Power and Sample size for time to event analysis

Melania Pintilie

Advanced Statistical Methods for Clinical Trials

Goals

- To understand the 2 types of error in statistics
- To understand the intricacies of sample size calculation for time to event analysis with and without competing risks
- To know what information is necessary
- To familiarize with simulations in R
- To know how and what to report

Type I and II errors

- Significance level (α): The probability of a type I error, that the null hypothesis will be rejected and it will be concluded that there is an effect when in fact there is none. This is usually set to 0.05.
- Power (1-β): where β is the probability of a type II error, the chance of missing a clinically significant difference. B is usually set as 0.2. Power is usually expressed in percentages like 80%.
- Clinically important difference: The smallest difference in means/proportions/etc that the clinicians feel is worth detecting. This difference can be smaller or larger than *observed* differences from other trials.

Power vs. sample size

- The same technical approach (the formulae): one is assumed and the other is calculated
  - Assuming the α level, other parameters (like the effect size) and the power , the sample size can be calculated
  - Assuming the α level, other parameters (like the effect size) and the sample size, the power can be calculated.
- For a randomized clinical trial – sample size
- To test a covariate (like a marker) for which we have the cohort - power
- Sometimes the approach is dictated by what is easier to calculate

The extent of calculation

- A sample size calculation is not usually a single calculation but a **set** of calculations, which can be presented in a table or graph.
- It is important to be aware of how the required sample size varies according to variations in input parameters, **especially those for which the values are more uncertain**.

The extent of calculation

- If the sample size is limited by either
  - Financial constraints
  - Available pool of patients or
  - Time frame for recruiting

Then calculations should focus on what difference can be detected with the sample size available, and/or the power you have to detect a particular effect of interest.

Working on sample size/power

- This is the most ungratifying work for a statistician. One could spend a week on it but only one sentence would be included in the protocol/grant.
- In spite of this it is an important piece in the design of a study. A wrong calculation could render the study:
  - Useless if under-powered. This means that a good treatment will go to waste
  - Too long if over-powered. This means that more patients will be given an inefficient treatment.

Needs to be performed by at least 2 different methods.

1) **"By hand" using the formulae**
2) **Software**
3) **Simulations**

Sample size calculation -survival

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

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

HR = hazard ratio to be detected

σ = standard deviation of the covariate

$z_{1-\alpha/2}$ = quantile of the standard normal

Note that the formula for the number of events can be applied regardless whether the covariate is continuous or categorical.

Understanding the denominator $\sigma ln(HR)$ - normalized effect

- $ln(HR)$ is in fact the coefficient
- Suppose that age is the covariate of interest
- Say age_m=age in month and age_y=age in years with σ_m=standard deviation for age_m and σ_y =standard deviation for age_y

$$age_m = 12\ x\ age_y$$
$$\sigma_m = 12\ x\ \sigma_y$$
$$coef_m = \frac{coef_y}{12}$$
$$\sigma_m coef_m = 12\ \sigma_y \frac{coef_y}{12} = \sigma_y\ coef_y$$

$\sigma ln(HR)$ is invariant to the unit of measure

$\sigma ln(HR)$ is the effect when the covariate is normalized

Calculating the probability of event (death)

a = accrual time, *f* = follow-up time, *λ* = the hazard rate for the overall

$$P_{ev} = \int_0^a P(death\ and\ accrued\ at\ time\ t)dt =$$
$$\int_0^a P(death \mid accrued\ at\ time\ t)x\ P(accrued\ at\ time\ t)dt =$$
$$1 - \frac{1}{a}\int_0^a S(a + f - t)dt = 1 - \frac{1}{a}\int_f^{a+f} S(u)du$$

Assuming uniform accrual

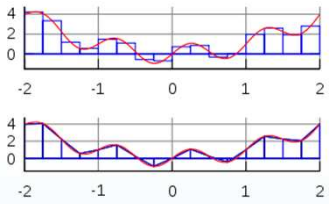
Calculating the integral

No assumption for survival distribution and applying Simpson's rule:

$$P_{ev} = 1 - \frac{1}{6}\{S(f) + 4S(0.5a + f) + S(a + f)\}$$

Assuming exponential distribution

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



Assuming other distributions: It could be difficult to solve. One could approximate the integral using the rectangular or the trapezoidal rule.

