

Faculty of Health Sciences

Basal statistik

Overlevelsesanalyse

Lene Theil Skovgaard

11. maj 2020



Overlevelsesanalyse

- ▶ Levetider og censurerede observationer
- ▶ Kaplan-Meier kurver
- ▶ Log-rank test
- ▶ Cox regression
- ▶ Modelkontrol og udvidede modeller

Home pages:

<http://publicifsv.sund.ku.dk/~sr/BasicStatistics>

E-mail: ltsk@sund.ku.dk

*: Siden er lidt teknisk



Forklarende variable = Kovariater

Outcome	Dikotom	Kategorisk	Kvantitativ	Kategoriske og kvantitative
Dikotom parret	2*2-tabeller Mc Nemar	χ^2 -test svært, mixed models		Logistisk regression Mixed models
Kategorisk	Kontingenstabeller/ χ^2 -test		Generaliseret logistisk regression	
Ordinale	svært, f.eks. proportional odds modeller			
Kvantitativ parret	Mann-Whitney Wilcoxon signed rank	Kruskal-Wallis Friedman		Robust multipel regression
Normalfordelte residualer	T-test uparret/parret	Variansanalyse ensidet/tosidet		Kovariansanalyse Multipel regression
Censureret	Log-rank test		Cox regression	
Korrelerede kvantitative Nf. residualer	Varianskomponent-modeller Mixed models		Modeller for gentagne målinger	



Eksempel: Lungecancer blandt krigsveteraner

137 krigsveteraner med diagnosen inoperabel lungecancer, af **4 forskellige typer** (1:squamous, 2:smallcell, 3:adeno, 4:other) følges frem til død (vi glemmer lige et øjeblik, at 9 af dem faktisk endnu er i live ved tidspunktet for analysen)

De er randomiseret til enten

- ▶ standard behandling ($\text{treat}=0$, $n=69$)
- ▶ ny test behandling ($\text{treat}=1$, $n=68$)

Vi vil gerne sammenligne disse to behandlinger ...og så har vi nogle kovariater (udover behandling og type), som vi vender tilbage til



Valg af tidsskala

Der skal vælges en *tidsskala* for levetiden, dvs. et nulpunkt.

Typiske muligheder er:

- ▶ alder
- ▶ kalendertid
- ▶ tid siden diagnose
- ▶ **tid siden randomisering** (vælges her)

Man kan f.eks.

- ▶ få diagnosen som 40-årig, indgå i undersøgelsen fra dette tidspunkt og dø som 72-årig....
- ▶ få diagnosen som 50-årig, randomiseres til undersøgelsen som 55-årig og dø som 65-årig....



Valg af tidsskala, fortsat

Ofte er der flere muligheder, og **valget er ikke altid helt oplagt**:

Der skal være en grund til at tro, at den valgte tidsskala afspejler risikoen ens for alle individer (på nær diverse kovariater), altså at der som funktion af tiden ses samme *form* på risikoen.

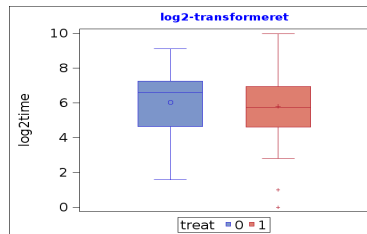
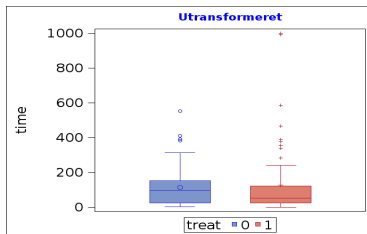
Eksempelvis kunne man vælge

- ▶ For kroniske sygdomme:
Alder: Fordi behandlingen formodentlig ikke ændrer radikalt ved prognosen
- ▶ For stærkt behandlingskrævende sygdomme:
Tid siden behandling: iden man kan formode, at en vellykket behandling på afgørende vis ændrer prognosen



Sammenligning af behandlinger

Hvis alle var fulgt indtil død, kunne vi lave T-test el.lign.



T-test på log-skala: $P=0.50$,

Mann-Whitney: $P=0.35$

men men men.....

Hvad er nyt?

Outcome: Tid fra et veldefineret nulpunkt og indtil en bestemt begivenhed (sædvanligvis død) indtræffer.

Her: Levetid i dage siden randomisering, angivet som y_i .

Det er her, der sker noget nyt, fordi det ikke er alle, der er døde endnu

Vi kalder det for censurerede observationer

Her er der kun 9 af sådanne, men ofte er der mange flere

Kovariater: Behandling, cancertype, alder, etc. etc.

Ikke noget *essentielt* nyt her....



Notation for outcome

t_i : Det tidspunkt (regnet fra nulpunktet),
der er tilknyttet det i 'te individ,
og som angiver *enten* død *eller* censureringstidspunkt.

d_i : Status ved tid t_i :

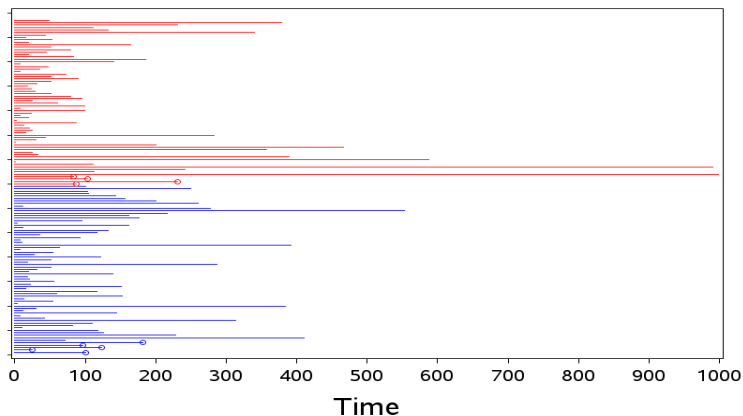
- ▶ 0: censureret, dvs. stadig i live
- ▶ 1: død

y_i : Tid til død for i 'te patient er så ikke observeret for alle:

- ▶ Hvis en patient dør til tid t_i , så er $y_i = t_i$ og $d_i = 1$
- ▶ Hvis man mister kontakten til en patient til tid t_i ,
så er $y_i > t_i$ og $d_i = 0$
- ▶ Hvis analysen laves til tid t_i , *inden* patient i dør,
så er $y_i > t_i$ og $d_i = 0$



Illustration af data



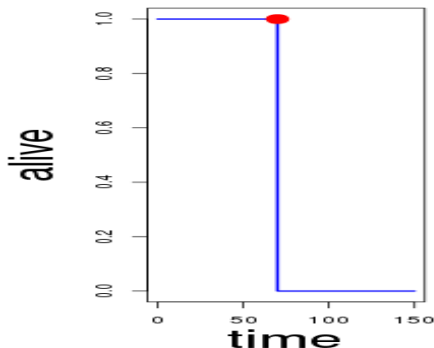
treat=0, treat=1

Cirklene angiver censurerede observationer



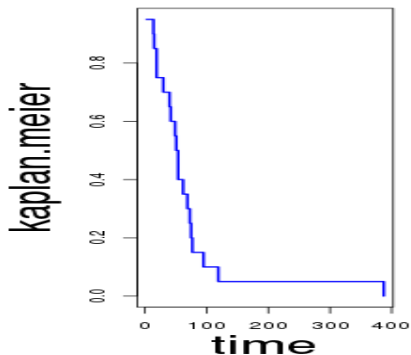
Hvad ved vi?

