

RESEARCH ARTICLE

Open Access



Two-stage optimal designs with survival endpoint when the follow-up time is restricted

Guogen Shan^{1*†}  and Hua Zhang^{2*†}

Abstract

Background: Survival endpoint is frequently used in early phase clinical trials as the primary endpoint to assess the activity of a new treatment. Existing two-stage optimal designs with survival endpoint either over estimate the sample size or compute power outside the alternative hypothesis space.

Methods: We propose a new single-arm two-stage optimal design with survival endpoint by using the one-sample log rank test based on exact variance estimates. This proposed design with survival endpoint is analogous to Simon's two-stage design with binary endpoint, having restricted follow-up.

Results: We compare the proposed design with the existing two-stage designs, including the two-stage design with survival endpoint based on the nonparametric Nelson-Aalen estimate, and Simon's two-stage designs with or without interim accrual. The new design always performs better than these competitors with regards to the expected total study length, and requires a smaller expected sample size than Simon's design with interim accrual.

Conclusions: The proposed two-stage minimax and optimal designs with survival endpoint are recommended for use in practice to shorten the study length of clinical trials.

Keywords: Clinical trials, Exact variance, One-sample log-rank test, Restricted follow-up, Simon's two-stage design

Background

A multiple-stage design is often preferable in early phase clinical trials to investigate the activity of a new treatment. Such design is able to protect patients better as compared to the traditional one-stage design by allowing a trial to be stopped earlier when the new treatment is indeed ineffective. For this reason, early stopping for futility is always allowed in these trials. Among multiple-stage designs, a two-stage design is widely used in phase II clinical trials whose sample size is relatively smaller than that in the following phase III trial to confirm the effectiveness of the new treatment(s).

When the outcome is binary (e.g., response VS non-response), Simon's two-stage minimax and optimal

designs are widely used in practice [1–8]. When the required number of patients in the first stage are enrolled, a trial generally has to be suspended temporarily to allow these patients completing the treatment schedule. After that, data analysis is performed to make the decision whether a trial proceeds to the second stage or not, based on the result from the first stage. This suspension during the clinical trial could lead to a longer study time as compared to the modified Simon's two-stage design with interim accrual [9]. Recently, adaptive version of Simon's two-stage design has been proposed to improve the flexibility of trials [3, 4, 10–12]. In such trials, the second stage sample size depends on the outcome from the first stage.

In some other trials (e.g., cytostatic therapies), a survival endpoint is served as the primary outcome to measure the activity of a new treatment. Feldman et al. [13] reviewed seven single-arm phase II trials for patients with refractory germ cell tumors, and recommended a 12-week progression-free survival as compared to the commonly used response rate, to test the activity of novel agents. For

*Correspondence: guogen.shan@unlv.edu; zerozhua@126.com

†Guogen Shan and Hua Zhang contributed equally to this work and considered co-first author.

¹Epidemiology and Biostatistics Program, Department of Environmental and Occupational Health, School of Community Health Sciences, University of Nevada Las Vegas, Las Vegas 89154, NV, USA

²School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, Zhejiang, China



such trials, a multiple-stage design with survival endpoint would be appropriate for use in practice. Lin et al. [14] proposed group sequential designs for a trial with survival endpoint by deriving the asymptotic joint distribution of the Nelson-Aalen estimates at different time points. Based on Lin et al.'s work, Case and Morgan [9] developed a two-stage optimal design evaluating survival probabilities with restricted follow-up. They proposed two-stage optimal designs with the smallest expected duration of accrual or the smallest expected total study length. Later, Kwak and Jung [15] proposed a new two-stage optimal design based on the one-sample log-rank test without follow-up restriction. Power of their proposed design was computed under the average of the cumulative hazard function under the null hypothesis and that under the alternative hypothesis. In addition, the asymptotic variance estimate of the one-sample log-rank test was used in type I error rate and power calculation. Recently, Belin et al. [16] proposed a two-stage design based on the design setting as in Kwak and Jung [15], but having restricted follow-up as in Case and Morgan [9].

For a trial with a survival endpoint as the primary outcome, the survival probability at the clinically meaningful follow-up time is often the parameter of interest, (e.g., the survival probability at 1 year). We develop a new single-arm two-stage optimal design by using the one-sample log-rank test with exact mean and variance estimates [17, 18]. A trial is allowed to be stopped in the first stage due to futility to protect patients when the treatment under investigation is indeed ineffective. Although exact mean and variance estimates of the one-sample log-rank test are used for sample size calculation, the joint distribution of the test statistic for the first stage and that for the two stages combined is assumed to asymptotically follow a bivariate normal distribution. For this reason, the actual power of the identified study design may not be guaranteed [19]. We propose adjusting the nominal power level in design search to guarantee that the new designs meet the power requirement. The proposed two-stage minimax and optimal designs with survival endpoint are compared with the design by Belin et al. [16] and Simon's two-stage designs with or without interim accrual.

The rest of this article is organized as follows. In Section [Methods](#), we present the type I error rate and power calculation for a two-stage design with survival endpoint by using the one-sample log-rank test, and provide a detailed search method for two-stage minimax and optimal designs. In Section [Results](#), we compare the performance of the new proposed two-stage designs with the existing Belin's design with survival endpoint and Simon's two-stage design with binary endpoint. At the end of that section, we revisit two trials to illustrate the application of the proposed two-stage designs with

survival endpoint. Lastly, we provide some comments in Section [Discussion](#).

Methods

Suppose $S(t)$ is the survival function of the survival time T . In a single-arm study, the survival probability of a new treatment at the clinically meaningful follow-up time t_c , $S(t_c)$, is compared to the estimated historical survival probability, $S_0(t_c)$. Then the hypotheses are presented as

$$H_0 : S(t_c) \leq S_0(t_c) \text{ against } H_1 : S(t_c) > S_0(t_c). \quad (1)$$

In this article, the survival function $S(t)$ is assumed to follow the Weibull distribution with the shape parameter k and the scale parameter λ , specifically,

$$S(t) = \exp^{-t/\lambda^k},$$

where $k > 0$ and $\lambda > 0$. The widely used exponential distribution is a special case of the Weibull distribution when $k = 1$.

Under the Weibull distribution for survival outcome, suppose the failure rate under the null hypothesis is the same as that under the alternative hypothesis (the same shape parameter k), but scale parameters are different with λ_0 and λ_1 under the null hypothesis and the alternative hypothesis, respectively. Then, $\Delta = (\lambda_0/\lambda_1)^k$ is the hazard ratio (HR), which is always less than 1 under the alternative. The hypotheses in Eq. (1) can be specifically rewritten as

$$H_0 : \Delta \geq 1 \text{ against } H_1 : \Delta < 1. \quad (2)$$

When a new study is assumed to have a different failure rate as historical data, the HR is then calculated as $\Delta = \frac{\lambda_0^{k_0}}{\lambda_1^{k_1}} \times \frac{k_1 t^{k_1-1}}{k_0 t^{k_0-1}}$, where k_0 and k_1 are the shape parameter under the null hypothesis and that under the alternative hypothesis, respectively.