Wrapping up: power/sample size for survival

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

$$P_{ev} = 1 - \frac{1}{6}\{S(f) + 4S(0.5a + f) + S(a + f)\}$$

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

A very simple power calculation

If we know or can estimate the number of deaths in the data set.

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 HIF1α gene) is prognostic for survival.

A very simple power calculation

If we know or can estimate the number of deaths in the data set.

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.
With similar follow-up, the P_{ev} will be similar.

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

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

Example 1

- 289 patients diagnosed with early stage (1/2) lung cancer. All had surgery and no other treatment. 120 are dead.
- Aim: to test a covariate which classifies half of the patients as high risk and the other half as low risk.
- The covariate is obtained from several molecular markers. The test is expensive. The tissue exists in the tissue bank for all 289 patients.
- Is it possible to minimize the cost by testing fewer patients?

Details of the calculation

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

$\alpha = 0.05, z_{1-\alpha/2} = 1.96$
$(\text{two} - \text{sided})$

$\beta = 0.2, z_{1-\beta} = 0.84$
$(\text{power} = 80\%)$

$n_{ev} = 120$

$$\sigma = \sqrt{\frac{1}{2} x \frac{1}{2}} = \frac{1}{2}$$

```
> alpha=0.05
> beta=0.2
> zalpha=qnorm(1-alpha/2)
> zbeta=qnorm(1-beta)
> sigma=1/2
> nev=120
>
> coef=(zalpha+zbeta)/(sigma*sqrt(nev))
> exp(coef)
[1] 1.667786
```

If all data are used.

Details of the calculation

It is expected that the HR should be ~2.
A lower effect is not of interest.
Can we test fewer patients?

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

$\alpha = 0.05, z_{1-\alpha/2} = 1.96$
$(\text{two} - \text{sided})$
$\beta = 0.2, z_{1-\beta} = 0.84$
$(\text{power} = 80\%)$
HR= 2

$$\sigma = \sqrt{\frac{1}{2} x \frac{1}{2}} = \frac{1}{2}$$

```
> HR=2
> sqrtnev=(zalpha+zbeta)/(sigma*log(HR))
> sqrtnev^2
[1] 65.34566
> 289*66/120
[1] 158.95
```

YES.
Need 66 deaths=160 patients

HR is an abstract concept
Express the effect as difference in survival

$$HR = \frac{h(high\ risk)}{h(low\ risk)} = \frac{h_2}{h_1} = \frac{\lambda_2}{\lambda_1}$$

$S_i(t) = e^{-\lambda_i t}$ Exponential

$$\frac{S_1(t) + S_2(t)}{2} = S(t) \qquad S(5y) = 0.656$$

$$\begin{cases} \frac{\lambda_2}{\lambda_1} = 2 \\ \frac{e^{-5\lambda_1} + e^{-5\lambda_2}}{2} = 0.656 \end{cases}$$

$$\begin{cases} S_1(5y) = 0.75 \\ S_2(5y) = 0.56 \end{cases}$$ 19% difference in survival at 5 years

Solving this system of equations:
- **Algebraic when possible**
- **Guessing (tedious)**
- **Numeric**
- **A combination of the above**

Simulation

- Assuming exponential
- We can get the parameter from the dataset
- We need 2 parameters: one for each group

$$\begin{cases} S_1(5y) = 0.75 \\ S_2(5y) = 0.56 \end{cases} \qquad \begin{cases} \lambda_1 = 0.058 \\ \lambda_2 = 0.116 \end{cases} \qquad n_1 = n_2 = 80$$

The plan:

- Generate 80 time points ~ Exp(0.058)
- Generate 80 time points ~Exp(0.116)
- Generate 160 time points for censoring
- Generate the status variable based on the event times and the censoring time
- Create the covariate: values of 0 for the first 80 and 1 for the next 80
- Test the covariate
- Repeat many times

