



Sample size re-estimation without un-blinding for time-to-event outcomes in oncology clinical trials

Lihong Huang^{1,Δ}, Jianling Bai^{1,Δ}, Hao Yu¹, Feng Chen^{1,2,✉}

¹Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 211166, China;

²Ministry of Education Key Laboratory for Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 211166, China.

Abstract

Sample size re-estimation is essential in oncology studies. However, the use of blinded sample size reassessment for survival data has been rarely reported. Based on the density function of the exponential distribution, an expectation-maximization (EM) algorithm of the hazard ratio was derived, and several simulation studies were used to verify its applications. The method had obvious variation in the hazard ratio estimates and overestimation for the relatively small hazard ratios. Our studies showed that the stability of the EM estimation results directly correlated with the sample size, the convergence of the EM algorithm was impacted by the initial values, and a balanced design produced the best estimates. No reliable blinded sample size re-estimation inference can be made in our studies, but the results provide useful information to steer the practitioners in this field from repeating the same endeavor.

Keywords: oncology study, clinical trial, sample size re-estimation, expectation-maximization algorithm

Introduction

Given the life-threatening nature and the unmet medical needs of many types of cancer, the drug development process in the field of oncology should be accelerated. To this end, many clinical trial approaches have been proposed. The sample size required for a clinical trial should be sufficiently large to provide a reliable answer to the questions addressed^[1], and the method by which the sample size is calculated should be provided in the protocol. This method is the most basic requirement for planning all studies because it is critical to the success of a study and pertains to budget considerations. For fixed sample size designs, there is a

risk that expected trial outcomes may not obtain adequate power because some uncertainty is usually associated with the parameters in the planning phase. This deficit can be remedied by re-estimating the parameters during the interim analysis and modifying the initially planned sample size if necessary^[2-3].

Although various methods have been proposed and used for sample size, a basic debate remains: Should the interim data be examined with the treatment group in a blinded or un-blinded manner? Regulatory authorities certainly favor blinded sample size reassessment (BSSR) (CPMP Working Party on Efficacy of Medicinal Products^[4]; ICH-E9 Expert Working Group^[1]; European Medicines Agency/Committee for Medicinal

^ΔThese authors contributed equally to this study.

[✉]Corresponding author: Feng Chen, PhD, SPH building Room 412, 101 Longmian Avenue, Nanjing, Jiangsu 211166, China. Fax: +86 25-86868435, E-mail: fengchen@njmu.edu.cn.

Received 06 September 2016, Revised 08 November 2016, Accepted 20 January 2017, Epub 15 March 2017

CLC number: R311, Document code: A

The authors reported no conflict of interests.

This is an open access article under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited.

Products for Human Use^[5]) because it better preserves study integrity. Fortunately, the work of Gould and Shih^[6] has shown that un-blinding is not necessary to efficiently estimate within-group variance.

Nevertheless, previous studies of sample size re-estimation addressed various statistical and practical aspects of this approach, such as BSSR in non-inferiority and equivalence trials with normally distributed outcome variables and hypotheses formulated in terms of the ratio and difference of means^[7], BSSR in multi-armed clinical trials when the outcome variable is normally distributed^[8], BSSR with negative binomial counts in superiority and non-inferiority trials^[9], and BSSR with count data in multiple sclerosis^[10], etc.

However, significantly longer progression-free survival (PFS), overall survival (OS) and time to progression (TTP) are well known to be widely used in oncology studies as primary endpoints to evaluate the efficacy of treatment. Thus, the response variable of most oncology clinical trials is survival time. To the best of our knowledge, the use of BSSR in this context has rarely been reported. Here, we tried to develop an EM approach for BSSR with exponentially distributed outcomes. The performance and applicability of this procedure are described based on several simulation studies.

