Int J Clin Exp Med 2017;10(1):1164-1171
www.ijcem.com /ISSN:1940-5901/IJCEM0041634

Original Article

Correctness of efficacy estimation in early termination clinical trial

Lihong Huang[1], Fang Shao[1], Hao Yu[1], Jianling Bai[1], Feng Chen[1,2], Ping Yu[3]

[1]Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, P. R. China; [2]Ministry of Education Key Laboratory for Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing, P. R. China; [3]Jingan District Center for Disease Control and Provention, Shanghai, P. R. China

Received October 11, 2016; Accepted October 25, 2016; Epub January 15, 2017; Published January 30, 2017

Abstract: Background: Early termination of a clinical trial is well accepted when there is enough evidence for a significant effect based on an interim analysis; however, an interim analysis does have its risks. The main purpose of this study is to examine whether efficacy can be estimated correctly in early termination stage through formula derivation and simulation methods. Methods: We derived inequalities for the relationship between the estimated effects and the actual effects for normal and binary endpoints. Several simulations were also provided to support our findings. Results: The derived inequalities show that the difference between two statistical significant sample means/rates is greater than that between two pre-defined population means/rates. The simulation results show the ratios of the average of the estimated mean/rate differences in the statistical significant early termination stage and the pre-defined effect size (δ_1/δ) are always greater than 1 even if the type I error rate is controlled by Peto, O'Brien-Fleming or Pocock method. Therefore, the actual population benefits may be smaller than the estimation when a clinical trial early stopped for benefits because of significant interim analysis. Conclusions: The effect estimations of both normal and binary endpoints are overestimated in the early termination period. This finding is supported by formula derivation and simulations whose results are comsistent.

Keywords: Clinical trial, early termination, efficacy estimation

Introduction

The drug development process has become a highly competitive and costly enterprise. It is an important option in a clinical trial, if necessary, it can be early stopped. Several factors can influence the decision to stop an ongoing clinical trial including ethical issues, changes in accepted clinical practice that make the continuation of a clinical trial unwise, or reaching a positive or negative statistical endpoint earlier than anticipated. There are a number of negative reasons for discontinuing a trial prematurely [1], such as serious adverse events, inability to recruit or enroll an adequate number of patients, financial considerations, protocol found to be impractical or unworkable, etc. On the contrary, some positive reasons are also realized, for example, the unexpected benefit prompted a recommendation for early termination in order to extend the benefit to a large group of patients as soon as possible. The Beta-blocker Heart Attack Trial (BHAT) in 1988 was the very famous successful early termination clinical trial which was terminated nine months earlier than scheduled due to an observed treatment benefit [2].

In clinical trials, interim analyses are often performed before the completion of the trial, in which the intention is to possibly terminate the trial early. International Conference on Harmonization (ICH) E9 guideline [3] provides the definition of interim analysis, that is any analysis intended to compare treatment arms with respect to efficacy or safety at any time prior to formal completion of a trial. From sponsor's point of view, the utilization of interim analyses is to compare treatment arms, terminate the development of ineffective or unsafe drugs, or accelerate the regulatory approval process in the pharmaceutical industry [4]. To facilitate the application of the interim analysis and early termination, E9 guideline indicates

that the goal of such an interim analysis is to stop the trial early if the superiority of the treatment under study is clearly established, if the demonstration of a relevant treatment difference has become unlikely, if unacceptable adverse effects are apparent, or if sample size re-estimation is needed.

Therefore, early termination after interim analysis for efficacy reason is acceptable from both regulations and sponsors. However, boundaries for monitoring efficacy require more evidence to terminate a trial early than safety monitoring. Armin Koch mentioned that practical consequences of spending tiny portions of α in many interim analyses are too often not appropriately acknowledged [5]. Although there are many methods being applied to define early termination type I error boundary, no data to date is available about their actual impact on the efficacy estimation in early termination stage. It is widely accepted that any decision of terminating a large scale multi-center clinical trial early is a very complex process in which various elements should be considered. Obviously, the actual treatment effect is the most important element.

