Sample Size Formulas for Non-Inferiority Clinical Trials with Time-to-Event Data

Han Dong, Chen Zheng
Southern Medical University, Guangzhou, China
Corresponding author: Chen Zheng, email: zheng-chen@hotmail.com

ABSTACT

In many pharmaceutical studies, non-inferiority clinical trials are usually conducted because of the difficulty of finding a therapy that has more superior efficacy than a recognized effective one. When planning the non-inferiority clinical trials with a time-to-event endpoint, the calculation of sample size is one of the most fundamental steps. A proper sample size provides reasonable power to detect a clinically meaningful difference among groups. Currently exponential survival time assumption is usually made for planning a study. But, in cases that the hazard rate is not constant, the exponential assumption may be unsuitable. Thus, we introduce the sample size calculation formulas of another two distributions, Weibull and Gompertz distribution, in this paper to explore which distribution is more proper if hazard rate is not constant. The Weibull and Gompertz distribution, in which proportional hazard ratio holds, is more appropriate and flexible than exponential distribution for describing survival data because both include the shape parameter in addition to the scale parameter, which uniquely identifies the exponential distribution. Monte Carlo simulations are conducted to detect the applicability of three distribution sample formulas, in which various incidence rates, as well as increasing, constant, and decreasing hazard rates for Weibull and increasing hazard rates for Gompertz distribution, are taken into consideration. A two-tailed 95% confidence interval of the Cox proportional hazard model is used for inference. The simulation illustrates that exponential sample size formula may underestimate the sample size needed in cases where the hazard rate is increasing, and, moreover, may overestimate the sample size needed in cases where the hazard rate is decreasing. The calculated Gompertz sample size may be overlarge or too small depending on different parameter combinations. The empirical power from the Weibull distribution formula is around 0.8 in both scenarios. Therefore, we conclude that a Weibull distribution survival time assumption is suitable even if straightforward evidence supports the other two distributions of the survival time. An example is accordingly provided for illustration.

Key Words: Exponential distribution; Gompertz distribution; Survival analysis; Weibull distribution

1. Background

In many pharmaceutical studies, researchers would like to develop a new therapy that is superior to the existing one. However, it is relatively difficult to find a therapy of more superior efficacy than recognized effective one. So in many times, the researchers may want to prove that an experimental therapy is not inferior to the standard, usually in this case the experimental therapy has other advantages. Hence the planning methods of non-inferiority clinical trials (NiCTs) were studied in last decades.

A good clinical trial require a careful and efficient planning method. In the planning period of a study, estimation of sample size is one of the most fundamental steps. A proper sample size provides reasonable power to detect a clinically meaningful difference among groups. Different types of study and primary endpoints lead to different sample size estimation methods. Only the parallel study and

time-to-event data were considered in this paper.

There were some researches published recently (Rothmann et al. 2003, Jung et al. 2005, Crisp and Curtis 2008, Jung and Chow 2012, Chow et al. 2003) on the sample size determination in NiCTs with a time-to-event endpoint. Therein, the focus of Rothmann's work (Rothmann et al. 2003) was to examine issues of retaining a proportion of active-control effect. Sample size formulae for NiCTs were discussed by Chow et al. (Chow et al. 2003) and Crisp et al.(Crisp and Curtis 2008), which are the expanding of Lachin's work (Lachin and Foulkes 1986) from superior study to NiCTs, have been adopted in the softwares, such as nQuery 7.0, PASS 12.0, etc. In Lachin's paper, they took the follow-up period and enrollment period with nonuniform enrollment time into consideration in the superiority studies. Jung et al. (Jung et al. 2005, Jung and Chow 2012) proposed more accurate sample size formulae based on non-inferiority log-rank test and a generalized log-rank test, respectively.

Because of the importance of sample size estimation, not only the methods of estimation but also the assumed distributions should be chosen with cautiousness. Exponential survival time is assumed in all above papers. Although an exponential distribution may provide a reasonable approximation to the distribution of survival times over relatively short intervals, it typically does not adequately characterize the distribution of survival times because of its property of constant hazard over time (Heo et al. 1998).

However, the sample size determination for NiCTs with more flexibly distributed time-to-event endpoint has not been discussed in the literatures. Thus, we would like to verify the sample size formulas performance of another two distributions, in which proportional hazard assumption holds. The sample size formulas of three distributions are shown in Section 2. Simulation setting and results are described in Section 3 and 4. Section 5 give a real clinical trial as an example. We take a conclusion at end of this paper.

