Sample size calculation for single-arm non-inferiority trials of time-to-event data

Yeryung Won*, Inkyung Jung**
* Department of Biostatistics and Computing, Yonsei University Graduate School
** Department of Biomedical Systems Informatics, Yonsei University College of Medicine

Introduction

Background

A non-inferiority test in phase 3 trial requires approximately 4 times larger sample size than the typical superiority test. In order to reduce cost, it might be helpful to implement the test in phase 2 to provide evidence of non-inferiority before a phase 3 trial is carried out. Additionally, if there is reliable historical control data, non-inferiority can be tested in a single-arm study to compare the new drug with the historical control.

When an endpoint is time-to-event, however, sample size calculation for a one-sample non-inferiority test has not been studied so far. Therefore, we propose non-parametric and parametric sample size formulas necessary to test non-inferiority of time-to-event data in a single-arm study.

Method

Sample size calculation for one-sample non-inferiority non-parametric test

- $S(x)$: survival function for a test drug ($\approx S_1(x)$: survival distribution under H_1)
- $\Lambda(x)$: cumulative hazard function for a test drug
- $S_0(x)$: survival function for historical control
- $\Lambda_0(x)$: cumulative hazard function for historical control
- δ : noninferiority margin ($\delta > 0$)
- X_i : observed failure time for the i^{th} subject
- Δ_i : failure indicator for the i^{th} subject
- $N_i(x) = \Delta_i I\{X_i \le x\}$
- $Y_i(x) = I\{X_i \ge x\}$
- $G(x)$: survival distribution of censoring time C

- Hypothesis :

$$H_0 : S(x) - S_0(x) \le -\delta \;\; vs \;\; S(x) - S_0(x) > -\delta$$

The test drug is noninferior to the historical control when the difference between the survival rate of the test drug and that of historical control is greater than the margin.

- Test statistic and its asymptotic distribution:

Exact mean($= \sigma_1^2 - \sigma_0^2$) and variance($= \sigma^2$) of W under H_1 were used. Because $\hat{\sigma}^2$ is consistently estimated by σ_0^2 under H_1, $W - \sqrt{n}(\sigma_1^2 - \sigma_0^2)$ converges to normal distribution with mean 0 and variance σ^2 and following equation is satisfied.

$$\frac{W}{\hat{\sigma}} - \frac{\sqrt{n}(\sigma_1^2 - \sigma_0^2)}{\sigma_0^2} \to N\left(0, \frac{\sigma^2}{\sigma_0^2}\right) \; as \; n \to \infty$$

where

$$W = n^{-\frac{1}{2}} \Sigma_{i=1}^n \int_0^\infty \{dN_i(x) - Y_i(x) d\Lambda_0^*(x)\}, \hat{\sigma} = n^{-1} \Sigma_{i=1}^n \int_0^\infty Y_i(x) d\Lambda_0^*(x),$$

$$\sigma_0^2 = \int_0^\infty G(x) S_1(x) d\Lambda_0^*(x), \sigma_1 = \int_0^\infty G(x) S_1(x) d\Lambda_1(x)$$

The null hypothesis is rejected if $\mathrm{L} < -\mathrm{z}_{1-\alpha}$. Here, * indicates that $S_0(x)$ is replaced by $S_0(x) - \delta$.

- Sample size

$$\mathrm{n} = \frac{\left(\sigma_0 \mathrm{z}_{1-\alpha} + \sigma \mathrm{z}_{1-\beta}\right)^2}{(\sigma_1^2 - \sigma_0^2)^2}$$

Sample size calculation for one-sample non-inferiority parametric test

- λ : a hazard rate of the test drug
- λ_0 : a hazard rate of historical control
- δ : noninferiority margin ($\delta > 0$)

- Hypothesis :

$$\mathrm{H}_0 : (\lambda - \lambda_0) \ge \delta \;\; \mathrm{vs} \;\; (\lambda - \lambda_0) < \delta$$

The test drug is noninferior to the historical control when the difference between the hazard rate of the test drug and that of historical control is lower than the margin.

- Test statistic and its asymptotic distribution :

$$\mathrm{Z} = \frac{\left(\hat{\lambda} - \lambda_0\right) - \delta}{\sqrt{\frac{\sigma^2\left(\hat{\lambda}\right)}{n}}} \to N(0,1) \; as \; n \to \infty$$

The null hypothesis is rejected if $Z < -z_{1-\alpha}$.

