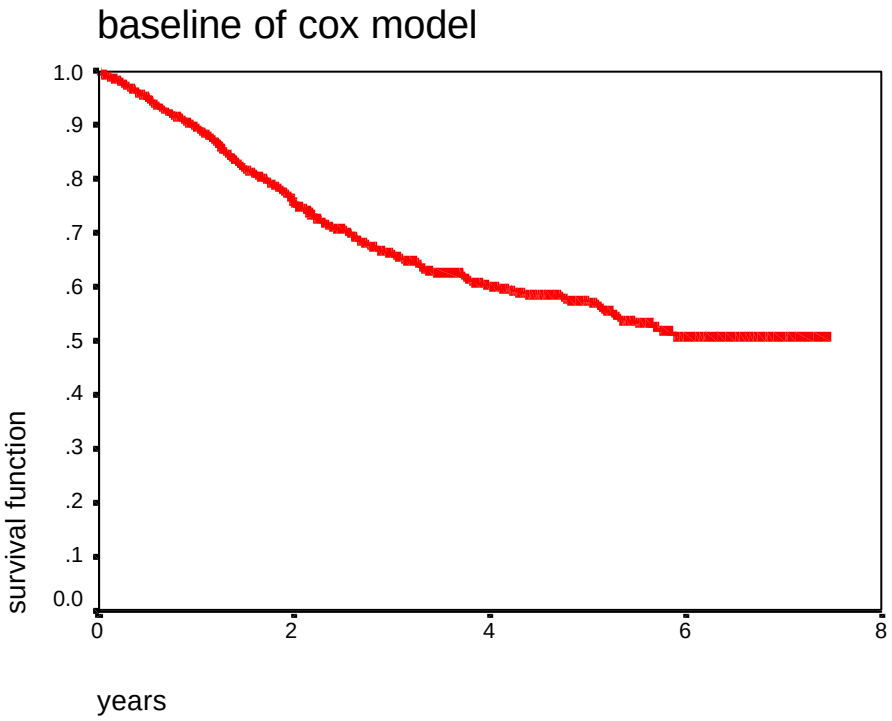
$$PL(ß) = \prod_{d_i = 1} \frac{\exp(X_i ß)}{\sum_{t_j \geq t_i} \exp(X_j ß)}$$

$\{j | t_j \geq t_i\}$ is called "risk set"

- Estimating the baseline cum. hazard $\hat{H}_0(t) = \sum_{t_i \leq t} \frac{d_i}{\sum_{t_j \geq t_i} \exp(X_j ß)}$

Continued example

ß	DIAM	0.2002
	FIGO	0.5419
	KARN	0.1715



Drawbacks of Cox model

- special mathematics required (counting processes, martingales etc.)
- baseline not easily available and very jumpy
- check for non-PH and extensions to non-PH models get complicated

Advantage

- generally accepted
- clinicians love it !

Survival as GLM (in the spirit of Kooperberg, et al., JASA, 1995)

- make m short intervals of length ?
- t_j ' (j&1)? , starting point of j-th interval
- define
  - **S** μ_{ij} ' $h(t_j|X_i)$? and ?$_{ij}$ ' ln(μ_{ij})
  - **S** Y_{ij} ' 1 if person i at risk in interval j
  - **S** d_{ij} ' 1 if person i dies in interval j
- then, Poisson approximation yields

  log-lik l ' $\sum_i \sum_j Y_{ij}(d_{ij}?_{ij}\&\exp(?_{ij}))$

Easy to implement strategy for small to moderate size data:

- make separate records for each Y_{ij}

  (T_i/? records per individual)

- define d_{ij} to be the outcome
- make model for $\mu_{ij}(t_j, X_i)$ (next sheet)
- aggregate data if possible
- apply Poisson regression

Side step: building blocks

- k covariates lead to
  X: n×k matrix (person×covariate)
- p base functions in time lead to
  Z: m×p matrix (interval×base function)
- T: n×m matrix of all conceivable ?'s

In the example

- X as given, but it could be extended by transformation and interaction
- Z generated by base functions t, $\ln(t)$, t^2

Cox' PH model is just the additive model

$$?_{ij} \sim a + X_i ß + Z_j \quad (\dim = 1+k+p)$$

with X_i and Z_j respective rows of X and Z.

Estimating equations

$$\sum_{i,j} Y_{ij}(d_{ij} - \exp(?_{ij})) \sim 0$$

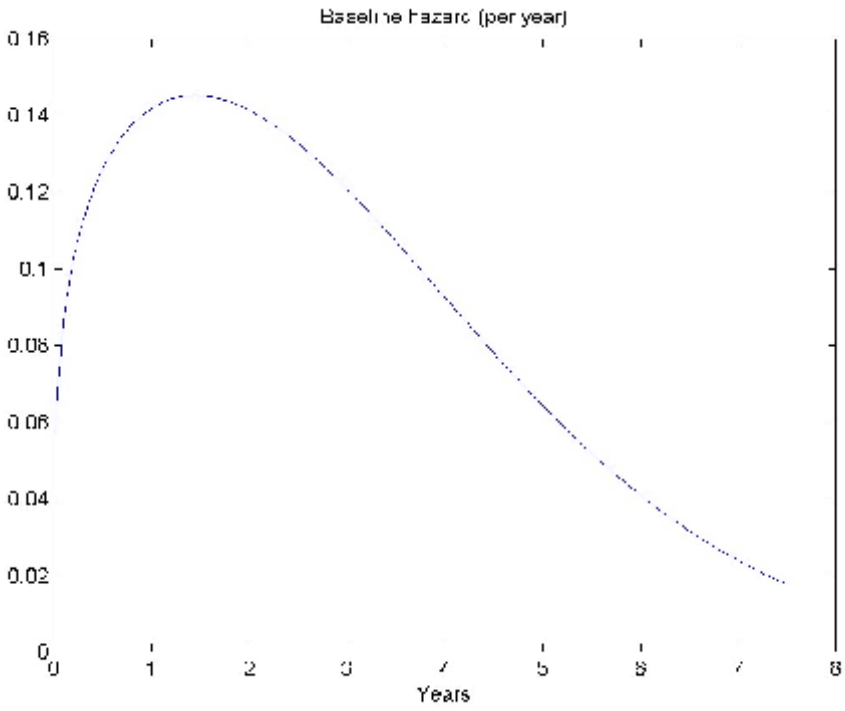
$$\sum_{i,j} Y_{ij}(d_{ij} - \exp(?_{ij}))X_{iu} \sim 0 \quad \text{for } u=1,..,k$$

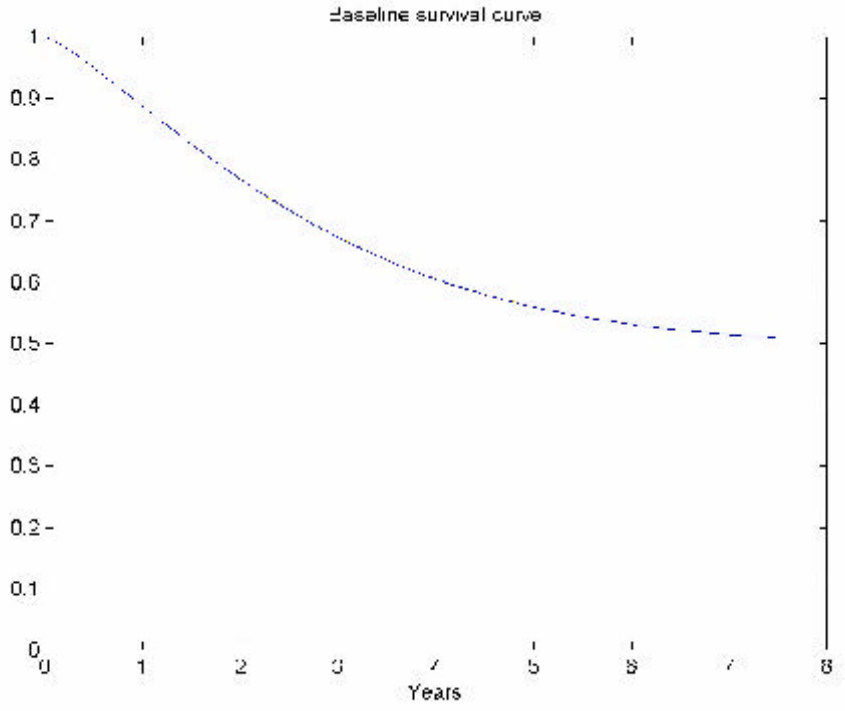
$$\sum_{i,j} Y_{ij}(d_{ij} - \exp(?_{ij}))Z_{jv} \sim 0 \quad \text{for } v=1,..p$$

