Contemporary Clinical Trials Communications 7 (2017) 224–230

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Decision on performing interim analysis for comparative clinical trials

Kyongsun Pak[a,*], Susanna Jacobus[b], Hajime Uno[c]

[a] Kitasato University, School of Pharmacy, Department of Clinical Medicine (Biostatistics), 5-9-1 Shirokane, Minato-ku, Tokyo 108-0072, Japan
[b] Dana Farber Cancer Institute – ECOG-ACRIN Biostatistics Center, Boston, MA, USA
[c] Dana Farber Cancer Institute, Department of Medical Oncology, Division of Population Sciences, Boston, MA, USA

ARTICLE INFO

Keywords:
Interim analysis
Blinded analysis
Early termination for superiority
Early termination for futility
Data monitoring

ABSTRACT

In randomized-controlled trials, interim analyses are often planned for possible early trial termination to claim superiority or futility of a new therapy. While unblinding is necessary to conduct the formal interim analysis in blinded studies, blinded data also have information about the potential treatment difference between the groups. We developed a blinded data monitoring tool that enables investigators to predict whether they observe such an unblinded interim analysis results that supports early termination of the trial. Investigators may skip some of the planned interim analyses if an early termination is unlikely. We specifically focused on blinded, randomized-controlled studies to compare binary endpoints of a new treatment with a control. Assuming one interim analysis is planned for early termination for superiority or futility, we conducted extensive simulation studies to assess the impact of the implementation of our tool on the size, power, expected number of interim analyses, and bias in the treatment effect. The numerical study showed the proposed monitoring tool does not affect size or power, but dramatically reduces the expected number of interim analyses when the effect of the treatment difference is small. The tool serves as a useful reference when interpreting the summary of the blinded data throughout the course of the trial, without losing integrity of the study. This tool could potentially save the study resources and budget by avoiding unnecessary interim analyses.

1. Introduction

In randomized-controlled trials, interim analyses are often planned to review the efficacy or safety of the therapeutic interventions. Early termination of the trial may occur due to evidence of superiority or futility of the new therapy based on the interim analysis. To conduct interim analyses, we need to access the data prior to the completion of the trial. Particularly for blinded studies, interim analysis requires unblinding of the treatment allocation and conducting a formal between-group comparison [1,2]. Although unblinded data provide complete information of the observed data, blinded data also contain information about the treatment difference between the groups. For instance, when the observed response rate in the pooled sample is very low at the time of the interim analysis, we know the response rates in both groups are very low. Therefore, there is little chance a significant difference between the groups would be observed and, consequently, a formal comparison is a wasteful expenditure of alpha. Even when response rates are not that small, if the control rate can be reasonably estimated based on previous studies, the blinded data yields a decent estimate of the treatment difference.

There are several data monitoring tools [3–5] that use blinded data originating in the Bayesian approach for safety monitoring in single arm studies proposed by Thall and Simon [6]. For example, Ball [3] focused on the adverse event rate in the pooled sample and proposed a decision rule based on the posterior distribution of it using the Bayesian approach. On the other hand, our focus in this paper is a blinded data monitoring tool predicting the result of a formal unblinded interim analysis for superiority or futility of a new therapy. The proposed tool works with the hypothesis testing approach. Specifically, we assume that the alpha spending function approach [7] is used as a stopping guideline for superiority in the formal interim analysis. For futility, we assume that the result of stochastic curtailment method is used as a guideline of early stopping [8]. We performed extensive numerical studies to assess the impact of the implementation of the data monitoring tool on the type I error rate, power, expected sample size, expected number of interim analyses to be performed and bias in the treatment effect for both superiority and futility. We illustrated the practical application of our tool, using data from a clinical trial conducted by the ECOG-ACRIN Cancer Research Group. With our tool, investigators may skip some of the planned interim analyses when the result of an interim analysis at that time point is unlikely to support early termination of the trial for superiority or futility. Therefore, this

* Corresponding author.
E-mail address: pakk@pharm.kitasato-u.ac.jp (K. Pak).

http://dx.doi.org/10.1016/j.conctc.2017.08.001
Received 27 February 2017; Received in revised form 7 August 2017; Accepted 8 August 2017
Available online 16 August 2017

tool could ultimately avoid unnecessary spending of study resources while maintaining scientific integrity of the trial.

2. Methods

In this paper, we specifically focus on randomized controlled trials comparing binary endpoints, namely response rates, between a new therapy and a control. In the trial, interim analyses are planned for early termination for superiority or futility or both.

