

Opgavebesvarelse vedr. overlevelsesanalyse

Opgaven går ud på at vurdere effekten af azathioprine på overlevelsen hos 216 patienter med primær biliær cirrhose (PBC), ref. Christensen et al. (1985).

Data stammer fra et klinisk forsøg (clinical trial), hvor de 216 patienter er randomiseret til behandling med azathioprine eller placebo, hvorefter de følges over tid indtil død eller censurering. Variablene er:

dead: Indikator for død (0: Censurering, 1: Død)

months: Antal måneder fra randomisering til udgang (død eller censurering)

azathiop: Behandling (0: Placebo, 1: Azathioprine)

bilirubin: Serum bilirubin ($\mu\text{mol/L}$) målt ved baseline

Indlæsning af data er foretaget ved

```
DATA PBCdata;
  INFILE "http://staff.pubhealth.ku.dk/~lts/basal/data/pbc.txt"
        URL /* fordi filen læses direkte fra Lenes hjemmeside */
        FIRSTOBS=2; /* fordi data skal indlæses fra og med linje 2 */
  INPUT Dead Months Azathiop Bilirubin;

  *- Dummy-variabel Indgangstid tilføjes, så programmets opsætning    -*;
  *- også kan bruges, hvis der faktisk er forsinket indgang i studiet -*;
  Indgangstid = -1;

  LABEL
    Dead = "Død (0=censur, 1=død)"
    Months = "Måneder fra randomisering"
    Azathiop = "Azathioprine (0~Placebo, 1~Azathioprine)"
    Bilirubin = "Baseline serum bilirubin"
  ;
RUN;
```

Dummy-variablen *Indgangstid* er tilføjet, for at de senere viste programbidder også skal kunne bruges til inspiration i tilfælde af, at man har data med forsinket indgang, fordi når man vælger en relevant tidsskala, har man meget ofte forsinket indgang i den tidsskala...

Stort set alle de programbidder, der er anvendt i denne opgaveløsning, kan også findes i programeksemplet *bagest* i forelæsningsnoterne.

1. *Da der er tale om et randomiseret studie, vurderes behandlingseffekten uden inddragelse af diverse kovariater.*
Vurder effekten af azathioprine ud fra

- (a) *et test for, om der er forskel på overlevelsen i de to grupper*

Vi skal se på tid til død for de to grupper under hensyntagen til, at der forekommer censureringer (*dead*=0):

```

PROC PHREG DATA=PBCdata;
  MODEL months*dead(0) = azathiop
    / ENTRY=Indgangstid TIES=DISCRETE RL;
RUN;

```

Vi finder nedenstående output, der ikke viser nogen signifikant forskel på de to grupper ($P=0,4936$ for log-rank test, fundet under Testing Global Null Hypothesis: $BETA=0$ ud for Score). Hazard ratio for de azathioprine-behandlede ($azathiop=1$) i forhold til de placebo-behandlede ($azathiop=0$) estimeres til 0,873 ($=\exp(-0,13578)$), med konfidensinterval (0,592; 1,288).

The PHREG Procedure

Model Information

Data Set	WORK.PBCDATA	
Entry Time Variable	Indgangstid	
Dependent Variable	Months	Måneder fra randomisering
Censoring Variable	Dead	Død (0~censur, 1~død)
Censoring Value(s)	0	
Ties Handling	DISCRETE	
Number of Observations Read	216	
Number of Observations Used	216	

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
216	103	113	52.31

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	961.641	961.174
AIC	961.641	963.174
SBC	961.641	965.809

Testing Global Null Hypothesis: $BETA=0$

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	0.4671	1	0.4943
Score	0.4687	1	0.4936
Wald	0.4680	1	0.4939

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
Azathiop	1	-0.13578	0.19849	0.4680	0.4939

Analysis of Maximum Likelihood Estimates

Parameter	Hazard Ratio	95% Hazard Ratio Confidence Limits
Azathiop	0.873	0.592 1.288

Hvis man i SAS programmet har defineret **azathiop** som en **CLASS** variabel i **PHREG**:

```
PROC PHREG DATA=pbcr;
CLASS azathiop;
MODEL months*dead(0) = azathiop / TIES=DISCRETE RL;
RUN;
```

vil estimatet for behandlingseffekten vende den “forkerte vej”, fordi SAS som standard vælger den højeste værdi som reference, så den azathioprine-behandlede gruppe bliver referencegruppen. Det giver en hazard ratio for de placebo-behandlede (**azathiop**=0) i forhold til de azathioprine-behandlede (**azathiop**=1) på $1,145 = \exp(0,136)$, med konfidensinterval (0,776; 1,690). I så fald bliver man nødt til selv at tage reciprokverdier af alt sammen for at beregne hazard ratio for de azathioprine-behandlede i relation til de placebo-behandlede ($1/1,145 = 0,873$, med konfidensintervallet $(1/1,690; 1/0,776) = (0,592; 1,289)$).

Hvis man vil bruge **CLASS**, så kan det anbefales at specificere referenceværdien i sit **CLASS**-statement. Det gøres ved eksempelvis at skrive **CLASS azathiop(REF="0")**; . Det kan også anbefales at få SAS til at bruge samme præsentation som i **PROC GLM** ved at bruge option **PARAM=GLM**, så outputtet kommer til at ligne det, man er vant til at se, og man dermed umiddelbart kan se, hvad SAS har brugt som referencegruppe. Det hele gøres ved at skrive “**CLASS azathiop(REF="0") / PARAM=GLM;**” i stedet for “**CLASS azathiop;**” i ovenstående programbid.