To the best of our knowledge, there was not much published literature mentioned the efficacy estimation in early terminated clinical trial. Therefore, the main purpose of this study is to examine whether efficacy can be estimated correctly in early termination stage through formula derivation and simulation methods. Inequalities for the relationship between the estimated effects and the actual effects are derived for normal and binary endpoints. Simulation results are also provided to support the findings.

Methods

Normal and binary endpoints are considered in this research. Formula derivation and simulation methods were used to evaluate the treatment effects estimated from early termination clinical trial.

Normal endpoint

Consider the issue of comparing the means of a continuous response to two randomized treatment groups. Let μ_1 and μ_2 (assume $\mu_1 > \mu_2$) be the population means of the two treament groups respectively, $\bar{x}_1$ and $\bar{x}_2$ be the sample means, n_1 and n_2 be the sample sizes, σ be the variance which is equal in two treatment groups.

When σ is unknown, it is known that

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

S_w is the pooled variance. The expectation regarding the difference between two predefined population means and two statistical significant sample means (i.e. satisfy the condition $t \geq t_\alpha$, t_α is the α percentile of the t distribution and t is the statistic of the two sample means comparison), which can be derived as follows:

$$E\left[((\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)) \,|\, t \geq t_\alpha\right]$$

$$= S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, E(t \,|\, t \geq t_\alpha)$$

$$> S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, E(t \,|\, t > 0)$$

$$> S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, E(t),$$

$$\because E(t) = 0, \therefore S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, E(t) = 0.$$

$E\left[((\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)) \,|\, t \geq t_\alpha\right] > 0$, so we get the relationship inequality $E[\bar{x}_1 - \bar{x}_2 \,|\, t \geq t_\alpha] > (\mu_1 - \mu_2)$. It indicates that the difference between two sample means is greater than 2 pre-defined population means under the condition $t \geq t_\alpha$. That means the actual population benefit may be smaller than that in the estimation when a clinical trial early stopped for benefit because of significant interim analysis.

Next, let we discuss this issue in another way. At the clinical trial design stage, when the sample allocation ratio is 1, the per-group sample size, N, is initially planned to detect δ at a level of significance α (two-sided test) and with power 1-β, where δ is the postulated effect size that is thought to be sufficiently conservative. Thus the sample size for each group can be estimated according to the following formula, where $Z_{1-\varepsilon}$ is the (1-ε) × 100 percentile of the standard normal distribution.

$$N = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 \times \sigma^2}{\delta^2}$$

It is assumed that the clinical trial could be early terminated at the time point when the per-group sample size is kN (k is the proportion of sample size, which ranges from 0% to 100%). Suppose that the power (90%) and the variance (σ^2) are consistence with the original study design. Then we can get the following equation:

$$\frac{2 \times (z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\delta_1^2} = kN = k\frac{2 \times (z_{1-0.05/2} + z_{1-\beta})^2 \sigma^2}{\delta^2}$$
$$\Rightarrow \left| \frac{z_{1-\alpha/2} + z_{1-0.1}}{\delta_1} \right| = \left| \frac{\sqrt{k}\,(z_{1-0.05/2} + z_{1-0.1})}{\delta} \right| \quad (1)$$
$$\Rightarrow \delta_1 = \frac{(z_{1-\alpha/2} + 1.282)\delta}{3.242\sqrt{k}}$$

Where δ is the postulated effect size at the design stage, δ is the estimated effect size at the early termination stage, and the two-sided test significance level is 0.05 designed for the original study.

