An Optimal Design Strategy for Phase III Clinical Trials with Time-To-Event Endpoint

1 Introduction

For clinical trial designs with time-to-event endpoint, it is well-known that the "effective" sample size is not number of patients, but number of events based on log-rank test which is most powerful under the proportional hazard (PH) assumption. The unique challenge is how to choose the optimal combination of two important design elements: the sample size (i.e. number of patients) and the length of study duration because there are numerous possible combinations that can correspond to the same number of events. A trial design with shorter study duration and smaller sample size is more favorable in terms of drug development, but they can hardly be achieved simultaneously in the time-to-event framework with fixed number of events. For example, smaller sample size would only prolong the study duration because there are less chance to obtain the required number of events. On the other hand, a shorter study duration requires a larger sample size. It then becomes a game of weighting which factor is more important.

From financial perspective, larger sample size generally raises the cost of the trial, but longer study duration also increases the trial maintenance cost and most importantly, it delays market access, thus reducing potential revenue. This typically has larger impact than trial cost to drug developers. However, before the investigational drug can reach market, designs with very large sample size and short study duration may trigger questions from regulatory agencies and health technology assessment (HTA) bodies on data maturity despite statistically significant results, creating barriers for regulatory approval. For trials with rapid enrollment and short follow up duration (relative to the median time of event), the reported Kaplan Meier (KM) curve for the time-to-event endpoint such as overall survival (OS) may not capture the majority of the survival curve, thus making it hard to assess the long term benefit of survival and check the proportional hazard assumption. In additional, such data packages also lack long term safety data monitoring, making it difficult for the benefit and risk assessment. Given the potential undesirable impact of immature data, drug developers usually choose designs that ensure data maturity at the time of the primary readout. In

practice, this requires many rounds of trial and error before the final decision is made. Due to limited regulatory or established guidance on what constitutes mature data, the solution is often based on individual experience and judgement.

To find the optimal trial design without these iterations, we propose a general framework that formulates the trial design process into an optimization problem. Specifically, we create an objective function that represents the total net revenue (i.e. total revenue - cost), and the goal is to maximize the objective function. In the meanwhile, data maturity requirement will serve as constraints to the design parameters to eliminate combinations that yield immature data at time of primary analysis. The solution of such optimization problem is thus the desired study design which optimizes the pre-specified goal while ensuring data maturity required for drug approval.

The rest of the paper is organized as follows. In Section 2, we briefly review traditional design process for clinical trials with time-to-event endpoint. In Section 3, we present a framework of formulating the trial design procedure to an optimization problem. In Section 4, we introduce a way to find the optimal solution for the proposed optimization problem. In Section 5, we demonstrate the application of our proposal through a hypothetical example of phase III oncology clinical trials. We end this paper in Section 6 with a few concluding remarks. Technical details are provided as online supplementary material.

2 Current approaches for trial design with time-to-event endpoint

Let us assume there are in total N patients enrolled into the clinical trial. For patient i, denote A_i the time from study start to the time of enrollment; denote T_i the time from study entry to the time of onset of event; and C_i is the time from study entry to loss to follow up. Assuming each patient is followed up until a time of event or the end of study, whichever comes first, then at primary analysis time S (relative from study start), we observe $U_i = \min(\max(S - A_i, 0), T_i, C_i)$ where this patient either experienced event if $U_i = T_i$; or is loss to follow up if $U_i = C_i$, or is administratively censored if $U_i = S - A_i$. Specifically we denote the event indicator $\delta_i = 1[U_i = T_i]$, event count $N_i(x) = 1[U_i \leq x, \delta_i = 1]$ and at risk count $Y_i(x) = 1[U_i \geq x]$. For a parallel design with two treatment arms, let $Z_i = 0, 1$ indicate two treatment groups and p_0, p_1 indicate the corresponding randomization probability. Suppose the cumulative distribution function of $T_i|Z_i = j$ is $F_j(t)$, and of $C_i|Z_i = j$ is $G_j(t)$, $j = 0, 1$. Further denote the hazard and density function of $T_i|Z_i = j$ as $\lambda_j(t)$ and $f_j(t)$.

Currently, most designs for time-to-event endpoint assume proportional hazard, i.e. $\lambda_1(t)/\lambda_0(t)$ is a constant, denote as HR. The logrank test statis-

tic (Z) used for formal hypothesis test of $H_0 : HR = 1$ has the following form:

$$Z = \frac{\sum_k \left(X_{(k)} - \frac{\sum Y_1(t_k)}{\sum Y_0(t_k) + \sum Y_1(t_k)} \right)}{\sqrt{\sum_k \frac{\sum Y_0(t_k) \cdot \sum Y_1(t_k)}{(\sum Y_0(t_k) + \sum Y_1(t_k))^2}}},$$

where $\{t_k\}$ is the complete set of event times in the trial, $X_{(k)}$ is the assigned treatment for the patient who fails at t_k, and $\sum Y_j(t_k) = \sum_{i:X_i=j} Y_i(t_k)$ is the number of patients in arm j at risk at time t_k.

