PREDICTING THE COURSE OF CLINICAL TRIALS WITH TIME-TO-EVENT ENDPOINT

Yang Ding
Kenvue Inc.
tomding.biostat@gmail.com

November 14, 2024

ABSTRACT

In randomized controlled clinical trials with time-to-event endpoints, statistical power is typically driven by the number of observed events at a specified observation time. The number of subjects enrolled is prospectively planned to reach the target number of observed events, ensuring that the sample size is neither too large, which would incur excessive costs, nor too small, which would delay the observation of the corresponding number of events. In this paper, we examined the quantitative relationships among enrollment time, individual maximum follow-up time, observation time, expected number of events, number of subjects enrolled, parametric survival distribution, and random censoring distribution. We constructed a probability model for observing the events of interest under various scenarios. Based on this model, each of the four primary study design parameters (the expected number of observed events, the number of subjects enrolled, the observation time, and the censoring parameter) can be derived analytically given the other three parameters. Furthermore, we use simulation studies to evaluate the performance of the proposed analytical approach. Additionally, an R package, *ClinTrialPredict*, was developed to facilitate the calculation and geneate simulation dataset based on the calculated design settings to conduct simulation studies.

***K*eywords** clinical trial prediction · simulation study

1 Introduction

In randomized controlled clinical trials with time-to-event endpoint, statistical power is typically driven by the number of observed events of interest at a specified observation time, accounting for the fact that a subjects time-to-event endpoint may not be fully observed. For instance, in oncology studies, primary endpoint like overall survival or progression-free survival are frequently used, and the number of deaths or disease progressions is calculated prospectively to achieve a predefined level of power. Throughout the study, a subjects event time may be right-censored, due to early dropout or administrative censoring, prior to the occurrence of the event of interest. Increasing the number of subjects enrolled can reduce the time needed to reach the target event count. However, increasing enrollment also adds to the studys financial burden. Therefore, careful planning of sample size is essential for the success of such trials.

Rubinstein, Gail and Santner [1] investiged the required trial length to assure a desired power, assuming the enrollment time of subjects follows the Poisson process and both survival time and lost-to-follow up time follows the exponential distribution. Bagiella and Heitjan [2] generalized the model proposed by Rubinstein, Gail and Santner [1], and constructed predictive intervals for the trial length needed, using the Bayesian simulation method. Ying [3] extends

the exponential-distribution based model proposed by Bagiella and Heitjan [2] to a more flexible Weilbull-distribution based model. Wan [4,5] proposed a general framework to model the censoring proportions, under the parametric assumptions for both survival distribution and random censoring distribution. The difficulties of predicting the expected number of events are: 1) it is a combination of various process, including enrollment at a given period, event occuring, random censoring, administrative censoring; 2) the probability model varies at different observation timings. Few studies have discussed the scenarios comprehensively.

In this paper, we examined the quantitative relationships among enrollment time, individual maximum follow-up time, observation time, expected number of events, number of subjects enrolled, parametric survival distribution, and random censoring distribution. We constructed a probability model for observing the events of interest under various scenarios. Based on this model, each of the four primary study design parameters (the expected number of observed events, the number of subjects enrolled, the observation time, and the censoring parameter) can be derived analytically given the other three parameters. Furthermore, we use simulation studies to evaluate the performance of the proposed analytical approach. Additionally, an R package, *ClinTrialPredict*, was developed to facilitate the calculation and geneate simulation dataset based on the calculated design settings to conduct simulation studies.

This paper is structured as follows: in section 2, we discussed the general settings and probability model under the various scenarios; in section 3, we discussed the application of the proposed analytical method for calculating design parameters; in section 4, we conductd the simulation studies to eveluate the performance of the proposed analytical method; in section 5, we gave the conclusive discussions.