Vi vil opfatte tiden som *dynamisk*, i den forstand, at vi for hvert tidspunkt t ser på, hvilken information vi har om den enkelte, nemlig status for overlevelse: $S_i(t) = I(y_i > t)$



Kaplan-Meier: Overlevelseskurve

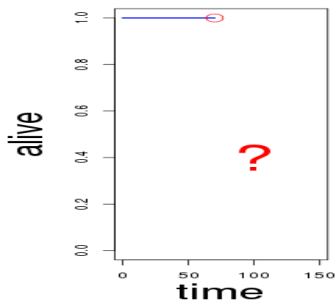
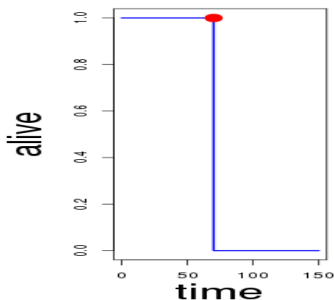
Hvis alle er observeret til død, er overlevelseskurven $S(t)$ blot gennemsnittet af disse, altså til hvert tidspunkt blot fraktionen, der stadig er i live. Den kaldes Kaplan-Meier kurven



Men når vi har censureringer?

så ved vi ikke, hvad der sker med funktionen $S_i(t) = I(y_i > t)$ efter det tidspunkt t_i , hvor vi sidst ser patienten i live.

Vi ved så kun, at deres levetid er større end t_i ,
men ikke *hvor meget* større.



Vigtigt - Uafhængige censureringer?

En **meget vigtig** antagelse om censureringer:

De skal være **uafhængige** af prognosen, dvs.

Det faktum, at en patient censureres ved tid t_i må på *ingen måde* være relateret til den fremtidige risiko for at dø

Censureringer, der muligvis *ikke* opfylder dette kriterium:

- ▶ Patienter, der overflyttes til et andet hospital:
De har formentlig en dårligere prognose
- ▶ Patienter, der flytter udenlands pga nyt job:
De har formentlig en bedre prognose

Og hvad, hvis

- ▶ patienten bliver kørt over?



Kaplan-Meier kurver, I

Når der er censureringer, kan vi ikke tage gennemsnittet af de individuelle S_i 'er, fordi vi ikke kender dem alle.

Men hvis censureringerne er uafhængige, kan vi estimere den del, vi ikke kender til blot at være "*ligesom dem, vi ser*", altså antage, at disse patienter dør i samme takt som de øvrige

Det er netop det, der ligger i uafhængig censurering.

Så hvis det, at man bliver kørt over, tyder på, at man har det dårligt (har fået recidiv), så har man nok på dette tidspunkt en værre prognose end de, der *ikke* blev kørt over....?



Håndtering af censureringer

Hvorfor kan vi ikke bare...

- ▶ Smide de censurerede individer helt væk?

Vi kender jo ikke deres levetid, så der er vel tale om *missing values*?

Nej, for vi ved *noget* om deres levetid, nemlig at den er mindst t_i , og hvis vi smider dem ud, laver vi derfor *bias* i overlevelseskurven

- ▶ Lade som om alle t_i 'er er dødsfald?

Fordi så ser det jo ud som om de dør alt for hurtigt, så her bliver også *bias*

Vi er nødt til at håndtere censureringer på en *dynamisk* måde, dvs. ved at se på situationen til et givet tidspunkt og et kort tidsrum herefter.



Kaplan-Meier kurver, II

når der er censureringer:

- ▶ Vi deler tidsaksen op ved hvert tidspunkt $t_1 < t_2 < t_3 \dots$ (her sker der noget, enten et dødsfald, eller en censurering)
- ▶ For hvert interval, f.eks. (t_3, t_4) ser vi på, hvor mange, der var *i risiko* i starten (n_3) og hvor stor en del af disse, der overlevede tidsintervallet, nemlig brøken

$$R_4 = \begin{cases} 1 & \text{hvis } t_4 \text{ er en censurering} \\ \frac{n_3-1}{n_3} & \text{hvis } t_4 \text{ er et dødsfald} \end{cases}$$

- ▶ Kaplan-Meier kurven fås nu ved successiv sammengangning af disse brøker:

$$1 \times R_1 \times R_2 \times R_3 \times R_4 \times \dots$$



Praktisk håndtering af censurering

Til programmering er outcome 2-delt:

- t_i Observationstiden for i 'te patient (time)
som kan være *enten* død (status=1), *eller* en
censurering (status=0).
- d_i Indikator for død (status), dvs. 1, hvis patienten
faktisk er død, og 0, hvis vedkommende er censureret

Begge disse variable skal angives som (todelt) outcome

I SAS gøres dette ved at benytte skrivemåden

`time*status(0)=...` i model-sætningen,

idet 0'et efter status angiver **koden for censurering**



Kaplan-Meier kurver i SAS

kan tegnes med proceduren PHREG:

```
proc phreg plots(overlay=row cl)=survival data=a1;  
  model time*status(0)= ;  
  strata treat;  
run;
```

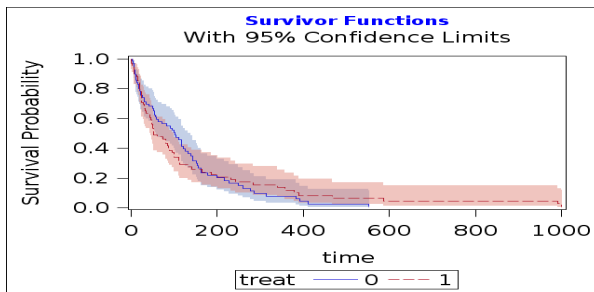
- ▶ Grupperingen styres af variabelen i strata, og bemærk, at selve modellen er *tom* (ingen kovariater)
- ▶ `plots(overlay=row cl)=survival` angiver, at vi ønsker survival kurver, lagt oveni hinanden, og med konfidensgrænser.

Figuren ses s. 20



Kaplan-Meier kurver, $S(t)$

fra PHREG i SAS (se kode s. 19):

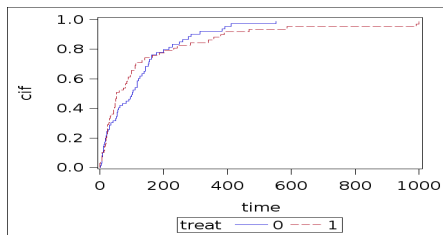


Til ethvert tidspunkt kan vi aflæse den forventede procentdel af patienter, der stadig er i live - for hver behandlingsgruppe

Kumulativ incidens

Til et hvert tidspunkt angiver dette sandsynligheden for at have fået et event, dvs. (i dette tilfælde) at være død. Denne er blot

$$CI(t) = 1 - S(t)$$



Kode til figuren ses s. 84

Pas på i tilfælde af *competing risks* (se s. 79-80)



Er der forskel på de to behandlinger?

De to overlevelseskurver (s. 20) krydser lidt frem og tilbage over hinanden. De eksperimentelt behandlede dør hurtigst til at starte med, men efter ca. 100 dage ser det ud til at vende (bedømt på hældningen af kurverne).

Hvordan kan vi sammenligne?:

- ▶ **log-rank test:**
giver en P-værdi for test af identitet
- ▶ **Cox proportional hazard model:**
giver estimat for forskel (ratio) mellem dødsrater
– under *forudsætning* af proportionale rater
(kommer om lidt...)