Simon's two-stage designs with binary endpoint

In Simon's two-stage optimal designs, a trial is allowed to be stopped in the first stage when the number of responses is insufficient. Suppose X_1 and X are the number of responses out of n_1 and n participants from the first stage and the two stages combined, respectively. The sample size in the second stage is $n_2 = n - n_1$. The null hypothesis is rejected when $X_1 > r_1$ and $X > r$, where r_1 and r are the critical values for the number of responses from the first stage and both stages, respectively.

In a pancreatic cancer trial with a combination of Gemcitabine and external beam radiation as the new treatment [9], the clinically meaningful follow-time is 1 year, $t_c = 1$. The unacceptable one-year survival rate is $S_0(1) = 35\%$, and the new treatment is considered as promising for further investigation when $S_1(1) = 50\%$ or more. To attain

90% power of the study at the significance level of 10%, Simon’s two-stage minimax design [1] is calculated as:

$$(n_1, r_1, n, r) = (43, 14, 72, 30),$$

with the expected sample size under the null hypothesis $ESS_0 = n_1 + (1 - PET)n_2 = 59.3$, where PET is the probability of early termination under the null hypothesis which is defined as $PET = p(X_1 \leq r_1 | S_0(1) = 35\%) = 43.65\%$. Suppose this is a 3 year study with the patient accrual rate of $\theta = 24$ patients per year. Then the enrollment time for the first stage and the second stage is calculated $t_1 = n_1/\theta$ and $t_2 = n_2/\theta$, respectively. The expected total study length (ETSL) under the null hypothesis is calculated as

$$ETSL_0 = (t_1 + t_c) + (1 - PET)(t_2 + t_c) = 4.0 \text{ years}$$

The two-stage optimal design needs $ESS_0 = 53.2$ and $ETSL_0 = 3.6$ years (see Table 1). The maximum possible sample size for Simon’s optimal design $n = 81$ is much larger than $n = 72$ for Simon’s minimax design.

When Simon’s two-stage design allows interim accrual at the end of the first stage, the expected sample size under the null hypothesis is calculated as

$$ESS_0 = n_1 + \theta t_c + (1 - PET)(n_2 - \theta t_c),$$

and the expected total study length under the null hypothesis is

$$ETSL_0 = (t_1 + t_c) + (1 - PET) [(t_2 - t_c) + t_c] \\ = (t_1 + t_c) + (1 - PET)t_2$$

The results of Simon’s two-stage designs with interim accrual are presented in Table 1. As compared to the traditional Simon’s two-stage design without interim accrual, the modified design with interim accrual requires a shorter $ETSL_0$ but a larger ESS_0 .

Two-stage optimal designs with survival endpoint when the follow-up time is limited

In a two-stage design with sample sizes of n_1 in the first stage and n_2 in the second stage, the maximum possible sample size in the study is $n = n_1 + n_2$. Given the patient accrual rate of θ , the accrual time for the first stage is $t_1 = n_1/\theta$. When the trial goes to the second stage, the total accrual time of the study is $t_a = n/\theta$, and the total study time for all patients to complete the study is $t = t_a + t_c$.

We assume that patients are uniformly enrolled in the study, with the entering times of $\tau_1, \tau_2, \dots, \tau_n$. They have the survival times of T_1, T_2, \dots, T_n and the censoring times of C_1, C_2, \dots, C_n . At the end of the first stage t_1 , the observed time for the i -th patient is the smallest of the following three measurements: (1) event time; (2) censoring time; and (3) time that this patient is followed so far in the study, specifically,

$$O_i = \min(T_i, C_i, \max(0, t_1 - \tau_i)).$$

By using the observed time and the censoring information of the first n_1 patients, the one-sample log-rank test can be calculated as

$$Z_1 = \frac{W_1}{\hat{\sigma}_1},$$

where W_1 is a function of the difference between observed number of events and the expected number of events, and $\hat{\sigma}_1$ is its standard deviation estimate. Please find the detailed formula of Z_1 under the null hypothesis and the alternative hypothesis in Appendix.

The null hypothesis is rejected when a small test statistic is observed. Suppose the critical value for Z_1 is c_1 . When the calculated Z_1 is larger than or equal to c_1 , the trial is stopped for futility and no further investigation is warranted. Otherwise, the trial goes to the second stage with additional $n_2 = n - n_1$ patients treated by the new treatment. At the end of study when all n patients complete the study, the one-sample log-rank test is calculated as

$$Z = \frac{W}{\hat{\sigma}}.$$

It can be seen that Z_1 and Z are not independent from each other since the data of the first n_1 patients is used in both Z_1 and Z . The type I error (TIE) rate of the study is calculated as

$$TIE = P(Z_1 \leq c_1, Z \leq c | H_0),$$

where c is the critical value for Z .

Following Kwak and Jung [15], the joint distribution of (Z_1, Z) is a bivariate normal distribution asymptotically. Then, the TIE can be specifically written as

Table 1 The resectable pancreatic cancer clinical trial with $S_0(t_c = 1) = 35\%$, and $S_1(t_c = 1) = 50\%$ to attain 90% power at the significance level of 10%

	Survival endpoint						Simon’s design, interim accrual									
	The proposed method						Belin		No				Yes			
	n_1	n	c_1	c	ESS_0	$ETSL_0$	ESS_0	n	ESS_0	$ETSL_0$	n	ESS_0	$ETSL_0$	n	ESS_0	$ETSL_0$
Minimax	44	73	0.240	-1.281	61.3	3.1			59.3	4.0	72	69.8	3.5			
Optimal	41	79	-0.085	-1.279	58.7	2.9	59.1	69	53.2	3.6	81	67.4	3.2			

The survival function follows an exponential distribution

$$TIE = \int_{-\infty}^c \phi(t) \Phi \left(\frac{c_1 - \rho_0 t}{\sqrt{1 - \rho_0^2}} \right) dt, \tag{3}$$

where ϕ and Φ are the probability density function and the cumulative distribution function of the standard normal distribution, and ρ_0 is the correlation coefficient estimate between Z_1 and Z under the null hypothesis, see Appendix for the detailed formula for ρ_0 . The actual power of the study can be computed similarly with ρ_0 being replaced by the ρ estimate under the alternative hypothesis.

Optimal design search

Similar to the search for Simon’s two-stage design, the two-stage optimal design with survival endpoint has to be searched over all the possible sample sizes (n_1 and n) and critical values (c_1 and c), given the design parameters ($\alpha, \beta, t_c, S_0(t_c), S_1(t_c), \theta$).

Although the exact variances of Z_1 and Z are available for use in sample size determination, the exact joint distribution of Z_1 and Z is not that straightforward. For this reason, we utilize the limiting distribution of (Z_1, Z) in searching for the two-stage optimal design for a study with the design parameters ($\alpha, \beta, t_c, S_0(t_c), S_1(t_c), \theta$), then use a simulation study to calculate the actual TIE and power of the optimal design. The following three steps are used to search for the two-stage minimax and optimal designs.

Step 1: Given the total sample size n , the range of the first stage sample size n_1 is from 1 to $n - 1$. The critical value c_1 from -0.3 to 1.6 with an increment of 0.005 is used in the design search. Similar to Kwak and Jung [15], the range of c_1 is chosen based on the simulation studies for all the configurations studied in this article. The range of c_1 is modifiable in the software program for design search.