Based on asymptotic theory and arguments in [Schoenfeld, 1981], the logrank test statistic (Z) approximately follows normal distribution with mean $\log\text{HR}\sqrt{p_0 p_1 (Nd(t))}$ and variance 1. Interestingly, one can find that $d(t) = P(\delta_i(t) = 1)$, thus $Nd(t)$ can be interpreted as the number of required events. This effective sample size based on logrank test with statistical power $1 - \beta$ and Type I error rate α can then be calculated as follows:

$$E_a = \frac{(z_\alpha + z_{1-\beta})^2}{p_0 p_1 (\log(\text{HR}))^2}. \tag{1}$$

This means as long as the number of events E_a is observed, the statistical power of $1-\beta$ can be achieved. This is why in practice, the primary analysis timing of clinical trials with time-to-event endpoint is driven by the total number of events. Provided with accrual duration S_a and accrual rate $r(\cdot)$, the expected number of events observed at S is calculated as $E(S; S_a, r(\cdot)) = E^{(0)} + E^{(1)}$ where

$$\begin{aligned}
E^{(j)} &= N \Pr(\delta_i = 1 \mid Z_i = j) \Pr(Z_i = j) \\
&= N \Pr(A_i + T_i \le S, T_i \le C_i \mid Z_i = j) \Pr(Z_i = j) \\
&= N \int_0^{\min(S_a, S)} \int_0^{S-t} [1 - G_j(x)] f_j(x) dx dH(t) \Pr(Z_i = j) \\
&= \int_0^{\min(S_a, S)} \int_0^{S-t} [1 - G_j(x)] f_j(x) r(t) dx dt \Pr(Z_i = j)
\end{aligned} \tag{2}$$

and $H(t) = \Pr(A_i < t) = \frac{1}{N} \int_0^t r(a) da$ is the probability of being enrolled into the study before time t. The accrual rate $r(t)$ can be flexible, taking on any of the following forms:

$$r(t) = \begin{cases} r_u & \text{Uniform accrual rate} \\ \min(r_{max}, a_0 + a_1 t) & \text{Linear with maximum rate} \\ r_{t_k}, t_{k-1} < t \le t_k \text{ for } k = 1, 2, ... & \text{Piecewise accural rate} \end{cases}$$

Then, study duration S could be solved from equation (3)

$$E(S; S_a, r(\cdot)) = E_a. \tag{3}$$

Meanwhile, the total sample size is obtained immediately by

$$N = \int_0^{S_a} r(t)dt. \tag{4}$$

Notice that generally, this is a four-parameter (N, S, S_a, $r(\cdot)$) and two-equations (3 and 4) problem. For uniform accrual, as long as two of these four parameters are provided, the other two are then determined, so is the trial design. Popular statistical design software EAST (`http://www.cytel.com/software/east`) for example allows the user to input either $(r(\cdot), S_a)$, or $(S, S_a, r(\cdot)/N)$.

However, current softwares are not able to provide the optimal design under a pre-specified goal, be it 1) minimizing S and N simultaneously, or 2) minimizing total cost of trial or 3) maximizing net revenue given sufficiently mature data. Usually, investigators have to try and compare different input values of $(r(\cdot), S_a)$ to yield (N, S) for many rounds until the one that best meet the goal in their mind is found. Instead of this trial-and-error approach, we propose a trial design strategy that can directly yield the optimal solution of design parameters (N, S, S_a, $r(\cdot)$) to achieve the goal.

3 Method

The essence of our proposal is to provide a framework to determine the optimal combination of sample size and study duration while maintaining the data maturity. First, we introduce an objective function to quantify the goal that the investigator is trying to achieve with the design. From drug sponsor's perspective, that goal is usually based on financial evaluations, i.e. maximizing the expected net revenue (ENR) the drug can generate. The proposed objective function links both sample size N and study duration S to these financial terms directly and thus, by maximizing it, the solution is the trial design that yields the highest ENR. Next, we incorporate data maturity into the framework by placing constraints on the feasible combinations of (N, S) so that the resulting optimal design will guarantee mature data (defined by the user) at time of primary analysis.

3.1 Expected net revenue as an objective function

The ENR consists of two components, trial cost and expected total revenue. Trial cost usually increases with sample size and study duration, due to the routine costs of maintaining clinical site and data monitoring procedures. Therefore, we decompose the entire trial cost into three parts: a fixed cost c_0, a cost per each patient enrolled c_1, and a cost per each unit time c_2. Forecasting revenue once a drug product reaches the market is typically a complicated process in practice. In oncology, it may even require some

dynamic modeling of patient flows because most cancer treatments are prescribed by line of therapy. In this work, we capture the essence of total revenue and formulate it as the integration of revenue at time t, $b(t)$ (in US dollar) over total sales duration, which is the period from when the drug reaches the market to the time of loss of exclusivity (LOE). Let l denote the duration between date of trial start and LOE and l_0 denote the time between primary analysis and market access, then the approximate total duration of sales is $l - (S + l_0)$.