Binary endpoint

For a binary endpoint, let p_1 and p_2 be the sample rates, n_1 and n_2 be the sample sizes, and p_c be the combined sample rate. It is known that the following μ asymptotically obey the standard normal distribution.

$$\mu = \frac{p_1 - p_2}{\sqrt{P_c(1 - P_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

Let π_1 and π_2 be the population rates, and assume $\pi_1 > \pi_2$. Then the relationship of the difference between 2 pre-defined population rates and 2 statistical significant sample rates (i.e. satisfy the condition $\mu \geq \mu_\alpha$, so μ_α is the α percentile of the μ distribution and μ is the statistic of the two sample rates comparison) which can be derived as follows:

$$E[((p_1 - p_2) - (\pi_1 - \pi_2)) \mid \mu \geq \mu_\alpha]$$
$$\approx \sqrt{P_c(1 - P_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\, E(\mu \mid \mu \geq \mu_\alpha)$$
$$> \sqrt{P_c(1 - P_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\, E(\mu \mid \mu > 0)$$
$$> \sqrt{P_c(1 - P_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\, E(\mu)$$

Where $E(\mu)=0$, $\mu_\alpha > 0$.

Therefore, $E[((p_1 - p_2) - (\pi_1 - \pi_2)) \mid \mu \geq \mu_\alpha] > 0$, we get the relationship inequality $E[(p_1 - p_2) \mid \mu \geq \mu_\alpha] > (\pi_1 - \pi_2)$.

Base on the binary endpoint sample size estimation formula, $N = ((z_{1-\alpha/2} + z_{1-\beta})/\delta)^2[\pi_1(1-\pi_1) + \pi_2(1-\pi_2)]$, we can get the following equation according to the same method of the normal endpoint above.

$$\frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta_1^2}[\pi_1(1-\pi_1) + \pi_2(1-\pi_2)] = kN$$
$$= k\frac{(z_{1-0.05/2} + z_{1-\beta})^2}{\delta^2}[\pi_1(1-\pi_1) + \pi_2(1-\pi_2)]$$
$$\Rightarrow \left| \frac{z_{1-\alpha/2} + z_{1-0.1}}{\delta_1} \right| = \left| \frac{\sqrt{k}\,(z_{1-0.05/2} + z_{1-0.1})}{\delta} \right| \quad (2)$$
$$\Rightarrow \delta_1 = \frac{(z_{1-\alpha/2} + 1.282)\delta}{3.242\sqrt{k}}$$

Therefore, the relationship of the difference between two pre-defined population rates and two estimated sample rates is the same as the normal endpoint.

Simulations

A computer simulation is a computer program that attempts to simulate an abstract model of a particular system, which is used in this research to evaluate the performance of efficacy estimation in the early termination period. The main advantage of simulation is that we can specify the "true" treatment effect in a simulation, a value that is usually unknown within real clinical data.

For the simulation, we also considered both normal and binary endpoints. It was set α= 0.05, power =90% for original designed study, and assumed that the sample allocation ratio is 1. For the normal endpoint part, 3 mean differences between two groups were considered: 0.5, 1 and 2, and the standard deviation was always set as 5. For binary endpoint part, rate differences were set as 0.05, 0.1 and 0.2, respectively.

Based on the above assumptions, the overall sample sizes per group can be derived based on the pre-defined different population mean or rate. With sample size N, individual patients' data were simulated by using normal and binary distribution. Besides that, we assumed that

Table 1. δ_1/δ under different k and α values

α values	k values									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.029	3.38	2.39	1.95	1.69	1.51	1.38	1.28	1.20	1.13	1.07
0.005	3.98	2.82	2.30	1.99	1.78	1.63	1.51	1.41	1.33	1.26
0.001	4.46	3.15	2.57	2.23	1.99	1.82	1.69	1.58	1.49	1.41

