

# Competing risks

Melania Pintilie

Ontario Cancer Institute/PMH/UHN

University of Toronto



Princess Margaret Hospital  
University Health Network



# Outline

- Background (SAS/R)
- Competing risks (CR) introduction
- Theoretical background
- Non-parametric estimation of the probability of event
- Kaplan-Meier estimate in the presence of competing risks

# Outline

- Testing a covariate
- Fine and Gray model
- Cox proportional hazards model vs. Fine and Gray model
- Prediction
- Non-proportionality of the hazard of subdistribution
- Sample size

# Input/Output in SAS

```
filename f 'c:/your_directory/follic.csv';  
filename outf 'c:/your_directory/follic_out.csv';  
  
data follic; infile f dsd truncover delimiter=','  
  firstobs=2 obs=542;  
input age sex $ mal $ path1 $ hgb ldh clinstg  
  blktxcat relsite $ ch $ rt $ survtime stat dftime  
  dfcens resp $ maltime stnum;  
*****;  
proc export data=follic outfile=outf dbms=csv  
  replace;
```

# SAS macros

[http://www.uhnres.utoronto.ca/labs/hill/People\\_Pintilie.htm](http://www.uhnres.utoronto.ca/labs/hill/People_Pintilie.htm)

- cuminc – estimates probability of the event of interest
- compcif
  - estimates for the probability of event of interest
  - estimates the probability for the competing risk event
  - compares the probabilities of the event of interest between 2 groups (Pepe and Mori, 1993)
- compcp - compares the conditional probabilities of the event of interest between 2 groups
  - estimates for the probability of event of interest and the competing risk event
  - estimates for the conditional probability of event of interest
  - compares the conditional probabilities of the event of interest between 2 groups (Pepe and Mori, 1993)
- Use *%include* to import the macro

# Software: *cmprsk* in R

<http://www.r-project.org> (CRAN)

- `cuminc` = estimates and compares CIFs
- `plot.cuminc`
- `timepoints` = gives estimates at certain points
- `crr` = modeling
- `predict.crr` = based on the hazard of subdistribution
- `plot.predict.crr`

# R functions

[http://www.uhnres.utoronto.ca/labs/hill/People\\_Pintilie.htm](http://www.uhnres.utoronto.ca/labs/hill/People_Pintilie.htm)

- compCIF
- cifDM
- compCP
- CPvar
- btvarCP2
- plot.cp
- kly
- power

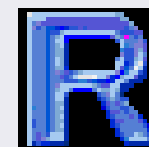
Use *source('name of file')* to read  
in these functions

# Details for R software

- basic: general functions like **mean**, **plot**, **sum**
- libraries: **survival** and **cmprsk**
- The libraries need to be installed once
- The libraries need to be loaded every time you start a new session
  - Menu: Packages-> Load
  - **library(survival)** or **library(cmprsk)**

# Details for R

- Change the directory to your working directory
  - Menu: File -> Change dir...
  - `getwd("C:/your_directory")`
- Save the workspace
  - Menu: File -> Save Workspace
  - `save.image("C:/your_directory/.RData")`
  - Will save all objects (datasets, variables) but need to load the libraries every time
- Useful commands: **help**, **help.search**, **class**



.RData

# Input/Output in R

```
follic=read.table('c:/your_directory/follic.csv',  
  sep=',',na.strings='.',header=T)
```

```
names(follic)=casefold(names(follic))
```

```
#####
```

```
write.table(follic,'follic_out.csv',sep=',')
```

```
## created in your home directory
```

# Outline

- Background (SAS/R)
- Competing risks (CR) introduction
- Theoretical background
- Non-parametric estimation of the probability of event
- Kaplan-Meier estimate in the presence of competing risks

# Definition

Competing risks type of event=

*the event whose occurrence either precludes the occurrence of another event under investigation or fundamentally alters the probability of occurrence of this other event.*

Gooley, TA; Leisenring, W; Crowley, J; Storer, BE,  
"Estimation of failure probabilities in the presence of competing risks: new representations of old estimators" *Statistics in Medicine* 1999 pp. 695-706

# Definition, Examples

*the event whose occurrence either precludes the occurrence of another event under investigation*

- **Death: due to disease (MI) and due to other causes**
- **First relapse: local relapse and distant relapse**

*fundamentally alters the probability of occurrence of this other event*

# Example

- Event of interest: Death to MI
- CR event: Death of other causes
- Alive

Ignore CR

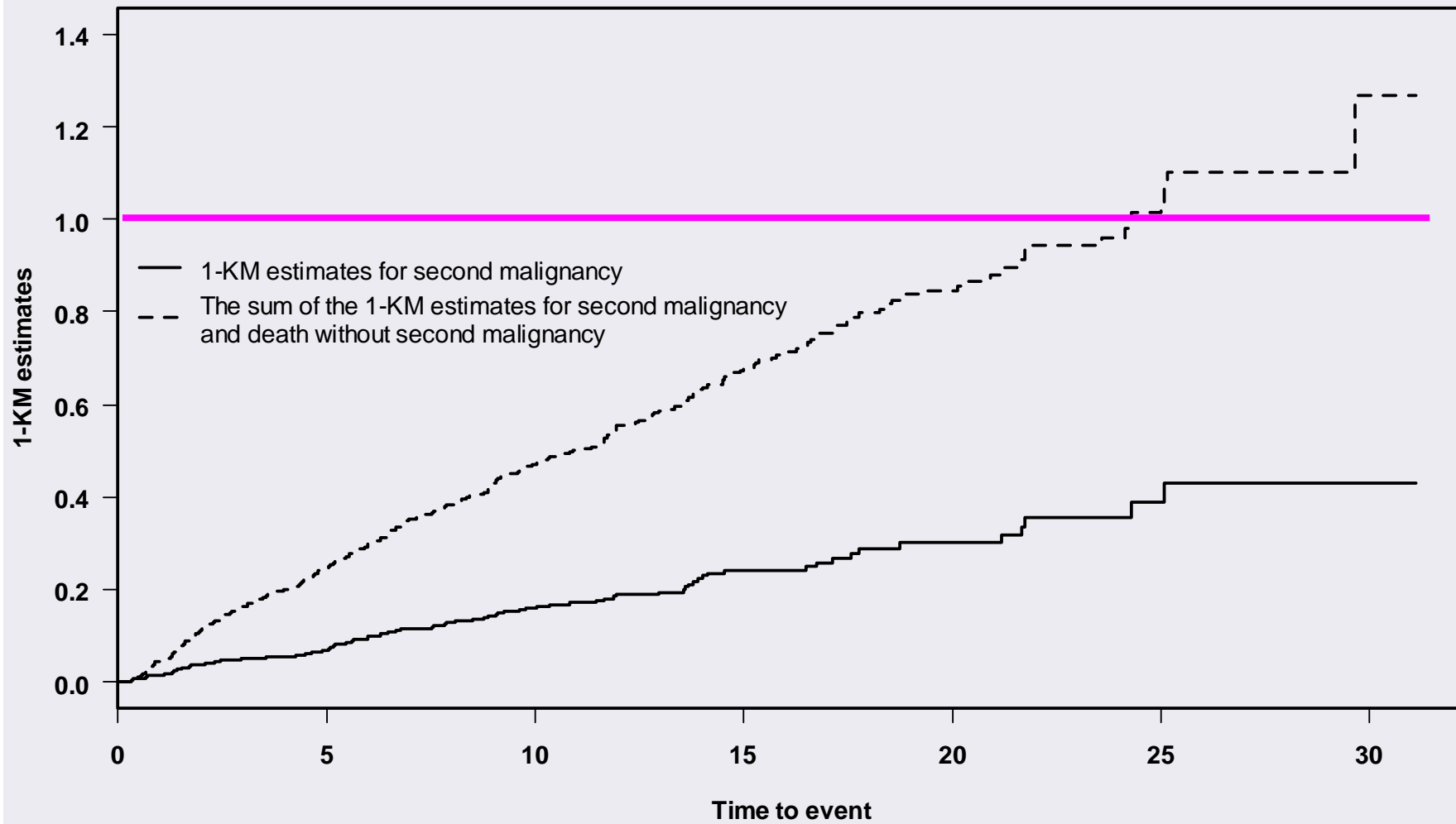
**Censor CR event**  
**Apply usual**  
**survival techniques**

Acknowledge CR

**Code CR event**  
**Apply specific**  
**techniques**

# Example follicular lymphoma

- Follicular lymphoma= type of cancer
- Early stage (disease is not spread), high cure rates, long term side effects of the treatment
- Event of interest is second malignancy
- Competing risk: death without second malignancy (most likely from lymphoma)



# Outline

- Background (SAS/R)
- Competing risks (CR) introduction
- Theoretical background
- Non-parametric estimation of the probability of event
- Kaplan-Meier estimate in the presence of competing risks

# Basic concepts (no CR)

- Survivor function (KM)

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

$$S(t) = e^{-\lambda t}$$

- **Distribution function (1-KM)**

$$F(t) = P(T \leq t) = 1 - S(t)$$

$$F(t) = 1 - e^{-\lambda t}$$

- Density function

$$f(t) = \frac{\partial F(t)}{\partial t}$$

$$f(t) = \lambda e^{-\lambda t}$$

- **Hazard function (modelled with Cox PH)**

$$h(t) = \lim_{\partial t \rightarrow 0} \left\{ \frac{P(t < T \leq t + \partial t \mid T > t)}{\partial t} \right\} = \frac{f(t)}{1 - F(t)}$$

$$h(t) = \lambda$$

# Basic concepts (CR)

- Subdistribution function

$$F_i(t) = P(T \leq t, C = i)$$

- Subdensity function

$$f_i(t) = \frac{\partial F_i(t)}{\partial t}$$

- Subhazard

$$\tilde{h}_i(t) = \lim_{\partial t \rightarrow 0} \left\{ \frac{P(t < T \leq t + \partial t, C = i | T > t)}{\partial t} \right\} = \frac{f_i(t)}{1 - F(t)}$$

- Hazard of the subdistribution

$$\gamma_i(t) = \lim_{\partial t \rightarrow 0} \left\{ \frac{P(t < T \leq t + \partial t, C = i | T > t \text{ or } T \leq t \ \& \ C \neq i)}{\partial t} \right\} = \frac{f_i(t)}{1 - F_i(t)}$$

# Basic concepts (CR) - Example

$$F(t_1, t_2) = 1 - e^{-\lambda_1 t_1 - \lambda_2 t_2} \quad \text{Observe : } T = \min \{T_1, T_2\}$$

Subdistribution function

$$F_1(t) = \frac{\lambda_1}{\lambda_1 + \lambda_2} (1 - e^{-(\lambda_1 + \lambda_2)t})$$

Subdensity

$$f_1(t) = \lambda_1 e^{-(\lambda_1 + \lambda_2)t}$$

CIF

Subhazard

$$\tilde{h}_1(t) = \lambda_1$$

Modelled

Hazard of subdistribution

$$\gamma_1(t) = \frac{f_1(t)}{1 - F_1(t)}$$

# Basic concepts-Subhazard

## Independent

$$(T_1, T_2) \sim \text{Exp}(\lambda_1, \lambda_2)$$

$$S(t_1, t_2) = e^{-\lambda_1 t_1 - \lambda_2 t_2}$$

Subhazard

$$\tilde{h}_1(t) = \lambda_1$$

## Dependent

$$(T_1, T_2) \sim \text{Exp}(\lambda_1, \lambda_2, \mu)$$

$$S(t_1, t_2) = e^{-\lambda_1 t_1 - \lambda_2 t_2 - \mu t_1 t_2}$$

Subhazard

$$\tilde{h}_1(t) = \lambda_1 + \mu t$$

# Outline

- Background (SAS/R)
- Competing risks (CR) introduction
- Theoretical background
- Non-parametric estimation of the probability of event
- Kaplan-Meier estimate in the presence of competing risks

# Estimation of the distribution function (no CR)

## Probability of event

$t_1 < t_2 < \dots < t_r$       Ordered time points of the events

$d_j$       Number of events of interest at time  $t_j$

$n_j$       Number at risk just before  $t_j$

$\hat{S}(t) = \prod_{t_j \leq t} \frac{n_j - d_j}{n_j}$       Kaplan-Meier, product-limit estimator

$\hat{F}(t) = 1 - \hat{S}(t) = \sum_{\text{all } j, t_j \leq t} \frac{d_j}{n_j} \hat{S}(t_{j-1})$       Estimates the probability of event

# Estimation of the subdistribution (CR) Probability of the event of interest

$t_1 < t_2 < \dots < t_r$       Ordered time points of **all types** of events

$d_{ev j}$       Number of events of interest at time  $t_j$

$n_j$       Number at risk just before  $t_j$

$S(t_j)$       Probability free of any event at time  $t_j$

$$\hat{F}_{ev}(t) = \sum_{\text{all } j, t_j \leq t} \frac{d_{ev j}}{n_j} \hat{S}(t_{j-1})$$

Kalbfleisch, JD; Prentice, RL, *The Statistical Analysis of Failure Time Data*, Wiley, New York, (1980)

Proof of previous equality:

$$\begin{aligned}
 1 - KM_{ev}(t_j) &= 1 - \prod_{i=1}^j \frac{n_i - d_{evi}}{n_i} \\
 &= 1 - \prod_{i=1}^{j-1} \frac{n_i - d_{evi}}{n_i} \times \frac{n_j - d_{evj}}{n_j} \\
 &= 1 - KM_{ev}(t_{j-1}) \times \frac{n_j - d_{evj}}{n_j} \\
 &= 1 - KM_{ev}(t_{j-1}) \times \left( 1 - \frac{d_{evj}}{n_j} \right) \\
 &= 1 - KM_{ev}(t_{j-1}) + KM_{ev}(t_{j-1}) \times \frac{d_{evj}}{n_j} \\
 &= \sum_{i=1}^j KM_{ev}(t_{i-1}) \times \frac{d_{evi}}{n_i}
 \end{aligned}$$

# Hodgkin's disease cohort

Hodgkin's disease is a type of cancer which affects the young population.

