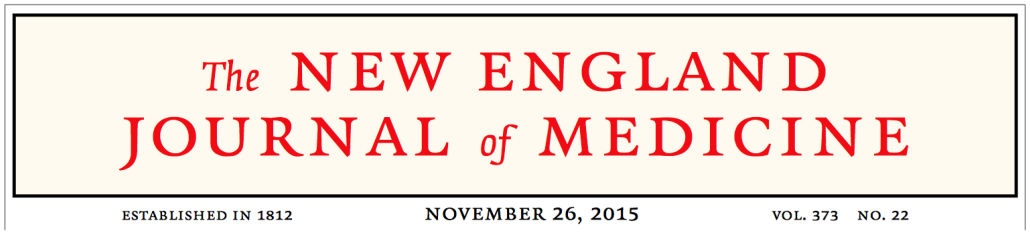
Power and sample size calculations

Dave Harrington

May 14 - 18, 2018

Background

An example



See publication by SPRINT writing committee.

Design of SPRINT

From the methods section of the paper:

> *We planned a 2-year recruitment period, with a maximum follow-up of 6 years, and anticipated a loss to follow-up of 2% per year. With an enrollment target of 9250 participants, we estimated that the trial would have 88.7% power to detect a 20% effect with respect to the primary outcome, assuming an event rate of 2.2% in the standard-treatment group.*

Type I error

Type I error (alpha error)

- Probability that trial will report a false positive, i.e., claim a significant result when there is no treatment effect
- Typically set no larger than 5%
- Depends on method of analysis, does not depend on sample size

Power

- Probability that the trial will report a true positive, i.e., claim a significant result when there is a treatment effect
- Should be 80% or greater
- Depends on sample size, method of analysis, and size of treatment effect.
- Power calculations relevant when study is designed
- Power calculations have little value after a study is complete.
    - Precision measured through confidence intervals

Overview of sample size for censored data

This unit focuses on the power of tests based on the exponential distribution and the log-rank test.

As in standard designs, the power depends on

- Type I error (significance level α)
- Difference of interest, Δ, under an alternative hypothesis H_A.

A notable difference from the usual scenario is that power depends on the **number of failures** that will be observed, not the total sample size.

Overview . . .

In practice, designing a survival study involves deciding how many patients or individuals to enter, as well as how long they should be followed.

Designs are usually either

- *fixed sample size*, with the sample size determined in advance, or
- *sequential*, which incorporate the possibility of stopping early for efficacy or futility

Collett, Chapter 12 covers sample size calculations.

Power calculations for normally distributed observations

Testing for differences between two means

Suppose data consist of:

- Group 1: $(Y_{11}, \ldots Y_{1n_1})$
- Group 0: $(Y_{01}, \ldots Y_{0n_0})$

Assume

$$Y_{1j} \sim N(\mu_1, \sigma^2), \;\; Y_{0j} \sim N(\mu_0, \sigma^2)$$

The usual objective is to test:

$$H_0 : \mu_1 = \mu_0 \Rightarrow \;\; H_0 : \triangle = 0, \text{ where } \triangle = \mu_1 - \mu_0$$

Power for a two sample normal

The standard test is based on the Z statistic:

$$Z = \frac{\overline{Y}_1 - \overline{Y}_0}{\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_0})}}$$

where s^2 is the pooled sample variance (assuming equal variances).

This test statistic has a $N(0, 1)$ distribution under H_0. If the sample sizes are equal in the two groups, $n_0 = n_1 = n/2$, then:

$$Z = \frac{\overline{Y}_1 - \overline{Y}_0}{\sqrt{s^2(\frac{1}{n/2} + \frac{1}{n/2})}} = \frac{\overline{Y}_1 - \overline{Y}_0}{2s/\sqrt{n}}$$

The steps for calculating sample size

1. Determine the critical value, c, for rejecting the null when it is true.
2. Calculate the probability of rejecting the null when the alternative is true, substituting c from above.
3. Write the expression in terms of sample size for a given power.