(b) *Kaplan-Meier kurver*

Nu vurderes forskellen på de to grupper ved hjælp af Kaplan-Meier kurver, som er nonparametriske estimater for overlevelseskurverne (nonparametrisk = ingen antagelser om fordelingen af dødstidspunkterne). De er her tegnet med punktvis konfidensgrænser, fordi det oftest er nødvendigt at have konfidensgrænserne med for at kunne vurdere, om der er klar forskel på grupperne

```
ODS GRAPHICS ON;
PROC PHREG DATA=PBCdata PLOTS(OVERLAY=ROW CL)=S;
  MODEL months*Dead(0) = / ENTRY=Indgangstid;
  STRATA Azathiop;
  BASELINE OUT=km
    SURVIVAL=KMcurves LOWER=LowerBound UPPER=UpperBound
    / METHOD=PL CLTYPE=LOGLOG;
RUN;

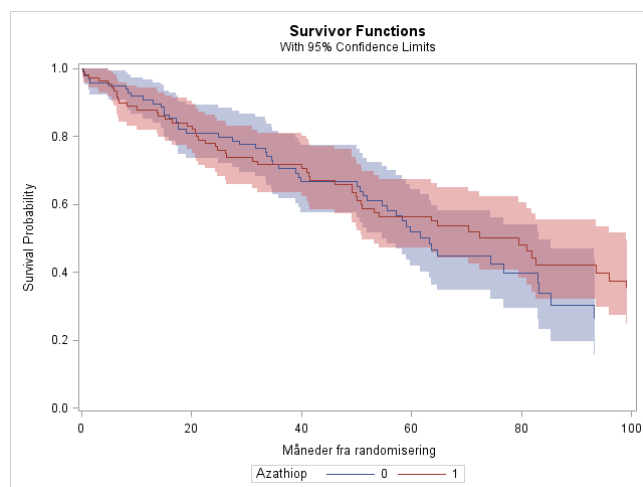
*- Til plot af Kaplan-Meier med konfidensgrænser, -*;
*- hvor man selv styrer akserne.                                -*;
DATA KMmCL; SET km;
  LABEL Curve = "Kaplan-Meier";
  IF azathiop=0 THEN DO;
    type=1; curve=kmcurves; OUTPUT;
    type=2; curve=lowerbound; OUTPUT;
    type=3; curve=upperbound; OUTPUT;
  END;
```

```

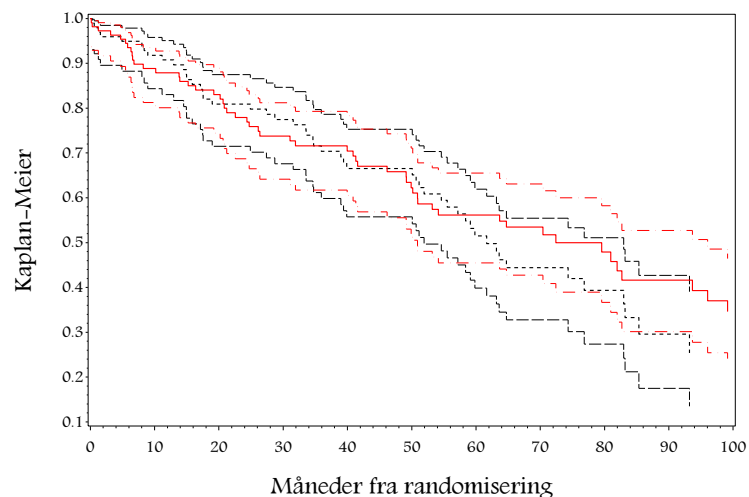
IF azathiop=1 THEN DO;
    type=4; curve=kmcurves; OUTPUT;
    type=5; curve=lowerbound; OUTPUT;
    type=6; curve=upperbound; OUTPUT;
END;
RUN;
SYMBOL1 R=1 C=BLACK V=NONE I=STEPLJ L=2 W=5 MODE=INCLUDE;
SYMBOL2 R=2 C=BLACK V=NONE I=STEPLJ L=4 MODE=INCLUDE;
SYMBOL3 R=1 C=RED V=NONE I=STEPLJ L=1 W=5 MODE=INCLUDE;
SYMBOL4 R=2 C=RED V=NONE I=STEPLJ L=41 MODE=INCLUDE;
AXIS1 LABEL=(H=2) VALUE=(H=1.75);
AXIS2 LABEL=(A=90 H=2) VALUE=(H=1.75);
PROC GPGLOT DATA=KMmCL;
    PLOT curve*months=type / HAXIS=AXIS1 VAXIS=AXIS2 NOLEGEND;
RUN; QUIT;

```

ODS GRAPHICS plot



Plot lavet med den specificerede SAS kode



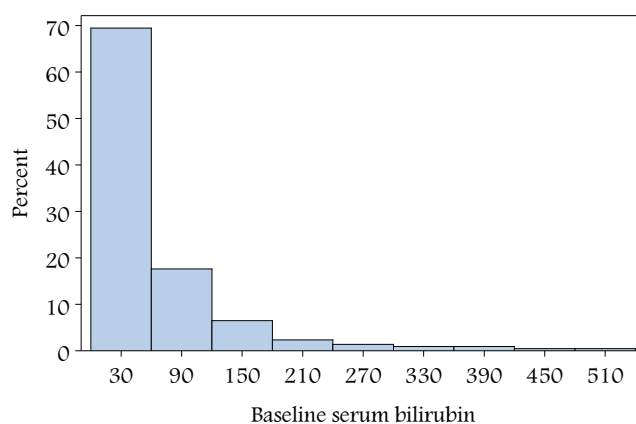
Vi ser, at der er adskillige overkrydsninger mellem de to overlevelseskurver og ingen iøjnefaldende forskelle undtagen, at de måske divergerer lidt mod slut. Kurverne giver ikke anledning til at betvivle rimeligheden af det ovenstående log-rank test i betragtning af de brede konfidensgrænser.

2. Serum bilirubin niveauet er generelt en meget stærk prognostisk faktor for patienter med leversygdom. Det er derfor relevant at undersøge betydningen af serum bilirubin for prognosen hos patienter med primær biliær cirrhose. Vurder først om serum bilirubin bør transformeres

- (a) ved at se på fordelingen af serum bilirubin – er der ekstreme observationer?

Et histogram over alle værdierne set for bilirubin viser tydeligt, at der er tale om en meget skæv fordeling:

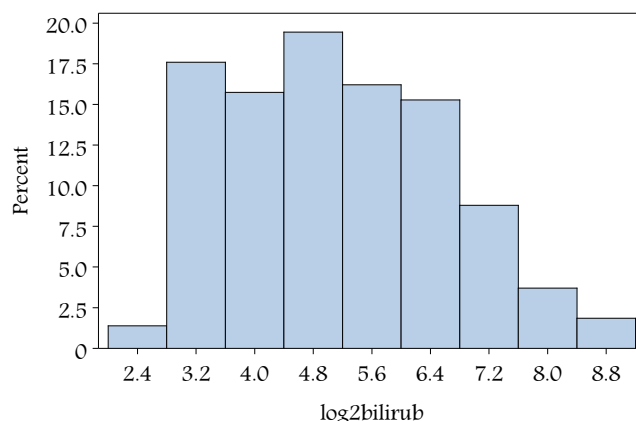
```
PROC UNIVARIATE DATA=PBCdata;  
  VAR bilirubin;  
  HISTOGRAM / HEIGHT=4;  
RUN;
```



Denne skævhed er ikke noget problem i sig selv, da der ikke er nogetsomhelst krav til fordelingen af en kovariat. **Men:** I en så skæv fordeling kan der være meget indflydelsesrige observationer blandt de få, meget høje værdier, og det kan derfor være meget ustabil at bibeholde denne skala. Logaritmetransformation kan bruges til at “trække” ekstremt høje værdier ind mod midten.

Vi ser derfor også på en logaritmetransformeret bilirubin. Her er valgt logaritmen med grundtal 2:

```
DATA PBCdata;  
  SET PBCdata;  
  log2bilirub=LOG2(bilirubin);  
RUN;  
PROC UNIVARIATE DATA=PBCdata;  
  VAR log2bilirub;  
  HISTOGRAM / HEIGHT=4;  
RUN;
```



Her ser vi en meget mere jævn fordeling uden meget ekstreme værdier. Det er derfor relevant at undersøge, om utransformeret bilirubin eller logaritmetransformeret bilirubin er mest relevant som prognostisk variabel – for det vigtigste er jo, at det er en passende model for sammenhængen mellem bilirubin og dødsraten.