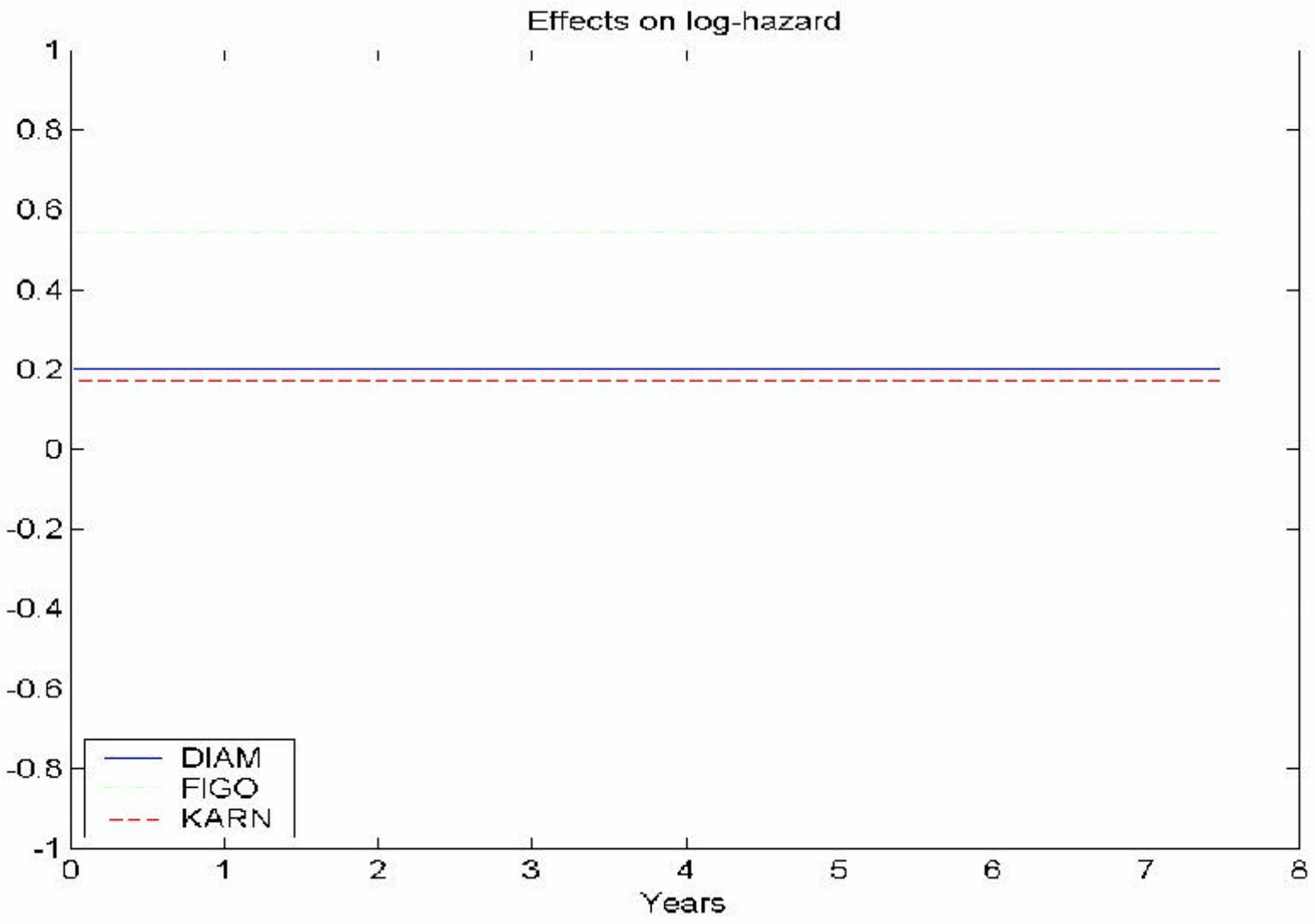
Information matrix $H \sim \sum_{i,j} Y_{ij}\exp(?_{ij})(1,X_i,Z_j)^{)}(1,X_i,Z_j)$

Continued example

- time divided in weeks
- estimated ß's as in Cox regression
- baseline hazard and survival given in graphs






Effects on log-hazard
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
1
2
3
4
5
6
7
8
Years
DIAM
FIGO
KARN

non-PH models

for each covariate time-dependent effect

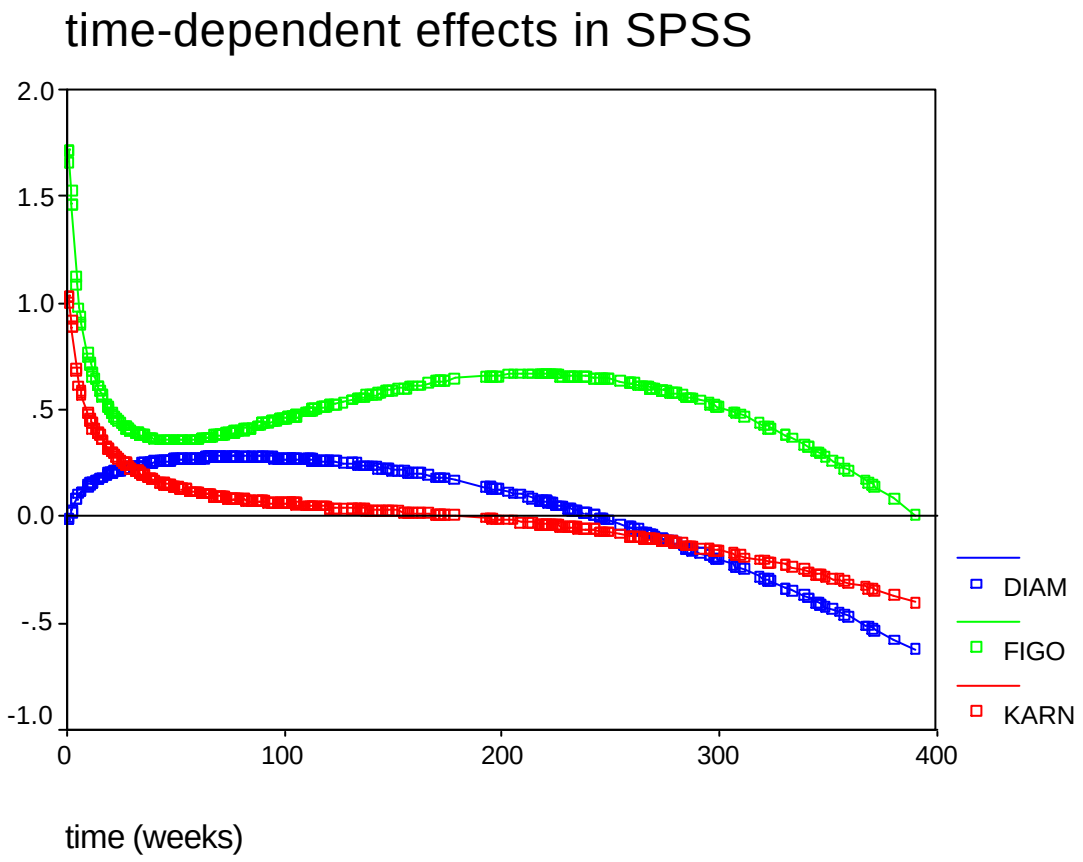
$$ß(t) ' ß_0 \% ß_1 (f_1(t) \% ß_2 (f_2(t) \% ..$$

Can be fitted by introducing pseudo-time-dependent covariates.

Baseline is hard to obtain.