Step 1

Set the significance level, α, the probability of rejecting the null hypothesis when it is true, and solve for c:

$$\alpha = P\left(|\overline{Y_1} - \overline{Y_0}| > c \mid H_0\right)$$

$$= P\left(\frac{|\overline{Y_1} - \overline{Y_0}|}{2s/\sqrt{n}} > \frac{c}{2s/\sqrt{n}} \mid H_0\right)$$

$$= P\left(|Z| > \frac{c}{2s/\sqrt{n}}\right) = 2\Phi\left(\frac{c}{2s/\sqrt{n}}\right)$$

$$\text{so } z_{1-\alpha/2} = \frac{c}{2s/\sqrt{n}}$$

$$\text{or } c = \frac{z_{1-\alpha/2}(2)(s)}{\sqrt{n}}$$

Note that z_γ is the value such that $\Phi(z_\gamma) = Pr(Z < z_\gamma) = \gamma$.

Step 2: power as a function of sample size

Calculate the probability of rejecting the null when H_A is true.

Begin by writing down the probability of a Type II error:

$$
\begin{aligned}
\beta &= P\left(\text{accept } H_0 \mid H_A\right) \\
\text{so power } = 1-\beta &= P\left(\text{reject } H_0 \mid H_A\right) \\
&= P\left(|\overline{Y}_1 - \overline{Y}_0| > c \mid H_A\right) \\
&= P\left(\frac{|\overline{Y}_1 - \overline{Y}_0| - \Delta}{2s/\sqrt{n}} > \frac{c-\Delta}{2s/\sqrt{n}} \mid H_A\right) \\
&= P\left(Z > \frac{c-\Delta}{2s/\sqrt{n}}\right) \\
&= P\left(Z > z_{1-\alpha/2} - \frac{\Delta}{2s/\sqrt{n}}\right) \\
&= 1 - \Phi\left(z_{1-\alpha/2} - \frac{\Delta}{2s/\sqrt{n}}\right)
\end{aligned}
$$

Notes on power . . .

The power is an increasing function of the standardized difference:

$$\mu_T(\triangle) = \frac{\triangle}{2s/\sqrt{n}}$$

This is the number of standard errors between the two means, under the assumption of equal variances.

1. As n increases, the power increases.
2. For fixed n, the power increases with $\triangle$.
3. For fixed n and $\triangle$, the power decreases with s.
4. Assigning equal numbers of patients to the two groups ($n_1 = n_0 = n/2$) is best for maximizing power.

Step 3: sample size as a function of power

From the calculation for power:

$$
\begin{aligned}
1-\beta &= 1-\Phi\left(z_{1-\alpha/2}-\frac{\Delta}{2s/\sqrt{n}}\right) \\
&\Longrightarrow \beta = \Phi\left(z_{1-\alpha/2}-\frac{\Delta}{2s/\sqrt{n}}\right) \\
&\Longrightarrow z_{1-\beta} = z_{1-\alpha/2}-\frac{\Delta}{2s/\sqrt{n}} \\
&\Longrightarrow z_{\beta}+z_{1-\alpha/2} = \frac{\Delta}{2s/\sqrt{n}} \\
&\Longrightarrow n = \frac{(z_{1-\alpha/2}+z_{1-\beta})^2 4}{s^2}
\end{aligned}
$$

Notes on sample size

1. Sample size increases as s decreases.
2. Sample size increases as power increases.
3. Sample size increases as α decreases.

An example

Derive the total sample size required for 90% power for detecting a difference of 0.5 standard deviations between means, based on a two-sided 0.05 level test.

$$\begin{aligned} \alpha &= 0.05 \\ z_{1-\frac{\alpha}{2}} &= 1.96 \\ \beta &= 0.10 \\ z_{1-\beta} = z_{0.90} &= 1.28 \end{aligned}$$

$$n = \frac{(1.96 + 1.28)^2 \, 4s^2}{\triangle^2} \approx \frac{42 \; s^2}{\triangle^2}$$

For a 0.5 standard deviation difference, $\Delta/s = 0.5$,

$$n \approx \frac{42}{(0.5)^2} = 168$$

Calculations in R

In practice, the test statistic is a t-statistic, not a Z-statistic.

The difference is small when sample sizes are large, but more important for smaller sample sizes.

