

RESEARCH ARTICLE

Optimal two-stage design of single arm Phase II clinical trials based on median event time test

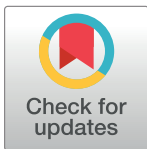
Yeonhee Park  *

Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, United States of America

* ypark56@wisc.edu

Abstract

The Phase II clinical trials aim to assess the therapeutic efficacy of a new drug. The therapeutic efficacy has been often quantified by response rate such as overall response rate or survival probability in the Phase II setting. However, there is a strong desire to use survival time, which is the gold standard endpoint for the confirmatory Phase III study, when investigators set the primary objective of the Phase II study and test hypotheses based on the median survivals. We propose a method for median event time test to provide the sample size calculation and decision rule of testing. The decision rule is simple and straightforward in that it compares the observed median event time to the identified threshold. Moreover, it is extended to optimal two-stage design for practice, which extends the idea of Simon's optimal two-stage design for survival endpoint. We investigate the performance of the proposed methods through simulation studies. The proposed methods are applied to redesign a trial based on median event time for trial illustration, and practical strategies are given for application of proposed methods.



OPEN ACCESS

Citation: Park Y (2021) Optimal two-stage design of single arm Phase II clinical trials based on median event time test. PLoS ONE 16(2): e0246448. <https://doi.org/10.1371/journal.pone.0246448>

Editor: Alan D Hutson, Roswell Park Cancer Institute, UNITED STATES

Received: September 7, 2020

Accepted: January 19, 2021

Published: February 8, 2021

Copyright: © 2021 Yeonhee Park. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting information](#) files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The primary objective of Phase II clinical trials is to test whether the therapeutic intervention actually works in treating a disease or indication. The therapeutic efficacy has been often quantified by response rate in Phase II settings, for example, overall response rate defined as the proportion of subjects who achieve a confirmed complete response or partial response determined by RECIST 1.1 [1] or survival probability at a certain year defined as the proportion of patients alive (and without recurrence) at certain year after the start of treatment. In practice, Simon's two-stage minimax and optimal designs are widely used for binary endpoint [2–7]. Simon's designs allow early trial termination due to futility while Fleming's design allows early trial termination due to futility and superiority in Phase II trials [8, 9]. Successful results in Phase II trials lead to proceed to confirmatory Phase III trials with more extensive development. Therefore, it is necessary to validate and use the surrogate endpoint for overall survival, which is the gold standard endpoint for Phase III study. When considering the survival

probability which enforces the survival time to the binary endpoint, we need much caution to analyze the results. Some designs address incomplete follow-up for some subjects by the time of the interim analysis to assess the survival probability [10–13]. Data augmentation can be used to address the timing of events for late-onset outcomes by imputing missing outcomes. [14, 15].

Surprisingly, a recent FDA report shows 22 case studies investigating disagreement in the results between early and confirmatory phases [16]. Some cases showed unexpected failures in Phase III study from the promising results on clinical outcomes in the Phase II study (e.g., *ini-parib*). Therefore, some investigators strongly desire to use time-to-event endpoint to evaluate the therapeutic efficacy of the drug for Phase II trials. They are interested in improvement of median survival compared to standard therapy. Finkelstein et al. [17], Sun et al. [18], Jung [19], Kwak and Jung [20], Wu et al. [21] propose designs with one-sample log-rank test which compares the survival distributions between prespecified null reference and the desired target. This idea comparing the survival distributions allows to conduct for hypothesis testing for the median survivals only when the survivals are assumed to follow exponential distribution. Wu [22] proposes a single-arm Phase II clinical trial design under a class of parametric cure models. Chu et al. [23] proposes the design for immunotherapy trials with random delayed treatment effect based on survival time.

In this paper, we propose new methods for a single arm Phase II study, which provide practical strategies for testing if there is an improvement of median survivals from a new drug compared to the standard therapy. The proposed median event time test uses the distribution of sample median to obtain the required sample size and threshold for the hypothesis testing at target type I and II errors. It provides the decision rule, which is very simple and straightforward, comparing the observed median survival with the identified threshold at the end of trial. This can be easily implemented with our shiny application and R codes built along with the method development. Moreover, this median event time test is extended in group sequential manner. It follows the idea of Simon's optimal two-stage design, in that the expected sample size is minimized under the null hypothesis, but considers survival endpoint to see the improvement based on the median event time test. This allows us to monitor futility of the drug based on median survival time and stop the trial early, which makes the design more efficient and practical.

The rest of the paper is organized as follows. In Methods Section, we provide a median event time test and propose an optimal two-stage design based on median event time test. In Results Section, simulation studies are presented, and a trial example is provided to illustrate the application of our methods. Lastly, we provide some comments in Discussion and concluding remarks in Conclusion.

Methods

Median event time test

Let Y_1, \dots, Y_n be random variables from an exponential distribution with mean μ . Then, the median ϕ is equivalent to $\mu \log 2$. Motivated by single arm Phase II studies whose primary endpoint is time-to-event, we formulate a hypothesis test based on median event time with $H_0: \phi = \phi_0$ versus $H_a: \phi = \phi_1$ for some values of ϕ_0 and ϕ_1 . In practice for intervention study, we use median time of standard drug or therapy and the expected median time from the intervention (for improvement) to specify the values of ϕ_0 and ϕ_1 , respectively. Let $\hat{\phi}_n(\mathbf{y})$ be a sample median event time obtained from a sample of size n . Then, for any $\lambda \geq 0$, the rejection region for the test is $\{\mathbf{y} : \hat{\phi}_n(\mathbf{y}) > \lambda\}$. For any α and β in unit interval $(0, 1)$, this test is the level α test

with power $1 - \beta$ such that $\Pr\{\hat{\phi}_n(\mathbf{y}) > \lambda|H_0\} = \alpha$ and $\Pr\{\hat{\phi}_n(\mathbf{y}) \leq \lambda|H_a\} = \beta$. Since the first conditional probability indicates the probability of rejecting the null hypothesis when the null hypothesis is true, the value of α indicates the type I error rate. Also, the second conditional probability indicates the probability of failure of rejecting null hypothesis when the null hypothesis is not true, and the value of β indicates the type II error rate. The distribution function of the sample median event time is derived in the following theoretical result.

Theorem 1 Let Y_1, \dots, Y_n be random variables from an exponential distribution with median ϕ . Then, the probability density function (pdf) of sample median of Y_1, \dots, Y_n is either

$$g_n(m) = n! \log 2 \{1 - \exp(-m \log 2 / \phi)\}^k \exp\{-m(k + 1) \log 2 / \phi\} / (k!k! \phi)$$

for $n = 2k + 1$ or

$$g_n(m) = \int_0^{2m} n! (\log 2)^2 \exp(-2m(\log 2) / \phi) \{1 - \exp\{-(\log 2)(m - r/2) / \phi\}\}^{k-1} \times \exp\{-(\log 2)(m + r/2)(n - k - 1) / \phi\} / \{(k - 1)!(n - k - 1)! \phi^2\} dr$$

for $n = 2k$.

