Applied Survival Analysis Using R
Chapter 11: Sample Size Determination for Survival Studies

Qi Guo

Department of Mathematical Sciences
The University of Texas at Dallas

April, 19 2019

Determinate the sample size

- Deciding how many subjects to include in a randomized clinical trial is a key component of its design.
- In survival analysis, there are two additional factors that one must specify regarding the *censoring mechanism* and the *particular* survival distributions.
  - Specify what parametric survival model one is using, allows for *determining the number of deaths* (or events) required to meet the power and other design specifications.
  - Administrative reasons, provide an estimate of the number of patients that need to be entered into the trial to *produce the required number of deaths*.

Hypothesis Test

- The hypothesis test is $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda = \lambda_A$, and the H_0 mean is $\mu_0 = 1/\lambda_0$ and H_A mean is $\mu_A = 1/\lambda_A$, so the treatment(hazard) ratio is $\Delta = \mu_A/\mu_0 = \lambda_0/\lambda_A$.
- The most direct way to derive a sample size formula is based on a *Wald test*, transform to let $\theta = \log(\mu) = -\log(\lambda)$, then we can express the log-likelihood function as:

$$\ell(\theta) = d \log \lambda - \lambda V = -\theta d - V e^{-\theta} \tag{1}$$

and in Chapter we have shown the m.l.e $\hat{\theta} = \log(V/d)$ and $var(\hat{\theta}) \approx 1/d$, where $d = \sum \delta_i$ and $V = \sum t_i$, the number of death and total patient time.

- So now we can use $\hat{\theta}$ as our test statistic, reject H_0 in favor of H_A if $\hat{\theta} > k$ for some constant k.

Number of deaths to achieve

- The significance level of the test, or *Type I error*, is $\alpha = Pr(\hat{\theta} > k | \theta = \theta_0)$
- Using a normalizing transformation,

$$Z = \frac{\hat{\theta} - \mu}{1/\sqrt{d}} \tag{2}$$

so

$$\alpha = Pr\left(Z > \frac{k - \theta_0}{1/\sqrt{d}}\right) \tag{3}$$

Hence

$$k = \theta_0 + \frac{z_\alpha}{\sqrt{d}} \tag{4}$$

Number of deaths to achieve

- The power of the test is given by:

$$1 - \beta = Pr(\hat{\theta} > k | \theta = \theta_A) = Pr\left(Z > \frac{k - \theta_A}{1/\sqrt{d}}\right) \tag{5}$$

so

$$-z_\beta = \sqrt{d}\left(\theta_0 + \frac{z_\alpha}{\sqrt{d}} - \theta_A\right) \tag{6}$$

Solving for d we have:

$$d = \frac{(z_\beta + z_\alpha)^2}{(\theta_A - \theta_0)^2} = \frac{(z_\beta + z_\alpha)^2}{(\log \Delta)^2} \tag{7}$$

since $\log(\Delta) = \log(\lambda_0) - \log(\lambda_A)$

- This gives us *the number of deaths* needed to achieve the specified power, *not the number of patients*.

In R

- Compute the number of deaths based on (7)

```
> expLogMeanDeaths <- function(Delta, alpha, pwr) {
 z.alpha <- qnorm(alpha, lower.tail=F)
 z.beta <- qnorm(1-pwr, lower.tail=F)
 num <- (z.alpha + z.beta)^2
 denom <- (log(Delta))^2
 dd <- num/denom
 dd   }
```

- We use the "qnorm" function to compute z_α and z_β.

```
expLikeRatio <- function(d, alpha, pwr) {
num <- qchisq(alpha, df=(2*d), lower.tail=F)
denom <- qchisq(pwr, df=(2*d), lower.tail=F)
Delta <- num/denom
Delta }
```

Estimate of the proportion of death

- We need to provide an estimate of the proportion π of patients who will die by the time of analysis. If all patients entered at the same time, we would simply have $\pi = 1 - S(t, \lambda)$, where t is the *follow-up* time.
- Patients actually enter over *an accrual period* of length a, after accrual to the trial has ended, they are followed for an *additional time* f.
- A patient who enters at time $t = 0$ will have failure probability $\pi(0) = 1 - S(a + f, \lambda)$, so a patient who enters at time a, $\pi(a) = 1 - S(f, \lambda)$

