Lecture 25

1. Another graphical check of proportional hazards assumption
2. Modeling non-proportional hazards
3. Criticism of proportional hazards regression
4. Competing risks: cumulative incidence

Another graphical check of proportional hazards assumption

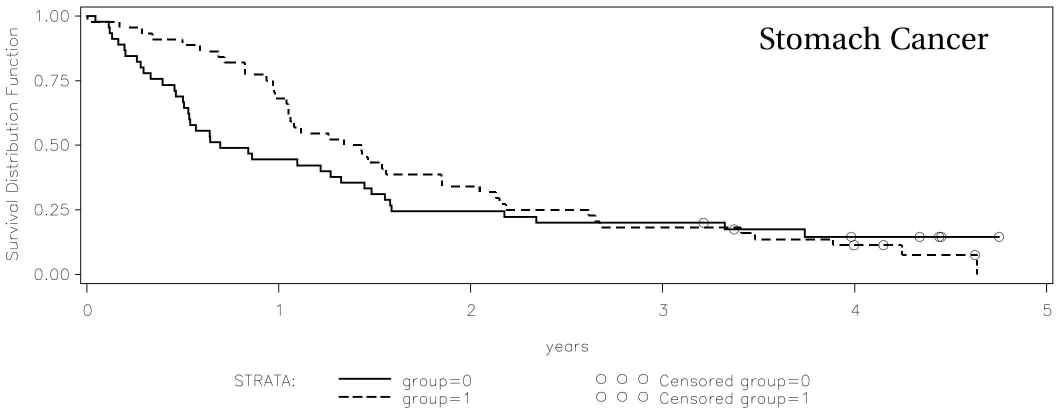
PHreg (but not TPHreg) will create graphics to check the proportional hazards assumption for individual predictors.

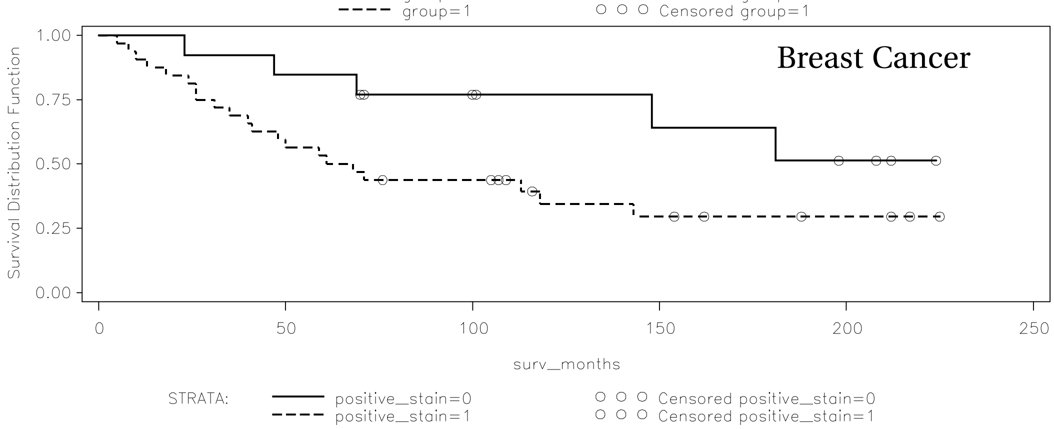
PHreg generates simulated processes under the null hypothesis that the predictor X has multiplicative effect on the baseline hazard: $\exp(\beta x) \times h_0(t)$

Actual process is plotted on top of these simulated processes:

If it lies within the range of variability, predictor may satisfy proportional hazards assumption

