Survival Analysis

Mingwei Fei

December 12, 2022

1 Sample Size

The $ln(HR)$ follows a normal distribution, we use this to calculate the sample size.

$$\begin{aligned} ln(\hat{\Delta}) &\sim N\left(ln(\Delta), \frac{1}{d_1} + \frac{1}{d_2}\right) \\ \left(\frac{1}{d_1} + \frac{1}{d_2}\right)^{-1} &= \left[\frac{(z_{\alpha/2} + z_\beta)^2}{(ln\Delta_0)^2}\right] \end{aligned}$$

where d_i is the number of observed events.

If hazard ratio set at 2.1, then

$$\begin{aligned} \left(\frac{1}{d_1} + \frac{1}{d_2}\right)^{-1} &= \left[\frac{(1.96 + 0.842)^2}{(ln2.1)^2}\right] = 14.26 \\ \frac{1}{d_1} + \frac{1}{d_2} &= \frac{1}{14.26} = 0.07, \qquad d_1 = d_2 = 28.5 \end{aligned}$$

The one-sided significance level 0.25, power is 0.8. Note that $Z_{\alpha/2}$ is the z score for the probability $1 - \alpha/2$, and z_β is the z score for the probability $1 - \beta$. Assume the overall event and censored rate is 20%, then the sample size is $57/0.2 = 285$. The total number in the paper is 276.

1.1 Non-inferiority margin Hazard ratio $\Delta_0 = 2.1$

The assumption is that control group (C) event rate 10% and treatment group (T) event rate 20% at 6 months. Assume survival function is an exponential distribution:

$$\begin{aligned} S_t(t) &= exp(-\lambda_1 t), \qquad t = 0.5, S_t = 0.8, -\lambda_1 = ln(0.8)/0.5 \\ S_c(t) &= exp(-\lambda_2 t), \qquad t = 0.5, S_c = 0.9, -\lambda_2 = ln(0.9)/0.5 \\ \Delta_0 &= \frac{\lambda_1}{\lambda_2} = \frac{ln(0.8)}{ln(0.9)} = 2.117 \end{aligned}$$

1.2 Hazard ratio actual = 0.55

The control group survival 76.8% and treatment group survival 86.2% at 6 months. Assume survival function is an exponential distribution:

$$\begin{aligned} S_t(t) &= exp(-\lambda_1 t), \qquad t = 0.5, S_t = 0.862, -\lambda_1 = ln(0.862)/0.5 \\ S_c(t) &= exp(-\lambda_2 t), \qquad t = 0.5, S_c = 0.768, -\lambda_2 = ln(0.768)/0.5 \end{aligned}$$

$$\begin{aligned} HR = &\frac{\lambda_1}{\lambda_2} \\ &= \frac{ln(0.862)}{ln(0.768)} = 0.56 \end{aligned}$$

2 Sample Size Formula

The test hypothesis is

$$\begin{aligned} H_0 :&\lambda_1 = \lambda_2 \\ H_1 :&\lambda_1 \neq \lambda_2 \end{aligned}$$

Or equivalently, in terms of hazard ratio, $\Delta = \lambda_1/\lambda_2$

$$\begin{aligned} H_0 :&\Delta = 1 \\ H_1 :&\Delta \neq 1 \end{aligned}$$

A much simpler and quite accurate approximation for a reasonably large number of events is based on the approximate normality of th natural logarithm of the estimated hazard ratio in each treatment group:

$$ln(\hat{\lambda}_i) \sim N(ln\lambda_i, \frac{1}{d_i})$$

where d_i is the number of observed events. Thus, the $ln\Delta = ln\lambda_1 - ln\lambda_2$ also follows a normal distribution with variance $\frac{1}{d_1} + \frac{1}{d_2}$.

$$ln(\hat{\Delta}) \sim N\left(ln(\Delta), \frac{1}{d_1} + \frac{1}{d_2}\right)$$

$$\left(\frac{1}{d_1} + \frac{1}{d_2}\right)^{-1} = \left[\frac{(z_{\alpha/2} + z_\beta)^2}{(ln\Delta_0)^2}\right]$$

The calculation of sample size follows

$$\begin{aligned} Z = \frac{ln(\hat{\Delta})}{\sigma}, \qquad \sigma = \sqrt{\frac{1}{d_1} + \frac{1}{d_2}}, \qquad \delta = ln(\Delta_0) \\ P(Z \geq Z_{1-\alpha/2}|H_0) \leq \alpha/2 \\ P(Z \leq Z_\beta|H_1 = \delta) \geq \beta \end{aligned}$$

So we set Z satisfy the below equation

$$\frac{ln(\hat{\Delta})}{\sigma} = Z_{1-\alpha/2}, \qquad \text{H}_{\ 0}$$

$$\frac{ln(\hat{\Delta}) - \delta}{\sigma} = Z_{\beta}, \qquad \text{H}_1$$

So we have

$$ln(\hat{\Delta}) = Z_{1-\alpha/2}\sigma, \qquad ln(\hat{\Delta}) = Z_{\beta}\sigma + \delta, \qquad Z_{1-\alpha/2}\sigma = Z_{\beta}\sigma + \delta$$

$$\sigma = \frac{\delta}{Z_{1-\alpha/2} - Z_{\beta}}, \qquad \frac{1}{d_1} + \frac{1}{d_2} = \frac{\delta^2}{(Z_{1-\alpha/2} + Z_{1-\beta})^2}$$

3 Hazard Rate Asymptotic Distribution

3.1 Likelihood Function

If T_i and C_i are independent, which means non-informative censoring. We look at the cumulative conditional probability at time T:

$$p(T \leq s + \epsilon | T \geq s) \approx p(T < s + \epsilon | T \geq s, C \geq s)$$

Note that the above probability is not the hazard rate, it is the cumulative hazard rate. The hazard rate is as below

$$h(t) = \frac{p(t)}{S(t)} = p(s \leq T \leq s + \epsilon | T \geq s)$$

The key of success is to construct likelihood function. We use conditional probability in the situation when there are hidden variables that we can't or don't need to estimate. When there are censoring time, we don't know exactly what those censoring times are.