2 Assumptions and models

2.1 General setting

In this research, we will consider a two-arm parallel group randomized clinical trial with a time-to-event endpoint (e.g., overall survival or progression-free survival), the number of events of interest required for interim or final analysis is determined before the study begins to achieve the pre-specified statistical power. Based on the target number of events to be observed, the sponsor will determine the number of subjects to enroll. Increasing the number of subjects enrolled can expedite the time of achieving the target number of events; however, enrolling an excessive number of subjects will incur substantial costs. Typical questions that sponser are interested when planning the sample size are:

- Question 1: What is the expected number of events to be observed at a given observation time (usually at an interim or the final analysis)?
- Question 2: How many subjects need to enroll to observe a target number of events at a given observation time?
- Question 3: How much observation time needed when the target number of events can be oberved?
- Question 4: What is the censoring distribution to achieve a target number of events at a given observation time?

Such clinical trials with survival endpoint typically include an enrollment period followed by a follow-up period. Each subject in the trial is assigned a maximum follow-up time. The outcomes for subjects can either be the occurrence of the event of interests, as demonstrated by Subject 1 and 2 in Figure 1 or being right censored, at a given observation time (usually at an interim or final analysis). There are three censoring reasons: 1) Subjects are censored due to a random censoring event, such as being lost to follow-up, prior to the occurrence of the event of interest, as demonstrated by Subject 3 in Figure 1; 2) subjects are censored because they reach their maximum follow-up time before the event of interest occurs, as demonstrated by Subject 4 in Figure 1; 3) subjects are censored at the time of interim or final analysis before the event of interest is observed, as demonstrated by Subject 5 in Figure 1.

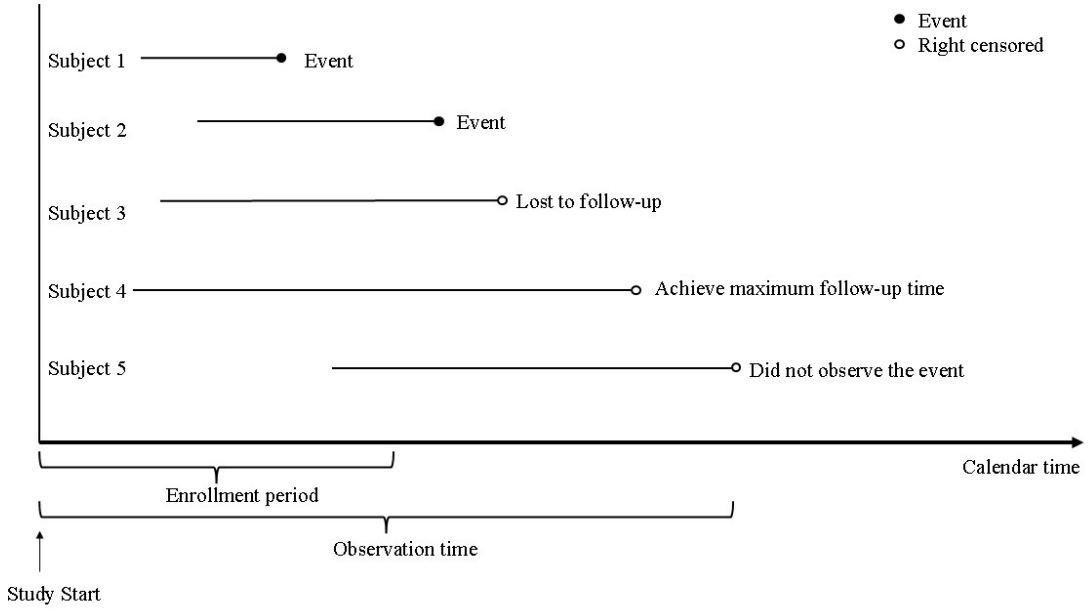

Figure 1: Schematic diagram of right censored subjects in a clinical trial with time-to-event endpoint

2.2 Assumptions and models

To quantitatively answer the questions listed in section 2.1, we will constrcuct the probability model with regard to the probability of observing the event of interets for the population. Firstly, let random variable A denote the enrollment time for a subject, and in this paper we will assume enrollment time follows a uniform distribution from 0 to s with probability density function:

$$f_A(a) = \frac{1}{s}, \tag{1}$$

where s is the enrollment period. Let N_0 and N_1 denote the number of subjects to be enrolled in control arm and experimental arm, respevtively.