Best to use software based on the t-distribution.

```
power.t.test(delta = .5, sd = 1, sig.level = 0.05,
             power = 0.9, type = "two.sample",
             alternative = "two.sided", strict = TRUE)
```

```
##
##      Two-sample t test power calculation
##
##               n = 85.031
##           delta = 0.5
##              sd = 1
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Censored data

Sample size based on the log-rank test

Recap of the log-rank

Consider a two-group survival problem, with equal numbers of individuals in the two groups (n_0 in group 0 and n_1 in group 1).

Let $\tau_1, ..., \tau_K$ represent the K ordered, distinct failure times, and at the j-th event time:

	Fail		
Group	Yes	No	Total
0	d_{0j}	$r_{0j} - d_{0j}$	r_{0j}
1	d_{1j}	$r_{1j} - d_{1j}$	r_{1j}
Total	d_j	$r_j - d_j$	r_j

where d_{0j} and d_{1j} are the number of events in group 0 and 1, respectively, at the j-th event time, and r_{0j} and r_{1j} are the corresponding numbers at risk.

The log-rank test statistic (Z-statistic version)

$$Z_{LR} = \frac{\sum_{j=1}^{K}(d_{1j} - e_j)}{\sqrt{\sum_{j=1}^{K} v_j}} \qquad \text{with } e_j = d_j \frac{r_{1j}}{r_j}$$

$$v_j = r_{1j} r_{0j} d_j (r_j - d_j)/[r_j^2 (r_j - 1)]$$

Distribution of the log-rank statistic

Suppose that the hazard rates in the two groups are $\lambda_0(t)$ and $\lambda_1(t)$, with hazard ratio

$$\theta = e^{\beta} = \frac{\lambda_1(t)}{\lambda_0(t)}$$

and suppose $H_0 : \beta = \log(\theta) = 0$, which is equivalent to $H_0 : \theta = 1$.

It is possible to show that if there are no ties, and the observed distribution is "near" H_0,

then

- $E(d_{1j} - e_j | d_{1j}, d_{0j}, r_{1j}, r_{0j}) \approx \log(\theta)/4$
- $v_j \approx 1/4$

Distribution of the log-rank statistic ...

At a value $\log(\theta)$ under the alternative:

$$Z_{LR} \approx \frac{\sum_{j=1}^{K} \log(\theta)/4}{\sqrt{\sum_{j=1}^{K} 1/4}}$$

$$= \frac{d \log(\theta)/4}{\sqrt{d/4}}$$

$$= \frac{\sqrt{d} \log(\theta)}{2}$$

$$\text{and } Z_{LR} \sim N(\frac{\sqrt{d} \log(\theta)}{2}, 1)$$

Power of the log-rank test

Using a similar argument to before, the power of a two-sided, level α logrank test is approximately:

$$\text{Power}(\theta) \approx 1 - \Phi\left[z_{1-\frac{\alpha}{2}} - \frac{\sqrt{d}\log(\theta)}{2}\right]$$

Power depends on only d and θ.

Possible to solve for required number of events to achieve a certain power at a specified value of θ...

Power of the log-rank test...

For Power$(\theta) = 1 - \beta$, d must satisfy

$$1 - \beta = 1 - \Phi\left(z_{1-\frac{\alpha}{2}} - \frac{\sqrt{d}\log(\theta)}{2}\right)$$

$$\Rightarrow z_\beta = z_{1-\frac{\alpha}{2}} - \frac{\sqrt{d}\log(\theta)}{2}$$

$$\Rightarrow d = \frac{4\left(z_{1-\frac{\alpha}{2}} - z_\beta\right)^2}{[\log(\theta)]^2}$$

$$\text{or } d = \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{[\log(\theta)]^2}$$

Example

Suppose investigators are planning a 2-arm study, and want to detect a hazard ratio of 1.5 with 90% power at a 2-sided significance level of $\alpha = 0.05$.

Required number of events:

$$
\begin{aligned}
d &= \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{[\log(\theta)]^2} \\
&= \frac{4(1.96 + 1.282)^2}{[\log(1.5)]^2} \\
&\approx \frac{42}{0.1644} = 256
\end{aligned}
$$

Events required for various hazard ratios

Hazard Ratio	**Power** 80%	90%
1.5	191	256
2.0	66	88
2.5	38	50
3.0	26	35

Most studies are designed to detect a hazard ratio of 1.5-2.0.

#Practical considerations

Practical considerations

- Deciding on θ
- Translating the number of failures to number of enrolled participants

