Optimal two-stage designs based on restricted mean survival time for a single-arm study

Guogen Shan

Department of Epidemiology and Biostatistics, School of Public Health, University of Nevada Las Vegas, Las Vegas, NV, 89154, USA

ABSTRACT

Restricted mean survival time is an alternative measure of treatment effect to hazard ratio in clinical trials with time-to-event outcome. The current methods have been focused on one-stage designs. In this article, we propose optimal two-stage designs for a single-arm study with the smallest expected sample size. We compare the performance of the new optimal two-stage designs with the existing one-stage design with regards to the expected sample size and the expected total study length. The simulation results indicate that the new two-stage designs can save the expected sample size substantially as compared to the one-stage design. We use a non-small cell lung cancer trial to illustrate the application of the proposed designs. The proposed optimal two-stage designs are recommended for use when time for patient accrual is longer than the pre-specified follow-up time.

1. Introduction

Restricted mean survival time (RMST) is an alternative measure of treatment effect to hazard ratio (HR) for a study with time-to-event data [1–3]. RMST is calculated as the area under the survival curve from time 0 to a pre-specified time τ (e.g., 1-year) [4]. Thus, it can be interpreted as the expected survival time for a patient being followed up to τ [5,6]. Although HR is traditionally used in designing clinical trials, its statistical inference depends on the assumption of proportional hazards over time. It was reported that almost 25% of cancer studies did not meet the proportional hazards assumption [7], but the HR-based models were still used in data analysis. In early phase cancer clinical trials with survival outcome as the primary endpoint, treatment effects are often compared by using the survival rates at the pre-specified time. The survival rate at τ provides information at that particular time, while RMST captures the survival information from baseline to τ with a better and comprehensive understanding of the treatment effect on average survival.

Recently, Trinquart et al. [7] compared statistical inferences using HR and the difference of RMST with 54 randomized cancer clinical trials selected from five leading oncology journals. They found that both measures generally agree with each other with regards to the group difference using HR or RMST [8,9]. They also reported that the computed HR is often larger than the ratio of RMST from those cancer studies. Later, Huang and Kuan [10] used simulation studies to compare the performance of the HR-based inference and the RSMT test [11]. Their simulation results indicated that the log-rank test is generally more

powerful than the RSMT test when the proportional hazard assumption is satisfied [8,12]. However, when that assumption is not met, the RSMT test is preferable.

In early phase cancer clinical trials, single-arm studies are widely used to quickly investigate the activity of a new treatment as compared to the treatment effect estimated from historical data. Current designs based on RMST have been focused on one-stage designs (e.g., the R package *ssrmst*) [1,4,13,14]. In the case when a new treatment under investigation is not as effective as expected, a study should be allowed for futility stopping earlier. Therefore, there is a need to develop new optimal single-arm two-stage designs based on RMST to investigate the activity of a new treatment. The new two-stage designs in this article are based on the same criteria as Simon's two-stage designs for binary outcome [15–21], having the smallest expected sample size under the null hypothesis [22–24]. In a two-stage design, investigators could temporarily suspend the patient enrollment after the enrollment of the last patient at the first stage, and wait for time τ to observe the time-to-event outcome of each participant. A two-stage design without interim accrual could be used when τ is short. Alternatively, investigators may consider the optimal two-stage designs with interim accrual to save study time [9,15,25].

