Power and Sample Size Calculations

Collett Chapter 10

Two sample problem

Given hazard ratio, what sample size is needed to achieve adequate power?

Given sample size, what hazard ratio are we powered to detect?

Why important ?

Under-power: miss scientifically meaningful differences

Over-power: waste of resources

Notation, Basic Setup

Parameter of interest: log hazard ratio

$$\theta = \log\{\lambda_1(t)/\lambda_0(t)\}$$

$H_0 : \theta = 0$ vs $H_A : \theta = \theta_R$

Test statistic Z, critical region C_α

$\Pr[\text{Type I error}] = \alpha = \Pr[Z \in C_\alpha | H_0]$

$\Pr[\text{Type II error}] = \beta = \Pr[Z \notin C_\alpha | H_A]$

(note different from β in previous slides)

$\text{Power} = 1 - \beta = \Pr[Z \in C_\alpha | H_A]$

Key result (Collett 10.1):

To have $1-\beta$ power to reject $H_0 : \theta = 0$ using a two-sided test at the $\alpha = 0.05$ level of significance

When $H_A : \theta = \theta_R$

The total number of deaths required equals

$$d = \frac{4(z_{1-\alpha/2} + z_{1-\beta})^2}{\theta_R^2} \qquad (10.1)$$

where z_q is the q quantile of the standard normal distribution

Example 10.1 Collett

Clinical trial to assess new treatment for patients with chronic active hepatitis

Under standard treatment, 41% of patients survive beyond 5 years

Expect new treatment to increase survival beyond 5 yrs to 60%

Assuming proportional hazards

$$S_1(5) = S_0(5)^{\exp(\theta_R)}$$

implying

$$\exp(\theta_R) = \frac{\log(S_1(5))}{\log(S_0(5))} = \frac{\log(0.60)}{\log(0.41)} = 0.5729$$

Thus $\theta_R = \log(.5729) = -0.557$

For $\alpha = 0.05$, 90% power (i.e., $\beta = 0.1$)

Applying equation (10.1)

$$d = \frac{4(1.96 + 1.645)^2}{0.557^2} = 135.5$$

Equation (10.1) tends to underestimate the required number of deaths, so Collett rounds up to 140

Design study so 136 (140) deaths total will be observed

Unequal Allocation

Equation (10.1) assumes equal number of individuals in each group

For unequal allocation, required total number of deaths is

$$d = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{\pi(1-\pi)\theta_R^2}$$

where π is the proportion in group 1

Derivation of (10.1)

Logrank test statistic $Z = U/\sqrt{V}$ where

$$U = \sum_j (d_{1j} - e_{1j}) \text{ and } V = \sum_j \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

Under H_0,

$$Z \sim N(0, 1) \text{ or } U \sim N(0, V)$$

Under H_A, for small θ_R,

$$Z \sim N(\theta_R \sqrt{V}, 1) \text{ or } U \sim N(\theta_R V, V)$$

For $\theta_R > \theta$

$$\begin{aligned} 1-\beta &= \Pr[Z < z_{\alpha/2}|\theta = \theta_R] + \Pr[Z > z_{1-\alpha/2}|\theta = \theta_R] \\ &\approx \Pr[Z > z_{1-\alpha/2}|\theta = \theta_R] \\ &= \Pr[Z - \theta_R\sqrt{V} > z_{1-\alpha/2} - \theta_R\sqrt{V}|\theta = \theta_R] \end{aligned}$$

Therefore

$$z_{1-\alpha/2} - \theta_R\sqrt{V} = z_\beta = -z_{1-\beta}$$

Implying

$$V = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{\theta_R^2}$$

Note when $n_j \gg d_j$

$$V = \sum_j \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)} \approx \sum_j \frac{n_{1j} n_{2j} d_j}{n_j^2}$$

Also for θ_R small and equal allocation $n_{1j} \approx n_{2j}$, implying

$$V \approx \sum_j \frac{n_{1j} n_{2j} d_j}{(n_{1j} + n_{2j})^2} \approx \sum_j \frac{n_{1j} n_{1j} d_j}{(2n_{1j})^2} = \sum_j d_j/4 = d/4$$

Therefore

$$\frac{d}{4} = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{\theta_R^2}$$

At later time points the approx $n_{1j} \approx n_{2j}$ may be poor, so that

$$\frac{n_{1j} n_{2j}}{n_j^2} < \frac{1}{4}$$

leading to d being underestimated

Right censoring

Major limitation of (10.1) is that it does not account for right censoring

If there were no right censoring, could apply (10.1) directly since $n = d$

Otherwise (Section 10.3 Collett)

$$n = \frac{d}{\Pr[\text{event}]}$$

where Pr[event] is the probability of observing an event (death) for an individual in the study