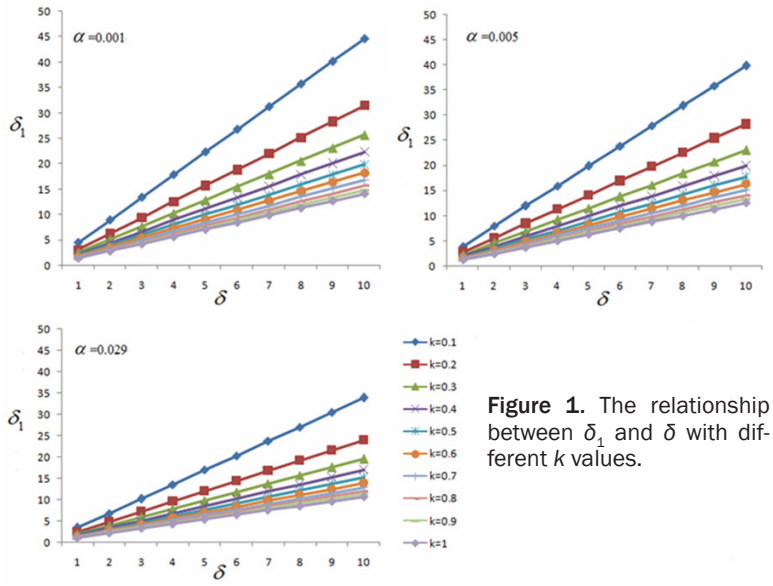

Figure 1. The relationship between δ_1 and δ with different k values.

there only one time planned interim analysis was carried out at the time point when the per-group sample size is kN (k is valued as 10%, 20%, 40%, 60% and 80%).

It is known that there are many methods being applied to define early termination type I error boundary, such as Pocock method [6], O'Brien & Fleming method [7], Peto method [8] and Lan-Demets α spending function method [9]. Assuming a total of 2 times analysis, including 1 time interim analysis, then the interim analysis type I error boundary α can be defined as 0.029, 0.005 and 0.001 according to Pocock, O'Brien-Fleming and Peto method, respectively. Therefore, three types of early termination boundaries were considered and they were 0.001, 0.005 and 0.029 according to Peto, O'Brien-Fleming and Pocock method.

All simulations were conducted by using the SAS 9.2.

Results

Table 1 displays the corresponding relationship between δ and δ_1 under different k and α values in detail according to formula (1) and (2). For given α values, it shows that δ_1 istotally larger than δ, even when k is equal to 1. As k increases, the difference between δ and δ_1 decreases. It also shows that with a k value, the smaller α value is, the greater gap between δ and δ_1 will be. **Figure 1** also presents the same relationship, i.e. the estimated effect at the early termination stage is overestimated than the real effect.

Simulation results for normal endpoint are presented in **Table 2**, and binary endpoint's results are in **Table 3**. "Termination proportion (%)" in **Tables 2** and **3** refers to the proportion of early termination times among the simulated 2 treatments parallel trials on 10000 runs which are satisfied the pre-defined early termination boundary (α), it is calculated as 100 × (trial numbers satisfied $P \leq \alpha$)/10000. δ_1/δ in **Tables 2** and **3** is interpreted as the ratio of the average of the estimated mean/rate differences in the statistical significant early termination stage and the postulated effect size at the design stage.

Despite the different size of postulated population means/rates and the early termination boundary, the values of "Termination proportion (%)" in both **Tables 2** and **3** increase with increasing proportion of sample size. Meanwhile, the δ_1/δ decreases with increasing proportion of sample size, but is always greater than 1, i.e. the actual effect size is overestimated in the early termination period, even after the optimal time of one interim analysis for the early term ination which was suggested as approximately two-thirds of the planned observations [10] (n/N=60%, 80%). Besides that, the precision of the 95% confidence intervals for δ_1/δ improves with the increasing proportion of sample size, ratio 1 which means $\delta_1=\delta$ is not included in all confidence intervals. About 3 different early termination boundaries, for a given proportion of sample size, efficacy estimation is closest to the population parameters under

Table 2. Simulation results for normal endpoint on 10000 runs (α=0.05, STD=5, power =90%)