Example of code for simulation

This is to make sure it works

```
library(survival)
lambda1=0.058
lambda2=0.116
n=80
timeev1=rexp(n,lambda1)
timeev2=rexp(n,lambda2)
timeev=c(timeev1,timeev2)

timecensor=runif((2*n),1,10)
time=apply(cbind(timeev,timecensor),1,min)
stat=(time==timeev)+0
x=c(rep(0,n),rep(1,n))

fitcox=coxph(Surv(time,stat)~x)
fitkm=survfit(Surv(time,stat)~x)
hr=exp(fitcox$coef)
est=summary(fitkm,times=5)$surv
pvwald=summary(fitcox)$waldtest[3]
pvlrt=summary(fitcox)$logtest[3]
pvscore=summary(fitcox)$sctest[3]
out=data.frame(hr,estlowrisk=est[1],esthighrisk=est[2],pvwald,pvlrt,pvscore)

out
```

Example of code for simulation

```
library(survival)
fsim=function(i,n,lambda1,lambda2,f,a)
{
 timeev1=rexp(n,lambda1)
 timeev2=rexp(n,lambda2)
 timeev=c(timeev1,timeev2)
 ……. …… ……. ……
  out=data.frame(coef,numev,
      estlowrisk=est[1],
esthighrisk=est[2],
    pvwald,pvlrt,pvscore)
 print(i)
 return(out)
}

fsim(n=80,lambda1=0.058,lambda2=
0.116,f=1,a=9)
```

```
set.seed(123)
nsim=10000  ## try it first with 10
a=sapply(c(1:nsim),fsim,n=80,lambda1=0.058,
        lambda2=0.116,f=1,a=9)
#a # when you try it with 10, check to see how it
looks
bcr=data.frame(apply(a,1,unlist))
head(bcr)

> sum(b$pvwald<=0.05)/nsim
[1] 0.7224
> sum(b$pvlrt<=0.05)/nsim
[1] 0.7294
> sum(b$pvscore<=0.05)/nsim
[1] 0.7285
> exp(mean(b$coef))
[1] 2.020777
> mean(b$estlowrisk)
[1] 0.7489374
> mean(b$esthighrisk)
[1] 0.5602127
> mean(b$numev)
[1] 56.9206
```

Follow-up is not uniform
The distribution may not be exponential

Instead of 66

Conclusion

- It is estimated that a random sample of 160 patients contains 66 deaths. With this number of events it is possible to detect a HR of 2 at the 0.05 level of significance with 80% power.
- A HR of 2 translates in a 19% survival difference at 5 years (56% - 75%).

Example2

- The standard treatment for early stage lung cancer is surgery.
- An experimental treatment (surgery+chemo) needs to be tested in a phase III randomized study.
- *dc* data set is a cohort of patients similar to the one which will be accrued.
- The accrual should not be longer than 6 years.
- The added follow-up at the end of accrual can be as long as 2 years but no longer.
- A difference of 10% at 5 years is considered clinically important.
- The accrual rate is about 50/year.

To do list

- Find out what is the survival at 5 years
- Check the assumption of exponential distribution
- Calculate the HR corresponding to a difference of 10%
- Calculate the number of events
- Calculate the number of patients

$$Weibull: S(t) = e^{-\left(\frac{t}{b}\right)^{1/a}} \qquad Exponential: S(t) = e^{-\left(\frac{t}{b}\right)}$$

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

```
> summary(dc$survtime[dc$stat==0])
Min. 1st Qu. Median Mean 3rd Qu. Max.
7.00  44.00  64.14  69.19 84.00  168.10
> fit=survfit(Surv(survtime,stat)~1,data=dc)
> summary(fit,times=60)
time n.risk n.event survival std.err lower 95%CI upper 95%CI
  60   128    91     0.656    0.0299     0.6          0.717
> fit=survreg(Surv(survtime,stat)~1,data=dc) > summary(fit)
  Value Std. Error      z     p
(Intercept)  4.9148     0.1023 48.041 0.000
Log(scale)  -0.0308     0.0792 -0.388 0.698

> S0=0.65             > HR=1.5
> S1=0.75             > sqrtnev=(zalpha+zbeta)/(sigma*log(HR))
> HR=log(S0)/log(S1) > sqrtnev^2
> HR                   [1] 190.968
[1] 1.497427
```


