Power analysis for the log-rank test

Changjiang Xu

August 14, 2023

1 Methods

Consider a survival study for comparing the survivor functions in two groups using the log-rank test. Let $S_1(t)$ and $S_2(t)$ denote the survivor functions of the control and treatment groups, respectively. The key assumption of the log-rank test is that the hazard functions are proportional, under which the hazard ratio $\Delta(t) = \log(S_2(t))/\log(S_1(t)) = \Delta$ is constant and used for the measure of effect size. If $\Delta < 1$, the survival in the treatment group is higher than the survival in the control group and the new treatment is superior to the standard treatment. Under the assumption of proportional hazards, the hypothesis test,

$$\text{H0} : S_1(t) = S_2(t) \text{ versus H1} : S_1(t) < S_2(t) \text{ (one-sided) or } S_1(t) \neq S_2(t) \text{ (two-sided)}$$

is equivalent to the test,

$$\text{H0} : \Delta = 1 \text{ versus H1} : \Delta < 1 \text{ (one-sided) or } \Delta \neq 1 \text{ (two-sided).}$$

To determine the required number of events, we need to specify the significance level α, power $1 - \beta$, and the clinically significant difference between the two treatments, i.e., the effect size Δ. The effect size can also be calculated from the expected survival probabilities S_1 and S_2 in the control and treatment groups as $\Delta = \log(S2)/\log(S1)$.

Let $r = N_2/N_1$ be the ratio of samples in the two groups, where N_1 and N_2 are the number of samples in each group, and z_a denote the a-th quantile of the standard normal distribution. For a given effect size Δ, level of significance α, and power $1 - \beta$, under the assumption of proportional hazards, the total number of events required for the study can be estimated by two methods (1) (Freedman (1982)) and (2) (Schoenfeld (1981)), respectively,

$$N_e = \frac{(z_{1-\alpha/k} + z_{1-\beta})^2(1 + r\Delta)^2}{r(1 - \Delta)^2}, \tag{1}$$

$$N_e = \frac{(z_{1-\alpha/k} + z_{1-\beta})^2(1 + r)^2}{r[\log(\Delta)]^2}, \tag{2}$$

where $k = 1$ for the one-sided test and $k = 2$ for the two-sided test.

The estimates of the sample size are based on the approximation of the probability of an event. Let p_e be the probability that a subject will get an event since study. The total sample size required is given by

$$N = \frac{N_e}{p_e}. \tag{3}$$

The number of subjects required in each group is given by

$$N_1 = \frac{N}{1 + r}, \quad N_2 = \frac{Nr}{1 + r}.$$

For the study without censoring, $p_e = 1$ because all subjects get an event at the end of the study. So the total number of subjects is equal to the total number of events, i.e., $N = N_e$, for the study without censoring. For the study with censoring, the probability of an event is approximated by (Freedman (1982)),

$$p_e = \frac{(1 - S_1) + r(1 - S_2)}{1 + r}, \tag{4}$$

where S_1 and S_2 are the expected survival probabilities at the minimum follow-up time in the control and treatment groups.

Freedman (1982) and Schoenfeld (1981) derived the formulas for the number of events based on the asymptotic distribution of the log-rank test statistic. In general, the Freedman method gives higher estimates than the Schoenfeld method. Freedman's formula predicts the highest power for the log-rank test when the sample-size ratio of the two groups equals the reciprocal of the hazard ratio ($r = 1/\Delta$). Schoenfeld's formula predicts highest powers when sample sizes in the two groups are equal ($r = 1$).