Log-rank testet for H_0 : ens dødelighed

Et **non-parametrisk** test til sammenligning af overlevelseskurver **uden antagelser om formen af disse**

- ▶ Til hvert tidspunkt ser man på forskellen mellem observeret og forventet antal døde (under H_0) i de to grupper.
- ▶ Forskellen mellem disse, passende normeret, giver et χ^2 -test med $G - 1$ frihedsgrader (G er antallet af grupper, her $G = 2$)
- ▶ Det er ikke så stærkt, hvis forskellen på kurverne ændrer sig markant over tid.

Her giver testet $P = 0.93$ (se s. 25), altså **ingen påviselig forskel**



Log-rank testet i praksis

- ▶ Testet fås som et såkaldt *score*-test (*teknisk*: det har at gøre med tangenthældningen på likelihoodfunktionen).
I SAS PHREG kræver det at man benytter `ties=discrete`:

```
proc phreg data=a1;  
  class treat(ref="0") / param=glm;  
  model time*status(0)=treat / ties=discrete rl;  
run;
```

- ▶ Det fås også direkte fra SAS LIFETEST, se s. 85



Output fra Log-rank test

The PHREG Procedure

Model Information

Data Set	WORK.A1
Dependent Variable	time
Censoring Variable	status
Censoring Value(s)	0
Ties Handling	DISCRETE

Number of Observations Used	137
-----------------------------	-----

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
137	128	9	6.57

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Score	0.0082	1	0.9277 <---log-Rank test

Der er absolut ikke signifikant forskel -
men hvad med kvantificering af forskellen?



Rater = hazards = intensiteter

- ▶ Vi har set, at levetiden ikke dur som outcome, fordi en sådan ikke altid observeres.
- ▶ I stedet vil vi gerne modellere “hvor hurtigt overlevelseskurven går ned ad bakke”, altså en mere *dynamisk* beskrivelse, i form af den umiddelbare risiko for død (*rater*, *hazards* eller *intensiteter*):

Givet, at jeg er i live lige nu, hvad er så min risiko for at dø indenfor det næste år?

Eller næste måned? Næste dag/næste time....

- ▶ **Hazard funktion $h(t)$:** den *umiddelbare* risiko for at dø i det næste korte interval Δt er $h(t)\Delta t$

$h(t)$ er en rate, ikke en sandsynlighed



Hazard funktionen

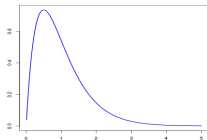
- ▶ er ikke så god at estimere, fordi den hopper op og ned (andel døde i hvert lille tidsrum er enten 0 eller et ret lille tal)
- ▶ estimer i stedet **kumuleret hazard, $H(t)$**
- ▶ Matematikken kan vise, at $H(t) = -\log(S(t))$
- ▶ Hældningen af $H(t)$ afspejler den umiddelbare risiko/rate/hazard $h(t)$ for at dø:
 - ▶ stejl hældning: stor risiko
 - ▶ lille hældning: lav risiko

Teknisk note: $H'(t) = -\frac{S'(t)}{S(t)}$, dvs. det er altså den *relative* ændring af overlevelseskurven, vi modellerer.

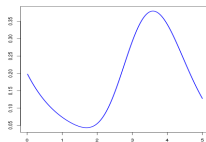


Eksempler på hazard-funktioner $h(t)$

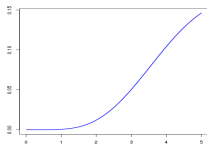
Umiddelbar farlig behandling:



Gavnlig behandling,
med senfølger:

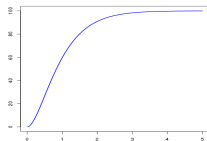


Alder som mest betydende:

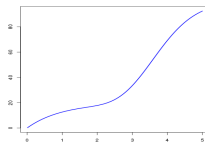


Kumulerede hazard-funktioner $H(t)$

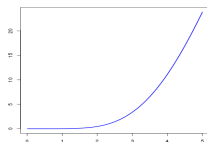
Umiddelbar farlig behandling:



Gavnlig behandling,
med senfølger:

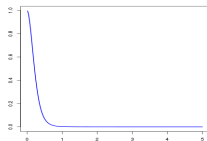


Alder som mest betydende:

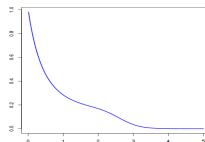


Overlevelseskurver $S(t)$

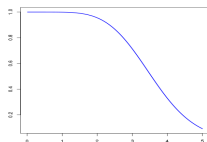
Umiddelbar farlig behandling:



Gavnlig behandling,
med senfølger:



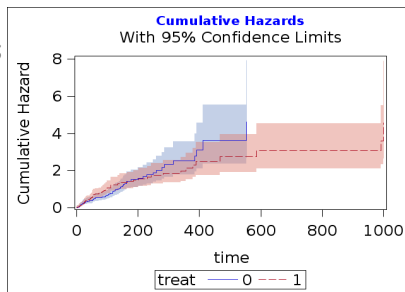
Alder som mest betydende:



Kumulerede hazard funktioner

i vores eksempel om krigsveteraner

```
proc phreg data=a1  
  plots(overlay=row cl)=cumhaz;  
  model time*status(0)= ;  
  strata treat;  
run;
```



Er de ens?

Hvis de ikke er ens, er de så **proportionale?** (de starter jo begge i 0)

Cox *proportional hazards* model

En såkaldt **Cox model** specificerer alle hazard-funktioner til at være proportionale.

Her har vi bare to grupper, så vi **antager**

$$h_1(t) = k \times h_0(t)$$

Men **baseline hazard funktionen** $h_0(t)$ specificeres ikke nærmere, dvs. vi antager ikke noget om afhængigheden af tiden

Vi vil gerne estimere proportionalitetskonstanten k ,
hvis modellen altså holder....., dvs. hvis der er additivitet på log-skala: $\log(h_1(t)) = \log(k) + \log(h_0(t))$



Estimat for behandlingseffekt

under antagelse af proportionale intensiteter, som antages, når `treat` sættes som kovariat i model-sætningen (se kode s. 24):

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
treat 1	1	0.01645	0.18129	0.0082	0.9277
treat 0	0	0	.	.	.

Parameterestimatet er her differensen mellem grupperne på log-hazard skala, og den skal tilbagetransformeres til en faktor, kaldet **hazard ratio** eller *rate ratio*.

Parameter		Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
treat 1	1	1.017	0.713 1.450	treat 1
treat 0	0	.	.	treat 0



Fortolkning af behandlingseffekten

Effekten af eksperimentel behandling ($\text{treat}=1$ vs. $\text{treat}=0$) er at øge hazard med faktoren 1.017, altså en forøgelse på 1.7%.

Dette er dog på ingen måde signifikant ($P=0.93$), idet konfidensintervallet er $CI=(0.71, 1.45)$, og altså indeholder effekter lige fra en forøgelse af hazard med 45% til en formindskelse af hazard med 29%.