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

```
> ac=5
> fup=2
> Pev0=1-(exp(-lambda0*fup)-exp(-lambda0*(ac+fup)))/(lambda0*ac)
> Pev1=1-(exp(-lambda1*fup)-exp(-lambda1*(ac+fup)))/(lambda1*ac)
> Pev=mean(c(Pev0,Pev1))
> N=191/Pev > N=191/Pev
> N
[1] 705.3517
```

Reasonable?
Can it be done?

Rate of accrual =100/year

6 years of accrual

N>600

Increasing the follow-up to 3 years N=596 which is a reachable target

Special cases

- For 3 groups of unequal size you need to calculate the σ appropriately and use the same technique. The assumption is that the effect between group 1 and 2 has the same magnitude for 2 and 3. (linearity).
- For continuous data (a marker) use the appropriate σ. You can also consider that the marker values are z-standardized and hence the $\sigma=1$. In this case HR has to reflect the fact that data is z-standardized.

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

Case cohort design.

- Does the radiation treatment (RT) given for Hodgkin's Disease (HD) predisposes to myocardial infarction(MI)?
- Many patients with HD (1000s).
- Few have MI ~ 5% (10s, max 100)
- Patients with HD and patients with MI can be easily identified through OHIP records.
- It is not known:
  - Who got RT
  - If the heart was irradiated
  - The dose given to the heart

Case cohort design

- Instead of collecting the information for all of them one could design a case cohort study:
  - Get a subset of the whole cohort
  - Add all patients with MI
- The analysis is relatively simple and yet powerful.
  - It takes advantage of all cases.
  - It uses the time to event aspect of the data.
- Disadvantages
  - One cannot estimate the probability of event since the cases are oversampled.

Case cohort design

J Clin Epidemiol Vol. 52, No. 12, pp. 1165–1172, 1999

0895-4356/99/$–see front matter
PII S0895-4356(99)00102-X

Analysis of Case-Cohort Designs

William E. Barlow,[1,2] Laura Ichikawa,[1] Dan Rosner,[1] and Shizue Izumi[3]*

[1]Center for Health Studies, Group Health Cooperative, Seattle, Washington; [2]Department of Biostatistics, University of Washington, Seattle, Washington; and [3]Department of Statistics, Radiation Effects Research Foundation, Hijiyama-koen, Hiroshima, Japan

ABSTRACT. The case-cohort design is most useful in analyzing time to failure in a large cohort in which failure is rare. Covariate information is collected from all failures and a representative sample of censored observations. Sampling is done without respect to time or disease status, and, therefore, the design is more flexible than a nested case-control design. Despite the efficiency of the methods, case-cohort designs are not often used because of perceived analytic complexity. In this article, we illustrate computation of a simple variance estimator and discuss model fitting techniques in SAS. Three different weighting methods are considered. Model fitting is demonstrated in an occupational exposure study of nickel refinery workers. The design is compared to a nested case-control design with respect to analysis and efficiency in a small simulation. In this example, case-cohort sampling from the full cohort was more efficient than using a comparable nested case-control design. J CLIN EPIDEMIOL 52;12:1165–1172, 1999. © 1999 Elsevier Science Inc. All rights reserved.

KEY WORDS. Cohort studies, nested case-control, occupational exposure, software, variance estimation, weighted sampling

Case cohort design - analysis

- Cox proportional hazards model
- The cases enter in the risk set only at the time of event
  - Sub-cohort but not cases: 0 to time of MI/last f-up
  - Cases: time of MI/last f-up-0.001(1 day) to time of MI/last f-up
- The variance needs to be calculated properly, but it is one item which can be extracted from the a Cox PH model: Jackknife variance, (approximated by dfbeta residuals)

Sample size and power

- Suppose that there are 2000 patients with HD
- Among them 100 have MI
- For simplicity we assume that none died before MI
- Dose is normally distributed with mean 0 and SD=1 (it was z-standardized).