The rest of the article is organized as follows. In Section 2, we provide the detailed steps to compute sample sizes for optimal two-stage designs based on RMST. In Section 3, we compare the performance of the existing one-stage design with the proposed two-stage designs with regards to the expected sample size and the expected total study length. Then, a non-small cell lung cancer trial is used as historical data to cal-
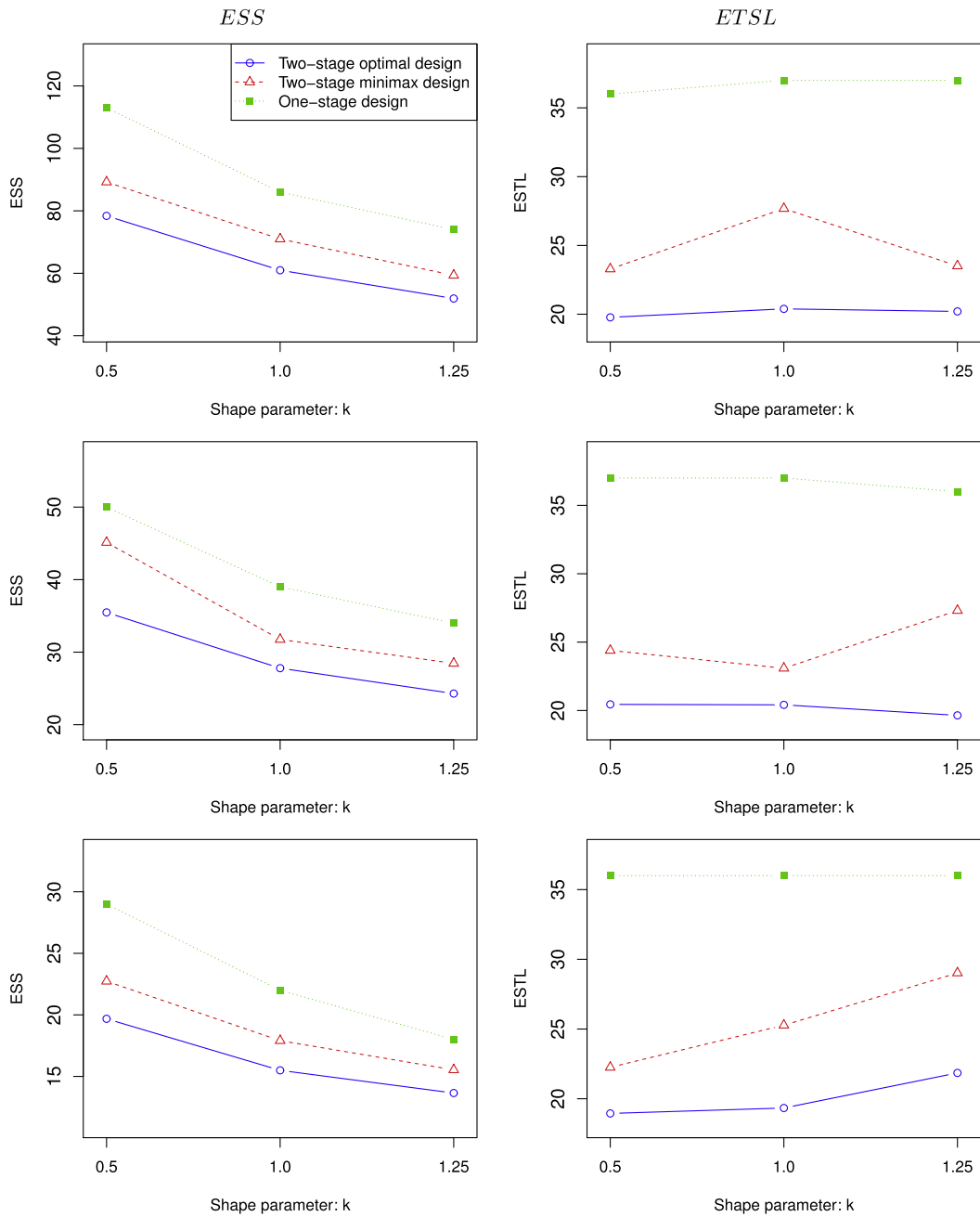
ESS

ETSL

Fig. 1. The ESS and the ETSL of the proposed two-stage designs as compared to the one-stage design, when $\mu_0 (\tau = 12) = 0.5\tau = 6$ months and the alternative expected survival months 0.60τ (top), 0.65τ (middle), and 0.70τ (bottom).

culate sample size for optimal two-stage designs. Lastly, we provide some comments in Section 4.

2. Methods

When the patient accrual time T_a is relatively longer than the restricted follow time (τ), a two-stage design can potentially save sample sizes and costs as compared to the one-stage design. Let $\mu(\tau)$ be the RMST with the pre-specified follow-up time of τ. In this article, we propose developing new optimal two-stage designs based on RMST in a single-arm study setting. Suppose the estimated RMST from previous studies is $\mu_0(\tau)$, and a new treatment is expected to have a larger RMST as $\mu_1(\tau)$, where $\mu_1(\tau) > \mu_0(\tau)$. Then, the hypotheses to test the effectiveness of a new treatment are presented as

$$H_0 : \mu(\tau) \leq \mu_0(\tau) \quad \text{against} \quad H_a : \mu(\tau) > \mu_0(\tau),$$

where the statistical power of a study is computed at $\mu_1(\tau)$.

For a study with N participants, we assume that they are uniformly enrolled during the accrual time $[0, T_a]$, where $T_a = N/\theta$, and θ is the patient accrual rate. Let E_i and C_i be the event time and the right censoring time for the i-th participant, respectively, where $i = 1, 2, \cdots,$ and N. The censoring data are $\Delta_i = I (E_i \leq C_i)$.

For a one-stage design, a study is ended when all participants are being followed by τ. Upon the completion of the study with the total study time of $T_{one} = T_a + \tau$, the observed time for the i-th participant is: $O_i = \min (E_i, C_i)$. It follows that the RMST $\mu(\tau)$ after the completion of a study can be estimated as:
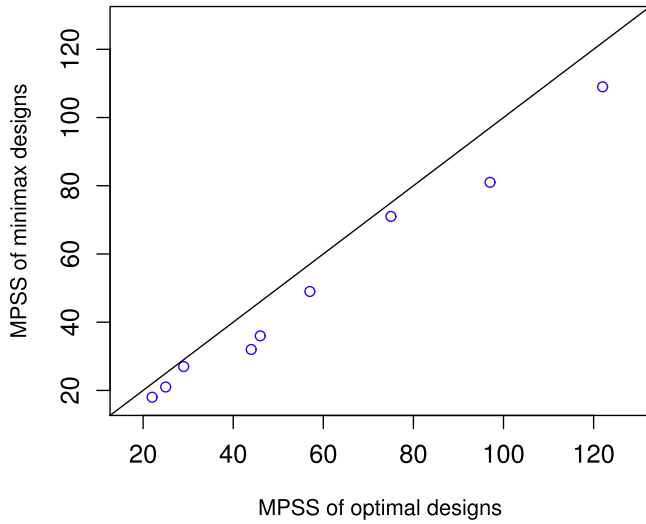
Fig. 2. The MPSS of the optimal two-stage designs and the MPSS of the minimax two-stage designs for the 9 configurations in Fig. 1 when $\mu_0(\tau = 12) = 0.5\tau = 6$ months.

$$\hat{\mu}(\tau) = \int_0^\tau \hat{S}(t)\,dt, \tag{1}$$

where $S(t)$ is the Kaplan–Meier (KM) estimator [26] for the survival function of T based on the observed data $\{(O_i, \Delta_i), i = 1, 2, \cdots, \text{and } N\}$. The $\hat{\mu}(\tau)$ can be computed by using the *survRM2* package from the statistical software R [27]. Specifically, the *rmst_1* function is used from the *survRM2* package for a single-arm study. The developed R software program is available per request from the author.