The proof is in Appendix A. Theorem 1 provides a cumulative distribution function (cdf) of median event time given by $G_n(\lambda) = \int_0^\lambda g_n(m) dm$, where g_n denotes the pdf of the sample median, for $\lambda \geq 0$. Since the cdf of sample median event time has no closed form, a numerical search over grid is required to identify sample size n and threshold λ such that the empirical errors $\hat{\alpha}(n, \lambda)$ and $\hat{\beta}(n, \lambda)$ are close to the nominal target values of type I and II error rate (i.e., α and β), respectively. The empirical errors are calculated as

$$\hat{\alpha}(n, \lambda) = \Pr\{\hat{\phi}_n > \lambda|H_0\} = 1 - G_n(\lambda|H_0) \quad \text{and} \quad \hat{\beta}(n, \lambda) = \Pr\{\hat{\phi}_n \leq \lambda|H_a\} = G_n(\lambda|H_a),$$

where $G_n(\cdot|H_0)$ and $G_n(\cdot|H_a)$ denote the cdf of sample median of event time under null and alternative hypotheses, respectively. It implies that the identified sample size n justifies to achieve $\{1 - \hat{\beta}(n, \lambda)\} * 100\%$ power based on one-sided test with a significance level of $\hat{\alpha}(n, \lambda)$ to detect improvement of the sample median time of the experimental drug against the median time of standard drug. This test states that at the end of trial, i.e., based on sample of size n , the null hypothesis is rejected if $\hat{\phi}_n > \lambda$ and we argue that experimental drug increases in median event time from ϕ_0 to ϕ_1 . We call this median event time test.

Let's consider a clinical trial with hypothesis testing of median progression free survivals (PFS) $\phi_0 = 10$ months versus $\phi_1 = 17$ months. Suppose that maximum number of patients for this study is 100. We consider a grid search over the integer n between 1 and 100 and a real number λ between 10 and 17 with the increment 0.1. For the target error rates of $\alpha = 0.05$ and $\beta = 0.2$, our numerical study results in $n = 42$ and $\lambda = 14.1$ months minimizing deviation of the empirical errors from the target rates, i.e., $\{\hat{\alpha}(n, \lambda) - \alpha\}^2 + \{\hat{\beta}(n, \lambda) - \beta\}^2$. Fig 1 shows results for numerical search of n and λ , and Table 1 summarizes the decision rule for testing median survival ϕ_0 versus ϕ_1 . In Table 1, scenarios with different value of ϕ_0 and ϕ_1 are considered. The choice of ϕ_0 and ϕ_1 is determined intentionally. To see the performance of method, we fixed ϕ_0 or ϕ_1 to vary ϕ_1 or ϕ_0 , respectively.

We have proposed median event time test for exponential survivals. The proposed method is extended for other parametric survival distributions. We provide the distribution function of the sample median event time for uniform and weibull survivals in Appendix B, which determines the decision rule.

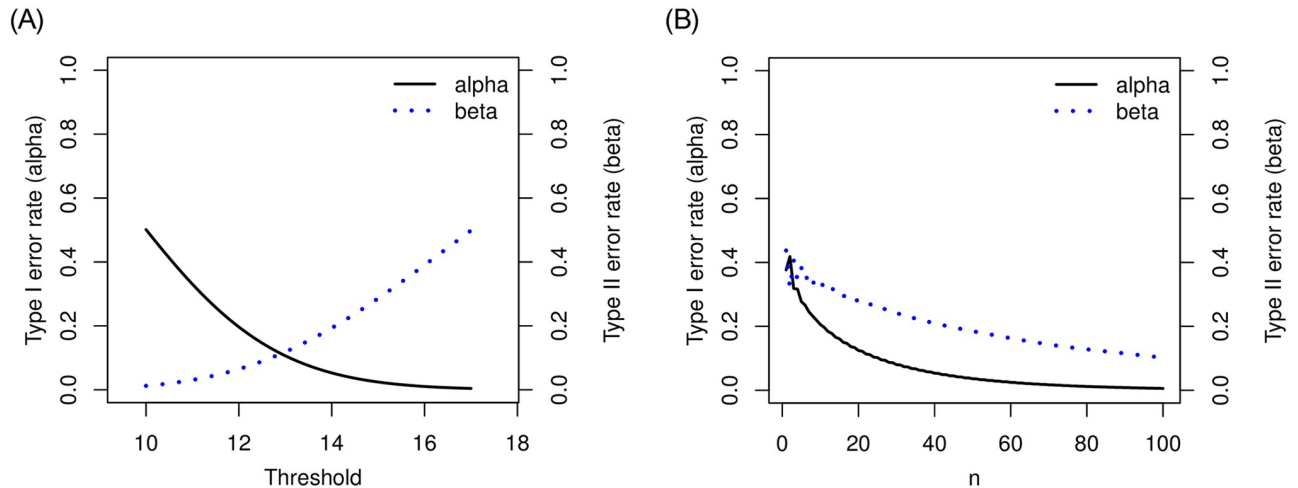


Fig 1. Plot of $\hat{\alpha}(n, \lambda)$ and $\hat{\beta}(n, \lambda)$ for median event time test with $\phi_0 = 10$ and $\phi_1 = 17$ months: Left panel shows the result when $n = 42$ and right panel shows the results when $\lambda = 14.1$ months.

<https://doi.org/10.1371/journal.pone.0246448.g001>

This proposed median event time test is attractive with three reasons. First, this is newly proposed for an exact median event time test using the observed median event time. Second, author provides a software to calculate sample size and identify the threshold for the test, which is open for public use (<https://yeonhee.shinyapps.io/METTshinyapp/>). It does not require complicated statistical analysis for decision but implements easily. The simple and straightforwardly interpretable rule can save lots of time for drug development. Third, it has potential to extend for many applications. For example, the decision rule can be used to monitor futility in group sequential designs (for Phase II or III studies), and the threshold λ provides a good candidate (or range) for the Bayesian monitoring rule. Specifically, it provides foundation for optimal two-stage design based on median event time test which is described later.

Before we move to next section, we provide a remark on the median event time test. To develop the exact median event time test, we considered hypothesized values for median of survivals $Y_i = \min(S_i, U_i)$, where S_i denote the time-to-event and U_i denote the administrative censoring time for the i th patient. Our setting in this section uses Y_i and does not require information for accrual rate and follow-up time. However, when we consider a hypothesis testing with median of survivals S_i , we should care the censoring information. Assume that the time to arrival of the patient and survival time (i.e., S_i) follow some exponential distributions. Then, U_i , which is the time to arrival of the last patient plus follow-up time minus time to arrival of the i th patient, follows another exponential distribution. We notice that minimum of

Table 1. Decision rules for hypothesis testing of median PFS ϕ_0 months versus ϕ_1 months at target error rates of $\alpha = 0.05$ and $\beta = 0.2$. Numerical search is done over the integer n between 1 and 100 and a real number λ between ϕ_0 and ϕ_1 with the increment 0.1.