Dette er den simplest tænkelige Cox-model med kun en enkelt kategorisk kovariat på 2 niveauer (2 grupper), *og raterne antages altså at være proportionale.*



*Proportionale rater/hazards

I tilfælde af to grupper har vi $h_1(t) = k \times h_0(t)$

og dermed parallellitet på log-skala, også for de *kumulerede* rater:

$$\log(H_1(t)) = \log(k) + \log(H_0(t))$$

Da der gælder $H(t) = -\log(S(t))$, svarer det til, at

$$\log(-\log(S_1(t))) = \log(k) + \log(-\log(S_0(t)))$$

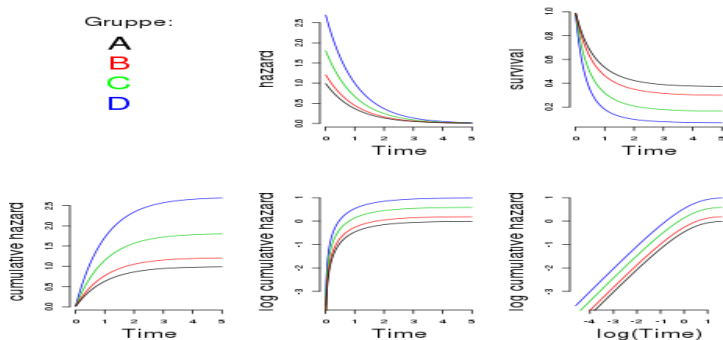
og man siger derfor, at der er parallellitet på loglogs-skala, eller på *cloglog*-skala. **Dette kan kontrolleres**

- ▶ Grafisk: se s. 37 og 40
- ▶ Numerisk: se s. 40-41



Eksempel på proportionale rater/hazards

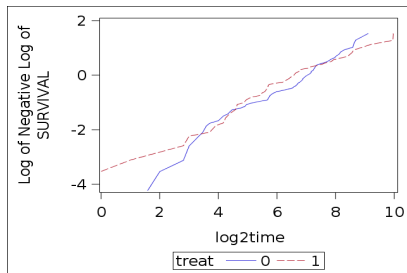
i 4 grupper (4 værdier af en kategorisk kovariat):



Figuren af $\log H(t)$ mod $\log(t)$ nederst til højre viser tydeligt parallelle kurver, dvs. **proportional hazards (PH)**

Har vi PH i vores eksempel?

For at lave den relevante figur, gemmer vi $\log(H(t))$ og regner videre, se s. 87:



Er disse to parallelle?

Ikke synderligt....

Check af proportionale rater

Grafisk: Her er to muligheder:

1. Check parallellitet af $H(t)$ 'erne i grupperne, på loglogs-skala, som vist s. 37
2. Benyt de såkaldte *kumulerede score-residualer* via assess-statement i PHREG, se s. 40

Numerisk: Her er også 2 muligheder:

1. Den ovenfor omtalte assess-sætning i PHREG kan benyttes til et numerisk test, baseret på simulationer, se s. 41
2. **En sværere metode:** Inkluder en tidsafhængig effekt af kovariaten, se mere s. 72ff



*Score residualer (Schoenfeld)

Til hvert tidspunkt for event, f.eks. t_i for individ i ser vi på vedkommendes kovariat x_i og spørger:

Er det "den forventede" person, der dør?

$$x_i - \bar{x}(t_i, \hat{\beta})$$

Disse kumuleres over tid og standardiseres (*standardized score process*) og kan afsløre, om betydningen af x er den samme over tid (svarende til proportional hazards):

Hvis f.eks. folk med høje værdier af kovariaten dør "*for tidligt*" i starten, vil processen starte med at være positiv, for derefter at synke ned mod 0.

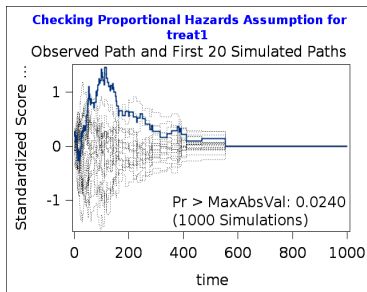


Grafisk check af proportionale rater

Benyt nedenstående sætning i PHREG (se også s. 88):

```
assess ph / resample seed=106165;
```

Tallet i seed er jeres valg....



Bemærkninger næste side



Numerisk check af proportionale rater

Supremum Test for Proportionals Hazards Assumption

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
treat1	1.4529	1000	106165	0.0240

Baseret på 1000 simulationer ser vores proces ud til at være ret aparte ($P=0.024$), så antagelsen om proportional hazards er ikke helt god.

Figuren s. 40 viser, at det i starten er patienter med høje værdier af variablen treat1 (dvs. personer med 1-taller for indikatoren for treat=1, altså de eksperimentelt behandlede), der dør for tidligt. Dette svarer også til, hvad vi har set på overlevelseskurven (s. 20) og kumuleret hazard funktion (s. 37).



Antagelsen om proportionale intensiteter er ikke opfyldt

Det betyder at:

- ▶ Forskellen på behandlingerne er *ikke* er konstant over tid.
- ▶ Hazard-funktionerne har ikke samme form for de to grupper

Hvad gør vi så?

- ▶ Overvejer, om vi har glemt nogle vigtige kovariater:
Fik vi svaret på det rigtige?
- ▶ Benytter tidsafhængig effekt af kovariater
Dette er svært, så vi udskyder det til s. 72ff



Har vi andre kovariater?

Jada, vi har adskillige:

celltype: 1:squamous, 2:smallcell, 3:adeno, 4:other

karno: Karnofsky performance score (0-100, 100 er godt):

- ▶ 10-30: hospitaliseret
- ▶ 40-60: delvist hospitaliseret
- ▶ 70-90: tager vare på sig selv

diagtime: Måneder fra diagnose til randomisering/behandling

age: Personens alder i år

prior: Er der givet tidligere behandling? 0:nej, 1:ja

Vi forsøger os med en model, der (ud over `treat`) tager højde for `celltype` og `karno`, dvs. en model med to kategoriske og en enkelt kvantitativ kovariat.



Regressionsmodeller i overlevelsesanalyse

i daglig tale: **Cox modeller:**

Da hazard-funktioner er positive, $h(t) > 0$, vil $\log(h(t))$ kunne antage alle værdier, og det er derfor naturligt at modellere effekten af kvantitative kovariater (x , f.eks. karno) på denne skala:

$$\log(h(t|x)) = \log(h_0(t)) + \beta x + \dots \text{flere}$$

således at effekten bliver *multiplikativ* på selve hazard-skalaen.

Her er $h_0(t)$ **baseline hazard**, svarende til $x = 0$

(altså en generaliseret form for intercept),

og β er regressionskoefficienten hørende til kovariaten x .

Bemærk, at β er uafhængig af tiden,

og at $\exp(\beta)$ netop er proportionalitetskonstanten k fra s. 32



Hazard-funktioner

For hver kombination af kovariater har vi til hvert tidspunkt en vis umiddelbar rate for død, som har det multiplikative udtryk:

$$h(t|x) = h_0(t) \cdot \exp(\beta x)$$

Hazard funktionen (raten) er altså en funktion af to (eller flere) variable: tiden og kovariaten (kovariaterne):

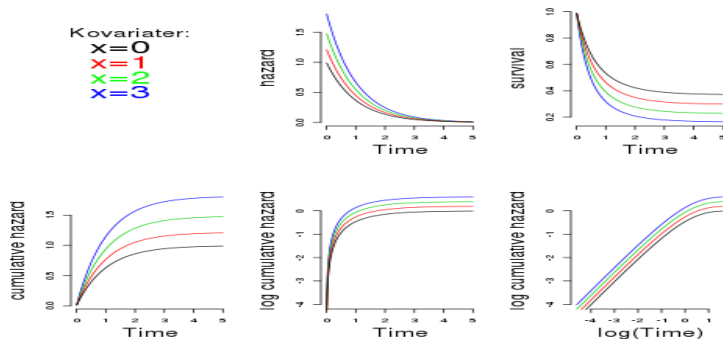
Proportional hazards: er en antagelse om afhængigheden af tiden, nemlig at denne afhængighed er den samme for alle kovariat-værdier, på nær en multiplikativ effekt (en faktor)

Linearitet: er en antagelse om afhængigheden af hver af de kvantitative kovariater, nemlig at der for hver givet tidspunkt er linearitet (på log-skala)