Secondly, let random variable T denote the survival time for a subject. In this paper, we will assume the proportional hazard between the two groups. Under the proportional hazard assumption, the hazard function, probability density function and survival function of survival time are:

$$\begin{aligned} h(t|X,\theta_0) &= h_0(t|\theta_0)\exp(\beta X), \\ f(t|X,\theta_0) &= h_0(t|X,\theta_0)\exp\left(-H_0(t|\theta_0)\exp(\beta X)\right), \\ S(t|X,\theta_0) &= \exp\left(-H_0(t|\theta_0)\exp(\beta X)\right), \end{aligned} \tag{2}$$

where $X = 0, 1$ denote control arm and experimental arm, respevtively; $h_0(t|\theta_0)$ denote the baseline hazard function and θ_0 denote the parameter vector in baseline hazard function; $H_0(t) = \int_0^t h_0(u)du$ is the cumulative baseline hazard function at time t; $\exp(\beta)$ denote the hazard ratio between experimental arm and control arm. In the next paragraph of this paper, we will use $f_T(.)$ to denote the probability density function of survival time.

Thirdly, let random variable C denote the random censoring time for a subject, and $f_C(c|\theta_c)$ is the probability density function of random censoring time, where θ_c denote the parameter vector of the density function. In this paper, we will assume the random censoring time and survival time are independently distributed, and the distribution of random censoring time is not conditional on treatment arm X.

Finally, let l denote the observation time since the enrollment start, and m denote the maximum follow-up time for a subject. Let δ denote the censoring indicator, where $\delta = 0$ indicates the event observed, and $\delta = 1$ indicates right censoring. Based on the settings in section 2.1, the event is observed if and only if $T < min(C, l - A, m)$. The essential point to answer the questions in section 2.1 is to derive $P(\delta = 0|X, \theta_0, \theta_c, \beta)$, the probability of observing an event given $(X, \theta_0, \theta_c, \beta)$. We will consider $P(\delta = 0|X, \theta_0, \theta_c, \beta)$ in below 5 scenarios:

Scenario 1: Observation time is within the enrollment period and less than the maximum follow-up time for a single subject ($l < s, l < m$).

$$\begin{aligned} P(\delta = 0|X, \theta_0, \theta_c, \beta) &= P(T \leq C, T + A \leq l, A \leq l) \\ &= \int_0^l \frac{1}{s} \int_0^{l-a} f_T(t|X, \theta_0, \beta) \int_t^\infty f_C(c|\theta_c) dcdtda \end{aligned} \tag{3}$$

Scenario 2: Observation time is within the enrollment period and greater or equal than the maximum follow-up time for a subject ($l < s, l \geq m$).

$$\begin{aligned} P(\delta = 0|X, \theta_0, \theta_c, \beta) =& P(T \leq C, T \leq m, A \leq l - m) + P(T \leq C, T + A \leq l, l - m \leq A \leq l) \\ =& \int_0^{l-m} \frac{1}{s} \int_0^m f_T(t|X, \theta_0, \beta) \int_t^\infty f_C(c|\theta_c) dcdtda + \int_{l-m}^l \frac{1}{s} \int_0^{l-a} f_T(t|X, \theta_0, \beta) \int_t^\infty f_C(c|\theta_c) dcdtda \end{aligned} \tag{4}$$

Scenario 3: Observation time is after the enrollment period and less than the maximum follow-up time for a subject ($l \geq s, l < m$).

$$\begin{aligned} P(\delta = 0|X, \theta_0, \theta_c, \beta) =& P(T \leq C, T + A \leq l, A \leq s) \\ =& \int_0^s \frac{1}{s} \int_0^{l-a} f_T(t|X, \theta_0, \beta) \int_t^\infty f_C(c|\theta_c) dcdtda \end{aligned} \tag{5}$$