- (b) *ved at inkludere serum bilirubin både utransformeret og passende transformeret på samme tid*

Dette kan gøres med følgende SAS kode, som også undersøger den estimerede betydning af den mindst mulige og den størst mulige fordobling inden for den variationsbredde i bilirubin, som vi har i vores data:

```
PROC PHREG DATA=PBCdata;
  MODEL months*dead(0) = bilirubin log2bilirub
    / ENTRY=Indgangstid RL;
  ESTIMATE "Mindst dbl"  Bilirubin 5      Log2Bilirub 1 / EXP ;
  ESTIMATE "Størst dbl"  Bilirubin 264   Log2Bilirub 1 / EXP ;
  Begge: TEST  Bilirubin=log2Bilirub=0;
RUN;
```

Vi finder nedenstående output, der klart viser, 1) at den logaritmetransformerede version af bilirubin er den med stærkest statistisk signifikans, mens den utransformerede kan udelades, og 2) at sammenhængen flader en smule *mere* ud end logaritme-transformationen (effekten af en fordobling er *mindre* i den høje ende, selv om det drejer sig om mere end 50 gange så mange enheder for bilirubin):

The PHREG Procedure

	Model Information	
Data Set	WORK.PBCDATA	
Entry Time Variable	Indgangstid	
Dependent Variable	Months	Måneder fra randomisering
Censoring Variable	Dead	Død (0~censur, 1~død)
Censoring Value(s)	0	
Ties Handling	BRESLOW	
Number of Observations Read	216	
Number of Observations Used	216	

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
216	103	113	52.31

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	965.850	874.486
AIC	965.850	878.486
SBC	965.850	883.756

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	91.3632	2	<.0001
Score	115.4179	2	<.0001
Wald	83.3765	2	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Bilirubin	1	-0.00184	0.00195	0.8910	0.3452
log2bilirub	1	0.80690	0.13978	33.3248	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
Bilirubin	0.998	0.994 1.002	Baseline serum bilirubin
log2bilirub	2.241	1.704 2.947	

Linear Hypotheses Testing Results

Label	Chi-Square	DF	Pr > ChiSq
Begge	83.3765	2	<.0001

Label	Estimate	Standard Error	z Value	Pr > z	Exponentiated
Mindst dbl	0.7977	0.1317	6.06	<.0001	2.2204

Label	Estimate	Standard Error	z Value	Pr > z	Exponentiated
Størst dbl	0.3207	0.4056	0.79	0.4291	1.3781

- (c) *(hvis du har mod på det og har tid til det) ved at modellere effekten af serum bilirubin ved hjælp af en lineær spline (se hvordan sidst i forelæsningsnoterne)*

Dette er svært... Vi bruger programbidder fra bagest i forelæsningsnoterne – vores forklarende variabel hedder oven i købet tilfældigvis næsten det samme.

Vi starter med at finde kvartilerne (25, 50 og 75 percentilerne) blandt de døde for at bruge dem som knækpunkter. Dette gøres ved hjælp af UNIVARIATE, der giver værdierne henholdsvis 30, 54 og 125 $\mu\text{mol/L}$. For at undgå senere tastefejl er kvartilerne gemt direkte som variablene xpctl25, xpctl50 og xpctl75 i et SAS datasæt pctldata, som kan bruges i det efterfølgende DATA-step, hvor spline dummy-variablene bliver tilføjet til datasættet PBCdata:

```

*- Find relevante percentiler blandt de døde.          -*;
*- Percentilerne gemmes direkte på datasættet PCTLDATA -*;
*- som xpctl5, xpctl25, xpctl50, xpctl75 og xpctl95     -*;
*- 33.3 og 66.7 percentiler tages med til senere brug -*;
PROC UNIVARIATE DATA=PBCdata PCTLDEF=3 NOPRINT; *- NOPRINT ~ Intet output -*;
  WHERE dead=1;
  VAR bilirubin;
  OUTPUT OUT=pctldata PCTLPRE=xpctl PCTLPTS=5 25 50 75 95 33.3 66.7;
RUN;
*- Beregne hjælpevariable, der skal bruges i PHREG -*;
DATA PBCdata;
  SET PBCdata;
  IF _N_=1 THEN SET pctldata; *- lægger percentilerne ind i PBCdata datasættet -*;
  *- Beregning af hjælpevariablene med den "smarte" kode -*;
  IF NOT MISSING(bilirubin) THEN DO;
    b_u25pctl=MIN(bilirubin-xpctl25,0);
    b_25_50pctl=MIN(MAX(xpctl25-xpctl50,bilirubin-xpctl50),0);
    b_50_75pctl=MAX(MIN(xpctl75-xpctl50,bilirubin-xpctl50),0);
    b_o75pctl=MAX(bilirubin-xpctl75,0);
  END;
RUN;
*- Estimation og test -*;
PROC PHREG DATA=PBCdata
  OUTEST=parms /* gemmer parameterestimerne på datasættet PARMS
                i variable med de samme navne som de tilsvarende
                variable i analysen (f.eks. b_u25pctl) */ ;
  MODEL months*dead(0) = b_u25pctl b_25_50pctl b_50_75pctl b_o75pctl
    / ENTRY=Indgangstid RL;
  TestLine: TEST b_u25pctl = b_25_50pctl = b_50_75pctl = b_o75pctl;
RUN;

```

Output fra PHREG viser, at der ikke er en lineær sammenhæng med baseline serum bilirubin! Testet for linearitet (TestLine) giver $p < .0001$.

The PHREG Procedure

Model Information	
Data Set	WORK.PBCDATA
Entry Time Variable	Indgangstid
Dependent Variable	Months
Censoring Variable	Dead
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	216
Number of Observations Used	216

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
216	103	113	52.31

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
	Without	With
Criterion	Covariates	Covariates
-2 LOG L	965.850	875.235
AIC	965.850	883.235
SBC	965.850	893.774

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	90.6144	4	<.0001
Score	115.8288	4	<.0001
Wald	80.9404	4	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
b_u25pctl	1	0.07888	0.02215	12.6786	0.0004
b_25_50pctl	1	0.01541	0.01680	0.8419	0.3589
b_50_75pctl	1	0.01315	0.00568	5.3528	0.0207
b_o75pctl	1	0.00282	0.00156	3.2653	0.0708

Analysis of Maximum Likelihood Estimates			
Parameter	Hazard Ratio	95% Hazard Ratio Confidence Limits	
b_u25pctl	1.082	1.036 1.130	
b_25_50pctl	1.016	0.983 1.050	
b_50_75pctl	1.013	1.002 1.025	
b_o75pctl	1.003	1.000 1.006	

Linear Hypotheses Testing Results			
Wald			
Label	Chi-Square	DF	Pr > ChiSq
TestLine	32.3930	3	<.0001

Det numeriske test bør dog aldrig stå alene – man skal altid tegne den lineære spline. Hvis p -værdien er lav, men det er sammen med en meget savtakket kurve omkring en generelt lineær trend, så kan en lineær sammenhæng stadig være den mest relevante model at bruge.