The well-known risk of drug development is that not all phase III trials will be successful. If a drug becomes a marketed product, all the total predicted revenue can be realized, but if it eventually fails in regulatory approval or proper reimbursement assessment, the total revenue becomes zero. We introduce $\mathcal{P}$ to represent the probability that the total revenue can be realized. Then, the objective function is the difference between the expected total revenue and trial cost expressed in equation (5).

$$\text{ENR(N,S)} := \mathcal{P} \cdot \underbrace{\int_0^{l-S-l_0} b(t)dt}_{\text{Revenue}} - \underbrace{\left(c_0 + c_1 N + c_2 S\right)}_{\text{Cost}} \quad \text{where (N,S)} \in C \tag{5}$$

In reality, drug sponsors have to overcome trial success, regulatory success and market access success before patients around the globe can have access to the effective new treatment. We propose several statistical terms to facilitate the estimation of this probability in the next section. For those factors outside of the trial data itself is beyond the scope of this paper and will not be discussed. Note that if $\mathcal{P}$ is set to zero, the objective function represents only the negative of trial cost. By maximizing it, the optimal design minimizes the trial cost.

We also add constraints denoted as C on the design parameters (N, S) to represent the additional requirements on mature data from regulatory as well as applicable health technology assessment (HTA) agencies. Different options of such requirement in terms of design parameters are further discussed in Section 3.3.

3.2 Probability of success $\mathcal{P}$

A natural choice of $\mathcal{P}$ is the probability of reaching statistically significant efficacy result. Once the number of events is fixed with the assumed treatment effect HR, $\mathcal{P}$ in equation (5) becomes the power $1 - \beta$. However, a statistically significant efficacy readout does not automatically translate into a regulatory approval. There are many other assessments performed by the regulatory agency before approval is granted. One of the most important evaluation related to efficacy is whether the treatment benefit is clinically

meaningful in light of the concurrent treatment paradigm in the proposed indication. A systematic review of more than 300 initial applications of new drugs between 2000 and 2012 reveals that only 73.5% of the applications are approved [Sacks, 2014] with one major source of delay or denial of approval being lack of clinically meaningful efficacy. We will next incorporate the clinical meaningfulness into the calculation of $\mathcal{P}$ to ensure that revenue is not generated unless trial is both statistically significant and clinically meaningful.

Unlike statistical significance, clinically meaningful results require the improvement of treatment effect big enough to be meaningful for patients in practice. One commonly used efficacy measure is median survival time, and the difference or ratio of medians between treatment and control arm typically represent relative treatment effect measures. Both American Society of Clinical Oncology (ASCO) and European Society for Medical Oncology (ESMO) also recognize these measures by incorporating them into their proposed value frameworks to evaluate the benefit of cancer therapies (Lowell et al. 2016; Cherny et al. 2017). We therefore propose the following two measures for clinically meaningful treatment effect:

A_1**:** $\hat{m_1} - \hat{m_0} > d_0$

A_2**:** $\hat{m_1}/\hat{m_0} > r_0$

where $\hat{m_j}$ is the estimated median survival time of arm j. With the assumption of exponential distribution of survival time with parameter λ_j, loss to follow-up time with parameter η_j for arm j, and uniform accrual, we are able to figure out the asymptotic distribution of $\hat{m}_j, j = 0, 1$. Then the probability of achieving either proposed clinical meaningful measures, $P(A_k), k = 1, 2$, can be written as follows with details in Appendix A:

$$\begin{aligned} P(A_1) &= 1 - \Phi\left(\frac{d_0 - \frac{\log 2}{\lambda_0}\left(\frac{1}{\mathrm{HR}} - 1\right)}{\frac{\log 2}{\lambda_0}\sqrt{\frac{1}{\mathrm{HR}^2 E^{(1)}} + \frac{1}{E^{(0)}}}}\right) \\ P(A_2) &= 1 - \Phi\left(\frac{\log\left(r_0 \mathrm{HR}\right)}{\sqrt{\frac{1}{E^{(1)}} + \frac{1}{E^{(0)}}}}\right), \end{aligned}$$

where $E^{(j)}, j = 0, 1$ is calculated from equation 2. By choosing either A_1 or A_2 as the measure of clinical meaningfulness, we define $\mathcal{P}_k = (1 - \beta)P(A_k)$ by taking both statistical significance and clinical meaningfulness into consideration.

3.3 Data maturity as constraints

Directly maximizing the ENR in equation (5) could possibly yield a design with very short study duration, which is more likely to be deemed immature by either regulatory agency or HTA bodies. This is because the design

with short study duration would lead to higher chance of producing KM curves with only the beginning portion of the survival curves, which is hard to assess long term benefit. Regulators as well as HTA agencies around the globe have now put more emphasis on data maturity, despite of statistically significant p-value and clinically meaningful treatment effect. Hence, we incorporate data maturity into our optimal trial design framework for this concern. In particular, data maturity requirements are placed as constraints on the design parameters so that any sets of (N, S) that can cause an immature KM curve are prohibited from being the optimal design.

To our knowledge, there has been no unanimous agreement or general guideline on the exact measure for data maturity. The first discussion on this topic appeared in a two-page survey by Shuster [Shuster, 1991], with additional discussions in [Altman, 1995], [Schemper, 1996], and [Clark et al., 2003]. Shuster introduced some commonly used measures of data maturity at that time and stated their necessity in statistical terms, although he remained skeptical on their utility. Since then, this thread of discussions on data maturity focused on calculating median follow up time.