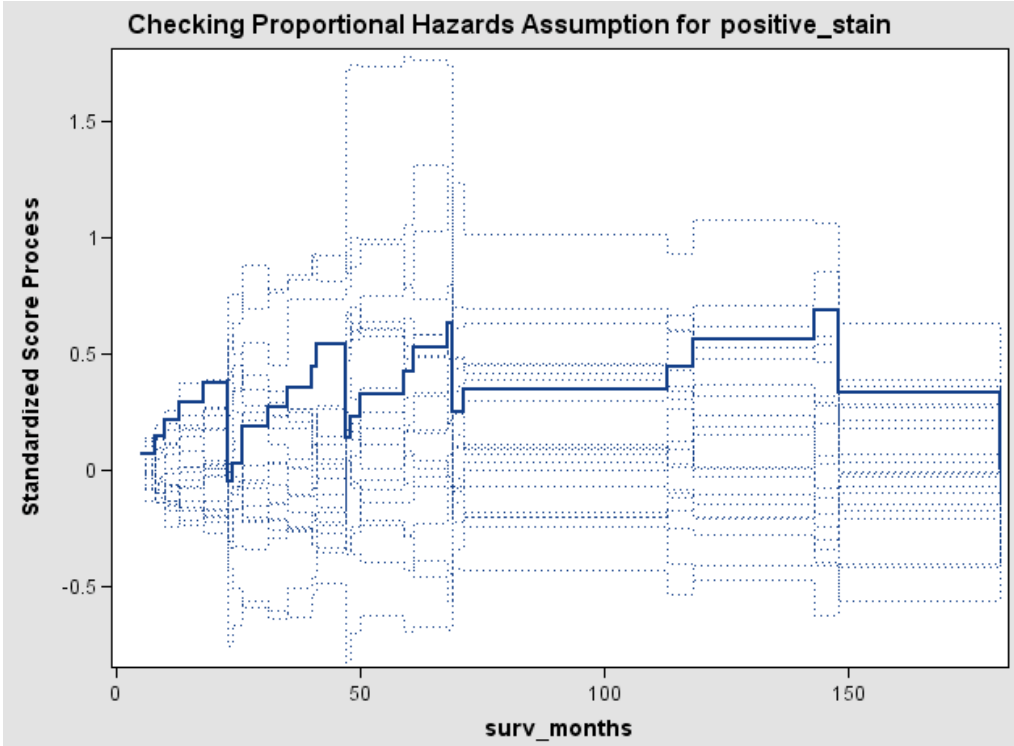
If not, evidence against proportional hazards for that predictor.



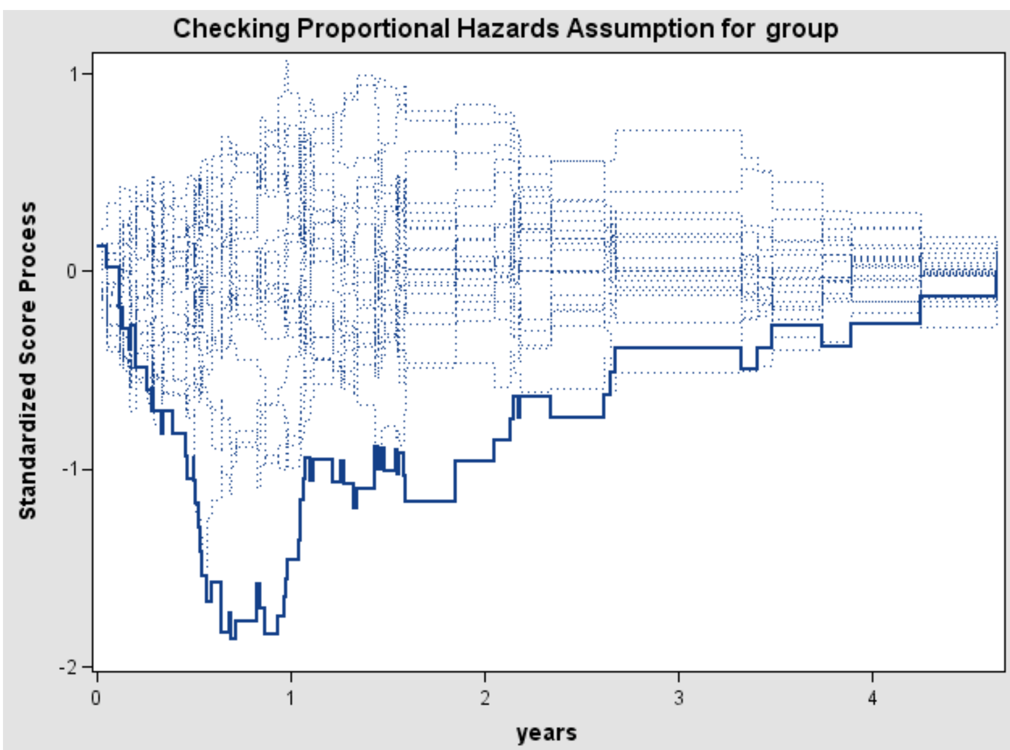


Use the same ODS (Output Delivery System) set up as for plots in Lifetest:

```
ods html;     or use PDF instead of HTML
ods graphics on;

Proc PHreg data=pubh.breast_cancer;
  model surv_months* died(0) = positive_stain  /
    risklimits ties=efron ;
  assess  ph   ;  produce plot for each predictor in model
run;
 ods graphics off;
 ods html close;
```



Actual path within envelope of simulated paths: proportional hazards seems OK.