ϕ_0	ϕ_1	n	λ	$\hat{\alpha}(n, \lambda)$	$\hat{\beta}(n, \lambda)$
10	17	42	14.1	0.049	0.2019
8	17	21	12.9	0.049	0.1974
8	14	38	11.5	0.048	0.2016
3	7	16	5.2	0.042	0.2007
3	6	24	4.7	0.047	0.2027
3	5	48	4.2	0.042	0.2030

<https://doi.org/10.1371/journal.pone.0246448.t001>

two exponential random variables (i.e., Y_i) follow some exponential distribution. Therefore, this setting for hypothesis testing with Y_i looks reasonable with the exponential survival assumption for S_i and the right censoring.

Optimal two-stage design based on median event time test

The proposed median event time is straightforward for clinicians to interpret and justify the sample size for the exact level of errors. It is critical, especially in rare disease trials, to obtain promising evidence with the target error rates and minimize the expected sample size. Moreover, from an ethical and practical viewpoint, it is desirable to stop the trial early if the therapeutic intervention is not effective. This motivates to propose a single arm Phase II trial with an interim planned when the total accrual reached n_1 patients. The final analysis will be performed after the follow-up of all planned number of n patients. A two-stage design using median event time test is proposed as follows. In the first stage (i.e., at interim), we determine go/no-go of the trial based on the observed median event time for those n_1 patients. When the observed median event time is less than or equal to the threshold λ , the trial is stopped for futility. Otherwise, the trial continues to enroll $(n - n_1)$ patients in the second stage. At final analysis, we argue that the drug is sufficiently promising to evaluate against the standard therapy if the observed median event time based on all trial data is larger than the threshold.

The expected sample size for the two-stage design above is $EN = n_1 + (1 - \text{PET})(n - n_1)$, where PET denotes the probability of early termination after the first stage. EN depends on n , n_1 and λ , and the optimal two-stage design using median event time test is proposed in that EN is minimized under the null hypothesis. In the following, we describe how to specify n , n_1 and λ for the optimal two-stage design. Let M be a maximum sample size for the study, e.g., this can be specified by clinicians or by sample size calculator for one-arm nonparametric statistics (<https://stattools.crab.org/Calculators/oneNonParametricSurvival.htm>). Let ϵ_1 and ϵ_2 denote the acceptable difference between the estimated type I and II error rates from the trial design and the target error rates. For each prespecified values of ϕ_0 , ϕ_1 , α and β ,

Step 1. For each n_1 between 1 and $M - 1$, search n and λ minimizing

$$\{\hat{\alpha}(n_1, n, \lambda) - \alpha\}^2 + \{\hat{\beta}(n_1, n, \lambda) - \beta\}^2, \text{ where}$$

$$\hat{\alpha}(n_1, n, \lambda) = \Pr(\hat{\phi}_{n_1} > \lambda, \hat{\phi}_n > \lambda | H_0) \quad (1)$$

$$\hat{\beta}(n_1, n, \lambda) = \Pr(\hat{\phi}_{n_1} \leq \lambda | H_a) + \Pr(\hat{\phi}_{n_1} > \lambda, \hat{\phi}_n \leq \lambda | H_a), \quad (2)$$

over $n_1 < n \leq M$ and $\phi_0 \leq \lambda \leq \phi_1$.

Step 2. Choose an optimal pair of (n, n_1, λ) minimizing EN among the pairs satisfying

$$|\hat{\alpha}(n_1, n, \lambda) - \alpha| < \epsilon_1 \text{ and } |\hat{\beta}(n_1, n, \lambda) - \beta| < \epsilon_2.$$

As seen in median event time test, we don't have closed forms for the marginal or joint distributions of sample median event times to calculate empirical probabilities in (1) and (2). The numerical search is used in the first step. The criteria given in the second step targets at minimizing the expected sample size, EN. This optimality criteria can be modified according to the study objective, e.g., the design minimizes expected total study length. Moreover, our design is flexible to use different thresholds, λ_1 for the interim analysis and λ_2 for the final analysis. Investigators fix a threshold λ_1 to stop early for futility and search threshold λ_2 satisfying the

target error rates (or they can search both thresholds):

$$\begin{aligned}\hat{\alpha}(n_1, n, \lambda_1, \lambda_2) &= \Pr(\hat{\phi}_{n_1} > \lambda_1, \hat{\phi}_n > \lambda_2 | H_0) \\ \hat{\beta}(n_1, n, \lambda_1, \lambda_2) &= \Pr(\hat{\phi}_{n_1} \leq \lambda_1 | H_a) + \Pr(\hat{\phi}_{n_1} > \lambda_1, \hat{\phi}_n \leq \lambda_2 | H_a) \\ |\hat{\alpha}(n_1, n, \lambda_1, \lambda_2) - \alpha| &< \epsilon_1 \text{ and } |\hat{\beta}(n_1, n, \lambda_1, \lambda_2) - \beta| < \epsilon_2.\end{aligned}$$

Although the proposed method does not assume specific survival distribution, the probability calculation requires to specify the survival distribution. We can borrow information from the previous research to specify the survival distribution (e.g., exponential, uniform, or weibull distribution). As an example to illustrate the optimal two-stage design based on median event time test, we assume that the survivals follow from an exponential distribution with median ϕ and will be right-censored for subjects who have not yet met the criteria at the date of the last valid disease assessment. Setting with median survivals $\phi_0 = 10$ and $\phi_1 = 17$ for the standard therapy and a new therapy, respectively, and target error rates $\alpha = 0.05$ and $\beta = 0.2$, we consider the maximum allowable sample size $M = 50$ (which is obtained from a sample size calculator for one-arm nonparametric statistics). Patients arrived according to a Poisson process with the accrual rate of 1.04 patients per month. We continue follow-up for 24 months after the last patient was enrolled. Empirical estimate of type I and II errors were obtained based on 1000 simulation trials. Then, our method with $\epsilon_1 = 0.02$ and $\epsilon_2 = 0.025$ attains the optimal results at $n_1 = 32$, $n = 42$ and $\lambda = 15.2$. It implies that our two-stage design accrue $n_1 = 32$ patient for the first stage. At interim, if the observed median event time based on these n_1 patients is less than or equal to $\lambda = 15.2$ months, the study will be early stopped for futility. Otherwise, additional $n - n_1 = 10$ patients will be accrued in stage 2, resulting in a total sample size of $n = 42$. At the end of trial, if the observed median event time based on all $n = 42$ patients is larger than $\lambda = 15.2$ months, we reject the null hypothesis and claim that the treatment is sufficiently promising. Different null and alternative median event times can be considered, and the results of the optimal two-stage design are summarized in [Table 2](#).

The proposed optimal two-stage design finds, in the first stage, the total sample size n and threshold λ for the given interim size n_1 and target error rates (i.e., α and β). In other words, both n and λ are searched. In case where study has a certain planned total sample size, the proposed design is tailored to identify the threshold λ in the first stage for the given interim size n_1 , total sample size, and target error rates. In other words, only λ is searched, and n is prespecified. As an example, the value of M obtained from one-arm nonparametric statistics in [Table 2](#) is used for the prespecified total sample size n in scenarios. The results are also summarized in [Table 2](#).