Scenario 4: Observation time l is after the enrollment period and greater or equal than the maximum follow-up time, but less than the sum of enrollment time and maximum follow-up time ($l > s, l > m, l < m + s$).

$$\begin{aligned} P(\delta = 0|X, \theta_0, \theta_c, \beta) =& P(T \leq C, T \leq m, A \leq l - m) + P(T \leq C, T + A \leq l, l - m \leq A \leq s) \\ =& \int_0^{l-m} \frac{1}{s} \int_0^m f_T(t|X, \theta_0, \beta) \int_t^\infty f_C(c|\theta_c) dcdtda + \int_{l-m}^s \frac{1}{s} \int_0^{l-a} f_T(t|X, \theta_0, \beta) \int_t^\infty f_C(c|\theta_c) dcdtda \end{aligned} \tag{6}$$

Scenario 5: Observation time l is greater or equal than the sum of enrollment time and maximum follow-up time and ($l \geq m + s$)

$$\begin{aligned} P(\delta = 0|X, \theta_0, \theta_c, \beta) =& P(T \leq C, T \leq m, A \leq s) \\ =& \int_0^s \frac{1}{s} \int_0^m f_T(t|X, \theta_0, \beta) \int_t^\infty f_C(c|\theta_c) dc dt da \end{aligned} \tag{7}$$

3 Application

In this section, we will assume the parametric form of baseline hazard function of survival time T is based on the weibull distribution, in which the baseline hazard function is

$$h_0(t|\theta_0) = h_0(t|\alpha_0, \nu_0) = \alpha_0 \nu_0^{-\alpha_0} t^{\alpha_0 - 1} \tag{8}$$

From (2) and (8), the probability density function of survival time T is

$$f_T(t|X, \theta_0) = f_T(t|X, \alpha_0, \nu_0) = \frac{\alpha_0}{[\nu_0 \exp(X\beta/\alpha_0)]^{\alpha_0}} t^{\alpha_0 - 1} \exp\left(-\frac{t^{\alpha_0}}{[\nu_0 \exp(X\beta/\alpha_0)]^{\alpha_0}}\right) \tag{9}$$

Therefore, the survival time T follows the weibull distribution with shape and scale parameters $(\alpha_0, \nu_0 \exp(X\beta/\alpha_0))$. We will assume random censoring time follows an exponential distribution with $\theta_1 = \gamma$. Therefore, the probability density function of random censoring time C is:

$$f_C(c|\theta_2) = f_C(c|\gamma) = \gamma \exp(-\gamma c) \tag{10}$$

The probability of observing the event of interests can be numerically evaluated, by solving tripple integrations in (3), (4), (5), (6) and (7) for each of the five scenarios.

Let random variable D_0 and D_1 denote the number of events observed in control group and experimental group, respectively. At a given time, D_i follows a binomial distribution $Bin\left(N_i, P(N_i, P(\delta = 0|X = i))\right)$. The expected number of events observed for both arms at a given time is:

$$E(D_0 + D_1) = N_0 \times P(\delta = 0|X = 0) + N_1 \times P(\delta = 0|X = 1) \tag{11}$$

Therefore, question 1 and 2 in section 2.1 can be answered. It is easy to indentify the probability of observing an event $P(\delta = 0|X)$ is a non-decreasing function of observing time l. And we can calculate time needed under the expcted number of events observed, denoted as l^*:

$$l^* = min\{l : E(D_0 + D_1) = N_0 \times P(\delta = 0|X = 0) + N_1 \times P(\delta = 0|X = 1)\} \tag{12}$$

l^* can be eveluated numerically. Therefore, question 3 in section 2.1 can be answered. Similarly, the random censoring rate γ can also be eveluated numerically, and question 4 in section 2.1 can be answered. We have developed R package **ClinTrialPredict** to implement these calculations and simulate datasets with specified design parameters.