For a two-stage design, we propose two commonly used optimal designs which control for the type I error and power:

(1) the optimal two-stage design with the smallest expected sample size (ESS) under the null hypothesis;
(2) the minimax two-stage design having the smallest ESS among the ones with the smallest maximum possible sample size (MPSS).

These two optimal designs are commonly utilized in cancer clinical trials, and often known as the optimal design and the minimax design, respectively. For a study with time-to-event endpoint, the outcome is a long-term measure. Optimal two-stage designs with interim accrual which could significantly reduce study time. Meanwhile, designs without interim accrual are able to observe the time-to-event outcome of all participants at the interim analysis.

When the first stage time T_1 is too short (e.g., $T_1 < \tau$), the computed area under the survival curve is from 0 to T_1, not from 0 to τ. Then, that RMST estimator from the first stage can not be considered as an estimator for RMST in Equation (1). Therefore, the first stage time T_1 must be longer than τ, such as 1.1τ.

For a study with N_1 patients in the first stage, the total time in the first stage is $N_1/\theta + \tau$ for a study without interim accrual, and its RMST at the first stage is estimated as

$$\hat{\mu}_1(\tau) = \int_0^\tau \hat{S}_1(t)\,dt,$$

where $S_1(t)$ is the Kaplan–Meier estimator for the survival function of T based on $\{(O_i, \Delta_i), i = 1, 2, \cdots, \text{and } N_1\}$.

For a study with interim accrual, participants enrolled close to T_1 may be censored due to a short follow-up time. Suppose v_i is the entering time of the i-th participant. Then, the observed time for the i-th participant is $O_i^b = \min\left(E_i, C_i, \max\left(0, T_1 - v_i\right)\right)$, and the censoring outcome

is $\Delta_i^b = I\left(E_i \leq C_i\right) I\left(E_i + v_i \leq T_1\right)$. Then, its RMST at the first stage for a study with interim accrual is estimated as $\int_0^\tau \hat{S}_1^b(t)\,dt$, where $S_1^b(t)$ is the KM estimator for the survival function of T based on $\{(O_i^b, \Delta_i^b), i = 1, 2, \cdots, \text{and } N_1\}$.

To compute the required sample size for a study with design parameters (τ, $\mu_0(\tau)$, $\mu_1(\tau)$, α, β, and κ), the very first step is to determine the scale parameters of the Weibull distribution, where κ is the shape parameter of the Weibull distribution. The scale parameter of the Weibull distribution under the null hypothesis, λ_0, can be determined by solving the equation $\hat{\mu}(\tau) = \mu_0(\tau)$ in a one-stage design setting, where $S(t)$ in the $\hat{\mu}(\tau)$ follows the Weibull distribution with the shape parameter of κ and the scale parameter λ_0. Similarly, the scale parameter under the alternative hypothesis λ_1 is solved from the equation $\hat{\mu}(\tau) = \mu_1(\tau)$.

For a one-stage optimal design, data are simulated from the Weibull distributions with κ as the shape parameter, and λ_0 and λ_1 as scale parameters. For each simulated data, we use the *survRM2* package to calculate $\hat{\mu}(\tau)$ which is based on the KM estimator, and $\hat{\mu}(\tau)$ is used as the test statistic for type I error and power calculation. The $(1-\alpha)$ upper quantile of all the test statistics from the null data is the threshold to control for the type I error. Power is then computed as the proportion of data under the alternative distribution whose test statistics are above that threshold. The optimal one-stage design is defined as the one with the smallest N and the estimated power above $1 - \beta$. Suppose N_{one} is the required sample size for a one-stage design.

We search for optimal two-stage designs with the MPSS N from $0.9N_{one}$ to $1.2N_{one}$. In practice, the sample size boundaries should be adjusted when an optimal two-stage design is identified near the boundary of the sample size space. For a given N, the lower bound of the first stage sample size N_1 is set as $1.1\tau\theta$, where the constant 1.1 is added to make sure at least 10% participants being followed by τ. The upper bound of N_1 is set as $N - 1$. For each given N_1 and N, the test statistics $\hat{\mu}_1(\tau)$ for the first stage and $\hat{\mu}(\tau)$ for both stages combined, are computed under the null and alternative hypotheses. The type I error rate is calculated as

$$TIE(r_1, r) = P\left(\hat{\mu}_1(\tau) \geq r_1 \text{ and } \hat{\mu}(\tau) \geq r | H_0\right),$$

where r_1 and r are the thresholds for the first stage and both stages combined, respectively. We use the range of $\hat{\mu}_1(\tau)$ for the possible r_1 values (uniformly distributed), and a similar approach is used to choose the r values from the computed $\hat{\mu}(\tau)$ values. For any set of (r_1, r) with $TIE(r_1, r) < \alpha$, we compute power of a two-stage design as:

$$Power(r_1, r) = P\left(\hat{\mu}_1(\tau) \geq r_1 \text{ and } \hat{\mu}(\tau) \geq r | H_a\right).$$