2.1. Typical procedure of interim analysis

Usually, the interim analysis is implemented at the time when the pre-planned information fraction is reached. For a binary outcome, the total information will be defined as the planned total sample size. Assume that, during the accumulating the preset sample size M, there are N $(\leq M)$ participants and T $(\leq N)$ responders in the two arms at the time of the interim analysis. Let (T_1, T_0) denote the numbers of responders in the arm of the new therapy and control respectively, and then $T = T_1 + T_0$. When unblinding the data, we can observe (T_1, T_0), and formal comparison would be implemented. Depending on the resulting test statistic, or the corresponding p-value or conditional power, we decide whether to stop or continue the trial.

2.2. Blinded data monitoring tool

Before breaking the blinded treatment assignment code, we may monitor (N, T) from the blinded data. Assume that each T_1 and T_0 follows a binomial distribution with a parameter p_1 for the new therapy and p_0 for the control therapy, respectively. The probability mass function of T, $Pr(T = t)$, can be expressed with a mixture of the aforementioned two binomials. Given the allocation ratio during the study q: $(1 - q)$ for the new therapy and control respectively, where $q \in (0,1)$, $Pr(T = t)$ is expressed that

$$Pr(T = t) = \binom{N}{t}\{qp_1 + (1-q)p_0\}^t\{q(1-p_1) + (1-q)(1-p_0)\}^{N-t}.$$

With the blinded treatment allocation, if we have enough certainty about p_0 and if the allocation ratio is close to q, we would be able to predict the response rate of the new therapy p_1. Specifically, if p_0 is a known value, the maximum likelihood estimator of p_1 is obtained by

$$\widehat{p_1} = \frac{T - N(1-q)p_0}{Nq}.$$

Then the standardized test statistics for testing the null hypothesis H_0: $p_1 = p_0$ is given by $Z_b = (\widehat{p_1} - p_0)/\sqrt{\widehat{\mathrm{Var}}(\widehat{p_1})}$, where $\widehat{\mathrm{Var}}(\widehat{p_1}) = N\hat{r}(1-\hat{r})/(Nq)^2$ and $\hat{r} = q\widehat{p_1} + (1-q)p_0$. Utilizing the observed Z_b at the interim analysis point, we can predict whether or not the unblinded interim analysis result will meet the stopping criteria for superiority or futility. For superiority, one can then obtain the threshold values of the total number of responders T with respect to each number of subjects N, with which the p-value of the test would meet the pre-specified stopping criteria corresponding to the information time at the interim analysis. For futility, one might use a conditional probability as criteria for stopping.

2.3. Illustrative example

To illustrate the aforementioned decision criteria, we consider a specific numerical example of a randomized controlled trial comparing

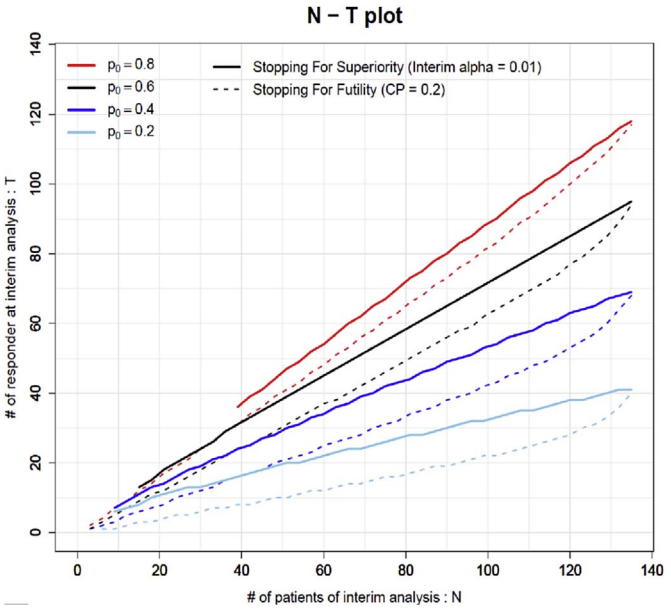

Fig. 1. N-T plot with $\mathbf{p}_0$= 0.8, 0.6, 0.4 and 0.2 for early stopping for superiority and futility.

response rates between the new and the control therapy. The accrual goal is 135 patients and the mixture proportion of allocation is q: $(1 - q) = \frac{2}{3}$: $\frac{1}{3}$ for the new therapy and the control, respectively.

First, we consider the case for interim analysis expecting early termination only for superiority and consuming type I error rate $\alpha = 0.01$ at the interim analysis. Under this scenario, the solid curves in Fig. 1 show the thresholds of N and T with various values of p_0. For example, the blue solid curve corresponds the case that $p_0 = 0.4$. Using the observed (N, T) with blindness maintained, these curves can be a reference to predict how likely the interim analysis result would meet the stopping criteria, if conducted. Specifically, in this example, when the observed (N, T) is above the blue curve, we can expect that the result of the interim analysis will support early stopping for superiority for the new therapy. Therefore, if we think that p_0 is very likely to be 0.4, we would conclude that an interim analysis should not be missed at this point.