Can be formulated in GLIM-context by tensorproducts

Results from SPSS



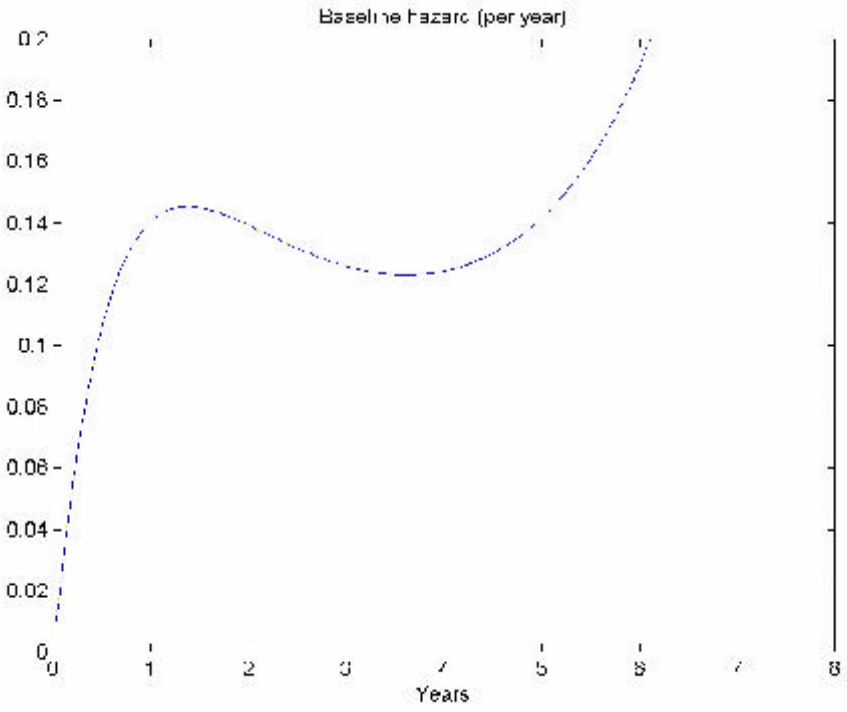
“full” model

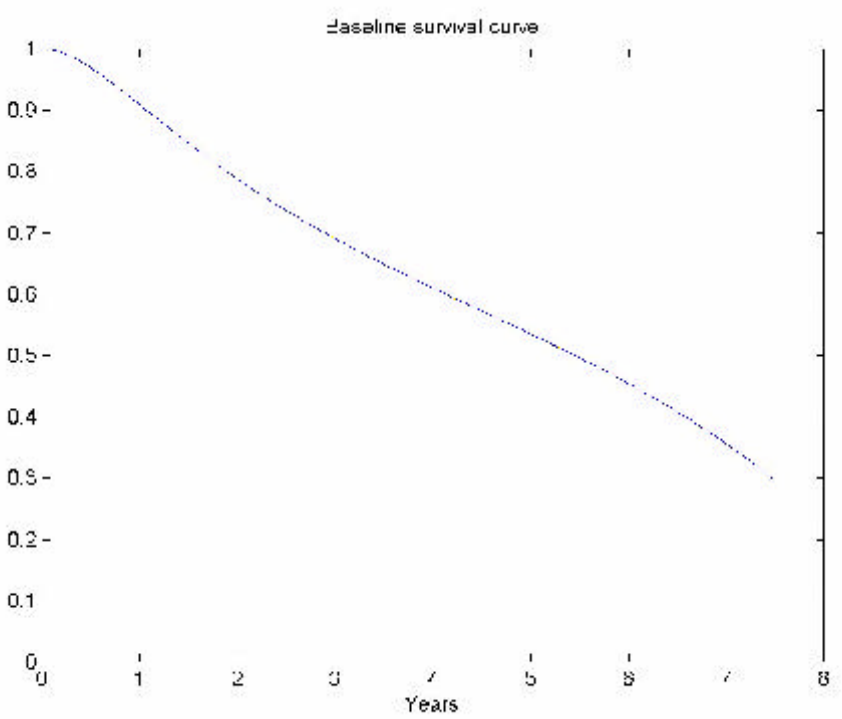
$$?_{ij} = a + \sum_{u=1}^{k} X_{iu}\beta_u + \sum_{v=1}^{p} Z_{jv}?_v + \sum_{u=1,v=1}^{k,p} ?_{uv} X_{iu} Z_{jv}$$

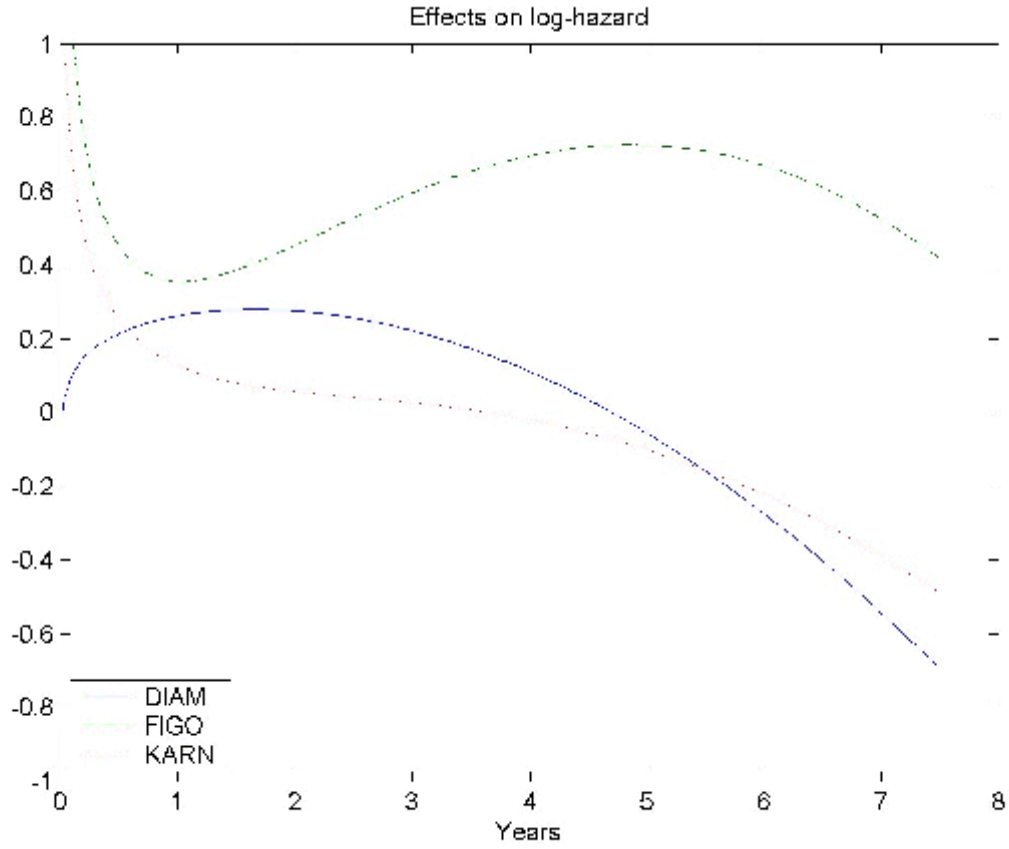
or as time-dependent coefficient model

$$?_{ij} = a + \sum_{v=1}^{p} Z_{jv}?_v + \sum_{u=1}^{k} X_{iu}\left(\beta_u + \sum_{v=1}^{p} ?_{uv} Z_{jv}\right)$$

dim=1+k+p+k*p


Baseline hazard (per year)
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0 1 2 3 4 5 6 7 8
Years


Baseline survival curve
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8
Years


Effects on log-hazard
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
DIAM
FIGO
KARN
0
1
2
3
4
5
6
7
8
Years

Parsimony is important for the predictive value of any model.

(Smits & van Houwelingen , 1999, 2000)

Parsimony achieved by reduced-rank interaction between X and Z

$$?_{ij} ' a \% X_i ß_0 \% Z_j ?_0 \% \sum_{w'1}^{r} (X_i ß_w)(Z_j ?_w)$$

< special case $ß_0 ' ß_1$ (“r=1/2”)

< model bi-linear

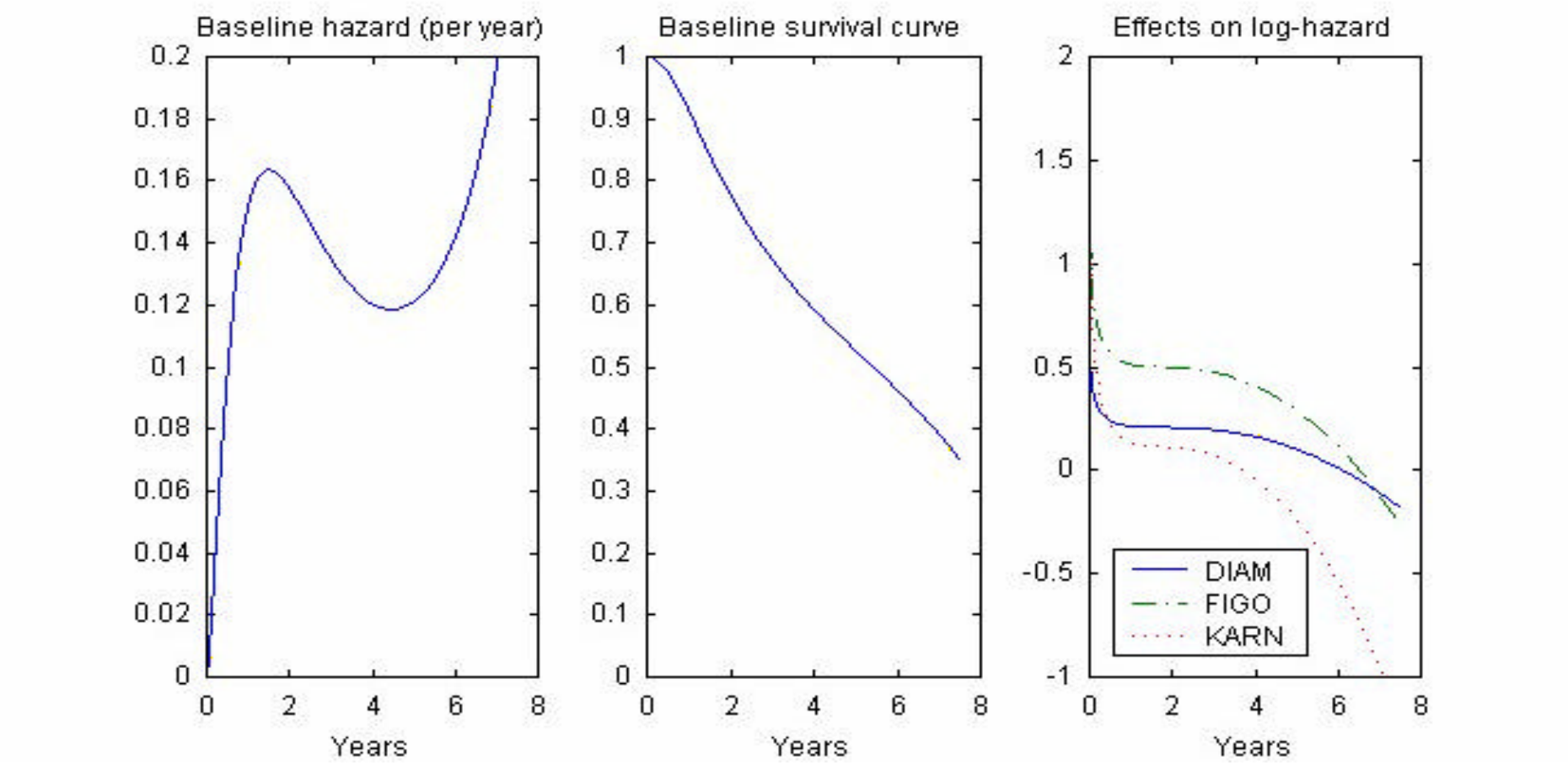
< model not identifiable (redundancy r^2)