Estimate of the proportion of death

- Patients enter uniformly between times 0 and a, so that the patient entry follows a *Uniform(0,a)distribution*.
- Then the probability of death π is obtained by averaging over these times, so that a patient that enters at time t is followed for additional time $a + f - t$.

$$\pi = \int_0^a \frac{1}{a} Pr(death \mid enter\ at\ time\ t) dt \tag{8}$$

or just $\mu = a + f - t$

$$\pi = \int_0^a \frac{1}{a}(1 - S(a + f - t; \lambda))dt = 1 - \frac{1}{a}\int_a^{a+f} S(\mu; \lambda)d\mu \tag{9}$$

- Assuming an exponential distribution, $S(\mu; \lambda) = e^{-\lambda\mu}$

$$\pi = 1 - \frac{1}{a\lambda}\{e^{-\lambda f} - e^{-\lambda(a+f)}\} \tag{10}$$

In R

- Compute the probability of death:

```
prob.death <- function(lambda, accrual, followup) {
probDeath <- 1 - (1/(accrual*lambda))* (exp(-lambda*followup) -
exp(-lambda*(accrual + followup)))
probDeath}
```

- $H_1 : \lambda = 0.10$, $a = 2$ and $f = 3$ years

```
> prob.death(lambda=0.10, accrual=2, followup=3)
[1] 0.3285622
```

- Previously we found that 38 deaths were needed, then Then the number of patients needed is approximately $38/0.3285622 = 115.6$

Prerequisites

- Let p denote the proportion of patients randomized to the control group, denote by n the total number of patients in the trial, and we have $n_0 = np$ and $n_1 = n(1-p)$ control and experimental patients, respectively.
- When the trial has been completed, we will observe d_0 and d_1 deaths in the control and experimental groups, and total patient times of $V_0 = \sum t_{0i}$ and $V_1 = \sum t_{1i}$.
- The maximum likelihood estimates of the hazards $\hat{\lambda_0} = d_0/V_0$ and $\hat{\lambda_1} = d_1/V_1$.
- Compare the two distributions, use $\delta = \log \Delta = \log \lambda_0 - \log \lambda_1$

Number of death to achieve

- Based on maximum likelihood theory

$$var(\hat{\delta}) = \sigma^2 = \frac{1}{n_0\pi_0} + \frac{1}{n_1\pi_1} = \frac{1}{np(1-p)} \cdot \frac{p\pi_0 + (1-p)\pi_1}{\pi_0\pi_1} \quad (11)$$

- where π_0 and π_1 are the probabilities of death in the control and treatment groups, define a new

$$\tilde{\pi} = \left(\frac{p\pi_0 + (1-p)\pi_1}{\pi_0\pi_1}\right)^{-1} = \left(\frac{p}{\pi_1} + \frac{1-p}{\pi_0}\right)^{-1} \quad (12)$$

- $\tilde{\pi}$ is a weighted harmonic mean of π_0 and π_1

Number of death to achieve

- Approximate the weighted mean $\bar{\pi} = p\pi_0 + (1-p)\pi_1$, so

$$var(\hat{\delta}) = \sigma^2 = \frac{1}{np(1-p)} \cdot \tilde{\pi}^{-1} \tag{13}$$

- Expressing the test in terms of $\delta = \log \Delta$, we reject $H_0 : \delta = 0$ in favor of $H_A : \delta > 0$ if $\hat{\delta} > k$ for some constant k.
- By the similar argument before, we get n:

$$n = \frac{(z_\alpha + z_\beta)^2}{\delta^2 p(1-p)\tilde{\pi}} \tag{14}$$

- The required number of patients is the number of deaths

$$d = \frac{(z_\alpha + z_\beta)^2}{\delta^2 p(1-p)} \tag{15}$$

Approximate

- Given a survival function $\hat{S}(t)$ there are a number of ways of evaluating the integral in (9)
- Approximate the integral by evaluating $\hat{S}(t)$ for a patient entering at time 0, $a/2$, and a, use some results from elementary integral calculus.
- *The trapezoidal rule*:

$$\pi_s \approx 1 - \frac{1}{6}\left\{\hat{S}(a+f) + 4\hat{S}(\frac{a}{2}+f) + \hat{S}(f)\right\} \tag{16}$$