2. Sample size formulae

In the NiCTs, alternative hypothesis is that the efficacy of experimental group ($x = 1$) is identical to or just a little worse than that of control group ($x = 0$). Suppose our primary endpoint is time to a negative event, such as death, progression of cancer, etc. i.e. the higher the hazard rate, the worse the efficacy. The hazard ratio is defined as $\Delta = \lambda_1(t) / \lambda_0(t)$, where $\lambda_1(t)$ and $\lambda_0(t)$ is the hazard rate of experimental and control group, respectively. If the Δ is constant over time, the proportional hazard assumption holds. Given the non-inferiority margin of hazard ratio Δ_0, we intend to test $H_0 : \Delta \geq \Delta_0$ against $H_1 : \Delta < \Delta_0$, where Δ_1 ($\Delta_1 < \Delta_0$) is the true hazard ratio under H_1. N_0 and N_1 indicate the sample size of control and experimental group, respectively. The total sample size $N = N_0 + N_1$. Only the balanced design is considered in this paper, i.e. $N_0 = N_1$. Duration of enrolment and follow-up is denoted by R and T_f respectively, and the enrolment time is uniform distributed with $[0, R]$.

Currently, the most widely used method for sample size calculation in NiCTs when the survival time is exponential distributed (Crisp and Curtis 2008, Chow et al. 2003) is

$$N_0 = N_1 = \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{\ln(\Delta_0) - \ln(\Delta_1)} \right)^2 \left[E_0^{-1} + E_1^{-1} \right], \tag{1}$$

where

$$E_x = \frac{\lambda_x}{\lambda_x + \phi} \left[1 - \frac{e^{-T_f(\lambda_x + \phi)} - e^{-(T_f + R)(\lambda_x + \phi)}}{R(\lambda_x + \phi)} \right], \tag{2}$$

ϕ denotes the hazard rate of identical exponential censoring distribution in both

groups. Z_α denotes $(1-\alpha)$ quartile of the standard the normal distribution. It is often assumed that the efficacy of experimental group is the same as that of control group, i.e. $\Delta_1 = 1$ and $\ln(\Delta_1) = 0$.

When the distribution of survival time is Weibull or Gompertz, the E_x in Equation (1) is replaced by

$$E_x^W = \int_{T_f}^{T_f+R} \frac{1}{R} \int_0^z k\lambda_x u^{k-1} e^{-\lambda_x u^k} e^{-\phi_x u} du dz \text{ , or} \tag{3}$$

$$E_x^G = \int_{T_f}^{T_f+R} \frac{1}{R} \int_0^z \theta_x \exp(\alpha^G u) \exp\left\{\theta_x / \alpha^G [1 - \exp(\alpha^G u)]\right\} \exp(-\phi_x u) du dz , \tag{4}$$

respectively, where k and λ_x is the shape and scale parameter of Weibull distribution. α^G and θ_x is the shape parameter and scale parameter of Gompertz distribution. The numerical algorithm is used in the calculation of E_x^W and E_x^G. We assumed that k and α^G is identical in two groups, so that the proportional hazard assumption holds.

3. Simulation

Monte Carlo simulation studies were undertaken to verify the performance of sample size formulae under the exponential, Weibull and Gompertz distribution when the survival time is exponential, Weibull or Gompertz distribution. Pilot studies are simulated for estimating the parameter before calculating the sample size. Hence, 200 simulated trials of 500 subjects with Weibull or Gompertz failure time and exponential censoring time are simulated, respectively, for estimating the parameter for determining the sample size of each distribution.

10000 independent trials were simulated for various combinations of several parameter. The increasing ($k = 2$), constant ($k = 1$) and decreasing ($k = 0.5$) hazard rates and several incident intensities of control group were considered ($\lambda_0 = 0.5, 0.7, 0.8, 1.0, 2.0$) with ($\phi = 0.2, 0.5$) and without ($\phi = 0$) censoring rate, respectively. The shape ($\alpha^G = 0.05, 0.1, 0.5, 1.0$) and scale ($\theta_0 = 0.05, 0.1, 0.5, 1.0$) is set for Gompertz simulation. $\Delta_0 = 1.2$, or 1.5 is taken into consideration while all the Δ_1's were set to be 1. For each combination, if the upper limit of 95% two-tailed confidence interval of hazard ratio from Cox proportional hazard regression model is less than the non-inferiority margin, the non-inferiority is concluded.