If the computed power is above the nominal level $(1-\beta)$, its associated sample sizes and threshold values (N_1, N, r_1, r) are saved as a candidate for the desired optimal design. The ESS under the null hypothesis is computed as

$$ESS = N_1 + (1 - PET)(N - N_1),$$

where $PET = P\left(\hat{\mu}_1(\tau) < r_1 | H_0\right)$ is the probability of early termination (PET) at the first stage due to futility. The ETSL for a study with interim accrual is calculated as

$$ETSL = \frac{N_1}{\theta} + (1 - PET)\left(\frac{N - N_1}{\theta} + \tau\right), \tag{2}$$

and the ETSL for a study without interim accrual is

$$ETSL = \left(\frac{N_1}{\theta} + \tau\right) + (1 - PET)\left(\frac{N - N_1}{\theta} + \tau\right). \tag{3}$$
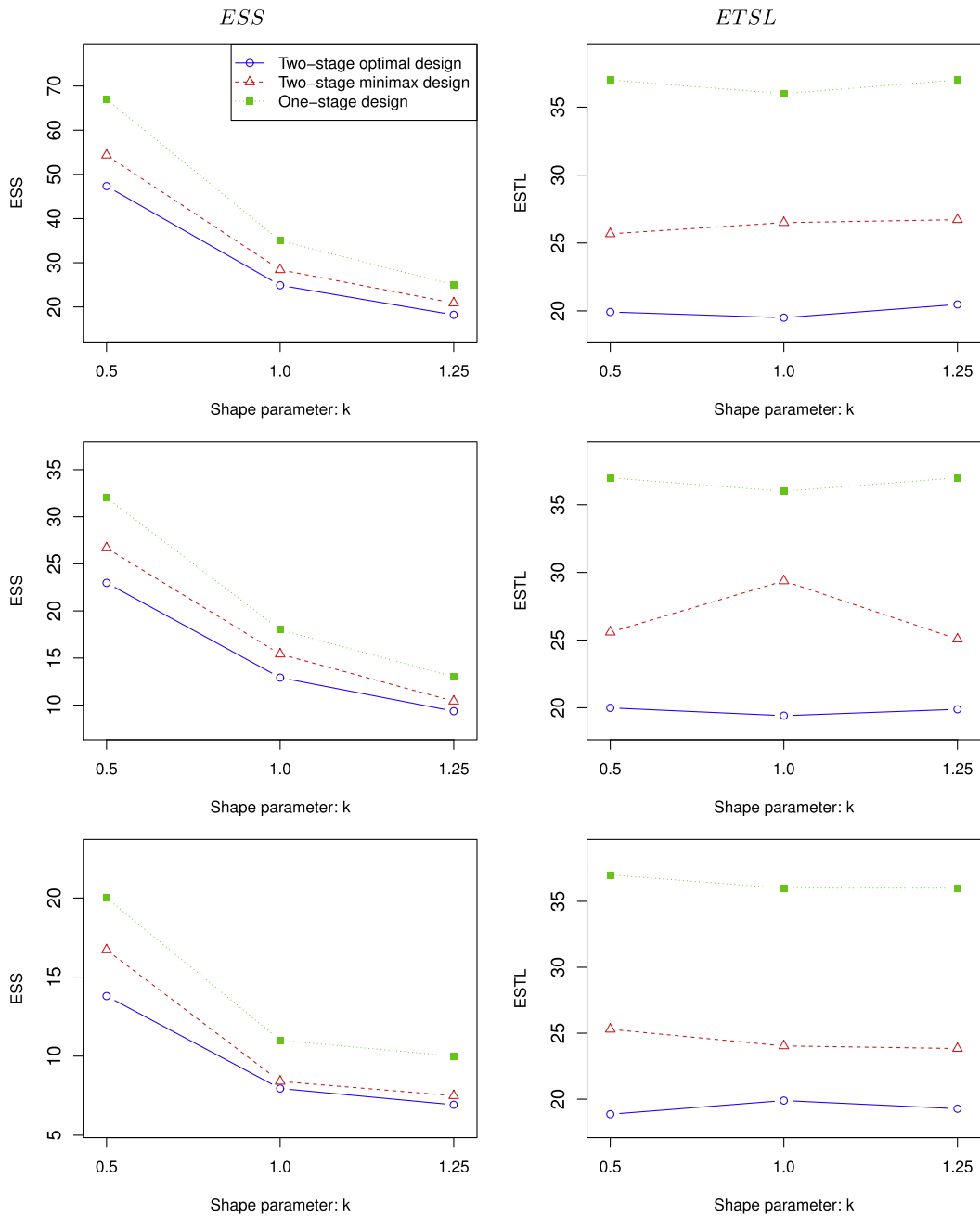
Fig. 3. The ESS and the ETSL of the proposed two-stage designs as compared to the one-stage design, when $\mu_0 (\tau = 12) = 0.2\tau = 2.4$ months and the alternative expected survival months 0.30τ (top), 0.35τ (middle), and 0.40τ (bottom).

It should be noted that the PETs in Equation (2) and Equation (3) are computed using different $\hat{\mu}_1 (\tau)$ distributions. The ESS and the ETSL are highly correlated: a study with a large ESS often has a long ETSL. When the PET is high, the computed ESS is close to the first stage sample size and the ETSL is close to T_1 for the first stage.