Questions:

- How many of the 2000 we should sample. Usually a multiple of the number of MIs (100).
- What is the effect that we could detect considering the type I error 0.05 and power 80%.

To do list

- Find out which distribution of the time to MI
- Find out the accrual and follow-up time
- Find out some possible effect sizes

To do list – the distribution

- Why you need it – need h_0 : $h(t \mid x) = h_0(t)\exp(\beta x)$
- If you have the time to MI and the follow-up time for those without MI you can find out which distribution it is.
- If you do not, very likely you will assume exponential distribution
- The parameter lambda of the exponential distribution can be the baseline hazard when the covariate is z-standardized. Lambda can be obtained from the estimate of percentage MI at a time point.

Follow-up and effects

- Most of time the data set is a large administrative one and the dates of diagnosis for HD are known.
- Also the last time the data set was updated is known.
- The PI will have an idea of approximate effects s/he is looking at. Your job is to convert them into HRs for the standardized covariate. For this, you need to know the approximate mean and the SD of the dose.
- Very likely you will try several effect sizes.

Simulation code

```
fsimcc=function(i,n,HR,m,sd,h0,k,acc,fup)
{
 x=rnorm(n,mean=m,sd=sd)
 coef=log(HR)
 h=h0*exp(coef*x)
 timeev=rexp(n,h)
 timecensor=runif(n,fup,fup+acc)
 time=apply(cbind(timeev,timecensor),1,min)
 stat=(time==timeev)+0
 numev=sum(stat)

 whichsubcohort=sample(c(1:n),(k*numev))
 idsubcohort=c(1:n) %in% whichsubcohort
 idcases=(stat==1) & !idsubcohort
 time0=(time-0.001)*idcases

 fit=coxph(Surv(time,stat)~x)
 jk=resid(fit,type='dfbeta')
 subfit=coxph(Surv(time0,time,stat)~x,subset=(idcases | idsubcohort))
 subjk=resid(subfit,type='dfbeta')

 coef=fit$coef
 subcoef=subfit$coef
 vv=fit$var
 jk2=t(jk)%*%jk
 subvv=subfit$var
 subjk2=t(subjk)%*%subjk

out=data.frame(numev,coef,vv,jk2,subcoef,subvv,subjk2)
 print(i)
 return(out)
}
```

Running and results – check type I error

```
set.seed(123)
nsim=1000
date()
a=sapply(c(1:nsim),fsimcc,n=2000,HR=1,m=0,sd=1,h0=0.005,k=3,acc=10,fup=5)
date()
bcc=data.frame(apply(a,1,unlist))

> sum(bcc$pv<=0.05)/nsim
[1] 0.049
> sum(bcc$subpv<=0.05)/nsim
[1] 0.043
> exp(mean(bcc$coef))
[1] 1.002061
> exp(mean(bcc$subcoef))
[1] 1.002868
```


Running and results

```
set.seed(345)
nsim=1000
date()
a=sapply(c(1:nsim),fsimcc,n=2000,HR=1.4,m=0,sd=1,h0=0.005,k=3,acc=10,
fup=5)
date()
bccd2=data.frame(apply(a,1,unlist))

> sum(bccd2$pv<=0.05)/nsim
[1] 0.927
> sum(bccd2$subpv<=0.05)/nsim
[1] 0.843
> exp(mean(bccd2$coef))
[1] 1.39996
> exp(mean(bccd2$subcoef))
[1] 1.406848
```

N~ 400 instead of 2000

- library(cmprsk)
- plot.cuminc=getS3method('plot','cuminc')

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

Examples

- To study the time to failure (any type) in a group of cancer patients
  - Relapse
  - Deaths
- To study cause specific survival
  - Deaths of the disease
  - Other deaths
- Study late side effects of chemotherapy

- In competing risks setting we can observe only one event of multiple possible events.
- Example
  - Death of disease vs. other deaths
  - Relapse vs. death before relapse
- Number of events is important for the sample size/power calculation
- Cannot ignore the existence of CR events (cannot censor the CR events).