As expected, proportional hazards does not hold: actual path outside envelope.

Modeling non-proportional hazards: strata

Clinical trial to compare treatments A and B enrolls both children and adults.

Predictors include baseline disease stage, adult status (=1 for adults, =0 children).

Try a proportional hazards model:

```
Proc TPHreg ;
   class stage;
   model survival_months* died(0) = treatment  stage adult;
   assess PH;
```

Bad news: the predictor adult violates proportional hazards assumption.

Other predictors are OK.

Proportional hazards regression can accommodate groups with different baseline hazards.

Stratify on adult status: fit two baseline hazards, $h_C(t)$ for children, $h_A(t)$ for adults

Remove adult from model.

```
Proc TPHreg ;
   class stage;
   model survival_months* died(0) = treatment  stage;
   strata adult;
```

- No test for difference between children and adults (age effect)
- Predictors in model assumed to have *identical* effects in all strata

Criticism of proportional hazards regression

- Widely used to analyze both randomized trials and observational studies
- Assume: censoring is independent of events (chance that someone is censored not affected by chance of event)
- Assume: every predictor has additive effect on log hazard ratio
- Model essentially uses the rank order of failure times, not actual times
- Observational studies: model assumptions very difficult to verify
- Randomized trials may not need the adjustment

Freedman (2008) Survival analysis: A primer. *The American Statistician,* 62: 110–119

Competing risks

So far, we have studied distribution of times to a single event. What about more than one event?

Example: patients are treated for cancer, then either the cancer returns (relapse) or the patient dies without cancer (in remission).

Only one of these events will happen. If a patient dies in remission they do not relapse. If a patient relapses, they have not already died.

Patients are at risk for both events, which will happen first?

Model for competing risks

Each patient has two (potential) times to event:

- T_R = time to relapse
- T_D = time to death in remission

We get to observe $T = \min\{T_R, T_D\}$ and the type of event, and the indicator C for censoring if neither event has occurred.

Overall hazard is the sum of the hazard for each event: $h(t) = h_R(t) + h_D(t)$

Example: survival (days) after bone marrow transplant.

Outcomes: 0=alive, 1=dead before relapse, 2=relapse

ID	survival_ days	outcome
35	1	1
34	55	2
28	74	2
30	86	1
18	104	2
36	107	1
33	109	2
37	110	2
26	122	1
29	122	2

What is the combined chance of either event within 1 year after treatment?

Answer:

Cumulative event rate through 1 year, treating both R and D as a single event:

$$1 - \hat{S}(1 \text{ year})$$

```
Proc Lifetest data=bmt ;
   time survival_days * outcome(0);
```

Note that both outcome=1 and outcome=2 are treated as *the* event.

What is the combined chance of either event within 1 year after treatment?

survival_ days	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.00	1.0000	0	0	0	38
1.00	0.9737	0.0263	0.0260	1	37
55.00	0.9474	0.0526	0.0362	2	36
74.00	0.9211	0.0789	0.0437	3	35
86.00	0.8947	0.1053	0.0498	4	34
104.00	0.8684	0.1316	0.0548	5	33
107.00	0.8421	0.1579	0.0592	6	32
109.00	0.8158	0.1842	0.0629	7	31
110.00	0.7895	0.2105	0.0661	8	30
122.00	.	.	.	9	29
122.00	0.7368	0.2632	0.0714	10	28
129.00	0.7105	0.2895	0.0736	11	27
172.00	0.6842	0.3158	0.0754	12	26
192.00	0.6579	0.3421	0.0770	13	25
194.00	0.6316	0.3684	0.0783	14	24
226.00*	.	.	.	14	23
230.00	0.6041	0.3959	0.0795	15	22
276.00	0.5767	0.4233	0.0805	16	21
332.00	0.5492	0.4508	0.0812	17	20
383.00	0.5217	0.4783	0.0817	18	19
418.00	0.4943	0.5057	0.0819	19	18