Survival (of all causes, early stage disease) at 15 years = 71%

## General question

- The long term side effects of treatment, specifically the risk for other malignancies

# Description of events

- 384 alive, no recurrence (R), no second malignancy (M)
- 99 dead (D) without R or M
- 91 M -> 93 M
- 254 R -> 289 R
- 37 R and M
  - 2 malignancies before recurrence

## Types of events:

- Second malignancy (M), coded as 1
- Recurrence (R), coded as 2
- Death (D), coded as 3

# *cuminc* macro in SAS

```
data hd;set hd;  
time=min(maltime,dftime);  
cens=(mcens=1 and  maltime<dftime)+  
+ 2*(rcens=1 and  dftime<=maltime)+  
+ 3*(stat=1 and  rcens=0 and  mcens=0);  
group=0;  
run;  
%cuminc(ds=hd,time=time,cenvble=cens,interest=1,  
        group=group);  
run;
```

# cuminc macro in SAS - output

-----group=0-----

	Number of Number time left	Number type 1 events	Total number of events	CIF for type 1 events	Variance for CIF type 1 events	CP for type 1 events	Variance for CP type 1 events
9.476	510	0	1	0.017560	0.000020	0.028136	0.000051
9.728	506	0	1	0.017560	0.000020	0.028190	0.000052
9.758	505	0	1	0.017560	0.000020	0.028244	0.000052
<b>9.944</b>	<b>503</b>	<b>1</b>	<b>1</b>	<b>0.018758</b>	<b>0.000022</b>	0.030172	0.000055
10.084	501	0	1	0.018758	0.000022	0.030231	0.000056
10.182	500	1	1	0.019960	0.000023	0.032166	0.000059

# Deriving the time and censor variables

```
> ## create time and censoring variables  
> time=apply(cbind(hd$dftime,hd$maltime),1,min)  
> cens=(hd$mcens==1 & hd$maltime<hd$dftime)+  
+ 2*(hd$rcens==1 & hd$dftime<=hd$maltime)+  
+ 3*(hd$stat==1 & hd$rcens==0 & hd$mcens==0)  
> table(cens)
```

```
cens  
  0    1    2    3  
384  93 289  99
```

# Estimating the probabilities

```
> fit=cuminc(time,cens)
```

```
> fit
```

**Estimates and Variances:**

**\$est**

**10 20 30**

**1 1** 0.01875841 0.0958677 0.2134645

**1 2** 0.32093295 0.3376243 0.3376243

**1 3** 0.05735474 0.1060812 0.1659892

**\$var**

**10 20 30**

**1 1** 2.162134e-05 0.0001324540 0.0007336042

**1 2** 2.530717e-04 0.0002643918 0.0002643918

**1 3** 6.340779e-05 0.0001210085 0.0005296034

```
> timepoints(fit,times=c(5,10,15,20))
```

**\$est**

**5 10 15 20**

**1 1** 0.004634889 0.01875841 0.04009511 0.0958677

**1 2** 0.285780189 0.32093295 0.33098900 0.3376243

**1 3** 0.020862345 0.05735474 0.09132882 0.1060812

**\$var**

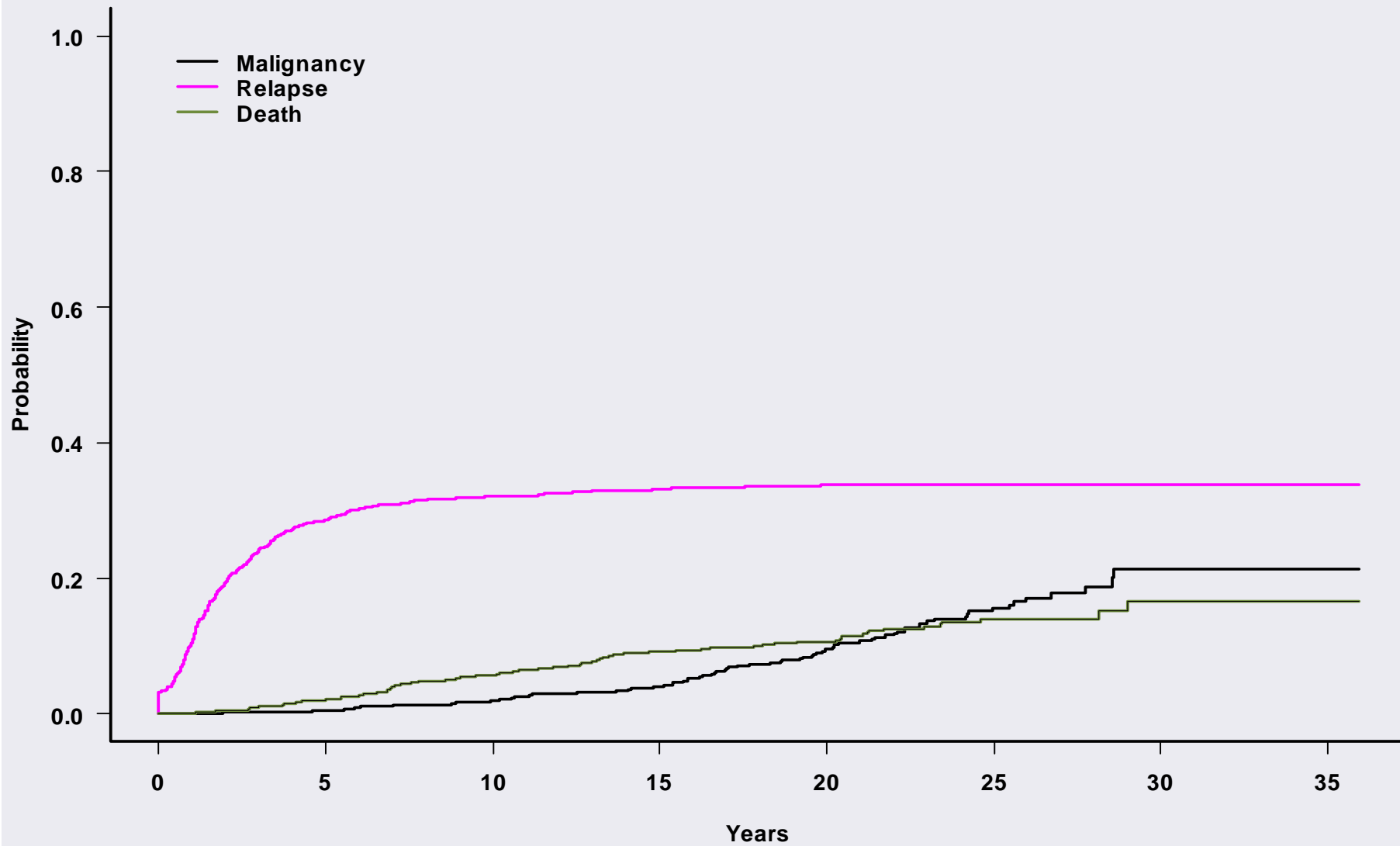
**5 10 15 20**

**1 1** 5.353571e-06 2.162134e-05 4.698782e-05 0.0001324540

**1 2** 2.365398e-04 2.530717e-04 2.581295e-04 0.0002643918

**1 3** 2.371237e-05 6.340779e-05 1.002201e-04 0.0001210085

```
> plot.cuminc(fit,lty=1,col=c(cc,cc1,cc2),  
  curvlab=c('Malignancy','Relapse','Death')  
  )
```



# The variance and confidence intervals

- Aalen's variance estimator
- Delta method
- *cuminc* and SAS macro *cuminc* calculate the Aalen's estimator
- Confidence intervals should be based on  $1 - \hat{F}_i(t) = \hat{S}_i$

$$\left[ \hat{S}_i \right]^{\exp(\pm A)} \quad A = \frac{z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{S}_i)}}{\hat{S}_i \log(\hat{S}_i)}$$

# Outline

- Background (SAS/R)
- Competing risks (CR) introduction
- Theoretical background
- Non-parametric estimation of the probability of event
- Kaplan-Meier estimate in the presence of competing risks

# Non-parametric estimation

## Kaplan-Meier vs. CIF

$$\hat{F}_{ev}(t) = \sum_{all\ j, t_j \leq t} \frac{d_{ev\ j}}{n_j} \hat{S}(t_{j-1})$$

← Based on all types of events

$$1 - KM_{ev}(t) = \sum_{all\ j, t_j \leq t} \frac{d_{ev\ j}}{n_j} KM_{ev}(t_{j-1})$$

Based on events of interest only

# KM in the presence of CR

KM estimates exponent of the negative cumulative subhazard

$$(T_1, T_2) \sim \text{Exp}(\lambda_1, \lambda_2, \mu) \quad S(t_1, t_2) = e^{-\lambda_1 t_1 - \lambda_2 t_2 - \mu t_1 t_2}$$

$$\text{Observe : } T = \min\{T_1, T_2\}$$

Subhazard

$$\tilde{h}_1(t) = \lambda_1 + \mu t$$

Cumulative subhazard

$$\tilde{H}_1(t) = \lambda_1 t + \frac{\mu t^2}{2}$$

**1-KM estimates**

$$1 - e^{-\tilde{H}_1(t)}$$

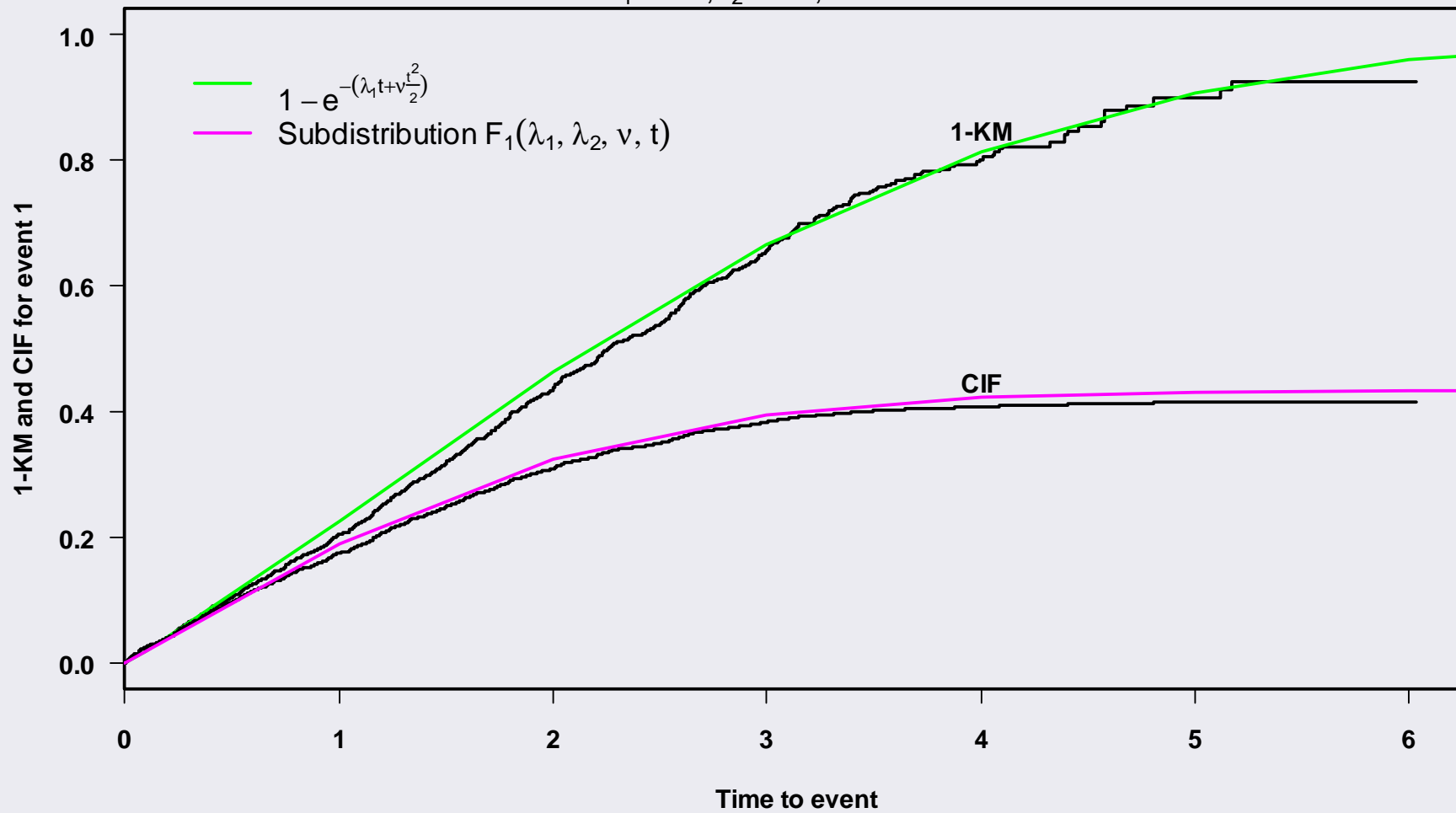
# CIF estimates the subdistribution

$$F_1(t) = \frac{1 - e^{-(\lambda_1 t + \lambda_2 t + \mu t^2)} + G(t, \lambda_1, \lambda_2, \mu)}{2}$$

$$G(t, \lambda_1, \lambda_2, \mu) = (\lambda_1 - \lambda_2) \sqrt{\pi/\mu} e^{\frac{(\lambda_1 - \lambda_2)^2}{4\mu}} \left( \Theta(\alpha_1) - \Theta(\alpha_2) \right)$$

$$\alpha_1 = \sqrt{2\mu} \left( t + \frac{\lambda_1 + \lambda_2}{2\mu} \right) \quad \alpha_2 = \frac{\lambda_1 + \lambda_2}{\sqrt{2\mu}}$$

