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

November 8, 2020

1 Introduction

In clinical trials, the use of time-to-event endpoint creates some difficulties at design stage even under the proportional hazard (PH) assumption. While traditional design strategies focus on finding the minimum sample size given fixed power and significance level for testing a hypothesis, a trial with time-to-event endpoint can be adapted to the same framework. The difference though, is that the "effective" sample size is not the number of patients, but the number of events instead (based on logrank test). The real difficulty 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 numerously many possible combinations that can yield the exact same number of events.

Any trial 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. By the same token, a shorter study duration requires a larger sample size. It then becomes a game of weighting which factor is more important. Larger sample size generally raises the cost of the trial, while longer study duration increases the maintenance cost as well as delays market access, which influence the potential revenue. Therefore, in practice, a balance between sample size and study duration is often required during the design stage and several iterative discussions are typically involved before the final design parameters are agreed.

Even though there are technically numerous combinations of sample size and study duration that could achieve the same statistical power, designs with very large sample size and short study duration often trigger questions from regulatory agencies despite statistically significant results and may

thus hinder the regulatory approval. This is because a short study duration, relative to the median time of event, can result in a Kaplan Meier (KM) curve appearing to contain only the beginning part of an entire survival curve. Such curves typically are deemed immature by regulatory agencies and health technology assessment (HTA) bodies due to several reasons: 1) difficulty in assessing proportional hazards assumption which is crucial in the interpretation of clinical results such as hazard ratio (HR); 2) unreliability or large variability in the observed efficacy measure; 3) insufficient long term safety data. Given the potential undesirable impact of immature data, drug developers usually choose designs that ensures data maturity at the time of the primary read out, which further complicates the trial-and-error search for the most appropriate study design. Due to limited regulatory or established guidance on what constitutes mature data, the solution is often based on individual experience rather than probabilistic evaluations.

To find the optimal trial design without this iterative decision making process, we propose a general framework that formulate the trial design to an optimization problem. The investigator's primary goal, such as maximizing the total revenue, will be clearly depicted by the objective function. Meanwhile, the data maturity requirements serve as constraints to eliminate the 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 the data maturity to facilitate the drug approval.

The rest of the paper is organized as follows. In Section 2, we briefly review traditional sample size calculation 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 analysis time t (relative from study start), we observe $U_i(t) = \min(\max(t - A_i, 0), T_i, C_i)$ where this patient either experienced event if $U_i(t) = T_i$; or is loss to follow up if $U_i(t) = C_i$, or is administratively censored if $U_i(t) = t - A_i$. Specifically we denote the event indicator $\delta_i(t) = 1[U_i(t) = T_i]$, event count $N_i(t, x) = 1[U_i(t) \leq x, \delta_i(t) = 1]$ and at risk count $Y_i(t, x) = 1[U_i(t) \geq x]$. 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. Based on asymptotic theory (Tsiatis 1982), the numerator (U) of the logrank test statistic (Z) has the following form and follows normal distribution.

$$\begin{aligned} U &= \sum_{i=1}^{N} \int_0^t \left(Z_i - \frac{\sum_{l=1}^{N} Z_l Y_l(t,x)}{\sum_{l=1}^{N} Y_l(t,x)} \right) dN_i(t,x) \\ &\sim \mathcal{N}(N \log(HR) h(t), N h(t)) \end{aligned} \tag{1}$$

where

$$h(t) = \int_0^t \mu(t,x)(1 - \mu(t,x)) V(t,x) dx;$$

and

$$\begin{aligned} \mu(t,x) &= \frac{E(Y_i(t,x)|Z_i = 1)p_1}{E(Y_i(t,x))} \\ &= \frac{P(U_i(t) \geq x)|Z_i = 1)p_1}{P(U_i(t) \geq x)|Z_i = 1)p_1 + P(U_i(t) \geq x)|Z_i = 0)p_0}; \\ V(t,x) &= \sum_{j=0}^{1} \lambda_j(x) P(U_i(t) \geq x)|Z_i = j) p_j. \end{aligned}$$

Therefore, the logrank test statistic (Z) follows normal distribution with mean $\text{logHR}\sqrt{Nh(t)}$ and variance 1. Based on the arguments from [Schoenfeld, 1981], $\mu(t,x) \to p_1$, $1 - \mu(t,x) \to p_0$, let $d(t) = \int_0^t V(t,x)dx$, then the mean becomes $\text{logHR}\sqrt{p_0 p_1 (N d(t))}$. Interestingly, one can find that $d(t) = P(\delta_i(t) = 1)$, thus $Nd(t)$ can be interpreted as the number of required events. The effective sample size based on logrank test with statistical power $1 - \beta$ and Type I error rate α can then be calculated as follows [Schoenfeld, 1981]:

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

Under the framework of time-to-event analysis, the study stops when the expected number of events E aggregated from both arms $Z \in \{0, 1\}$ reaches the desired effective sample size E_a. Provided with accrual duration S_a and accrual rate $r(\cdot)$, the expected number of events observed at S is determined by

$$\begin{aligned}
E(S; S_a, r(t)) &= \sum_{i=1}^{N} \Pr(\delta_i = 1) \\
&= \sum_{j=0}^{1} N \Pr(\delta_i = 1 \mid Z_i = j) \Pr(Z_i = j) \\
&= \sum_{j=0}^{1} N \Pr(A_i + T_i \leq S, T_i \leq C_i \mid Z_i = j) \Pr(Z_i = j) \\
&= \sum_{j=0}^{1} N \int_0^{\min(S_a, S)} \int_0^{S-t} [1 - G_j(x)] f_j(x) dx dH(t) \Pr(Z_i = j) \\
&= \sum_{j=0}^{1} \int_0^{\min(S_a, S)} \int_0^{S-t} [1 - G_j(x)] f_j(x) r(t) dx dt \Pr(Z_i = j)
\end{aligned}$$

where $H(t) = \Pr(a < t) = \frac{1}{N} \int_0^t r(a) da$ is the probability of being enrolled into the study before time t. The accrual rate $r(t)$ is flexible, which could be

$$r(t) = \begin{cases} r_u & \text{Uniform accrual rate} \\ \min(r_{max}, r_0 + r_1 t) & \text{Linear with maximum rate} \\ r_{t_k}, t_{k-1} < t < t_k \text{ for } k = 1, 2, ... & \text{Piecewise accural rate} \end{cases}$$

Then, study duration S could be solved from equation

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

Meanwhile, the total sample size is obtained immediately by

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

Notice that basically, this is a four-parameter (N, S, S_a, $r(\cdot)$) and two-equation (equation 3 and 4) problem. No matter which two of these four parameters are provided, the other two are then determined, so is the trial design. For instance, in EAST (`http://www.cytel.com/software/east`), user could input either $(r(\cdot), S_a)$ or (S, S_a) (where a uniform accrual rate is

assumed), and then by solving equation 3, S or r_u is found, and finally the solution for N.

However, current softwares are not able to provide the optimal combination of (N, S) under a pre-specified goal, be it 1) minimizing S and N simultaneously, or 2) minimizing total cost of trial or 3) maximizing revenue given sufficiently mature data. Usually, investigators have to try and compare different inputs 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 (N, S) in terms of the investigator's goal.

3 Method

The essence of our proposal is to provide an optimization 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 S and study duration N to these financial terms directly and thus, by maximizing it, the solution is the trial design that yields the most ENR. Second, 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 analysis.

3.1 Expected net revenue as an objective function

The ENR is considered to consist 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 final 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, 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

$$\text{ENR(N, S)} := \mathcal{P} \cdot \underbrace{\int_0^{l-S-l_0} b(t)dt}_{\text{Revenue}} - \underbrace{\big(c_0 + c_1 N + c_2 S\big)}_{\text{Cost}}, \quad b(t), c_0, c_1, c_2 \geq 0. \tag{5}$$

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 by setting $\mathcal{P}$ to zero, the objective function represents only the negative trial cost. By maximizing it, the optimal design is the most budget saving one.

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 the treatment is big enough for the patients in practice. One commonly used meaningful measure is the median survival time, and the difference or ratio of medians between treatment and control arm are regarded as meaningful treatment efficacy measures. The ASCO value framework

(cite here) also recognized these measures by allowing bonus points if median survival is xxx. 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 parametric assumption of the survival time, loss to follow-up time, and administration censoring time, we are able to figure out the asymptotic distribution of $\hat{m}_j, j = 0, 1$. Then the probability of achieving either clinical meaningful results, $P(A_k), k = 1, 2$, is figured by (details see Appendix C)

$$\begin{aligned} P(A_1) &= 1 - \Phi\left(\frac{d_0 - \log 2\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_0}\right)}{\log 2\sqrt{\frac{1}{\lambda_1^2 E^{(1)}} + \frac{1}{\lambda_0^2 E^{(0)}}}}\right); \\ P(A_2) &= 1 - \Phi\left(\frac{\log\left(\frac{r_0\lambda_1}{\lambda_0}\right)}{\sqrt{\frac{1}{E^{(1)}} + \frac{1}{E^{(0)}}}}\right). \end{aligned}$$

where

$$E^{(j)} = \int_0^{\min(S_a, S)} \int_0^{S-t} [1 - G_j(x)] f_j(x) r(t) dx dt \Pr(Z_i = j), \quad j = 0, 1.$$

By choosing either A_1 or A_2 as the measure of clinical meaningfulness, the probability of a success trial is thus updated to $\mathcal{P} = (1 - \beta)P(A_k)$ where both statistical significance and clinical meaning have been taken into consideration. The objective function of ENR is eventually

$$\text{ENR}_k(N, S) = (1 - \beta)P(A_k) \cdot \int_0^{l-S-l_0} b(t)dt - \big(c_0 + c_1 N + c_2 S\big), \quad k = 1, 2. \tag{6}$$

3.3 Data maturity as constraints

Directly maximizing the ENR in equation (6) could possibly yield a design with very short study duration, which is more likely to be deemed immature by either regulatory agency or health technology assessment (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 validate the PH assumption. 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 the 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 about data maturity appeared in a two-page survey by Shuster [Shuster, 1991], and the discussion continues 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, however he remained skeptical on their utility. Since then, this thread of discussions on data maturity focused on calculating median follow up time. We here propose the following measures of data maturity:

1. the minimum follow-up time longer than a certain period, for instance, the median event time of control arm;
2. events observed are sufficient in terms of total patients;
3. a high probability that median estimates are available for both arms;
4. median follow-up time based on reverse KM method for all patients is sufficiently long.