- Sample size

$$\mathrm{n} = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2}{\left((\lambda - \lambda_0) - \delta\right)^2} \sigma^2(\lambda)$$

where $\sigma^2(\lambda) = \frac{\lambda^2}{E(d|\lambda)}$ and $E(d|\lambda) = 1 + \frac{\exp(-\lambda\tau) - \exp\left(-\lambda(\tau - t_f)\right)}{\lambda t_f}$.

Here, t_a is accrual time, t_f is follow-up time and τ is total time.

Simulation study

Sample size calculation for one-sample non-inferiority non-parametric test

Data generation

- $T \sim weib\left(\rho = \frac{\mathrm{x}}{\{-\log(S_1(x))\}^{\frac{1}{\kappa}}}, \kappa\right)$ under H_1, $weib\left(\rho = \frac{\mathrm{x}}{\{-\log(S_0(x) - \delta)\}^{\frac{1}{\kappa}}}, \kappa\right)$ under H_0 : failure time
- $C \sim Unif\left(t_f, \tau\right)$: censoring time
- $A \sim Unif(0, t_a)$: accrual time
- $\mathrm{c} = \mathrm{I}(\mathrm{T} > \mathrm{C} \,|A + T > \tau)$: censoring indicator
- $X = \min\{T, C, \tau - A\}$: observed failure time
- $\mathrm{d} = \mathrm{I}(c \ne 0)$: event indicator

Result

We evaluated the power and type 1 error. Noninferiority is proved if the lower limit of the 95% CI around the difference between survival rate of test drug and that of historical control is above the margin. α, β are preset as 0.1, 0.05.

Distribution	S_0	S_1	Margin	β	α	Sample size	Power	Size
Weibull ($\rho, \kappa = 0.5$)	0.95	0.9	0.05	0.1	0.05	70	0.934	0.073
	0.95	0.9	0.1	0.1	0.05	37	0.87	0.044
Exponential ($\kappa = 1$)	0.95	0.9	0.05	0.1	0.05	62	0.901	0.081
	0.95	0.9	0.1	0.1	0.05	33	0.906	0.082
Weibull ($\rho, \kappa = 1.5$)	0.95	0.9	0.05	0.1	0.05	55	0.85	0.071
	0.95	0.9	0.1	0.1	0.05	30	0.765	0.045

The simulation result showed that the sample size decreased as the margin increased. The empirical power had similar value to the nominal level when $\kappa = 1$, but it was different from the nominal level when $\kappa = 0.5$ and it became much smaller than nominal level when $\kappa = 1.5$. Furthermore, empirical type 1 error(size) was not close to nominal level.

Sample size calculation for one-sample non-inferiority parametric test

Data generation

- $T \sim exp(\rho)$ given $\hat{\lambda} = 2$
- $C \sim Unif\left(t_f, \tau\right)$: censoring time
- $A \sim Unif(0, t_a)$: accrual time
- $\mathrm{c} = \mathrm{I}(\mathrm{T} > \mathrm{C} \,|A + T > \tau)$: censoring indicator
- $X = \min\{T, C, \tau - A\}$: observed failure time
- $\mathrm{d} = \mathrm{I}(c \ne 0)$: event indicator

Result

We evaluated the power. The simulation result showed that that the sample size decreased as the margin increased. Parametric Z test is underpowered overall in the sense that empirical power is smaller than the nominal level especially when the margin is large.

Distribution	λ_0	λ_1	Margin	β	Sample size	Power
Exponential ($\kappa = 1$)	0.45	0.5	0.05	0.2	133	0.721
	0.5	0.5	0.05	0.2	646	0.756
	0.45	0.5	0.075	0.2	86	0.681
	0.5	0.5	0.075	0.2	287	0.713
	0.45	0.5	0.05	0.1	185	0.838
	0.5	0.5	0.05	0.1	895	0.87
	0.45	0.5	0.075	0.1	118	0.835
	0.5	0.5	0.075	0.1	398	0.848

Conclusion & Discussion

- We suggested new sample size formulas for single-arm non-inferiority trials with survival endpoints.
- Whether the difference between a survival rate and margin is considered as a constant can be an issue.
- Formula 1 did not maintain type 1 error and underpowered when the distribution was Weibull and formula 2 was underpowered overall.
- Given that there has been no relevant research for sample size formula in current design, these suggestions are meaningful and expected to be useful when sample size calculation is needed in the similar research.
- We recommend formula 1 when failure time is exponentially distributed for sample size determination.
- A general accrual pattern and random loss to follow-up can be explored in the future studies.