466.00	0.4668	0.5332	0.0818	20	17
487.00	0.4394	0.5606	0.0815	21	16
526.00	0.4119	0.5881	0.0809	22	15
530.00*	.	.	.	22	14
609.00	0.3825	0.6175	0.0803	23	13
662.00	0.3531	0.6469	0.0793	24	12
996.00*	.	.	.	24	11
1111.00*	.	.	.	24	10
1111.00*	.	.	.	24	10
1167.00*	.	.	.	24	9
1182.00*	.	.	.	24	8
1199.00*	.	.	.	24	7
1330.00*	.	.	.	24	6
1377.00*	.	.	.	24	5
1433.00*	.	.	.	24	4
1462.00*	.	.	.	24	3
1496.00*	.	.	.	24	2
1602.00*	.	.	.	24	1
2081.00*	.	.	.	24	0

NOTE: The marked survival times are censored observations.

Where should we truncate the graph?

There is no option to plot $1 - \hat{S}(t)$.

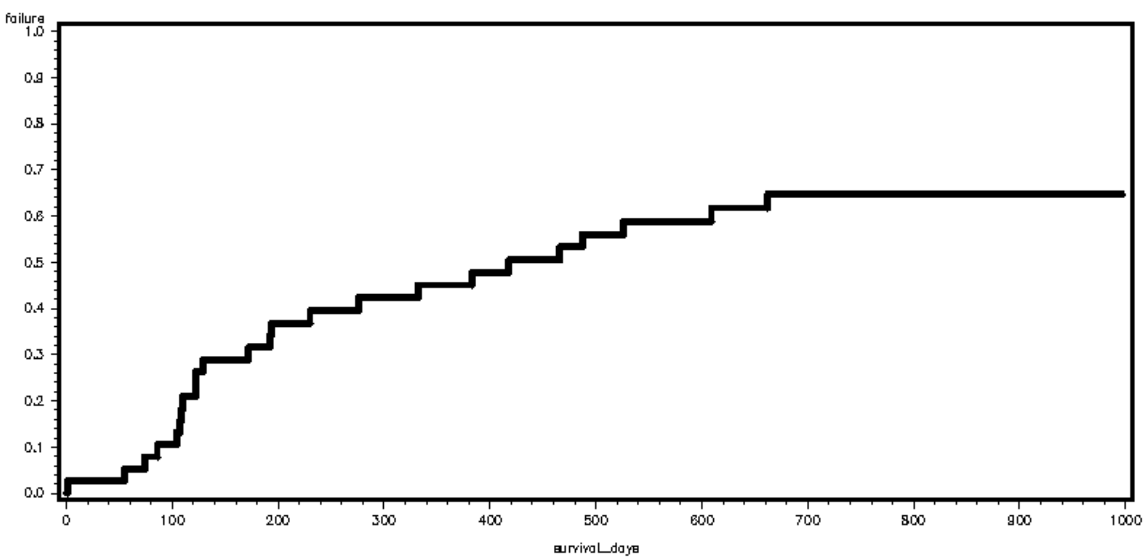
Output $\hat{S}(t)$ to data set from Proc Lifetest, fill in the missing values for censored observations and plot.

```
 proc lifetest data=b  outsurv=c;
   time survival_days * outcome(0);

data failure_plot;
   set c;
   if survival_days LE 1000;
   failure = 1 - survival;
   retain last_failure;
   if survival_days=0 then last_failure=0;
   else  if failure=. then failure=last_failure;
     else last_failure=failure;
```

```
Goptions reset=all  ...
    symbol1  interpol= stepLJ    line=1 width=2;
     step function, values at left end of step, join vertically

  Proc Gplot data=failure_plot;
    plot failure*survival_days / vaxis=0 to 1 by .1;
```

Need to set vertical axis.

Kaplan-Meier cumulative incidence for both relapse and death



What is the chance of relapse within 1 year after treatment?