Next, we consider the case of early termination for futility based on the conditional power less than 0.2. The dashed lines in Fig. 1 show the corresponding N-T curves for futility with various p_0's. Again, consider the case that $p_0 = 0.4$. The observed (N, T) below the blue dashed curve indicates the conditional power will likely be below 0.2. Thus, if we are confident with a $p_0 = 0.4$, we would determine the interim analysis should not be missed for potential futility stop. On the other hand, if the observed (N, T) is above the blue dashed curve, it may be an option to skip the scheduled interim analysis, if there are no other concerns on the study.

This tool can also be used for the cases that both superiority and futility stoppings are of interest. In those cases, we will use both solid and dashed curves in Fig. 1. When the observed (N, T) is in between solid and dashed lines, the interim analysis result will likely not meet stopping criteria for either superiority or futility. We may skip performing the interim analysis and continue the trial, unless there are other concerns in the study.

In this manner, the proposed blinded monitoring process is helpful for identifying whether it is a good time to conduct interim analysis, preserving the integrity of the study. Appendix 1 provides the computer programs to generate N-T plots with a documented example.

3. Simulation studies

We performed extensive numerical studies to assess the impact of the implementation of the proposed blinded data monitoring tool with respect to a randomized controlled design using 2:1 allocation (new therapy:control) and comparison of a binary outcome. We assume that the investigators plan to do one interim analysis when the outcome of 60 patients are available among the planned 135 patients. The overall type I error rate is set to 0.05. We consider that 0.01 of the alpha would be spent at the interim analysis, and then the critical p-value for the second analysis is derived to be 0.0446 by the Hybittle-Peto method [9,10]. The true response rates of the new therapy p_1 were set to 0.40, 0.50, 0.60, 0.65, 0.70 and 0.80, and the true response rate of the control group p_0 was set to 0.40. The binary outcomes are generated from the binomial distribution with the success probability p_1 for the new therapy and p_0 for the control arm, respectively. Iterating 10000 times, we assessed the overall type I error rate, power, expected sample size, the probability of conducting the interim analysis, the probability of stopping the trial at the interim analysis, and the bias of the treatment effect.

In these simulations, three patterns of early termination criteria were evaluated— 1) only for superiority, 2) only for futility, and 3) both for superiority and futility. Within each pattern, four scenarios as outlined below were considered.

- First scenario (Without using our tool)

As a conventional procedure of randomized controlled trials, an interim analysis is conducted at a pre-specified information time. If the result of the interim analysis is significant, we stop the study for superiority of the new therapy. Otherwise, the trial will be continued until full accrual, and the second analysis conducted with the planned sample size.

- Second scenario (Using our tool with a correct p_0)

At the pre-specified time point in the aforementioned scenario, the decision of whether or not to conduct an interim analysis is made by the proposed blinded data monitoring tool with a correctly specified parameter for the control arm. When our tool suggests an interim analysis should be conducted, the interim data will be unblinded and the interim analysis will be performed as in the first scenario. On the other hand, when our tool suggests skipping the interim analysis, the data will be analyzed only at the end of the trial. In the latter case, since we haven't spent alpha for the interim analysis, the nominal level of the type I error rate is used at the final analysis.

- Other scenarios (Using our tool with mis-specified p_0)

We take the same procedure as described in the second scenario, but consider the case when p_0 is misspecified when creating N-T plot. Specifically, we consider a case where we underestimate p_0 (i.e., $p_0 = 0.30$ in the third scenario 'lower' p_0) and a case where we overestimate p_0(i.e, $p_0 = 0.50$ in the fourth scenario).

4. Results

4.1. For case of early termination for superiority

Table 1 presents the simulation results in the case of early termination for superiority at the one interim analysis which is done with 60 patients. In this table, we report power or size (overall type I error rate), the expected sample size during the study (E[M]), the probability to conduct an interim analysis (IA), the probability to terminate the trial at the interim analysis for superiority of the new therapy (Sig.IA), and the proportion of Sig.IA/IA, among 10000 sets of samples. Note that the expected sample size would also be an indicator of the expected study duration. When the expected sample size is close to 135, the study would be continued until the time of the final analysis. We also evaluated the bias of the treatment effect via $E[\widehat{p_1} - \widehat{p_0}] - (p_1 - p_0)$. Scenario 1-1 shows the results of the case using the conventional strategy and the other three scenarios show the results using the proposed blinded data monitoring tool under various conditions of p_0.

4.2. For case of early termination for futility

In this setting, we consider early stopping for futility, instead of superiority. Specifically, at the planned interim analysis time point, we calculate the conditional probability. If it is below 0.2, we stop the trial. The incorporation of a futility stopping rule affects the overall type I error, but we do not adjust for it in this numerical study. Therefore, the critical p-value at the final analysis is 0.05. In Table 2, we report the probability to terminate the trial at the interim analysis for futility of the new therapy (Fut.IA), and the proportion of Fut.IA/IA, in addition to power or size (overall type I error rate), the expected sample size during the study E[M], IA and the bias of the treatment effect under each four data monitoring scenario.