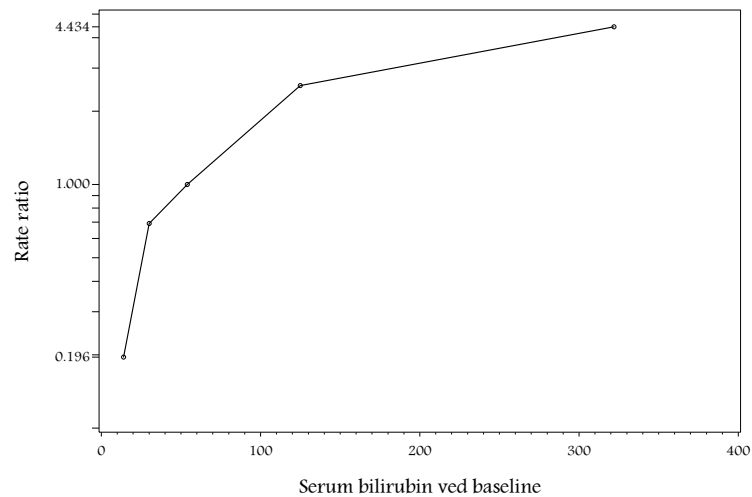
```

*- Beregning af punkter til plot af lineær spline -*;
DATA plot;
  MERGE pctldata parms;
  LABEL
    bili = "Serum bilirubin ved baseline"
    rr = "Rate ratio"
  ;
  *- Man behøver kun beregne støttepunkter i enderne og ved knækkene -*;
  DO bili=xpctl15, xpctl25, xpctl50, xpctl75, xpctl95;
    *- De nye dummy-variable skal hedde noget lidt anderledes, fordi -*;
    *- de "gamle" navne jo "er brugt" til PHREG parameterestimerne -*;
    b_u25=MIN(bili-Xpctl25,0);
    b_25_50=MIN(MAX(Xpctl25-Xpctl50,bili-Xpctl50),0);
    b_50_75=MAX(MIN(Xpctl75-Xpctl50,bili-Xpctl50),0);
    b_o75=MAX(bili-Xpctl75,0);
  
```

```

*- Beregning af det prognostiske index -*;
pi = b_u25pctl*b_u25 + b_25_50pctl*b_25_50
    + b_50_75pctl*b_50_75 + b_o75pctl*b_o75;
*- Beregning af rate ratio (hazard ratio) -*;      rr = EXP(pi);
*- Læg observationen ud på PLOT datasættet -*;      OUTPUT;
END;
RUN;
*- Tegning af lineær spline -*;
AXIS1 LABEL=(H=2)          VALUE=(H=1.75);
AXIS22 LABEL=(H=2 A=90) VALUE=(H=1.75)
      LOGBASE=10 INTERVAL=PARTIAL MAJOR=(H=1.5) MINOR=(H=1.5);
SYMBOL1 C=BLACK V=CIRCLE I=JOIN L=1 MODE=INCLUDE;
PROC GLOT DATA=plot;
  PLOT rr*bili / HAXIS=AXIS1 VAXIS=AXIS22;
  GOPTIONS DEVICE=; *- SKAL med for SAS Enterprise, er ellers unødvendig .*;
RUN; QUIT;

```



Hvis man ikke havde et godt bud på en transformation af serum bilirubin, kunne man i stedet præsentere de estimerede rate ratioer for de forskellige intervaller. Hvis man har valgt en passende skalering, kan man aflæse de interval-specifikke rate ratioer direkte under **Hazard Ratio**. Her ville det have været passende, hvis bilirubin divideret med 10 havde været brugt som variabel fra start af. Vi dog også beregne effekten pr 10 $\mu\text{mol/L}$ ud fra ovenstående output ved at gange parameterestimerne (**Parameter estimate**) med 10, og så tage anti-logaritmen $\exp$ bagefter:

- i. For serum bilirubin under 30 $\mu\text{mol/L}$ øges dødsraten med 120% (dvs. med en faktor på 2.2) for hver 10 $\mu\text{mol/L}$ forskel.
- ii. For serum bilirubin mellem 30 og 54 $\mu\text{mol/L}$ øges dødsraten med ca. 16,7% for hver 10 $\mu\text{mol/L}$ forskel.
- iii. For serum bilirubin mellem 54 og 125 $\mu\text{mol/L}$ øges dødsraten med ca. 14,1% for hver 10 $\mu\text{mol/L}$ forskel.
- iv. For serum bilirubin over 125 $\mu\text{mol/L}$ øges dødsraten med ca. 2,9% for hver 10 $\mu\text{mol/L}$ forskel.

Der er dog ingen grund til at bruge de interval-specifikke effekter i den aktuelle situation, for her kan en logaritmetransformation løse problemet. (Hvis man kører hele proceduren med en lineær spline igennem med `log2bilirub` som variabel i stedet for `bilirubin`, finder man en næsten perfekt lineær sammenhæng.)

3. *Estimer baseline serum bilirubins betydning for dødsraten.*

Alt under spørgsmål 2 viser, at vi bør se på logaritmetransformeret bilirubin. Vi vælger `log2`-transformeret bilirubin, fordi så får vi i outputtet vist effekten af en fordobling af serum bilirubin direkte som effektestimatet.

```
PROC PHREG DATA=PBCdata;
  MODEL months*dead(0) = log2bilirub
    / RL ENTRY=Indgangstid ;
RUN;
```

The PHREG Procedure

Model Information	
Data Set	WORK.PBCDATA
Entry Time Variable	Indgangstid
Dependent Variable	Months Måneder fra randomisering
Censoring Variable	Dead Død (0~censur, 1~død)
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	216
Number of Observations Used	216

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
216	103	113	52.31

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	965.850	875.447
AIC	965.850	877.447
SBC	965.850	880.082

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	90.4025	1	<.0001
Score	100.0062	1	<.0001
Wald	89.6345	1	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
log2bilirub	1	0.69366	0.07327	89.6345	<.0001

Analysis of Maximum Likelihood Estimates			
Parameter	Hazard Ratio	95% Hazard Ratio Confidence Limits	
log2bilirub	2.001	1.733 2.310	

En dobbelt så høj baseline serum bilirubin giver en 2,0 (95% kofidensinterval: 1,7 til 2,3) gange større dødsrate, altså en fordobling af dødsraten.

4. *Vurder effekten af azathioprine justeret for effekten af bilirubin.*
 Nu inddrager vi behandlingsvariablen, sammen med den logaritmetransformerede bilirubin:

```
PROC PHREG DATA=PBCdata;
  MODEL months*dead(0) = log2bilirub Azathiop
    / RL ENTRY=Indgangstid ;
RUN;
```

The PHREG Procedure

Model Information	
Data Set	WORK.PBCDATA
Entry Time Variable	Indgangstid
Dependent Variable	Months
Censoring Variable	Dead
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	216
Number of Observations Used	216

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
216	103	113	52.31

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	965.850	869.133
AIC	965.850	873.133
SBC	965.850	878.403

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	96.7161	2	<.0001
Score	103.5930	2	<.0001
Wald	92.2362	2	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
log2bilirub	1	0.74246	0.07731	92.2356	<.0001
Azathiop	1	-0.52028	0.20607	6.3743	0.0116

Analysis of Maximum Likelihood Estimates			
	Hazard	95% Hazard Ratio	
Parameter	Ratio	Confidence Limits	
log2bilirub	2.101	1.806	2.445
Azathiop	0.594	0.397	0.890

Her finder vi en *signifikant* effekt af azathioprine med $p=0,0116$. Den estimerede effekt af behandling med azathioprine bliver her en dødsrate på kun 59,4% (95% konfidensinterval fra 39,7% til 89,0%) af dødsraten blandt placebo-behandlede med samme baseline serum bilirubin.

Det var lidt af en forskel fra den ujusterede analyse...!