Data simulated from bivariate exponential  $S(t_1, t_2) = e^{-(\lambda_1 t_1 + \lambda_2 t_2 + \nu t_1 t_2)}$   
 $\lambda_1 = 0.2, \lambda_2 = 0.3, \nu = 0.11$



# Example: Hodgkin's Disease

Median age at onset is 30.

Early stage disease (1,2) is almost curable  
(15% at 20 years, cause specific mortality).

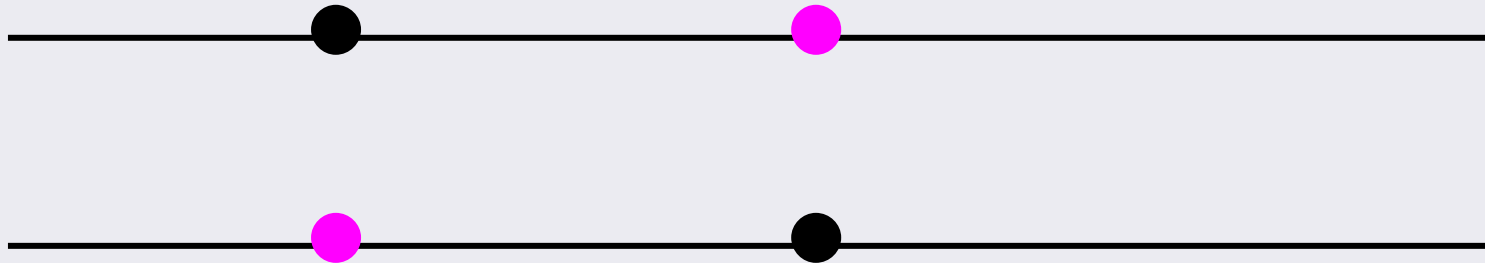
This cohort: 865 patients, diagnosed between  
1968-1986.

Median follow-up is 20 years.

# Example: Hodgkin's Disease

## The events

- 291 relapses
- 128 second malignancies
- 318 deaths



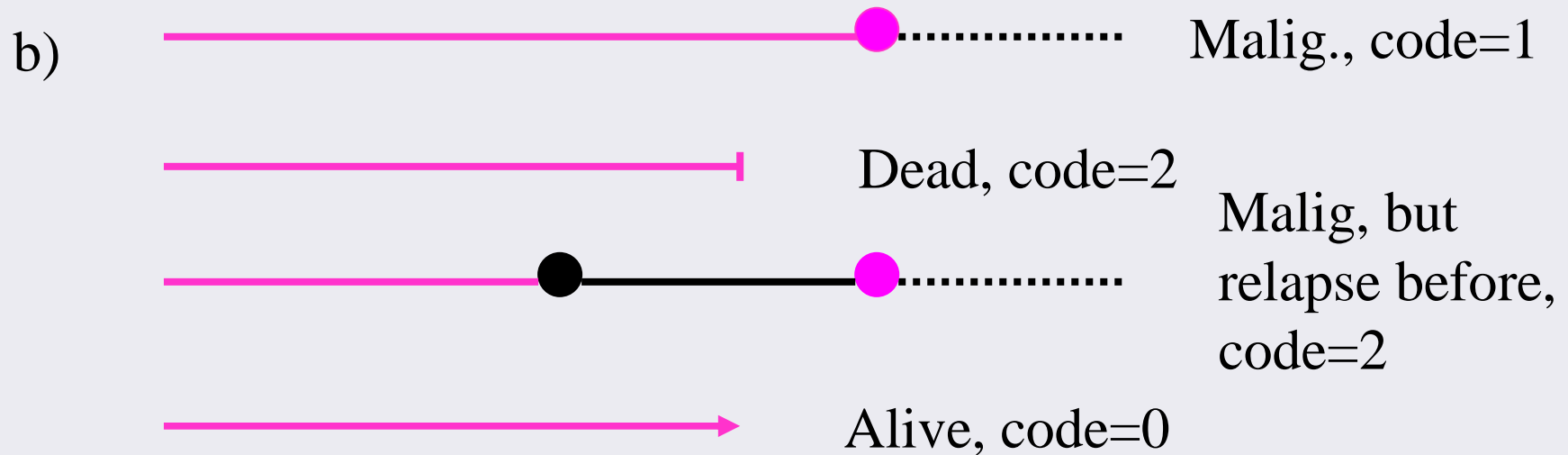
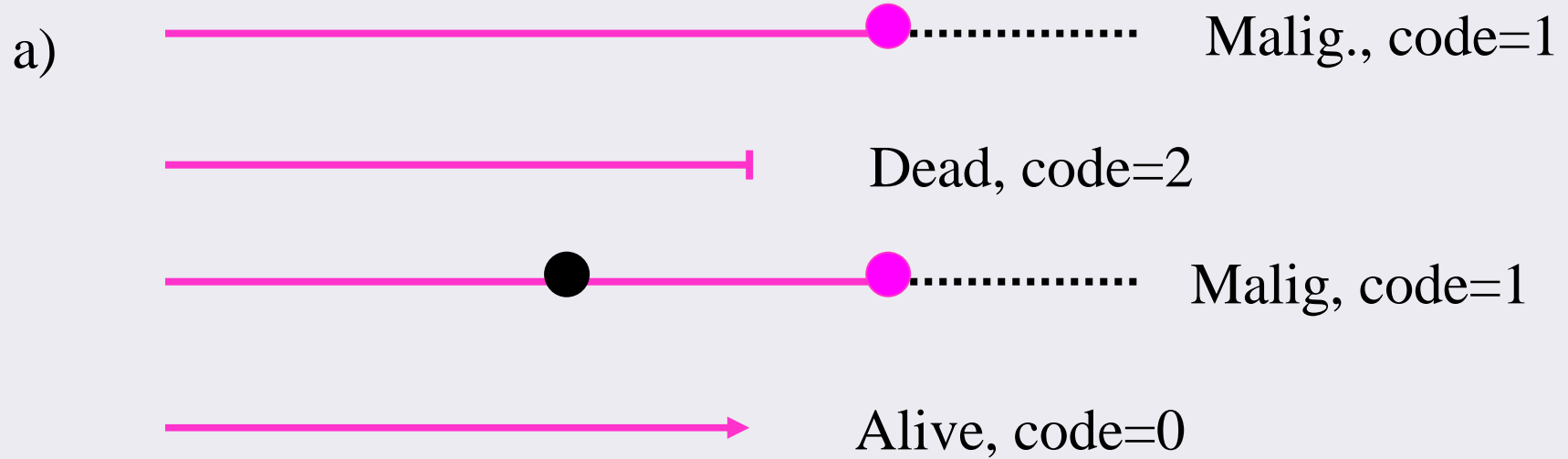
# Example: Hodgkin's Disease

1. Estimate the probability of second malignancy
  - a) CR: Death without malignancy

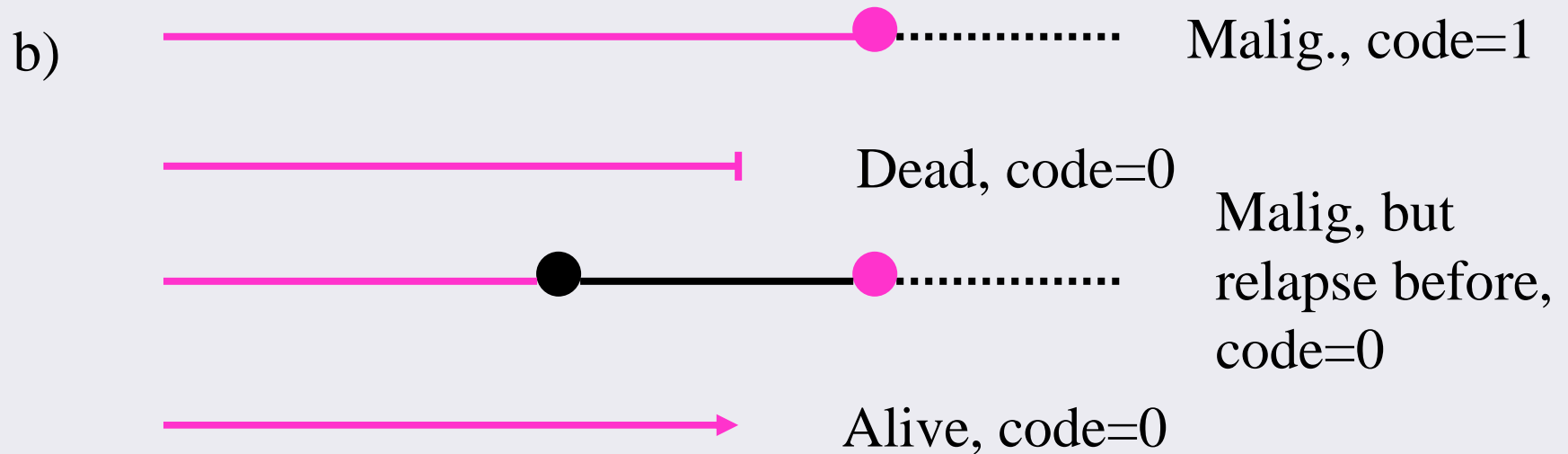
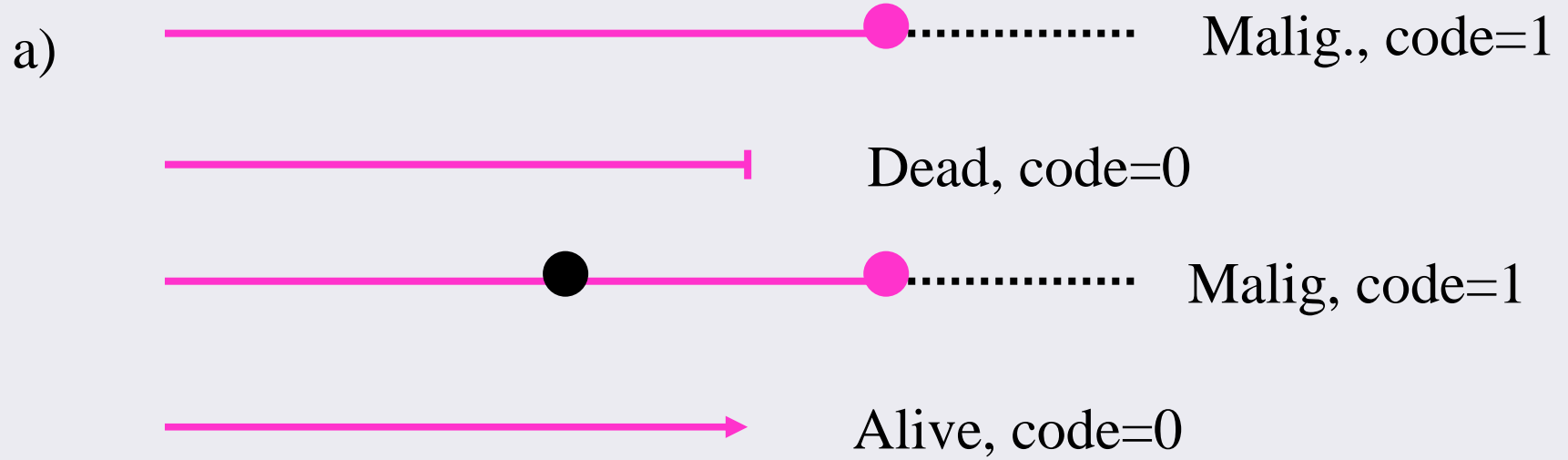
When relapse, second malignancy cannot be detected as easily.