4.3. For case of early termination for superiority and futility

We now consider the case with both superiority and futility stopping rules. Specifically, at the interim analysis, we will stop the trial for superiority if the observed p-value is less than 0.01, or for futility if the conditional probability is less than 0.2. Table 3 gives the results of the four data monitoring scenarios including the parameters power and size, E[M], IA, Sig.IA, Fut.IA, the proportion of Sig.IA + Fut.IA to IA (Sig.Fut.IA/IA) and the bias of the treatment effect.

The resulting three tables indicate that when the effect of the treatment difference is small, the chance to conduct interim analysis for superiority becomes dramatically reduced, and that for futility becomes increased by using our blinded-data monitoring tool. Furthermore, the trends of the probability to terminate the trial at the interim analysis based on superiority and/or futility conducted are not dependent the blinded or unblinded data monitoring strategies. Generally, the treatment effect simply estimated from the study data will be biased, when a stopping boundary is imposed. We find that the bias of the estimated treatment effect will be reduced by using the proposed blinded data monitoring tool, compared to the scenario when the interim analysis is precisely conducted. Interestingly, even in the cases that the anticipated rates on the outcome in the control therapy are not close to the truth, similar operational characteristics are observed. Therefore, using our blinded data monitoring tool, we can reduce the chance to conduct unnecessary interim analysis and wasting study resources, especially when there is little benefit for early stopping in the trial.

Notably, there is no gain in power by using the proposed blinded data monitoring tool, compared with the conventional method. Also, using our blinded data monitoring tool, the expected sample size will be slightly increased, compared to the conventional methods. This is because the interim analysis that meets the stopping criteria is sometimes skipped and the final analysis is then conducted with the whole planned sample size. We also find that, when the anticipated response rate on the outcome in the control therapy to create N-T plot for early stopping for superiority is underestimated (Scenario 1–3 in Table 1), the impact on the sample size is fairly small. On the other hand, when we overestimate the rate on the outcome in the control for early stopping for

Table 1
Superiority Stopping for the Binary Outcome (Fisher's Exact Test); With True Response Rate of Control Therapy p0 = 0.40; N-T plot constructed with interim alpha 0.01.

Scenario Number	Content of Scenario	Control Parameter p0 Used for N-T plot	True Response Rate of New Therapy p1	Power or Size (%)		E[M]	IA (%)	Sig.IA (%)	Sig.IA/IA (%)	Bias
1–1	Not Using N-T plot	N/A	0.80	Power	99.5	86.0	100.0	65.4	65.4	0.026
			0.70		90.1	113.5	100.0	28.7	28.7	0.024
			0.65		76.0	122.6	100.0	16.6	16.6	0.018
	(Interim analysis is surely done)		0.60		54.4	128.3	100.0	9.0	9.0	0.011
			0.50		16.5	133.4	100.0	2.2	2.2	0.004
			0.40	Size	4.2	134.5	100.0	0.6	0.6	0.002
1–2	N-T plot With Propoer p0	0.40	0.80	Power	99.5	88.6	94.7	61.9	65.4	0.024
			0.70		90.3	120.5	66.3	19.3	29.1	0.015
			0.65		77.1	129.4	45.6	7.5	16.5	0.008
	Do Interim analysis if T > 34		0.60		56.9	133.4	25.9	2.1	8.2	0.002
			0.50		18.0	134.9	4.8	0.1	1.9	0.001
			0.40	Size	4.1	135.0	0.3	0.0	0.0	0.001
1–3	N-T plot With Lower p0	0.30	0.80	Power	99.5	86.0	100.0	65.4	65.4	0.026
			0.70		90.1	114.0	97.8	28.0	28.6	0.023
			0.65		76.3	123.6	92.7	15.2	16.4	0.016
	Do Interim analysis if T > 28		0.60		55.0	129.7	81.9	7.0	8.6	0.009
			0.50		17.0	134.4	44.8	0.8	1.7	0.002
			0.40	Size	4.0	134.9	11.8	0.1	0.7	0.001
1–4	N-T plot With Higher p0	0.50	0.80	Power	99.5	112.6	45.3	29.8	65.9	0.007
			0.70		90.4	132.1	11.1	3.9	35.0	0.002
			0.65		77.2	134.4	4.2	0.8	20.2	0.000
	Do Interim analysis if T > 40		0.60		57.2	134.9	1.3	0.2	13.2	0.000
			0.50		18.0	135.0	0.1	0.0	0.0	0.001
			0.40	Size	4.1	135.0	0.0	0.0	–	0.001

E[M]: the expected sample size; IA(%): the probability to conduct interim analysis among 10000 samples; Sig.IA(%): the probability to terminate the trial at the interim analysis for superiority of the new therapy among 10000 samples; Sig.IA/IA(%): the proportion of Sig.IA to IA; Bias: the bias of the treatment effect $E[\widehat{p_1} - \widehat{p_0}] - (p_1 - p_0)$.

superiority (Scenario 1–4 in Table 1), the expected sample size is increased because most of the planned interim analyses are skipped and those studies are continued until the planned end.