3. Results

We compare the performance of the one-stage design and the two-stage optimal designs with regards to the ESS and the ETSL under the null hypothesis. The considered RMST values under the null hypothesis are: $\mu_0 (\tau) = 0.2\tau$, 0.5τ, and 0.7τ, where the restricted follow-up time τ is set as 12 months. The difference between $\mu_1 (\tau)$ and $\mu_0 (\tau)$ is 0.10τ, 0.15τ, and 0.20τ. The shape parameter of the Weibull distribution is assumed to be the same under the null and the alternative hypotheses,

with the values of κ =0.5, 1, and 1.25. The exponential distribution is a special case of the Weibull distribution with $\kappa = 1$. The type I error rate is set as $\alpha = 0.05$, and the nominal level of power is 80% ($\beta = 0.2$). Since the null hypothesis is rejected for a large RMST value, it is a one-sided test with $\alpha = 0.05$.

We use Monte Carlo simulation studies to generate random samples to identify the designs that meet both the type I error and power requirements as described in the Methods section. Fig. 1 shows the one-stage design and the two-stage designs when the RMST is $\mu_0 (12) = 0.5\tau = 6$ months. The required one-stage sample sizes decrease as the shape parameter of the Weibull distribution κ increases. The ESS of two-stage optimal designs is much smaller than the sample size of the one-stage design, with savings from 24% to 32% having the average saving of 29%. The ESS of the minimax two-stage design is between the optimal two-stage design and the one-stage design, and it saves the ESS
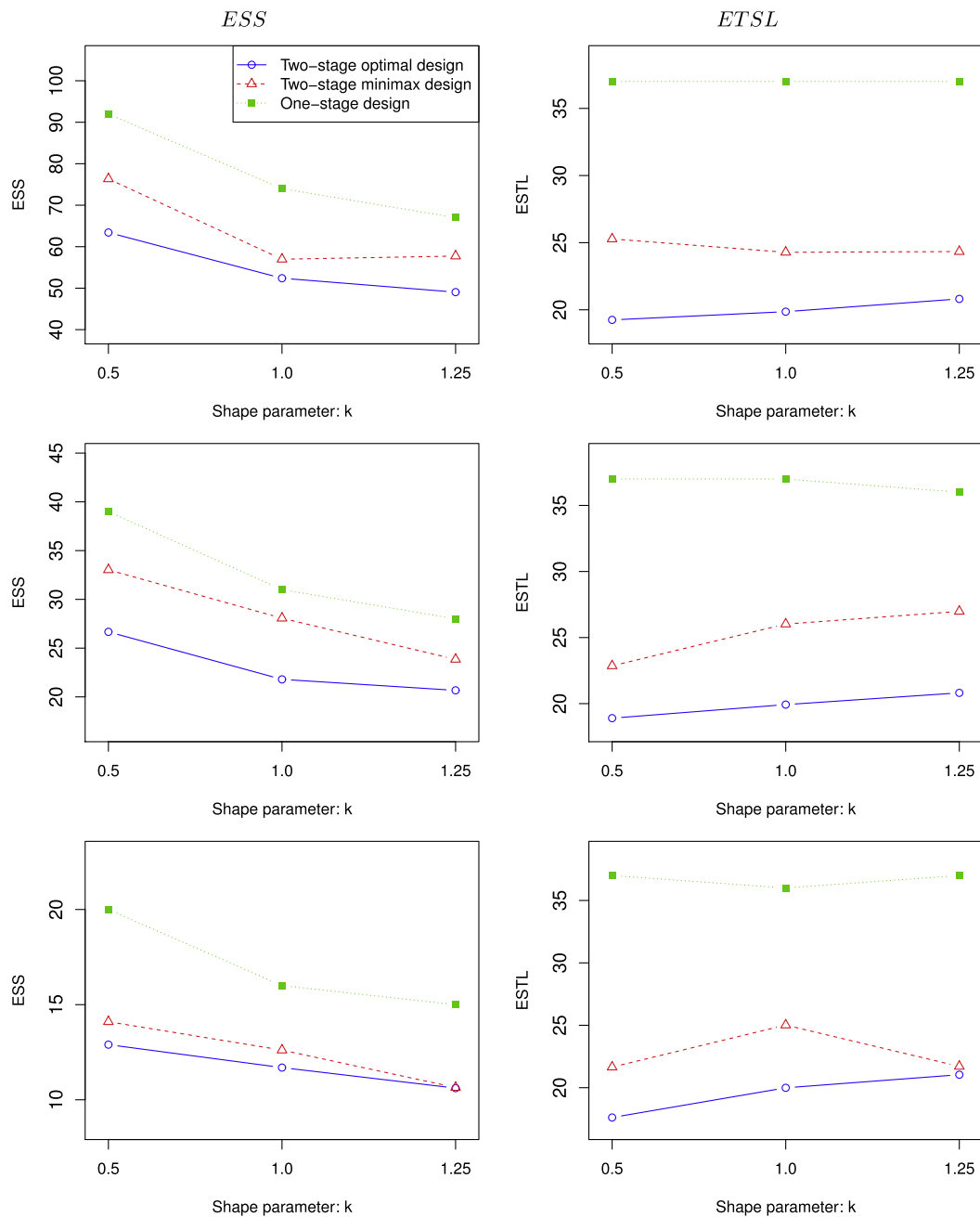
Fig. 4. The ESS and the ETSL of the proposed two-stage designs as compared to the one-stage design, when $\mu_0 (\tau = 12) = 0.7\tau = 8.4$ months and the alternative expected survival months 0.80τ (top), 0.85τ (middle), and 0.80τ (bottom).

in the average of 17% with the range from 10% to 22% as compared to the one-stage design. The savings of the ETSL for the proposed two-stage designs are similar to the ESS savings.