- b) CR: Death without malignancy  
Relapse before malignancy

# For CIF



# For 1-KM



# Example: Hodgkin's Disease

Estimate the probability of second malignancy

## 1-KM estimates:

exp (-cumulative hazard of the marginal)

$$(1-KM)_a \cong (1-KM)_b$$

## CIF

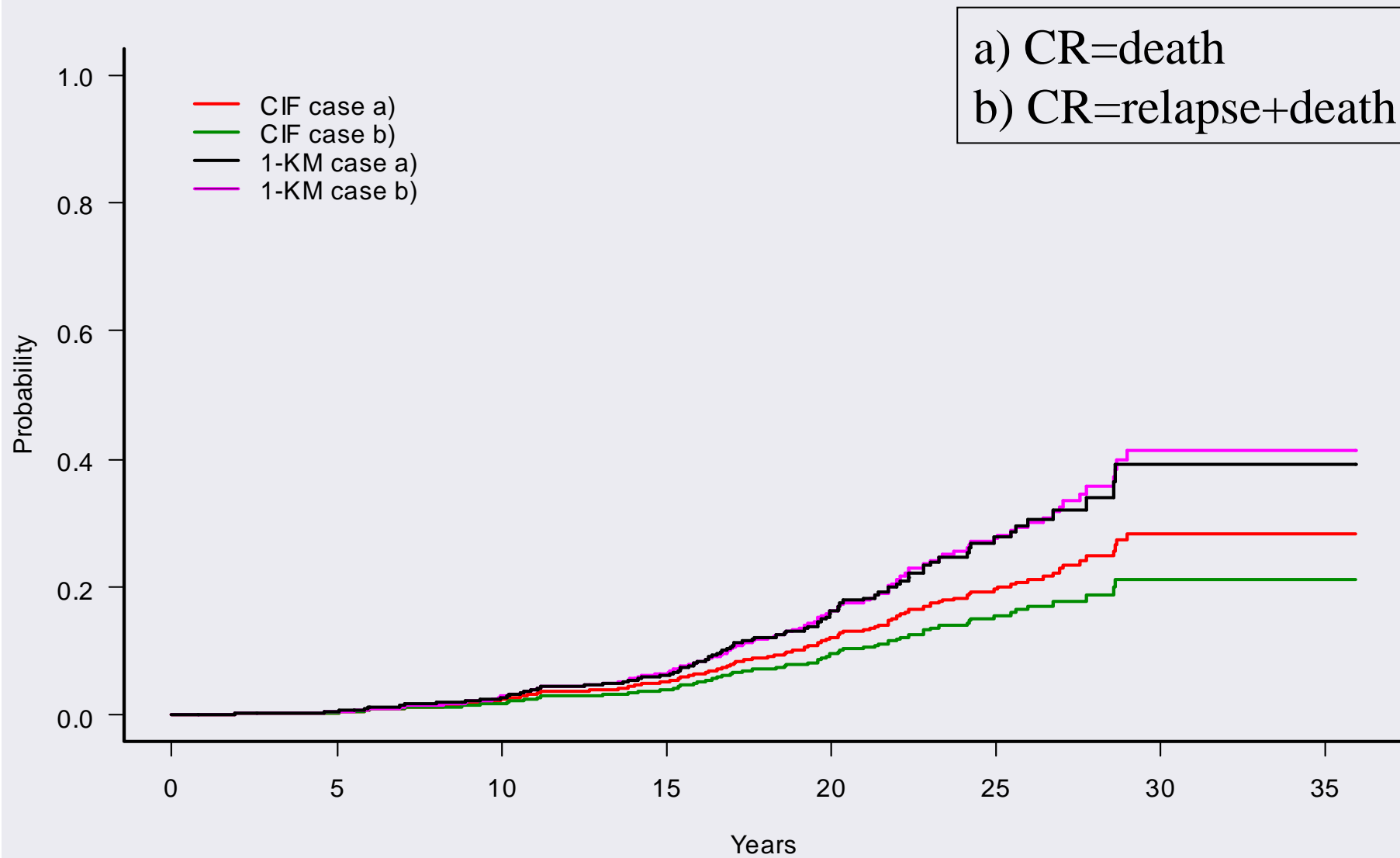
Estimates the observed probability

$$CIF_a > CIF_b$$

$$(1-KM)_{a/b} > CIF_a > CIF_b$$

- |                     |
|---------------------|
| a) CR=death         |
| b) CR=relapse+death |

The interpretation assumes independence between the event of interest and the competing risks.



# Outline

- Testing a covariate
- Fine and Gray model
- Cox proportional hazards model vs. Fine and Gray model
- Prediction
- Non-proportionality of the hazard of subdistribution
- Sample size

# Testing covariates in the presence of CR

- Gray's test , a modified log-rank test
- Pepe and Mori's test , just for 2 categories
- Fine and Gray model, a modified Cox proportional hazards model

# Logrank test (no CR)

Comparison between 2 groups only (for simplicity)

$t_1 < t_2 < \dots < t_r$       Ordered time points of the events

$d_j$       Number of events at time  $t_j$

$n_j$       Number at risk just before  $t_j$

$d_{1j}$       Number of events at time  $t_j$  in group 1

$n_{1j}$       Number at risk just before  $t_j$  in group 1

The numerator:

$$U = \sum_{j=1}^r \left( d_{1j} - n_{1j} \frac{d_j}{n_j} \right)$$

# Modified logrank test (CR)

## Comparison between 2 groups

$t_1 < t_2 < \dots < t_r$       Ordered time points of the events

$d_j$	Number of events of interest at time $t_j$	$d_{1j}$	Number of events of interest at time $t_j$ in group 1
$n_j$	Number at risk just before $t_j$	$n_{1j}$	Number at risk just before $t_j$ in group 1

The numerator:

$$U = \sum_{j=1}^r \left( d_{1j} - R_{1j} \frac{d_j}{R_j} \right)$$

$$R_{1j} = n_{1j} \frac{1 - \hat{F}_1(t_{j-1})}{\hat{S}_1(t_{j-1})}$$

$$R_j = R_{1j} + R_{2j}$$

Gray, R. J. (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *The Annals of Statistics*. **16**, 1141-1154.

# Gray's test

Modified logrank test

$$U\Sigma^{-1}U^t \sim \chi_{m-1}^2$$

Gray, R. J. (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *The Annals of Statistics*. **16**, 1141-1154.

# Probability of malignancy by treatment group

Gray's test, a modified log rank test

cens: 0 = no event, 1=malignancy, 2=relapse or death

**fit=cuminc(time,cens,hd\$trtgiven)**

Tests:

	stat		pv	df
1	16.40404	<b>5.117600e-05</b>		1
2	19.13056	<b>1.220746e-05</b>		1

Estimates and Variances:

**\$est**

		10	20	30
CMT	1	0.040519784	0.16744795	0.3926927
RT	1	0.009967435	0.06693621	0.1636329
CMT	2	0.278230540	0.32971910	0.3667854
RT	2	0.418732349	0.48999606	0.5546442

**\$var**

# Probability of malignancy by treatment group

## Graph

```
forplotM=list(list(fit$'CMT 1'$time,fit$'CMT 1'$est),  
list(fit$'RT 1'$time,fit$'RT 1'$est))  
forplotCR=list(list(fit$'CMT 2'$time,fit$'CMT 2'$est),  
list(fit$'RT 2'$time,fit$'RT 2'$est))
```

```
pvM=signif(fit$Tests[1,2],2)  
pvCR=signif(fit$Tests[2,2],2)  
textM=paste("Gray's test p-value=",pvM)  
textCR=paste("Gray's test p-value=",pvCR)
```

```
est15=round(timepoints(fit,times=15)$est,3)*100  
legendM=paste(c('CMT','RT'),', n=',a,'M at 15y=',est15[1:2],'%')  
legendCR=paste(c('CMT','RT'),', n=',a,'CR at 15y=',est15[3:4],'%')
```

# Probability of malignancy by treatment group

## Graph

```
par(mfrow=c(1,2))  
plot.cuminc(forplotM,curvlab=legendM,  
col=c(cc,cc1),lwd=2,  
xlab='Time to malignancy',ylab='Probability of  
malignancy')  
text(0,0.8,adj=0,font=2,textM)  
  
plot.cuminc(forplotCR,curvlab=legendCR,  
col=c(cc,cc1),lwd=2,  
xlab='Time to CR event',  
ylab='Probability of CR event')  
text(0,0.8,adj=0,font=2,textCR)
```

# Pepe and Mori's test

Compares the CIF's of the two groups

$$s = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \sum_{\text{all } t_j} \left\{ W(t_j) \left[ \hat{F}_1(t_j) - \hat{F}_2(t_j) \right] (t_{j+1} - t_j) \right\}$$

$$W(t_j) = \frac{(N_1 + N_2) \hat{C}_1(t_{j-1}) \hat{C}_2(t_{j-1})}{N_1 \hat{C}_1(t_{j-1}) + N_2 \hat{C}_2(t_{j-1})}$$

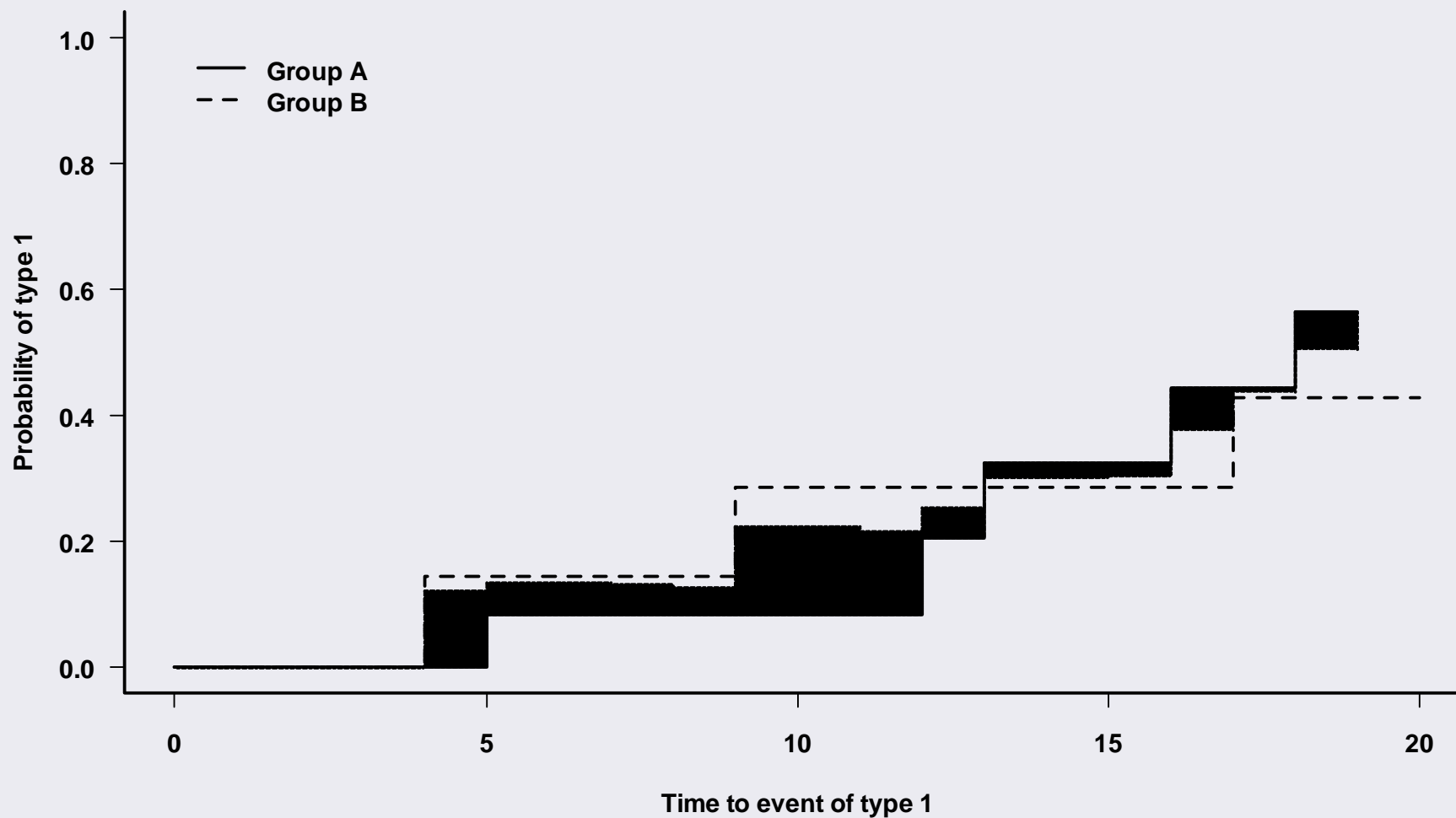
$1 - \hat{C}_i(t)$  = estimator of the censoring distribution of group  $i$  at time  $t$

$\hat{F}_i(t)$  = CIF of group  $i$  at time  $t$

$N_i$  = total number in group  $i$

$$s \sim N(0, \sigma)$$

# The influence of weight on the Pepe and Mori's test



# SAS- Pepe and Mori's test compcif macro

```
%compcif(ds=hd,time=time,cens=cens,group=sex,val1='F',val2='M');
```

Time	Number at risk	Number events of interest	Number competing risks	CIF event of interest	CIF competing risk
0.00	402	0	12	0.000000	0.029851
0.06	390	0	0	0.000000	0.029851
0.21	390	0	0	0.000000	0.029851
0.22	390	0	1	0.000000	0.032338
0.23	389	0	1	0.000000	0.034826
0.25	388	0	1	0.000000	0.037313
0.33	387	0	1	0.000000	0.039801

.....