4 Simulation

In this section, we conducted simulation study to assess the robustness of the analytical method in section 3 and 4, under different settings. We will consider a hypothetical two-arm randomized control clinical trial. The enrollment time s is set to 24 months; The single subject's maximum follow-up time m is evaluated at 6 months and 24 months, separately. For baseline hazard function, we will evaluate $\alpha_0 = 0.8, 1, 1.2$ separately, representing the baseline hazard is decreasing, constant and increasing; Hazard ratio between two arms is evaluated at $0.8, 1, 1.5$; The random censoring parameter γ is evaluated at $0.1, 0.2, 0.3$. All simulations are based on 10000 runs.

Table 1 and 2 compares the expected number of events observed from the analytical calculation with the mean of number of events observed from simulation datasets, under difference design settings. In table 1, the enrollment time s is less than the maximum follow-up time. While in table 2, the enrollment time is longer than the maximum follow-up time. The observation time l equals 8,13,20 in table 1 represents the Scenario 1, 4, and 5, respectively. And the observation time l equals 8,13,20 in table 1 represents the Scenario 1, 3, and 4, respectively. The results based on the analytical method and simulation method are consistent.

In table 3, the observation time is calculated based on the analytical method, and the mean of number of events observed from simulation datasets based on the calculated observation time is compared with desigated number of events observed. The results based on the analytical method and simulation method are consistent.

In table 4, the censoring parameter γ is calculated based on the anaytical method, and the mean of number of events observed from simulation datasets based on the calculated censoring parameter γ is compared with desigated number of events observed. The results based on the analytical method and simulation method are consistent.

Table 1: Comparing the expected number of events with mean number of events in simulation study ($N.0 = 200$, $N.1 = 200$, $s = 12$, $m = 6$)

Baseline hazard of survival time	Hazard ratio ($\exp(\beta)$)	Censoring rate (γ)	l=8		l=13		l=20	
			E(D)	Mean(D)	E(D)	Mean(D)	E(D)	Mean(D)
Weilbull(0.8, 20)	0.8	0.1	54.0	54.0	98.4	98.5	110.8	110.9
		0.2	46.2	46.2	82.2	82.3	90.6	90.7
		0.3	40.3	40.3	70.3	70.3	76.0	76.0
	1.0	0.1	48.9	48.9	89.2	89.3	100.6	100.8
		0.2	41.7	41.7	74.4	74.4	82.1	82.2
		0.3	36.3	36.3	63.5	63.5	68.7	68.8
	1.5	0.1	41.4	41.5	75.8	75.9	85.7	85.8
		0.2	35.3	35.3	63.1	63.1	69.8	69.8
		0.3	30.7	30.7	53.7	53.8	58.3	58.3
Weibull(1, 20)	0.8	0.1	39.8	39.9	75.4	75.5	87.5	87.6
		0.2	32.9	33.0	60.8	60.8	68.8	68.9
		0.3	27.8	27.8	50.2	50.2	55.6	55.7
	1.0	0.1	35.9	36.0	68.1	68.2	79.1	79.3
		0.2	29.6	29.7	54.8	54.8	62.1	62.2
		0.3	25.0	25.0	45.1	45.2	50.1	50.2
	1.5	0.1	30.3	30.4	57.6	57.7	67.1	67.2
		0.2	25.0	25.0	46.3	46.3	52.6	52.6
		0.3	21.1	21.1	38.1	38.1	42.4	42.4
Weilbull(1.2, 20)	0.8	0.1	29.5	29.6	57.8	57.8	68.9	69.0
		0.2	23.7	23.7	45.1	45.1	52.5	52.5
		0.3	19.4	19.4	36.1	36.1	41.0	41.1
	1.0	0.1	26.5	26.6	52.0	52.0	62.1	62.2
		0.2	21.2	21.3	40.5	40.5	47.2	47.2
		0.3	17.4	17.4	32.4	32.4	36.9	37.0
	1.5	0.1	22.3	22.4	43.9	43.9	52.5	52.5
		0.2	17.9	17.9	34.2	34.2	39.9	39.9
		0.3	14.6	14.7	27.3	27.3	31.1	31.2