5. *Hvordan ville du afrapportere resultaterne i en artikel?*

Her bør man tage i betragtning,

- (a) at Cochrane centeret har vist, at justerede analyser i gennemsnit overestimerer de faktiske behandlingseffekter (sikkert fordi folk tror på behandlingseffekten og derfor tror, at nu har de fundet den “rigtige” model, når den estimerede effekt tilfældigvis bliver stærkere).
- (b) at randomisering ikke sikrer, at man får helt ens fordelinger af confoundere, kun at der ikke er nogen systematiske forskelle til fordel for en bestemt behandlingsgruppe
- (c) hvad man selv skrev i protokollen (så man undgår fisketure)

6. *Lav modelkontrol for Cox modellen, primært for effekten af den randomiserede behandling, men hvis der er tid, også den grafiske modelkontrol for effekten af baseline serum bilirubin. Modelkontrollen foretages som*

(a) *grafisk modelkontrol*

Vi skal her se på antagelsen om proportionale intensiteter, hvilket gøres ved at tegne “log minus log” af overlevelseskurverne. Hvis der er proportionale intensiteter, skal disse kurver være parallelle.

Vi ser først på behandlingseffekten, hvor vi bruger nedenstående SAS kode. Ud over at gemme “log minus log” af overlevelseskurverne på et datasæt ved brug af LOGLOGS= option i BASELINE, tegnes også den kumulerede rate ved at aktivere ODS GRAPHICS og bruge option PLOTS(OVERLAY=ROW CL)=CUMHAZ i PHREG-kaldet. Den kumulerede rate skal bruges til at finde passende opdelingstidspunkter for det numeriske test under punkt 6 b.

```

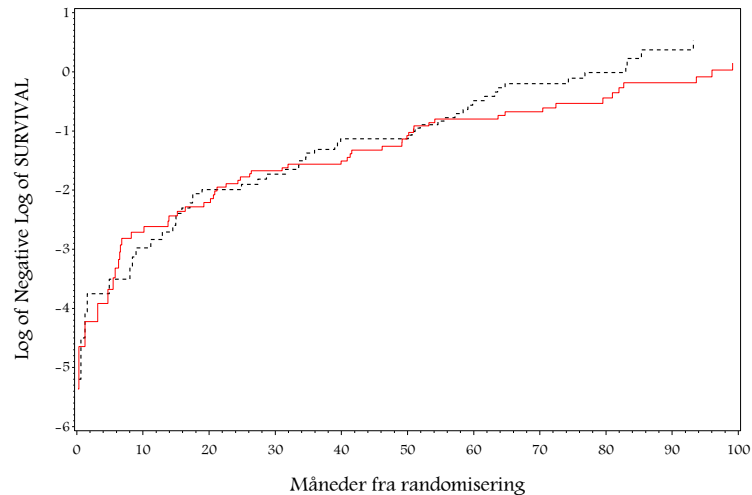
ODS GRAPHIC ON;
PROC PHREG DATA=PBCdata PLOTS(OVERLAY=ROW CL)=CUMHAZ;
  MODEL months*dead(0) = log2bilirub
    / RL ENTRY=Indgangstid ;
  STRATA azathiop;
  BASELINE OUT=TjekA LOGLOGS=LCumRate / METHOD=CH;
RUN;
AXIS1 LABEL=(H=2)          VALUE=(H=1.75);
AXIS2 LABEL=(A=90 H=2)    VALUE=(H=1.75);

```

```

SYMBOL1 C=BLACK V=NONE I=STEPLJ L=2 MODE=INCLUDE;
SYMBOL2 C=RED V=NONE I=STEPLJ L=1 MODE=INCLUDE;
  *- Almindelig lineær tidsakse -*;
PROC GPLOT DATA=tjekA;
  PLOT LCumRate*months=azathiop / HAXIS=AXIS1 VAXIS=AXIS2 NOLEGEND;
RUN; QUIT;

```



De to kurver ses at krydse over hinanden flere gange i det første $1\frac{1}{2}$ år. I perioden fra $1\frac{1}{2}$ til 5 år er der lidt variende *lodret* afstand mellem kurverne, mens kurverne og bliver mere og mere forskellige efter 5 år.

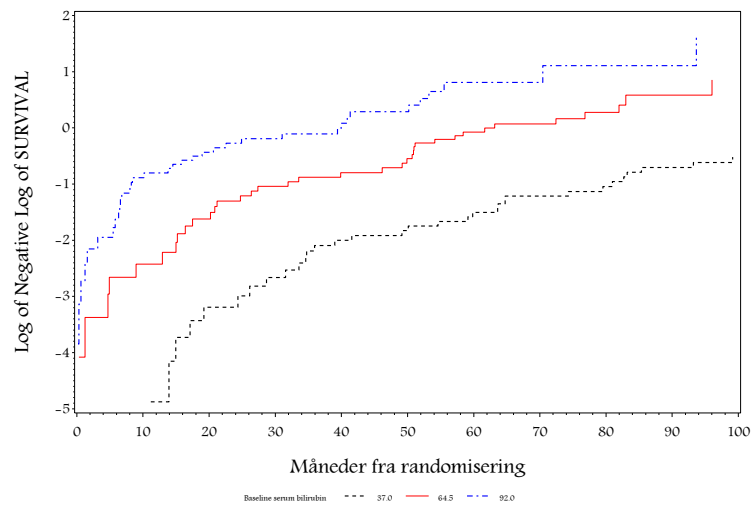
Ofte kan det være næsten umuligt at bedømme den *lodrette* afstand mellem kurverne i den først del af perioden, fordi raterne ofte er særligt høje i begyndelsen, så kurverne er meget stejle. I så fald er man nødt til også at tegne med eksempelvis en logaritmisk tidsakse. Det er ikke nødvendigt her, fordi kurverne krydser frem og tilbage over hinanden, og overkrydsningerne kan stadig ses, selvom kurverne er relativt stejle.

En grafisk vurdering for bilirubin kan foretages ved følgende kode:

```

PROC PHREG DATA=PBCdata;
  MODEL months*dead(0) = log2bilirub azathiop
    / RL ENTRY=Indgangstid ;
  STRATA bilirubin (37, 92); *- opdeler ved 33.3 og 66.7 percentilerne -*;
  BASELINE OUT=tjekB LOGLOGS=LCumRate / METHOD=CH;
RUN;
AXIS1 LABEL=(H=2) VALUE=(H=1.75);
AXIS2 LABEL=(A=90 H=2) VALUE=(H=1.75);
SYMBOL1 C=BLACK V=NONE I=STEPLJ L=2 MODE=INCLUDE;
SYMBOL2 C=RED V=NONE I=STEPLJ L=1 MODE=INCLUDE R=1;
SYMBOL3 C=BLUE V=NONE I=STEPLJ L=41 MODE=INCLUDE;
  *- Almindelig lineær tidsakse -*;
PROC GPLOT DATA=tjekB;
  PLOT LCumRate*months=bilirubin / HAXIS=AXIS1 VAXIS=AXIS2;
RUN; QUIT;

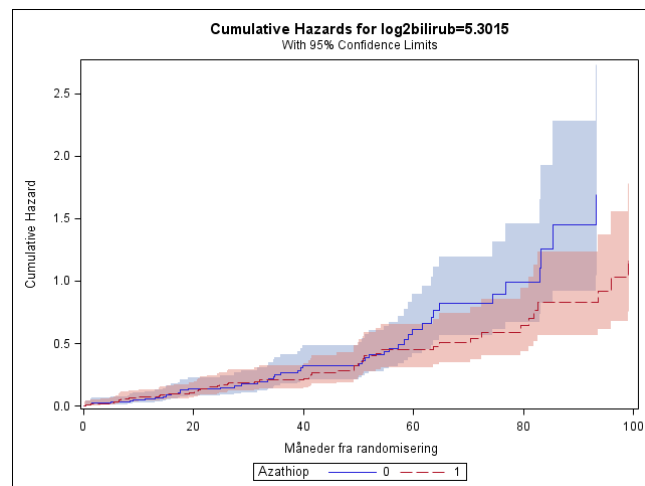
```