We here summarize the following measures of mature data commonly used in practice:

C1. $S - S_a \geq t_0, \qquad t_0 > 0$

C2. $E_a/N \geq e_0, \quad e_0 \in (0,1)$

C3. P($\hat{m}_j^{KM}$ estimable for j=0,1)$> p_0, \quad p_0 \in (0,1)$

C4. $m_{fu} \geq m_0, \quad m_0 > 0$

where $\hat{m}_j^{KM}$ denotes the estimated KM median for arm j and m_{fu} is the true median of follow up time $R_i(t) = \min(t - A_i, C_i), i = 1, ..., N$ [Schemper, 1996] defined as $m_{fu} = \{m : P(R_i(S) \leq m) = 0.5\}$.

Measure $C1$ requires the minimum follow-up time is t_0. Measure $C2$ requires a the proportion of observed events E_a in total sample N is at least e_0. As E_a is fixed under assumed HR, it is equivalent to setting an upper bound for the total sample size N. Measure $C3$ represents high probability that median KM estimates are available for both arms. As the lowest point of the KM curve happens at the time of last event, the event {KM median is achieved} is equivalent to {survival probability at last event time no bigger than 50%}. We therefore have

$$P(\hat{m}_j^{KM} \text{ estimable } j = 0,1) = P(\min_x \hat{S}_0^{KM}(x) \leq 0.5) P(\min_x \hat{S}_1^{KM}(x) \leq 0.5) \quad (6)$$

where $\min_x \hat{S}_j^{KM}(x)$ indicates the lowest point of KM curve for arm j. Equation (6) can be obtained by simulations. Ensuring measure $C4$ is equivalent

to require $P(R_i(S) \leq m_0) \leq 0.5$ where

$$P(R_i(S) \leq m_0) = P(\min(S - A_i, C_i) \leq m_0) = 1 - P(S - A_i > m_0)P(C_i > m_0)$$
$$= 1 - \min\left(\max\left(0, \frac{1}{N}\int_0^{S-m_0} r(t)dt\right), 1\right)\sum_{j=0}^{1}(1 - G_j(m_0))Pr(Z_i = j).$$

By incorporating both probability of success and data maturity, the optimal trial design is eventually the solution of the following maximization problem:

$$\begin{aligned} \max_{N,S} \quad & \text{ENR}_k(\text{N,S}) = (1-\beta)P(A_k)\textstyle\int_0^{l-S-l_0} b(t)dt - \big(c_0 + c_1 N + c_2 S\big) \quad (7) \\ \text{s.t.} \quad & (\text{N,S}) \in C \subseteq \ \{\text{C1, C2, C3, C4}\}. \end{aligned}$$

C is the user specified set of constraints, which is an arbitrary combination of C1, C2, C3, C4. By specifying one or more threshold values t_0, e_0, p_0, and m_0, the corresponding constraints will be activated so that the optimal design will not violate these data maturity requirements.

4 Optimization

We have demonstrated that the optimal trial design is the solution of the optimization problem (7). In this section, we will introduce a general algorithm to solve it. Since it is a nonlinear constrained optimization problem without convex feature, general optimization techniques can not be applied. However, incorporating the special feature of trial design, we are able to simplify the problem to a simple line search.

We first explore the feasible set of this optimization problem. As aforementioned, even though the trial consists of four design parameters: sample size N, study duration S, accrual duration S_a, and accrual rate $r(\cdot)$, the common approach is given $r(\cdot)$ to solve for (N, S). This is because the maximum accrual rate is typically provided by the clinical trial operations team and the result of the feasibility assessment. Therefore it can not be freely adjusted upward. So we will display the feasible sets of parameter (N, S) given $r(\cdot)$. Such visualization procedure helps to narrow down the search area efficiently which eases the optimization procedure.

For illustration purpose, we set the exponential event time with medians 10 and 20 months for control and treatment arm, respectively (i.e. $HR = 0.5$), with 1 to 1 allocation ratio. Additionally, let $\beta = 0.1$, $\alpha = 0.025$, thus the number of events E_a calculated from equation (1) is 88. With further assumption of no loss to follow-up, and uniform accrual rate (i.e. $r(\cdot) = r_u$), the corresponding feasible sets are shown in Figure 1. Given the natural constraint that the accrual period S_a is bounded by 0 and study duration S, the feasible sets of (N, S) will not cover all N-S plain, but only

a portion of it, which is the shaded area. If the accrual rate is fixed at a number e.g. 10 patients per months, then the feasible set further reduces to a single red curve as shown in Figure 1. Fixing sample size N, when accrual rate r_u increases from 10 to 30, study duration S decreases since events accumulates much faster leading to shorter time to reach 88 events. It is easy to see that given N, smaller S always leads to larger ENR based on equation (7). Therefore, the optimal design only lies on the curve where r_u is the maximum feasible accrual rate.