- The most accurate method is to evaluate the integral numerically

$$\pi_r = \sum_{t_{(i)}:f<t_{(i)}\leq a+f} [\hat{S}(a+f-t_{(i)}) \cdot (t_{(i)} - t_{(i-1)})] \tag{17}$$

In R

- Using the data "gastricXelox" from Chapter 3
- Extract the failure times and survival probabilities:

```
1 library(survival)
2 result.km <- survfit(Surv(timeMonths, delta) ~ 1,
3 conf.type="log-log")
4 timesXe <- result.km$time
5 survXe <- result.km$surv
```

- Set up the accrual and follow-up times, and select the portion of the failure times in the interval from f to $a + f$

```
1 accrual <- 12
2 followup <- 6
3 times.use <- c(followup, timesXe[timesXe >= followup &
4 timesXe <= accrual + followup])
5 surv.use <- summary(result.km, times=times.use)$surv
```

- Finally, we use the "diff" function to get the widths of the rectangles.

In R

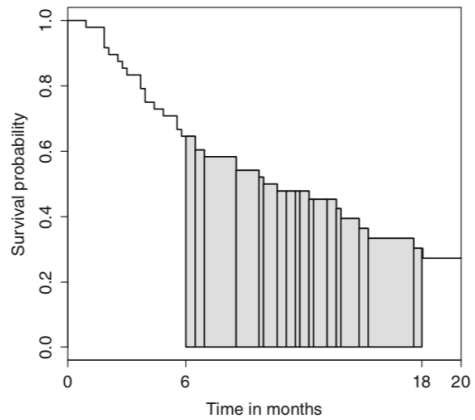

Fig. 11.2 Portion of the Kaplan-Meier plot (Fig. 3.6) showing the rectangles needed for the integral in Eq. 11.2.2

```
> surv.simpson <- summary(result.km,
times=c(followup, accrual/2 + followup, accrual+followup))
> surv.simpson$surv
[1] 0.6458333 0.4782609 0.3034080
#We evaluate the probability of death:
> pi.simpson <- 1 - (1/6)*(surv.simpson[1] + 4*surv.simpson[2] +
surv.simpson[3])
> pi.simpson
[1] 0.5229525
```

- So the estimate of the probability of death is $\pi_s = 0.523$

Simulation

1. Suppose that we are computing *power* for a two arm randomized clinical trial comparing a *standard therapy* to an *experimental therapy*
2. Select a parametric distribution that approximates the survival of patients given the standard therapy, and specify a hazard ratio to detect.
3. Model the entry of patients over a specified accrual period, randomization to *either of the two arms*, and follow them until death or until they are censored
4. Compute a test statistic and p-value using, typically, *a log-rank test*.
5. Repeat this process a large number of times, and observe the proportion of times we reject the null hypothesis of *no treatment difference*, this is the *estimated power*.

Example

Example 11.1

Determine if an *experimental agent* can *increase* the time to death from prostate cancer among patients diagnosed with advanced localized prostate caner. First specify the eligibility criteria, and to use the data in "`prostateSurvival`" to determine the survival distribution to select a Weibull distribution that matches the data. Specifically, consider men aged 66 to 74 with newly diagnosed, poorly differentiated, stage T2 prostate cancer, and find a Weibull distribution that matches the survival proportions of these patients at four and eight years.

Example

```
> attach(prostateSurvival)
> prost.66to74.poor <- prostateSurvival[{{grade == "poor"} &
{{ageGroup == "70-74"} | {ageGroup == "66-69"}} &
{stage == "T2"}},]
> library(survival)
> status.prost <- as.numeric(prost.66to74.poor$status == 1)
> result.prost.survfit <- survfit(Surv(survTime, status.
prost) ~ 1, data=prost.66to74.poor)
> summary(result.prost.survfit, time=c(48, 96))
time  n.risk  n.event  survival  std.err  lower95% CI  upper95% CI
 48     154       16     0.931    0.0171        0.898        0.965
 96      26       17     0.717    0.0565        0.615        0.837
```

- Find a Weibull distribution that matches the survival probabilities at these two times.

```
library(Hmisc)
Weib.p <- Weibull2(c(4,8),c(0.931,0.717))
```

Example

- Take a *peek* for the survival probability base on Weibull distribution

```
> Weib.p<- function(times = NULL, alpha = 0.003302123763290 6,
gamma = 2.21819823268731) {
exp(-alpha * (times^gamma))
}
```