Materials and methods

Derived EM algorithm

The sample-size formula for an oncology clinical trial can be simplified if it is expressed as the number of deaths required rather than the number of patients. Suppose that a two-sided test will be performed with a significance level of $\alpha/2$ and a power of $1-\beta$ for a hazard ratio Δ . Let $z_{1-\alpha/2}$ and $z_{1-\beta}$ be the $1-\alpha/2$ and $1-\beta$ percentiles of the normal distribution, respectively, and let P_A and P_B be the proportion of the patients randomized to treatments A and B, respectively. Then, the total number of deaths required is given by the following expression^[11]:

$$(Z_{1-\alpha/2} + Z_{1-\beta})^2 / (P_A P_B \ln^2 \Delta) \tag{1}$$

In general, only Δ in the above formula is unknown and needs to be estimated based on previous studies. However, if Δ is much larger or smaller than estimated, the sample size will need to be re-estimated without breaking the randomization codes. Gould and Shih proposed a EM algorithm-based procedure for blinded variance estimation for normally distributed endpoints^[6,12]; we extended this EM algorithm for Δ estimation to exponentially distributed endpoints

without un-blinding the treatment group at the interim stage.

As the treatments are not identified, any interim observation, x_i , $i = 1, \dots, n$, could be in either treatment group such that the treatment assignments are "missing at random". Let τ_i denote the treatment group indicator, e.g., $\tau_i = 1(0)$ indicates that sample i is in treatment group 1 (group 2); τ_1, \dots, τ_n are independent random variables with $P(\tau_i = 1) = \theta$. The density function of the exponential distribution is $f(t) = \lambda e^{-\lambda t}$, and λ is the scale parameter. Given τ_i , x_i ($i = 1, \dots, n$) is distributed as follows, with density

$$f(x_i | \tau_i, \lambda_1, \lambda_2) = \lambda_1^{\tau_i} \lambda_2^{(1-\tau_i)} \exp[-\tau_i \lambda_1 x_i - (1-\tau_i) \lambda_2 x_i] \tag{2}$$

Therefore, the expression for the expected τ_i given x_i is

$$\begin{aligned} P(\tau_i = 1 | x_i) &= \frac{P(\tau_i = 1)P(x = x_i)}{P(\tau_i = 0)P(x = x_i) + P(\tau_i = 1)P(x = x_i)} \\ &= \frac{\theta \lambda_1 \exp(-\lambda_1 x_i)}{\theta \lambda_1 \exp(-\lambda_1 x_i) + (1-\theta) \lambda_2 \exp(-\lambda_2 x_i)} \\ &= \frac{1}{1 + \frac{1-\theta}{\theta} \frac{\lambda_2}{\lambda_1} \exp(\lambda_1 - \lambda_2) x_i} \end{aligned} \tag{3}$$

Then, the log-likelihood of the interim observations is

$$\begin{aligned} L(\lambda_1, \lambda_2 | x_i, \tau_i) &= \sum_{i=1}^n \{ \tau_i \ln \lambda_1 + (1-\tau_i) \ln \lambda_2 + [-\tau_i \lambda_1 x_i - (1-\tau_i) \lambda_2 x_i] \} \end{aligned} \tag{4}$$

After taking the partial derivative of the above log-likelihood, the maximum likelihood estimates of λ_1/λ_2 are

$$\Delta = \lambda_1 / \lambda_2 = \frac{\sum_{i=1}^n \tau_i}{\sum_{i=1}^n (1-\tau_i)} / \frac{\sum_{i=1}^n \tau_i x_i}{\sum_{i=1}^n (1-\tau_i) x_i} \tag{5}$$

The EM algorithm for estimating λ_1/λ_2 proceeds as follows. For example, if θ is assumed to be 0.5, then the "E" step consists of substituting the "initial" estimates of λ_1 and λ_2 into formula (3) to obtain provisional values for the expected value of τ_i . "M" consists of obtaining maximum likelihood estimates of λ_1/λ_2 according to (5) after replacing τ_i in (5) with the provisional expectations. The "E" and "M" steps are repeated until the value of λ_1/λ_2 stabilizes.

Simulations

We use simulations to evaluate performances of the proposed procedures. Specifically, various Δ estimators were compared for different scenarios. The censoring was not considered in the following simulations because the derived EM algorithm was used to re-estimate the number of events. For the EM algorithm, the recursive computation was continued until successive estimates of λ_1 and λ_2 differed by less than 0.001. Besides, we also designed a simulated clinical trial to investigate the power and illustrate BSSR procedure.