< model fitted by alternating regression

model	par.'s	*dim.*	log-lik *l*	*l-dim*
additive	$a,ß,?$	7	-1616.4	-1623.4
r=1/2	$a,ß_0 ' ß_1,?_0,?_1$	10	-1609.1	-1619.1
r=1	$a,ß_0,ß_1,?_0,?_1$	12	-1607.0	-1619.0
r=2	$a,ß_0,ß_1,ß_2,?_0,?_1,?_2$	15	-1604.5	-1619.5
full	$a,ß,?,?$	16	-1604.2	-1620.2

r=1 is winner from AIC point of view

covariate effects slightly smoother than in full rank model



GLIM-approach is

- flexible
- parsimonious
- explicit
- computer intensive

Interesting to compare with

- neural networks
- HARE (Kooperberg)
- frailty and mixture models

Frailty (F) models (Aalen, 1994)

PH + random patient effect $\boldsymbol{Z}$

$$S(t|x,Z) = \exp(-\boldsymbol{Z} RR(x) H_0(t))$$

Popular:

Z has G-distribution with mean one and variance d (G(d,d))

$$S(t|x) = (1+\mathrm{d} RR(x) H_0(t))^{-1/\mathrm{d}}$$

$$h(t|x) = \frac{RR(x) h_0(t)}{1+\mathrm{d} RR(x) H_0(t)}$$

one-dimensional extension of PH !

Cure (C) models (Boag, 1949 and many others)

Mixture models

$$S(t|x) = p(x) + (1-p(x))S_D(t|x)$$

$p(x)$ the probability of "cure"

$$\ln(p(x)/(1-p(x)) = \beta_{C0} + x\beta_{C1}$$

$S_D(t|x)$: surv. fct. of the "non-cured" modeled as PH-model

$$S_D(t|x) = \exp(-RR(x)H_{0D}(t))$$

Regression-dimension doubled! Competitor of rank=1 model.

Frailty and Cure models are fitted in discretized time as the GLIM-models with the same base functions t , $\ln(t)$ and t^2 .

Programs are written in Gauss and not generally available.

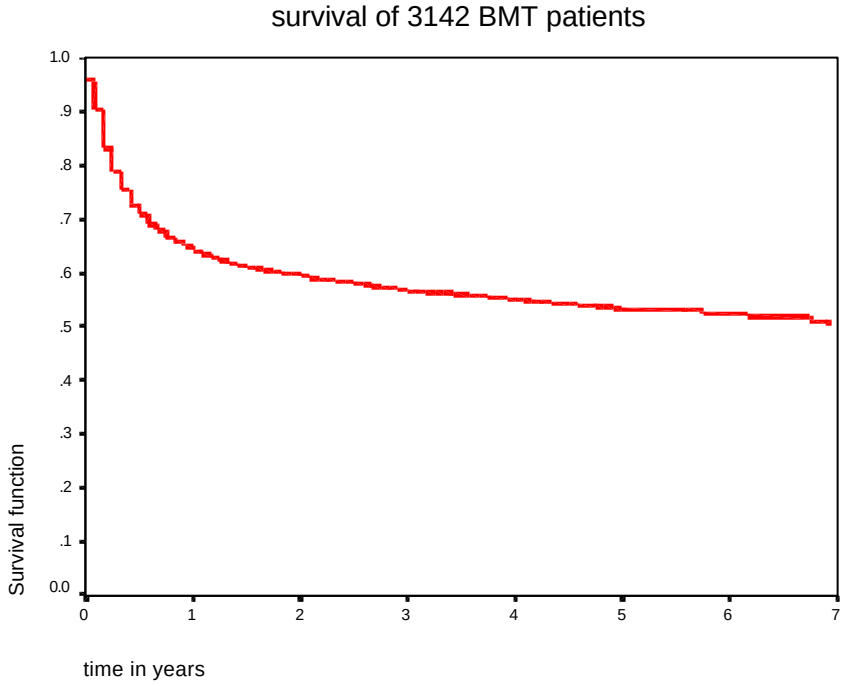
EBMT-data survival after Bone Marrow Transplant

(Gratwohl et al, 1998)

3142 patients 1220 events

time discretized in months,

maximal follow-up to 101 months



risk factors

type

HLA identical	2411
unrelated	731

bmtstage

chronic phase	2301
acute phase	417
blast crisis/other	424

patient's sex

male	1873
female	1269

sexcom

other	2364
female **ö** male	778

age mean 35, st.dev=11, range 0-60

Models fitted (dim(covariate space)=6)

	log-lik	df	model $?^2$	AIC
no cov.	-5486.7	4	0	0
PH	-5347.8	10	277.8	266
frailty model	-5347.8	11	277.8	264
rank=½	-5344.4	13	284.6	267
rank=1	-5338.5	18	296.4	**268**
full rank	-5328.9	28	315.6	268
cure model	-5344	17	285.4	259

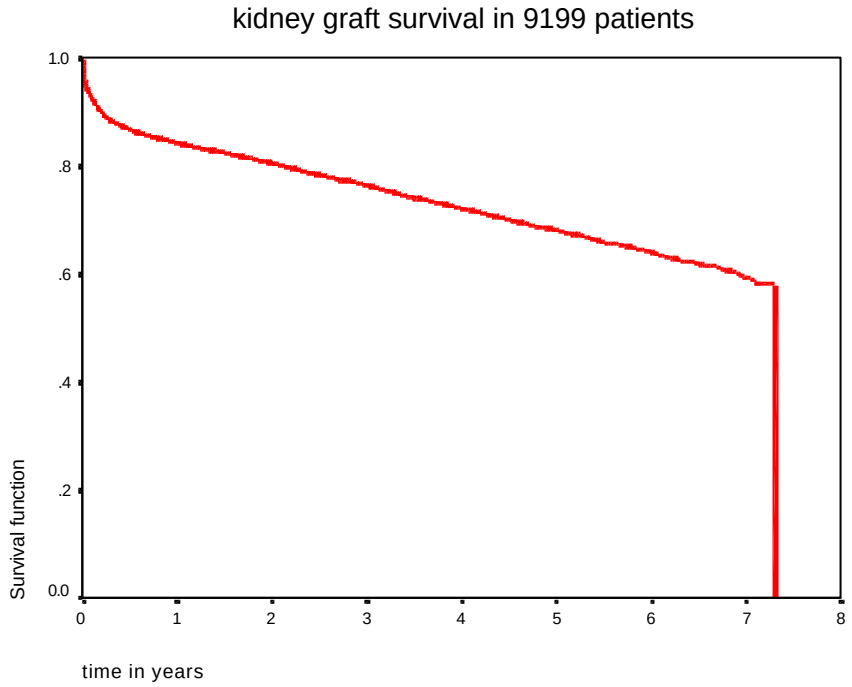
AIC=model $?^2$-2? (df) R^2 . 0.09

Kidney graft survival data

9199 patients 2620 events

time discretized in months

maximal follow-up 89



risk factors

age recipient	
16-54	6582
0-15	233
55-64	1932
65+	452

age of donor	
16-54	6888
0-15	645
55-64	1243
65+	423

cold ischaemic period	
0-18 hrs	2259
19-24 hrs	3476
24-36 hrs	3083
> 36 hrs	381

perc. reactive antibody	
0-5%	5141
6-84%	3342
85+%	716

cause of donor's death	
TC	3651
CVA	4134
other	1414

sex-match	
M to M	3503
M to F	2199
F to M	2117
F to F	1380

number of HLA-mismatches	
0	1712
1	1560
2	2820
3	2421
4+	686

sequence	
first	7832
repeat	1367

Models fitted (dim(covariate space)=21)

	log-lik	df	model $?^2$	AIC
no cov.	-14320.1	4	0	0
PH	-14031.7	25	576.8	534.8
frailty model	-14031.7	26	576.8	532.8
rank=½	-14027.2	28	585.8	537.8
rank=1	-14013	48	614.2	526.1
full rank	-13991.3	88	657.6	489.6
cure model	-14000	47	640.2	**546.2**

$R^2 . 0.06$

Waiting list for kidney transplantation (Smits et al, 1998)