Eksempel på hazard-funktioner

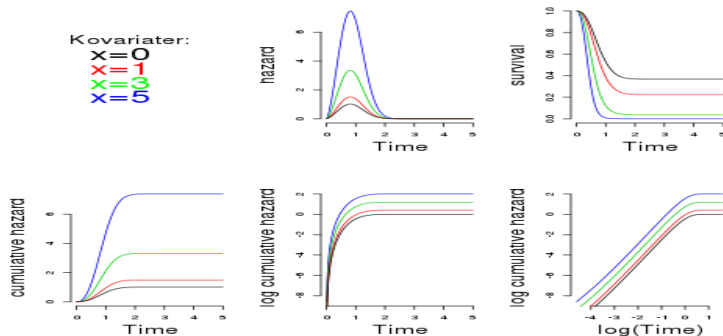
for 4 forskellige ækvidistante værdier (0, 1, 2 og 3) af en kvantitativ kovariat x :



Her er både **proportional hazards** og **linearitet**
 (det første ses ved parallellitet, og det sidste fordi afstanden mellem kurverne er ækvidistant)

Eksempel på hazard-funktioner

for 4 forskellige værdier af en kvantitativ kovariat x :

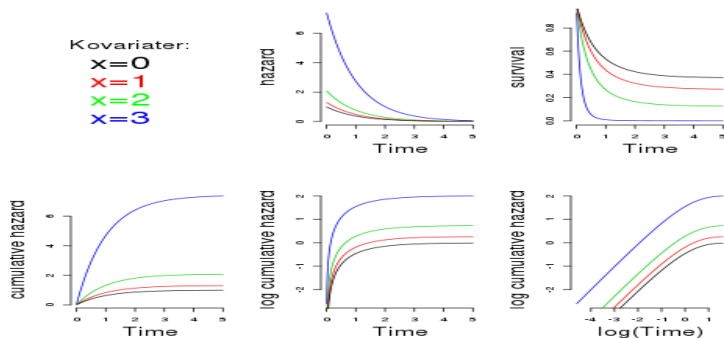


Her er både **proportional hazards** og **linearitet** (det første ses ved parallellitet, og det sidste fordi afstanden mellem kurverne afspejler de ikke-ækvivalente kovariat-værdier)



Eksempel på hazard-funktioner

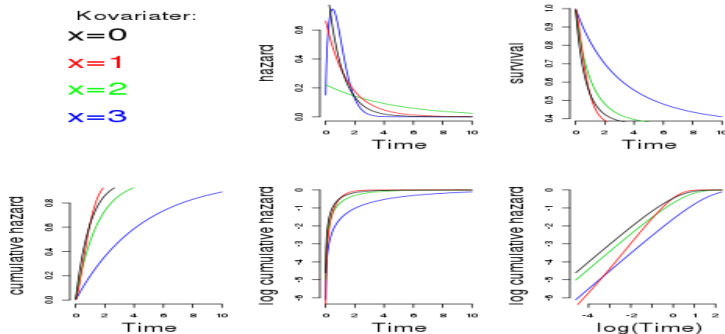
for 4 forskellige ækvidistante værdier (0, 1, 2 og 3) af en kvantitativ kovariat x :



Her er **proportional hazards**, men **ikke** linearitet (kurverne er parallelle, men **ikke** ækvidistante)

Eksempel på hazard-funktioner

for 4 forskellige ækvidistante værdier (0, 1, 2 og 3) af en kvantitativ kovariat x :



Her er *hverken* proportional hazards, *eller* linearitet

Model med 3 kovariater

- ▶ treat: Behandlingen
- ▶ celltype: De 4 cancertyper
- ▶ karno: kvantitativ Karnofsky score

```
proc phreg data=a1;  
  class treat(ref="0") celltype / param=glm;  
  model time*status(0)=treat celltype karno / rl;  
  
  estimate "effekt af 10 karno" karno 10 / exp cl;  
  hazardratio karno / unit=10;  
  hazardratio celltype / diff=all;  
  assess var=(karno) ph / resample seed=106165;  
  baseline out=ud loglogs=logCumrate survival=_ALL_;  
run;
```



Bemærkninger til modelsætningerne s. 50

- ▶ `estimate`-sætning er *velkendt*, her blot for at få effekten af 10 enheder af `karno`.
Denne kan her også skrives som den første af de to `hazardratio`-sætninger
- ▶ `hazardratio celltype / diff=all`:
Her beder vi om parvise sammenligninger mellem de 4 celletyper, se output s. 54
- ▶ `assess`-sætning er udbygget med `var=(karno)`, som checker linearitetsantagelsen for den kvantitative kovariat `karno`, se output s. 65-68



Output fra model med 3 kovariater

Kode s. 50 giver outputtet:

The PHREG Procedure

Type 3 Tests

Effect	DF	Wald	Pr > ChiSq
		Chi-Square	
karno	1	36.2597	<.0001
treat	1	1.6449	0.1997
celltype	3	17.3015	0.0006

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
karno		1	-0.03111	0.00517	36.2597	<.0001
treat	1	1	0.25731	0.20063	1.6449	0.1997
treat	0	0	0	.	.	.
celltype	adeno	1	1.14768	0.29493	15.1425	<.0001
celltype	large	1	0.39296	0.28223	1.9386	0.1638
celltype	smallcel	1	0.81962	0.26881	9.2968	0.0023
celltype	squamous	0	0	.	.	.



Output fra model med 3 kovariater, fortsat

Tilbagetransformerede estimer:

Analysis of Maximum Likelihood Estimates

Parameter		Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
karno		0.969	0.960	0.979	
treat	1	1.293	0.873	1.917	treat 1
treat	0	.	.	.	treat 0
celltype	adeno	3.151	1.768	5.617	celltype adeno
celltype	large	1.481	0.852	2.576	celltype large
celltype	smallcel	2.270	1.340	3.844	celltype smallcel
celltype	squamous	.	.	.	celltype squamous

Se fortolkning s. 55-56



Output fra model med 3 kovariater, fortsat

Output fra hazardratio-sætningerne fra kode s. 50:

Hazard Ratios for karno

Description	Point Estimate	95% Wald Confidence Limits	
karno Unit=10	0.733	0.662	0.811

Hazard Ratios for celltype

Description	Point Estimate	95% Wald Confidence Limits	
celltype adeno vs large	2.127	1.185	3.818
celltype adeno vs smallcel	1.388	0.820	2.351
celltype adeno vs squamous	3.151	1.768	5.617
celltype large vs smallcel	0.653	0.390	1.093
celltype large vs squamous	1.481	0.852	2.576
celltype smallcel vs squamous	2.270	1.340	3.844



Fortolkning af output fra de forrige sider

- ▶ Der er stærkt signifikant negativ effekt af karno-score ($P < 0.0001$)
- ▶ Der er stærkt signifikant forskel på cancertyperne, med adeno og smallcell som de farligste (de har de største positive afvigelser fra referencegruppen squamous)
- ▶ Behandlingseffekten er fortsat ikke signifikant, $P=0.20$

For at forstå selve estimerterne, skal vi se på rate ratioerne (hazard ratios), som det ses s. 53.



Fortolkning af rate ratioer

- ▶ Effekten af 1 enhed på karno-skalaen er et fald i dødsraten på 3.1%, $CI=(2.1\%, 4.0\%)$, se dog også næste side, hvor vi har effekten af 10 enheder
- ▶ Umiddelbart er den eksperimentelle behandling dårlig, med en forøget rate på 29.3%, men CI viser, at det ligeså godt kunne være en formindsket rate (12.7% lavere), eller muligvis helt op til næsten dobbelt så stor rate (91.7% højere)
- ▶ adeno-typen estimeres til at være ca. 3 gange så “farlig” som referencen squamous, dog med betydelig usikkerhed, svarende til faktorer mellem 1.77 og 5.62.