So in the presence of censoring, we only observe $(T_i, \delta_i), i = 1, ..n$. Let us suppose that T_i is the survival time, which may not be observed and we observe instead $U_i = min(T_i, C_i)$, where C_i is the potential censoring time.

$$\delta_i = \begin{cases} 1 & T_i \leq C_i, \qquad \text{Uncensored} \\ 0 & T_i > C_i, \qquad \text{Censored} \end{cases}$$

3.1.1 Likelihood under Censoring

The likelihood under censoring can be constructed using both the density and distribution functions or the hazard and cumulative hazard functions. Both are equivalent. The loglikelihood will be a mixture of probabilities and densities, depending on whether the observation was censored or not.

Let us suppose that T_i has distribution $f(x, \theta_0)$, where f is known but θ_0 is unknown. The likelihood construction must be with respect to the bivariate, random variable (U_i, δ_i).

We observe (U_i, δ_i) where $U_i = min(T_i, C_i)$ and δ_i is the indicator variable. In this section we treat C_i as if they were deterministic, we consider the case that they are random later.

We first observe that if $\delta_i = 1$, then the log-likelihood of the individual observation U_i is $logf(U_i, \theta)$, since

$$\begin{aligned} P(U_i = x|\delta_i = 1) = P(T_i = x|T_i \leq c_i) &= \frac{f(x;\theta)}{1 - S(x,\theta)} dx \\ &= \frac{h(x)S(x,\theta)}{1 - S(x,\theta)} dx \end{aligned}$$

where $S(x, \theta)$ is the survival function $1 - F(T_i \leq x)$.

On the other hand, if $\delta_i = 0$, the log likelihood of the individual observation $U_i = c_i|\delta_i = 0$ is simply one, since if $\delta_i = 0$, then $U_i = c_i$ (it is given). Of course it is clear that $p(\delta_i = 1) = 1 - S(c_i, \theta)$ and $P(\delta_i = 0) = S(c_i; \theta)$. Thus altogether the joint density of U_i, δ_i is

$$\begin{aligned} p(U_i, \delta_i) &= \left(\frac{f(x;\theta)}{1 - S(c_i,\theta)} (1 - S(c_i, \theta)) \right)^{\delta_i} (1 \times S(c_i, \theta))^{1-\delta_i} \\ &= f(x,\theta)^{\delta_i} [S(c_i, \theta)]^{1-\delta_i} \\ &= h(u_i)^{\delta_i} S(u_i) \\ p(\theta) &= \prod_{i=1}^{n} h(u_i)^{\delta_i} S(u_i) \end{aligned}$$

Therefore by using

$$\begin{aligned} f(U_i, \theta) &= h(U_i, \theta) S(U_i, \theta) \\ H(U_i, \theta) &= -logS(U_i, \theta) \end{aligned}$$

The joint log-likelihood of $(U_i, \delta_i)_{i=1}^{n}$ is

$$\begin{aligned} ln(\theta) &= \sum_{i=1}^{n} (\delta_i log f(\theta) + (1-\delta_i) log(1-F(\theta))) \\ &= \sum_{i=1}^{n} \delta_i \left[log h(T_i,\theta) - H(T_i,\theta) \right] - \sum_{i=1}^{n} (1-\delta_i) H(c_i,\theta) \\ &= \sum_{i=1}^{n} \delta_i log h(U_i,\theta) - \sum_{i=1}^{n} (1-\delta_i) H(U_i,\theta) \end{aligned}$$

We can get the MLE of θ by score function, Fisher Information to get the variance.

3.2 Exponential Distribution

Suppose that $T_1, T_2, ...T_n$ are i.i.d $Exp(\lambda)$ and subject to noninformative right censoring. The exponential distribution

$$f(x,\lambda) = \lambda exp(-\lambda x)$$

The survival function

$$S(x,\lambda) = 1 - F(x,\lambda) = 1 - \int_0^x \lambda exp(-\lambda x) dx = exp(-\lambda x)$$

The likelihood function

$$ln(\lambda) = \prod_{i=1}^{n} \lambda^{\delta_i} exp(-\lambda u_i) = \lambda^r exp(-\lambda W)$$

where $r = \sum_{i=1}^{n} \delta_i$ are the number of failures; $W = \sum_{i=1}^{n} u_i$ is total followup time.

The Score function and observed information

$$\begin{aligned} \frac{\partial ln(\lambda)}{\partial \lambda} &= \frac{r}{\lambda} - W \\ -\frac{\partial^2 ln(\lambda)}{\partial \lambda \, \partial \lambda} &= \frac{r}{\lambda^2} \end{aligned}$$

$\hat{\lambda}$ approximately follows $N(\lambda, \lambda^2/r)$ for large n.

By delta method,

$$log(\hat{\lambda}) \sim N(log(\lambda), r^{-1})$$

The variance of log hazard ratio r^{-1} is free of the unknown parameter λ. Similarly, we see that the log of odds ratio is used more common than odds ratio.