5. Application

We illustrate how to utilize our tool using the data from newly diagnosed multiple myeloma patients who participated in a clinical trial conducted by ECOG-ACRIN [11]. The primary objective of this trial was to evaluate the 4-month response rates of the combination therapy with thalidomide and dexamethasone (therapy A), compared with the standard therapy with dexamethasone alone (therapy B). A total of 199 eligible patients were randomized to therapy A (n = 99) and therapy B (n = 100). The study showed that the response rate in the therapy A group is significantly higher than therapy B. Note that this study was designed, anticipating that the 4-month response rate in therapy B group is 60%. However, the observed response rate in therapy B was 39% in this trial.

Here, we consider that the stopping criteria are p < 0.01 for superiority and conditional power < 0.2 for futility. In Fig. 2, there are three panels. Panel (1) shows the case that we anticipate that the response rate in the control group (therapy B) is $p_0 = 0.60$. The other two panels (2) and (3) are for $p_0 = 0.40$ and 0.20, respectively. The black solid curves show the reference boundary for superiority and the dashed curve for futility. The gray line in each panel indicates the observed N-T curve of the myeloma trial data, the three red dots on the gray line highlight the points at $N = 50$,100 and 199. At these time points, the observed p-values of Fisher's exact test were 0.023, 0.0089 and 0.0018, respectively, the conditional powers were 0.999, 0.989 and 1.00, respectively.

Depending on the anticipation of the response rate in the control group, one of these N-T plots will be used. If it is uncertain, several N-T plots may be used. For example, if investigators expect $p_0 = 0.40$ is the true response rate in the control group, Fig. 2 (2) will be used. Suppose data from 100 patients are available at a potential interim analysis time point. They may decide to perform the interim analysis at that time point, as Fig. 2 (2) indicates the interim analysis will likely support early termination for superiority. With this example, if the interim analysis had been conducted, the trial would then have stopped with smaller number of patients than the planned sample size. On the other hand, when the expectation of the response rate in the control group is much higher, investigators may use N-T plot in Fig. 2 (1). In that example, they may decide to perform the interim analysis since the N-T plot suggests the interim analysis will likely support early stopping for futility. In this manner, the N-T plot can be used to decide if it will be

Table 2
Futility Stopping for the Binary Outcome (Fisher's Exact Test); With True Response Rate of Control Therapy p0 = 0.40; N-T plot constructed with conditional power 0.20.

Scenario Number	Content of Scenario	Control Parameter p0 Used for N-T plot	True Response Rate of New Therapy p1	Power or Size (%)		E[M]	IA (%)	Fut.IA (%)	Fut.IA/IA (%)	Bias
2–1	Not Using N-T plot	N/A	0.80	Power	99.4	134.8	100.0	0.2	0.2	−0.001
			0.70		89.3	133.0	100.0	2.7	2.7	−0.005
			0.65		75.8	130.6	100.0	5.8	5.8	−0.009
	(Interim analysis is surely done)		0.60		55.9	126.4	100.0	11.5	11.5	−0.015
			0.50		17.3	110.5	100.0	32.7	32.7	−0.028
			0.40	Size	2.1	89.5	100.0	60.6	60.6	−0.029
2–2	N-T plot With Propoer p0	0.40	0.80	Power	99.6	135.0	0.0	0.0	–	0.000
			0.70		90.4	135.0	0.3	0.0	0.0	−0.001
			0.65		77.2	135.0	1.4	0.1	7.1	−0.001
	Do Interim analysis if T < 25		0.60		57.1	134.5	4.9	0.7	14.3	−0.001
			0.50		17.8	128.2	25.5	9.0	35.5	−0.007
			0.40	Size	2.8	105.1	65.3	39.8	60.9	−0.017
2–3	N-T plot With Lower p0	0.30	0.80	Power	99.6	135.0	0.0	0.0	–	0.000
			0.70		90.4	135.0	0.0	0.0	–	−0.001
			0.65		77.2	135.0	0.0	0.0	–	−0.001
	Do Interim analysis if T < 18		0.60		57.2	135.0	0.0	0.0	0.0	−0.001
			0.50		18.0	135.0	0.7	0.2	27.3	0.000
			0.40	Size	4.0	132.0	6.6	4.0	60.6	−0.001
2–4	N-T plot With Higher p0	0.50	0.80	Power	99.6	135.0	0.7	0.0	0.0	0.000
			0.70		90.3	134.8	11.6	0.3	2.2	−0.001
			0.65		76.8	133.9	25.8	1.5	5.9	−0.003
	Do Interim analysis if T < 31		0.60		56.6	130.8	45.5	5.6	12.3	−0.008
			0.50		17.4	115.7	82.0	25.7	31.4	−0.021
			0.40	Size	2.1	91.1	97.5	58.5	60.0	−0.025