Output fra estimate-sætningen, s. 50

The PHREG Procedure
Estimate

Label	Estimate	Standard Error	z Value	Pr > z	Alpha
effekt af 10 karno	-0.3111	0.05167	-6.02	<.0001	0.05

Label	Lower	Upper	Exponentiated
effekt af 10 karno	-0.4124	-0.2099	0.7326

Label	Exponentiated Lower	Exponentiated Upper
effekt af 10 karno	0.6621	0.8107

Fortolkning:

- Effekten af en øgning på 10 enheder på karno-skalaen er et fald i dødsraten på 26.74%, CI=(18.93%, 33.79%)



Modelcheck

For kategoriske kovariater:

Her drejer det sig om `treat` og `celltype`:

Her skal antagelsen om *Proportional hazards* checkes:

- ▶ ved at sætte kovariaten i `strata` i stedet for i `model`-sætning, for derefter at tegne $\log H(t)$, se det tidligere check af `treat`, s. 37, og det nuværende på s. 59

For kovariaten `celltype` finder vi her figuren på s. 60

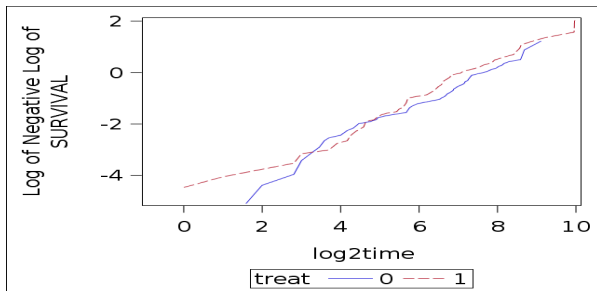
- ▶ benytte PH-option i `assess-statement`, se kode s. 50 med output s. 61 og 62



Modelkontroltegninger for treat

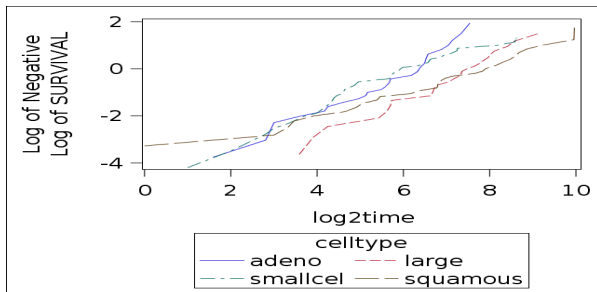
i denne nye model med 3 kovariater.

Da dette er en kategorisk variabel, sætter vi den i stedet i strata-statement for at kunne tegne kumulerede rater, på log-skala, se kode for tilsvarende figur på s. 87.



Modelkontroltegninger for celltype

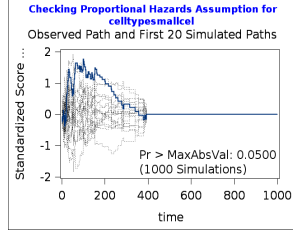
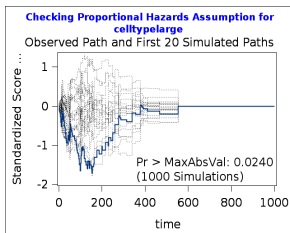
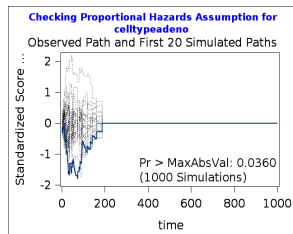
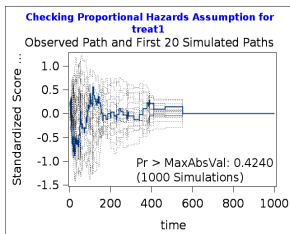
Til brug for denne figur, er det `celltype`, der sættes i `strata`-statement for at kunne tegne kumulerede rater, på log-skala, se tilsvarende kode s. 87



Det ses, at det netop er referencegruppen `squamous`, der opfører sig anderledes end de andre, med lavere hældning



Proportional hazards via *assess*, I



Proportional hazards via *assess*, II

Numeriske tests af modeltilpasning fra *assess*-sætning:

The PHREG Procedure

Supremum Test for Functional Form

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
karno	5.2973	1000	106165	0.3440

Supremum Test for Proportionals Hazards Assumption

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
karno	2.0301	1000	106165	<.0001
treat1	0.8825	1000	106165	0.4240
celltypeadeno	1.7714	1000	106165	0.0360
celltypelarge	1.7124	1000	106165	0.0240
celltypesmallcel	1.7677	1000	106165	0.0500



Bemærkning vedrørende Proportional hazards

Først for kovariaten `treat`:

Det numeriske test giver nu $P=0.42$, dvs.

vi har ikke mere anledning til at betvivle proportionaliteten

Bemærk, at vi tidligere (s. 40-41) havde $P=0.024$, og de proportionale intensiteter er altså opnået ved at tilføje vigtige ekstra kovariater



Bemærkning vedrørende Proportional hazards, II

Og så for kovariaten `celltype`:

Det numeriske test er her 3-delt, idet alle typer sammenlignes til reference-typen `squamous`. Det ses, at alle P-værdierne er signifikante (den ene dog lige på 0.05..., men alligevel).

Her har vi altså et problem, og fra figuren s. 60 ses det, at netop celletypen `squamous` adskiller sig fra de øvrige ved at have et fladere forløb, svarende til en større dødelighed fra start, efterfulgt af lavere dødelighed senere.

Havde vi valgt en anden referencegruppe, havde vi rimeligvis kun set problemer i forhold til sammenligningen med `squamous`.



Modelcheck for kvantitative kovariater

Her drejer det sig kun om karno:

Her skal der checkes to antagelser:

1. *Linearitet* på log(hazard)-skala Her kan man benytte
Kumulerede Martingal residualer:

For hver kovariatværdi x ser man på, i hvor høj grad personer med kovariatværdier $< x$ “levede for længe” (positive værdier) eller “døde for tidligt” (negative værdier).

Fås i SAS ved at skrive `var=(karno)` i en `assess`-sætning, se s. 50

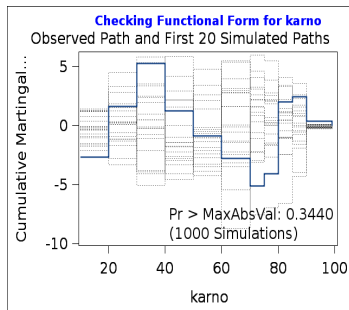
Metoden er dog ikke særligt sensitiv

2. *Proportional hazards*: checkes (næsten) som for kategoriske kovariater, nemlig via `assess`

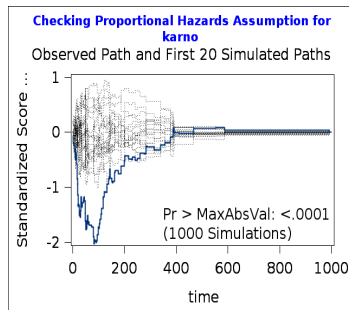


Check af modelforudsætninger for karno

Lineariteten:



Proportional hazards:



Linearitetsantagelsen for karno

- ▶ Figuren til venstre s. 66 og det tilhørende test s. 62 viser, at lineariteten er rimelig ($P=0.34$).
- ▶ Alternative checks af lineariteten kunne være:
 - ▶ Såvel karno som $\log(\text{karno})$ som kovariater
 - ▶ En lineær spline i karno, hvor vi f.eks. tillader knæk ved karno=40 og karno=80:

```
proc phreg data=a1;  
  class treat(ref="0") celltype / param=glm;  
  model time*status(0)=karno karno40 karno80 treat celltype / rl;  
  output out=tegn xbeta=spline;  
  Testlinearitet: TEST karno40=karno80=0;  
run;
```

Se mere s. 90, samt dele af output næste side

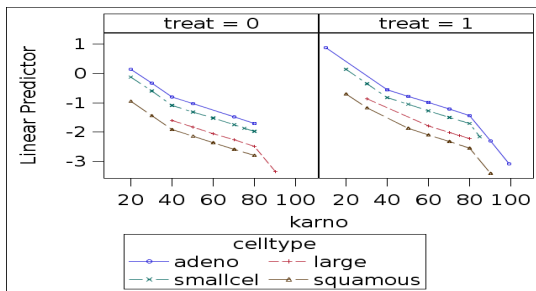


Effekt af karno som lineær spline

Test af *ingen knæk i spline'n*:

Linear Hypotheses Testing Results

	Wald		
Label	Chi-Square	DF	Pr > ChiSq
Testlinearitet	2.6054	2	0.2718



Check af proportional hazards for karno

Testet for proportionale intensiteter for karno (s. 62) viser, at denne antagelse er helt gal ($P < 0.0001$).

Denne manglende proportionalitet ses også på figuren til højre på s. 66, hvor den standardiserede score-process (se s. 39) ses at afvige fra de simulerede.

Dette betyder:

- ▶ Effekten af karno er *ikke* er konstant over tid.
- ▶ Hazard-funktionerne har ikke samme form for alle værdier af karno

Hvad gør vi så??



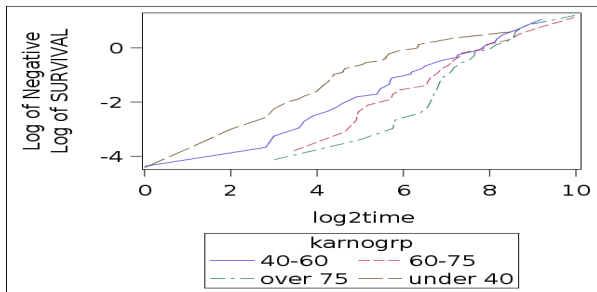
Vi prøver at forstå problemet

- ▶ På figuren s. 66 ses, at de med små værdier af karno i højere grad dør i starten af perioden i forhold til, hvad modellen siger
- ▶ Effekten af karno er i sig selv negativ, således at små værdier af karno altid vil formodes at dø først - de gør det bare endnu mere i starten
- ▶ Det betyder, at effekten af karno er størst i starten, men aftager efterhånden som tiden går



Hvis vi nu grupperer karno-kovariaten

og sætter den ind i strata:



Her ser vi ligeledes, at kurverne afviger en del i starten (svarende til stor effekt af karno), men efterhånden nærmer sig hinanden (svarende til *næsten ingen* effekt)

Tidsafhængig effekt af kovariater

Vi har altså set, at antagelsen om proportionale intensiteter **hazards** ikke holdt for kovariaten karno.

Det betyder, at den prognostiske værdi af karno ændrer sig med tiden, og vi fortolkede det som at effekten af karno er stærkest i starten.

Hvad gør vi så?

Vi kan opdele tiden i intervaller, og estimere en proportionalitetsfaktor for hvert interval for sig.

Vi kan f.eks. opdele tiden efter **kvartilerne for de døde**:
23, 62 og 147.5 dage



*Tidsafhængige kovariater i praksis

- ▶ For at få effekten af 10 karno-enheder direkte, skifter vi karno-variablen ud med $\text{karno10} = \text{karno}/10$, således at 1 enhed af karno10 svarer til 10 enheder af karno.
- ▶ Dernæst indfører vi separate effekter af karno10 ved direkte i PHREG at definere separate kovariater i de forskellige tidsrum:

```
proc phreg data=a1;
  class treat(ref="0") celltype / param=glm;
  model time*status(0)=treat celltype
        karno1 karno2 karno3 karno4 / rl;
  if time<=23 then karno1=karno10; else karno1=0;
  if 23<time<=62 then karno2=karno10; else karno2=0;
  if 62<time<=147.5 then karno3=karno10; else karno3=0;
  if 147.5<time then karno4=karno10; else karno4=0;
  NumCheck: test karno1=karno2=karno3=karno4;
run;
```



Output fra tidsafhængig effekt af karno

Analysis of Maximum Likelihood Estimates

Parameter		Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
treat	1	1.087	0.728	1.623	treat 1
treat	0	.	.	.	treat 0
celltype	adeno	2.978	1.653	5.366	celltype adeno
celltype	large	1.397	0.801	2.437	celltype large
celltype	smallcel	2.412	1.402	4.148	celltype smallcel
celltype	squamous	.	.	.	celltype squamous
karno1		0.598	0.498	0.717	
karno2		0.695	0.581	0.831	
karno3		0.926	0.729	1.175	
karno4		0.968	0.753	1.245	

Linear Hypotheses Testing Results

		Wald		
Label	Chi-Square	DF	Pr > ChiSq	
NumCheck	12.8609	3	0.0049	

Se fortolkning næste side



Fortolkning af tidsafhængig effekt

Effekten af karno aftager med tiden:

Rate-ratioerne nærmer sig 1, ganske som vi fandt det tidligere på kontroltegningen s. 71.

Det kunne have den simple forklaring, at karno var målt ved indgang i studiet.

Vi får et alternativt test for konstant effekt af karno, med $P = 0.0049$, altså (igen) forkastelse.



Kan der være interaktioner med cellype?

Man kunne forestille sig en differentieret effekt af behandlingen for de 4 cellyper, altså interaktionen `treat*celltype` (se kode s. 92):

Type 3 Tests

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
treat	1	0.0146	0.9037
celltype	3	18.8449	0.0003
treat*celltype	3	6.1418	0.1049
karno1	1	29.1285	<.0001
karno2	1	15.4592	<.0001
karno3	1	0.3617	0.5476
karno4	1	0.0206	0.8859

Der er ikke tilstrækkelig evidens for en interaktion ($P=0.10$)



Interaktioner med celltype, fortsat

Estimerer for behandlingen, opdelt efter celletype:

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square
celltype	adeno	1	0.96993	0.45621	4.5201
celltype	large	1	-0.19513	0.39677	0.2419
celltype	smallcel	1	0.37338	0.33907	1.2126
celltype	squamous	0	0	.	.
treat*celltype 1	adeno	1	-0.23554	0.42944	0.3008
treat*celltype 1	large	1	0.38605	0.40013	0.9309
treat*celltype 1	smallcel	1	0.54845	0.32845	2.7882
treat*celltype 1	squamous	1	-0.60258	0.40650	2.1974
treat*celltype 0	adeno	0	0	.	.
treat*celltype 0	large	0	0	.	.
treat*celltype 0	smallcel	0	0	.	.
treat*celltype 0	squamous	0	0	.	.
karno1		1	-0.50791	0.09411	29.1285
karno2		1	-0.36698	0.09334	15.4592
karno3		1	-0.07567	0.12582	0.3617
karno4		1	-0.01926	0.13426	0.0206



Problemstillinger, I: Forsinket indgang

Forsinket indgang = left-truncation

Når vi skal sammenligne behandlinger, kan patienterne naturligvis først tælle med, når de blev randomiseret til en af disse og derfor er en del af *risikogruppen*.