Replacing $M = 100$ in [Table 1](#) with the value obtained from one-arm nonparametric statistics, we obtained the same results. [Tables 1 and 2](#) show that decision rule of two-stage design yields smaller expected sample size under the null than the one-stage design except the case with $\phi_0 = 8$ and $\phi_1 = 17$ requires 2 or 3 more patients to be enrolled. When $\phi_0 = 3$ and $\phi_1 = 5$, the total sample size n for two-stage design is smaller than one-stage design, and the expected sample size under the null is much smaller. More patients can avoid from treating the ineffective drug under the two-stage design.

Results

We investigated the performance of the proposed optimal two-stage design based on median event time test. First, back to the setting we examined for [Table 2](#) with null median 10 months and alternative median 17 months, we are interested how the operating characteristics are changed with follow-up time and accrual rate. Given the specified rule from [Table 2](#)

Table 2. Summary results of optimal two-stage design for hypothesis testing of median survivals ϕ_0 versus ϕ_1 when survivals are assumed to follow an exponential distribution. Note M denotes the maximum sample size for the study calculated from one-arm nonparametric statistics. Notations such as EN_0 and PET_0 are used to denote EN and PET, respectively, under the null hypothesis.

			n is searched						
ϕ_0	ϕ_1	M	n_1	n	λ	$\hat{\alpha}$	$\hat{\beta}$	EN_0	PET_0
10	17	50	32	42	15.2	0.06	0.222	32.6	0.94
8	17	26	22	26	12.1	0.065	0.212	22.26	0.94
8	14	45	29	37	12.6	0.067	0.199	29.54	0.93
3	7	21	12	21	5.1	0.063	0.204	12.57	0.94
3	6	31	13	30	5	0.062	0.225	14.05	0.94
3	5	54	17	44	4.9	0.067	0.221	18.81	0.93
			$n = M$ is prespecified						
ϕ_0	ϕ_1	M	n_1	n	λ	$\hat{\alpha}$	$\hat{\beta}$	EN_0	PET_0
10	17	50	34	42	15.3	0.062	0.21	34.99	0.94
8	17	26	23	26	12.4	0.066	0.207	23.30	0.93
8	14	45	28	37	12.4	0.062	0.224	29.05	0.94
3	7	21	12	21	5.1	0.062	0.198	12.56	0.94
3	6	31	13	30	5	0.061	0.223	14.10	0.94
3	5	54	17	44	4.9	0.063	0.225	19.33	0.94

<https://doi.org/10.1371/journal.pone.0246448.t002>

(i.e., $n_1 = 32, n = 42$ and $\lambda = 15.2$), we considered follow-up times 6, 9, 12, 15, 18, 21, 24, 27, 30, 36 months and accrual rates 0.5, 0.6, 0.7, 0.8, 0.9, 1.04, 1.1, 1.2, 1.3, 1.5 patients per month. The results are described in Fig 2. The proposed design is robust to the follow-up time but impacted by the accrual rate. As more patients are accrued and we have more available events (i.e., the observed median survival is less), it is more likely to stop earlier due to futility, which increases type II error rate but decrease type I error rate. Therefore, when the study is designed by using the optimal two-stage design based on median event time test, we need to have more reliable information for accrual rate, for example, the previous history for accrual rate and close collaboration with clinicians investigating the study can provide the right design.

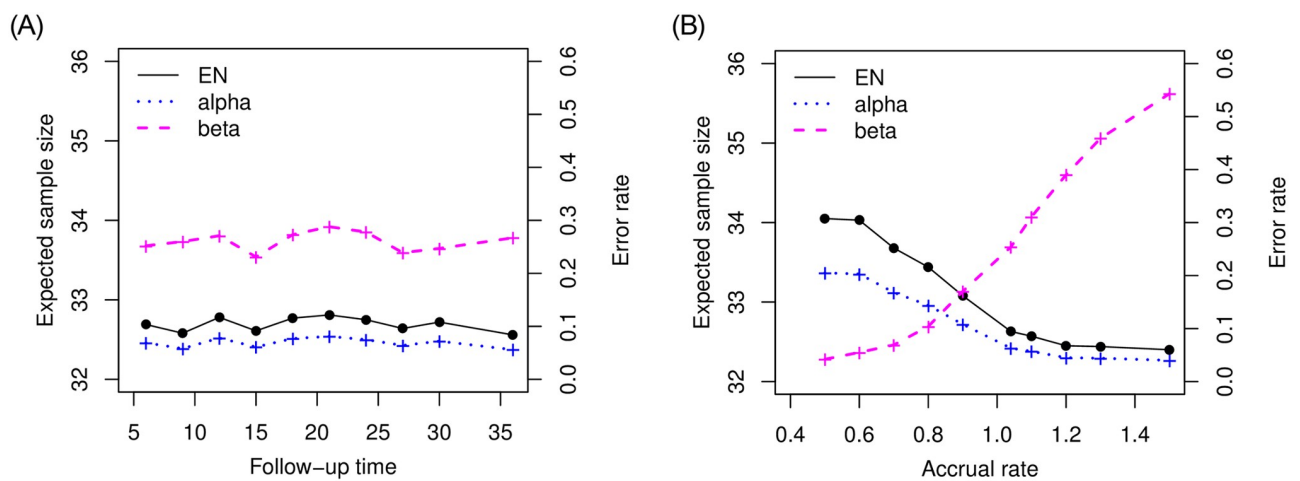


Fig 2. Plot of operating characteristics (i.e., expected sample size and error rates) in follow-up time and accrual rate when testing null median 10 months versus alternative median 17 months.

<https://doi.org/10.1371/journal.pone.0246448.g002>

Table 3. Summary results of optimal two-stage design for hypothesis testing of median survivals ϕ_0 versus ϕ_1 when survivals are assumed to follow non-exponential distributions (i.e., uniform or weibull distribution). Note M denotes the maximum sample size for the study calculated from one-arm nonparametric statistics, and EN_0 denotes EN under the null hypothesis.

				<i>n</i> is searched						<i>n</i> = <i>M</i> is prespecified				
	ϕ_0	ϕ_1	<i>M</i>	n_1	<i>n</i>	λ	$\hat{\alpha}$	$\hat{\beta}$	EN_0	n_1	λ	$\hat{\alpha}$	$\hat{\beta}$	EN_0
Uniform	10	17	50	17	50	10.9	0.052	0.211	18.72	18	11	0.061	0.197	19.95
	8	17	26	14	26	9.2	0.049	0.203	14.59	14	9.2	0.059	0.213	14.71
	8	14	45	14	42	8.7	0.064	0.206	15.79	14	8.8	0.054	0.209	15.67
	3	7	21	7	20	3.5	0.049	0.197	7.64	7	3.5	0.051	0.194	7.71
	3	6	31	7	30	3.4	0.053	0.204	8.22	7	3.4	0.053	0.2	8.27
	3	5	54	7	46	3.3	0.057	0.225	9.22	8	3.3	0.06	0.196	10.76
Weibull	10	17	50	18	21	12.1	0.047	0.205	18.14	16	11.2	0.067	0.224	18.28
	8	17	26	13	23	9.4	0.051	0.208	13.51	14	9.6	0.032	0.194	14.38
	8	14	45	14	24	9.4	0.052	0.201	14.52	14	9.1	0.067	0.199	16.08
	3	7	21	6	21	3.6	0.043	0.217	6.65	6	3.5	0.061	0.209	6.92
	3	6	31	7	16	3.7	0.052	0.196	7.47	7	3.6	0.046	0.196	8.10
	3	5	54	8	16	3.7	0.054	0.199	8.43	7	3.5	0.043	0.205	9.02

<https://doi.org/10.1371/journal.pone.0246448.t003>

Moreover, we examined the performance of the proposed design with non-exponential survivals. Table 3 provides the results when survivals are generated from either uniform or weibull distributions. We used uniform distribution with minimum value 0 and maximum value $2\phi_0$ to generate flat survivals and Weibull distribution with scale parameter $\phi_0/(\log 2)^{1/2}$ and shape parameters 2 to obtain an increasing hazard. Compared with the results in Table 2 where survivals are assumed to follow exponential distribution, we can see the different decision rule (n_1 and λ) and required sample size (n) for nonexponential survival times.