- Define functions using the "`Quantile2`"to specify the control survival function ("`Weib.p`") and the hazard ratio(0.75), indicate that the experimental agent would reduce the hazard of prostate cancer mortality by 25 %

```
ff <- Quantile2(Weib.p,hratio=function(x) 0.75)
plot(ff, xlim=c(0,8))
```

- Specify names for the two survival distributions and extract them from "`ff`" (Fig. 11.3).

```
rcontrol <- function(n) ff(n, what='control')
rintervention <- function(n) ff(n, what='intervention')
```

Example

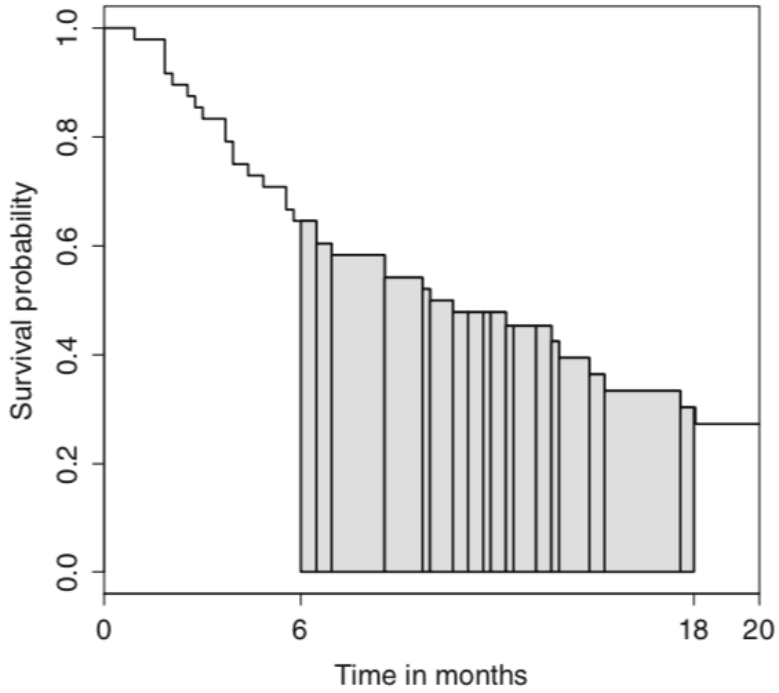

Fig. 11.2 Portion of the Kaplan-Meier plot (Fig. 3.6) showing the rectangles needed for the integral in Eq. 11.2.2

- Specify the censoring distribution "`rcens`" using the R function "`runif`" as follows (using time in years):

```
1 rcens <- function(n) runif(n, 5, 8)
```

Example

- Carry out the power simulation using "`spower`". Here we specify that there will be nc = 1500 patients enrolled in the *control arm* and ni = 1500 in the *intervention arm*, simulate nc = 1000 data sets, and carry out a log-rank test at the 0.025

```
> spower(rcontrol, rintervention, rcens, nc=1500, ni=1500,
test=logrank, nsim=1000, alpha=0.025)
[1] 0.827
```

- The result of the function, 0.827, is the power of the test, verify the significance level of the test by specifying "`rcontrol`" as both the control and intervention distribution

```
> spower(rcontrol, rcontrol, rcens, nc=1500, ni=1500, test=logrank,
[1] 0.028
```

- Empirical Type I error rate is 2.8 %

Example

- This way can *accommodate deviations* from the usual assumptions of uniform accrual, proportional hazards, and perfect patient compliance
- Suppose we expect that 10% of the patients on the intervention arm will be non-compliant, in that they to not take the experimental agent.
- Include that *noncompliance factor* in the simulation via the "`dropout`" argument in the "`Quantile2`" function.

```
> ff.dropout <- Quantile2(Weib.p,hratio=function(x) 0.75,
dropout=function(x) 0.10)
> rcontrol <- function(n) ff.dropout(n, what=?control?)
> rintervention <- function(n) ff.dropout(n, what=?intervention?)
> spower(rcontrol, rintervention, rcens, nc=350, ni=350,
test=logrank, nsim=1000, alpha=0.025)
[1] 0.734
```

- We see that the noncompliance has resulted in a loss of power, from 82.7% to 73.4%.