Table 2: Comparing the expected number of events with mean number of events in simulation study ($N.0 = 200$, $N.1 = 200$, $s = 12$, $m = 18$)

Baseline hazard of survival time	Hazard ratio ($\exp(\beta)$)	Censoring rate (γ)	l=8		l=13		l=20	
			E(D)	Mean(D)	E(D)	Mean(D)	E(D)	Mean(D)
Weilbull(0.8, 20)	0.8	0.1	55.2	55.2	109.7	109.8	143.1	143.2
		0.2	46.8	46.9	87.4	87.5	103.9	104.1
		0.3	40.6	40.6	72.7	72.7	81.8	81.9
	1.0	0.1	50.0	50.0	100.0	100.0	131.5	131.7
		0.2	42.3	42.3	79.3	79.4	94.8	94.9
		0.3	36.6	36.6	65.8	65.8	74.2	74.3
	1.5	0.1	42.4	42.5	85.4	85.5	113.5	113.6
		0.2	35.9	35.9	67.5	67.5	81.1	81.2
		0.3	31.0	31.0	55.8	55.8	63.2	63.2
Weibull(1, 20)	0.8	0.1	41.1	41.2	88.2	88.4	124.9	125.0
		0.2	33.6	33.6	66.6	66.7	84.1	84.2
		0.3	28.2	28.2	52.9	52.9	62.2	62.3
	1.0	0.1	37.1	37.2	80.1	80.3	114.5	114.7
		0.2	30.3	30.3	60.3	60.3	76.5	76.6
		0.3	25.3	25.3	47.7	47.7	56.3	56.4
	1.5	0.1	31.4	31.5	68.2	68.3	98.6	98.7
		0.2	25.6	25.6	51.1	51.1	65.3	65.4
		0.3	21.3	21.4	40.3	40.4	47.8	47.9
Weilbull(1.2, 20)	0.8	0.1	30.9	31.0	71.5	71.6	110.0	110.1
		0.2	24.4	24.4	51.4	51.3	69.1	69.1
		0.3	19.8	19.8	39.0	39.0	48.1	48.2
	1.0	0.1	27.8	27.8	64.8	64.8	100.7	100.8
		0.2	21.9	21.9	46.4	46.3	62.7	62.8
		0.3	17.7	17.8	35.1	35.1	43.5	43.6
	1.5	0.1	23.4	23.5	55.1	55.1	86.5	86.6
		0.2	18.4	18.5	39.2	39.2	53.5	53.5
		0.3	14.9	15.0	29.6	29.6	36.9	37.0

Table 3: calculate l

	m	s	alpha0.t	nu0.t	HR	gamma.c	d	l	meanD
1	6.00	12.00	0.80	20.00	0.80	0.10	20.00	4.18	20.10
2	6.00	12.00	0.80	20.00	0.80	0.10	50.00	7.56	50.00
3	6.00	12.00	0.80	20.00	0.80	0.10	100.00	13.28	100.10
4	18.00	12.00	1.20	20.00	1.20	0.10	20.00	7.00	20.10
5	18.00	12.00	1.20	20.00	1.20	0.10	50.00	11.63	50.00
6	18.00	12.00	1.20	20.00	1.20	0.10	100.00	22.79	100.10

	m	s	alpha0.t	nu0.t	HR	l	d	gamma.c	meanD
1	6.00	12.00	0.80	20.00	0.80	3.00	10.00	0.31	10.00
2	6.00	12.00	0.80	20.00	0.80	8.00	50.00	0.15	49.90
3	6.00	12.00	0.80	20.00	0.80	15.00	60.00	0.45	60.20
4	18.00	12.00	1.20	20.00	1.20	5.00	8.00	0.27	8.00
5	18.00	12.00	1.20	20.00	1.20	15.00	40.00	0.27	39.90
6	18.00	12.00	1.20	20.00	1.20	20.00	70.00	0.16	69.50