Parameters	n/N, %	$P \leq 0.001$ (Peto)			$P \leq 0.005$ (O'Brien-Fleming)			$P \leq 0.029$ (Pocock)		
		Termination proportion (%)	δ_1/δ Mean	δ_1/δ 95% CI	Termination proportion (%)	δ_1/δ Mean	δ_1/δ 95% CI	Termination proportion (%)	δ_1/δ Mean	δ_1/δ 95% CI
μ_2-μ_1=0.5 N=2200	10	1.31	3.475	(3.416, 3.533)	3.89	3.053	(2.998, 3.108)	12.98	2.523	(2.492, 2.554)
	20	3.57	2.475	(2.452, 2.499)	9.53	2.203	(2.186, 2.220)	24.01	1.875	(1.861, 1.888)
	40	11.32	1.801	(1.789, 1.812)	23.30	1.619	(1.610, 1.629)	46.14	1.405	(1.397, 1.413)
	60	22.91	1.504	(1.496, 1.512)	39.98	1.367	(1.360, 1.374)	64.53	1.217	(1.211, 1.224)
	80	37.00	1.337	(1.331, 1.343)	55.9	1.233	(1.227, 1.238)	78.23	1.124	(1.118, 1.129)
μ_2-μ_1=1 N=600	10	1.35	3.348	(3.288, 3.409)	4.03	2.951	(2.912, 2.989)	13.39	2.440	(2.413, 2.467)
	20	3.91	2.373	(2.346, 2.399)	10.19	2.117	(2.100, 2.134)	25.34	1.805	(1.792, 1.818)
	40	12.88	1.737	(1.726, 1.748)	26.38	1.557	(1.548, 1.566)	50.07	1.357	(1.349, 1.365)
	60	26.75	1.455	(1.448, 1.462)	44.67	1.328	(1.322, 1.335)	68.98	1.190	(1.184, 1.196)
	80	41.44	1.302	(1.296, 1.307)	60.77	1.201	(1.196, 1.207)	81.68	1.102	(1.097, 1.108)
μ_2-μ_1=2 N=150	10	1.09	3.155	(3.068, 3.242)	3.81	2.848	(2.802, 2.894)	12.82	2.371	(2.336, 2.407)
	20	3.40	2.351	(2.321, 2.382)	9.34	2.108	(2.089, 2.127)	25.14	1.794	(1.781, 1.807)
	40	12.17	1.739	(1.727, 1.752)	25.56	1.561	(1.552, 1.570)	48.98	1.364	(1.356, 1.372)
	60	25.30	1.462	(1.455, 1.470)	43.96	1.328	(1.322, 1.335)	68.93	1.189	(1.183, 1.196)
	80	39.98	1.304	(1.298, 1.310)	60.47	1.199	(1.193, 1.204)	81.59	1.102	(1.096, 1.107)

STD is standard deviation, CI is confidence interval.

Table 3. Simulation results for binary endpoint on 10000 runs (α=0.05, power =90%)