We compared the proposed two-stage designs (called METT2E, METT2U, and METT2W for optimal two-stage designs assuming exponential, uniform, and weibull survivals, respectively) with three designs: (1) a restricted KJ design, called r-KJ [12], which tests for whole survival curves based on one-sample log-rank test statistics proposed by Kwak and Jung [20]; (2) a two-stage design minimizing expected sample size, called OES [11]; (3) a two-stage design minimizing expected total study length, called OETSL [11]. Both OES and OETSL use the normalized Z-statistic to test and determine decision rules at each stage. Because r-KJ, OES, and OETSL require the clinically meaningful time point, in our simulations we used 6 months. The null and alternative values of the 6 months survival probability were determined by the survival distribution. We assumed the follow-up time is 24 months and accrual rate is 1.04 patients per month, which is the same as the setting of Tables 2 and 3. Table 4 provides the comparison results of EN_0 and PET_0 for several hypothesis testings based on median survival times ϕ_0 and ϕ_1 . As seen in Tables 2 and 3, both METT2E and METT2U are likely to enroll more patients in the trial (i.e., n is close to M), compared to METT2W in most cases. The number of patients in the first stage (i.e., n_1) obtained from METT2U or METT2W is smaller than that of METT2E. Thus, METT2W yields smaller expected sample size than METT2E and METT2U. Since r-KJ does not restrict to certain survival distribution and both OES and OETSL assume the weibull survivals, the results of METT2W are comparable with r-KJ, OES, and OETSL. In most cases, METT2W uses smaller expected sample size and stops the trial early for futility when therapeutic intervention is not effective. METT2E and METT2U also yielded smaller expected sample size and larger probability to stop trial for futility under the null hypothesis compared to r-KJ.

Table 4. Comparison results of the proposed designs (METT2E, METT2U, and METT2W) with r-KJ, OES, and OETSL. Note M denotes the maximum sample size for the study calculated from one-arm nonparametric statistics, EN_0 and PET_0 denote EN and PET, respectively, under the null hypothesis.

			EN_0					
ϕ_0	ϕ_1	M	METT2E	METT2U	METT2W	r-KJ	OES	OETSL
10	17	50	32.6	18.72	18.14	72.38	49.08	49.00
8	17	26	22.26	14.59	13.51	34.06	25.74	25.54
8	14	45	29.54	15.79	14.52	54.52	31.52	31.38
3	7	21	12.57	7.64	6.65	12.83	7.00	6.00*
3	6	31	14.05	8.22	7.47	17.58	8.41	6.90
3	5	54	18.81	9.22	8.43	29.71	10.60	9.77
			PET_0					
ϕ_0	ϕ_1	M	METT2E	METT2U	METT2W	r-KJ	OES	OETSL
10	17	50	0.94	0.95	0.95	0.57	0.64	0.61
8	17	26	0.94	0.95	0.95	0.56	0.69	0.62
8	14	45	0.93	0.94	0.95	0.59	0.69	0.63
3	7	21	0.94	0.95	0.96	0.54	0.76	0.94*
3	6	31	0.94	0.95	0.95	0.54	0.76	0.18
3	5	54	0.93	0.94	0.95	0.56	0.72	0.40

* Clinical time point of 5.7 months is used.

<https://doi.org/10.1371/journal.pone.0246448.t004>

Application: Trial illustration

We provide an application of the proposed methods with a trial NCT00780494. This is a Phase II, single arm, single-institution study of bevacizumab in combination with carboplatin and capecitabine for patients with unresectable or metastatic gastroesophageal junction or gastric cancers. The study was started on February 2009 and completed on December 2017 to accrue enrollment of 35 participants. It enrolled two patients per month and follow up at time of study completion for 12 months. The study objective is to investigate the efficacy of the addition of bevacizumab to standard chemotherapy based on progression-free survival (PFS), which is defined as the duration of time from the start of treatment to time of disease progression or death. The median PFS is 5 months with the standard treatment therapy and the study team hypothesized the addition of bevacizumab to standard chemotherapy improve PFS by 90%, i.e., the median PFS is 9.5 months (https://clinicaltrials.gov/ProvidedDocs/94/NCT00780494/Prot_SAP_000.pdf).

We applied the proposed methods to redesign the trial based on median PFS. Target error rates are $\alpha = 0.05$ and $\beta = 0.2$, and the maximum sample size is 35 obtained from one-arm non-parametric test. First, we considered a case where investigators want a trial without interim monitoring to collect the data with all 35 patients. The exact median event time test was applied, and provided the decision rule with threshold 7.5 months for a sample of size 29. This yielded 80.48% power. Second, we supposed that investigators want an interim for futility monitoring. Assuming the exponential survivals, we applied optimal two-stage design based on median event time test. We set $\epsilon_1 = \epsilon_2 = 0.005$ to get closer empirical error rates to the target rates, and it determined $n_1 = 33$, $n = 34$ and $\lambda = 8$ months with $\hat{\alpha} = 0.047$ and $\hat{\beta} = 0.196$. The rule implies that an additional patient will be enrolled after the first stage, and it is inappropriate from a practical point of view. Rather than considering two-stage design with a threshold, two-stage design with different thresholds λ_1 and λ_2 would be suggested to obtain the reasonable decision rule. Table 5 provides the summary results of the optimal two-stage design, which search n_1 , n , and λ_2 for a fixed threshold λ_1 for the first stage. When survivals are

Table 5. Summary results of optimal two-stage design with a fixed threshold λ_1 when a maximum sample size M is 35.

	n_1	n	λ_1	λ_2	$\hat{\alpha}$	$\hat{\beta}$	EN_0	PET_0
Exponential	17	32	5.5	8.5	0.051	0.198	17.77	0.949
	20	29	6	8.7	0.050	0.201	20.45	0.950
Uniform	16	24	5.5	5.8	0.054	0.202	16.43	0.946
	18	31	6	5.6	0.051	0.200	18.66	0.949
Weibull	15	31	5.5	5.9	0.052	0.202	15.83	0.948
	16	34	6	5.7	0.054	0.204	16.97	0.946

<https://doi.org/10.1371/journal.pone.0246448.t005>

assumed to follow the exponential distribution and threshold for the first stage is 5.5 months, the optimal two-stage design based on median event time test determined $n_1 = 17$, $n = 32$, $\lambda_2 = 8.5$ months. Specifically, the interim analysis will be performed based on the first 17 enrolled patients by comparing the observed median PFS with 5.5 months. If the observed median PFS is smaller than or equal to 5.5 months, the study is stopped early for futility. Otherwise, additional 15 patients are enrolled to have the total sample size of 32. At the final analysis, the observed median PFS with 32 patients is compared with the threshold 8.5 months, and we claim the addition of bevacizumab to the standard therapy improves PFS if the observed median PFS is larger than the threshold. This trial decision rule yielded power of 80.2%. When the true median PFS is 5 months, the expected sample size for this trial (i.e., EN) is 17.77 and the probability of early stopping (i.e., PET) is 94%. The required sample size and decision rule are changed for the different choice of λ_1 .