2 R function *logRank.power* for estimating the sample sizes

logRank.power(S1, S2, hazard.ratio, censoring = TRUE, r = 1, alpha = 0.05, power = 0.8, alternative = c("one.sided", "two.sided"), method = c("Freedman", "Schoenfeld"), verbose = TRUE)

- S1 and S2, the expected survival probabilities at the minimum follow-up time in the control and treatment groups.
- hazard.ratio, the expected hazard ratio of the treatment to the control groups. The hazard ratio, a measure of effect size, is equal to $\log(S2)/\log(S1)$.
- censoring: a logical indicating the study with or without censoring.
- r, the ratio of subjects in the two groups, i.e., $r = N2/N1$.
- alpha, the significant level.
- power, the test power.
- alternative, if alternative = "one.sided", H1: $\Delta < 1$. If alternative = "two.sided", H1: $\Delta \neq 1$.
- method: Freedman or Schoenfeld approaches

3 Examples

We have two treatment groups; G1 and G2. The mice in G1 are expected to survive 3-6 months and the mice in G2 are expected to survive indefinitely (usually followed up to one year). We would like to estimate the sample size in each group with a power of 0.9 at a 0.05 alpha level.

```
source("R/logRank.power.R")
alpha <- 0.05
power <- 0.9
H1 <- "one.sided"
```

3.1 Given S1 and S2

Suppose the expected survivals in G1 (control group) and G2 (treatment group) at the minimum follow-up time given as:

```
S1 <- 0.4
S2 <- 0.8
```

3.1.1 Study without censoring

```
censoring <- FALSE
```

```r
method <- "Freedman"
logRank.power(S1 = S1, S2 = S2, censoring = censoring, alpha = alpha, power = power,
              alternative = H1, method = method)
```

```
## Hazard ratio (HR): 0.2435292
## Significant level:  0.05
## Power:  0.9
## Ratio of subjects in the two groups (N2/N1):  1
## Proportion of subjects in control group:  0.5
## Proportion of subjects in treatment group:  0.5
##
## Estimated sample sizes for log-rank test
## H0: HR = 1 versus H1:  HR < 1 ( one.sided alternative hypothesis)
## method:  Freedman
## Censoring:  FALSE
## Proportion of events:  1
## Total number of events:  23.1
## Total number of samples:  23.1
## The number of samples in control group:  11.6
## The number of samples in treatment group:  11.6
```

```
##            N.events           N.samples    N.sample.control N.samples.treatment
##                23.1                23.1                11.6                11.6
```

```r
method <- "Schoenfeld"
logRank.power(S1 = S1, S2 = S2, censoring = censoring, alpha = alpha, power = power,
              alternative = H1, method = method)
```

```
## Hazard ratio (HR): 0.2435292
## Significant level:  0.05
## Power:  0.9
## Ratio of subjects in the two groups (N2/N1):  1
## Proportion of subjects in control group:  0.5
## Proportion of subjects in treatment group:  0.5
##
## Estimated sample sizes for log-rank test
## H0: HR = 1 versus H1:  HR < 1 ( one.sided alternative hypothesis)
## method:  Schoenfeld
## Censoring:  FALSE
## Proportion of events:  1
## Total number of events:  17.2
## Total number of samples:  17.2
## The number of samples in control group:  8.6
## The number of samples in treatment group:  8.6
```

```
##            N.events           N.samples    N.sample.control N.samples.treatment
##                17.2                17.2                 8.6                 8.6
```

3.1.2 Study with censoring

```r
censoring <- TRUE

method <- "Freedman"
logRank.power(S1 = S1, S2 = S2, censoring = censoring, alpha = alpha, power = power,
              alternative = H1, method = method)
```

```
## Hazard ratio (HR): 0.2435292
## Significant level:  0.05
## Power:  0.9
## Ratio of subjects in the two groups (N2/N1):  1
## Proportion of subjects in control group:  0.5
## Proportion of subjects in treatment group:  0.5
##
## Estimated sample sizes for log-rank test
## H0: HR = 1 versus H1:  HR < 1 ( one.sided alternative hypothesis)
## method:  Freedman
## Censoring:  TRUE
## Proportion of events:  0.4
## Total number of events:  23.1
## Total number of samples:  57.9
## The number of samples in control group:  28.9
## The number of samples in treatment group:  28.9
```

```
##            N.events           N.samples    N.sample.control N.samples.treatment
##                23.1                57.9                28.9                28.9
```

```
method <- "Schoenfeld"
logRank.power(S1 = S1, S2 = S2, censoring = censoring, alpha = alpha, power = power,
              alternative = H1, method = method)
```

```
## Hazard ratio (HR): 0.2435292
## Significant level:  0.05
## Power:  0.9
## Ratio of subjects in the two groups (N2/N1):  1
## Proportion of subjects in control group:  0.5
## Proportion of subjects in treatment group:  0.5
##
## Estimated sample sizes for log-rank test
## H0: HR = 1 versus H1:  HR < 1 ( one.sided alternative hypothesis)
## method:  Schoenfeld
## Censoring:  TRUE
## Proportion of events:  0.4
## Total number of events:  17.2
## Total number of samples:  42.9
## The number of samples in control group:  21.5
## The number of samples in treatment group:  21.5
```

```
##            N.events           N.samples    N.sample.control N.samples.treatment
##                17.2                42.9                21.5                21.5
```