Firstly, the sample size requirement of the above EM algorithm was investigated. We considered the situation of equal samples for each distribution, population parameters were set as $\lambda_1 = 0.1, 0.12, 0.15, 0.2, 0.25, 0.3$ and $\lambda_2 = 0.1$, and several scenarios for the number of events per group differed from 5 to 800. The initial values were set to 1.0 and 0.8. In addition, the complete proportion of the total sample size in interim analysis is also an impact factor related to sample size. We defined $\Delta = 0.66, 2.0, 3.0$ to do investigation. It was set $\alpha = 0.05$, power = 90%, we calculated the number of events using (1) for each scenario with 100% complete proportion, and the estimated values are listed in column "n" in the first line of **Table 1**. The other "n" cells are calculated according to the defined complete proportion; for example, the 80% complete proportion means 80% of the total number of events are completed in interim stage, n is 103 for scenario $\lambda_1 = 0.10$ and $\lambda_2 = 0.15$ which is calculated by $129 \times 80\%$.

Secondly, as the negligible impact of initialization on the EM estimation procedure, we investigated its effects by calculating the Δ estimates according to the above EM procedure for 1,000 repeated simulations with different values of the initialization constants of λ^*_1 and λ^*_2 . We selected equal number of events per group ($n = 50$) for the internal pilot study.

Thirdly, the balanced design had been considered in most sample size estimation studies. Here, we investigated the impact of the sample allocation ratio on the above EM procedure. It was set with $\alpha = 0.05$, power =

90%, and we calculated the number of events using (1) for each allocation ratio and the estimated values listed in column "n" of **Table 2**. As some cells of "n" had the n_1 or n_2 values which were smaller than 20, considering the impact of the sample size, the number of events was expanded to at least 50 for one of the groups to obtain column "n*" in **Table 2**. Δ was then re-estimated based on "n" and "n*", respectively.

In a simulation study, we investigated the power and BSSR procedure with dummy randomized clinical trials in which the survival data of the two treatment groups were compared. We obtained the independent and identically distributed time-to-event observations from exponential distributions and the censored data from uniform distribution.

In our design, an initial sample size (n_0) was carried out based on assumed parameters. Although this sample constituted the final sample size for the fixed design, the sample size could be adjusted using BSSR based on half of the initially sample size, i.e., 50% information time. The following parameter values were considered: assumed median survival time $T = 10$ and 5 for the treatment group and the control group, respectively; therefore, the scale parameters λ of the exponential distributions were calculated to be 0.07 and 0.14 based on the formula $\lambda = \log(2)/T$. Two different scenarios were considered for the true median survival time, $T' = 10, 5$ and $T' = 12, 5$. Furthermore, the design was balanced and the significance level and target power were the usual $\alpha = 0.05$ (two-sided) and $1 - \beta = 0.9$.

The number of failures for the fixed design was estimated to be $45^{[13-14]}$ per group based on the assumed parameters, and the internal pilot study included 23 failures per group for an information time of 50%. The assumed censor rates were 0, 20%, and 40%, and then the corresponding event rates were 100%, 80%, and 60%, respectively. Accordingly, these values resulted in fixed sample sizes (n_0) of 45, 56, and 75, respectively, based on sample size = number of events/(event rate). The true parameters were used to generate simulated data with a fixed sample size (n_0). According to EM algorithm, the blinded re-estimated

Table 1 EM re-estimation of Δ with different completed proportions on 1,000 runs

Completed proportion	$\lambda_1 = 0.10, \lambda_2 = 0.15$			$\lambda_1 = 0.2, \lambda_2 = 0.1$			$\lambda_1 = 0.3, \lambda_2 = 0.1$		
	n	Δ Mean	Δ SD	n	Δ Mean	Δ SD	n	Δ Mean	Δ SD
100%	129	0.6661	0.15923	45	2.0259	0.80409	18	3.0880	1.82815
80%	103	0.6730	0.16825	36	2.0375	0.87898	14	3.1139	2.05301
60%	77	0.6811	0.18731	27	2.0468	0.96885	11	3.2532	2.57707
40%	52	0.6877	0.20607	18	2.1623	1.38557	7	3.5356	3.86211
20%	26	0.7027	0.2418	9	2.4899	2.50656	4	5.0993	9.37600

λ_1 and λ_2 are exponential distribution parameters, Δ is λ_1/λ_2 .