In this trial illustration, PFS was assumed to follow exponential distribution, and the decision rule was identified for the trial under the assumption. We further investigated with the nonexponential assumption. As seen in Table 5, decision rules, especially interim monitoring rules, for exponential versus nonexponential survivals are different, which implies that survival distribution assumption matters to design the clinical trials. In practice, earlier phase trials or experts' knowledge would be able to provide information of survival distribution for Phase II study. We can borrow the information to identify the appropriate decision rule for the study.

Discussion

Most existing methods generally assume exponential survival distribution to develop statistical methods or design based on median survival time for convenience [24, 25]. From our simulation studies, we found that operating characteristics of the design depend on the survival distribution, and the decision rule of median event time test is different according to the survival assumption. Type I or II error rate can be inflated when survival distribution is misspecified. It is critical for median event time test to specify survival distribution for robust clinical trial research. The specification of survival distributions can be determined by the relevant pilot study or historical trials.

We have proposed several designs for median event time test: single stage, two stage with single threshold, and two stage with two different thresholds. According to investigators' interest, study objective, and trial assumption, further simulation investigation may be required to explore more rules. Close collaboration with clinicians as well as statistical practice will guide the better and ethic design for the study.

Conclusion

We proposed methods to test if there is an improvement of median survivals from a new drug compared to the standard therapy. The proposed median event time test provides the required

sample size to control type I and II errors for the hypothesis testing based on the median event time. It also provides a decision rule, which is very simple and straightforward, comparing the observed median survival with the identified threshold by the test. Shiny application and R codes were also built along with the method development so that users can implement easily the hypothesis test based on median event time (<https://sites.google.com/view/yeonheepark/software>). This approach is extended for the trial with interim at which the study monitors futility. The proposed two-stage design based on median event time test is optimal in that the expected sample size is minimized under the null hypothesis. Early stopping for futility enhances ethics in patient care and expedites the discovery of new therapies. Moreover, our methods would reduce unexpected failures in confirmatory phase after the promising results in Phase II study and improve success rate for drug development.

Appendix

A. Proof of theoretical result in methods

Proof of Theorem 1 Let $F(y)$ and $f(y)$ be the cdf and pdf, respectively, of the random variable whose median is ϕ (i.e., mean $\mu = \phi/\log 2$). Then, we have $f(y) = \log 2 \exp(-y \log 2/\phi)/\phi$ and $F(y) = 1 - \exp(-y \log 2/\phi)$. Suppose that we have a sample of size $n = 2k + 1$ for some $k = 0, 1, 2, \dots$. Then, by Cramér [26], the pdf of the sample median is

$$\begin{aligned} g(m) &= \{n!/(k!k!)\}F(m)^k\{1 - F(m)\}^k f(m) \\ &= \{n! \log 2/(k!k!\phi)\}\{1 - \exp(-m \log 2/\phi)\}^k \exp\{-m(k+1) \log 2/\phi\}. \end{aligned}$$

We now derive the pdf of the sample median for a sample of size $n = 2k$ for some $k = 1, 2, \dots$. When n is an even number, the sample median is $\hat{m} = (Y_{(k)} + Y_{(k+1)})/2$, where $Y_{(k)}$ denotes the k th order statistic of the sample. Note that the joint pdf of $Y_{(k)}$ and $Y_{(k+1)}$ is

$$\begin{aligned} f(y_k, y_{k+1}) &= c f(y_k) f(y_{k+1}) F(y_k)^{k-1} \{1 - F(y_{k+1})\}^{n-k-1} \\ &= (c/\mu^2) \exp\{-(y_k + y_{k+1})/\mu\} \{1 - \exp(-y_k/\mu)\}^{k-1} \exp\{-y_{k+1}(n-k-1)/\mu\}, \end{aligned}$$

where $c = n!/((k-1)!(n-k-1)!)$ and $\mu = \phi/\log 2$, for $0 < y_k < y_{k+1} < \infty$. Let $R = Y_{(k+1)} - Y_{(k)}$ and $V = (Y_{(k)} + Y_{(k+1)})/2$. Then, V is our interest and we have an one-to-one transformation from $(Y_{(k)}, Y_{(k+1)})$ to (R, V) . Since the Jacobian for this transformation is -1 , the joint pdf of (R, V) is

$$f(r, v) = (c/\mu^2) \exp(-2v/\mu) [1 - \exp\{(v-r)/\mu\}]^{k-1} \exp\{-(v+r/2)(n-k-1)/\mu\},$$

for $0 < r < \infty$ and $r/2 < v < \infty$. Thus, the pdf of the sample median is

$$\begin{aligned} f_V(v) &= \int_0^{2v} f(r, v) dr \\ &= \int_0^{2v} c (\log 2/\phi)^2 \exp(-2v \log 2/\phi) [1 - \exp\{-\log 2(v-r)/\phi\}]^{k-1} \\ &\quad \times \exp\{-\log 2(v+r/2)(n-k-1)/\phi\} dr \end{aligned}$$

for $0 < v < \infty$.

B. Median event time test for other parametric survival distributions

We provide two theoretical results stating the distribution function of the sample median event time for uniform and weibull distributions.

Theorem 2

- Let Y_1, \dots, Y_n be random variables from a uniform distribution with lower bound parameter a and median ϕ (i.e., upper bound parameter is $2\phi - a$). Then,

$$g_n(m) = \frac{n!(m - a)^k(2\phi - a - m)^k}{k!k!\{2(\phi - a)\}^{2k+1}}$$

for $n = 2k + 1$ and

$$g_n(m) = \begin{cases} \int_0^{2(m-a)} \frac{c(m - a - r/2)^{k-1}(2\phi - a - m - r/2)^{n-k-1}}{\{2(\phi - a)\}^n} dr & a \leq m \leq \phi \\ \int_0^{-2(m-2\phi+a)} \frac{c(m - a - r/2)^{k-1}(2\phi - a - m - r/2)^{n-k-1}}{\{2(\phi - a)\}^n} dr & \phi < m \leq 2\phi - a \end{cases}$$

where $c = n!/\{(k - 1)!(n - k - 1)!\}$, for $n = 2k$.