Group 'F'	Group 'M'	Score	Chi-square	p-value
402	463	2.8606	1.9814	0.15924

# R – Pepe and Mori's test

## R function

```
source("c:/directory_R_functons/compCIF.txt")  
## to second malignancy in hd dataset  
time=apply(cbind(hd$dftime,hd$maltime),1,min)  
cens=(hd$mcens==1 & hd$maltime<hd$dftime)+  
  2*(hd$rcens==1 & hd$dftime<=hd$maltime)+  
  2*(hd$stat==1 & hd$rcens==0 & hd$mcens==0)  
table(cens)  
  
compCIF(time,cens,group=hd$trtgiven)
```

# Pepe and Mori vs. Gray test

## Gray's test

```
cuminc(time,cens,group=hd$trtgiven)
```

```
      stat          pv df
1 16.40404 0.0000512  1
```

## Pepe and Mori's test

```
compCIF(time,cens,group=hd$trtgiven)
```

```
chisquare      pvalue
12.69323 0.0003669812
```

# Outline

- Testing a covariate
- Fine and Gray model
- Cox proportional hazards model vs. Fine and Gray model
- Prediction
- Non-proportionality of the hazard of subdistribution
- Sample size

# Modelling

- No CR: hazard
- CR:
  - subhazard
  - hazard of subdistribution

# Cox proportional hazards model (no CR)

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

$r$  = number of events

$x$  = the covariate

$R_j$  = the risk set at time  $t_j$

$$PL(\beta) = \prod_{j=1}^r \left( \frac{\exp(\beta x_j)}{\sum_{i \in R_j} \exp(\beta x_i)} \right)$$

$$R_j = \{i; t_i \geq t_j\}$$

$$\exp(\beta) = \text{hazard ratio} = HR$$

Fold increase of the hazard due to one unit  
increase of the covariate  $x$

# CR: Cox PH

- Censor variable: 1=event of interest, 0=the rest
- Models the subhazard
- Ignores completely the competing risk
- Independence: models the hazard of the marginal distribution. The model is valid in a virtual world where CR does not exist (CR is ignored).

# Fine and Gray - modelling the hazard of the subdistribution

$$\gamma(t | x) = \gamma_0(t) \exp(\beta x)$$

$$PL(\beta) = \prod_{j=1}^r \left( \frac{\exp(\beta x_j)}{\sum_{i \in R_j} w_{ji} \exp(\beta x_i)} \right) \quad w_{ji} = \frac{\hat{G}(t_j)}{\hat{G}(\min(t_j, t_i))}$$

$$R_j(t) = \{i; T_i \geq t \text{ or } T_i \leq t \text{ and the subject had a CR event}\}$$

$$\hat{G}(t_j) = \text{probability of censoring}$$

Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk . **JASA** 94 (446): 496-509 JUN 1999

Observation	Time	Type of event	X	$\hat{G}$
S1	1	0	12	0.9
S2	2	2	10	0.9
S3	3	1	9	0.9
S4	4	1	13	0.9
S5	5	0	8	0.75
S6	6	2	9	0.75
S7	7	1	12	0.75
S8	8	0	10	0.5
S9	9	1	11	0.5
S10	10	0	8	0

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
3		$\frac{\hat{G}(3)}{\hat{G}(2)} = 1$	1	1	1	1	1	1	1	1
4		$\frac{\hat{G}(4)}{\hat{G}(2)} = 1$		1	1	1	1	1	1	1
7		$\frac{\hat{G}(7)}{\hat{G}(2)} = 0.83$				$\frac{\hat{G}(7)}{\hat{G}(6)} = 1$	1	1	1	1
9		$\frac{\hat{G}(9)}{\hat{G}(2)} = 0.56$				$\frac{\hat{G}(9)}{\hat{G}(6)} = 0.67$			1	1

## Cox PH

- Models the subhazard
- Does not account for competing risks
- It is a comparison in the virtual world where competing risks do not exist (indep)

## Fine and Gray model

- Models the hazard of subdistribution
- Accounts for competing risks
- Models what is observed

# Treatment effect (Malignancy) controlled for: age, sex and stage

```
x=cbind((hd$trtgiven=='CMT')+0,hd$age,  
(hd$sex=='M')+0, hd$clinstg-1)  
fitM=crr(time,cens,x,failcode=1)
```

```
fitM
```

```
convergence: TRUE
```

```
coefficients:
```

```
[1] 0.787500 0.002276 -0.364400 -0.013520
```

```
standard errors:
```

```
[1] 0.210700 0.006429 0.206900 0.240800
```

```
two-sided p-values:
```

```
[1] 0.00019 0.72000 0.07800 0.96000
```

```
Brussels 2008
```

# Treatment effect (Malignancy) controlled for: age, sex and stage

```
x=cbind(trt=(hd$strgiven=='CMT')+0,age=hd$age,  
male=(hd$sex=='M')+0, stage2=hd$clinstg-1)  
fitM=crr(time,cens,x,failcode=1)  
fitM  
coef=fitM$coef  
hr=exp(coef)  
variance=diag(fitM$var)  
pvalue=round(1-pchisq(coef^2/variance,1),6)  
  
data.frame(variables=colnames(x),coef,hr,variance,pvalue)
```

# Better output

```
> data.frame(variables=colnames(x),coef,hr,  
             variance,pvalue)
```

	variables	coef	hr	variance	pvalue
1	trt	0.787452391	2.1977902	4.439053e-02	0.000186
2	age	0.002275838	1.0022784	4.133094e-05	0.723339
3	male	-0.364390530	0.6946199	4.279222e-02	0.078152
4	stage2	-0.013520629	0.9865704	5.798378e-02	0.955223

# Class crr

**\$coef = the estimated regression coefficients**

\$loglik = log pseudo-likelihood evaluated at coef

\$score = derivatives of the log pseudo-likelihood evaluated at coef

\$inf = second derivatives of the log pseudo-likelihood

**\$var = estimated variance covariance matrix of coef**

**\$res = matrix of residuals giving the contribution to each score (columns) at each unique failure time (rows)**

\$uftime = vector of unique failure times

\$bfitj = jumps in the Breslow-type estimate of the underlying sub-distribution cumulative hazard (used by predict.crr())

\$tfs = the tfs matrix (output of tf(), if used)

\$converged = TRUE if the iterative algorithm converged.

# Class crr

$$PL(\beta) = \prod_{j=1}^r \left( \frac{\exp(\beta x_j)}{\sum_{i \in R_j} w_{ji} \exp(\beta x_i)} \right)$$

\$score

$$\frac{\partial \log(PL)}{\partial \beta} = \sum_{j=1}^r \left\{ x_j - \frac{\sum_{i \in R_j} w_{ji} x_i \exp(\hat{\beta} x_i)}{\sum_{i \in R_j} w_{ji} \exp(\hat{\beta} x_i)} \right\}$$

\$res

# Outline

- Testing a covariate
- Fine and Gray model
- Cox proportional hazards model vs. Fine and Gray model
- Prediction
- Non-proportionality of the hazard of subdistribution
- Sample size

# Examples (simulation using exponential distribution, independence, mean of HRs are shown)

$$\lambda_{ev\ 1}=0.1 , \lambda_{ev\ 2}=0.5$$

$$\lambda_{cr\ 1}=0.01 , \lambda_{cr\ 2}=0.01$$

CoxPH, HR=5.1

F&G, HR=5.0

$$\lambda_{ev\ 1}=0.2 , \lambda_{ev\ 2}=0.2$$

$$\lambda_{cr\ 1}=0.1 , \lambda_{cr\ 2}=0.5$$

CoxPH, HR=1.0

F&G, HR=0.5

$$\lambda_{ev\ 1}=0.1 , \lambda_{ev\ 2}=0.5$$

$$\lambda_{cr\ 1}=0.1 , \lambda_{cr\ 2}=0.5$$

CoxPH, HR=5.3

F&G, HR=2.3

$$\lambda_{ev\ 1}=0.1 , \lambda_{ev\ 2}=0.5$$

$$\lambda_{cr\ 1}=0.5 , \lambda_{cr\ 2}=0.1$$

CoxPH, HR=5.3

F&G, HR=9.0

# Example: Hodgkin's Disease

younger subcohort (age<50, n=1418)

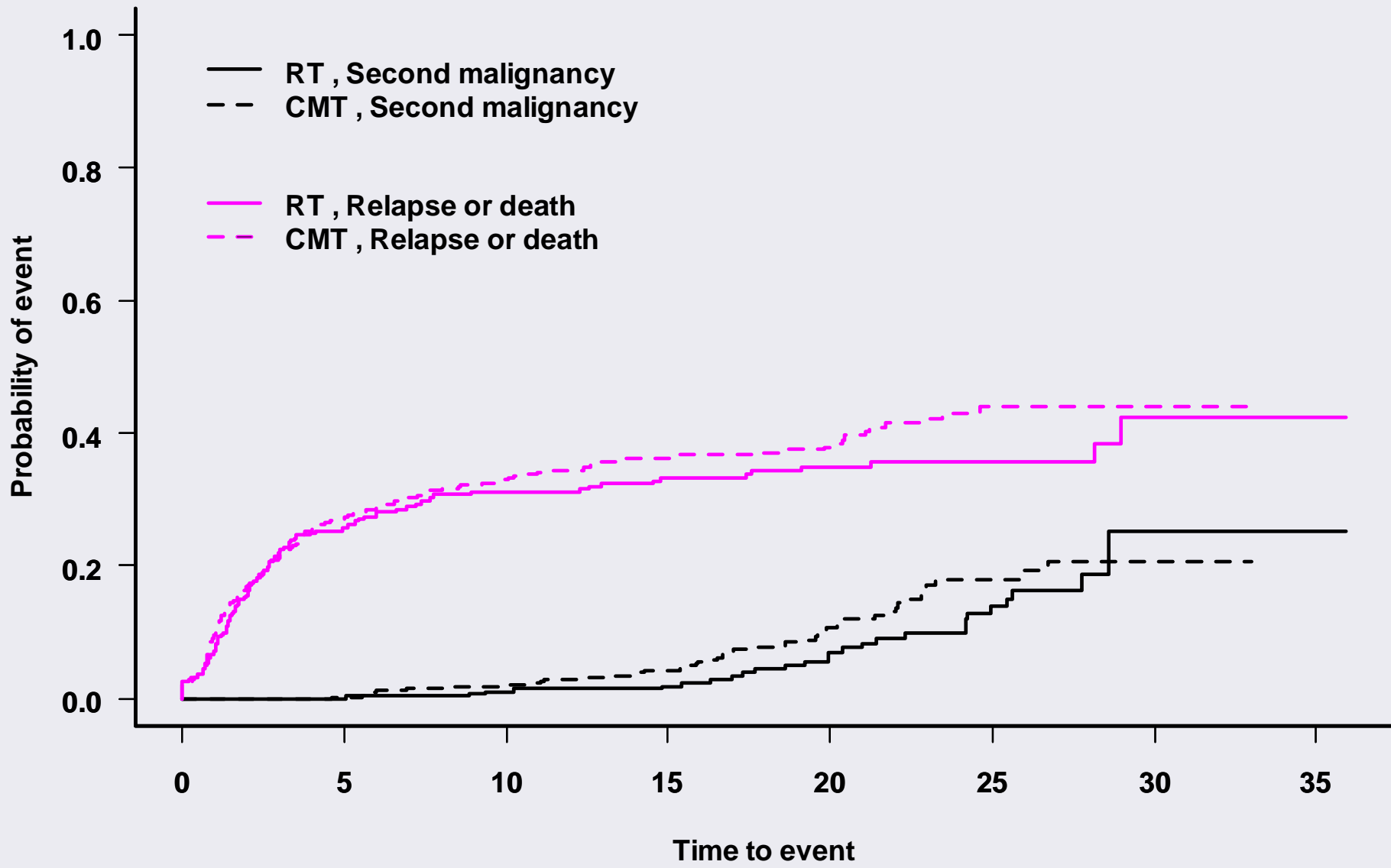
## Events

- 154 patients with second malignancy
- 534 deaths without second malignancy

## Treatment

- 562 Radiation (RT)
- 856 Chemotherapy+Radiation (CMT)

Question: Is the addition of chemotherapy associated with the risk for second malignancy



# Example: Hodgkin's Disease, Discussion

- The addition of Chemotherapy (CMT group) may increase the risk for second malignancy
- This is not a randomized study. It is conceivable that CMT was given for more advanced stage disease. These patients have a higher risk of death without second malignancy, probably due to disease. The CMT group has a higher risk of CR.
- Analysis when CR are accounted for may show a diminished effect of CMT on second malignancy
- Analysis when ignoring CR shows the effect in a virtual world where CR does not exist.