E[M]: the expected sample size; IA(%): the probability to conduct interim analysis among 10000 samples; Fut.IA(%): the probability to terminate the trial at the interim analysis for futility of the new therapy among 10000 samples; Fut.IA/IA(%): the proportion of Fut.IA to IA; Bias: the bias of the treatment effect $E[\widehat{p_1} - \widehat{p_0}] - (p_1 - p_0)$.

worthwhile conducting an interim analysis during the study, based on the expectations for the response rate in the control group.

6. Discussion

In randomized-controlled trials, monitoring which involves interim analyses requiring unblinding of accumulated data may risk inflation of type I error rate. Using our blinded data monitoring tool, we can obtain useful reference information of blinded data and use it to assess the appropriateness of conducting a formal unblinded interim analysis. According to the results of simulations, our data monitoring tool can potentially save study resources and budget by avoiding unnecessary interim analyses. From this aspect, the blinded analyses have remarkable characteristics in terms of saving alpha and operational burden to unblind the data. Note that, when the investigators plan to conduct interim analysis and utilize the proposed monitoring tool, they should pre-specify in the protocol that there is a possibility to reduce the number of the interim analysis using that tool. For those trials where skipping any scheduled planned interim analysis is undesirable, the proposed method should not be applied.

With our method, the choice of doing an interim analysis depends on setting the p_0 parameter of the control arm. Practically, the anticipated efficacy of the control therapy often differs from the observed results. Even in such cases, nevertheless, the power and type I error of our blinded monitoring tool remain consistent with cases when the parameter is correctly specified. In the cases when the knowledge of the control therapy is somewhat vague, we recommend considering several possible parameters for the response rate of the outcome in the control arm. Using the proposed graphical tool repeated for various control rate assumptions at the time of a given interim analysis provides a comprehensive analysis and enables investigators to make an informed decision on decide whether to conduct the formal interim analysis.

In this research, we evaluated how interim monitoring of binary endpoints with data blinded, based on conventional frequentist hypotheses testing methods, impacts the operating characteristics of study design as compared with standard unblinded interim analysis with extensive numerical studies. Our approach uses accessible reference information to produce a valuable monitoring tool for assessing the appropriateness of interim analyses in standard clinical trials. Future work will examine application to other types of outcomes, e.g. continuous quantitative measures using the mean value of blinded data. We may also apply the similar approach to time to event endpoints for assessing the appropriateness of conducting interim analyses, using the mixture of two exponential or Weibull distributions for blinded data.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Table 3
Superiority Stopping and Futility Stopping for the Binary Outcome (Fisher's Exact Test); With True Response Rate of Control Therapy p0 = 0.40; N-T plot constructed with interim alpha 0.01 and conditional power 0.20.

Scenario Number	Content of Scenario	Control Parameter p0 Used for N-T plot	True Response Rate of New Therapy p1	Power or Size (%)		E[M]	IA (%)	Sig.IA (%)	Fut.IA (%)	Sig.Fut.IA /IA (%)	Bias
3–1	Not Using N-T plot	N/A	0.80	Power	99.4	85.8	100.0	65.4	0.2	65.6	0.025
			0.70		89.5	111.5	100.0	28.7	2.7	31.3	0.019
			0.65		76.1	118.2	100.0	16.6	5.8	22.4	0.009
	(Interim analysis is surely done)		0.60		56.2	119.7	100.0	9.0	11.5	20.4	−0.004
			0.50		17.7	108.9	100.0	2.2	32.7	34.8	−0.024
			0.40	Size	2.3	89.3	100.0	0.4	60.6	61.0	−0.028
3–2	N-T plot With Propoer p0	0.40	0.80	Power	99.4	88.3	94.8	62.1	0.2	65.7	0.023
			0.70		89.8	119.8	65.2	18.8	1.6	31.2	0.011
			0.65		76.4	127.8	44.5	7.2	2.4	21.7	0.003
	Do Interim analysis if T < 25 or T > 34		0.60		56.5	131.2	27.2	2.1	3.0	18.9	−0.002
			0.50		17.7	129.3	22.7	0.5	7.1	33.5	−0.005
			0.40	Size	3.2	110.1	55.3	0.2	33.1	60.2	−0.014
3–3	N-T plot With Lower p0	0.30	0.80	Power	99.3	85.8	100.0	65.5	0.2	65.7	0.025
			0.70		89.2	112.4	97.6	27.8	2.4	30.9	0.018
			0.65		75.1	112.0	92.5	14.8	5.2	21.6	0.007
	Do Interim analysis if T < 18 or T > 28		0.60		54.2	123.4	81.6	6.4	9.1	19.0	−0.004
			0.50		16.7	124.1	45.3	0.9	13.6	32.1	−0.009
			0.40	Size	3.8	128.1	15.8	0.1	9.1	57.9	−0.003
3–4	N-T plot With Higher p0	0.50	0.80	Power	99.5	113.2	43.9	29.0	0.1	66.3	0.007
			0.70		90.2	130.8	17.2	5.3	0.4	33.0	0.002
			0.65		76.5	131.5	21.8	3.5	1.2	21.5	0.001
	Do Interim analysis if T < 31 or T > 40		0.60		55.8	129.7	36.2	3.0	4.1	19.4	−0.002
			0.50		16.6	116.6	74.3	1.6	22.9	3.0	−0.016
			0.40	Size	2.3	91.9	95.5	0.3	57.2	60.2	−0.024