Table 2 EM re-estimation of Δ with different sample allocation ratios on 1,000 runs

Parameters	Allocation ratio	n	Δ Mean	Δ SD	n^*	Δ Mean	Δ SD
$\lambda_1 = 0.2, \lambda_2 = 0.1$	3:1	$n_1 = 87, n_2 = 29$	1.9439	0.69021	$n_1 = 150, n_2 = 50$	1.9611	0.5554
	2:1	$n_1 = 66, n_2 = 33$	1.9666	0.70763	$n_1 = 100, n_2 = 50$	1.9779	0.60863
	1:1	$n_1 = 45, n_2 = 45$	2.0259	0.80409	$n_1 = 50, n_2 = 50$	2.0291	0.78344
	1:2	$n_1 = 33, n_2 = 66$	2.1372	1.0208	$n_1 = 50, n_2 = 100$	2.0970	0.84075
	1:3	$n_1 = 29, n_2 = 87$	2.2181	1.19095	$n_1 = 50, n_2 = 150$	2.1460	0.90141
$\lambda_1 = 0.3, \lambda_2 = 0.1$	3:1	$n_1 = 36, n_2 = 12$	2.8674	1.34612	$n_1 = 150, n_2 = 50$	2.9581	0.68016
	2:1	$n_1 = 28, n_2 = 14$	2.9109	1.38883	$n_1 = 100, n_2 = 50$	2.9712	0.77198
	1:1	$n_1 = 18, n_2 = 18$	3.0880	1.82815	$n_1 = 50, n_2 = 50$	3.0060	1.03945
	1:2	$n_1 = 14, n_2 = 28$	3.3216	2.46425	$n_1 = 50, n_2 = 100$	3.0630	1.10128
	1:3	$n_1 = 12, n_2 = 36$	3.6878	3.96318	$n_1 = 50, n_2 = 150$	3.0660	1.16034

λ_1 and λ_2 are exponential distribution parameters, Δ is λ_1/λ_2 , n^* is expanded sample size.

sample size (n_1) was based on the 23 (the half of 45) events from an assumed internal pilot study which is defined as 50% information time of the integrity trial^[15]. Subsequently, 1,000 trails were simulated from each scenario.

Results

Sample size requirement

Fig. 1 shows the EM re-estimation of Δ for different number of events; 1,000 simulation replications were performed for each situation. The estimation results stabilized as the number of events increased. All results were overestimated when the number of events per group was less than 30. The smaller Δ , the higher sample size was required. More than 100 sample sizes was needed when Δ was 1.5. For $\Delta = 1.0$, there was about 16.05% overestimation even if sample size reached 800, this result was similar with Xie's research (13.54%)^[16].

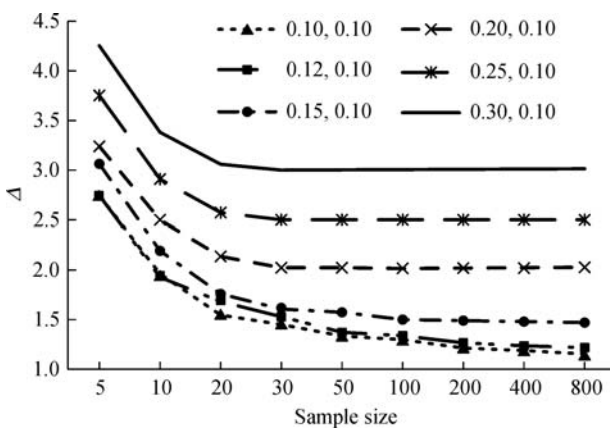


Fig. 1 EM re-estimation of Δ with different sample sizes

Table 1 also shows that the estimates were impacted by event numbers (sample size). For $\lambda_1 = 0.10$ and $\lambda_2 = 0.15$, the estimates were all acceptable even if the completed proportion was 20%, which nearly satisfied the sample size requirement of this EM algorithm. On the contrary, the estimates significantly deviated from the real values for $\lambda_1 = 0.3$ and $\lambda_2 = 0.1$ if the completed proportion was less than 80% due to the number of events was insufficient for the EM algorithm.