Easiest to think about this for the hazard ratio θ of two exponential distributions.

If $T_i \sim exp(\lambda_i)$, then

$$\text{Median}(T_i) = -\log(0.5)/\lambda_i$$

It follows that

$$\frac{\text{Median}(T_1)}{\text{Median}(T_0)} = \frac{\lambda_0}{\lambda_1} = e^{-\beta} = \frac{1}{\theta}$$

Doubling the median survival of a treatment group compared to a control group corresponds to halving the hazard.

Using R-year survival rates with an exponential distribution

Suppose the R-year survival rate is $S_1(R)$ in group 1 and $S_0(R)$ in group 0.

Under the exponential model:

$$S_i(R) = \exp(-\lambda_i R)$$

Hence,

$$\frac{\log(S_1(R))}{\log(S_0(R))} = \frac{-\lambda_1 R}{-\lambda_0 R} = \frac{\lambda_1}{\lambda_0} = e^{\beta} = \theta$$

Hence, doubling the hazard rate from group 1 to group 0 corresponds to doubling the log of the R-year survival rate.

Note that this result does not depend on R.

Example

Suppose the 5-year survival rate on treatment A is 20% and investigators want 90% power to detect an improvement to 30%.

The corresponding hazard ratio of treatment to control is:

$$\frac{ln(0.3)}{ln(0.2)} = \frac{-1.204}{-1.609} = 0.748$$

From the power formula for the log-rank, the number of events needed to detect this improvement with 90% power, based on a 2-sided 5% level test is

$$d = \frac{4(1.96 + 1.282)^2}{[\log(0.748)]^2} = 499$$

Translating to number of enrolled patients

Suppose a study enters all N patients at time 0, and will continue the study for F units of time.

Under H_0, the probability that an individual will fail during the study is:

$$P(fail) = \int_0^F \lambda_0 e^{-\lambda_0 t} dt$$
$$= 1 - e^{-\lambda_0 F}$$

Hence, if power calculations imply the study needs d failures, then

$$d = (N/2)(1 - e^{-\lambda_0 F}) + (N/2)(1 - e^{-\lambda_1 F})$$

Translating to number of enrolled...

The solution for N requires values of F and d.

1. Assume a HR θ, then calculate d.
2. Assume a follow-up period F, then calculate N.

Example

Suppose investigators wish to detect a 50% improvement in the median survival from 12 months to 18 months with 80% power at $\alpha = 0.05$, and plan to follow participants for 3 years (36 months).

Use the two medians to calculate λ_0, λ_1, and the hazard ratio, θ:

$$\text{Median}(T_i) = -\log(0.5)/\lambda_i$$

$$\text{so } \lambda_1 = \frac{-\log(0.5)}{M1} = \frac{0.6931}{18} = 0.0385$$

$$\lambda_0 = \frac{-\log(0.5)}{M0} = \frac{0.6931}{12} = 0.0578$$

$$\theta = \frac{\lambda_1}{\lambda_0} = \frac{0.0385}{0.0578} = \frac{12}{18} = 0.667$$

From the earlier table, the number of events required is $d = 191$ (same for $\theta = 1.5$ as it is for $\theta = 1/1.5 = 0.667$).

Example . . .

Now solve. . .

$$\begin{aligned} 191 &= (N/2)(1 - e^{-0.0578(36)}) + (N/2)(1 - e^{-0.0385(36)}) \\ &= (N/2)(0.875) + (N/2)(0.7500) = (N/2)(1.625) \\ \Rightarrow N &= 235 \end{aligned}$$

Round up to 236 and randomize 118 patients to each treatment group.

Another example

Even if accrual does not all happen at time 0, this formula can be surprisingly useful.

A clinical trial in esophageal cancer will randomize patients to radiotherapy alone (Rx A) versus radiotherapy plus chemotherapy (Rx B).

- The goal of the study is to compare the two treatments with respect to survival, using the log-rank test.
- From historical data, the median survival on Rx A for this disease is around 9 months.
- Want 90% power to detect an improvement in the median to 14 months.
- Past studies have been able to accrue approximately 50 patients per year.

Esophageal cancer

Start by estimating the number of events the trial will need for 90% power.