Parameters	n/N, %	$P \leq 0.001$ (Peto)			$P \leq 0.005$ (O'Brien-Fleming)			$P \leq 0.029$ (Pocock)		
		Termination proportion (%)	δ_1/δ Mean	δ_1/δ 95% CI	Termination proportion (%)	δ_1/δ Mean	δ_1/δ 95% CI	Termination proportion (%)	δ_1/δ Mean	δ_1/δ 95% CI
π_1=0.3 π_2=0.35 N=2000	10	1.30	3.388	(3.339, 3.438)	4.17	2.981	(2.939, 3.022)	13.32	2.491	(2.463, 2.519)
	20	3.26	2.419	(2.396, 2.443)	9.79	2.138	(2.121, 2.155)	25.50	1.821	(1.808, 1.834)
	40	12.41	1.757	(1.747, 1.767)	25.56	1.582	(1.574, 1.591)	48.36	1.383	(1.375, 1.391)
	60	25.40	1.477	(1.470, 1.484)	42.46	1.350	(1.344, 1.357)	66.90	1.207	(1.201, 1.213)
	80	39.18	1.322	(1.317, 1.328)	58.82	1.219	(1.213, 1.224)	80.61	1.113	(1.108, 1.119)
π_1=0.3 π_2=0.4 N=500	10	1.21	3.418	(3.366, 3.470)	4.04	2.991	(2.939, 3.042)	12.84	2.521	(2.491, 2.550)
	20	3.42	2.477	(2.452, 2.501)	9.24	2.197	(2.180, 2.214)	24.44	1.858	(1.844, 1.871)
	40	11.48	1.799	(1.788, 1.811)	24.02	1.613	(1.604, 1.623)	46.95	1.402	(1.394, 1.410)
	60	23.60	1.505	(1.498, 1.513)	40.73	1.369	(1.363, 1.376)	65.25	1.220	(1.213, 1.226)
	80	37.02	1.340	(1.334, 1.345)	57.31	1.229	(1.224, 1.235)	79.02	1.123	(1.117, 1.128)
π_1=0.3 π_2=0.5 N=150	10	0.57	3.374	(3.312, 3.437)	1.69	3.103	(3.064, 3.141)	9.03	2.552	(2.526, 2.578)
	20	4.02	2.308	(2.287, 2.328)	11.62	2.032	(2.018, 2.046)	27.06	1.752	(1.740, 1.763)
	40	14.78	1.680	(1.671, 1.690)	27.58	1.528	(1.520, 1.536)	54.11	1.322	(1.315, 1.329)
	60	29.55	1.417	(1.411, 1.424)	47.63	1.299	(1.293, 1.305)	71.36	1.171	(1.166, 1.177)
	80	44.95	1.272	(1.267, 1.277)	64.10	1.181	(1.176, 1.186)	84.50	1.088	(1.083, 1.093)

CI is confidence interval.

Pocock boundary which is the biggest one. Simulation results also show that the estimated effect is overestimated in early terminated stage even the type I error has been controlled by Pocock, O'Brien-Fleming or Peto method.

Discussion

The execution of an interim analysis has been widely accepted in the clinical trial. Sometimes the interim analysis results can determine to early stop a clinical trial for the significantly treatment superiority [11-15]. It seems that patients could get better treatment earlier if the clinical trial was early terminated. But it was recognized that early termination for efficacy purpose does have its risks [16, 17] and early termination should be treated more careful because more evidence is required.

Grothey A reported an early termination trial in 2011 regarding intravenous calcium and mag-

nesium for oxaliplatin-induced sensory neurotoxicity in adjuvant colon cancer [18]. Although the interim analysis results were significant, the researchers still had no confidence in the promotion of their research results since only 102 subjects were enrolled which was less than the target sample size. After that, Loprinzi CL designed another similar large sample clinical trial which was reported in 2013 [19] and the trial verified that calcium and magnesium cannot reduce the oxaliplatin-induced sensory neurotoxicity. From this example, we learned that the interim analysis results are often uncertain and may make the problem complicated.

Besides thereal example above, several Meta analyses focused on the treatment effects of the early terminated clinical trials in the past years. Montori VM published a systematic review regarding trials stopped early for benefit on JAMA in 2005, which indicated that RCTs (randomized clinical trials) stopped early for benefit often failed to adequately report relevant information about the decision to stop early and showed implausibly large treatment effects, particularly when the number of events is small [20]. As for another similar systematic review written by Bassler D published on JAMA in 2010, the author also indicated that truncated RCTs were associated with greater effect sizes than RCTs not stopped early and this difference was independent of the presence of statistical termination rules and was greatest in smaller studies [21].

However, there has not been a widely consistent conclusion so far for efficacy estimation in an early termination clinical trial based on theoretical analysis. Besides Meta analyses, the final conclusion needs more powerful evidence. In this article, we have found that the efficacy of both normal and binary endpoints are overestimated in the early termination period. This finding is supported by formula derivation and simulation methods whose results are consistent. It is obvious that overestimating the effect size can be destructive to the success of the clinical trial. It is also important to note that controlling the type I error rate does not address this potential bias of overestimation. In simulation part, three types of early termination boundaries were considered, they were 0.001, 0.005 and 0.029 according to Peto, O'Brien-Fleming and Pocock method, but none of them could control the overestimation tendency.