Initial values

The simulation results are presented in Table 3. Specifically, the estimates obtained from the EM procedure depended on the initialization. It shows that the Δ estimates exceeded 1 if $\lambda^*_1 > \lambda^*_2$ and were less than 1 if $\lambda^*_2 < \lambda^*_1$. Fortunately, the estimated Δ values were very close to the true values when the initial values satisfied $\lambda^*_1 > \lambda^*_2$ with the true values $\lambda_1 > \lambda_2$, and $\lambda^*_1 < \lambda^*_2$ with $\lambda_1 < \lambda_2$. The estimated number of the required events using (1) was equal for both $\Delta = \lambda_1/\lambda_2$ and $\Delta = \lambda_2/\lambda_1$. Therefore, the choice of $\lambda^*_1 > \lambda^*_2$ or $\lambda^*_1 < \lambda^*_2$ did not affect the re-estimation of the number of events. Table 3 also shows the overestimation results for $\Delta = 1$ ($\lambda_1 = 0.2, \lambda_2 = 0.2$), and the choice of initial values did not affect these overestimated results.

Sample allocation ratio

Table 2 results indicate that the EM estimates of Δ were impacted by the sample allocation ratio. The balanced design produced the best estimates. The unbalanced designed estimates, including 2:1, 1:2, 3:1 and 1:3, were slightly larger or smaller, and the sample size difference between the two groups directly correlated with the estimated values. More unbalance would get more biased estimation although the number of events increased to at least 50 for one group in "n*" column.

Table 3 EM re-estimation of Δ with different initial values on 1,000 runs

Initial values	$\lambda_1 = 0.2, \lambda_2 = 0.1$		$\lambda_1 = 0.1, \lambda_2 = 0.2$		$\lambda_1 = 0.2, \lambda_2 = 0.2$		$\lambda_1 = 0.3, \lambda_2 = 0.1$		$\lambda_1 = 0.1, \lambda_2 = 0.3$	
	Δ Mean	Δ SD	Δ Mean	Δ SD	Δ Mean	Δ SD	Δ Mean	Δ SD	Δ Mean	Δ SD
$\lambda_1^* = 1.0, \lambda_2^* = 0.2$	2.0299	0.78558	2.0118	0.75331	1.3324	0.51644	3.0169	1.05123	2.9939	0.99592
$\lambda_1^* = 0.8, \lambda_2^* = 0.2$	2.0301	0.78542	2.0121	0.75315	1.3197	0.51084	3.0169	1.05125	2.9938	0.99575
$\lambda_1^* = 0.4, \lambda_2^* = 0.2$	2.0283	0.78143	2.0109	0.75004	1.3494	0.54675	3.0020	1.03824	2.9794	0.98333
$\lambda_1^* = 0.2, \lambda_2^* = 0.4$	0.5694	0.21396	0.5704	0.21015	0.8178	0.22699	0.3814	0.15811	0.3810	0.15433
$\lambda_1^* = 0.2, \lambda_2^* = 0.8$	0.5692	0.21413	0.5703	0.21036	0.8093	0.22995	0.3803	0.15873	0.3799	0.15491
$\lambda_1^* = 0.2, \lambda_2^* = 1.0$	0.5693	0.21428	0.5705	0.21049	0.8260	0.22439	0.3803	0.15869	0.3799	0.15504

λ_1 and λ_2 are exponential distribution parameters, Δ is λ_1/λ_2 , λ_1^* and λ_2^* are initial values.