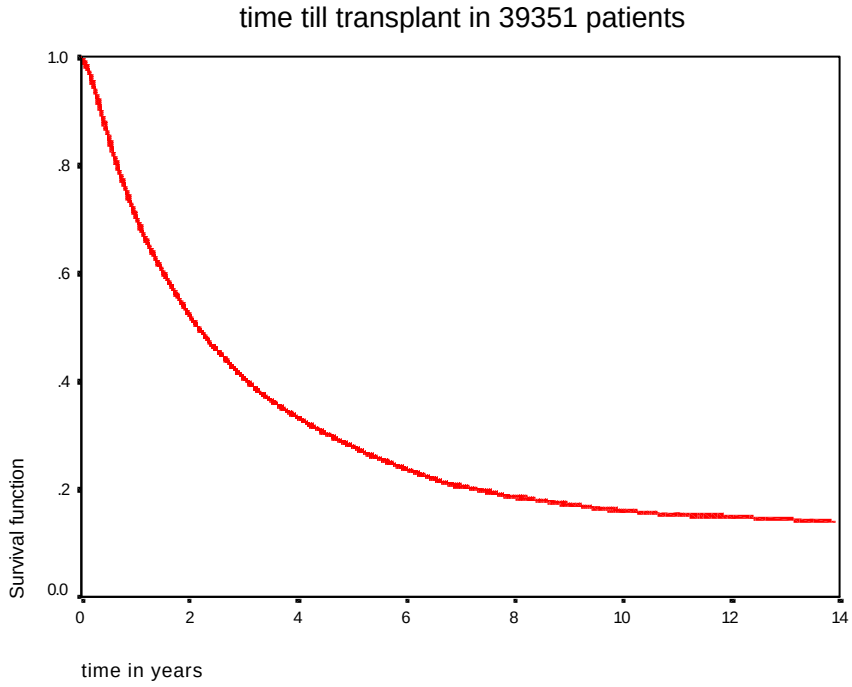
Event: transplantation

(competing risks: death, removal from waiting list)

38685 patients with 25060 events

time discretized in quarters

maximal follow-up 60 quarters



Risk factors

Perc. reactive antibody

0-5%	24606
5-85%	12126
85+%	1953

continuous (0,1,2)

age

<16yrs	1092
>=16yrs	37593

matchability

0=common HLA	8727
1.00	7609
2.00	8382
3.00	5362
4=rare type	8605

continuous (

0,1,2,3,4)

PERIOD

<= 1988	16661
>1988	22024

blood group

0	16151
B	4415
AB	1775
A	16344

GENDER

female	14991
male	23694

Models fitted dim(covariate space)=8

	log-lik	df	model $?^2$	AIC
no cov.	-92694	4	0	0
PH	-89324	12	6740	6724
frailty model	-89322	13	6744	6726
rank=½	-88824	15	7740	7728
rank=1	-88642.04	22	8104	8078
full rank	-88575	36	8238	**8174**
cure model	-88901	21	7586	7552

$R^2 . 0.20$

Simple-minded conclusions

- Frailty model no good at all

- Cure model useful for kidney graft failure data; no use for other examples

Serious question:

- Do the models really differ ?

Comparing similarities of models

- by studying correlation between prognostic indices Xß
- by graphical inspection
- by defining distances between models

Distance measures between two densities

< symmetrized Kullback-Leibler

$$d(f,g) = \int (\ln(g(x)) - \ln f(x))(g(x) - f(x))\,dx$$

< chi-square type of distance

$$d(f,g) = \int \frac{(f(x) - g(x))^2}{f(x) + g(x)}\,dx$$

The latter is slightly more well behaved and will be used here.

The "natural horizon" t_0 can easily be handled by having a point mass at t_0 corresponding with prediction beyond t_0.

Distance between models A and B over individuals is defined as

$$d(A,B) = \frac{1}{n} \sum_{i=1}^{n} d(f_{Ai}, f_{Bi})$$

The distances can be used to map the models via multi-dimensional scaling.

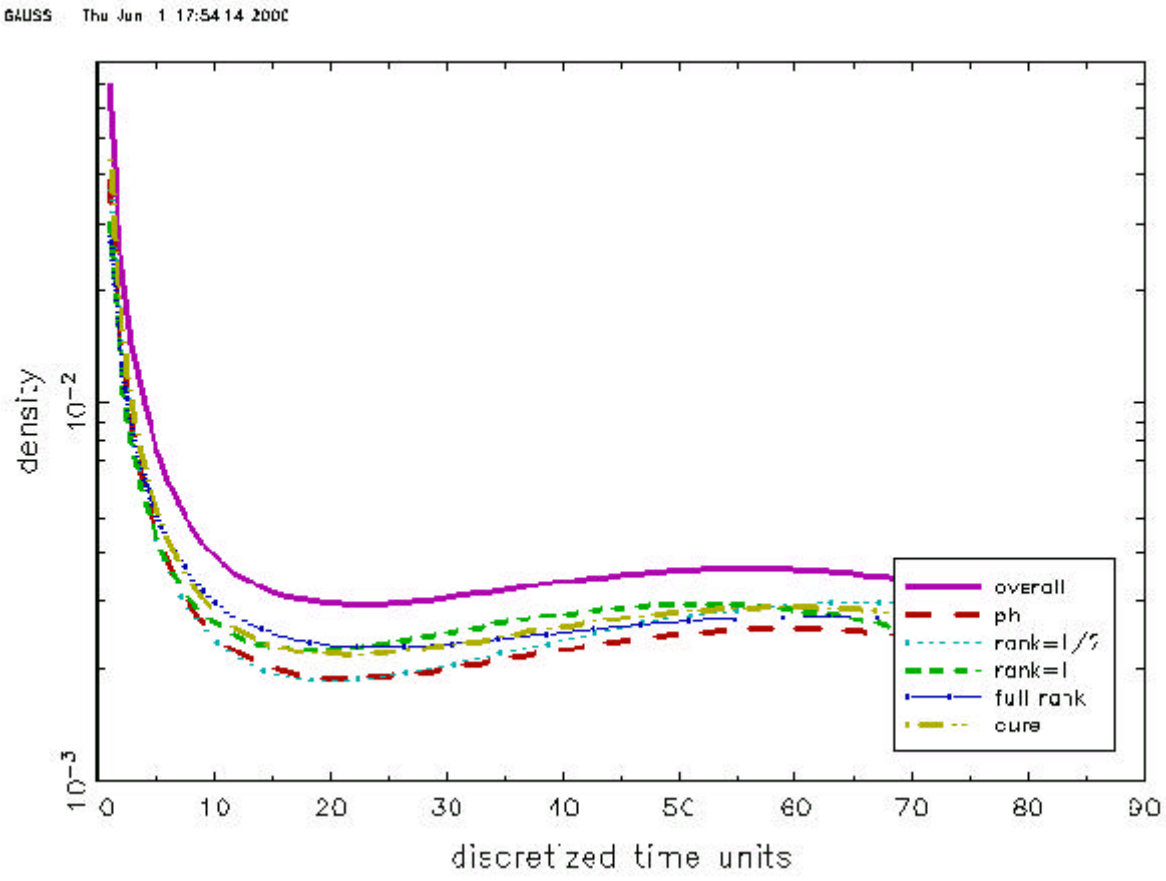
(Presentation in low dimension with same (squared) distances)

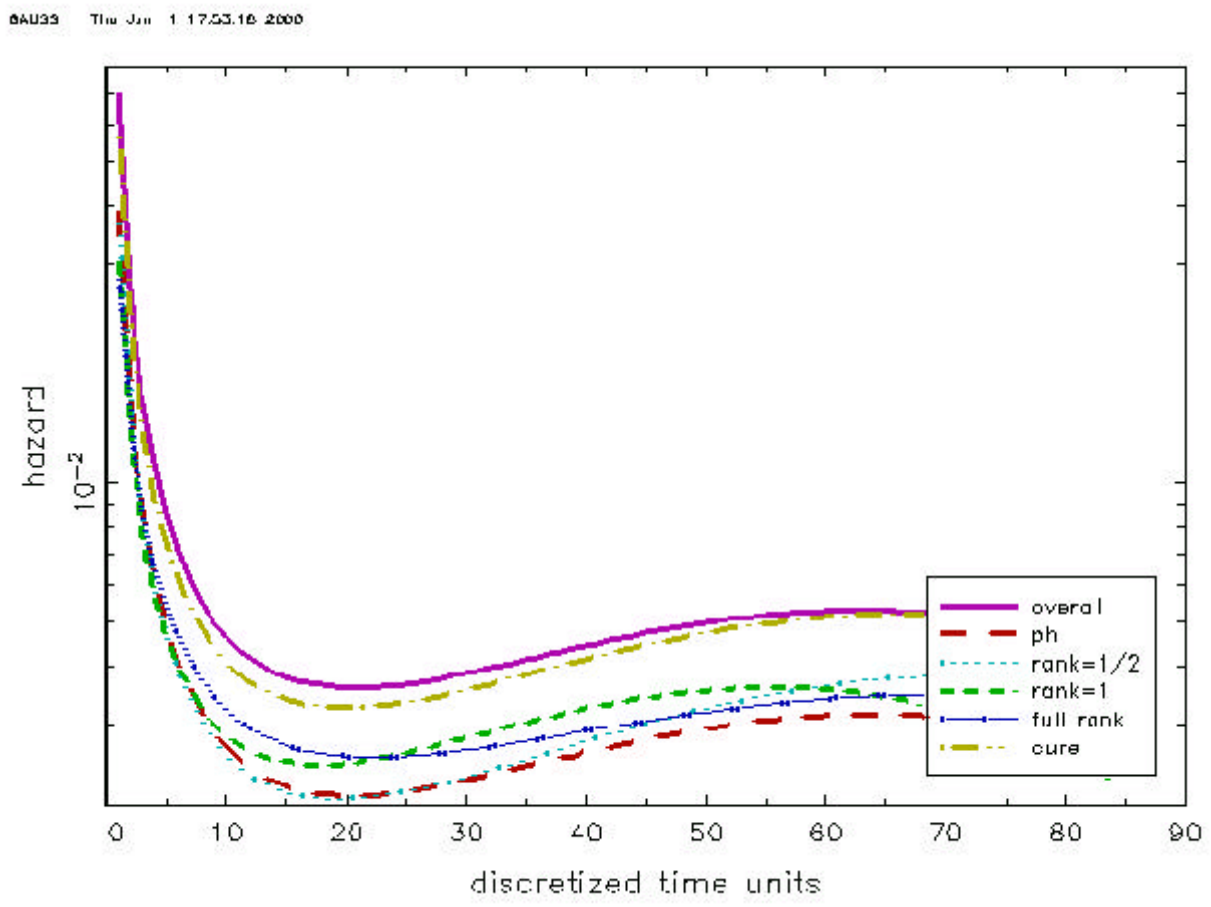
Model comparison for graft survival.
Correlations between “prognostic indices”