# Example: Hodgkin's Disease, Results

Covariate: Treatment: CMT vs. RT

For event of interest: second malignancy

Analysis	HR	p-value
Cox PH	1.5	0.02
Fine and Gray	1.3	0.12

For CR: death without malignancy

Analysis	HR	p-value
Cox PH	1.16	0.09
Fine and Gray	1.15	0.12

# Example Hodgkin's Disease, Conclusions

- The CMT group has a larger risk for second malignancy than the RT group
- The CMT group also has a larger risk for death without malignancy (possible death due to disease)
- There are fewer second malignancies observed in CMT because of the existence of CR
- What are the other causes? Disease? Other?  
Modelling and controlling for other factors.

# Outline

- Testing a covariate
- Fine and Gray model
- Cox proportional hazards model vs. Fine and Gray model
- Prediction
- Non-proportionality of the hazard of subdistribution
- Sample size

# Prediction when

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

$$S(t | x) = \exp(-H(t | x)) = \exp(-H_0(t) \exp(\beta x))$$

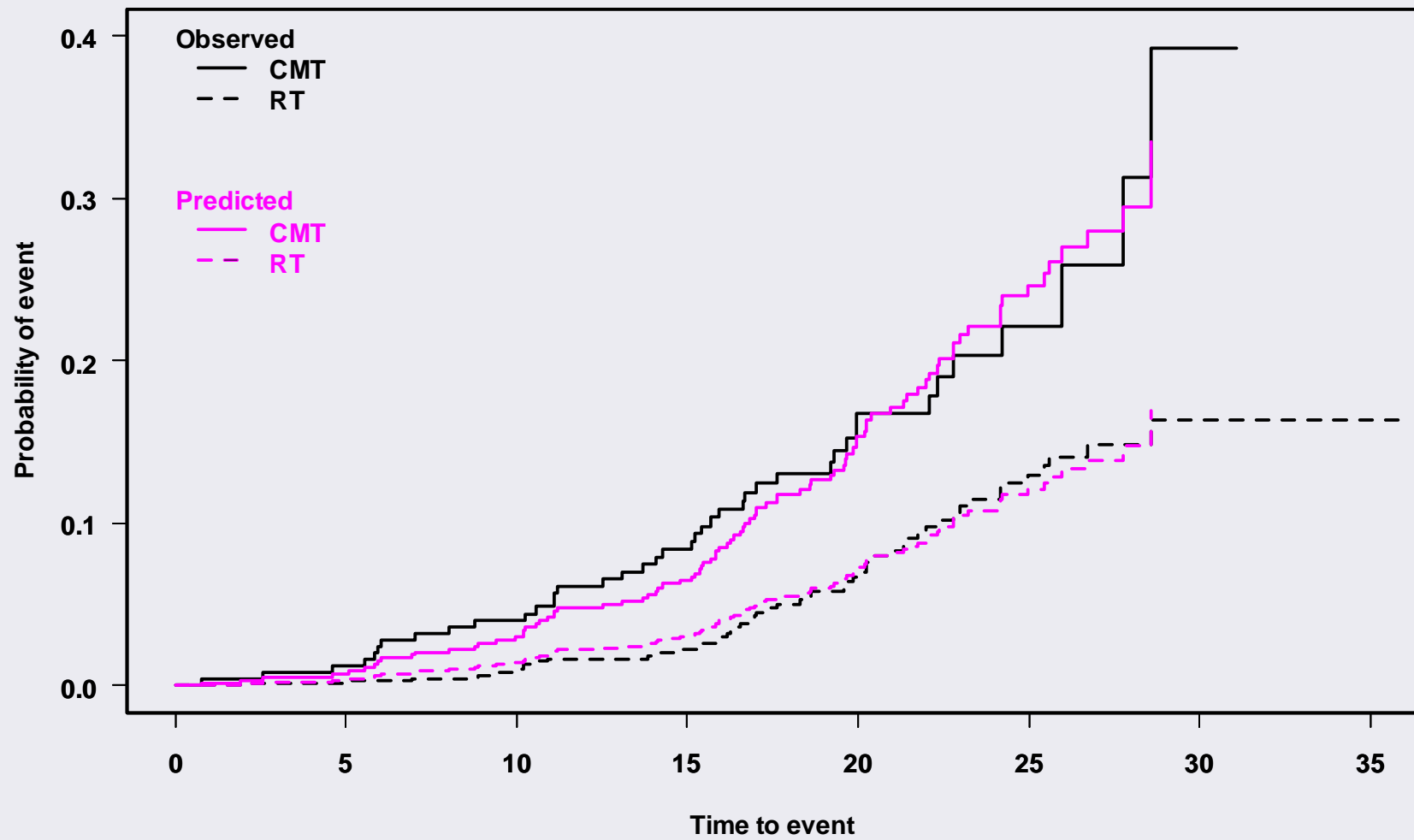
**The jumps in  $H_0$  are in  $\beta x_i$  of the crr object**

$$H_0(t) = \sum_{j=1}^r \left( \frac{d_j}{\sum_{i \in R_j} w_{ji} \exp(\beta x_i)} \right)$$

# Prediction

```
rt=(hd$trtgiven=='RT')+0
fit=cuminc(time,cens,hd$trtgiven)
forplot=list(list(fit$'CMT 1'$time,fit$'CMT 1'$est), list(fit$'RT 1'$time,fit$'RT 1'$est))
plot.cuminc(forplot,wh=c(2,3),ylim=c(0,0.4),xlim=c(0,35),
            xlab='Time to event',ylab='Probability of event')
```

```
fit=crr(time,cens,rt)
pfit=predict.crr(fit,cov1=c(0,1))
par(new=T)
plot.predict.crr(pfit,lty=c(1,2),ylim=c(0,0.4),col=cc1, xlim=c(0,35),xlab=' ',ylab=' ')
text(0,0.4,adj=0,'Observed')
legend(0,0.4,lty=c(1,2),bty='n',col=cc, legend=c('CMT','RT'))
text(0,0.3,adj=0,col=cc1,'Predicted')
legend(0,0.3,lty=c(1,2),bty='n',col=cc1,text.col=cc1, legend=c('CMT','RT'))
```



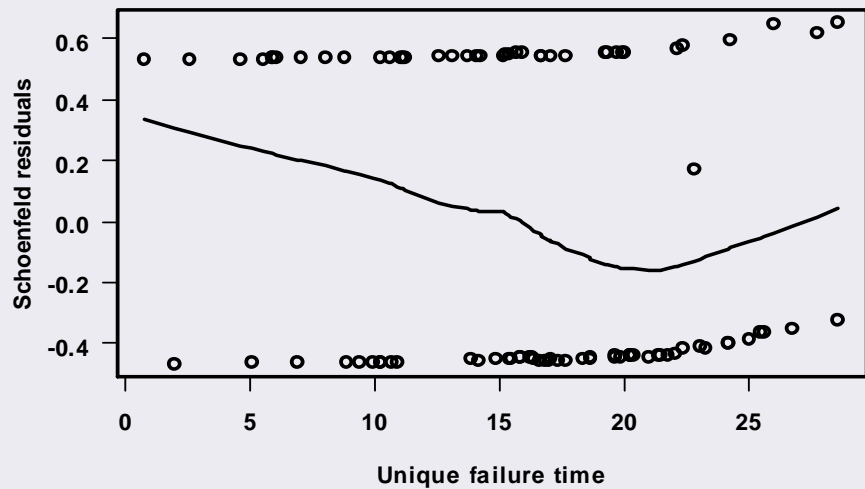
# Outline

- Testing a covariate
- Fine and Gray model
- Cox proportional hazards model vs. Fine and Gray model
- Prediction
- Non-proportionality of the hazard of subdistribution
- Sample size

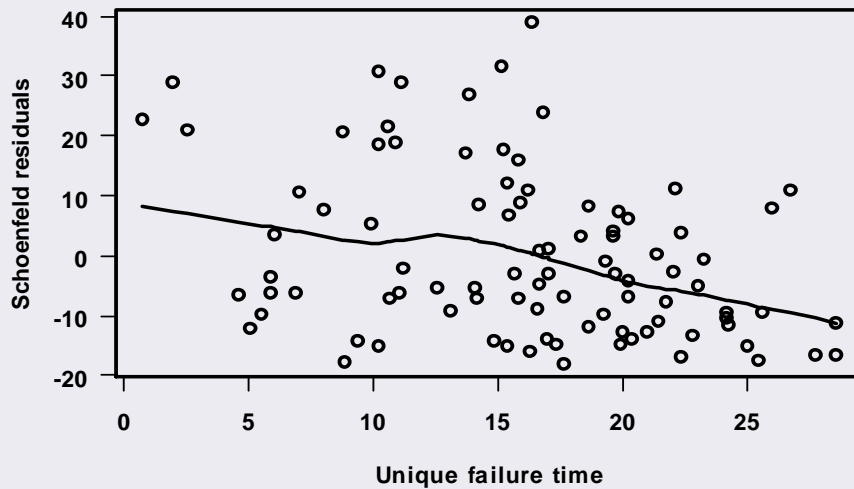
# Checking PH using Schoenfeld-like residuals

```
x=cbind((hd$strtgiven=='CMT')+0,hd$age,  
(hd$sex=='M')+0, hd$clinstg-1)  
fitM=crr(time,cens,x,failcode=1)  
  
plot(fitM$uftime,fitM$res[,1],  
      xlab='Unique failure time',ylab='Schoenfeld  
      residuals',main='Treatment')  
lines(lowess(fitM$uftime,fitM$res[,1]))
```

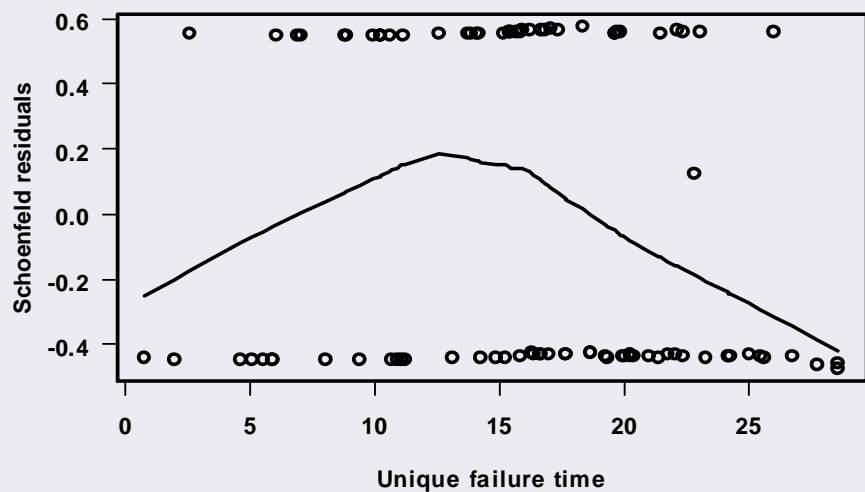
Treatment



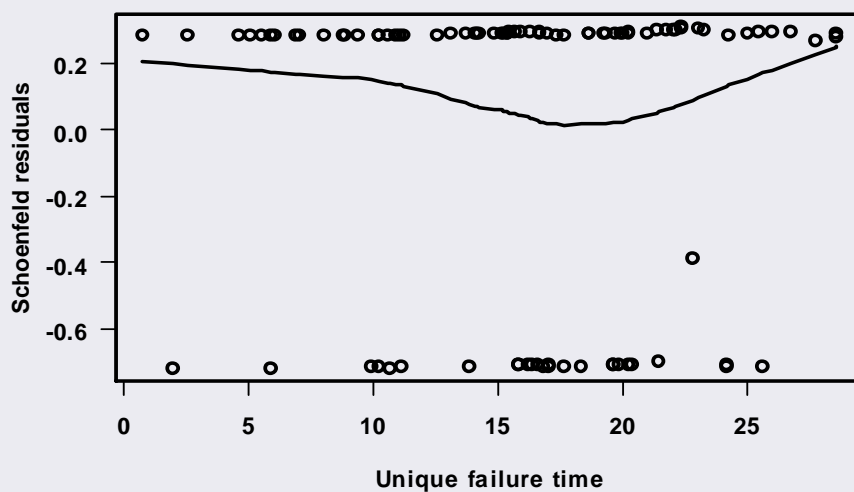
Age



Sex



Clinical stage



- Coefficient of treatment varies probably quadratically with time
- Coefficient of age varies linearly with time
- Coefficient of sex varies quadratically with time
- Coefficient of clinical stage does not vary with time

```
x=treatment , age, sex, clinical stage
```

```
xx=2 x treatment, age, 2 x sex, none for stage
```

```
xx=cbind(x[,1],x[,1],x[,2],x[,3],x[,3])
```

Function for creating time variables for the interaction with these variables: linearly or quadratic

```
tf=function(x)
```

```
{
```

```
y1=x
```

```
y2=x^2
```

```
one=rep(1,length(x))
```

```
A=cbind(y1,y2,y1,y1,y2)
```

```
return(A)
```

```
}
```

# Results

```
> fitMtdep=crr(time,cens,x,xx,tf)
```

```
> fitMtdep
```

coefficients:

```
[1] 3.225000 0.058760 -2.361000 -0.029560 -0.256600 0.005733 -0.003629
```

```
[8] 0.402500 -0.015220
```

standard errors:

```
[1] 1.2900000 0.0161600 1.1450000 0.2392000 0.1658000 0.0050660 0.0009406
```

```
[8] 0.1650000 0.0056190
```

two-sided p-values:

```
[1] 0.01200 0.00028 0.03900 0.90000 0.12000 0.26000 0.00011 0.01500 0.00670
```

Order of parameters (9):

Treatment at time 0, age at time 0, sex at time 0, clinical stage, linear component of the interaction between treatment and time, quadratic component of the interaction between treatment and time, the interaction between age and time, linear component of the interaction between sex and time, quadratic component of the interaction between sex and time.