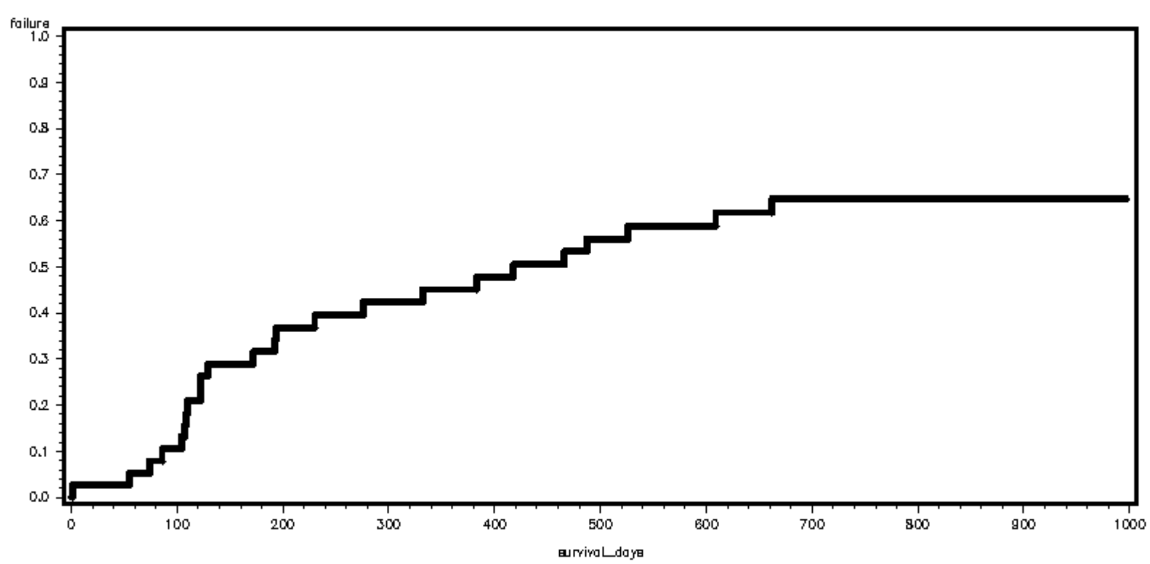
Common approach: treat both death and censored as censored.

Outcome: 0=alive, 1=dead before relapse, 2=relapse

```
proc lifetest data=b  outsurv=d;
  time survival_days * outcome(0,1) ;
```

Both 0 and 1 values of outcome are treated as censored.

Complement of Kaplan-Meier survival function for relapse, death-censored



What is the chance of relapse within 1 year after treatment?

survival_ days	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.00	1.0000	0	0	0	38
1.00*	.	.	.	0	37
55.00	0.9730	0.0270	0.0267	1	36
74.00	0.9459	0.0541	0.0372	2	35
86.00*	.	.	.	2	34
104.00	0.9181	0.0819	0.0453	3	33
107.00*	.	.	.	3	32
109.00	0.8894	0.1106	0.0522	4	31
110.00	0.8607	0.1393	0.0579	5	30
122.00	0.8320	0.1680	0.0626	6	29
122.00*	.	.	.	6	28
129.00	0.8023	0.1977	0.0671	7	27
172.00*	.	.	.	7	26
192.00	0.7715	0.2285	0.0713	8	25
194.00*	.	.	.	8	24
226.00*	.	.	.	8	23
230.00	0.7379	0.2621	0.0756	9	22
276.00*	.	.	.	9	21
332.00*	.	.	.	9	20
383.00	0.7010	0.2990	0.0804	10	19
418.00*	.	.	.	10	18

Problem with treating death (the other risk) as censored is that *it's no longer a risk.*

Estimated chance of relapse before 1 year = 26% from this Kaplan-Meier estimate applies to a hypothetical world in which death does not happen.

Although this approach appears to answer the question, the answer is not helpful.

We want to know the risk in the real world where death is a competing risk.

Cumulative Incidence Estimate

This estimates the chance of an relapse before 1 year *and before death.*

Essentially it is the proportion relapsing among those at risk of relapse or death.

If r_i is the number of relapses at time t_i among N_i subjects at risk,
then the cumulative incidence estimate is

$$C(t) = \sum_{t_i \le t} \hat{S}(t_i^-) \frac{r_i}{N_i}$$

where $\hat{S}(t_i^-)$ is the Kaplan-Meier estimate treating both relapse and death as the event, evaluated just before t_i

Suppose 100 patients are treated.

At year 1, 30 have died and 10 have relapsed but none are censored during the year. With no censoring, K-M reduces to sample proportion.

Cumulative incidence of relapse is $10/100$.

Cumulative incidence of death is $30/100$.

Using the Kaplan-Meier estimate, the answer depends on the ordering of the deaths and relapses.