```
# using years as time scale
alpha = 0.05; beta = 0.10
med.0 = 0.75; med.1 = 14/12

num.events = 4 * (qnorm(1 - alpha/2)+ qnorm(1 - beta))^2 /
  log(med.1/med.0)^2
num.events
```

```
## [1] 215.3
```

The trial has to enroll a minimum of 216 patients (with complete follow-up).

If the anticipated accrual rate (50 patients per year) is correct, the trial could enroll 250 patients over 4 years, then follow participants until 216 events are observed.

- How long should participants be followed?

A more realistic accrual pattern

In reality, not everyone will enter the study on the same day.

Instead, the accrual will occur in a "staggered" manner over a period of time.

The standard assumption about enrollment

Participants enter the study uniformly over an accrual period lasting A units of time, and that after the accrual period, follow-up will continue for another F units of time.

The translation d to N requires the probability that a participant is observed to have an event under this accrual and follow-up scenario.

$$\begin{aligned} P(\text{fail}) &= \int_0^A P(\text{fail}|\text{enter at } a) f(a) da \\ &= 1 - \frac{\int_0^A S(a+F)\ da}{A} \end{aligned} \tag{1}$$

A more realistic accrual pattern ...

Solve for d, where P_c is the proportion of failures in the control group and P_e is the proportion of failures in the experimental group:

$$\begin{aligned} d &= (N/2)P(\text{fail}; \lambda_0) + (N/2)P(\text{fail}; \lambda_1) \\ &= (N/2)P_c + (N/2)P_e \\ &= (N/2)(P_c + P_e) \end{aligned}$$

Solve for N based on the previous formula for d:

$$N = \frac{2\ d}{(P_c + P_e)}$$

$$N = \frac{8\ \left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{[\log(\theta)]^2} \times \frac{1}{(P_c + P_e)}$$

Calculating P_c and P_e from equation (1)

If failure times are exponential distributed, (1) implies:

$$P_i = 1 - \frac{exp(-\lambda_i F)(1 - exp(-\lambda_i A))}{\lambda_i A} \qquad \text{for } i = c, e \qquad (2)$$

Freedman suggested an approximation for P_c and P_e, by computing the probability of an event at the median duration of follow-up, $(A/2 + F)$:

$$P_i = P(\text{fail}; \lambda_i) = 1 - exp[-\lambda_i (A/2 + F)]$$

He showed that this approximation works pretty well for the exponential distribution (i.e., it gives values close to (2)).

Other approximations

Rubenstein, Gail, and Santer (1981) and Lachin and Foulkes (1986) have given more accurate approximations for calculating the probablilty of an event.

- These methods are more complicated and require software
- Software now widely available in R and other packages

R packages: `Hmisc`, `TrialSize`, `gsDesign`

- Many other R packages and programs here
- `cpower()` in `Hmisc` is a simple and robust program for computing power as a function of the ususal parameters
- `nSurv()` more complicated to use, but allows many more options

Commercial packages: EaST, nQuery, SAS, etc.

Using cpower()

Parameters:

- `tref`: time point at which mortalities estimated, usually given in years
- `n`: total sample size (both groups combined). If allocation is unequal so that there are not n/2 observations in each group, sample sizes can be specified in `nc` and `ni`.
- `mc`: tref-year mortality, control group, as a decimal. This is the value of the control survivor function at `tref`.
- `r`: relative % reduction in mc by intervention. A reduction from 50% to 40% mortality at time tref is a 20% reduction.
- `accrual`: duration of accrual period, in same units as `tref`
- `tmin`: minimum follow-up time

Using cpower()...

- `noncomp.c`: % non-compliant in control group (drop-ins)
- `noncomp.i`: % non-compliant in intervention group (drop-outs, non-adherers)
- `alpha`: type I error probability. A 2-tailed test is assumed.
- `nc`: number of subjects in control group
- `ni`: number of subjects in intervention group. `nc` and `ni` are specified exclusive of n.
- `pr`: set to FALSE to suppress printing of details

Using cpower() for esophageal cancer example

The parameters:

- `tref` is 9 months, or 0.75 years, and since that is the median in the control group, `mc` $= 0.5 = S_c(0.75)$
- `n` $= 150$ patients enrolled over 3 years
- `accrual` $= 3$, and let us set `tmin` $= 1$ year of follow-up
- `alpha` $= 0.05$
- Assume no non-compliance

Using cpower() for esophageal cancer example ...

Calculating `r`, using approximations ...

The median for the intervention is 14 months, so

- $\lambda_i = \log(2)/(14/12) = 0.59$
- $S_i(0.75) = \exp(-(0.59)(0.75)) = 0.64$

Mortality at 0.75 years is reduced from 50% to 36%, a proportionate reduction of $(14/50) = 0.28$, or 28%.

Using cpower() for esophageal cancer example ...

What if we allow 1 year of follow-up?