So it is true that a trial terminated early for benefits will tend to overestimate true effects, which happen because there is always variability in estimation of true effects [22, 23] Overestimation of treatment effects may occur if the decision to stop a trial coincides with a "random high" in the treatment effect. Such random fluctuations of the estimated treatment effect are often large and typically happen early in a trial's progress, and when assessing data over time, evidence of extreme benefit is more likely to be obtained at times.

We propose the following recommendations. 1) For confirmatory trial, it is inadvisable to early terminate even the effect of treatment group was significantly superior to the control group in an interim analysis, because this efficacy may be overestimated. It is recommended to continue the confirmatory trial after an interim analysis in order to get more safety and efficacy evidence. But adjusting the group proportion is acceptable in order that more subjects can be enrolled in superior treatment group. 2) For exploratory trial, as we all know, the rationale and design of confirmatory trials nearly always rest on a series of exploratory studies. But it is also not suggested to early terminate an exploratory trial because of the aforementioned reasons. It is better to continue the exploratory trial after the interim analysis to get more evidence for the stable parameters estimation which is needed for confirmatory trial design. 3) When to early terminate the trial? If the drug's efficacy and safety effects have been verified by the previous approved clinical trial, it is certain to terminate the trial early when the same results have been found in the interim analysis.

Conclusions

In this research, formula derivation and simulation methods were used to evaluate the treatment effects estimated from early termination clinical trial for normal and binary endpoints. The outputs show that the effect estimation of both normal and binary endpoints is overestimated in the early termination period and controlling the type I error rate does not address this potential bias of overestimation. The works are from the basic theoretical angle, which are powerful supplementary evidence to the previous published Meta analyses.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (81273184) the National Natural Science Foundation of China Grant for Young Scientists (81302512).

Disclosure of conflict of interest

None.

Authors' contribution

Lihong Huang is the first author, who made substantial contributions to derive inequalities, design and complete simulations; Fang Shao and Jian-ling Bai participated in the inequalities developments; Hao Yu and Ping Yu were involved in drafting the manuscript; Feng Chen gave the final approval of the version to be published. All authors have read and approved the final manuscript.

Address correspondence to: Dr. Feng Chen, Department of Biostatistics, School of Public Health, Nanjing Medical University, SPH Building Room 412, 101 Longmian Avenue, Nanjing 211166, Jiangsu, P. R. China; Ministry of Education Key Laboratory for Modern Toxicology, School of Public Health, Nanjing Medical University, SPH Building Room 412, 101 Longmian Avenue, Nanjing 211166, Jiangsu, P. R. China. Tel: +86-25-86868435; Fax: +86-25-86868435; E-mail: fengchen@njmu.edu.cn; Ping Yu, Jingan District Center for Disease Control and Provention, No. 195, Yonghe Road, Jingan, Shanghai 200072, P. R. China. Tel: +86-02156659090, Fax: +86-02156659090, E-mail: jxcdcyp@126.com

References

[1] Boland E, Monsod T, Delucia M, Brandt CA, Fernando S and Tamborlane WV. Limitations of conventional methods of self-monitoring of blood glucose: lessons learned from 3 days of continuous glucose sensing in pediatric patients with type 1 diabetes. Diabetes Care 2001; 24: 1858-1862.

[2] DeMets DL, Hardy R, Friedman LM and Lan KK. Statistical aspects of early termination in the beta-blocker heart attack trial. Control Clin Trials 1984; 5: 362-372.

[3] ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. Stat Med 1999; 18: 1905-1942.

[4] Delgado-Herrera L and Anbar D. A model for the interim analysis process: a case study. Control Clin Trials 2003; 24: 51-65.

[5] Koch A. How much do we have to pay for how many interim analyses? Contemp Clin Trials 2005; 26: 113-116.

[6] Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika 1977; 64: 191-199.

[7] O'Brien PC and Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979; 35: 549-556.

[8] Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J and Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer 1976; 34: 585-612.

[9] Lan KG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983; 70: 659-663.