Example (simulated)

Table 4 shows the statistical power tested by the log-rank test and exponential regression for fixed design and sample sizes re-estimated adjusted design. For the true median survival times of $T = 12$ and 5 , the original sample sizes were 45 for a 100% event rate, 56 for an 80% event rate and 75 for a 60% event rate, as shown in column " n_0 ". Thus, these sample sizes were overfull to obtain 90% power for the log-rank test or exponential regression. Therefore, the power of the log-rank test and exponential regression exceeded 90% significantly. After a Δ re-estimation using the above EM algorithm, the number of events was re-calculated. Based on the pre-defined event rates (60% , 80% and 100%), the corresponding sample sizes were re-calculated and were shown in column " n_1 ". From the output, the re-estimated sample sizes " n_1 " were near the actual requirement, the power for the adjusted design was closer to 90% for a 100% event rate.

For true median survival times of $T = 10$ and 5 , " n_0 " was the correct required sample size, and the values of re-estimated sample sizes " n_1 " were similar to the " n_0 " values; the power for both fixed design and adjusted design was close to 90% for a 100% event rate. The variation details of the simulated EM estimated values of λ_1 and λ_2 for each scenario are presented in **Fig. 2**.

Notably, the power for both the log-rank test and exponential regression inversely correlated with the event rate in each scenario because the sample size calculation was based on the event numbers/(event rate). Lower event rates required larger samples, which increased the power of the test.

Discussion

The topic of sample size reassessment during an ongoing trial has become very popular in recent years. The inflation of the type I error rate and the loss of power has long been an intractable problem for sample

size adjustment in an internal pilot study with adaptive design. In the ICH E9 guidelines, it is reflected by the following requirement for planned sample size adjustment: "The step taken to preserve blindness and consequences, if any, for the type I error and the width of confidence intervals should be explained". The calculation of the actual type I error rate for the blinded case was previously derived for the t -test situation^[17], and the results showed that the nominal inflation in the type I error rate did not significantly differ for the blinded sample size recalculation in the unrestricted design. Over the last decade, a number of studies had addressed the problem of BSSR for an ongoing clinical trial. These documents noted that the assumptions of sample size calculation should be reassessed with the blinded data and that the effect on the type I error rate can be controlled^[18]. To our knowledge, this work is restricted to the normally or binomially distributed data.

Cancer is currently one of the major diseases affecting human health, and anti-cancer drug becomes the focus of research in the pharmaceutical industry. Survival time-related outcomes such as PFS and OS are generally used as primary endpoints in oncology studies to assess efficacy. Thus, oncology trials usually require long follow-up periods and include a planned interim analysis, and sample size re-estimation is also essential. In this paper, we tried to extend BSSR method for exponentially distributed survival data based on EM algorithm.

The blinded data of an ongoing trial is from a mixture of two populations; one for the treatment group and the other for the control group. EM method described here is applied to estimate model parameters of the mixture distributions and therefore assess the hazard ratio. The derived EM estimation only can be used to re-estimate the number of events for oncology trials, and the censoring is ignored. Normally, it is acceptable in the clinical trial because the censor rate is always unknown and only assumed in the design stage. With the re-

Table 4 Simulated power for fixed and adjusted designs

Scenario	True parameters					Event rate	n_0	Fixed design power (%)		Re-estimated		Adjusted design power (%)	
	λ_1	λ_2	T_1	T_2	Δ'			LogRank	Exponential regression	Δ	n_1	LogRank	Exponential regression
1						100%	45	89.7	89.3	0.512	46	90.2	91.6
2	0.07	0.14	10	5	0.5	80%	56	92.1	92.9	0.510	58	92.5	92.1
3						60%	75	94.8	94.0	0.510	77	95.5	94.9
4						100%	45	98.0	98.2	0.439	31	91.1	93.2
5	0.06	0.14	12	5	0.42	80%	56	99.3	99.5	0.430	36	93.4	94.0
6						60%	75	99.6	99.5	0.448	55	97.2	98.9

λ_1 and λ_2 are exponential distribution parameters, T_1 and T_2 are true median survival times, Δ' is true hazard ratio, Δ is re-estimated hazard ratio, n_0 is original sample size, n_1 is re-estimated sample size.