(b) *suppleret med et numerisk test*

Hvis man ønsker et numerisk test, må man vælge nogle passende tidspunkter, hvor proportionalitetsfaktoren kan få lov til at ændre sig. En god måde at vælge disse opdelingstidspunkter er at undersøge, om de kumulerede rater er nogenlunde lineære i nogle intervaller. Hvis de kumulerede rater er nogenlunde lineære, så er raterne nogenlunde konstante i de samme intervaller. Hvis raterne er konstante i begge grupper i et interval, er de selvfølgelig også proportionale i hele det interval.

Vi tjekker derfor den kumulerede rate i de to grupper (plot fra SAS koden vist under spørgsmål 6a:



Plottet indikerer, at de kumulerede rater begge kan tilnærmes med rette linjer med et enkelt knæk ved ca. 50 måneder. Vi tester derfor, om den samme proportionalitetsfaktor kan bruges i de to intervaller med opdeling ved 50 måneder:

```

PROC PHREG DATA=PBCdata;
  MODEL Months*Dead(0) = log2Bilirub  Az_0_50  Az_Fr50
    / ENTRY=Indgangstid RL;
  IF Months<50 THEN Az_0_50=Azathiop; ELSE Az_0_50=0;
  IF 50<=Months THEN Az_Fr50=Azathiop; ELSE Az_Fr50=0;
  Testprop: TEST  Az_0_50  =  Az_Fr50;
RUN;

```

The PHREG Procedure

Model Information	
Data Set	WORK.PBCDATA
Entry Time Variable	Indgangstid
Dependent Variable	Months
Censoring Variable	Dead
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	216
Number of Observations Used	216

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
216	103	113	52.31

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	965.850	866.655
AIC	965.850	872.655
SBC	965.850	880.559

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	99.1948	4	<.0001
Score	106.3621	4	<.0001
Wald	94.7777	4	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio	95% Confidence Limits
log2bilirub	1	0.73903	0.07728	91.4434	<.0001	2.094	1.800 2.436
Az_0_50	1	-0.29173	0.25291	1.3306	0.2487	0.747	0.455 1.226
Az_Fr50	1	-0.95151	0.34423	7.6404	0.0057	0.386	0.197 0.758

Linear Hypotheses Testing Results

Wald			
Label	Chi-Square	DF	Pr > ChiSq
Testprop	2.4604	1	0.1168

I stedet for at basere opdelingen på stykkevis linearitet for den kumulerede rate, kan man eksempelvis vælge at opdele tidsintervallerne, så der bliver ca. lige mange døde i hvert tidsinterval, fordi så er der ca.

lige stor statistisk styrke i hvert interval (jf. styrkeberegningen). For de aktuelle data kan bruges en opdeling ved 19 og 50 måneder, som brugt i nedenstående test for proportionalitetsantagelsen for bilirubin:

```
PROC PHREG DATA=PBCdata;
  MODEL Months*Dead(0) = Azathiop
                        log2Bili0_19 log2Bili19_50 log2BiliFr50
  / ENTRY=Indgangstid RL;
  IF Months<19 THEN log2Bili0_19=log2Bilirub; ELSE log2Bili0_19=0;
  IF 19<=Months<50 THEN log2Bili19_50=log2Bilirub; ELSE log2Bili19_50=0;
  IF 50<=Months THEN log2BiliFr50=log2Bilirub; ELSE log2BiliFr50=0;
  Testprop: TEST log2Bili0_19=log2Bili19_50=log2BiliFr50;
RUN;
```

med outputtet:

The PHREG Procedure

Model Information	
Data Set	WORK.PBCDATA
Entry Time Variable	Indgangstid
Dependent Variable	Months
Censoring Variable	Dead
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	216
Number of Observations Used	216

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
216	103	113	52.31

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	965.850	866.967
AIC	965.850	874.967
SBC	965.850	885.506

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	98.8828	4	<.0001
Score	105.7865	4	<.0001
Wald	92.9478	4	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio	95% Confidence Limits
Azathiop	1	-0.50679	0.20618	6.0418	0.0140	0.602	0.402 0.902
log2Bili0_19	1	0.82555	0.11904	48.0981	<.0001	2.283	1.808 2.883
log2Bili19_50	1	0.58916	0.12939	20.7340	<.0001	1.802	1.399 2.323
log2BiliFr50	1	0.81331	0.15332	28.1375	<.0001	2.255	1.670 3.046

Linear Hypotheses Testing Results			
	Wald		
Label	Chi-Square	DF	Pr > ChiSq
Testprop	2.1604	2	0.3395

Det ser således ud til, at proportionalitetsantagelsen (=konstant effekt) for serum bilirubin holder rimeligt godt henover hele perioden.

Man kunne alternativt have gjort det hele automatiseret ved at lægge variablene med tertilerne for dødstidspunkterne ind på vores PBC datasæt:

```
PROC UNIVARIATE DATA=PBCdata PCTLDEF=3 NOPRINT;
  WHERE dead=1;
  VAR months;
  OUTPUT OUT=TidsTertil PCTLPRE=Tid PCTLPTS=33.3 66.7;
RUN;
*- Læg tidstertilerne (variablene Tid33_3 og Tid66_7) ind på -*;
*- PBCdata, og lad SAS tage sig af resten "helt automatisk". -*;
DATA PBC_m_TidsTertil;
  IF _N_=1 THEN SET TidsTertil;
  SET PBCdata;
RUN;
PROC PHREG DATA=PBC_m_TidsTertil;
  MODEL Months*dead(0) = Azathiop BiliTidligt BiliMellem BiliSent
    / RL ENTRY=t_entry; /* ENTRY= hvis der er forsinket indgang */
  IF Months<=Tid33_3 THEN BiliTidligt=log2bilirub; ELSE BiliTidligt=0;
  IF Tid33_3<Months<=Tid66_7 THEN BiliMellem=log2bilirub; ELSE BiliMellem=0;
  IF Tid66_7<Months THEN BiliSent=log2bilirub; ELSE BiliSent=0;
  TestProp: TEST BiliTidligt = BiliMellem = BiliSent;
RUN;
```

Reference:

Christensen E, Neuberger J, Crowe J, Altman DG, Popper H, Portmann B, et al. Beneficial effect of azathioprine and prediction of prognosis in primary biliary cirrhosis. Final results of an international trial. *Gastroenterology* 1985;89(5):1084-91.