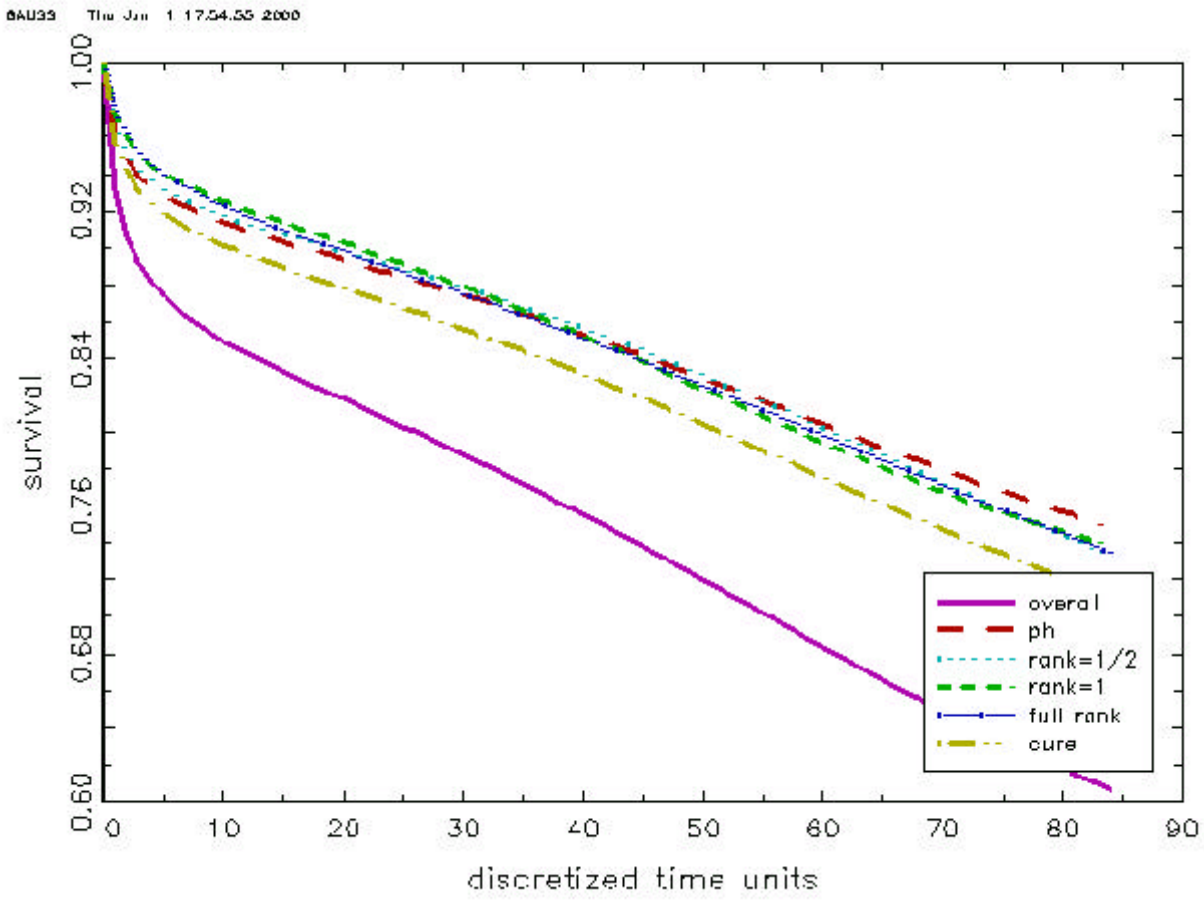
PH	rank =1/2	rank=1		full rank				cure	
1.0								C	D
1.0	1.0								
1.0	1.0	1.0							
0.2	0.2	0.2	1.0						
0.8	0.8	0.8	0.0	1.0					
0.2	0.2	0.2	0.0	0.7	1.0				
0.2	0.1	0.2	-0.5	0.7	0.8	1.0			
0.0	0.0	0.0	0.2	0.5	0.9	0.6	1.0		
0.5	0.5	0.5	0.0	0.5	0.2	0.2	0.1	1.0	
1.0	1.0	1.0	0.1	0.8	0.2	0.2	0.1	0.7	1.0

Comparison of models based on first 84 months

One random example



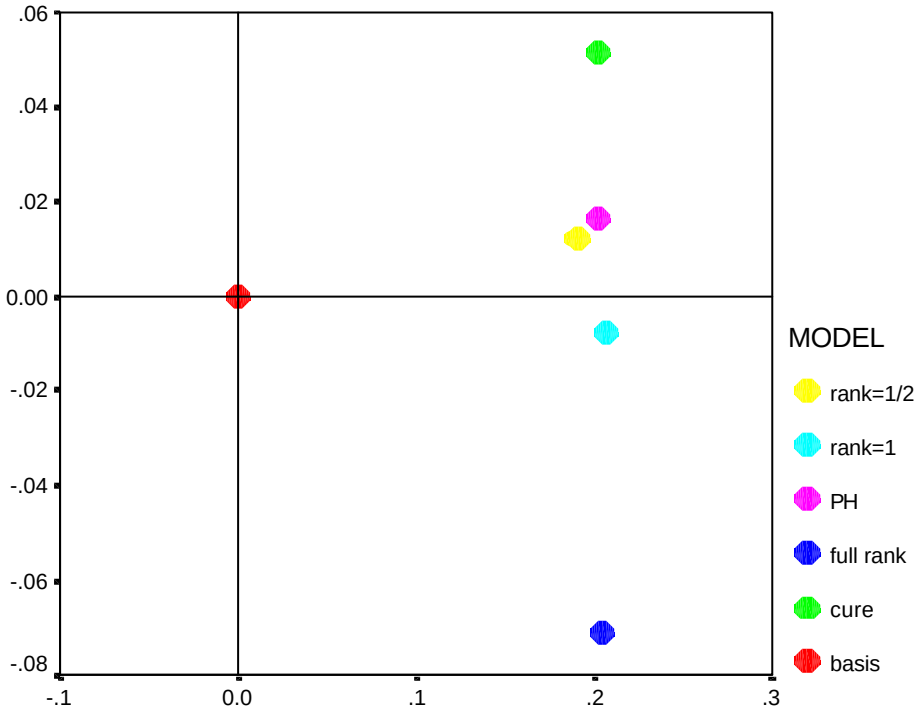

hazard
10^{-2}
0
10
20
30
40
50
60
70
80
90
discretized time units
overal
ph
rank=1/2
rank=1
full rank
cure

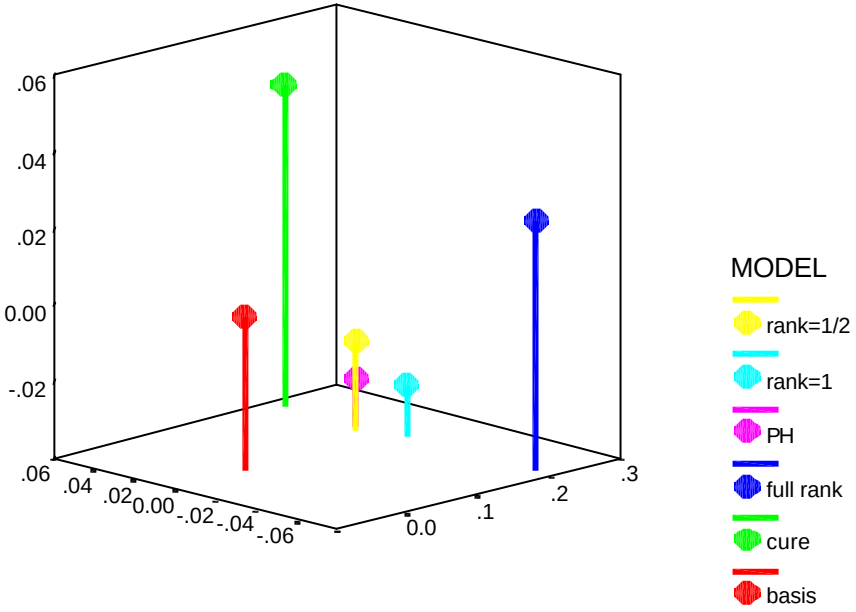

survival
1.00
0.92
0.84
0.76
0.68
0.60
0
10
20
30
40
50
60
70
80
90
discretized time units
overal
ph
rank=1/2
rank=1
full rank
cure