Each of these definitions reflects a different aspect of data maturity. We provide all these as options and either one or more could be specified for a trial design. In the following paragraphs, we will discuss how to associate these measures with the design parameters.

In measure 1, the follow-up time is $S - S_a$ which directly put constraint on the minimum difference between study duration S and accrual duration S_a. Measure 2 requires a minimum proportion of observed events E_a in total sample N, i.e. E_a/N. As E_a is fixed under proportional hazard assumption, it is equivalent to an upper bound for the total sample size N.

The key of measure 3 is to obtain the probability that the KM median could be obtained within each arm. 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%}. Denote the probability that median estimates are available for both arms as $P(\hat{m}_j \text{ estimable } j = 0, 1)$, and we have

$$\begin{aligned} & P(\hat{m}_j \text{ estimable } j = 0, 1) = P(\hat{m}_0 \text{ estimable })P(\hat{m}_1 \text{ estimable }) \\ = \; & P(\min_x \hat{S}_0^{KM}(x) \leq 0.5)P(\min_x \hat{S}_1^{KM}(x) \leq 0.5) \end{aligned} \tag{7}$$

where $\min_x \hat{S}_j^{KM}(x)$ indicates the lowest point of KM curve for arm j. To estimate it, we simulate a reasonably large number of trials by sampling event times, loss to follow up times, and enrollment times following the assumptions described in section 2, and count the proportion of the KM curves with the last point under 0.5.

By reverse KM method [Schemper, 1996], the median of $R_i(t) = \min(t - A_i, C_i), i = 1, ..., N$ represents the median follow-up m_{fu} which can be found by the following formula

$$m_{fu} = \{m : P(R_i(S) \leq m) = 0.5\}$$

where

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

If measure 4 claims a mature data with m_{fu} larger than some pre-specified number m_0, the design parameter (N, S) should be adjusted to guarantee $P(R_i(S) \leq m_0) \leq 0.5$.

To sum, the mathematical expressions for all four data maturity requirements are respectively

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

and the optimal trial design is eventually the solution of the following optimization problem,

$$\begin{aligned} \max \quad & \text{ENR}_k(\text{N,S}), \\ s.t. \quad & \text{C} \subseteq \text{ C1, C2, C3, C4,} \end{aligned} \tag{8}$$

where C is the constraint, 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 8. 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 shall not be applied straightforwardly. However, considering the special feature of trial design, we are able to simplify the original problem to a simple line search.

4.1 Feasible sets of N and S

We would 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 $r(\cdot)$, knowing two of them determines the rest. So we will mainly focus on N and S and display the feasible sets of (N, S) in a N-S plain. Such visualization procedure helps to narrow down the searching area efficiently which eases the optimization procedure.

For illustration purpose, we set the exponential event time with medians 10 and 20 (unit of time) 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 (2) is 88. Further let no loss to follow-up, and uniform accrual rate.

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. Basically, we need to search over this feasible area to find one that maximize the objective function 6, which is the solution.

In reality however, when designing a trial, we have to consider the capacity of sites in enrolling patients, or some other limitations. This means the feasible region should be further restricted by excluding those combinations with unfeasible $r(\cdot)$ or S_a. In many situations, when investigators have the idea of either S_a or r_u, the feasible sets are further narrowed to a single curve by losing a degree of freedom (colored curves in Figure 1). In panel (a), it shows that for a fixed N, a smaller S_a, which implies a larger accrual rate, shortens the whole study duration. It is consistent with the results from panel (b) where curve with larger r_u is underneath.

After considering the data maturity constraints, the feasible sets are further restricted, as shown is Figure 2. Each panel illustrates the impact of one data maturity measure. The shaded lines represents the feasible sets