If all deaths occur before the first relapse, then the chance is $10/70$.

If all relapses occur first, then chance is $10/100$.

Different event patterns give values in between.

Fitting cumulative incidence

Macro available from SAS website:

```
%macro CumIncid(
  data=,      /* specifies the name of the input data set           */
  out=,       /* specifies the name of the output data set that      */
              /* contains the cumulative incidence estimates         */
  time=,      /* specifies the time variable                         */
  status=,    /* specifies the variable whose values indicate whether */
              /* an observation corresponds to the event of interest, */
              /* the competing risk event, or is censored            */
  event=,     /* specifies the values of status= that correspond to  */
              /* the event of interest                               */
  compete=,   /* specifies the values of status= that correspond to  */
              /* the competing risk events                           */
  censored=,  /* specifies the values of status= that correspond to  */
              /* censored observations                               */
  strata=,    /* specifies the list of strata variables              */
  alpha=,     /* specifies the complement of the confidence level for */
              /* pointwise confidence limits                         */
  odsgraphics=,
              /* on or off, on for turning on the ODS graphics on,   */
              /* and the cumulative incidence curve will be displayed. */
```

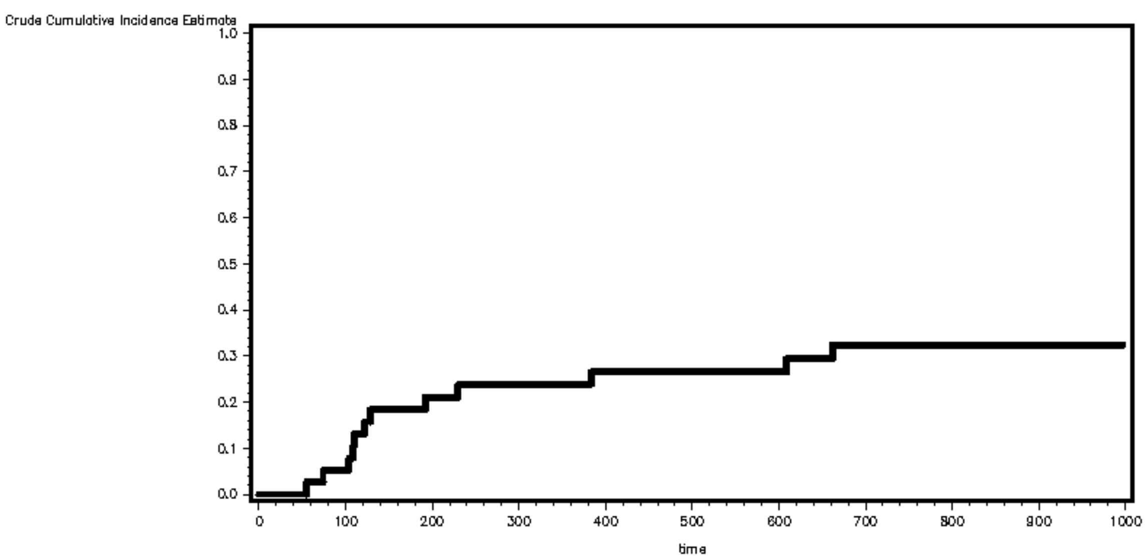


```
              /* odsgraphics=on is the default.                      */
  htmfile=,   /* output htm file without the .htm extension          */
  rtffile=,   /* output rtf file without the .rtf extension          */
  options=,   /* Specify one or more of following values after the   */
              /* OPTIONS=.                                           */
              /*                                                     */
              /* CSE    produces the cause-specific failure cumulative */
              /*        incidence estimates. This estimate is        */
              /*        precisely the complement of the Kaplan-Meier */
              /*        estimate of the survivor function with       */
              /*        competing risk events being treated as       */
              /*        censored values. When this option is         */
              /*        specified, standard errors and lower and     */
              /*        upper confidence limits are also produced.   */
              /*                                                     */
              /* PLOTCL                                              */
              /*        plots the pointwise confidence limits for    */
              /*        individual cumulative incidence curves.      */
              /* NOPRINT                                             */
              /*        suppress print of the cumulative incidence   */
              /*        estimates                                    */
       );
```