Samlet program

```
*-----*;
*- AXIS-statements til senere brug      -*;
*- Køres først! Så er de klar til senere. -*;
*-----*;
*- Almindelig X-akse -*;
AXIS1 LABEL=(H=2)      VALUE=(H=1.75);
*- Almindelig Y-akse -*;
AXIS2 LABEL=(A=90 H=2) VALUE=(H=1.75);
*- Log-skaleret Y-akse -*;
AXIS22 LABEL=(H=2 A=90) VALUE=(H=1.75)
      LOGBASE=10 INTERVAL=PARTIAL MAJOR=(H=1.5) MINOR=(H=1.5);

*-----*;
*- Indlæsning af data -*;
*-----*;
DATA PBCdata;
  INFILE "http://staff.pubhealth.ku.dk/~lts/basal/data/pbc.txt"
        URL /* fordi filen læses direkte fra Lenes hjemmeside */
        FIRSTOBS=2; /* fordi data skal indlæses fra og med linje 2 */
  INPUT Dead Months Azathiop Bilirubin;

*- Dummy-variabel Indgangstid tilføjes, så programmets opsætning -*;
*- også kan bruges, hvis der faktisk er forsinket indgang i studiet -*;
Indgangstid = -1;

LABEL
  Dead = "Død (0=censur, 1=død)"
  Months = "Måneder fra randomisering"
  Azathiop = "Azathioprine (0~Placebo, 1~Azathioprine)"
  Bilirubin = "Baseline serum bilirubin"
;
RUN;

*-----*;
*- log rank test -*;
*-----*;
PROC PHREG DATA=PBCdata;
  MODEL months*dead(0) = azathiop / ENTRY=Indgangstid TIES=DISCRETE RL;
RUN;

*-----*;
*- Kaplan-Meier med konfidensgrænser -*;
*-----*;
ODS GRAPHICS ON;
PROC PHREG DATA=PBCdata PLOTS(OVERLAY=ROW CL)=S;
  MODEL months*Dead(0) = / ENTRY=Indgangstid;
  STRATA Azathiop;
  BASELINE OUT=km
    SURVIVAL=KMcurves LOWER=LowerBound UPPER=UpperBound
    / METHOD=PL CLTYPE=LOGLOG;
RUN;
```

```

*- Til plot af Kaplan-Meier med konfidensgrænser, -*;
*- hvor man selv styrer akserne.                                -*;
DATA KMmCL; SET km;
  LABEL Curve = "Kaplan-Meier";
  IF azathiop=0 THEN DO;
    type=1; curve=kmcurves; OUTPUT;
    type=2; curve=lowerbound; OUTPUT;
    type=3; curve=upperbound; OUTPUT;
  END;
  IF azathiop=1 THEN DO;
    type=4; curve=kmcurves; OUTPUT;
    type=5; curve=lowerbound; OUTPUT;
    type=6; curve=upperbound; OUTPUT;
  END;
RUN;
SYMBOL1 R=1 C=BLACK V=NONE I=STEPLJ L=2 W=5 MODE=INCLUDE;
SYMBOL2 R=2 C=BLACK V=NONE I=STEPLJ L=4 MODE=INCLUDE;
SYMBOL3 R=1 C=RED V=NONE I=STEPLJ L=1 W=5 MODE=INCLUDE;
SYMBOL4 R=2 C=RED V=NONE I=STEPLJ L=41 MODE=INCLUDE;
PROC GPLOT DATA=KMmCL;
  PLOT curve*months=type / HAXIS=AXIS1 VAXIS=AXIS2 NOLEGEND;
RUN; QUIT;

*-----*;
*- Histogrammer over bilirubin -*;
*-----*;
PROC UNIVARIATE DATA=PBCdata;
  VAR bilirubin;
  HISTOGRAM / HEIGHT=4;
RUN;
DATA PBCdata;
  SET PBCdata;
  log2bilirub=LOG2(bilirubin);
RUN;
PROC UNIVARIATE DATA=PBCdata;
  VAR log2bilirub;
  HISTOGRAM / HEIGHT=4;
RUN;

*-----*;
*- Cox model med både bilirubin og log2(bilirubin) -*;
*-----*;
PROC PHREG DATA=PBCdata;
  MODEL months*dead(0) = bilirubin log2bilirub
    / ENTRY=Indgangstid RL;
  ESTIMATE "Mindst dbl" Bilirubin 5 Log2Bilirub 1 / EXP ;
  ESTIMATE "Størst dbl" Bilirubin 264 Log2Bilirub 1 / EXP ;
  Begge: TEST Bilirubin=log2Bilirub=0;
RUN;

```

```

*-----*;
*- Lineær spline -*;
*-----*;
*- Find relevante percentiler blandt de døde.          -*;
*- Percentilerne gemmes direkte på datasættet PCTLDATA -*;
*- som xpctl5, xpctl25, xpctl50, xpctl75 og xpctl95    -*;
*- 33.3 og 66.7 percentiler tages med til senere brug -*;
PROC UNIVARIATE DATA=PBCdata PCTLDEF=3 /* PCTLDEF=3 giver observerede percentiler */
                                NOPRINT; *- NOPRINT ~ Intet output -*;

    WHERE dead=1;
    VAR bilirubin;
    OUTPUT OUT=pctldata PCTLPRE=xpctl PCTLPTS=5 25 50 75 95 33.3 66.7;
RUN;
*- Beregne hjælpevariable, der skal bruges i PHREG -*;
DATA PBCdata;
    SET PBCdata;
    IF _N_=1 THEN SET pctldata; *- lægger percentilerne ind i PBCdata datasættet -*;
*- Beregning af hjælpevariablene med den "smarte" kode -*;
    IF NOT MISSING(bilirubin) THEN DO;
        b_u25pctl=MIN(bilirubin-Xpctl25,0);
        b_25_50pctl=MIN(MAX(Xpctl25-Xpctl50,bilirubin-Xpctl50),0);
        b_50_75pctl=MAX(MIN(Xpctl75-Xpctl50,bilirubin-Xpctl50),0);
        b_o75pctl=MAX(bilirubin-Xpctl75,0);
    END;
RUN;
*- Estimation og test -*;
PROC PHREG DATA=PBCdata
    OUTEST=parms; /* gemmer parameterestimerne på datasættet PARMS i variable med
                   de samme navne som de tilsvarende variable i analysen (f.eks. b_u25pctl) */
    MODEL months*dead(0) = b_u25pctl b_25_50pctl b_50_75pctl b_o75pctl
        / ENTRY=Indgangstid RL;
    TestLine: TEST b_u25pctl = b_25_50pctl = b_50_75pctl = b_o75pctl;
RUN;
*- Beregning af punkter til plot af lineær spline -*;
DATA plot;
    MERGE pctldata parms;
    LABEL
        bili = "Serum bilirubin ved baseline"
        rr = "Rate ratio"
    ;
*- Man behøver kun beregne støttepunkter i enderne og ved knækkene -*;
DO bili=xpctl5, xpctl25, xpctl50, xpctl75, xpctl95;
    *- De nye dummy-variable skal hedde noget lidt anderledes, fordi -*;
    *- de "gamle" navne jo "er brugt" til PHREG parameterestimerne -*;
    b_u25=MIN(bili-Xpctl25,0);
    b_25_50=MIN(MAX(Xpctl25-Xpctl50,bili-Xpctl50),0);
    b_50_75=MAX(MIN(Xpctl75-Xpctl50,bili-Xpctl50),0);
    b_o75=MAX(bili-Xpctl75,0);
    *- Beregning af det prognostiske index -*;
    pi = b_u25pctl*b_u25 + b_25_50pctl*b_25_50
        + b_50_75pctl*b_50_75 + b_o75pctl*b_o75;