Analysis

- To estimate the probability of event: a modification of the K-M technique = cumulative incidence function (CIF)
- The probability of event is not a proper distribution – sub-distribution
- Modelling: a modification of the Cox proportional hazards model = Fine and Gray model (models the hazard of the sub-distribution=sub-distribution hazard)
- The effects are expressed similarly – hazard ratio, hazard=sub-distribution hazard.

Sample size calculation (CR)

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

HR= refers to the ratio of the sub-distribution hazards

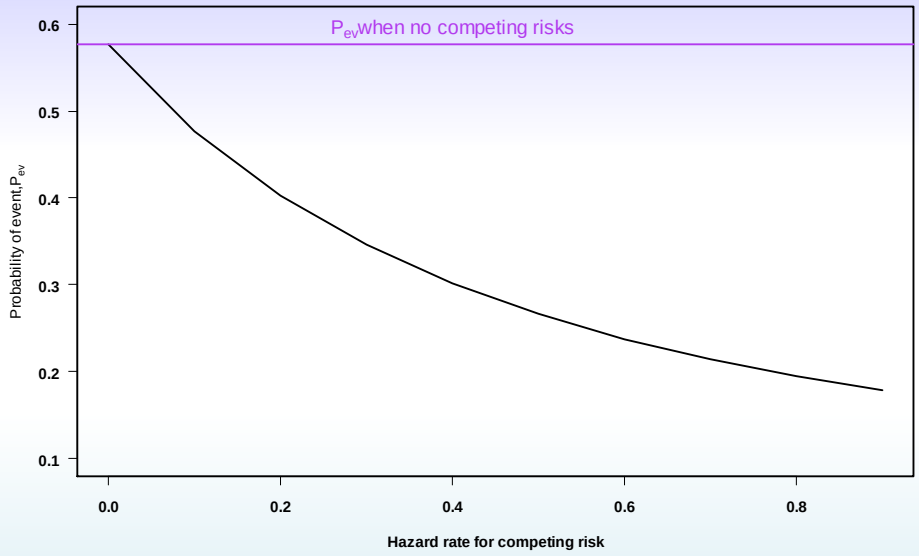
$$P_{ev} = \frac{\lambda_{ev}}{\lambda_{ev} + \lambda_{cr}} \times \left(1 - \frac{e^{-(\lambda_{ev}+\lambda_{cr})\times f} - e^{-(\lambda_{ev}+\lambda_{cr})\times(f+a)}}{(\lambda_{ev} + \lambda_{cr}) \times a}\right)$$

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

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

Assumption: Exponential distribution, independence

Decrease of P_{ev} as λ_{cr} increases



A general power calculation

- A randomized clinical trial (stomach)
- Standard RT vs. another type of RT
- Local failure (stomach)
- Parameters and assumptions:
  - Time to local failure ~ Exponential: $\lambda_S=0.4$, $\lambda_E=0.2$
  - Time to CR (death or other failures): Exp: $\lambda_{CR}=0.1$
  - The intervention has no effect on CR
  - Accrual time over 5 years. 1 year extra follow-up
  - Accrual rate = 40 patients/year
  - Equal number of patients in the 2 arms

Instead of giving the 2 hazards rates one may have the HR of the sub-distributions.

A general power calculation

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

$\alpha = 0.05, \quad z_{1-\alpha/2} = 1.96$

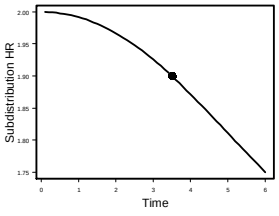
$\beta = 0.2, \quad z_{1-\beta} = 0.84$

$\sigma = 1/2$

$$HSD_S = ln\left(1 - \frac{\lambda_S}{\lambda_S + \lambda_{CR}}\left(1 - e^{-(\lambda_S+\lambda_{CR})t}\right)\right)$$

$$HSD_E = ln\left(1 - \frac{\lambda_E}{\lambda_E + \lambda_{CR}}\left(1 - e^{-(\lambda_E+\lambda_{CR})t}\right)\right)$$

$$HR = \frac{HSD_S}{HSD_E}$$



Basically, we do not have proportional hazards.