For each combination of n_1 and c_1 , the critical value c can be determined as the largest c value such that $TIE(c) \leq \alpha$ from Eq. (3). Power of the study is then computed by using Eq. (4) in Appendix. If power is above

the nominal level, this set of sample sizes and critical values, (n_1, c_1, n, c) , is saved as a candidate for the optimal two-stage design. Among all the sets satisfying the power requirement, the one with the smallest ESS_0 is the optimal two-stage design when the total sample size is n , and it is denoted as $B(n) = (n_1, c_1, n, c)$ whose expected sample size is $ESS_0(n)$.

Step 2: The design search starts with a relatively small n (e.g., 5) with an increment of 1, and $B(n)$ could be a empty set when n is small. The two-stage minimax design is the one with the smallest n , $n_{minimax}$ such that $B(n)$ is not empty. The optimal two-stage design is the one with the smallest ESS_0 . The search may be stopped at n_u when its $ESS_0(n_u)$ is 10% more than the smallest ESS_0 from the identified optimal designs with n from $n_{minimax}$ to n_u : $ESS_0(n_u) \geq 110\% \times \min\{ESS_0(n) : n_{minimax} \leq n \leq n_u\}$.

Step 3: Once the minimax and optimal two-stage designs are identified from Step 1 and Step 2, we use a simulation study to calculate the actual TIE and power based on 100,000 simulations. We find that the actual TIE of the optimal design $B(n) = (n_1, c_1, n, c)$ is always guaranteed, while power may not be preserved in some cases. If the simulated power of the two designs meet the nominal levels, they are the final two-stage minimax and optimal designs. Otherwise, we search for the designs again with the power nominal level being increased by 1%, ($\alpha, \beta - 1\%$) in Step 1 and Step 2 again. This process is stopped when both minimax and optimal two-stage designs meet the power requirement.

Results

We first compare the proposed two-stage minimax and optimal designs with survival endpoint when the follow-up time is restricted, with the designs developed by Belin et al. [16] (referred to as Belin’s design). They developed a two-stage optimal design as a modification of the design by Kwak and Jung [15] by adding restricted follow-up in the study design [9]. In Belin’s design, power of the study is computed at the average of the cumulative hazard functions under the null and the alternative, that is less than

Table 2 Comparison between the proposed two-stage minimax and optimal designs with survival endpoint and Belin’s two-stage optimal design with survival endpoint, when the follow-up time is restricted to the clinically meaningful follow-up time $t_c = 1$ year

Power	θ	Minimax design					Optimal design					Belin	
		n_1	n	c_1	c	ESS_0	n_1	n	c_1	c	ESS_0	n	ESS_0
90%	15	28	52	-0.10	-1.64	39.1	26	56	-0.30	-1.64	37.5	53	42.3
95%	15	36	65	-0.09	-1.64	49.5	33	70	-0.29	-1.64	47.3	65	52.6
90%	30	30	52	0.30	-1.64	43.6	30	55	-0.04	-1.64	42.2	53	44.6
95%	30	40	65	0.19	-1.64	54.3	40	69	-0.20	-1.64	52.2	65	54.8
90%	50	34	52	0.51	-1.64	46.5	32	54	0.32	-1.63	45.7	52	47.0
95%	50	44	65	0.46	-1.64	58.2	42	68	0.17	-1.64	56.7	64	57.5

The null survival probability at 1 year is $S_0(t_c) = 50\%$, and the hazard ratio is 2. Patient accrual rate θ is set as 15, 30, or 50 per year

Table 3 Simulated TIE and power of the proposed two-stage minimax and optimal designs in Table 2

Power	θ	Minimax design		Optimal design	
		TIE	Power	TIE	Power
90%	15	0.040 (0.036,0.044)	0.907 (0.901,0.913)	0.037 (0.033,0.041)	0.903 (0.898,0.909)
95%	15	0.041 (0.037,0.045)	0.957 (0.953,0.961)	0.038 (0.035,0.042)	0.955 (0.951,0.959)
90%	30	0.040 (0.037,0.044)	0.911 (0.905,0.916)	0.039 (0.035,0.043)	0.910 (0.904,0.916)
95%	30	0.042 (0.038,0.046)	0.959 (0.955,0.963)	0.040 (0.036,0.044)	0.958 (0.954,0.962)
90%	50	0.041 (0.037,0.045)	0.911 (0.905,0.916)	0.040 (0.037,0.044)	0.909 (0.903,0.914)
95%	50	0.042 (0.038,0.046)	0.960 (0.956,0.963)	0.041 (0.037,0.045)	0.959 (0.955,0.963)

The 95% confidence intervals for the parameters of interest are computed using 1000 simulations where 10,000 designs are simulated in each simulation

the cumulative hazard functions under the alternative at which value the actual power should be computed. This leads to an decreased effect size in sample size calculation; thus, the computed sample size may be over-estimated. As a result of the over-estimated sample size, the actual power is often above the nominal level.

Table 2 shows the comparison between the proposed designs with Belin’s design, when the survival distribution follows an exponential distribution. Belin et al. [16] investigated the performance of two-stage optimal designs with restricted follow-up under exponential distributions only (the shape parameter $k = 1$ in the Weibull distribution). The clinically meaningful follow-up time t_c is assumed to be 1 year. Under the null hypothesis, the survival rate at $t_c = 1$ is $S_0(t_c) = 50%$ ($\lambda_0 = 1.44$) as studied in Table 2. The hazard ratio is assumed to be 0.5, which is $\Delta = \lambda_0/\lambda_1 = 0.5$. Then the scale parameter under the alternative is $\lambda_1 = 2.88$. The nominal power level is set as either 90% or 95%. The accrual rate θ is 15, 30, or 50. The ESS_0 of the proposed minimax or optimal designs is often less than that of the Belin’s design, that may be due to the fact

that power of Belin’s design is computed outside the alternative hypothesis space. The simulated TIE and power of the developed two-stage minimax and optimal designs are shown in Table 3. In Table 3, we also report the 95% confidence interval for the TIE and power based on 1,000 simulated TIE and power values, where each simulated TIE and power are computed using 10,000 simulations. It can be seen that the proposed designs control for TIE and power.

We further compare the proposed two-stage minimax and optimal designs with survival endpoint, with Simon’s two-stage designs with or without interim accrual for a trial with binary endpoint, see Table 4 when the survival distribution follows the Weibull distribution with a common shape parameter of $k = 0.5$. The significance level is set as 5%, and the nominal power level is 80%. The null survival probability at the clinically meaningful follow-up time $t_c = 1$, $S_0(t_c) = 10%$ and 60% are studied in Table 4. We consider a medium to large effect size as $S_1(t_c) - S_0(t_c) = 10%$, 15%, and 20%. For each configuration of $S_0(t_c)$ and $S_1(t_c)$, the

Table 4 Comparison between the proposed two-stage minimax design with survival endpoint and Simon’s two-stage minimax design with binary endpoint with or without interim accrual, when $\alpha = 5%$, $\beta = 20%$, and the shape parameter $k = 0.5$ in the Weibull distribution