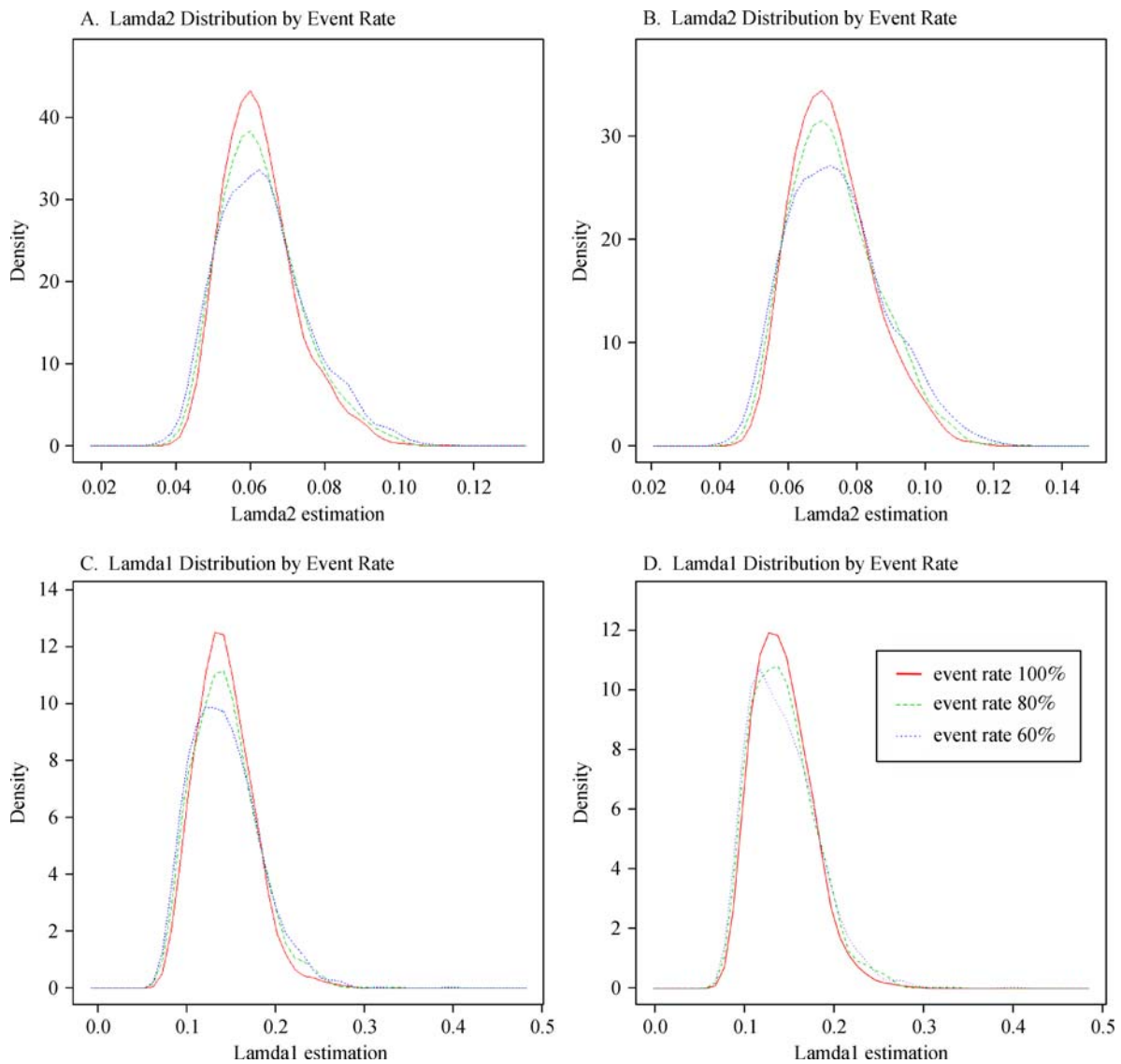


Fig. 2 Kernel Density of λ_1 and λ_2 estimation. A and C are for Scenario 1-3 in Table 4, B and D are for Scenario 4-6 in Table 4.

estimated number of events in the interim stage, the total censor rate of the two treatments which is available without breaking blind, could be used to do sample size re-estimation, because sample size = the re-estimated number of events/(1-interim stage censor rate). On the other hand, only exponential distribution has been considered in our research, and Weibull distribution needs to be investigated in the future since it is more widely applied in survival analysis.