Data BMT outcomes: 0=alive, 1=dead before relapse, 2=relapse

ID	survival_ days	outcome
35	1	1
34	55	2
28	74	2
30	86	1

```
%include ‘‘.. path to macro..’’

%CumIncid(data=BMT, out=ci_relapse, time=survival_days,
    status=outcome, event=2, compete=1, censored=0,
    strata=group,alpha=0.05, odsgraphics=off,options=.);
```

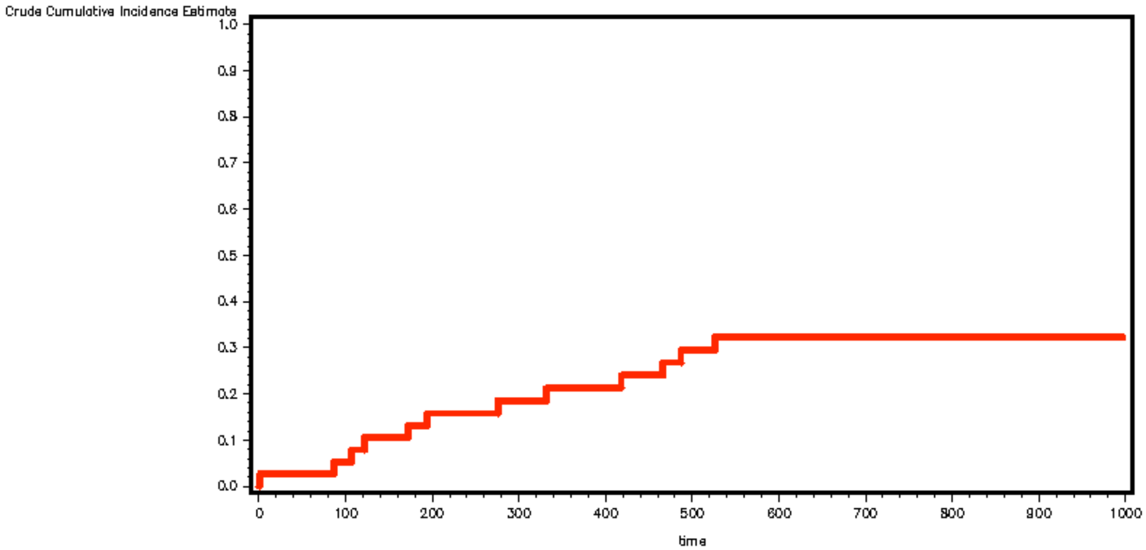
What is the cumulative incidence of relapse within 1 year after treatment?

t2	censor	CumInc	StdErr CumInc	Lower95Pct CumInc	Upper95Pct CumInc
0.00	.	0.00000	.	.	.
1.00	1	.	.	.	.
55.00	0	0.02632	0.025967	0.00380	0.18203
74.00	0	0.05263	0.036224	0.01366	0.20281
86.00	1	.	.	.	.
104.00	0	0.07895	0.043744	0.02665	0.23387
107.00	1	.	.	.	.
109.00	0	0.10526	0.049784	0.04166	0.26598
110.00	0	0.13158	0.054836	0.05814	0.29780
122.00	0	0.15789	0.059153	0.07577	0.32905
122.00	1	.	.	.	.
129.00	0	0.18421	0.062886	0.09435	0.35967
172.00	1	.	.	.	.
192.00	0	0.21053	0.066135	0.11374	0.38968
194.00	1	.	.	.	.
226.00	1	.	.	.	.
230.00	0	0.23799	0.069299	0.13449	0.42113
276.00	1	.	.	.	.
332.00	1	.	.	.	.
383.00	0	0.26545	0.072032	0.15595	0.45181
418.00	1	.	.	.	.

To get cumulative incidence of death:

Outcomes: 0=alive, 1=dead before relapse, 2=relapse

```
%include ‘‘.. path to macro..’’

%CumIncid(data=BMT, out=ci_death, time=survival_days,
    status=outcome, event=1, compete=2, censored=0,
    strata=group,alpha=0.05, odsgraphics=off,options=.);
```