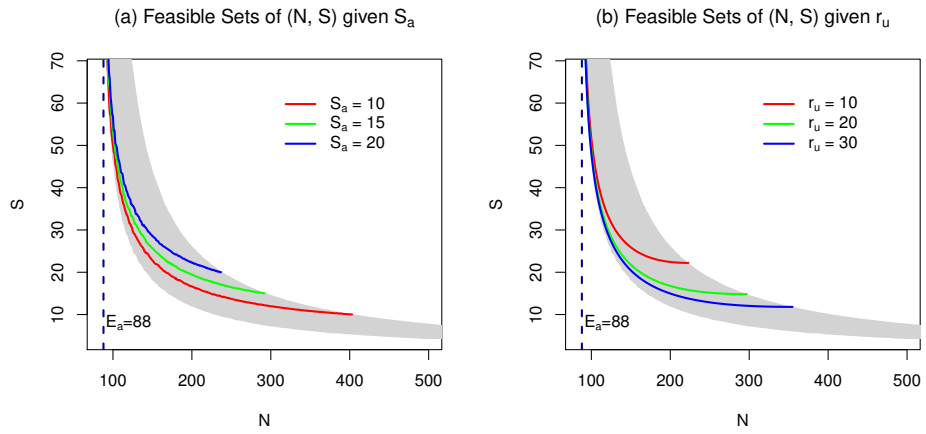

Figure 1: Feasible sets of (N, S). Shaded area represents possible combinations of (N, S) under the natural constraint $0 < S_a \leq S$. (a) Colored curves are the feasible sets when S_a is given. (b) Colored curves are the feasible sets when r_u is given.

in figure 1 that are no longer available under the specific maturity requirement. Basically, the eliminated sets are of large sample size but short study duration. It makes perfect sense since such trial designs are considered as immature in general.

The general procedure of finding the optimal solution when either $r(\cdot)$ or S_a is provided includes 3 steps. First, we figure out the feasible sets of (N, S), which is a single curve. Then, get rid of unqualified proportion according to the pre-specified constraints C. Finally, we could conduct a line search along the remaining curve to find the combination that maximize the objective function. This is not a heavy task as N only takes integers which are further bounded by E_a and data maturity constraints.

5 Case Study

In this section, we will demonstrate how to use the proposed framework for practical problem solving. We introduce a hypothetical phase III oncology clinical trials conducted under the traditional event-driven design, and compare results with the alternatives from our proposed framework.

5.1 XXX trial on acute myeloid leukemia

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.

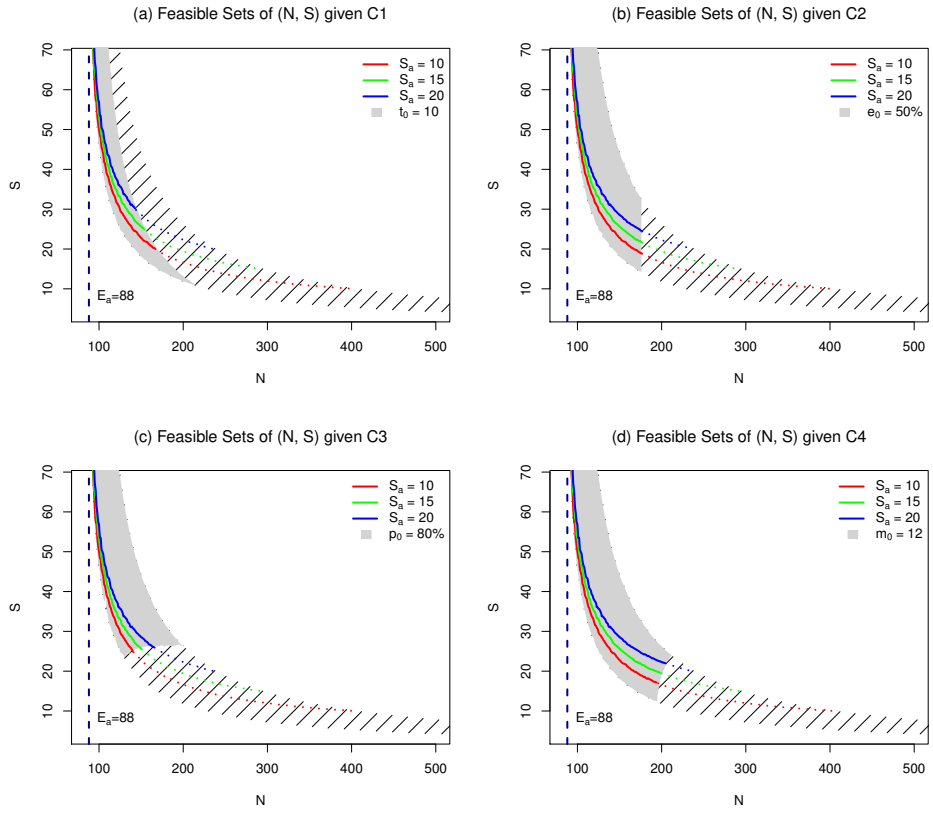

Figure 2: Feasible sets of (N, S) under different data maturity constraints. Shaded area represents all feasible sets satisfying the given data maturity requirements, while the shaded lines cover the region that are no longer available. Colored curves are the feasible sets when S_a is given. (a) Data 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.

With a type I error rate of 0.025 (1-sided), and power of 90%, the trial would require 372 events with the assumption of median OS of 10.5 months for the investigational drug vs 7.5 months for the standard therapy arm, i.e. HR=0.71. The loss to follow up is set to be 5% per year for both arm.

The sponsor can open as much as 40 sites with an estimate of enrolling 1 patient per site per month. So the fastest enrollment rate is 40 patients per month. With this information, the traditional approach allows the sponsor to input enrollment rate of 40, and sample size, then enrollment duration, and study duration could be determined accordingly. Table 1 below shows 8 possible design options.