HR~1.9

Required number of events = 76

A general power calculation

Apply this for each arm

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

Probability of event for the duration of the study is 0.6 for standard and 0.4 for the experimental. Approximately the probability of event is 0.5

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

It follows that the total number of patients needed to be able to observe 76 events is 152. Since the accrual rate is 40/year then the assumed accrual time of 5 years is reasonable but long.
As a consequence, some patients will not have as much follow-up as assumed:

- Instead of 1-6 years
- Will be 1-4.8 years

Recommendation: to increase the follow-up from 1 years to 2 years.
P=0.56 => N=134 => with 152 a slightly larger power.

How to estimate the hazard rates for standard arm

- In the literature the estimates for relapse rate will very likely be based on the Kaplan-Meier methodology (censoring the competing risks).
- $\lambda = -\frac{RFR_{KM}}{t}$ where t is the time at which the estimate was calculated.

How to estimate the hazard rates for standard arm

If the estimates are based on the cumulative incidence function approach (CIF or $\hat{F}$)

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

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

Example of code for simulation when CR

```
fsimcr=function(i,yearrate,
   lambda1ev,lambda2ev,
   lambda1cr,lambda2cr,f,a)
{
 ntotal=a*yearrate
 n=ceiling(ntotal/2)
 timeev1=rexp(n,lambda1ev)
 timeev2=rexp(n,lambda2ev)
 timeev=c(timeev1,timeev2)
 timecr1=rexp(n,lambda1cr)
 timecr2=rexp(n,lambda2cr)
 timecr=c(timecr1,timecr2)

 timecensor=runif((2*n),f,(f+a))
 time=apply(cbind(timeev,timecr,timecensor),
        1,min)
 event=(time==timeev)+2*(time==timecr)
 x=c(rep(1,n),rep(0,n))
 numev=sum(event==1)
 numcr=sum(event==2)

 fitcrr=crr(time,event,x)
  fitcif=cuminc(time,event,x)
  coef=fitcrr$coef
  est=timepoints(fitcif,times=5)$est[1:2]
  pvwald=summary(fitcrr)$coef[5]
  pvgray=fitcif$Tests[3]
  out=data.frame(n,coef,numev,numcr,
    estlowrisk=est[1],esthighrisk=est[2],
     pvwald,pvgray)
  print(i)
  return(out)
}
```

Example of code for simulation when CR

```
set.seed(123)
nsim=10000
date()
a=sapply(c(1:nsim),fsimcr,yearrate=40,lambda1ev=0.4,lambda2ev=0.2,lambda1cr=0.1,lambda2cr=0.1,f=
2,a=3.8)
date()
bcr=data.frame(apply(a,1,unlist))
head(bcr)

> sum(bcr$pvwald<=0.05)/nsim
[1] 0.8478
> sum(bcr$pvgray<=0.05)/nsim
[1] 0.8468
> exp(mean(bcr$coef))
[1] 1.909454
> summary(bcr$estlowrisk)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
0.2545 0.4687 0.5162   0.5169 0.5634   0.7817   243
> summary(bcr$esthighrisk)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
0.4859 0.6835  0.7245   0.7228 0.7637    0.9325   2716
> summary(bcr$numev)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
61.00 81.00    85.00    84.95 89.00   106.00
```


How can one decide the number of simulations?

- Decide on a number of simulations which is reasonable time wise.
- Run it twice, each time using a different seed.
- If the results are quite different than you have to increase the number of simulations

Conclusion

- The sample size calculation could be time consuming but very important
- Need to take the time to do it correctly.
  - We want to minimize the number of patients exposed to an inferior treatment.
  - We do not want to discard a potential superior treatment
- The results should be translated such that the PI understands the magnitude of the effect size to be detected