4. Results

The simulated powers of Weibull survival time are shown in Figure 1, in which the horizontal axis of each subfigure is the parameter combination index, which is the sequential number ordered by λ_0. The index of the points of identical shape on censoring rate represents a unique λ_0 with various ϕ.

Because the Weibull distribution is its own true distribution in Figure 1, the empirical power is around the predetermined power very closely. In cases that the k and λ_0 is relatively small (Figure 1-A), if censoring rate is higher (say 30%), the sample size calculated by exponential distribution is larger than that by Weibull distribution by 20 percent. The largest empirical power of exponential distribution is above 0.95, outclassing the presetting 0.8. So, in this scenario exponential formula may waste a lot of subjects. As a result of exponential sample size approaching to Weibull sample size with the increasing of λ_0 and decreasing of censoring rate, the powers are getting closer to the predetermined power.

In cases that $k=1$, the formulae under exponential distribution and Weibull distribution are completely the same and the power is around 0.8 in each combination (Figure 1-B). When k is larger than 1 (Figure 1-C), if censoring rate is higher (say

30%), exponential formula underestimates the sample size needed. Similar to the situation that $k<1$, with decreasing of the censoring rate, the powers are getting closer to the predetermined power.
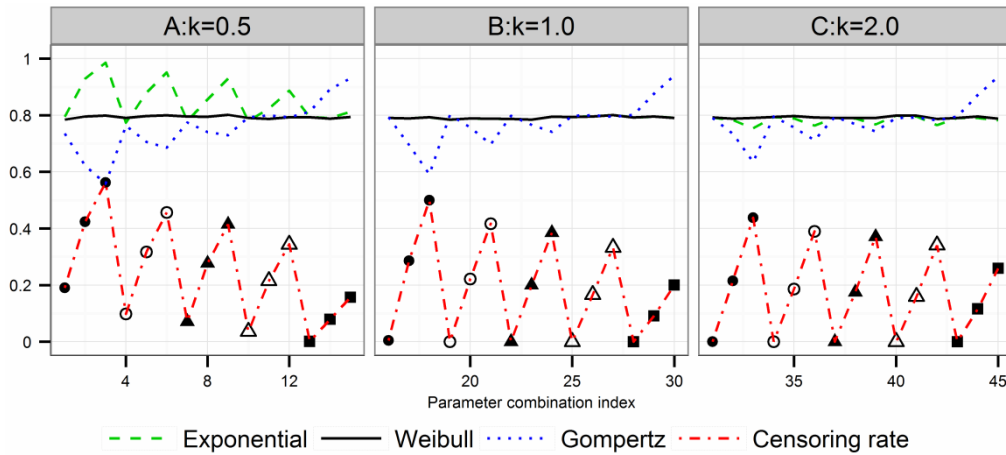

Figure 1. Part of the simulation results for the Weibull survival time ($\Delta_0 = 1.5$)

● $\lambda_0 = 0.5$ ○ $\lambda_0 = 0.7$ ▲ $\lambda_0 = 0.8$ △ $\lambda_0 = 1.0$ ■ $\lambda_0 = 2.0$

The trend of empirical powers of Gompertz distribution sample size formula in these scenarios is increasing from around 0.7 to over 0.9 with the increasing of λ_0. In case that λ_0 is fixed, the distance of Gompertz power to Weibull increases as the censoring rate becomes larger.

Figure 2 shows the Gompertz survival time, of which the index is ordered by ϕ and θ_0. The index of the same shape points on censoring rate represents unique ϕ and θ_0 with various α^G. Weibull sample size is very close to that of Gompertz, which is around 0.8. When the censoring rate is higher, the empirical power of exponential sample size is below the predetermined power 0.8. The smallest is around 0.2 in the case of the largest censoring rate.
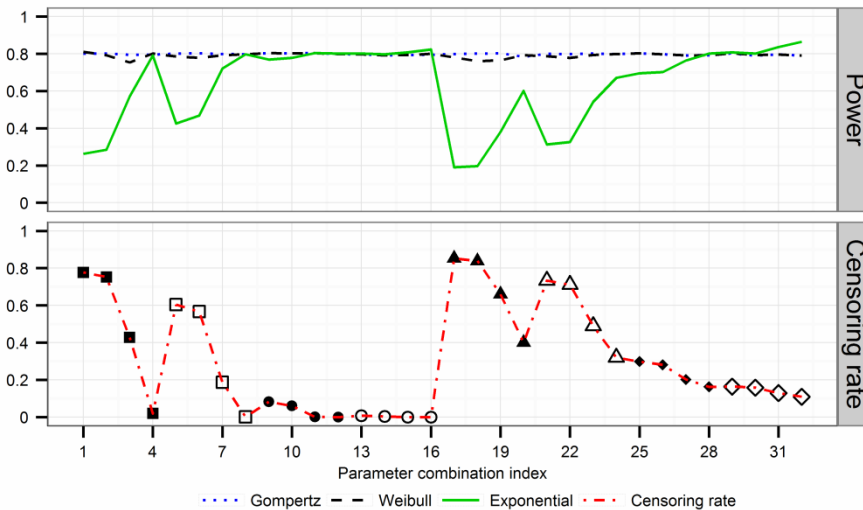