Distances between models

	basis	PH	r=1/2	r=1	full	cure
basis	0.000	0.042	0.037	0.044	0.047	0.045
PH	0.042	0.000	0.002	0.003	0.011	0.007
r=1/2	0.037	0.002	0.000	0.004	0.010	0.007
r=1	0.044	0.003	0.004	0.000	0.008	0.009
full	0.047	0.011	0.010	0.008	0.000	0.016
cure	0.045	0.007	0.007	0.009	0.016	0.000

Bi- and triplots





Results for waiting list data

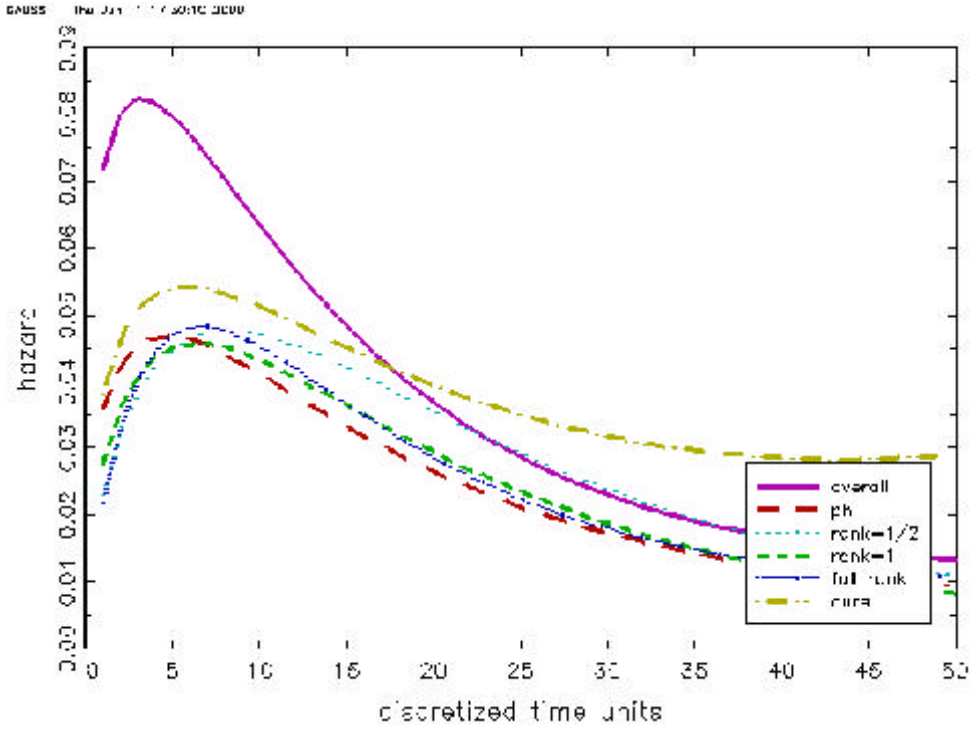
Correlations between “prognostic indices”

PH	rank =1/2	rank=1		full rank				cure	
1.0								C	D
1.0	1.0								
1.0	1.0	1.0							
0.9	0.9	0.9	1.0						
0.8	0.8	0.9	1.0	1.0					
0.6	0.6	0.8	0.9	0.9	1.0				
-0.2	-0.2	0.0	0.2	0.4	0.6	1.0			
0.5	0.5	0.7	0.8	0.7	0.9	0.3	1.0		
0.0	0.0	0.2	0.3	0.5	0.6	**0.9**	0.2	1.0	
1.0	1.0	1.0	0.9	**0.9**	0.7	0.0	0.6	0.2	1.0

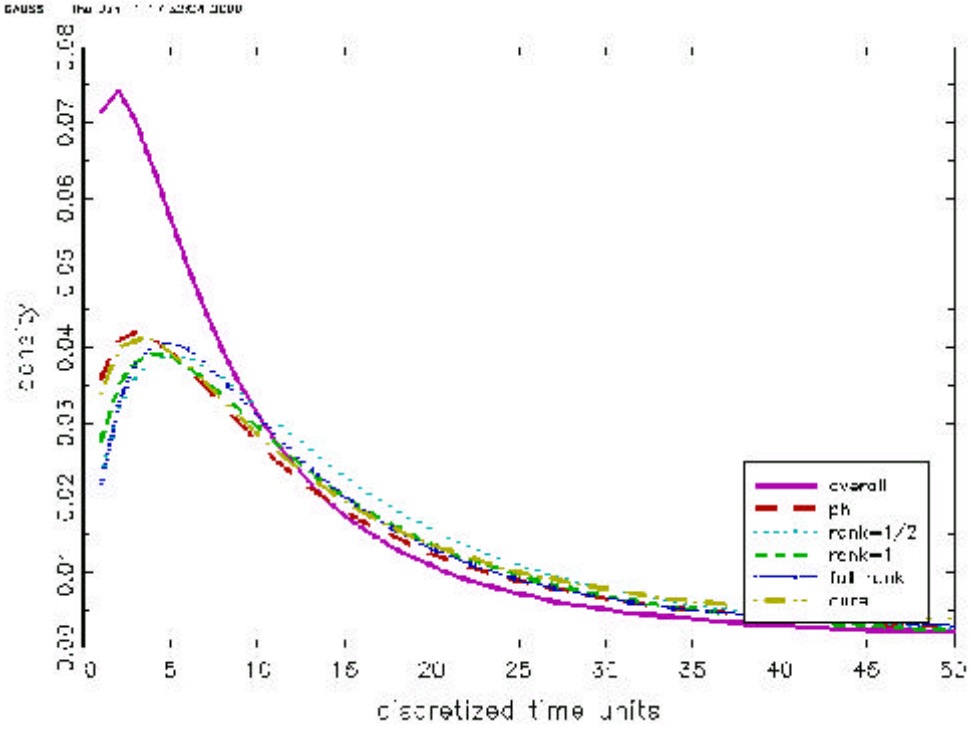
Comparison of models based of first 50 quarters