$S_0(t_c)$	$S_1(t_c)$	Survival endpoint		Simon’s two-stage minimax designs							
		n_1	n	No interim accrual				Interim accrual			
		ESS_0	$ETSL_0$	n_1	n	ESS_0 (%)	$ETSL_0$ (%)	ESS_0 (%)	$ETSL_0$ (%)	ESS_0 (%)	$ETSL_0$ (%)
0.1	0.2	37	63	50.5	2.5	45	78	60.6 (17%)	3.8 (35%)	74.3 (32%)	3.3 (26%)
0.1	0.25	19	33	26.2	2.5	22	40	28.8 (9%)	3.5 (30%)	37.5 (30%)	3.1 (21%)
0.1	0.3	11	21	15.6	2.3	15	25	19.5 (20%)	3.8 (39%)	24.5 (36%)	3.3 (30%)
0.6	0.7	87	162	126.6	3.2	139	142	139.2 (9%)	4.0 (20%)	184.5 (31%)	3.9 (19%)
0.6	0.75	33	70	49.4	2.8	30	62	43.8 (-13%)	3.6 (20%)	55.7 (11%)	3.1 (9%)
0.6	0.8	17	39	26.0	2.6	13	35	20.8 (-25%)	3.1 (16%)	28.5 (9%)	2.8 (5%)

% is for the ESS_0 or the $ETSL_0$ percentage saving of the new proposed two-stage design as compared to Simon’s two-stage design, which is computed as (Simon-New)/Simon. When the percentage saving is positive, the new design requires a smaller ESS_0 or a shorter $ETSL_0$ as compared to the existing Simon’s design. The patient accrual rate θ is determined by the sample size from Simon’s minimax design with no interim accrual as $\theta = n_{minimax}/3$

scale parameters λ_0 and λ_1 in the Weibull distribution can be calculated, the ESS_0 and $ETSL_0$ of the proposed minimax design and Simon's minimax design are computed. Patient accrual rate θ is calculated by assuming it is a 3 year study when Simon's two-stage minimax design is used. In the table, percentage (%) is for the ESS_0 or the $ETSL_0$ percentage saving of the proposed two-stage design with survival endpoint as compared to Simon's two-stage design, which is computed as (Simon-New)/Simon. When the percentage saving is positive, the new design requires a smaller ESS_0 or a shorter $ETSL_0$ as compared to the existing Simon's design. When the null survival probability $S_0(t_c)$ is low, say 10%, the proposed two-stage design with survival endpoint saves sample size as compared to Simon's two-stage minimax design. This trend is reversed when $S_0(t_c) = 60\%$. In Table 4, we also present the results of Simon's two-stage minimax design with interim accrual. It can be seen that the new design always requires a smaller ESS_0 than Simon's design with interim accrual. The new design always saves the $ETSL_0$ as compared to Simon's design with or without interim accrual. The saving becomes smaller as the null survival probability goes up from 10% to 60%. Similar results are observed in Table 5 for the two-stage optimal designs.

We further compare the new two-stage minimax design with Simon's two-stage minimax design with the shape parameter k from 0.25 to 2 in Fig. 1 for a trial to attain 90% power at the significance level of 5%. When $S_0(t_c)$ is low, the new design needs a smaller expected sample size than Simon's minimax design, and this trend is reversed when $S_0(t_c)$ is high, e.g., 40%, and 75%. The saving of the new design often decreases as k goes up. The new design always requires a shorter expected total study length than Simon's minimax design. Similar

results are observed in Fig. 2 where the new two-stage optimal design is compared with Simon's optimal design. We also compare the new design with Simon's two-stage minimax and optimal designs with interim accrual in Fig. 3 and Fig. 4, respectively. The results indicate that the new design performs better than Simon's design with interim accrual with regards to both ESS_0 and $ETSL_0$.

Examples

We revisit the cancer trial discussed by Case and Morgan [9] in "Simon's two-stage designs with binary endpoint" subsection to investigate the effectiveness of a combination of Gemcitabine and external beam radiation for patients with resectable pancreatic cancer. The clinically meaningful follow-up time is assumed to be 1 year, $t_c = 1$. The survival probability under the null and the alternative are $S_0(1) = 35\%$, and $S_1(1) = 50\%$, respectively. The survival function follows an exponential distribution. This trial is designed to attain 90% power at the significance level of 10%. We compute the detailed two-stage designs with survival endpoint, including sample sizes and critical values for each stage in Table 1. The ESS_0 of the new design is slightly larger than that of Simon's design, but much smaller than that of Simon's design with interim accrual. The $ETSL_0$ of the new design is always shorter than that of Simon's designs with or without interim accrual, and the study time saving is substantial.

We also consider a second clinical trial evaluating the activity of a combination of irinotecan and cisplatin for patients with refractory or recurrent non-small cell lung cancer [20]. The response rates are 10% and 25% under the null and the alternative hypotheses. Suppose the clinically meaningful follow-up time is 1 year. For Simon's

Table 5 Comparison between the proposed two-stage optimal design with survival endpoint and Simon's two-stage optimal design with binary endpoint with or without interim accrual, when $\alpha = 5\%$, $\beta = 20\%$, and the shape parameter $k = 0.5$ in the Weibull distribution

$S_0(t_c)$	$S_1(t_c)$	Survival endpoint				Simon's two-stage optimal designs					
		n_1	n	ESS_0	$ETSL_0$	No interim accrual		Interim accrual			
						n_1	n	ESS_0 (%)	$ETSL_0$ (%)	ESS_0 (%)	$ETSL_0$ (%)
0.1	0.2	26	72	45.1	2.2	30	89	50.8 (11%)	3.3 (35%)	67.6 (33%)	3.0 (27%)
0.1	0.25	15	37	24.0	2.2	18	43	24.7 (3%)	3.1 (29%)	34.9 (31%)	2.8 (22%)
0.1	0.3	10	23	15.0	2.2	10	29	15.0 (0%)	3.1 (29%)	21.6 (30%)	2.8 (21%)
0.6	0.7	66	179	109.2	2.7	53	173	91.4 (-20%)	3.3 (18%)	124.0 (12%)	2.9 (9%)
0.6	0.75	27	76	46.1	2.6	27	67	39.4 (-17%)	3.2 (18%)	53.9 (14%)	2.9 (10%)
0.6	0.8	15	41	25.1	2.5	11	43	20.5 (-23%)	3.1 (17%)	28.9 (13%)	2.8 (7%)

% is for the ESS_0 or the $ETSL_0$ percentage saving of the new proposed two-stage design as compared to Simon's two-stage design, which is computed as (Simon-New)/Simon. When the percentage saving is positive, the new design requires a smaller ESS_0 or a shorter $ETSL_0$ as compared to the existing Simon's design. The patient accrual rate θ is determined by the sample size from Simon's minimax design with no interim accrual as $\theta = n_{minimax}/3$