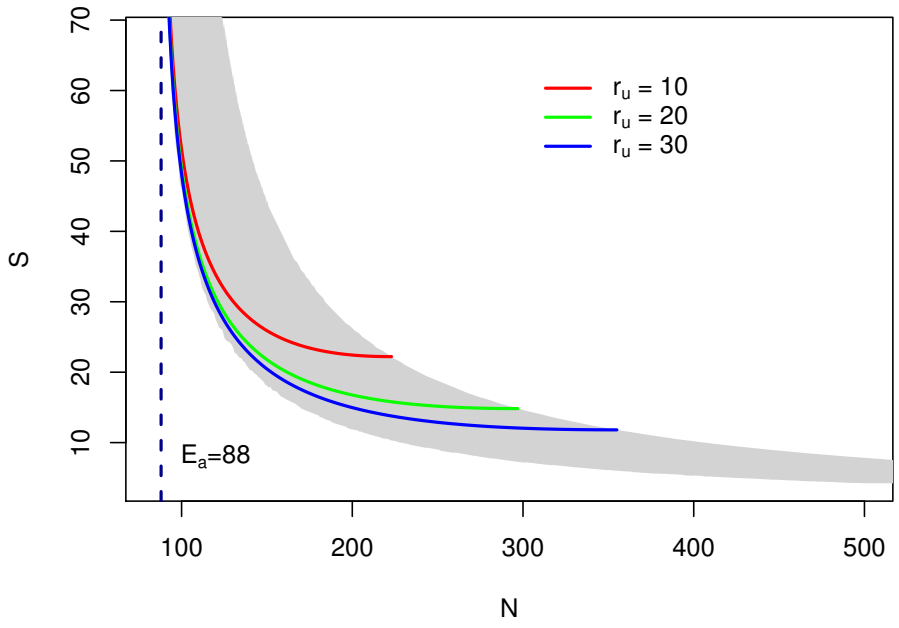

Figure 1: Feasible sets on N-S plane. Shaded area represents possible combinations of (N, S) under the natural constraint $0 < S_a \leq S$. Colored curves are the feasible sets with provided r_u.

After considering the data maturity constraints $C1 - C4$, the feasible sets are further restricted, as shown in Figure 2 where each panel illustrates the impact of one data maturity measure. The shaded lines represents the feasible sets in figure 1 that are no longer available under the specific data maturity requirement. Mostly, the eliminated sets are of large sample size but short study duration since such trial designs may lead to immature results in general.

The procedure of finding the optimal design based on equation (7) includes three steps. First, we figure out the feasible sets of (N, S), which is

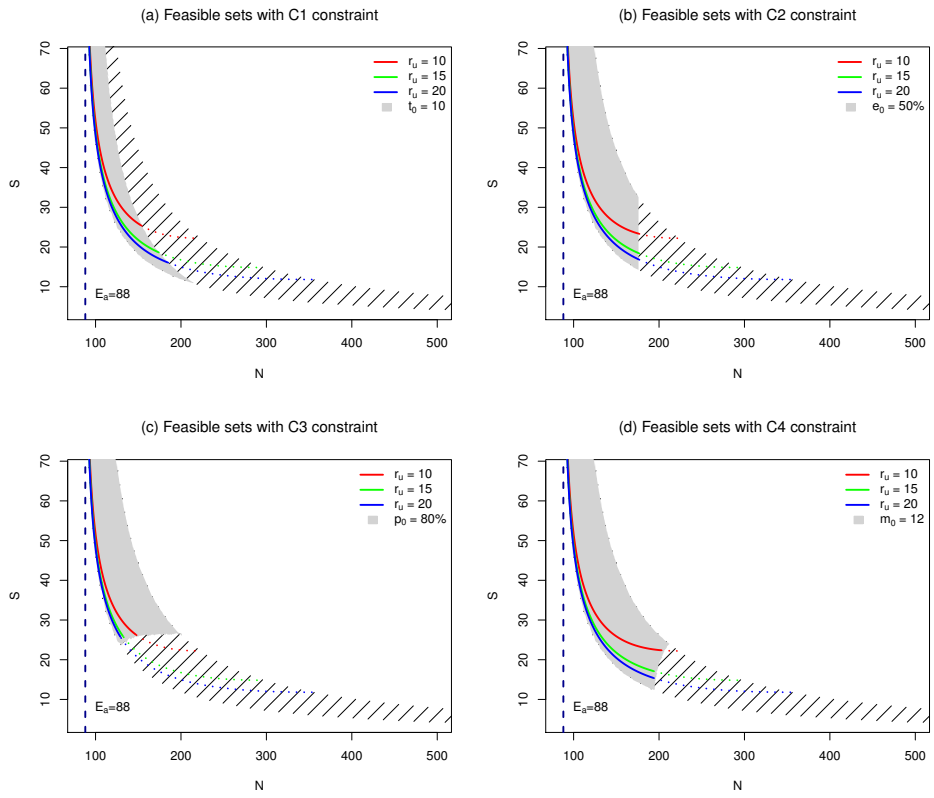

Figure 2: Feasible sets of (N, S) under different data maturity constraints. Shaded area represents feasible sets satisfying the given data maturity requirements, while the shaded lines cover the region that are no longer available in Figure 1. Colored curves are the feasible sets when r_u is given. (a) Data maturity requirement C1 is activated with $t_0 = 10$ month. (b) Data maturity requirement C2 is activated with $e_0 = 50\%$. (c) Data maturity requirement C3 is activated with $p_0 = 80\%$. (d) Data maturity requirement C4 is activated with $m_0 = 12$ month.

a single curve given the maximum accrual rate. Second, we remove unqualified part of the curve according to the pre-specified constraints C. Finally, we conduct a line search along the remaining curve to find the design parameter combination that maximizes the objective function. This is not a heavy task as N only takes integers and is further bounded by E_a and data maturity constraints.