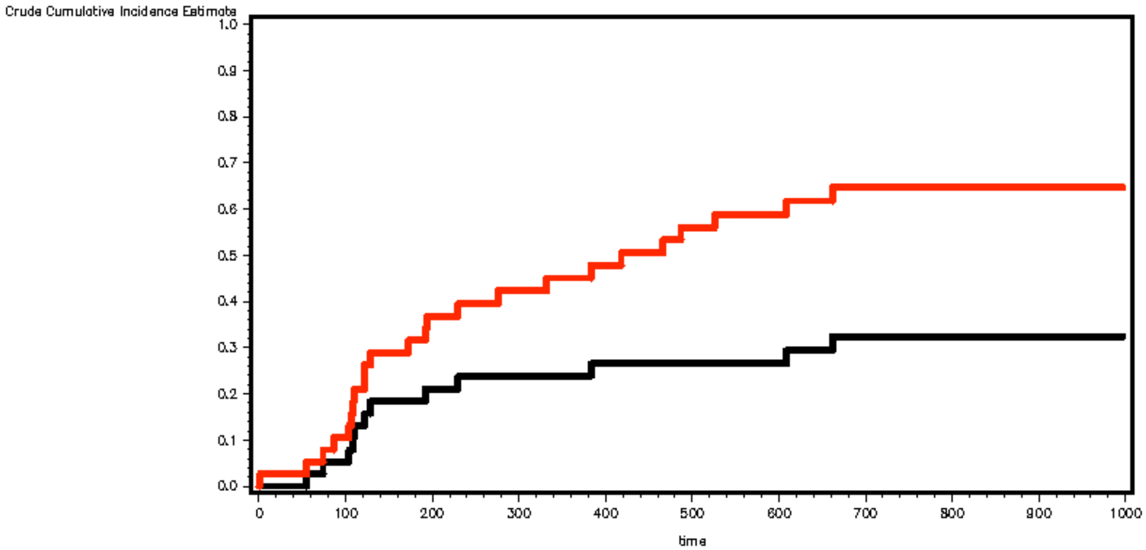


What is the cumulative incidence of death within 1 year after treatment?

t2	censor	CumInc	StdErr CumInc	Lower95Pct CumInc	Upper95Pct CumInc
0.00	.	0.00000	.	.	.
1.00	0	0.02632	0.025967	0.00380	0.18203
55.00	1	.	.	.	.
74.00	1	.	.	.	.
86.00	0	0.05263	0.036224	0.01366	0.20281
104.00	1	.	.	.	.
107.00	0	0.07895	0.043744	0.02665	0.23387
109.00	1	.	.	.	.
110.00	1	.	.	.	.
122.00	0	0.10526	0.049784	0.04166	0.26598
122.00	1	.	.	.	.
129.00	1	.	.	.	.
172.00	0	0.13158	0.054836	0.05814	0.29780
192.00	1	.	.	.	.
194.00	0	0.15789	0.059153	0.07577	0.32905
226.00	1	.	.	.	.
230.00	1	.	.	.	.
276.00	0	0.18535	0.063275	0.09494	0.36189
332.00	0	0.21281	0.066827	0.11500	0.39382
383.00	1	.	.	.	.
418.00	0	0.24027	0.069898	0.13586	0.42494

Each cumulative incidence is relative to the other competing risks, so it is important to report them together.

With two risks, a good plot shows CI for risk 1 and (CI for risk 1 + CI for risk 2). This gives each CI and the death-and-relapse-free probability at the top.



Merge the 2 cumulative incidence output datasets, fill in censored values, sum:

```
data ci_plot;
  merge ci_relapse(rename=(t2=time CumInc = relapse_CI StdErrCumInc=relapse_SE))
        ci_death  (rename=(t2=time CumInc = death_CI   StdErrCumInc=death_SE));
  by time;
  if time LE 1000;
  array ci[2] relapse_CI death_CI;
  array last_ci[2] last_relapse_CI last_death_CI;
  retain last_relapse_CI last_death_CI;
  if time=0 then do j = 1 to 2; last_ci[j]=0; end;
  else do j = 1 to 2;
    if ci[j]=. then ci[j]=last_ci[j];
       else last_ci[j]=ci[j];
  end;
  sum_CI = relapse_CI+ death_CI;
  keep time relapse_CI death_CI sum_CI;
```

```
Goptions reset=all  ...
    symbol1 interpol= stepLJ line=1 width=2;
    symbol2 interpol= stepLJ line=1 color=red width=2;

 Proc Gplot data=ci_plot;
   plot relapse_CI*time
        sum_CI*time
          / overlay  vaxis=0 to 1 by .1;
```

Data source and reference:

Klein and Moeschberger (2003) *Survival Analysis, second edition.* Springer.