Due to the rare indication and high cost of standard therapy, the overall trial is very expensive. Specifically, assume the cost per patient is \$300,000, and cost per month for site maintenance is \$100,000 with a fixed upfront cost of \$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 the *Cost* part in equation 5. For instance, suppose we plan the study with a sample size of 400 and a study duration of 55.53 months, then the total trial cost could add up to \$130.6 million, which is the most budget-friendly one among all 8 scenarios. The reason is, in scenario 1, it sacrifices study duration for samples since samples are more "expensive" in terms of study cost comparing to study time.

Further assume that the total duration from trial start to date of LOE is 15 years, time between final analysis and market access is approximately 10 months, and for each month the average total revenue it can generate is \$10 million. Also we require the minimum follow-up time at least larger than the median of control arm, i.e. 7.5 month, for a mature data at time of analysis. Meanwhile, assume the clinical meaningful results request $\hat{m}_1/\hat{m}_0 > 1.2$, where 1 indicates the investigational drug. Then in terms of ENR from equation 6, we noticed that scenario 1 is no longer the optimal design. Instead, scenario 7, which requires 580 patients, 21.68 months of study, and generates a total of \$1061.1 million net revenue, outperforms the rest. However, the minimum follow-up time in scenario 7 is only $21.68 - 14.5 = 7.18 < 7.5$, which violates the data maturity requirement. Therefore, the desired optimal design among the proposed 8 scenarios turns out to be scenario 6, which has the second largest ENR, \$1060.7 million, while assuring the maturity of data.

In our method, there is no need to try multiple scenarios, choose the best one, and then check data maturity constraint. Instead we are able to obtain the optimal solution immediately. Following the optimization process described in section 4, we first need to determine the feasible sets for a fixed

Table 1: Possible study designs given $r_a = 40$

Scenario	1	2	3	4	5	6	7	8
r_f	40	40	40	40	40	40	40	40
N	400	420	440	460	500	540	580	620
S	55.53	39.43	33.33	29.74	25.54	23.16	21.68	20.73
Sa	10	10.5	11	11.5	12.5	13.5	14.5	15.5
Cost(M/$)	130.6	134.9	140.3	146.0	157.6	169.3	181.2	193.1
ENR(M/$)	829.0	959.5	1005.1	1029.4	1052.7	1060.7	1061.1	1057.1

Table 2: Optimal study design in terms of ENR under different financial settings

	Value(M/$)	Sample Size	Study Duration	Accrual Duration	Cost(M/$)	ENR(M/$)
	0.3	562	22.27	14.05	175.8	1061.6
Cost/Pts	0.6	505	25.18	12.63	310.5	902.8
	0.9	478	27.51	11.95	438.0	755.9
	0.1	562	22.27	14.05	175.8	1061.6
Cost/Mo	0.3	564	22.20	14.10	180.9	1057.2
	0.5	566	22.13	14.15	185.9	1052.7
	10	562	22.27	14.05	175.8	1061.6
Sales/Mo	20	574	21.86	14.35	179.4	2302.2
	30	574	21.86	14.35	179.4	3542.9

$r_a = 40$, which is the trajectory displayed in a grey curve in figure 3. Then, we remove the sets that do not satisfy the data maturity constraint, which is the dotted line in the figure. Eventually, we search on the remaining feasible sets and find one that provide the maximal ENR. It turns out that the maximal ENR could be achieved under this hypothetical setting is \$1061.6 million, \$0.9 million more than the scenario 6, with a sample size of 562, a study duration of 22.27 months, 14.05 months of accrual, and the minimum follow-up time is 8.22 months, which agrees with the data maturity requirement.

The ENR with different financial settings is illustrated in table 2. Different values of cost per patient, cost per month, and sales per month are examined, with r_f fixed at 40. If not specified otherwise, cost per patient, cost per time, and sales per month remain the original set up, i.e. \$0.3 million, \$0.1 million, and \$10 million, respectively. Obviously, the larger cost per month, the less total net revenue, or on the other hand, the more you sale per month, the more you earn. Furthermore, among three different parameters, sales/month has the greatest impact on the final net revenue. It is noteworthy that in the table, same trial designs are provided for Sales/Mo equals to 20 and 30, as it could no longer pursue higher ENR by shortening the study duration without violating the data maturity constraint.