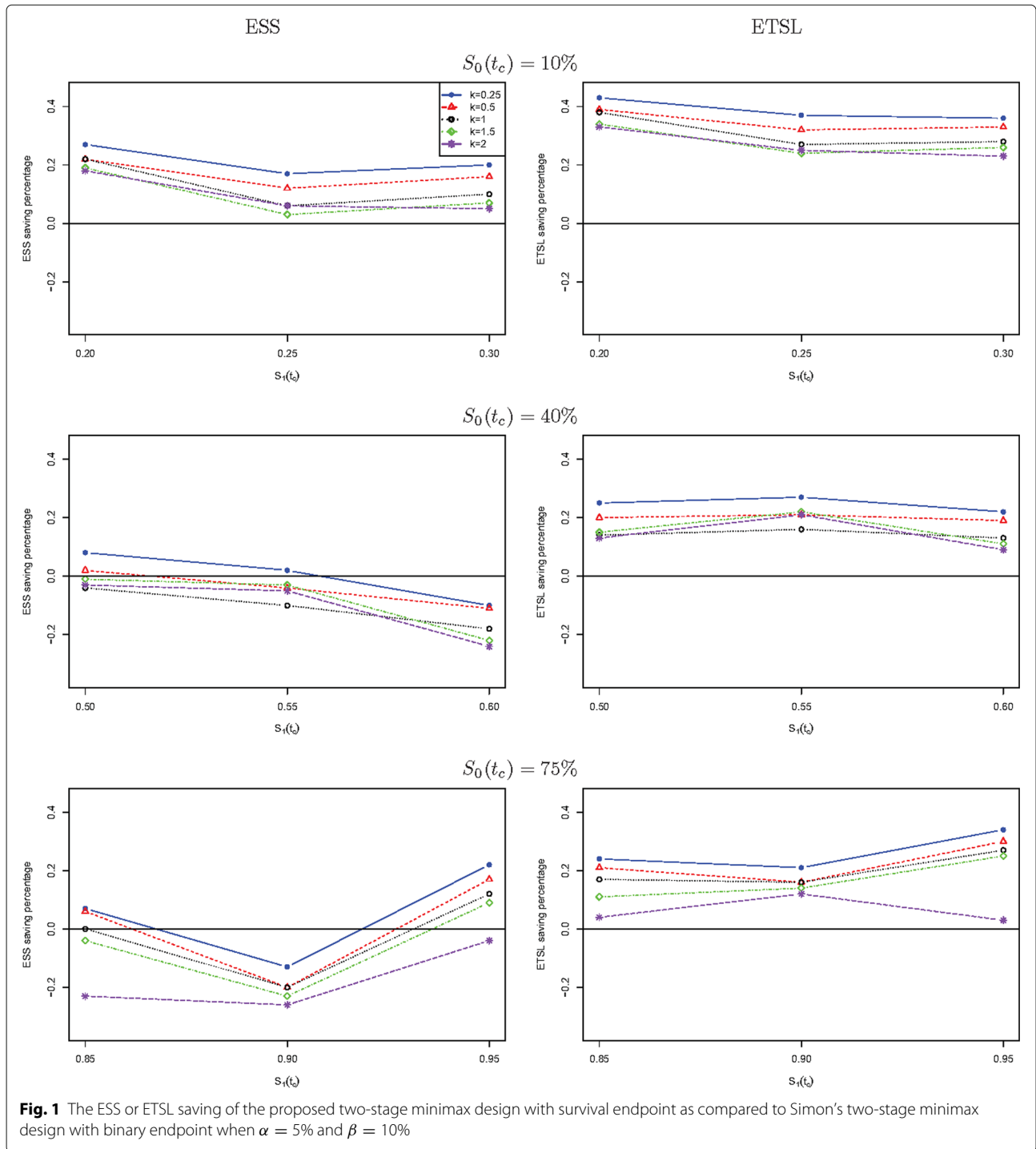
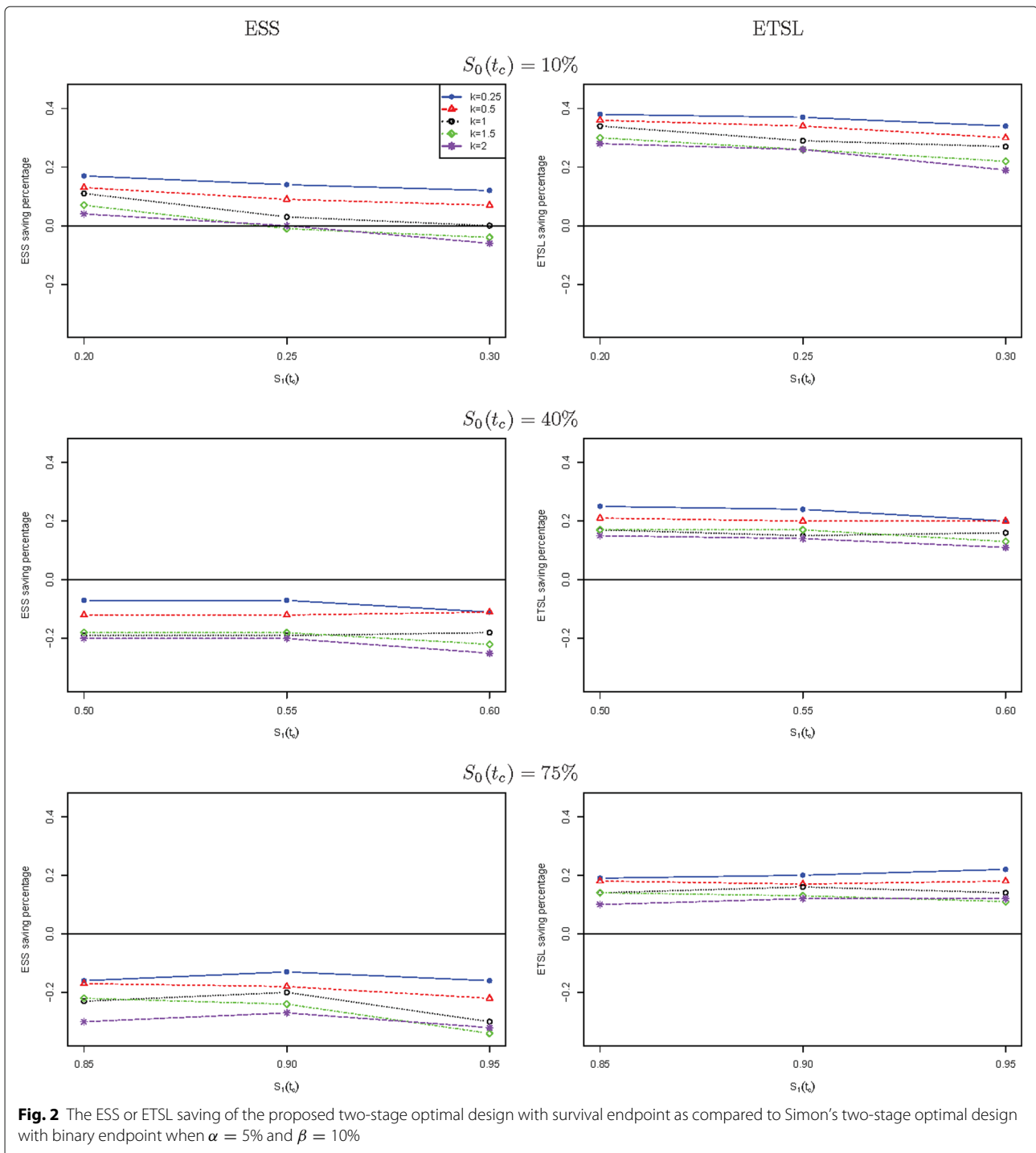


Fig. 1 The ESS or ETSL saving of the proposed two-stage minimax design with survival endpoint as compared to Simon’s two-stage minimax design with binary endpoint when $\alpha = 5\%$ and $\beta = 10\%$

two-stage optimal design when $\alpha = 5\%$ and $\beta = 20\%$, the maximum possible sample size is $n = 43$ and the expected sample size under the null hypothesis is $ESS_0 = 24.7$, see Table 5 for the case with $S_0(t_c) = 10\%$ and $S_1(t_c) = 25\%$. The proposed new two-stage optimal design with survival endpoint needs a slightly smaller ESS_0 as 24.0, and can

save the expected total study length by almost 1 year (2.2 VS 3.1 from Simon’s design). A 95% two-sided confidence interval of the response rate was reported in the original research article by Takiguchi et al. [20]. The hypothesis is one sided in both Simon’s design and the proposed design. Therefore, a 90% two-sided confidence interval for the



response rate or the survival rate should be reported when $\alpha = 5\%$.