We present the MPSS of the optimal and minimax two-stage designs in Fig. 2 using the 9 configurations from Fig. 1 when $\mu_0 (12) = 0.5\tau = 6$ months. It can be seen that the MPSS of the minimax designs is generally smaller than that of the optimal designs, with the average saving of 15%. We observe similar results for the ESS and the ETSL comparisons between the one-stage design and the two-stage designs in Fig. 3 when $\mu_0 (12) = 0.2\tau = 2.4$ months, and in Fig. 4 when $\mu_0 (12) = 0.7\tau = 8.4$ months.

We use a phase II refractory or recurrent non-small cell lung cancer trial [28] to further study the proposed single-arm two-stage optimal designs. Data from the phase II study reported by Takiguchi et al. [28] are used as pilot data to estimate the RMST given $\tau = 12$ months. The

RMST is estimated as $\mu_0 (\tau) = 0.32\tau$, which can be calculated by using the digital software, Engauge Digitizer, to capture the measures on the progression-free survival (PFS) curve. Suppose we have a new treatment for refractory or recurrent non-small cell lung cancer with the estimated RMST of 0.42τ which is $0.1\tau = 1.2$ months longer than the treatment with second-line chemotherapy consisting of weekly irinotecan and cisplatin in the reported phase II study. The calculated one-stage sample size is 62 to attain 80% power at $\alpha = 0.05$ based on RMST. Alternatively, one may use the HR model for sample size calculation, and it is computed as 60 patients for the one-stage design [8].

We assume that $N_{one} = 62$ patients are enrolled uniformly over the 2-year time period. The optimal and minimax two-stage designs are presented in Table 1 for studies with interim accrual. The expected sample sizes of optimal two-stage designs are much less than that of the one-stage design: ESS = 34.10 for the optimal design and ESS = 38.30

Table 1
Optimal two-stage designs with interim accrual based on the pilot data from the non-small cell lung cancer trial, with the design parameters: $\tau = 12$ months, $\mu_0(\tau) = 0.32\tau = 3.8$ months, $\mu_1(\tau) = 0.42\tau = 5.0$ months, $\alpha = 0.05$, and 80% power. The sample size for the one-stage design is $N_{one} = 62$ patients.

r_1	r	N_1	N	PET	ESS	ETSL
Patient accrual time: 2 years						
5.54	4.55	34	67	0.99	34.10	13.25
5.18	4.59	38	60	0.99	38.30	15.01
Patient accrual time: 3 years						
5.35	4.54	23	70	0.98	24.01	14.21
5.07	4.60	24	58	0.96	25.53	15.38
Patient accrual time: 5 years						
5.07	4.50	14	67	0.91	18.63	19.14
4.26	4.58	29	55	0.74	35.64	37.66

Table 2
Optimal two-stage designs without interim accrual based on the pilot data from the non-small cell lung cancer trial, with the design parameters: $\tau = 12$ months, $\mu_0(\tau) = 0.32\tau = 3.8$ months, $\mu_1(\tau) = 0.42\tau = 5.0$ months, $\alpha = 0.05$, and 80% power. The sample size for the one-stage design is $N_{one} = 62$ patients.

r_1	r	N_1.	N	PET	ESS	ETSL
Patient accrual time: 2 years						
4.31	4.44	34	72	0.80	41.60	30.52
4.07	4.58	40	60	0.68	46.34	33.76
Patient accrual time: 3 years						
4.23	4.43	25	77	0.73	38.87	37.80
4.09	4.61	41	58	0.68	46.39	42.77
Patient accrual time: 5 years						
4.18	4.46	23	71	0.70	37.40	51.91
4.19	4.59	48	54	0.78	49.30	62.47

for the minimax design. The ETSL is shorter in the optimal design as compared to the minimax design. As expected, the MPSS of the minimax design is less than that of the optimal design. Given $N_{one} = 62$ patients in the one-stage design, we also investigate the effect of patient accrual time on the ESS and the ETSL, with $T_a = 3$ years or 5 years in Table 1. It can be seen that the ESS of the optimal design decreases as T_a is longer. When a study has a longer accrual time, the lower bound of the first stage sample size becomes smaller which leads to a smaller ESS of the optimal design. But the ETSL is getting longer in general. We notice that the PETs of these designs are very high for this example. In the aforementioned simulation studies, the PET is near 75% for the configurations in Fig. 1 with $\mu_0(\tau) = 6$ months.

We also present the optimal and miniax two-stage designs with no interim patient accrual for this example in Table 2. When a study is designed with no interim patient accrual, patient enrollment is temporarily suspended when N_1 patients at the first stage are enrolled. In this case, data analysis is going to be conducted at the time $T_1 + \tau$ to follow all patients by τ to observe the time-to-event outcome. The ETSL is calculated by using Equation (3). As expected, designs without interim accrual in Table 2 have longer expected study lengths than the designs with interim accrual in Table 1. The ESS of the optimal design is consistent as the patient accrual time T_a increases, although the ETSL becomes longer.