Figure 2. The simulation results for the Gompertz survival time ($\Delta_0 = 1.5$)

■ $\theta_0 = 0.05$ □ $\theta_0 = 0.10$ ● $\theta_0 = 0.50$ ○ $\theta_0 = 1.00$

5. Example

The sample size comparison is illustrated by a randomized, open-label, phase III, parallel clinical study (Kang et al. 2009). This clinical trial was conducted to compare the efficacy on the patients with gastric cancer of Capectitabine/cisplatin (XP, experimental arm, $x = 1$) versus 5-fluorouracil/cisplatin (FP, standard/control arm, $x = 0$) as first-line therapy. The primary outcome of the trial was that progression-free survival (PFS) of XP is non-inferior to FP, as measured by the hazard ratio $\Delta_{XP/HP}$ with a non-inferiority margin of 1.40.

As we can get from the paper (Kang et al. 2009), the median PFS was 5.6 months and 5.0 months for XP and FP, respectively. The parameter combinations with random censoring and without random censoring is also considered. The quantities for estimation the sample size is listed in Table 1.

Table 1. The sample size computation for different parameters combinations.

k	λ_0	Δ_0	ϕ	N per group	Events per group
0.5	0.310	1.40	0	178	139
			0.05	222	139
1.0	0.139	1.40	0	144	139
			0.05	192	139
1.5	0.062	1.40	0	140	139
			0.05	180	139

Notes: $\Delta_1 = 1$, $T_f = 24$, $R = 1$, $\alpha = 0.05$, power=0.8

Under exponential assumption, 144 and 191 patients are required for $\Delta_{XP/FP} = 1.40$ without and with random censoring, respectively. In case of decreasing hazard rates, the sample size needed is larger than that of the constant hazard rate for the sake of the small incidence rate. If the hazard rate is increasing, the event incidence rate is a little higher so that subjects required decrease to 139 and 182.

6. Conclusion

Usually at the planning stage of a study, the exponential assumption will be made for sample size determination and trial procedure planning. As we found from this paper that exponential and Gompertz sample size formula may waste the samples or underestimate the sample size needed. The Weibull sample size formula works well in different distributions, in which the proportional hazard assumption holds. Therefore, we suggest that it is better to make a Weibull distribution survival time assumption even if straightforward evidence supports the other two distributions of the survival time.

Reference

[1] Rothmann M, Li N, Chen G, et al.(2003) "Design and analysis of non-inferiority mortality trials in oncology". *Statistics In Medicine*,22,239-264.
[2] Jung S, Kang S J, Mccall L M, et al.(2005) "Sample size computation for two-sample noninferiority log-rank test". *Journal of Biopharmaceutical Statistics*,15,969-979.

[3] Crisp A, Curtis P.(2008) "Sample size estimation for non-inferiority trials of time-to-event data". *Pharmaceutical statistics*,7,236-244.

[4] Jung S H, Chow S C.(2012) "On Sample Size Calculation for Comparing Survival Curves Under General Hypothesis Testing". *Journal of Biopharmaceutical Statistics*,22,485-495.

[5] Chow S, Shao J, Wang H. (2003) *Sample Size Calculation in Clinical Research*. CRC PressI Llc,

[6] Lachin J M, Foulkes M A.(1986) "Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification". *Biometrics*,507-519.

[7] Heo M, Faith M S, Allison D B.(1998) "Power and sample size for survival analysis under the Weibull distribution when the whole lifespan is of interest". *Mechanisms Of Ageing And Development*,102,45-53.

[8] Kang Y, Kang W, Shin D, et al.(2009) "Capecitabine/cisplatin versus 5-fluorouracil/cisplatin as first-line therapy in patients with advanced gastric cancer: a randomised phase III noninferiority trial". *Annals Of Oncology*,20,666-673.