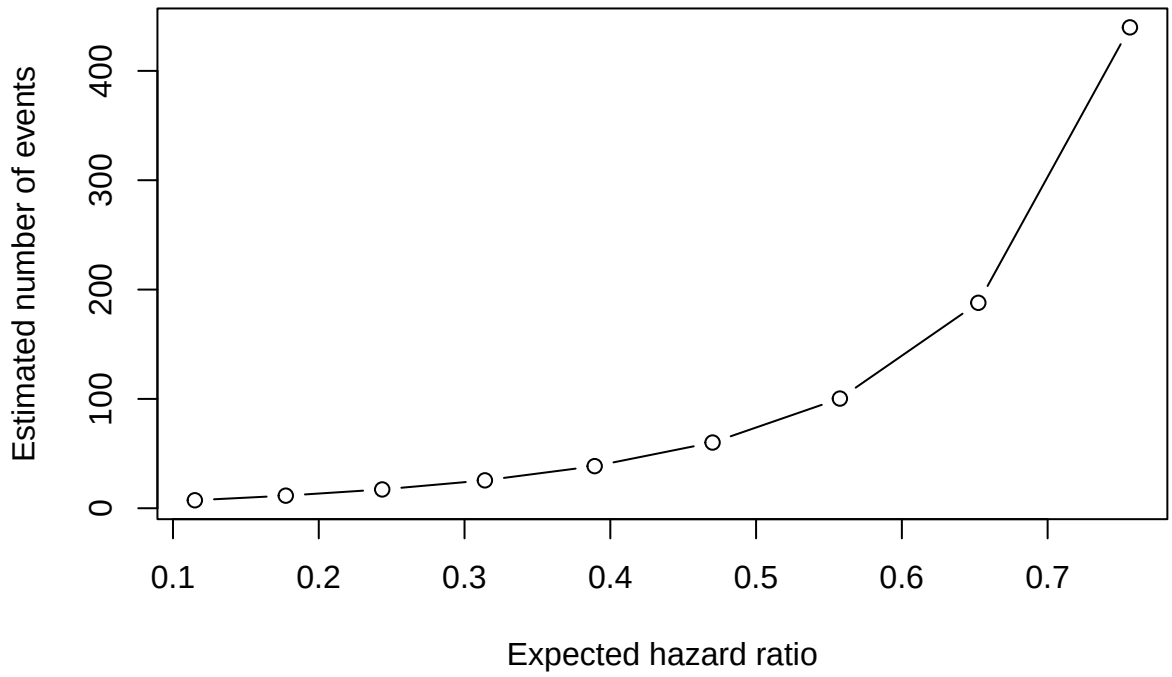
3.2 Given S1 and various S2

```
S1 <- 0.4
S2.set <- seq(0.5, 0.9, by = 0.05)

censoring <- TRUE
method <- "Schoenfeld" #"Freedman"
sizes <- NULL
for (S2 in S2.set){
    s <- logRank.power(S1 = S1, S2 = S2, censoring = censoring, alpha = alpha, power = power,
                       alternative = H1, method = method, verbose = F)
```