4. Discussion

For a two-stage design, when the first stage sample size N_1 is very close to the upper bound $(N - 1)$, the sample size difference between the two-stage optimal designs and the one-stage design is negligible. On the other side, when the first stage sample size N_1 is very close to the lower boundary and the PET is high, its ESS is small. For this reason, it could be a practically useful to design a two-stage design with a new constraint on the sample size ratio between the two stages:

$N_1/(N - N_1)$, to potentially reduce the issue of the obtained designs near the parameter boundaries.

In the simulation studies, we add the restriction of the time for the first stage with at least 10% more than τ. Given a small to medium sample size in an early phase clinical trial, it is important to measure the treatment effect with as many completed cases as possible. For this reason, we would suggest researchers to provide optimal or minimax designs with multiple t_1 restrictions to investigators (e.g., $1.2\,\tau$, $1.4\,\tau$). The developed R software program can be modified to search for the optimal designs by adding that constraint. In the discussed lung cancer example, when we increase the percentage of patients completed the follow-up (e.g., $1.3\tau\theta$, $1.5\tau\theta$), the optimal two-stage designs are often identified with N_1 being very close to the lower boundary, and the first stage sample size of the minimax designs towards the lower boundary as that percentage increases. As pointed out of the reviewers, the proposed two-stage designs can be applied to studies whose patient accrual time T_a is longer than the follow-up time τ. When a study's follow-up time is very long (e.g., trials for Alzheimer's Disease [29,30]), it is possible that all patients are already enrolled while no patient has been followed by τ. In such cases, other new two-stage designs should be developed.

Conditional power after the first stage may be considered as an alternative criteria to the first stage threshold r_1 for futility stopping at the end of the first stage. Jennison and Turnbull [31] reviewed several approaches for conditional power calculation for group sequential designs. Conditional power is computed under the assumption that future patients are similar to the already enrolled patients. A study is stopped early for futility due to a low conditional power. When only a few patients complete the follow-up, the information from existing patients may not be reliable enough to test that important assumption.

When the estimated RMST from historical data is not reliable (e.g., a very few studies with different populations), a two-arm or multiple-arm randomized study may be utilized to compare the new treatment(s) and the control. Uno et al. [1,27] developed one-stage designs for a randomized clinical trial comparing one treatment with the control based on RMST. It would be a straightforward extension of their one-stage designs to a study with multiple arms. However, it is computational intensive to extend their work to randomized two-stage designs with two or more arms. We consider this as future work. Another interesting topic would be adaptive designs based on RMST, where the second stage sample size and the censoring distribution can be modified using the observed results from the first stage [21]. In order to develop such adaptive designs, effective search algorithms are critical to search for the optimal adaptive designs [14,32].

Acknowledgments

References

[1] H. Uno, J. Wittes, H. Fu, S.D. Solomon, B. Claggett, L. Tian, et al., Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies, Ann. Intern. Med. 163 (2) (2015 jul) 127–134.

[2] L. Tian, L. Zhao, L.J. Wei, Predicting the restricted mean event time with the subject's baseline covariates in survival analysis, Biostatistics 15 (2) (2014 apr) 222–233.

[3] C. Zhang, Y. Wu, G. Yin, Restricted mean survival time for interval-censored data, Stat. Med. 39 (26) (2020 aug) 3879–3895 Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8699.

[4] H. Uno, M. Horiguchi, M.J. Hassett, Statistical test/estimation methods used in contemporary phase III cancer randomized controlled trials with time-to-event outcomes, Oncol. 25 (2) (2020 feb) 91–93.

[5] K.L. Calkins, C.E. Canan, R.D. Moore, C.R. Lesko, B. Lau, An application of restricted mean survival time in a competing risks setting: comparing time to ART initiation by injection drug use, BMC Med. Res. Methodol. 18 (1) (2018 mar) 27 Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0484-z.

[6] P. Royston, M.K.B. Parmar, Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome, BMC Med. Res. Methodol. 13 (1) (2013 dec).

[7] L. Trinquart, J. Jacot, S.C. Conner, R. Porcher, Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials, J. Clin. Oncol. 34 (15) (2016 may) 1813–1819 Available from: http://www.ncbi.nlm.nih.gov/pubmed/26884584.

[8] G. Shan, H. Zhang, Two-stage optimal designs with survival endpoint when the follow-up time is restricted, BMC Med. Res. Methodol. 19 (74) (2019) https://doi.org/10.1186/s12874-019-0696-x.

[9] G. Shan, H. Zhang, T. Jiang, Adaptive two-stage optimal designs for Phase II clinical studies that allow early futility stopping, Seq. Anal. 38 (2) (2019) 199–213 (In press), https://doi.org/10.1080/07474946.2019.1611307.

[10] B. Huang, P.F. Kuan, Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event end point, Pharmaceut. Stat. 17 (3) (2018 may) 202–213 Available from: http://www.ncbi.nlm.nih.gov/pubmed/29282880.

[11] G. Shan, E. Bayram, J.Z.K. Caldwell, J.B. Miller, J.J. Shen, S. Gerstenberger, Partial correlation coefficient for a study with repeated measurements, Stat. Biopharm. Res. (2020) In press. Available from: https://www.tandfonline.com/doi/abs/10.1080/19466315.2020.1784780.