```
library(Hmisc)
cpower(tref = 0.75, n = 150, mc = 0.5,
       r = 28, accrual = 3, tmin = 1,
       noncomp.c = 0, noncomp.i = 0,
       alpha = 0.05, pr = TRUE)
```

```
##
## Accrual duration: 3 y  Minimum follow-up: 1 y
##
## Total sample size: 150
##
## Alpha= 0.05
##
## 0.75-year Mortalities
##      Control Intervention
##         0.50         0.36
##
## Hazard Rates
##      Control Intervention
##      0.92420      0.59505
##
## Probabilities of an Event During Study
##      Control Intervention
##      0.86581      0.74288
##
## Expected Number of Events
##      Control Intervention
##         64.9         55.7
##
## Hazard ratio: 0.64386
## Standard deviation of log hazard ratio: 0.18261

##   Power
## 0.67402
```

Example: Esophageal cancer

Power of initial guess is too low.

- 67% power because only $65 + 56 = 121$ events

We can increase enrollment time or follow-up time or both.

`cpower()` can be used to search for a design.

Searching for a design

```
accrual.period = c(3)
followup.period = c(1,2,3,4)
num.accrual.periods = length(accrual.period)
num.followup.periods = length(followup.period)
perc.reduct = 28
p = matrix(0,nrow = num.accrual.periods, ncol = num.followup.periods)
for(jj in 1:num.followup.periods){
  p[1,jj]= cpower(tref = 0.75, n = 150, mc = .50, r = perc.reduct,
       accrual = accrual.period, tmin = followup.period[jj],
       noncomp.c = 0, noncomp.i = 0,
       alpha = 0.05, pr = FALSE)
  }
p
```

```
##          [,1]    [,2]    [,3]    [,4]
## [1,] 0.67402 0.72517 0.74768 0.75829
```

A more thorough search...

```
accrual.period = c(3,5,7,9)
enrollment.total = 50*accrual.period
followup.period = c(1,3,5,7)
num.accrual.periods = length(accrual.period)
num.followup.periods = length(followup.period)
perc.reduct = 28
p = matrix(0,nrow = num.accrual.periods, ncol = num.followup.periods)
for(ii in 1:num.accrual.periods){
   for(jj in 1:num.followup.periods){
     p[ii,jj]= cpower(tref = 0.75, n = enrollment.total[ii],
                      mc = .50, r = perc.reduct,
                      accrual = accrual.period[ii],
                      tmin = followup.period[jj],
                      noncomp.c = 0, noncomp.i = 0,
                      alpha = 0.05, pr = FALSE)
   }
}
power.table = matrix(0, nrow = num.accrual.periods + 1,
                     ncol = num.followup.periods + 1)
power.table[1,] = c(0, followup.period)
power.table[,1] = c(0, accrual.period)
for(ii in 1:num.accrual.periods){
  for(jj in 1:num.followup.periods){
  power.table[ii + 1, jj + 1] = p[ii,jj]
  }
}

power.table
```

```r
rownames(power.table) <- c("yrs follow-up", rep("-", 4))
colnames(power.table) <- c("yrs enrollment", rep("-", 4))
round(power.table, digits = 3)
```

```
##                yrs enrollment     -     -     -     -
## yrs follow-up                0 1.000 3.000 5.000 7.000
## -                            3 0.674 0.748 0.764 0.768
## -                            5 0.900 0.928 0.934 0.935
## -                            7 0.975 0.982 0.984 0.984
## -                            9 0.994 0.996 0.997 0.997
```

Studies often need to be larger than anticipated.

Take a closer look at the design with 5 years of enrollment and 3 years of follow-up.