[10] Togo K and Iwasaki M. Optimal timing for interim analyses in clinical trials. J Biopharm Stat 2013; 23: 1067-1080.

[11] Fernandes RM, van der Lee JH and Offringa M. Data monitoring committees, interim analysis and early termination in paediatric trials. Acta Paediatr 2011; 100: 1386-1392.

[12] White R, Chileshe M, Dawson L, Donnell D, Hillier S, Morar N, Noguchi L and Dixon D. Fostering community understanding of sufficient benefit and early stopping for a phase 2B HIV prevention clinical trial in Africa. Clin Trials 2011; 8: 103-111.

[13] Booth CM, Ohorodnyk P, Zhu L, Tu D and Meyer RM. Randomised controlled trials in oncology closed early for benefit: trends in methodology, results, and interpretation. Eur J Cancer 2011; 47: 854-863.

[14] Fernandes RM, van der Lee JH and Offringa M. A systematic review of the reporting of Data Monitoring Committees' roles, interim analysis and early termination in pediatric clinical trials. BMC Pediatr 2009; 9: 77.

[15] Stegert M, Kasenda B, von Elm E, You JJ, Blumle A, Tomonaga Y, Saccilotto R, Amstutz A, Bengough T, Briel M. An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. J Clin Epidemiol 2016; 69: 152-160.

[16] Casazza G and Casella F. Can we trust in trials stopped early for benefit? Intern Emerg Med 2012; 7: 559-561.

[17] Briel M, Bassler D, Wang AT, Guyatt GH and Montori VM. The dangers of stopping a trial too early. J Bone Joint Surg Am 2012; 94 Suppl 1: 56-60.

[18] Grothey A, Nikcevich DA, Sloan JA, Kugler JW, Silberstein PT, Dentchev T, Wender DB, Novotny PJ, Chitaley U, Alberts SR and Loprinzi CL. Intravenous calcium and magnesium for oxaliplatin-induced sensory neurotoxicity in

adjuvant colon cancer: NCCTG N04C7. J Clin Oncol 2011; 29: 421-427.

[19] Loprinzi CL, Qin R, Dakhil SR, Fehrenbacher L, Flynn KA, Atherton P, Seisler D, Qamar R, Lewis GC and Grothey A. Phase III randomized, placebo-controlled, double-blind study of intravenous calcium and magnesium to prevent oxaliplatin-induced sensory neurotoxicity (N08CB/Alliance). J Clin Oncol 2014; 32: 997-1005.

[20] Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schunemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R, Lo B, Thompson CA, Zhou Q, Mills E and Guyatt GH. Randomized trials stopped early for benefit: a systematic review. JAMA 2005; 294: 2203-2209.

[21] Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, Heels-Ansdell D, Walter SD, Guyatt GH, Flynn DN, Elamin MB, Murad MH, Abu Elnour NO, Lampropulos JF, Sood A, Mullan RJ, Erwin PJ, Bankhead CR, Perera R, Ruiz Culebro C, You JJ, Mulla SM, Kaur J, Nerenberg KA, Schunemann H, Cook DJ, Lutz K, Ribic CM, Vale N, Malaga G, Akl EA, Ferreira-Gonzalez I, Alonso-Coello P, Urrutia G, Kunz R, Bucher HC, Nordmann AJ, Raatz H, da Silva SA, Tuche F, Strahm B, Djulbegovic B, Adhikari NK, Mills EJ, Gwadry-Sridhar F, Kirpalani H, Soares HP, Karanicolas PJ, Burns KE, Vandvik PO, Coto-Yglesias F, Chrispim PP and Ramsay T. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA 2010; 303: 1180-1187.

[22] Schou IM and Marschner IC. Meta-analysis of clinical trials with early stopping: an investigation of potential bias. Stat Med 2013; 32: 4859-4874.

[23] Berry SM, Carlin BP and Connor J. Bias and trials stopped early for benefit. JAMA 2010; 304: 156; author reply 158-159.