[12] G. Shan, Two-stage optimal designs based on exact variance for a single-arm trial with survival endpoints, J. Biopharm. Stat. 30 (5) (2020 mar) 797–805 Available from: http://www.ncbi.nlm.nih.gov/pubmed/32129130.

[13] I.R. Weir, L. Trinquart, Design of non-inferiority randomized trials using the difference in restricted mean survival times, Clin. Trials 15 (5) (2018 oct) 499–508.

[14] G. Shan, G.E. Wilding, A.D. Hutson, S. Gerstenberger, Optimal adaptive two-stage designs for early phase II clinical trials, Stat. Med. 35 (8) (2016 apr) 1257–1266 Available from:, https://doi.org/10.1002/sim.6794.

[15] G. Shan, A.D. Hutson, G.E. Wilding, Two-stage k-sample designs for the ordered alternative problem, Pharmaceut. Stat. 11 (4) (2012) 287–294 Available from:, https://doi.org/10.1002/pst.1499.

[16] G. Shan, C. Ma, A.D. Hutson, G.E. Wilding, An efficient and exact approach for detecting trends with binary endpoints, Stat. Med. 31 (2) (2012) 155–164 Available from:, https://doi.org/10.1002/sim.4411.

[17] G. Shan, C. Ma, A.D. Hutson, G.E. Wilding, Randomized two-stage phase II clinical trial designs based on barnard's exact test, J. Biopharm. Stat. 23 (5) (2013 aug) 1081–1090 Available from:, https://doi.org/10.1080/10543406.2013.813525.

[18] G. Shan, G.E. Wilding, A.D. Hutson, Computationally Intensive Two-Stage Designs for Clinical Trials, Wiley StatsRef: Statistics Reference Online, 2017, pp. 1–7 Available from:, https://doi.org/10.1002/9781118445112.stat07986.

[19] G. Shan, Exact Statistical Inference for Categorical Data, first ed., Academic Press, San Diego, CA, 2015 Available from: http://www.worldcat.org/isbn/0081006810.

[20] G. Shan, A better confidence interval for the sensitivity at a fixed level of specificity for diagnostic tests with continuous endpoints, Stat. Methods Med. Res. 26 (1) (2017 feb) 268–279 Available from: http://view.ncbi.nlm.nih.gov/pubmed/25049287.

[21] G. Shan, Exact confidence limits for the response rate in two-stage designs with over or under enrollment in the second stage, Stat. Methods Med. Res. 27 (4) (2018 apr) 1045–1055 Available from: http://view.ncbi.nlm.nih.gov/pubmed/27389669.

[22] G. Shan, C. Dodge-Francis, G.E. Wilding, Exact Unconditional Tests for Dichotomous Data when Comparing Multiple Treatments with a Single Control, Therapeutic Innovation & Regulatory Science, 2020, pp. 411–417 https://doi.org/10.1007/s43441-019-00070-w.

[23] G. Shan, Accurate confidence intervals for proportion in studies with clustered binary outcome, Stat. Methods Med. Res. 29 (10) (2020 oct) 3006–3018 Available from: http://journals.sagepub.com/doi/10.1177/0962280220913971.

[24] J. Gronsbell, C. Hong, L. Nie, Y. Lu, L. Tian, Exact inference for the random-effect model for meta-analyses with rare events, Stat. Med. 39 (3) (2020 sep) 252–264 Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/sim.8500https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8500https://onlinelibrary.wiley.com/doi/10.1002/sim.8500.

[25] G. Shan, W. Wang, Advanced statistical methods and designs for clinical trials for COVID-19, Int. J. Antimicrob. Agents 57 (1) (2021 jan) 106167 Available from: https://pubmed.ncbi.nlm.nih.gov/33408032/.

[26] E.L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc. 53 (282) (1958) 457–481.

[27] H. Uno, Vignette for survRM2 Package: Comparing Two Survival Curves Using the Restricted Mean Survival Time, 2017 Available from: https://cran.r-project.org/web/packages/survRM2/vignettes/survRM2-vignette3-2.html.

[28] Y. Takiguchi, T. Moriya, Y. Asaka-Amano, T. Kawashima, K. Kurosu, Y. Tada, et al., Phase II study of weekly irinotecan and cisplatin for refractory or recurrent non-small cell lung cancer. Lung cancer (Amsterdam, Netherlands), Lung Canc. 58 (2) (2007 nov) 253–259 Available from: http://view.ncbi.nlm.nih.gov/pubmed/17658654.

[29] J. Cummings, G. Lee, A. Ritter, M. Sabbagh, K. Zhong, Alzheimer's disease drug development pipeline, Alzheimer's Dementia: Transl. Res. Clin. Interv. 5 (2019) 272–293 2019 jan.

[30] G. Shan, S. Banks, J.B. Miller, A. Ritter, C. Bernick, J. Lombardo, et al., Statistical advances in clinical trials and clinical research. Alzheimer's & Dementia, Transl. Res. Clin. Interv. 4 (2018) 366–371.

[31] C. Jennison, B.W. Turnbull, Group Sequential Methods (Chapman & Hall/CRC Interdisciplinary Statistics), first ed., Chapman and Hall/CRC, 1999 Available from: http://www.worldcat.org/isbn/0849303168.

[32] G. Shan, Accurate confidence intervals for proportion in studies with clustered binary outcome, Stat. Methods Med. Res. 29 (10) (2020 apr) 3006–3018 Available from: http://journals.sagepub.com/doi/10.1177/0962280220913971.