```
##
## Accrual duration: 5 y  Minimum follow-up: 3 y
##
## Total sample size: 250
##
## Alpha= 0.05
##
## 0.75-year Mortalities
##      Control Intervention
##         0.50         0.36
##
## Hazard Rates
##      Control Intervention
##      0.92420      0.59505
##
## Probabilities of an Event During Study
##      Control Intervention
##      0.98661      0.94649
##
## Expected Number of Events
##      Control Intervention
##        123.3        118.3
##
## Hazard ratio: 0.64386
## Standard deviation of log hazard ratio: 0.12869

##   Power
## 0.92803
```

What is the effect of non-compliance?

ITT: analyze according to assigned treatment, not treatment received.

Main justification:

- p-values are calculated assuming no treatment difference (the null hypothesis)
- Under that assumption, assigned treatment does not affect outcome
- p-values will be correct (valid) when comparing the two groups according to treatment assignment

Example may help make this clear.

Simple trial, no difference, non-random crossover

Suppose two treatments (A and B) are equally effective.

100 participants randomized to each treatment.

ITT table:

Response	Treatment A	Treatment B
Success	40	40
Failure	60	60

Now assume, after randomization:

- 10 participants with good prognosis (future responders) switch from A to B
- 10 participants with bad prognosis (future non-responders) switch from B to A

Simple trial, no difference, non-random crossover...

Two treatments still equally effective.

Table for the as-treated groups:

Response	Treatment A	Treatment B
Success	30	50
Failure	70	50

An as-treated analysis would imply that B is more effective than A.

Simple trial, difference, random crossover

ITT can be biased when there is a real treatment effect (random crossovers).

Suppose B is more effective than A, so for 100 in each group:

Response	Treatment A	Treatment B
Success	30	50
Failure	70	50

Assume 10 randomly chosen participants from each group switch treatments after randomization but before starting treatment.

Table with only patients who do not switch

Response	Treatment A	Treatment B
Success	27	45
Failure	63	45

Attrition did not change measured success rates, but. . .

- does reduce the effective sample size

What happens when 'switchers' are put back in?

- 10 $A \to B$, 5 respond, 5 do not
- 10 $B \to A$, 3 respond, 7 do not

ITT table with assigned treatment, real response

A gets 5 responders (who received B)

B gets 3 responders (who received A)

Response	Treatment A	Treatment B
Success	32	48
Failure	68	52

Apparent success rate:

- A: 32% after crossover vs. 30% before crossover
- B: 48% after crossover vs 50% before crossover

Response proportions have moved closer together.

Non-random attrition can also cause bias in the analysis because of missing data.

Accounting for non-compliance

If some patients do not take their assigned treatments, power of the study will decrease. This issue has two sides:

Drop-outs (d_e)

- Patients who cannot tolerate the medication stop taking it.
- Their hazard rate would become the same as the placebo group (if included in study) at that point.

Drop-ins (d_c)

- Patients assigned to less effective therapy may not get symptom relief and seek other therapy, or request to cross over.

A conservative remedy: adjust P_e and P_c as follows:

$$P_e^* = P_e(1 - d_e) + P_c d_e$$

$$P_c^* = P_c(1 - d_c) + P_e d_c$$

Esophageal cancer, again

- Treatments for cancer are often toxic, and patients have difficulty adhering to regimens.
- `cpower()` can incoporate non-adherence through the parameters `noncomp.c, noncomp.i`.
- What happens to power with 20% non-adherence on the intervention arm?