If no right censoring, Pr[event] = 1

For study with equal allocation and type I censoring at time τ

$$\Pr[\text{event}] = 1 - \frac{S_0(\tau) + S_1(\tau)}{2}$$

Suppose in chronic active hepatitis example $\tau = 5$ years

$$\Pr[\text{event}] = 1 - \frac{.41 + .6}{2} = .495$$

Thus

$$n = \frac{135.5}{.495} \approx 274$$

total number of patients needed to be enrolled to have 90% power to detect an increase in survival at 5 years from 0.41 to 0.60

SAS

Proc POWER computes sample size or power using a variety of assumptions

One method is to assume exponential distribution

For control arm

$$S_0(5) = .41 = \exp\{-5\lambda_0\}$$

implying hazard equals $\lambda_0 = -\log(.41)/5 = 0.178$

```
proc power;
      twosamplesurvival test=logrank
         hazardratio = 0.57
         refsurvexphazard = 0.178
         followuptime = 5
         totaltime = 5
         ntotal = .
         power = .9;
   run;
```

The POWER Procedure
Log-Rank Test for Two Survival Curves

Fixed Scenario Elements

Method	Lakatos normal approximation
Form of Survival Curve 1	Exponential
Form of Survival Curve 2	Exponential
Follow-up Time	5
Total Time	5
Reference Survival Exponential Hazard	0.178
Hazard Ratio	0.57
Nominal Power	0.9
Number of Sides	2
Number of Time Sub-Intervals	12
Group 1 Loss Exponential Hazard	0
Group 2 Loss Exponential Hazard	0
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed N Total

Actual Power	N Total
0.901	274

What if no censoring ?

```
proc power;
      twosamplesurvival test=logrank
         hazardratio = 0.57
         refsurvexphazard = 0.178
         followuptime = 10000
         totaltime = 10000
         ntotal = .
         power = .9;
   run;
```

The POWER Procedure
Log-Rank Test for Two Survival Curves

Fixed Scenario Elements

Method	Lakatos normal approximation
Form of Survival Curve 1	Exponential
Form of Survival Curve 2	Exponential
Follow-up Time	10000
Total Time	10000
Reference Survival Exponential Hazard	0.178
Hazard Ratio	0.57
Nominal Power	0.9
Number of Sides	2
Number of Time Sub-Intervals	12
Group 1 Loss Exponential Hazard	0
Group 2 Loss Exponential Hazard	0
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed N Total

Actual Power	N Total
0.901	140

What to do in more complicated situations ?

(1) Derive formula, (2) Explore options in SAS/R, (3) Simulation study

To implement (3)

1. Simulate a single data set under particular alternative
2. Evaluate test statistic for simulated data set. Record whether reject H_0 or not.
3. Repeat Steps 1 and 2 multiple times (e.g., 1000)
4. Compute proportion of simulated data sets where H_0 rejected. This is an estimate of power.

To make sure your simulation is working correctly, check the following:

- Simulate data sets under the null. Then the proportion of simulated data sets where H_0 rejected should not exceed the specified type I error rate α
- As you move away from H_0, the estimated power should increase towards 1

```
# simulator to approx power for two sample problem in survival

library(survival)

onesim <- function(){

        # randomly assign exponential failure times
        t0 <- rexp(n0,lambda0)
        t1 <- rexp(n1,lambda0*hr)

        # censor if beyond tau
        x0 <- pmin(t0,5); delta0 <- 1*(t0<5)
        x1 <- pmin(t1,5); delta1 <- 1*(t1<5)

        x <- c(x0,x1)
        delta <- c(delta0,delta1)
        group <- c(rep(0,n0),rep(1,n1))

        # return 1 if logrank test rejects, 0 o/w
        1*(survdiff(Surv(x,delta)~group)$chisq > qchisq(.95,1))
}

n <- 274                # total sample size
n0 <- n1 <- n/2         # balanced allocation
lambda0 <- 0.178        # hazard in control arm
tau <- 5                # time of type I censoring
nsims <- 10000          # total simulations

getpower <- function(){
        power <- matrix(NA,1, nsims)
        for (ii in 1:nsims) power[ii] <- onesim()
        power <- sum(power)/nsims
        print(paste("Empirical power",power,"for hr",hr,"based on nsims=",nsims))
}

hr <- 1; getpower()     # check type I error
hr <- 0.8; getpower()
hr <- 0.57; getpower()

# output

> source("H:/bios680/power.r")
[1] "Empirical power 0.0495 for hr 1 based on nsims= 10000"
[1] "Empirical power 0.272 for hr 0.8 based on nsims= 10000"
[1] "Empirical power 0.8967 for hr 0.57 based on nsims= 10000"
```