One random example

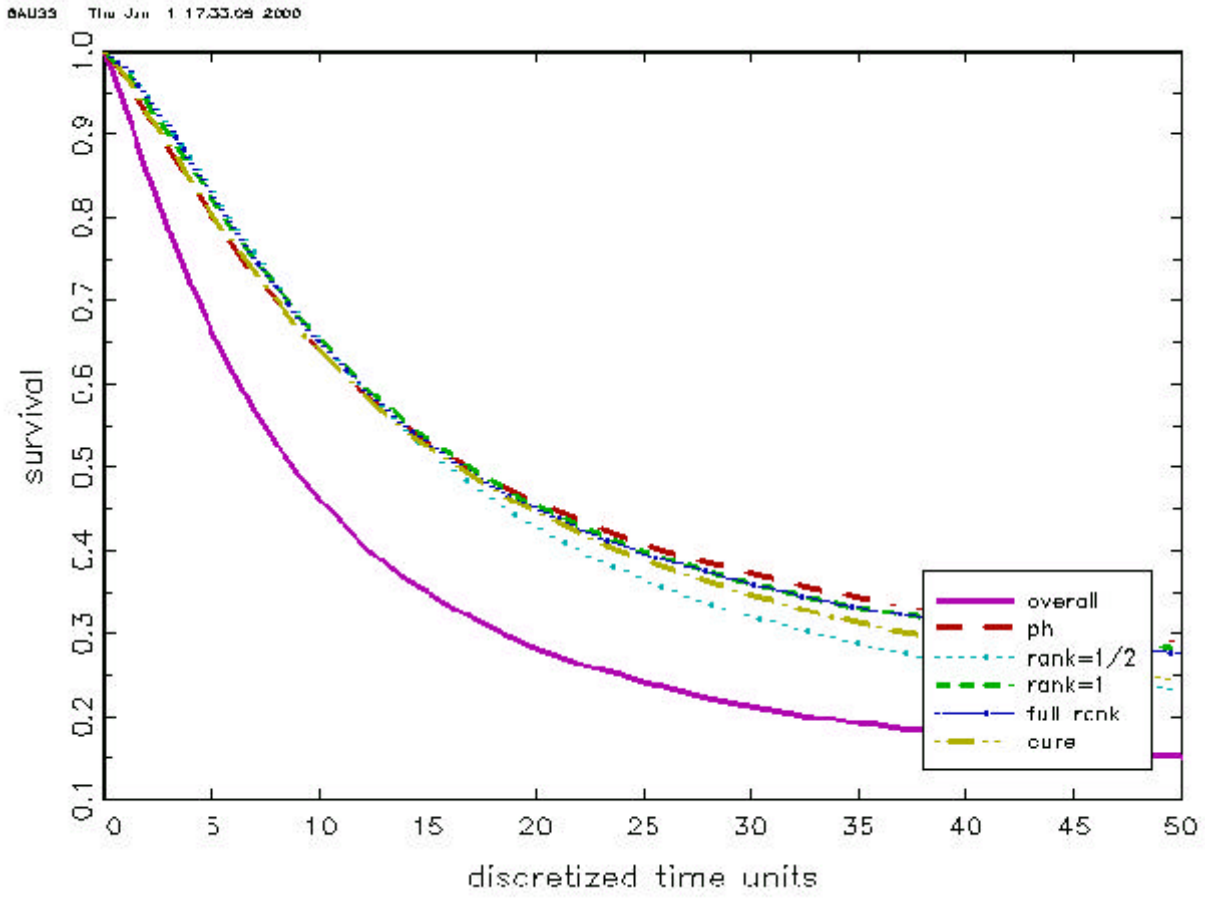
hazards



densities


density
discretized time units
overall
ph
rank=1/2
rank=1
full rank
cure

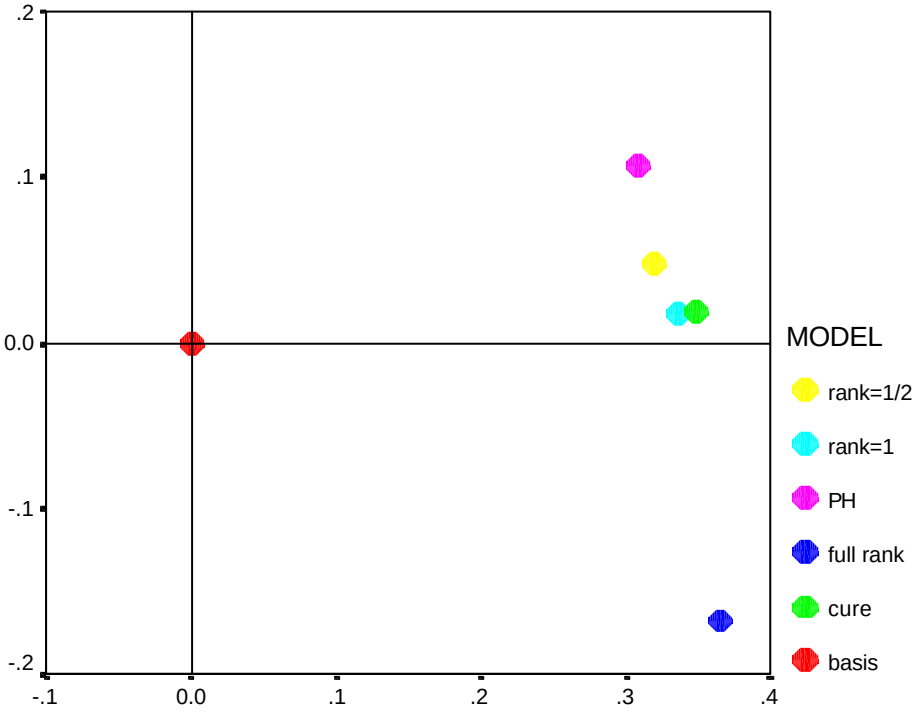
survivor functions

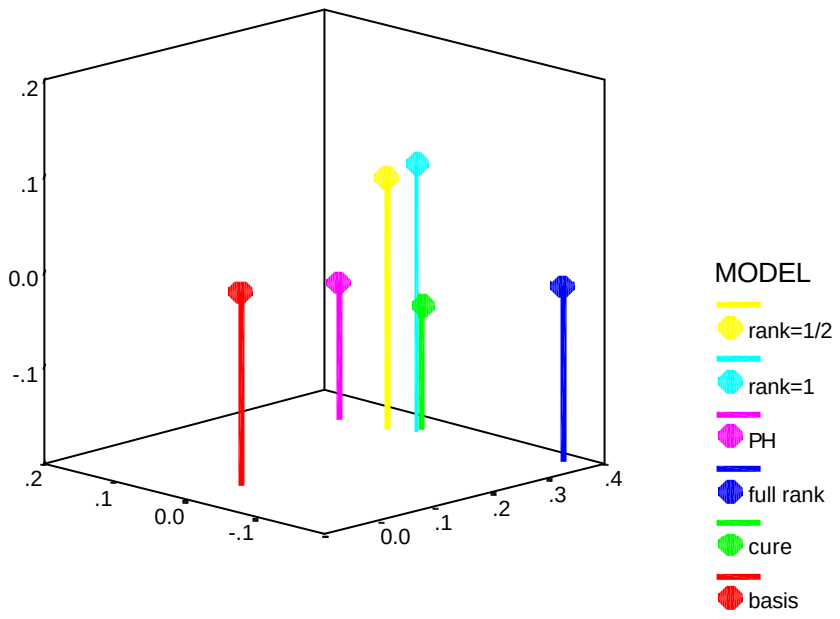


Distances between models

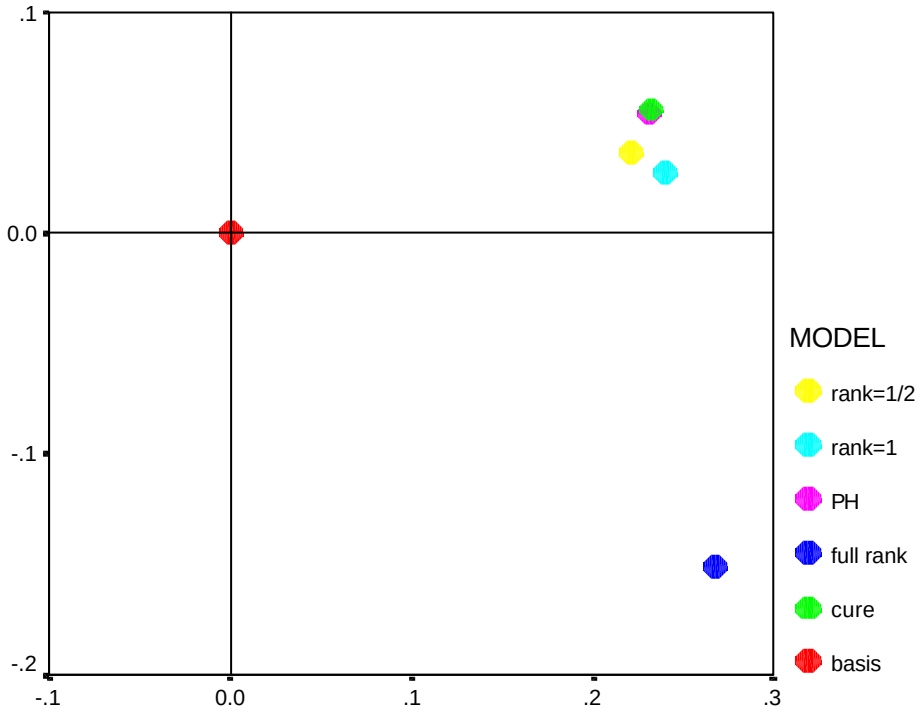
	basis	PH	rank=1/2	rank=1	full rank	cure
basis	0.00	0.11	0.11	0.12	0.16	0.13
PH	0.11	0.00	0.02	0.03	0.08	0.03
rank=1/2	0.11	0.02	0.00	0.01	0.06	0.03
rank=1	0.12	0.03	0.01	0.00	0.05	0.03
full	0.16	0.08	0.06	0.05	0.00	0.05
cure	0.13	0.03	0.03	0.03	0.05	0.00

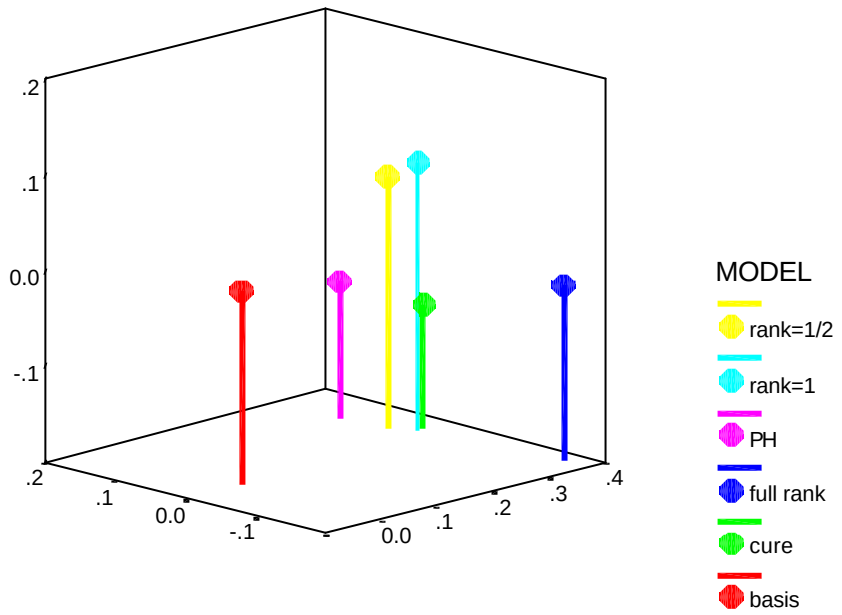
Bi- and Tri-plots of distance between models





Bi- and tri-plots for EBMT data





Conclusion from results of model comparison

- Many events are needed to detect differences
- Cure models is intrinsically different from GH-model of rank=1
- Cure model can be low-dimensional alternative to time-dependent effect modeling
- Supermodel that combines Cure-model and GH-model could be of interest