```

```

    *- Beregning af rate ratio (hazard ratio) -*;      rr = EXP(pi);
    *- Læg observationen ud på PLOT datasættet -*;      OUTPUT;
END;
RUN;
*- Tegning af lineær spline -*;
SYMBOL1 C=BLACK V=CIRCLE I=JOIN L=1 MODE=INCLUDE;
PROC GPLOT DATA=plot;
    PLOT rr*bili / HAXIS=AXIS1 VAXIS=AXIS2;
    GOPTIONS DEVICE=; *- SKAL med for SAS Enterprise, ellers helt unødvendig -*;
RUN; QUIT;

*-----*;
*- Estimation af effekten af serum bilirubin alene -*;
*-----*;
PROC PHREG DATA=PBCdata;
    MODEL months*dead(0) = log2bilirub / RL ENTRY=Indgangstid ;
RUN;

*-----*;
*- Estimation af effekten af azathioprine justeret -*;
*- for effekten af baseline serum bilirubin          -*;
*-----*;
PROC PHREG DATA=PBCdata;
    MODEL months*dead(0) = log2bilirub Azathiop / RL ENTRY=Indgangstid ;
RUN;

*-----*;
*- Grafisk tjek af proportionalitetsantagelsen for azathioprine -*;
*-----*;
PROC PHREG DATA=PBCdata PLOTS(OVERLAY=ROW CL)=CUMHAZ;
    MODEL months*dead(0) = log2bilirub / RL ENTRY=Indgangstid ;
    STRATA azathiop;
    BASELINE OUT=TjekA LOGLOGS=LCumRate / METHOD=CH;
RUN;
SYMBOL1 C=BLACK V=NONE I=STEPLJ L=2 MODE=INCLUDE;
SYMBOL2 C=RED    V=NONE I=STEPLJ L=1 MODE=INCLUDE;
    *- Almindelig lineær tidsakse -*;
PROC GPLOT DATA=TjekA;
    PLOT LCumRate*months=azathiop / HAXIS=AXIS1 VAXIS=AXIS2 NOLEGEND;
RUN; QUIT;

*-----*;
*- Grafisk tjek af proportionalitetsantagelsen for serum bilirubin -*;
*-----*;
PROC PHREG DATA=PBCdata PLOTS(OVERLAY=ROW CL)=CUMHAZ;
    MODEL months*dead(0) = log2bilirub azathiop / RL ENTRY=Indgangstid ;
    STRATA bilirubin (37, 92); *- opdel v 33.3 og 66.7 percentiler blandt døde -*;
    BASELINE OUT=TjekB LOGLOGS=LCumRate / METHOD=CH;
RUN;

```

```

SYMBOL1 C=BLACK V=NONE I=STEPLJ L=2 MODE=INCLUDE;
SYMBOL2 C=RED V=NONE I=STEPLJ L=1 MODE=INCLUDE R=1; *- SAS husker R=2 fra KM plottet -*;
SYMBOL3 C=BLUE V=NONE I=STEPLJ L=41 MODE=INCLUDE;
  *- Almindelig lineær tidsakse -*;
PROC GPLOT DATA=TjekB;
  PLOT LCumRate*months=bilirubin / HAXIS=AXIS1 VAXIS=AXIS2;
RUN; QUIT;

*-----*;
*- Numerisk tjek af proportionalitetsantagelsen for azathioprine -*;
*-----*;
*- Numerisk tjek af proportionalitetsantagelsen for azathioprine -*;
*- med opdeling ved 50 måneder. -*;
PROC PHREG DATA=PBCdata;
  MODEL Months*Dead(0) = log2Bilirub Az_0_50 Az_Fr50
    / ENTRY=Indgangstid RL;
  IF Months<50 THEN Az_0_50=Azathiop; ELSE Az_0_50=0;
  IF 50<=Months THEN Az_Fr50=Azathiop; ELSE Az_Fr50=0;
  Testprop: TEST Az_0_50 = Az_Fr50;
RUN;

*-----*;
*- Numerisk tjek af proportionalitetsantagelsen for log2(bilirubin) -*;
*-----*;
*- I stedet for at kigge efter stykkevis linearitet kan man vælge -*;
*- at fordele antallet af events ca. ligeligt i hvert tidsinterval, -*;
*- fordi antallet af events bestemmer den statistiske styrke. -*;
PROC UNIVARIATE DATA=PBCdata PCTLDEF=3 NOPRINT;
  WHERE dead=1;
  VAR months;
  OUTPUT OUT=TidsTertil PCTLPRE=Tid PCTLPTS=33.3 66.7;
RUN;
PROC PRINT DATA=Tidstertil; RUN;

*- Numerisk tjek af proportionalitetsantagelsen (= samme effekt af -*;
*- log2(bilirubin) i hele perioden) med opdeling ved 19 og 50 måneder -*;
*- (tæt på tertilerne blandt dødstidspunkterne). -*;
PROC PHREG DATA=PBCdata;
  MODEL Months*Dead(0) = Azathiop
    log2Bili0_19 log2Bili19_50 log2BiliFr50
    / ENTRY=Indgangstid RL;
  IF Months<19 THEN log2Bili0_19=log2Bilirub; ELSE log2Bili0_19=0;
  IF 19<=Months<50 THEN log2Bili19_50=log2Bilirub; ELSE log2Bili19_50=0;
  IF 50<=Months THEN log2BiliFr50=log2Bilirub; ELSE log2BiliFr50=0;
  Testprop: TEST log2Bili0_19 = log2Bili19_50 = log2BiliFr50;
RUN;

```

```

*- Automatiseret numerisk test for proportionalitet -*;
PROC UNIVARIATE DATA=PBCdata PCTLDEF=3 NOPRINT;
  WHERE dead=1;
  VAR months;
  OUTPUT OUT=TidsTertil PCTLPRE=Tid PCTLPTS=33.3 66.7;
RUN;
*- Læg tidstertilene (variablene Tid33_3 og Tid66_7) ind på -*;
*- PBCdata, og lad SAS tage sig af resten "helt automatisk". -*;
DATA PBC_m_TidsTertil;
  IF _N_=1 THEN SET TidsTertil;
  SET PBCdata;
RUN;
PROC PHREG DATA=PBC_m_TidsTertil;
  MODEL Months*dead(0) = Azathiop BiliTidligt BiliMellem BiliSent
    / RL ENTRY=Indgangstid; /* ENTRY= hvis der er forsinket indgang */
  IF Months<=Tid33_3 THEN BiliTidligt=log2bilirub; ELSE BiliTidligt=0;
  IF Tid33_3<Months<=Tid66_7 THEN BiliMellem=log2bilirub; ELSE BiliMellem=0;
  IF Tid66_7<Months THEN BiliSent=log2bilirub; ELSE BiliSent=0;
  TestProp: TEST BiliTidligt = BiliMellem = BiliSent;
RUN;

```