- Let Y_1, \dots, Y_n be random variables from a uniform distribution with upper bound parameter b and median ϕ (i.e., lower bound parameter is $2\phi - b$). Then,

$$g_n(m) = \frac{n!(m - 2\phi + b)^k(b - m)^k}{k!k!\{2(b - \phi)\}^{2k+1}}$$

for $n = 2k + 1$ and

$$g_n(m) = \begin{cases} \int_0^{2(m-2\phi+b)} \frac{c(m - r/2 - 2\phi + b)^{k-1}(b - m - r/2)^{n-k-1}}{\{2(b - \phi)\}^n} dr & 2\phi - b \leq m \leq \phi \\ \int_0^{2(b-m)} \frac{c(m - r/2 - 2\phi + b)^{k-1}(b - m - r/2)^{n-k-1}}{\{2(b - \phi)\}^n} dr & \phi < m \leq b \end{cases}$$

where $c = n!/\{(k - 1)!(n - k - 1)!\}$, for $n = 2k$.

Proof. We first consider a uniform random variable with lower bound parameter a and median ϕ . Then, we have pdf $f(y) = 1/\{2(\phi - a)\}$ for $a \leq y \leq 2\phi - a$ and cdf

$$F(y) = \begin{cases} 0 & y < a \\ (y - a)/\{2(\phi - a)\} & a \leq y \leq 2\phi - a \\ 1 & y > 2\phi - a \end{cases}$$

By Cramér [26], if sample size is $n = 2k + 1$ for some $k = 0, 1, 2, \dots$, the pdf of the sample median is

$$g_n(m) = \frac{n!(m - a)^k(2\phi - a - m)^k}{k!k!\{2(\phi - a)\}^{2k+1}}.$$

When the sample size is $n = 2k$ for some $k = 1, 2, \dots$, we use the joint pdf of $Y_{(k)}$ and $Y_{(k+1)}$, where $Y_{(k)}$ denotes the k th order statistic of the sample, given by

$$f(y_k, y_{k+1}) = c(y_k - a)^{k-1}(2\phi - a - y_{k+1})^{n-k-1}/\{2(\phi - a)\}^n,$$

where $c = n!/\{(k - 1)!(n - k - 1)!\}$, for $a \leq y_k < y_{k+1} \leq 2\phi - a$. Let $R = Y_{(k+1)} - Y_{(k)}$ and $V = (Y_{(k)} + Y_{(k+1)})/2$. The one-to-one transformation from $(Y_{(k)}, Y_{(k+1)})$ to (R, V) gives the joint pdf

of (R, V) is

$$f(r, v) = c(v - a - r/2)^{k-1} (2\phi - a - v - r/2)^{n-k-1} / \{2(\phi - a)\}^n$$

for $0 < r \leq 2(\phi - a)$ and $r/2 + a \leq v \leq -r/2 + 2\phi - a$. Thus, the pdf of the sample median is the marginal density of V , which is

$$f_v(v) = \begin{cases} \int_0^{2(v-a)} \frac{c(v - a - r/2)^{k-1} (2\phi - a - v - r/2)^{n-k-1}}{\{2(\phi - a)\}^n} dr & a \leq v \leq \phi \\ \int_0^{-2(v-2\phi+a)} \frac{c(v - a - r/2)^{k-1} (2\phi - a - v - r/2)^{n-k-1}}{\{2(\phi - a)\}^n} dr & \phi < v \leq 2\phi - a \end{cases}$$

Similarly, for a uniform random variable with upper bound parameter b and median ϕ , we have pdf $f(y) = 1/\{2(b - \phi)\}$ for $2\phi - b \leq y \leq b$ and cdf

$$F(y) = \begin{cases} 0 & y < 2\phi - b \\ (y - 2\phi + b) / \{2(b - \phi)\} & 2\phi - b \leq y \leq b \\ 1 & y > b \end{cases}$$

It is easy to verify that

$$g_n(m) = \frac{n!(m - 2\phi + b)^k (b - m)^k}{k!k! \{2(b - \phi)\}^{2k+1}}, \quad n = 2k + 1 \text{ for some } k = 0, 1, 2, \dots$$

and

$$g_n(m) = \begin{cases} \int_0^{2(m-2\phi+b)} \frac{c(m - r/2 - 2\phi + b)^{k-1} (b - m - r/2)^{n-k-1}}{\{2(b - \phi)\}^n} dr & 2\phi - b \leq m \leq \phi \\ \int_0^{2(b-m)} \frac{c(m - r/2 - 2\phi + b)^{k-1} (b - m - r/2)^{n-k-1}}{\{2(b - \phi)\}^n} dr & \phi < m \leq b \end{cases},$$

$n = 2k$ for some $k = 1, 2, \dots$

Theorem 3 Let Y_1, \dots, Y_n be random variables from a weibull distribution with shape parameter τ and median ϕ (i.e., scale parameter is $\phi/(\log 2)^{1/\tau}$). Then,

$$g_n(m) = \frac{n! \tau (\log 2)^{1/\tau}}{k!k! \phi} \{(\log 2)^{1/\tau} m / \phi\}^{\tau-1} (1 - \exp[-\{(\log 2)^{1/\tau} m / \phi\}^\tau])^k \times \exp[-\{(\log 2)^{1/\tau} m / \phi\}^\tau (k + 1)]$$

for $n = 2k + 1$ and

$$g_n(m) = \int_0^{2m} c(\tau \log 2 / \phi^\tau)^2 \{(m - r/2)(m + r/2)\}^{\tau-1} (1 - \exp[-\{(\log 2)^{1/\tau} (m - r/2) / \phi\}^\tau])^{k-1} \exp[-\{(\log 2)^{1/\tau} (m - r/2) / \phi\}^\tau - \{(\log 2)^{1/\tau} (m + r/2) / \phi\}^\tau (n - k)] dr,$$

where $c = n! / \{(k - 1)!(n - k - 1)!\}$, for $n = 2k$.

Proof. Let Y be a weibull random variable with scale parameter τ and median ϕ . Then, we have pdf $f(y) = \tau(\log 2)^{1/\tau} \{(\log 2)^{1/\tau} y / \phi\}^{\tau-1} \exp[-\{(\log 2)^{1/\tau} y / \phi\}^\tau] / \phi$ and cdf $F(y) = 1 - \exp$

$[-\{(\log 2)^{1/\tau} y/\phi\}^\tau]$. By Cramér [26], if sample size is $n = 2k + 1$ for some $k = 0, 1, 2, \dots$, the pdf of the sample median is

$$g_n(m) = \frac{n! \tau (\log 2)^{1/\tau}}{k! k! \phi} \{(\log 2)^{1/\tau} m / \phi\}^{\tau-1} (1 - \exp[-\{(\log 2)^{1/\tau} m / \phi\}^\tau])^k \times \exp[-\{(\log 2)^{1/\tau} m / \phi\}^\tau (k + 1)].$$

When the sample size is $n = 2k$ for some $k = 1, 2, \dots$, by proof of Theorem 1, the pdf of the sample median is

$$g_n(m) = \int_0^{2m} c (\tau \log 2 / \phi^\tau)^2 \{(m - r/2)(m + r/2)\}^{\tau-1} (1 - \exp[-\{(\log 2)^{1/\tau} (m - r/2) / \phi\}^\tau])^{k-1} \exp[-\{(\log 2)^{1/\tau} (m - r/2) / \phi\}^\tau] - \{(\log 2)^{1/\tau} (m + r/2) / \phi\}^\tau (n - k) dr,$$

where $c = n! / \{(k - 1)!(n - k - 1)!\}$, for $0 < m < \infty$.