Discussion

In the design search process, we search for the minimax and optimal designs when both designs have power above

the nominal level. In practice, when one type of design is of interest (e.g., the two-stage minimax design), we would suggest searching for the design such that power of this particular type design is above the nominal level. The written R program computes the designs to have both the minimax design and the optimal design meet the nominal

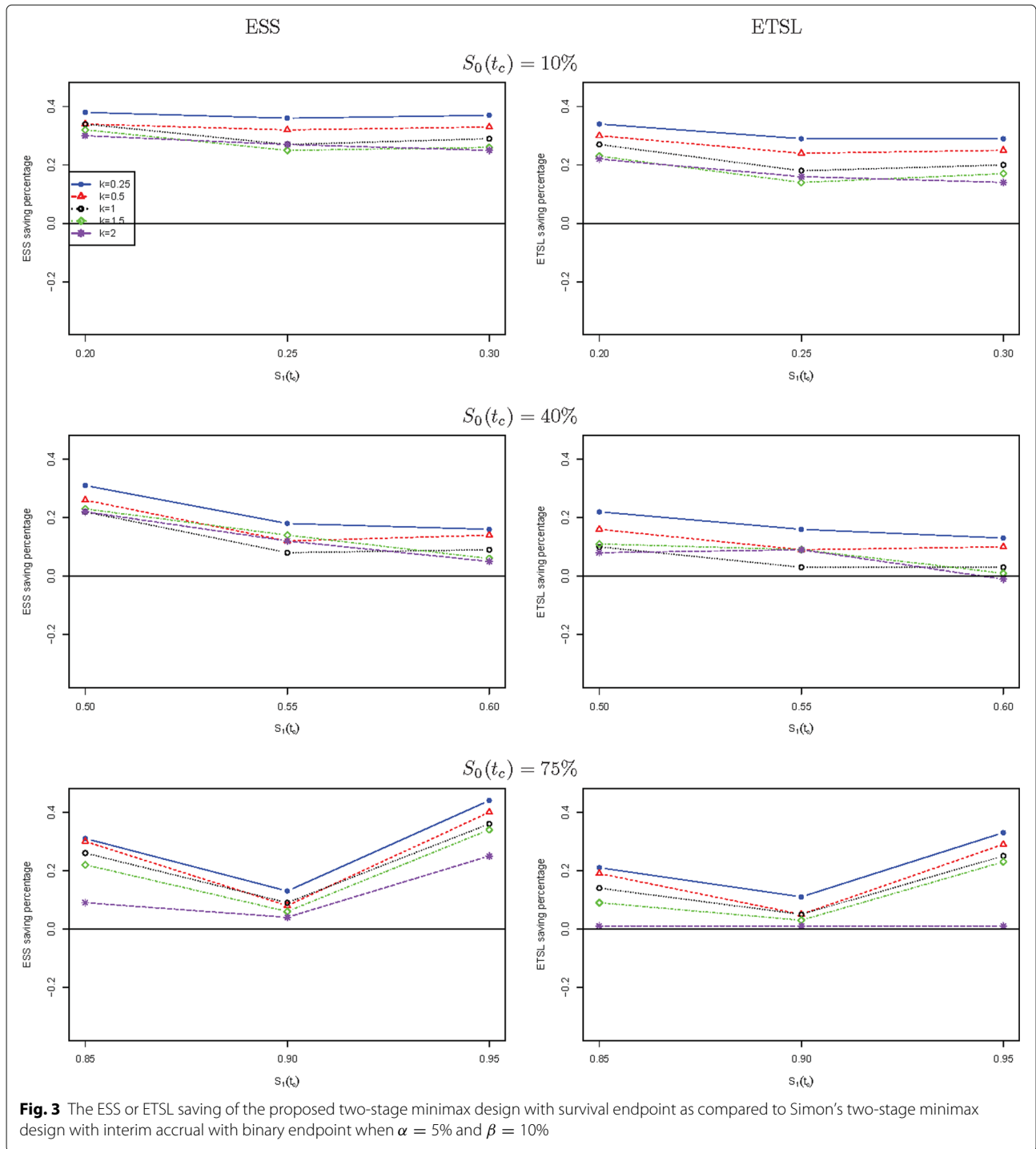


Fig. 3 The ESS or ETSL saving of the proposed two-stage minimax design with survival endpoint as compared to Simon's two-stage minimax design with interim accrual with binary endpoint when $\alpha = 5\%$ and $\beta = 10\%$

power level, which is available upon request from the first author.

Conclusions

The commonly used Simon's two-stage design has to suspend the enrollment temporarily after n_1 patients enrolled in the first stage [5, 11, 21–28]. The research

team has to wait a while (t_c) until all n_1 patients complete the study. The calculated test statistic from the first stage is then compared to the pre-determined critical value to make a go or no-go decision to the second stage. Meanwhile, the proposed two-stage designs with survival endpoint do not have to suspend the trial, thus the comparison between the proposed design with