E[M]: the expected sample size; IA(%): the probability to conduct interim analysis among 10000 samples; Sig.IA(%): the probability to terminate the trial at the interim analysis for superiority of the new therapy among 10000 samples; Fut.IA(%): the probability to terminate the trial at the interim analysis for futility of the new therapy among 10000 samples; Sig.Fut.IA/IA(%): the proportion of (Sig.IA + Fut.IA) to IA; Bias: the bias of the treatment effect $E[\widehat{p_1} - \widehat{p_0}] - (p_1 - p_0)$.

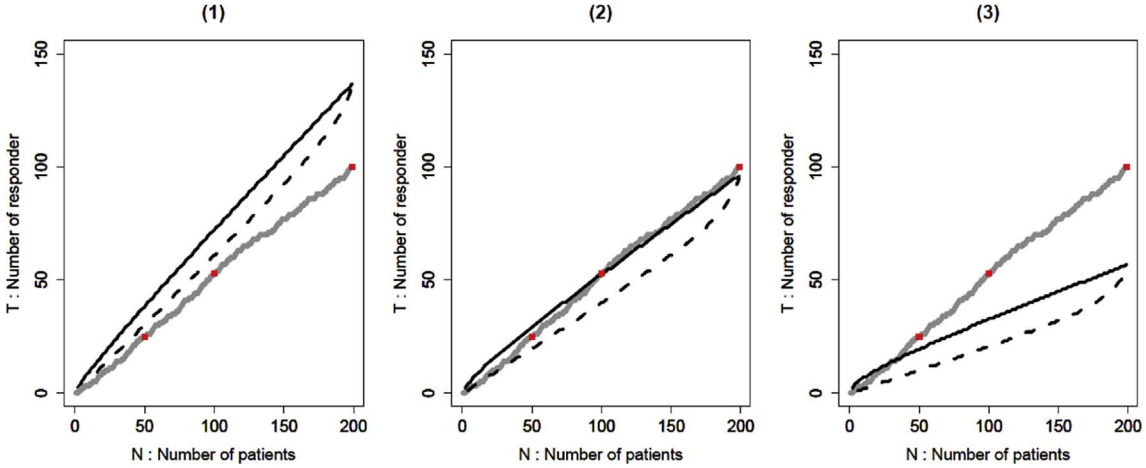

Fig. 2. N-T plot for E1A00 study; Black and solid curves are the expected N-T plots for superiority, black dashed curves are that for futility, and gray curves are the observed N-T plot for the case expected that (1) $\boldsymbol{p}_0$= 0.60, (2) $\boldsymbol{p}_0$= 0.40 and (3) $\boldsymbol{p}_0$= 0.20.

Acknowledgements

The illustrating data used in this paper was from the study that was coordinated by the ECOG-ACRIN Cancer Research Group (Robert L. Comis, MD and Mitchell D. Schnall, MD, PhD, Group Co-Chairs), which was supported by the National Cancer Institute of the National Institutes of Health under the following award numbers: CA180820, CA180794. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. The authors thank Dr. Beth Ann Griffin and Dr. Xuemin Fang for insightful comments and suggestions.

Appendix A

To obtain N-T plot for early termination for superiority or futility, computer programs in R [R Foundation for Statistical Computing] are available. We briefly describe the implementation with R (IAbin package). This package is available from the Comprehensive R Archive Network Web Site (http:// … TBD).