Our studies show that estimates from the EM method are highly variable, which coincides with the literature on interim analysis of treatment effects with blinded data and a normally distributed endpoint. Besides, the relatively small hazard ratios ($\Delta < 1.2$) are overestimated according to this EM method.

Specifically, for the hazard ratios ($\Delta \geq 1.2$), the EM estimation shown here is subject to a sample size requirement and the stability of the estimation results directly correlated with the number of events. The initial values for the parameters directly affect the convergence of the algorithm and the estimated results. The simulation shows that the initial values should be carefully selected, and the calculation of the optimum initial values requires further research. Moreover, the EM estimation described here is more suitable for a balanced design.

Due to the variation and overestimation of smaller hazard ratios, no reliable inference can be made on sample size re-estimation in our studies. The results from this paper provide useful information to steer the practitioners in this field from repeating the same endeavor. This should be of some relief to health authorities.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (81273184), and the National Natural Science Foundation of China Grant for Young Scientists (81302512). We would like to thank the referees for their comments that greatly helped us improve the manuscript.

References

- [1] ICH. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group[J]. *Stat Med*, 1999, 18(15): 1905–1942.
- [2] Posch M, Bauer P. Interim analysis and sample size reassessment[J]. *Biometrics*, 2000, 56(4): 1170–1176.
- [3] Proschan MA. Sample size re-estimation in clinical trials[J]. *Biom J*, 2009, 51(2): 348–357.
- [4] CPMP. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products. CPMP Working Party on Efficacy of Medicinal Products Note for Guidance III/3630/92-EN[J]. *Stat Med*, 1995, 14(15): 1659–1682.
- [5] EMEA/CHMP. Reflection Paper on Methodological Issues in Confirmatory Clinical Trials with Flexible Design and Analysis Plan[J]. Available online at <http://www.emea.eu.int>. 2006.
- [6] Gould AL, Shih WJ. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance[J]. *Commun Stat Theory Methods*, 1992, 21(10): 2833–2853.
- [7] Friede T, Kieser M. Blinded sample size reassessment in non-inferiority and equivalence trials[J]. *Stat Med*, 2003, 22(6): 995–1007.
- [8] Kieser M, Friede T. Blinded sample size Reestimation in multiarmed clinical trials[J]. *Drug Inf J*, 2000, 34(2): 455–460.
- [9] Friede T, Schmidli H. Blinded sample size reestimation with negative binomial counts in superiority and non-inferiority trials[J]. *Methods Inf Med*, 2010, 49(6): 618–624.
- [10] Friede T, Schmidli H. Blinded sample size reestimation with count data: methods and applications in multiple sclerosis[J]. *Stat Med*, 2010, 29(10): 1145–1156.
- [11] Schoenfeld DA. Sample-size formula for the proportional-hazards regression model[J]. *Biometrics*, 1983, 39(2): 499–503.
- [12] Chang M. Adaptive design theory and implementation using SAS and R[M]. Boca Raton: Chapman & Hall/CRC, 2008.
- [13] Desu MM, Raghavarao D. Sample size methodology[M]. Boston: Academic Press; 1990.
- [14] Bain LJ, Engelhardt M. Statistical analysis of reliability and life-testing models: theory and methods[M]. 2nd ed. New York: M. Dekker; 1991.
- [15] Birkett MA, Day SJ. Internal pilot studies for estimating sample size[J]. *Stat Med*, 1994, 13(23-24): 2455–2463.
- [16] Xie J, Quan H, Zhang J. Blinded assessment of treatment effects for survival endpoint in an ongoing trial[J]. *Pharm Stat*, 2012, 11(3): 204–213.
- [17] Friede T, Kieser M. Sample size recalculation in internal pilot study designs: a review[J]. *Biom J*, 2006, 48(4): 537–555.
- [18] Kieser M, Friede T. Simple procedures for blinded sample size adjustment that do not affect the type I error rate[J]. *Stat Med*, 2003, 22(23): 3571–3581.