# A more friendlier output

	variables	coef	hr	variance	pvalue
1	trt	3.225440994	25.16466917	1.663698e+00	0.012397
2	age	0.058756941	1.06051744	2.612291e-04	0.000278
3	male	-2.361206604	0.09430636	1.311654e+00	0.039237
4	stage2	-0.029562088	0.97087060	5.719961e-02	0.901627
5	trt lin	-0.256555129	0.77371234	2.747765e-02	0.121691
6	trt quadr	0.005732799	1.00574926	2.566835e-05	0.257830
7	age lin	-0.003628604	0.99637797	8.847645e-07	0.000114
8	sex lin	0.402505853	1.49556768	2.724015e-02	0.014738
9	sex quadr	-0.015222971	0.98489231	3.157711e-05	0.006748

# Results

Parameters	Coefficient	p-value
Treatment	3.23	0.012
Age	0.059	0.00028
Sex	-2.36	0.039
Clinical stage	-0.030	0.90
Treatment x T	-0.26	0.12
Treatment x T <sup>2</sup>	0.0057	0.26
Age x T	-0.0036	0.00011
Sex x T	0.40	0.015
Sex x T <sup>2</sup>	-0.015	0.0067

# Results

w/o quadratic term for treatment

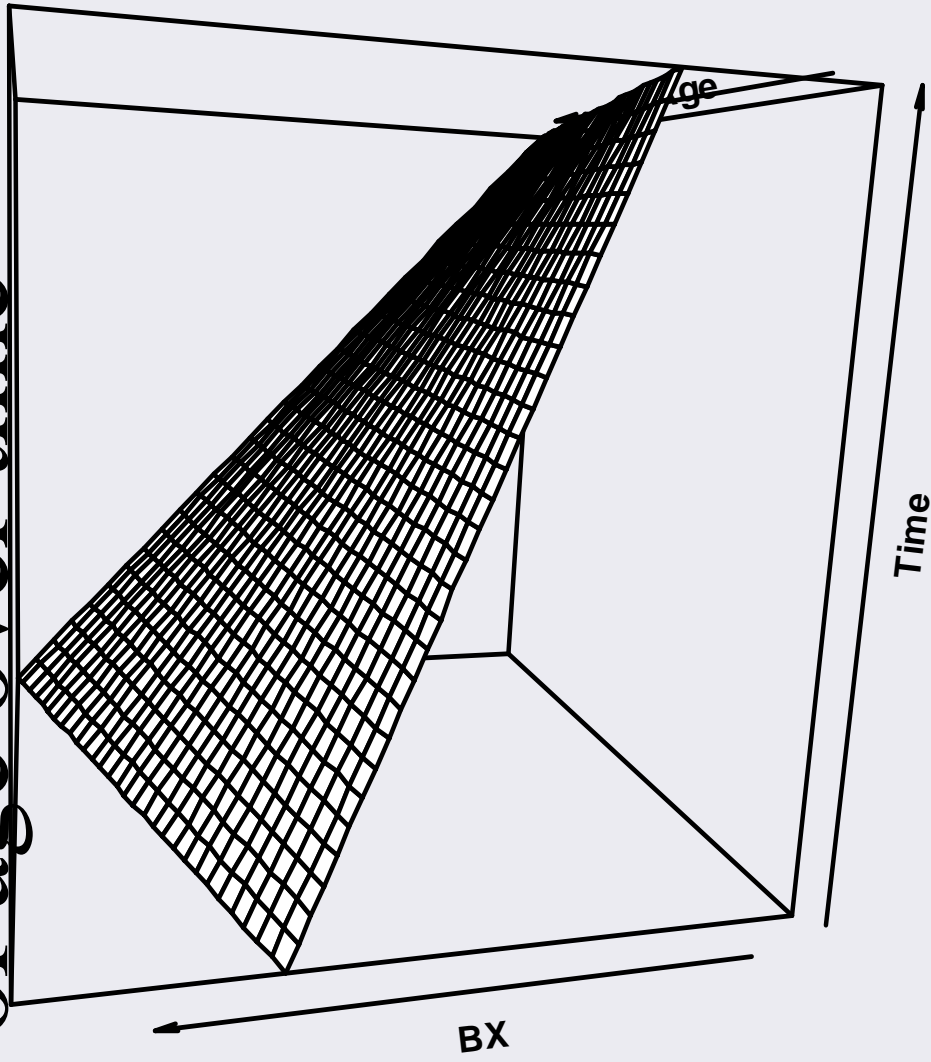
Parameters	Coefficient	p-value
Treatment	2.1	0.001
Age	0.057	0.00024
Sex	-2.36	0.040
Clinical stage	-0.036	0.88
Treatment x T	-0.081	0.029
Age x T	-0.0035	0.00009
Sex x T	0.40	0.014
Sex x T <sup>2</sup>	-0.015	0.0061

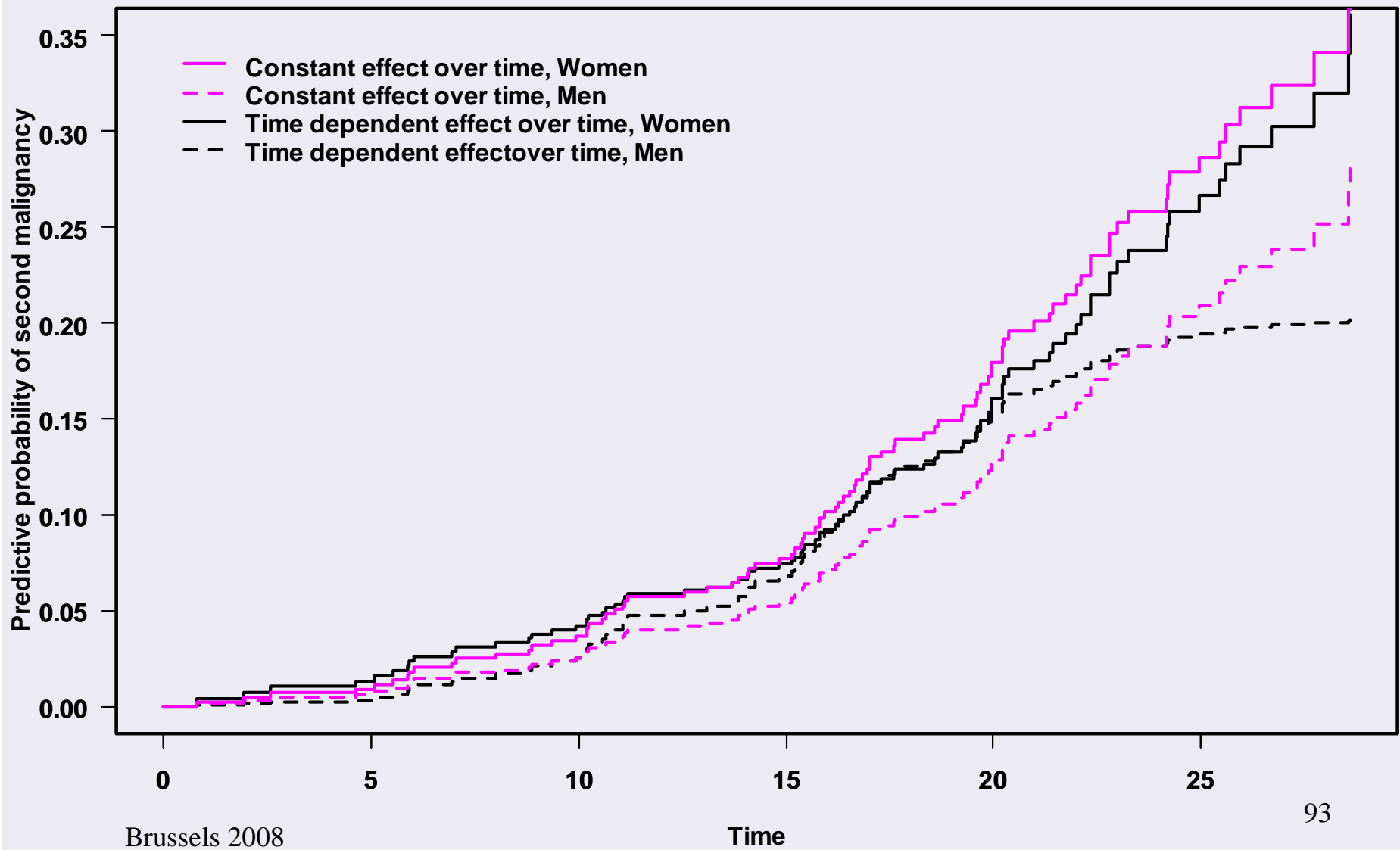
# Comparison between with and w/o time varying coefficients

Parameters	Coef.	p-value
Treatment	0.79	0.00019
Age	0.0023	0.72
Sex	-0.36	0.078
Clinical stage	-0.014	0.96

Parameters	Coefficient	p-value
Treatment	2.1	0.001
Age	0.057	0.00024
Sex	-2.36	0.040
Clinical stage	-0.036	0.88
Treatment x T	-0.081	0.029
Age x T	-0.0035	0.00009
Sex x T	0.40	0.014
Sex x T <sup>2</sup>	-0.015	0.0061

# Effect of age over time





# Outline

- Testing a covariate
- Fine and Gray model
- Cox proportional hazards model vs. Fine and Gray model
- Prediction
- Non-proportionality of the hazard of subdistribution
- **Sample size**

# Sample size calculation

- $\alpha$ , Type I error = probability to find a significance when none exists
- $\beta$ , Type II error = probability of not finding an existing difference
- $\text{power} = 1 - \beta$ , probability to detect a specific effect

# Sample size calculation (no CR)

$$\sqrt{n_{ev}} = \frac{(z_{1-\alpha/2} + z_{1-\beta})}{\sigma \ln(HR)}$$

$$P_{ev} = 1 - \frac{e^{-\lambda f} - e^{-\lambda(f+a)}}{\lambda a}$$

$$N = \frac{n_{ev}}{P_{ev}}$$

**Assumption:** Exponential distribution

HR = hazard ratio  
to be detected  
 $\sigma$  = standard deviation  
of the covariate  
 $z_{1-\alpha/2}$  = quantile of the  
standard normal

a = accrual time  
f = follow-up time  
 $\lambda$  = the hazard rate  
for the overall