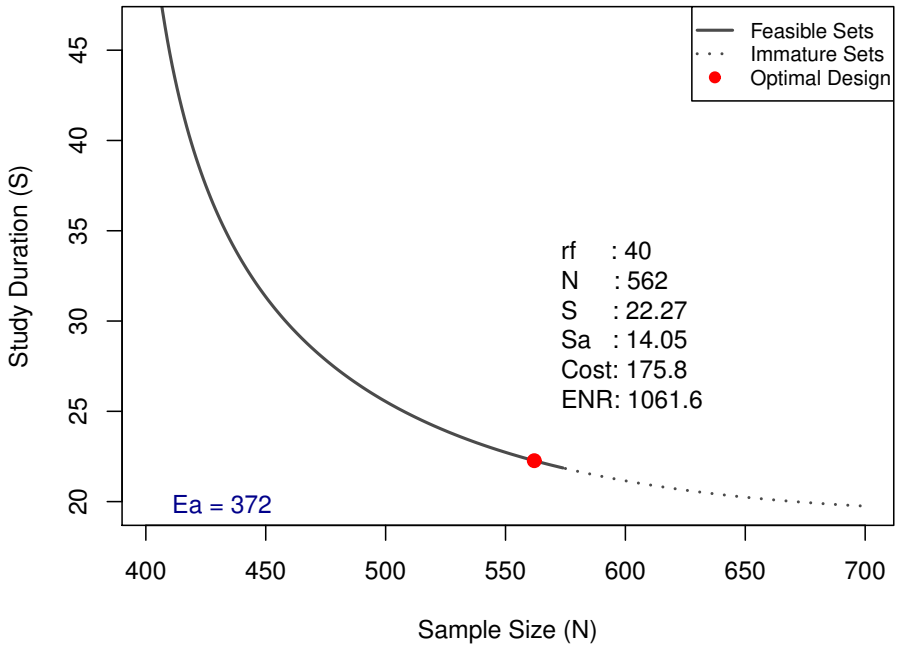

Figure 3: Feasible sets, optimal design, and the corresponding design parameters shown in N-S plane.

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 Monotonous relationship between N and S given S_a or r_a

Given r_a and target event number E_a, according to equation (**??**) and (**??**), study duration S and total sample size N for a single arm satisfies the following equation:

$$r_a \frac{\lambda}{\lambda+\eta}\left[\min(\frac{N}{r_a}, S) - \frac{e^{-(\lambda+\eta)S}}{\lambda+\eta}(e^{(\lambda+\eta)\min(\frac{N}{r_a},S)} - 1)\right] - E_a = 0. \quad \text{(A.1)}$$

First considering the case that study duration last longer than the accrual period, i.e. $\min(N/r_a, S) = N/r_a$. With some simple algebra, we obtain that

$$\begin{aligned} S(N|r_a, E_a) &= -\frac{1}{\lambda+\eta}\log\left[\frac{(\lambda+\eta)(\frac{N}{r_a} - \frac{E_a}{r_a}\frac{\lambda+\eta}{\lambda})}{e^{(\lambda+\eta)\frac{N}{r_a}} - 1}\right] \\ &= -\frac{1}{\lambda+\eta}\log\left(f(N|r_a, E_a)\right). \end{aligned} \quad \text{(A.2)}$$

Let $f(N|r_a, E_a) = \frac{aN - bE_a}{e^{aN}-1}$, where $a = (\lambda+\eta)/r_a, b = (\lambda+\eta)^2/(\lambda r_a)$, and take derivative with respect to N, we get

$$\frac{\partial f}{\partial N} = \frac{a}{(e^{aN}-1)^2}\left[(1 - aN + bE_a)e^{aN} - 1\right]. \quad \text{(A.3)}$$

Due to the fact that $S > N/r_a$, it could be derived that

$$\begin{aligned} -\frac{1}{\lambda+\eta}\log(f) > N/r_a \quad &\Rightarrow \quad f < e^{-aN} \\ &\Rightarrow \quad \frac{aN - bE_a}{e^{aN}-1} < e^{-aN} \\ &\Rightarrow \quad (1 - aN + bE_a)e^{aN} > 1 \\ &\Rightarrow \quad \frac{\partial f}{\partial N} > 0. \end{aligned}$$

Therefore, $f(N)$ is a monotonically increasing function with respect to N and due to the fact that $-\log()$ is a monotonically decreasing function, S is a monotonically decreasing function of N.

In the case that $\min(N/r_a, S) = S$, or equivalently, $S \leq N/r_a$, N is no longer involved in the equation (A.1) and S is only determined by r_a. This result is consistent to the horizontal dashed line in panel (b), Figure 1.

Similarly, when S_a is given, equation (A.1) could also be represented as

$$\frac{N}{S_a}\int_0^{\min(S_a,S)}\int_0^{S-a}\left[1-G(x)\right]f(x)dxda - Ea = 0,$$

and no matter which one of S_a and S is smaller, N decreases when S increases. The reason is the integrand is always positive so that integration is an increasing function of S, and thus, N is a decreasing function of S. Here we only show a single arm case, but multiple arms will not change the relationship. So we have proved that S is monotonically decreasing function with respect to N once either S_a or r_a is observed.