Supporting information

S1 File.

(PDF)

Author Contributions

Conceptualization: Yeonhee Park.

Funding acquisition: Yeonhee Park.

Investigation: Yeonhee Park.

Methodology: Yeonhee Park.

Software: Yeonhee Park.

Validation: Yeonhee Park.

Visualization: Yeonhee Park.

Writing – original draft: Yeonhee Park.

Writing – review & editing: Yeonhee Park.

References

1. Schwartz LH, Litière S, de Vries E, Ford R, Gwyther S, Mandrekar S, et al. RECIST 1.1—Update and clarification: From the RECIST committee. *European journal of cancer*. 2016; 62:132–137. <https://doi.org/10.1016/j.ejca.2016.03.081> PMID: 27189322
2. Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled clinical trials*. 1989; 10(1):1–10. [https://doi.org/10.1016/0197-2456\(89\)90015-9](https://doi.org/10.1016/0197-2456(89)90015-9) PMID: 2702835
3. Shan G, Zhang H, Jiang T, Peterson H, Young D, Ma C. Exact p-values for Simon's two-stage designs in clinical trials. *Statistics in biosciences*. 2016; 8(2):351–357. <https://doi.org/10.1007/s12561-016-9152-1> PMID: 28003856
4. Shan G, Wilding GE, Hutson AD, Gerstenberger S. Optimal adaptive two-stage designs for early phase II clinical trials. *Statistics in medicine*. 2016; 35(8):1257–1266. <https://doi.org/10.1002/sim.6794> PMID: 26526165
5. Shan G, Chen JJ, Ma C. Boundary problem in Simon's two-stage clinical trial designs. *Journal of biopharmaceutical statistics*. 2017; 27(1):25–33. <https://doi.org/10.1080/10543406.2016.1148716> PMID: 26881325

6. Kim J, Schell MJ. Modified Simon's minimax and optimal two-stage designs for single-arm phase II cancer clinical trials. *Oncotarget*. 2019; 10(42):4255. <https://doi.org/10.18632/oncotarget.26981> PMID: 31303960
7. Edelmann D, Habermehl C, Schlenk RF, Benner A. Adjusting Simon's optimal two-stage design for heterogeneous populations based on stratification or using historical controls. *Biometrical Journal*. 2020; 62(2):311–329. <https://doi.org/10.1002/bimj.201800390> PMID: 31553076
8. Fleming TR. One-sample multiple testing procedure for phase II clinical trials. *Biometrics*. 1982; p. 143–151. <https://doi.org/10.2307/2530297> PMID: 7082756
9. Mander A, Thompson S. Two-stage designs optimal under the alternative hypothesis for phase II cancer clinical trials. *Contemporary clinical trials*. 2010; 31(6):572–578. <https://doi.org/10.1016/j.cct.2010.07.008> PMID: 20678585
10. Case LD, Morgan TM. Design of phase II cancer trials evaluating survival probabilities. *BMC Medical Research Methodology*. 2003; 3(1):6. <https://doi.org/10.1186/1471-2288-3-6> PMID: 12697051
11. Huang B, Talukder E, Thomas N. Optimal two-stage phase II designs with long-term endpoints. *Statistics in Biopharmaceutical Research*. 2010; 2(1):51–61. <https://doi.org/10.1198/sbr.2010.09001>
12. Belin L, De Rycke Y, Broët P. A two-stage design for phase II trials with time-to-event endpoint using restricted follow-up. *Contemporary clinical trials communications*. 2017; 8:127–134. <https://doi.org/10.1016/j.conctc.2017.09.010> PMID: 29696201
13. Shan G, Zhang H. Two-stage optimal designs with survival endpoint when the follow-up time is restricted. *BMC medical research methodology*. 2019; 19(1):74. <https://doi.org/10.1186/s12874-019-0696-x> PMID: 30943896
14. Jin IH, Liu S, Thall PF, Yuan Y. Using data augmentation to facilitate conduct of phase I–II clinical trials with delayed outcomes. *Journal of the American Statistical Association*. 2014; 109(506):525–536. <https://doi.org/10.1080/01621459.2014.881740> PMID: 25382884
15. Yuan Y, Lin R, Li D, Nie L, Warren KE. Time-to-event Bayesian optimal interval design to accelerate phase I trials. *Clinical Cancer Research*. 2018; 24(20):4921–4930. <https://doi.org/10.1158/1078-0432.CCR-18-0246> PMID: 29769209
16. FDA. 22 Case studies where phase 2 and phase 3 trials had divergent results. 2017;.
17. Finkelstein DM, Muzikansky A, Schoenfeld DA. Comparing survival of a sample to that of a standard population. *Journal of the National Cancer Institute*. 2003; 95(19):1434–1439. <https://doi.org/10.1093/jnci/djg052> PMID: 14519749
18. Sun X, Peng P, Tu D. Phase II cancer clinical trials with a one-sample log-rank test and its corrections based on the Edgeworth expansion. *Contemporary clinical trials*. 2011; 32(1):108–113. <https://doi.org/10.1016/j.cct.2010.09.009> PMID: 20888929
19. Jung SH. *Randomized phase II cancer clinical trials*. CRC Press; 2013.
20. Kwak M, Jung SH. Phase II clinical trials with time-to-event endpoints: optimal two-stage designs with one-sample log-rank test. *Statistics in medicine*. 2014; 33(12):2004–2016. <https://doi.org/10.1002/sim.6073> PMID: 24338995
21. Wu J, Chen L, Wei J, Weiss H, Chauhan A. Two-stage phase II survival trial design. *Pharmaceutical Statistics*. 2020; 19(3):214–229. <https://doi.org/10.1002/pst.1983> PMID: 31749311
22. Wu J. Single-arm phase II trial design under parametric cure models. *Pharmaceutical statistics*. 2015; 14(3):226–232. <https://doi.org/10.1002/pst.1678> PMID: 25846141
23. Chu C, Liu S, Rong A. Study design of single-arm phase II immunotherapy trials with long-term survivors and random delayed treatment effect. *Pharmaceutical Statistics*. 2020;. <https://doi.org/10.1002/pst.1976> PMID: 31930622
24. Thall PF, Wooten LH, Tannir NM. Monitoring event times in early phase clinical trials: some practical issues. *Clinical trials*. 2005; 2(6):467–478. <https://doi.org/10.1191/1740774505cn121oa> PMID: 16422307
25. Zhou H, Chen C, Sun L, Yuan Y. Bayesian optimal phase II clinical trial design with time-to-event endpoint. *Pharmaceutical Statistics*. 2020;. <https://doi.org/10.1002/pst.2030> PMID: 32524679
26. Cramér H. *Mathematical methods of statistics*. Princeton university press. 2016; 9.