The arguments in the functions are determined in the design stage of a clinical trials, in which the endpoint is a binary, say, response or non-response. Here, *p0* is a value of the expected response rate in the control therapy. If the several possible rate of the outcome in the control therapy are considered, *p0* can be a vector, i.e., *p0 = c(0.20, 0.30, 0.40)*. *M* is an expected total sample size in both new therapy and control arms, and *q* is an allocation ratio of the new therapy arm ($0 < q < 1$). *alpha1* is a critical alpha at an interim analysis.

The following command gives a matrix names N, T, Z_score and P_val and automatically draws a N-T plot for early stopping for superiority with respect to p0:

plotNT.sup(*p0*, *M = 100*, *q = 0.5*, *alpha1 = 0.01*)

For early stopping for futility, the *plotNT.fut* function is used:

plotNT.fut(*p0*, *M = 100*, *q = 0.5*, *alpha1 = 0.01*, *cp1 = 0.2*)

Here, *cp1* is a critical conditional power at an interim analysis. With above two commands, we can obtain Fig. A for both early stopping for superiority and futility, with respect to *p0 = c(0.20, 0.40, 0.60)*.

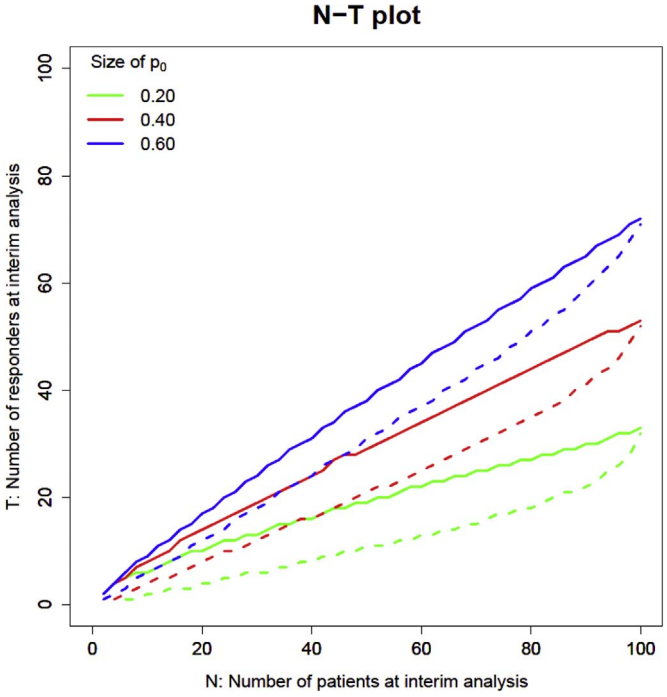

Fig. A. Example of N-T plot by functions in IAbin package.

Appendix B. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.conctc.2017.08.001.

References

[1] S.S. Ellenberg, T.R. Fleming, D.L. DeMets, Data Monitoring Committees in Clinical Trials: a Practical Perspective, John Wiley & Sons, Chichester, 2002.
[2] International Conference on Harmonization (ICH), Guidance for Industry: E9 Statistical Principles for Clinical Trials, Food and Drug Administrations, Rockville, MD, September 1998.
[3] G. Ball, Continuous safety monitoring for randomized controlled clinical trials with blinded treatment information; Part 4: one method, Contemp. Clin. Trials 32 (2011) S11–S17.
[4] S. Wen, G. Ball, J. Dey, Bayesian monitoring of safety signals in blinded clinical trial data, Ann. Public Health Res. 2 (2) (2015) 1019.
[5] B. Yao, L. Zhu, Q. Jiang, H. Amy Xia, Safety monitoring in clinical trials, Pharmaceutics 5 (2013) 94–106.
[6] P.F. Thall, R. Simon, Practical Bayesian guidelines for phase IIB clinical trials, Biometrics 50 (2) (1994) 337–349.
[7] K.K.G. Lan, D.L. DeMets, Discrete sequential boundaries for clinical trials, Biometrika 70 (3) (1983) 659–663.
[8] C. Jennison, B.W. Turnbull, Group Sequential Methods with Applications to Clinical Trials, John Wiley & Sons, Chichester, 2000.
[9] J.L. Haybittle, Repeated assessment of results in clinical trials of cancer treatment, Br. J. Radiol. 44 (1971) 93–797.
[10] R. Peto, M.C. Pike, et al., Design and analysis of randomized clinical trials requiring prolonged observation of each patient: Part I. introduction and design, Br. J. Cancer 34 (1976) 585–612.
[11] S.V. Rajkumar, E. Blood, D. Vesole, R. Fonseca, P.R. Greipp, Phase III clinical trial of thalidomide plus dexamethasone compared with dexamethasone alone in newly diagnosed multiple myeloma: a Clinical Trial Coordinated by the Eastern Cooperative Oncology Group, J. Clin. Oncol. 24 (3) (2006) 431–436.