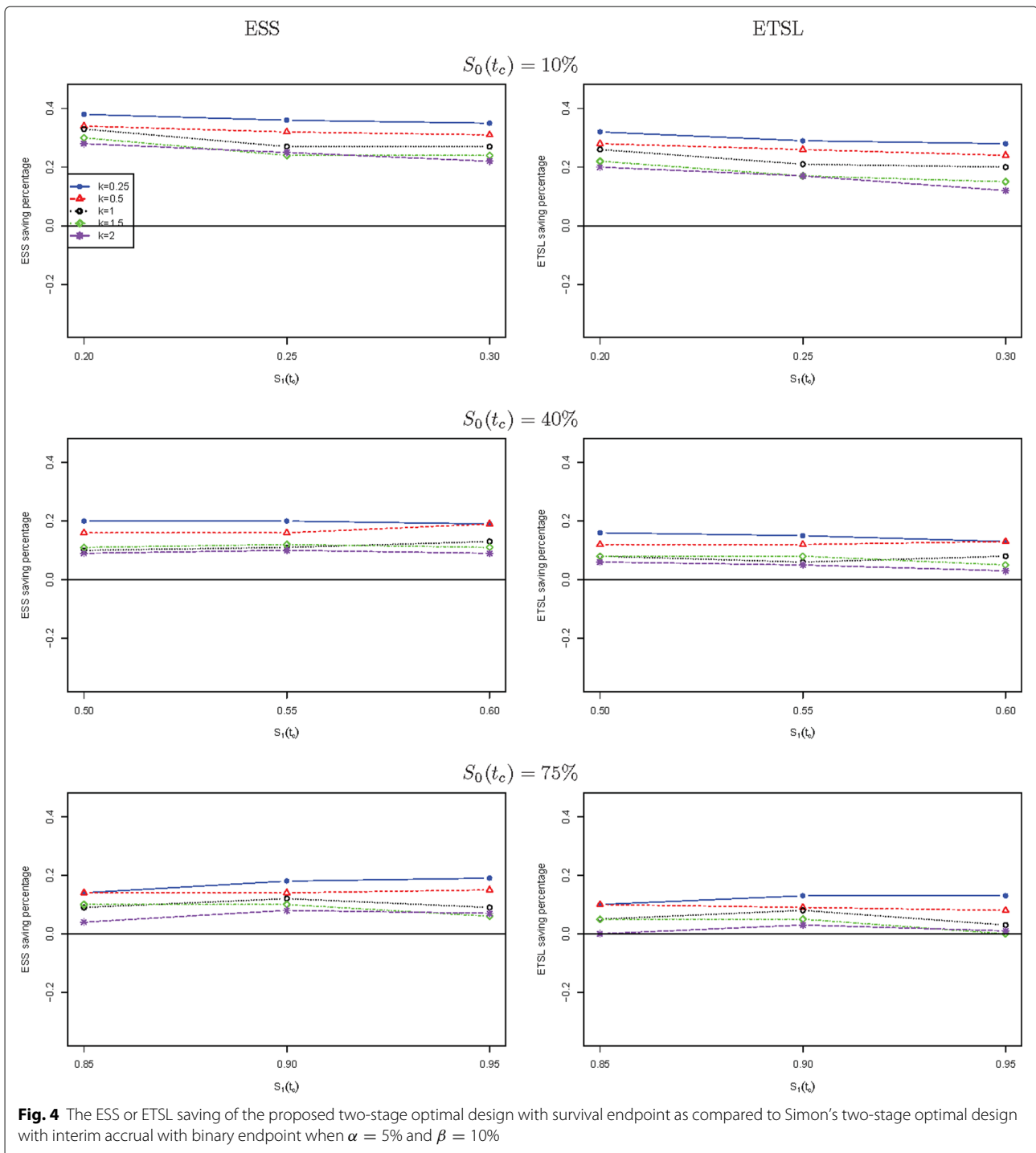


Fig. 4 The ESS or ETSL saving of the proposed two-stage optimal design with survival endpoint as compared to Simon’s two-stage optimal design with interim accrual with binary endpoint when $\alpha = 5\%$ and $\beta = 10\%$

Simon’s two-stage design with no interim accrual is not very appropriate. Due to the popularity of Simon’s two-stage design, we include this design as reference. Simon’s two-stage design with interim accrual is a reasonable competitor for the proposed two-stage design with survival endpoint.

Appendix
Test statistics of Z_1 and Z

At the end of the first stage t_1 , the observed time for the i -th patient is $O_i = \min(T_i, C_i, \max(0, t_1 - \tau_i))$, where $C_i = t_c$ with restricted follow-up, and $i = 1, 2, \dots, n_1$. Let $N_i(t) = I(T_i \leq \min(C_i, \max(0, t - \tau_i)))I(T_i \leq t)$ and

$Y_i(t) = I(T_i \geq t, T_i \geq t_c)$ be the event process and the at-risk process, respectively. The one-sample log-rank test at the end of the first stage is expressed as:

$$Z_1 = \frac{O - E}{\sqrt{E}},$$

where $O = \sum_{i=1}^n \int_0^\infty dN_i(t)$ are $E = \sum_{i=1}^n \int_0^\infty Y_i(t) d\Lambda_0(t)$ are the observed number of events and the expected number of events, respectively. The one-sample log-rank test can be alternatively written as

$$Z_1 = \frac{W_1}{\hat{\sigma}_1},$$

where $W_1 = (O - E)/\sqrt{n}$ and $\hat{\sigma} = E/n$, and $\hat{\sigma}_1^2$ is the variance estimate of W_1 . The one-sample log-rank test Z at the end of the study can be derived similarly by replacing $N_i(t)$ with $N_i(t) = I(T_i \leq C_i)I(T_i \leq t)$.

Mean and variance estimates of W_1 and W under the null hypothesis

The mean of W_1 or W under the null hypothesis is 0. The clinically meaningful follow-up time t_c is the upper bound follow-up time for each patient, then the censoring distribution is $G(t) = I(t \leq t_c)$. The censoring distribution for the first stage is $G_1(t) = U(0, t_1)I(t \leq t_c)$ due to a possible short follow-up time at the data analysis time t_1 . Then, the variances of W_1 and W are estimated as

$$\sigma_{01}^2 = Var(W_1) = - \int_0^{t_c} G_1(t) dS_0(t) \text{ and}$$

$$\sigma_{02}^2 = Var(W) = - \int_0^{t_c} G(t) dS_0(t).$$

It follows that the correlation between W_1 and W under H_0 is $\rho_0 = \sigma_{01}/\sigma_{02}$. The TIE in Eq. (3) can then be computed after the correlation coefficient ρ_0 being estimated.

Mean and variance estimates of W_1 and W under the alternative hypothesis

Under the alternative hypothesis, the mean values of W_1 and W are

$$E(W_1) = \sqrt{n_1}\omega_1 \text{ and } E(W) = \sqrt{n}\omega$$

where $\omega = p_1 - p_0$, $p_1 = \int_0^{t_c} G(t)S_1(t)d\Lambda_1(t)$, $p_0 = \int_0^{t_c} G(t)S_1(t)d\Lambda_0(t)$, and $\omega_1 = p_{1f} - p_{0f}$, $p_{1f} = \int_0^{t_c} G_1(t)S_1(t)d\Lambda_1(t)$, $p_{0f} = \int_0^{t_c} G_1(t)S_1(t)d\Lambda_0(t)$. Recently, Wu [17] derived the exact variance of W under the alternative hypothesis as

$$\sigma_{12}^2 = Var(W) = p_1 - p_1^2 - p_0^2 + 2p_0p_1 + 2p_{00} - 2p_{01},$$

where $p_{00} = \int_0^{t_c} G(t)S_1(t)\Lambda_0(t)d\Lambda_0(t)$ and $p_{01} = \int_0^{t_c} G(t)S_1(t)\Lambda_0(t)d\Lambda_1(t)$. The exact variance of W_1 , $\sigma_{11}^2 = Var(W_1)$, can be derived similarly. It follows that the correlation between W_1 and W under H_1 is $\rho_1 = \sigma_{11}/\sigma_{12}$, and power of a two-stage design is

$$Power = \int_{-\infty}^{\tilde{c}} \phi(t)\Phi\left(\frac{\tilde{c}_1 - \rho_1 t}{\sqrt{1 - \rho_1^2}}\right) dt, \tag{4}$$

where $\tilde{c}_1 = \frac{\sigma_{01}}{\sigma_{11}}\left(c_1 - \frac{\omega_1\sqrt{n_1}}{\sigma_{01}}\right)$, and $\tilde{c} = \frac{\sigma_{02}}{\sigma_{12}}\left(c - \frac{\omega_2\sqrt{n_2}}{\sigma_{02}}\right)$.