5 Case Study

In this section, we will demonstrate how to use the proposed framework to design a hypothetical phase III oncology clinical trial and compare results with the traditional approach. Suppose the study sponsor would like to conduct an open label randomized clinical trial in a specific subtype of acute myeloid leukemia comparing the investigational new drug with the standard of care treatment. The primary endpoint for this indication that is regulatory approvable is overall survival (OS). With a type I error rate of 0.025 (1-sided), and power of 90%, the trial would require 326 events under the assumption of proportional hazards and exponential distributions with median OS of 21.5 months for the investigational drug vs 15 months for the standard therapy arm, i.e. HR=0.70. The loss to follow up is also set to be exponentially distributed with 5% annual rate for both arms.

Suppose 140 sites are planned to be used with an estimated overall enrollment rate of 40 patients per month at full speed after 7 months linear ramp up. With this information, the traditional approach can generate several design options as shown in Table 1, with input of peak enrollment rate r_{max} and sample size N, and output of enrollment duration S_a and study duration S.

Due to the rare indication and high cost of standard therapy, the overall trial is very expensive. Specifically, we assume the cost per patient is c_1 =\$300,000 and cost per month for site maintenance is c_2 =\$100,000 with a fixed upfront cost of c_0 =\$5,000,000. To get a rough idea of the total trial cost, we also report the cost of each scenario in Table 1, which is calculated based on equation (5) with $\mathcal{P} = 0$. For instance, suppose we plan the study with a sample size of 500 and a study duration of 40.0 months, then the total trial cost could add up to \$159.0 million, which is the most budget-friendly option among all 8 potential designs. This is because design 1 has smallest sample size and longest duration and per patient cost is much higher than the monthly site maintenance cost.

We further assume that the total duration from trial start to date of loss of exclusivity (LOE) is 15 years, and time between final analysis and market access is approximately 10 months. Suppose after drug enters the market, the sale is linearly increasing for 6 years until it reaches the peak sale of \$50 million per month. The resulting ENR using equation (5) with $\mathcal{P} = 1$ are

Table 1: Possible study designs given $r_{max} = 40$ patients per month

Design option	1	2	3	4	5	6	7	8
N	500	550	600	650	700	750	800	850
S	40.0	35.6	32.8	30.9	29.6	28.7	28.1	27.7
Sa	16	17.3	18.5	19.8	21	22.3	23.5	24.8
Cost($M)	159.0	173.6	188.3	203.1	218.0	232.9	247.8	262.8
ENR($M)	5865.4	6047.9	6159.6	6230.1	6273.9	6299.3	6311.5	6313.7
P($\hat{m}_1 - \hat{m}_0 > 4$)	88.3%	88.3%	88.2%	88.2%	88.2%	88.1%	88.1%	88.1%
$\text{ENR}_1(\$M)$	5163.1	5318.6	5412.4	5470.4	5505.4	5524.7	5532.6	5532.1
P($\hat{m}_1/\hat{m}_0 > 1.27$)	86.7%	86.6%	86.6%	86.6%	86.6%	86.6%	86.6%	86.6%
$\text{ENR}_2(\$M)$	5062.2	5216.8	5310.4	5368.4	5403.5	5423.0	5431.0	5430.6
C1: $S - S_a$	24.0	18.4	14.3	11.1	8.6	6.4	4.6	3.0
C2: E_a/N	0.65	0.59	0.54	0.50	0.47	0.43	0.41	0.38
C3: $P(\hat{m}_j^{KM}$estimable j=0,1)	100.0%	99.9%	99.2%	97.6%	94.0%	89.8%	87.8%	84.9%
C4: m_{fu}	29.4	24.5	21.1	18.6	16.7	15.2	14.0	13.0

shown in Table 1.

The aforementioned revenue can't be realized without statistical significant and clinically meaningful trial results. At the planned final analysis, the minimum observed HR to achieve statistical significance is 0.805. Suppose a median OS of 15 months is observed for the standard therapy arm, then a median of 18.64 months for the treatment arm (i.e. difference of 3.64 months) approximately corresponds to statistical significance. Suppose according to the key opinion leaders in the medical community, a minimum of 4 (> 3.64) months in median difference is required for the treatment to be meaningful to patients and treating physicians, then this translates to an observed treatment median of at least 9.5 months. Table shows the resulting ENR_1 and ENR_2 for $d_0 = 4$ and $r_0 = 1.27$.