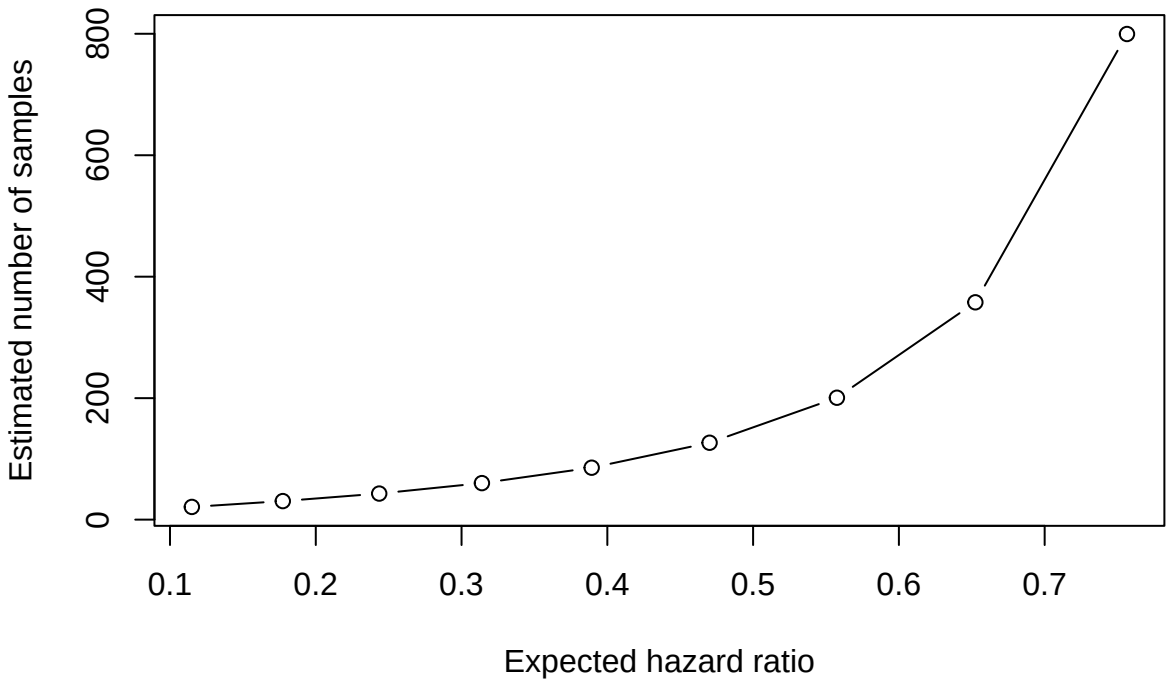
```
    sizes <- rbind(sizes, s)
}

HR <- log(S2.set)/log(S1)
sizes <- as.data.frame(sizes)
cbind(HR = round(HR, 2), S1 = S1, S2 = S2.set, sizes)
```

```
##       HR  S1   S2 N.events N.samples N.sample.control N.samples.treatment
## s   0.76 0.4 0.50    439.8     799.6            399.8                399.8
## s.1 0.65 0.4 0.55    187.9     357.8            178.9                178.9
## s.2 0.56 0.4 0.60    100.3     200.7            100.3                100.3
## s.3 0.47 0.4 0.65     60.1     126.6             63.3                 63.3
## s.4 0.39 0.4 0.70     38.5      85.5             42.8                 42.8
## s.5 0.31 0.4 0.75     25.5      60.1             30.0                 30.0
## s.6 0.24 0.4 0.80     17.2      42.9             21.5                 21.5
## s.7 0.18 0.4 0.85     11.5      30.5             15.3                 15.3
## s.8 0.11 0.4 0.90      7.3      20.9             10.5                 10.5
```

```
plot(HR, sizes$N.events, xlab = "Expected hazard ratio", ylab = "Estimated number of events", type = "b
```




```
plot(HR, sizes$N.samples, xlab = "Expected hazard ratio", ylab = "Estimated number of samples", type =
```



Reference

Freedman, L. S. 1982. "Tables of the Number of Patients Required in Clinical Trials Using the Logrank Test." *Statistics in Medicine* 1 (2): 121–29. https://doi.org/https://doi.org/10.1002/sim.4780010204.

Schoenfeld, David. 1981. "The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions." *Biometrika* 68 (1): 316–19. http://www.jstor.org/stable/2335833.