```r
library(Hmisc)
cpower(tref = 0.75, n = 50*5, mc = 0.5,
       r = 28, accrual= 3, tmin = 3, noncomp.c=0, noncomp.i=20,
       alpha=0.05, pr=FALSE)
```

```
##   Power
## 0.77438
```

What about attrition

Attrition happens when participants leave a study; treatment and outcome assessments cannot be made.

Attrition is also called *loss to follow-up* and is common in long term follow-up for studies of chronic diseases.

`cpower()` cannot adjust for attrition, but `nSurv()` in the package `gsDesign` can.

A lab exercise uses `nSurv()` to reproduce the design of the SPRINT trial summarized at the beginning of this unit.

Derivations

Heuristic proof of the distribution of the log-rank

$$
\begin{aligned}
E(d_{1j}|d_{1j}, d_{0j}, r_{1j}, r_{0j}) &= P(d_{1j} = 1|d_j = 1, r_{1j}, r_{0j}) \\
&= \frac{r_{1j}\lambda_0\theta}{r_{1j}\lambda_0\theta + r_{0j}\lambda_0} \\
&= \frac{r_{1j}\theta}{r_{1j}\theta + r_{0j}} \\
&= \frac{r_{1j}}{r_{1j} + r_{0j}} + \log(\theta)\left[\frac{r_{1j}r_{0j}}{(r_{1j} + r_{0j})^2}\right]
\end{aligned}
$$

But $e_j = r_{1j}/(r_{1j} + r_{0j})$, so:

$$
E(d_{1j}|d_{1j}, d_{0j}, r_{1j}, r_{0j}) - e_j = \log(\theta)\left[\frac{r_{1j}r_{0j}}{(r_{1j} + r_{0j})^2}\right]
$$

Heuristic proof of the distribution...

If $n_0 = n_1$, then near H_0, $r_{1j} \approx r_{0j}$, hence,

$$E(d_{1j}|d_{1j}, d_{0j}, r_{1j}, r_{0j}) - e_j = \log(\theta)/4$$

Similarly, with no ties,

$$v_j = r_{1j} r_{0j} / r_j^2 \approx 1/4$$

Alternative derivation using the partial likelihood

The partial likelihood is:

$$l(\beta) = \log\left[\prod_{j=1}^{n}\left(\frac{e^{\beta \mathbf{Z_j}}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta \mathbf{Z}_\ell}}\right)^{\delta_j}\right]$$

$$= \sum_{j=1}^{n} \delta_j \left[\beta \mathbf{Z_j} - \log\left(\sum_{\ell \in \mathcal{R}(\tau_\mathbf{j})} \mathbf{e}^{\beta \mathbf{Z}_\ell}\right)\right]$$

The partial derivative of log-likelihood (the score statistic) is:

$$U(\beta) = \frac{\partial}{\partial \beta}\ell(\beta)$$

$$= \sum_{j=1}^{n} \delta_j \left[\mathbf{Z_j} - \frac{\sum_{\ell \in \mathcal{R}(\tau_\mathbf{j})} \mathbf{Z}_\ell \ \mathbf{e}^{\beta \mathbf{Z}_\ell}}{\sum_{\ell \in \mathcal{R}(\tau_\mathbf{j})} \mathbf{e}^{\beta \mathbf{Z}_\ell}}\right]$$

Partial likelihood derivation ...

The negative second partial derivative of the log-likelihood) is:

$$-\frac{\partial^2}{\partial\beta^2}\ell(\beta) = \sum_{j=1}^{n}\delta_j\left[\frac{\sum_{\ell\in\mathcal{R}(\tau_j)}e^{\beta\mathbf{z}_\ell}\sum_{\ell\in\mathcal{R}(\tau_j)}\mathbf{z}_\ell\mathbf{e}^{\beta\mathbf{z}_\ell} - (\sum_{\ell\in\mathcal{R}(\tau_\mathbf{j})}\mathbf{z}_\ell\mathbf{e}^{\beta\mathbf{z}_\ell})^2}{\sum_{\ell\in\mathcal{R}(\tau_j)}e^{\beta\mathbf{z}_\ell}}\right]$$

The logrank statistic (with no ties) is equivalent to the score statistic for testing $\beta = 0$:

$$Z_{LR} = \frac{U(0)}{\sqrt{I(0)}}$$

By a Taylor series expansion:

$$U(0) \cong U(\beta) - \beta\frac{\partial U}{\partial\beta}(0)$$

$$E[U(0)] \cong \beta d/4 \qquad \text{and} \quad I(0) \cong d/4$$