Hvis vi benytter en tidsskala, der hedder *tid siden diagnose*, kan vi have et problem, idet nogle kan være døde efter diagnose og inden randomisering, således at de, der indgår i studiet er selekterede, måske forskelligt i de to behandlingsgrupper. *Hvad gør vi så?*

Vi konstruerer en variabel (f.eks. indgang), der indeholder det tidspunkt, som patienten skal tælle med fra, og så skriver vi f.eks.

```
model time*status(0)=treat celltype / entry=indgang;
```



Problemstillinger, II: Competing risks

Competing risks = konkurrerende dødsårsager/afgangsårsager:

Hvis man f.eks. studerer tid til tilbagefald efter behandling, har man det problem, at folk stadig kan dø, *inden* de får tilbagefald. **Hvad gør man så?**

1. Man analyserer det sammensatte endepunkt: tid til *enten* tilbagefald *eller* død, hvad der så end kommer først
2. Man opfatter død som censureringer

I første tilfælde er der intet problem, men i tilfælde 2 får man problemer:

Man kan *stadig* godt

- Udregne hazard-funktioner, og estimere i Cox-modeller for tilbagefald=recidiv, ved blot at opfatte død som en censurering



Competing risks, fortsat

men man kan *ikke*

- ▶ fortolke overlevelseskurver (Kaplan-Meier), se s. 20 fordi det ikke giver nogen mening at antage, at de censurerede patienter (som jo er døde) ville få recidiv i samme takt som de uncensurerede, se s. 15.
- ▶ beregne kumulerede incidenser, altså sandsynligheder for at opleve recidiv (som vist s. 21)

I stedet kan man evt. benytte de såkaldte **Fine and Gray modeller**, der direkte modellerer kumulative incidenser.

Pas på: Hvis en behandling slår folk ihjel hurtigt, når de jo ikke at få recidiv!



Ofte stillede spørgsmål: FAQ

- ▶ *Hvorfor bruger vi ikke bare logistisk regression af f.eks. 1-års mortaliteten?*

Fordi vi mister en masse information, og fordi vi mister muligheden for at håndtere censureringer

- ▶ *Hvorfor bruger vi ikke bare normalfordelingsmodeller på tiden fra randomisering til død?*

Fordi fordelingen ofte er ganske anderledes end en normalfordeling, og igen fordi vi mister muligheden for at håndtere censureringer

Overlevelsesanalyse er den stærkeste måde at håndtere overlevelsesdata på!



APPENDIX

med SAS-programbidder svarende til nogle af slides

- ▶ Kaplan-Meier kurver, s. 83
- ▶ Kumulerede incidenser, s. 84
- ▶ Log-rank test, s. 85
- ▶ Kumulerede hazards, s. 86
- ▶ Check af proportional hazards, s. 87-88
- ▶ Model med 3 kovariater og modelkontrol, s. 89-90
- ▶ Tidsafhængige kovariater, s. 91
- ▶ Interaktioner, s. 92



Kaplan-Meier kurver i SAS

Slide 20

Figur af 2 Kaplan-Meier overlevelseskurver, svarende til 2 værdier af variablen treat, med konfidensgrænser.

Grupperingen styres af variablen i strata, og bemærk, at selve modellen er tom (ingen kovariater)

```
proc phreg plots(overlay=row cl)=survival data=a1;  
  model time*status(0)= ;  
  strata treat;  
run;
```



Kumulerede incidenser, med figur

Slide 21

```
proc phreg data=a1;  
  model time*status(0)= ;  
  strata treat;  
  baseline out=ud survival=_ALL_;  
run;
```

```
data ud1;  
set ud;
```

```
cif=1-survival;  
log2time=log2(time);  
run;
```

```
proc sgplot data=ud1;  
step Y=cif X=time / group=treat;  
run;
```



Log-Rank test

Slide 24 med PHREG:

```
title 'log-Rank test';  
proc phreg data=a1;  
  class treat(ref="0") / param=glm;  
  model time*status(0)=treat / ties=discrete rl;  
run;
```

eller med LIFETEST:

```
proc lifetest notable plots=(survival(atrisk) logsurv) data=a1;  
  time time*status(0);  
  strata treat;  
run;
```



Figur af kumulerede hazards

Slide 31

Grupperingen styres af variabelen i strata, og bemærk, at selve modellen er tom (ingen kovariater)

```
proc phreg data=a1  
  plots(overlay=row cl)=cumhaz;  
  model time*status(0)= ;  
  strata treat;  
run;
```



Check af Proportional hazards, figur

Slide 37

```
proc phreg plots(overlay=row cl)=(survival cumhaz) data=a1;  
  model time*status(0)= ;  
  strata treat;  
  baseline out=ud loglogs=logCumrate;  
run;  
  
data ud1;  
set ud;  
  
log2time=log2(time); run;  
  
proc sgplot data=ud1;  
series Y=logCumrate X=log2time /  
      group=treat;  
run;
```



Numerisk check af proportionale rater

Slide 40-41

```
proc phreg data=a1;  
  class treat(ref="0") / param=glm;  
  model time*status(0)= treat / ties=discrete rl;  
  assess ph / resample seed=106165;  
run;
```



Model med 3 kovariater

Slide 50

```
proc phreg data=a1;  
  class treat(ref="0") celltype / param=glm;  
  model time*status(0)=karno treat celltype / rl;  
  estimate "effekt af 10 karno" karno 10 / exp cl;  
  assess var=(karno) ph / resample seed=106165;  
  hazardratio celltype / diff=all;  
  hazardratio karno / unit=10;  
run;
```



Kovariat modelleret som spline

Slide 67-68

Definitionerne af de to ekstra kovariater er

```
karno40=(karno>40)*(karno-40);  
karno80=(karno>80)*(karno-80);
```

og de ekstra sætninger i koden har følgende betydning:

- ▶ `output out=tegn xbeta=spline;`
Variablen `spline` i datasættet `tegn` indeholder den estimerede effekt af `karno`, altså den knækkede linie, for hver af de indgående grupper. Disse kan så efterfølgende tegnes vha f.eks. `sgplot`
- ▶ Testlinearitet: `TEST karno40=karno80=0;`
Vi tester, om begge de to ekstra led kan undværes samtidig, altså om effekten af `karno` kan beskrives ved simpel linearitet



Tidsafhængige kovariater

Slide 72ff

```
proc phreg data=a1;
  class treat(ref="0") celltype / param=glm;
  model time*status(0)=treat celltype
        karno1 karno2 karno3 karno4 / rl;
  if    time<=23 then karno1=karno10; else karno1=0;
  if 23<time<=62 then karno2=karno10; else karno2=0;
  if 62<time<=147.5 then karno3=karno10; else karno3=0;
  if 147.5<time      then karno4=karno10; else karno4=0;
NumCheck: test karno1=karno2=karno3=karno4;
run;
```



Interaktion

i model med tidsafhængige kovariater.

Slide 76

```
proc phreg data=a1;
  class treat(ref="0") celltype / param=glm;
  model time*status(0)=treat celltype treat*celltype
        karno1 karno2 karno3 karno4 / rl;
  if time<=23 then karno1=karno10; else karno1=0;
  if 23<time<=62 then karno2=karno10; else karno2=0;
  if 62<time<=147.5 then karno3=karno10; else karno3=0;
  if 147.5<time then karno4=karno10; else karno4=0;
  NumCheck: test karno1=karno2=karno3=karno4;
run;
```