In order to capture data maturity, the last four rows of Table 1 show the four proposed measures $C1 - C4$ for each design option. Suppose we require the minimum follow-up time is at least the median of control arm, i.e. 15 months to represent mature data at time of primary analysis. Meanwhile, assume the clinical meaningful results request $\hat{m}_1/\hat{m}_0 > 1.27$. Then in terms of ENR_2 from equation 7, we noticed that scenario 1 is no longer the optimal design. Instead, scenario 2, which has 550 patients, 35.6 months of study, and $S - S_a = 18.4 > 15$, yields a total of \$5216.8 million net revenue and outperforms the rest. Even though scenario 3 to 8 provide higher ENR_2, but they fail to satisfy the data maturity requirement.

In our proposed framework, there is no need to calculate all parts separately and try multiple options before choosing the best design. Instead we are able to obtain the optimal design given the objective function and specified data maturity constraints. Following the optimization process described in section 4, we first determine the feasible design sets for the given enrollment function as that is typically fixed and provided by the operations team as a result of the feasibility assessment. The grey curve in figure

3 represents all potential designs. Then, we remove the sets that do not satisfy the data maturity constraint $C1 : S - S_a \geq 15$, which is the dotted line in the figure. Finally, we search the remaining feasible sets and find the optimal design that provides the maximal ENR_2. It turns out that under this hypothetical setting, a trial design with sample size 590, study duration 33.3 months, 18.2 months of accrual, could achieve \$5295.1 million of ENR_2, \$78.3 million more than that in previous scenario 2. Meanwhile, the minimum follow-up time is 15 months, which agrees with the data maturity requirement.

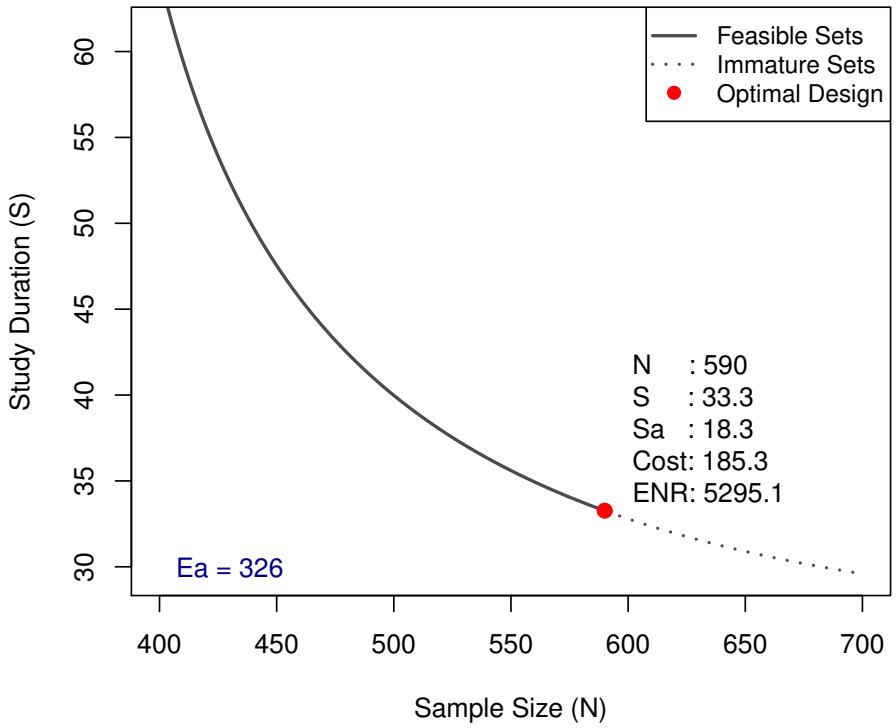

Figure 3: Feasible sets, optimal design, and the corresponding design parameters shown in N-S plane.

6 Conclusions

In this paper, we put forward a novel trial design strategy to directly yield the desired study design in investigator's mind without iterative discussions. By incorporating cost and revenue information, the trial design procedure

could be formulated into an optimization problem which the solution is the optimal design in terms of ENR. An important feature of the proposed method that sets it apart from traditional approach is it puts everything into a statistical framework, including clinical meaningful results and data maturity requirements. The output trial design thus automatically satisfies the user specific data maturity requirements which increases the odds in drug approval. This feature makes it particularly appealing in real world applications.

References

[Schoenfeld, 1981] Schoenfeld, David. "The asymptotic properties of non-parametric tests for comparing survival distributions." Biometrika 68.1 (1981): 316-319.

[DiMasi, 2016] DiMasi JA, Grabowski HG, Hansen RA. Innovation in the pharmaceutical industry: new estimates of R&D costs. Journal of Health Economics 2016;47:20-33.

[Shuster, 1991] Shuster, J. J. "Median follow-up in clinical trials." Journal of Clinical Oncology 9.1 (1991): 191-192.

[Altman, 1994] Altman, Douglas G., et al. "Dangers of using "optimal" cut-points in the evaluation of prognostic factors." JNCI: Journal of the National Cancer Institute 86.11 (1994): 829-835.

[Altman, 1995] Altman, D. G., et al. "Review of survival analyses published in cancer journals." British journal of cancer 72.2 (1995): 511-518.

[Clark et al., 2003] Clark, T. G., et al. "Survival analysis part I: basic concepts and first analyses." British journal of cancer 89.2 (2003): 232.