Abbreviations

ESS: Expected sample size; ETSL: Expected total study length; PET: Probability of early termination; TIE: Type I error

Acknowledgment

We would like to thank Dr. Jianrong Wu and Dr. Lisa Belin for sharing their R codes with us. Authors would like to thank Associate Editor and two referees, for their valuable comments and suggestions that helped to improve this manuscript.

Funding

Shan's research is partially supported by grants from the National Institute of General Medical Sciences from the National Institutes of Health: P20GM109025. Zhang's work is supported by the Zhejiang Provincial Natural Science Foundation of China (grant no. LY19F020003) and the National Natural Science Foundation of China (grant no. 61672459).

Availability of data and materials

Not applicable. This is a manuscript to develop novel statistical approaches, therefore, no real data is involved.

Authors' contributions

The idea for the paper was originally developed by GS. GS and HZ computed the required sample size for a two-stage design with a survival endpoint. GS and HZ drafted the manuscript and approved the final version.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 26 July 2018 Accepted: 26 February 2019

Published online: 03 April 2019

References

1. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*. 1989;10(1):1–10.
2. Fleming TR. One-sample multiple testing procedure for phase II clinical trials. *Biometrics*. 1982;38(1):143–51.
3. Shan G, Wilding GE, Hutson AD, Gerstenberger S. Optimal adaptive two-stage designs for early phase II clinical trials. *Stat Med*. 2016;35(8):1257–66. <https://doi.org/10.1002/sim.6794>.
4. Shan G, Zhang H, Jiang T. Minimax and admissible adaptive two-stage designs in phase II clinical trials. *BMC Med Res Methodol*. 2016;16(1):90. <https://doi.org/10.1186/s12874-016-0194-3>.
5. Shan G. Exact confidence limits for the response rate in two-stage designs with over- or under-enrollment in the second stage. *Stat Methods Med Res*. 2018;27(4):1045–55.
6. Shan G, Hutson AD, Wilding GE. Two-stage k-sample designs for the ordered alternative problem. *Pharmaceut Statist*. 2012;11(4):287–94. <https://doi.org/10.1002/pst.1499>.

7. Wilding GE, Shan G, Hutson AD. Exact two-stage designs for phase II activity trials with rank-based endpoints. *Contemp Clin Trials*. 2012;33(2): 332–41. <https://doi.org/10.1016/j.cct.2011.10.008>.
8. Shan G, Wilding GE, Hutson AD. Computationally Intensive Two-Stage Designs for Clinical Trials. In: Balakrishnan N, Colton T, Everitt B, Piegorisch W, Ruggeri F, Teugels JL, editors. *Wiley StatsRef: Statistics Reference Online*; 2017. p. 1–7. <https://doi.org/10.1002/9781118445112.stat07986>.
9. Case DD, Morgan TM. Design of Phase II cancer trials evaluating survival probabilities. *BMC Med Res Methodol*. 2003;3:6. <https://doi.org/10.1186/1471-2288-3-6>.
10. Berry DA. Adaptive clinical trials: the promise and the caution. *J Clin Oncol*. 2011;29(6):606–9. <https://doi.org/10.1200/jco.2010.32.2685>.
11. Shan G, Chen JJ. Optimal inference for Simon's two-stage design with over or under enrollment at the second stage. *Commun Stat Simul Comput*. 2017;1–11. <https://doi.org/10.1080/03610918.2017.1307398>.
12. Shan G, Wang W. Exact one-sided confidence limits for Cohen's kappa as a measurement of agreement. *Stat Methods Med Res*. 2017;26(2):615–32. <https://doi.org/10.1177/0962280214552881>.
13. Feldman DR, Patil S, Trinos MJ, Carouso M, Ginsberg MS, Sheinfeld J, Bajorin DF, Bosl GJ, Motzer RJ. Progression-free and overall survival in patients with relapsed/refractory germ cell tumors treated with single-agent chemotherapy: endpoints for clinical trial design. *Cancer*. 2012;118(4):981–6.
14. Lin DY, Shen L, Ying Z, Breslow NE. Group sequential designs for monitoring survival probabilities. *Biometrics*. 1996;52(3):1033–41.
15. Kwak M, Jung S-HH. Phase II clinical trials with time-to-event endpoints: optimal two-stage designs with one-sample log-rank test. *Stat Med*. 2014;33(12):2004–16.
16. Belin L, De Rycke Y, Broët P. A two-stage design for phase II trials with time-to-event endpoint using restricted follow-up. *Contemp Clin Trials Commun*. 2017;8:127–34.
17. Wu J. Sample size calculation for the one-sample log-rank test. *Pharm Stat*. 2015;14(1):26–33.
18. Huang B, Talukder E, Thomas N. Optimal Two-Stage Phase II Designs with Long-Term Endpoints. *Stat Biopharm Res*. 2010;2(1):51–61.
19. Whitehead J. One-stage and two-stage designs for phase II clinical trials with survival endpoints. *Stat Med*. 2014;33(22):3830–43.
20. Takiguchi Y, Moriya T, Asaka-Amano Y, Kawashima T, Kurosu K, Tada Y, Nagao K, Kuriyama T. Phase II study of weekly irinotecan and cisplatin for refractory or recurrent non-small cell lung cancer. *Lung Cancer (Amst, Neth)*. 2007;58(2):253–9.
21. Shan G, Ma C. Unconditional tests for comparing two ordered multinomials. *Stat Methods Med Res*. 2016;25(1):241–54. <https://doi.org/10.1177/0962280212450957>.
22. Zhang H, Shan G. Letter to the Editor: A novel confidence interval for a single proportion in the presence of clustered binary outcome data (SMMR, 2019). *Stat Methods Med Res*. 2019. <https://doi.org/10.1177/0962280219840056>.
23. Shan G, Kang L, Xiao M, Zhang H, Jiang T. Accurate unconditional p-values for a two-arm study with binary endpoints. *J Stat Comput Simul*. 2018;88(6):1200–10.
24. Shan G, Zhang H, Jiang T. Efficient confidence limits for adaptive one-arm two-stage clinical trials with binary endpoints. *BMC Med Res Methodol*. 2017;17(1):22. <https://doi.org/10.1186/s12874-017-0297-5>.
25. Shan G, Banks S, Miller JB, Ritter A, Bernick C, Lombardo J, Cummings JL. Statistical advances in clinical trials and clinical research. *Alzheimer Dement (NY)*. 2018;4:366–71.
26. Shan G. Exact confidence limits for the probability of response in two-stage designs. *Statistics*. 2018;52(5):1086–95. <https://doi.org/10.1080/02331888.2018.1469023>.
27. Shan G. *Exact Statistical Inference for Categorical Data*, 1st edn. San Diego: Academic Press; 2015. <http://www.worldcat.org/isbn/0081006810>.
28. Wilding GE, Consiglio JD, Shan G. Exact approaches for testing hypotheses based on the intra-class kappa coefficient. *Stat Med*. 2014;33(17):2998–3012. <https://doi.org/10.1002/sim.6135>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