# Sample size calculation (CR)

$$\sqrt{n_{ev}} = \frac{(z_{1-\alpha/2} + z_{1-\beta})}{\sigma \ln(HR)}$$

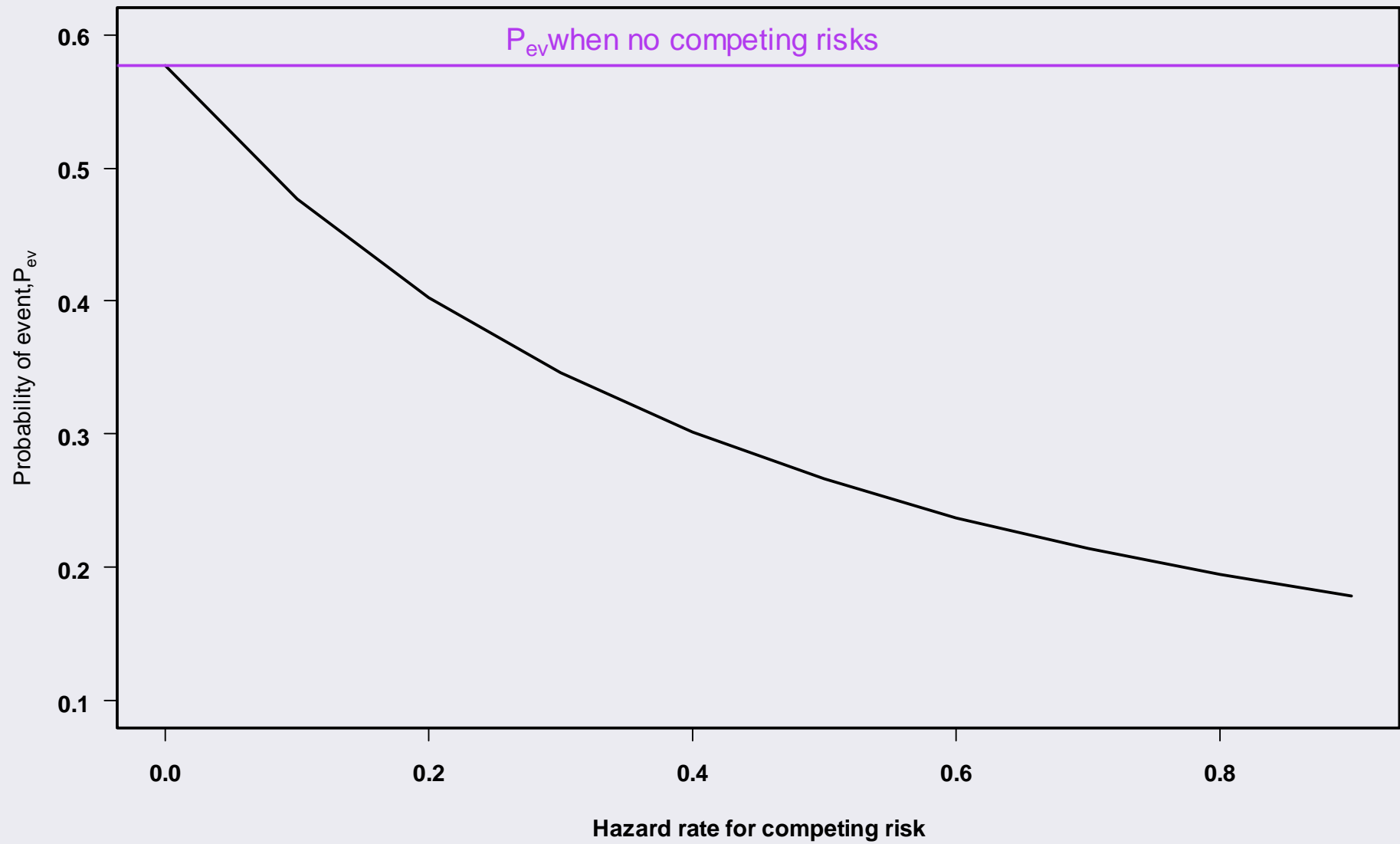
$$P_{ev} = \frac{\lambda_{ev}}{\lambda_{ev} + \lambda_{cr}} \times \left( 1 - \frac{e^{-(\lambda_{ev} + \lambda_{cr}) \times f} - e^{-(\lambda_{ev} + \lambda_{cr}) \times (f+a)}}{(\lambda_{ev} + \lambda_{cr}) \times a} \right)$$

$$N = \frac{n_{ev}}{P_{ev}}$$

$$P_{ev} = 1 - \frac{e^{-\lambda_{ev} f} - e^{-\lambda_{ev} (f+a)}}{\lambda_{ev} a}$$

**Assumption:** Exponential distribution, independence

### Decrease of $P_{ev}$ as $\lambda_{cr}$ increases



# A very simple power calculation

If we know the number of events in data set or the crude rate of events.

A randomized study accrued between 1992-2000. During the study tumour tissue was collected and kept for future use.

Advantages:

- Uniform group of patients
- The allocation of treatment is based on randomization not on patient characteristics

Of interest is whether a specific marker (say expression of HIF1 $\alpha$  gene is prognostic for survival.

# A very simple power calculation

If we know the number of events in data set or the crude rate of events.

A genetic signature was developed for a group of patients with lung cancer. The treatment is fairly uniform for this disease. It is desirable to validate the genetic signature. Then with similar follow-up the  $P_{ev}$  will be similar

$$\sqrt{n_{ev}} = \frac{(z_{1-\alpha/2} + z_{1-\beta})}{\sigma \ln(HR)} \quad \text{PASS could be used:} \quad P_{ev} = \frac{n_{ev}}{N}$$

Essential to use the correct  $P_{ev}$ , crude rate of the events of interest.

# PASS, Regression Cox

- Calculate power under the assumptions:
  - $\alpha=0.05$
  - $N=211$
  - Covariate x continuous with  $STD=1$
  - $HR=1.3$ , coefficient=0.262
  - $P_{ev} = 0.539$

# PASS, Regression Cox

**PASS: Regression: Cox**  
File Run Analysis Graphics PASS Window Help

Symbols 2 Background Axes Options Abbreviations 3D Reports Template  
Plot Text Symbols 1  
Data Plot Setup

Find (Solve For): **Beta and Power**

Hypothesis Test: Two-Sided

N (Sample Size): 211

Alpha (Significance Level): 0.05

Beta (1-Power): 0.20

B (Log Hazard Ratio): 0.262

S (Std Deviation of X1): 1

R-Squared Other X's: 0.0

P (Overall Event Rate): 0.539

**P<sub>ev</sub>**

Template Id: \_\_\_\_\_

Reset Guide Me

**FIND (SOLVE FOR):**  
Select the parameter to be solved for in terms of the other parameters.  
Note that this is the parameter displayed on the vertical axis of the plot.

# All the necessary information needs to be estimated

$\alpha, \beta$ , HR and  $\sigma$  (standard deviation for the covariate)

**The necessary NUMBER of EVENTS**

$\lambda_{ev}$ ,  $\lambda_{cr}$ , a (accrual time), f (follow-up time)

**Probability of event during the study**

Calculate the total number of patients

# Calculating the probability of the event of interest during the study period

Assumption: exponential distribution and independence between event of interest and competing risks.

$$P_{ev} = \frac{\lambda_{ev}}{\lambda_{ev} + \lambda_{cr}} \times \left( 1 - \frac{e^{-(\lambda_{ev} + \lambda_{cr}) \times f} - e^{-(\lambda_{ev} + \lambda_{cr}) \times (f+a)}}{(\lambda_{ev} + \lambda_{cr}) \times a} \right)$$

# Calculating the probability of the event of interest during the study period

The KM ( $\hat{S}$ ) estimates for the event of interest and competing risks at a time point.

$$\lambda_{ev} = -\frac{\log(\hat{S}_{ev}(t_0))}{t_0}$$

$$\lambda_{cr} = -\frac{\log(\hat{S}_{cr}(t_0))}{t_0}$$

# Calculating the probability of the event of interest during the study period

The CIF ( $\hat{F}$ ) estimates for the event of interest and competing risks at a time point.

$$\lambda_{ev} = \hat{F}_{ev}(t_0) \times \frac{-\log(\hat{S}(t_0))}{t_0(1 - \hat{S}(t_0))}$$

$$\hat{S}(t_0) = 1 - \hat{F}_{ev}(t_0) - \hat{F}_{cr}(t_0)$$

$$\lambda_{cr} = \hat{F}_{cr}(t_0) \times \frac{-\log(\hat{S}(t_0))}{t_0(1 - \hat{S}(t_0))}$$

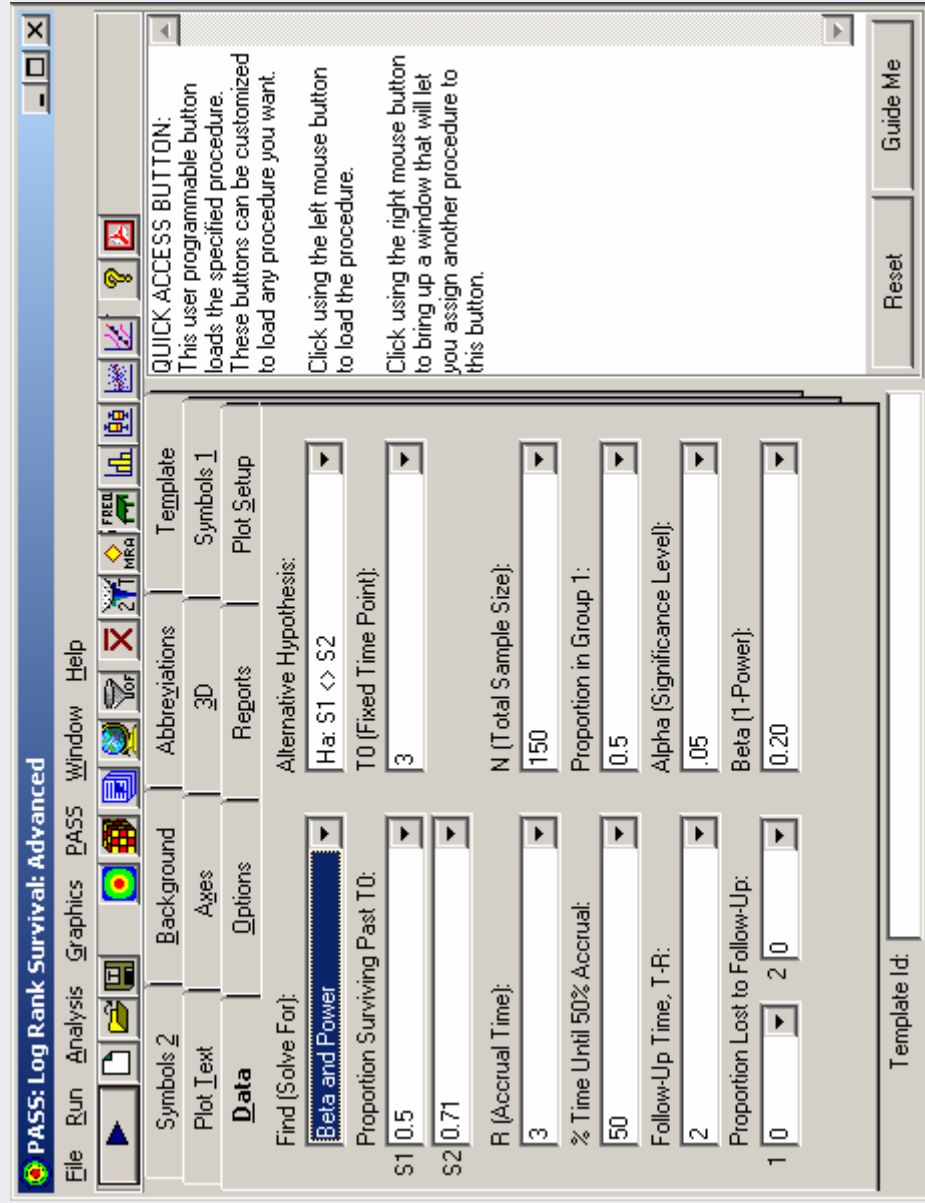
# Example: Sample size

- Goal: The experimental treatment improves distant failure in cervix cancer
- $\alpha=0.05$
- $N=150$ , accrual=3 years, follow-up=2 years
- $HR=2$  (hazard of the marginal)
- Standard:  $\hat{S}_{ev}(3)=0.5$ ,  $\hat{S}_{cr}(3)=0.4$

```
> power(N=150,a=3,f=2,pi=0.5,theta=2,t0=3,  
+ KMev0=0.5,KMcr0=0.4,KMcr1=0.4)
```

```
[1] 0.6162274
```

# Example: Sample size, ignoring CR



The power = 0.79

# Conclusions

- CIF for estimation of probability of event
- Modelling using Cox PH or F & G model. Each has its own interpretation.
- Proper power calculation needs to consider the competing risks regardless of which model is planned to be used.

# References

- Kalbfleisch, JD; Prentice, RL, *The Statistical Analysis of Failure Time Data*, Wiley, New York, (1980)
- Collett, D, *Modelling Survival Data in Medical Research*, Chapman and Hall, London, (1995)
- Crowder, M. *Classical competing risks*. Chapman & Hall/CRC. 2001.
- Gray, R. J. (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *The Annals of Statistics*. **16**, 1141-1154.
- Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, Marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*. **12**, 737-751.
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk . **JASA** 94 (446): 496-509 JUN 1999
- Gooley, TA; Leisenring, W; Crowley, J; Storer, BE, "Estimation of failure probabilities in the presence of competing risks: new representations of old estimators" *Statistics in Medicine* 1999 pp. 695-706