[Schemper, 1996] Schemper, Michael, and Terry L. Smith. "A note on quantifying follow-up in studies of failure time", Controlled clinical trials 17.4 (1996): 343-346.

[Sacks, 2014] Sacks, Leonard V., et al. "Scientific and regulatory reasons for delay and denial of FDA approval of initial applications for new drugs, 2000-2012." Jama 311.4 (2014): 378-384.

[Simone et al., 2008] Mathoulin-Pelissier, Simone, et al. "Survival end point reporting in randomized cancer clinical trials: a review of major journals." Journal of Clinical Oncology 26.22 (2008): 3721-3726.

[Kim & Tsiatis, 1990] Kim, Kyungmann, and Anastasios A. Tsiatis. "Study duration for clinical trials with survival response and early stopping rule." Biometrics (1990): 81-92.

A Calculating the probability of clinical meaningful treatment effect

Following the notation and assumption in section 2, the likelihood function for observed data $(U_i, \delta_i, Z_i), i = 1, 2, ..., N_0 + N_1$ is

$$\begin{aligned} L(\lambda_0, \lambda_1; U_i, \delta_i, Z_i) &= \prod_{j=0}^{1} \prod_{i=1}^{N_j} \Big([f(U_i; \lambda_j)]^{\delta_i} [S(U_i; \lambda_j)]^{1-\delta_i} \Big)^{1[Z_i=j]} \\ &= \prod_{j=0}^{1} \prod_{i=1}^{N_j} \Big([\lambda(U_i; \lambda_j)]^{\delta_i} [S(U_i; \lambda_j)] \Big)^{1[Z_i=j]}. \end{aligned}$$

Plug in the exponential survival and hazard function, we obtain the log-likelihood (for a single observation)

$$\ell(\lambda_0, \lambda_1; U_i, \delta_i, Z_i) = \sum_{j=0}^{1} 1[Z_i = j](log(\lambda_j)\delta_i - \lambda_j U_i).$$

Thus, the Fisher's information is

$$\mathcal{I}(\lambda_j) = -E\left[\frac{\partial^2 \ell}{\partial \lambda_j^2}\right] = \frac{E[\delta_i | Z_i = j]}{\lambda_j^2} = \frac{Pr(\delta_i = 1 | Z_i = j)}{\lambda_j^2},$$

and for N_j observations, the corresponding Information is

$$\mathcal{I}_n = N_j \mathcal{I}(\lambda_j) = \frac{N_j Pr(\delta_i = 1 | Z_i = j)}{\lambda_j^2} = \frac{E^{(j)}}{\lambda_j^2}.$$

Therefore, it follows from the property of MLE so that the estimated variable

$$\hat{\lambda}_j \to \mathcal{N}(\lambda_j, \mathcal{I}_n^{-1}) = \mathcal{N}\left(\lambda_j, \frac{\lambda_j^2}{E^{(j)}}\right)$$

in distribution as $E^{(j)} \to \infty$. For exponential distribution, the estimated median follows

$$\hat{m}_j = \frac{log(2)}{\hat{\lambda}_j}$$

and by delta method,

$$\hat{m}_j \to \mathcal{N}\left(\frac{log(2)}{\lambda_j}, \frac{(log(2))^2}{\lambda_j^2 E^{(j)}}\right).$$

Therefore, $P(A_1)$ could be figured out analytically given the true parameter values by

$$\hat{m}_1 - \hat{m}_0 \to \mathcal{N}\left(\frac{\log 2}{\lambda_1} - \frac{\log 2}{\lambda_0}, (\log 2)^2 \left(\frac{1}{\lambda_1^2 E^{(1)}} + \frac{1}{\lambda_0^2 E^{(0)}}\right)\right)$$

$$\Rightarrow P(A_1) = P(\hat{m}_1 - \hat{m}_0 > d_0) = 1 - \Phi\left(\frac{d_0 - \log 2\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_0}\right)}{\log 2\sqrt{\frac{1}{\lambda_1^2 E^{(1)}} + \frac{1}{\lambda_0^2 E^{(0)}}}}\right).$$

Similarly, for $P(A_2)$, we first calculate the asymptotic distribution for $\log(\hat{m}_j)$ by delta method

$$\log \hat{m}_j \to \mathcal{N}\left(\log\left(\frac{log(2)}{\lambda_j}\right), \frac{1}{E^{(j)}}\right).$$

Thus,

$$\log \hat{m}_1 - \log \hat{m}_0 \to \mathcal{N}\left(\log\left(\frac{\lambda_0}{\lambda_1}\right), \left(\frac{1}{E^{(1)}} + \frac{1}{E^{(0)}}\right)\right)$$

$$P(A_2) = P(\hat{m}_1/\hat{m}_0 > r_0) = P(\log \hat{m}_1 - \log \hat{m}_0 > \log(r_0)) = 1 - \Phi\left(\frac{\log\left(\frac{r_0\lambda_1}{\lambda_0}\right)}{\sqrt{\frac{1}{E^{(1)}} + \frac{1}{E^{(0)}}}}\right).$$

By replacing λ_1 with λ_0HR, we obtain the formulas in section 3.2.