B Relationship between N-S curves and r_a/S_a

In this section, we would like to provide the mathematical proof for what have been shown in figure 1. First consider the case that r_a is given, following equation (**??**) and (**??**), we obtained

$$E_a = \begin{cases} \sum_{j=0}^{1} p_j \cdot \frac{\lambda_j}{\lambda_j+\eta_j}\left[N - \frac{e^{-(\lambda_j+\eta_j)S}}{\lambda_j+\eta_j}\left(e^{(\lambda_j+\eta_j)\frac{N}{r_a}}-1\right)\cdot r_a\right] & \frac{N}{r_a} \leq S \\ \sum_{j=0}^{1} p_j \cdot \frac{\lambda_j}{\lambda_j+\eta_j}\cdot r_a\left[S + \frac{1}{\lambda_j+\eta_j}e^{-(\lambda_j+\eta_j)S} - \frac{1}{\lambda_j+\eta_j}\right] & \frac{N}{r_a} > S \end{cases}.$$

Similar to equation (E.1), $\left(\exp\left\{(\lambda_j+\eta_j)\frac{N}{r_a}\right\}-1\right)\cdot r_a$ is a monotonous decreasing function with respect to r_a and thus, for the case $N/r_a \leq S$, when N is fixed, S is negatively related to r_a. In the other case, N no longer appears in the equation and hence, r_a only related to S in a negative way since $S+\exp\left\{-(\lambda_j+\eta_j)S\right\}/(\lambda_j+\eta_j)$ is a monotonously increasing function to S. For both case, an increase in r_a results in a decrease in S for every single N, which is equivalent to a overall moving downwards in the curve.

For the case that S_a is provided, according to equation (**??**), it is trivial that for any fixed S, sample size N has to increase as S_a increases, regardless the relationship between S_a and S, which justifies a right shift of the curve for a larger S_a as shown in the figure.

C Calculating the probability of clinical meaningful treatment effect

Denote the observed data from one arm is $(U_i, \delta_i), i = 1, 2, ..., N_j$, where

$$\begin{aligned} T_i &\sim exp(\lambda_j), \\ C_i &\sim exp(\eta_j), \\ A_i &\sim U(0, S_a), \\ U_i &= \min\{T_i, C_i, max(0, S - A_i)\}, \\ \delta_i &= 1[U_i = T_i], \end{aligned}$$

and j represents the j^{th} arm, and for simplicity, we do not add j on the observed data. Then we can use the following likelihood function to make inference on λ (due to right non-informative censoring)

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

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

$$\ell(\lambda_j; U_i, \delta_i) = log(\lambda_j)\delta_i - \lambda_j U_i.$$

Thus, the Fisher's information is

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

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

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

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

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,

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

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

D Objective function when $P(A)$ is specified

Here we check the property of the objective function after including the probability of meaningful treatment effect. For a fixed value of revenue

$$Rev(N, S) = Pr(A|N, S)(1-\beta)\cdot b\cdot(l - S - l_0) - (c_0 + c_1 N + c_2 S) \equiv C,$$

take derivatives with respect to N on both sides which results in

$$\frac{dPr(A|N,S)}{dN}(1-\beta)\cdot b\cdot(l-S-l_0) - Pr(A|N,S)(1-\beta)\cdot b\cdot\frac{dS}{dN} - c_1 - c_2\frac{dS}{dN} = 0$$

$$\Rightarrow \frac{dS}{dN} = \frac{\frac{dPr(A|N,S)}{dN}(1-\beta)\cdot b\cdot(l-S-l_0) - c_1}{Pr(A|N,S)(1-\beta)\cdot b + c_2}$$

$$
\begin{aligned}
\Rightarrow \frac{d^2S}{dN^2} &= \frac{\left[\frac{d^2Pr(A|N,S)}{dN^2}(1-\beta)\cdot b\cdot(l-S-l_0) - \frac{dPr(A|N,S)}{dN}(1-\beta)\cdot b\frac{dS}{dN}\right]}{(Pr(A|N,S)(1-\beta)\cdot b+c_2)^2} \\
&\quad \cdot(Pr(A|N,S)(1-\beta)\cdot b+c_2) \\
&\quad -\frac{\left[\frac{dPr(A|N,S)}{dN}(1-\beta)\cdot b\cdot(l-S-l_0)-c_1\right](1-\beta)\cdot b\cdot\frac{dPr(A|N,S)}{dN}}{(Pr(A|N,S)(1-\beta)\cdot b+c_2)^2} \\
&= \frac{\frac{d^2Pr(A|N,S)}{dN^2}(1-\beta)\cdot b\cdot(l-S-l_0)\cdot(Pr(A|N,S)(1-\beta)\cdot b+c_2)}{(Pr(A|N,S)(1-\beta)\cdot b+c_2)^2} \\
&\quad -\frac{2\left[\frac{dPr(A|N,S)}{dN}(1-\beta)\cdot b\cdot(l-S-l_0)-c_1\right](1-\beta)\cdot b\cdot\frac{dPr(A|N,S)}{dN}}{(Pr(A|N,S)(1-\beta)\cdot b+c_2)^2}.
\end{aligned}
$$

Along with the fact that

$$
Pr(A_1|N,S) = \Phi\left(\frac{\log 2\left(\frac{1}{\lambda_1}-\frac{1}{\lambda_0}\right)-d_0}{\log 2\sqrt{\frac{1}{\lambda_1^2 p_1 NPr(\delta^{(1)}=1)}+\frac{1}{\lambda_0^2 p_0 NPr(\delta^{(0)}=1)}}}\right) = \Phi\left(\sqrt{N}\cdot h(S)\right)
$$

$$
\frac{dPr(A_1|N,S)}{dN} \propto \frac{1}{\sqrt{N}}\varphi\left(\sqrt{N}\cdot h(S)\right) \to 0 \text{ when N is large,}
$$

where Φ and φ are the CDF and PDF of a standard normal distribution respectively, we obtain that

$$
\frac{d^2S}{dN^2} \approx 0.
$$

Previous steps also hold for $Pr(A_2)$ with similar proof. Hence, due to the second order derivatives of S to N is approximately 0, the objective function could still be treated as a straight line which helps to ease the optimization process.

E Demonstration of optimal solution under different r_a

For better illustration, we selected two scenarios with $r_a = 32, 42$, respectively, to make a wider gap between two feasible curves. Figure 4 shows that when $r_a = 42$, there exists multiple combinations of (N, S) that has both smaller N and S, which is highlighted in orange, but the optimal set is a set that has slightly larger N than the optimal set when $r_a = 32$. By contour plot, this is straightforward as the orange region is bounded by two contours of the objective function, their objective values must also be bounded between. Or it could be shown analytically by checking the slope of the feasible curve with different r_a but fixed N. According to equation A.2 and

A.3, we could derive the expression for the slope of feasible curve

$$\frac{\partial S}{\partial N} = -\frac{1}{\lambda+\eta}\left(\frac{1}{N-\frac{\lambda+\eta}{\lambda}E_a} - \frac{\lambda+\eta}{r_a\left(1-e^{-\frac{\lambda+\eta}{r_a}N}\right)}\right),$$

and this is a decreasing function with respect to r_a for fixed N since

$$\frac{d\left[r_a\left(1-e^{-\frac{\lambda+\eta}{r_a}N}\right)\right]}{dr_a} = 1-(1+c)e^{-c} > 0, \tag{E.1}$$

where $c = (\lambda+\eta)N/r_a$. Since the feasible curve is a convex function, which means the second derivative is positive or equivalently, slope is an increasing function to N, enlarging N ensures a bigger slope value. Therefore, the optimal set on the feasible curve with larger r_a has to have a larger N to maintain the same slope.

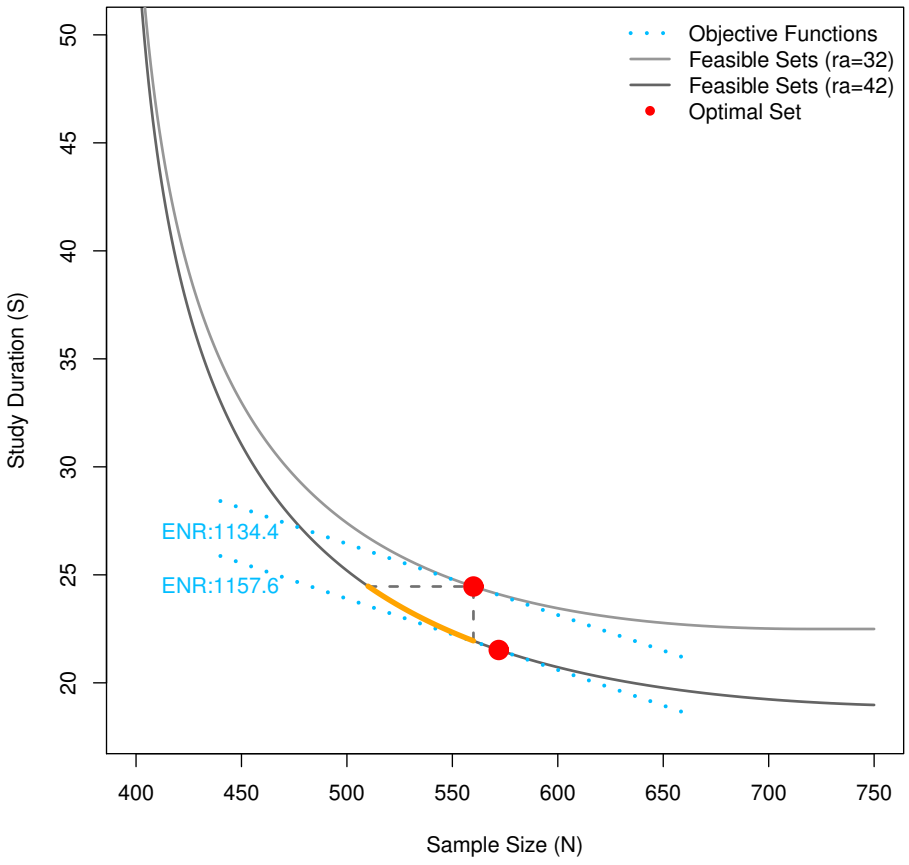

Figure 4: Two feasible curves for $r_a = 32, 42$ and the corresponding optimal sets as well as objective functions. The sets on the feasible curve which $r_a = 42$ and have smaller N and S than the optimal solution on curve with $r_a = 32$ is coded in orange.