Two-stage Design for Phase II Cancer Clinical Trials with Multiple Endpoints

Hui Gu[1], Yong Lin[1], Weichung Joe Shih[1], Yaqun Wang[1] and Kejian Liu[2]

[1] Department of Biostatistics, School of Public Health, Rutgers, 683 Hoes Lane West, Piscataway, NJ 08854

[2]Celgene Corporation, 300 Connell Drive, Berkeley Heights, NJ 07922

Abstract

The main purpose of a single-arm phase II cancer trial of a new regimen is to determine whether it has sufficient anti-tumor activity against a specific type of tumor to warrant its further clinical development. Such a research question can be answered under the framework of hypothesis testing. With the advent of targeted therapies that prolong disease stabilization, cancer patients typically experience stable disease (SD) rather than tumor shrinkage. It has been shown that patients with SD also achieve clinical benefits. Therefore, when evaluating the anti-tumor activity of a new treatment, clinicians are interested not only in overall response rate (complete or partial response(s)), but also in other types of measurements indicating clinical benefit. Taking two primary efficacy endpoints as an example, if the new treatment can improve on either endpoint(s), it may be promising for further evaluation. Therefore, "OR" logical relationship between the two primary efficacy endpoints is used when specifying the alternative hypothesis. In phase II cancer clinical trials, two-stage designs rather than single-stage ones are widely used for its possibility of early termination for futility to protect cancer patients. Motivated by two real cancer clinical trials, we propose a single-arm two-stage phase II cancer clinical trial design with two dichotomous alternative primary efficacy endpoints. Because of unknown correlation between two endpoints at the design stage, minimax rule is used to determine the optimal design, which minimizes the maximum of the expected sample size among all possible correlations, subject to the type I and II error constraints. Optimal designs for a variety of design parameters as well as the corresponding operating characteristics are provided.

Key Words: Two-stage optimal design, Phase II cancer trial, optimization, alternative primary endpoints

1. Introduction

The purpose of a phase IIA trial of a new anticancer drug is to determine whether the drug has sufficient activity against a specific type of tumor to warrant its further clinical development (phases IIB and phase III) (Simon, 1989). In terms of efficacy endpoints, historically, tumor response is an accepted endpoint to assess clinical benefit in phase II trials. Based on this single endpoint "tumor response", Gehan (1961), Fleming (1982), Simon (1989) have proposed two-stage (even multistage) trial design which allow early termination for futility to protect cancer patients. In recent years, our paradigm for understanding and treating cancer is changing. With the advent of targeted therapies that prolong disease stabilization, patients typically experience stable disease (SD) rather than tumor shrinkage (Mandrekar et al., 2010). It has been shown that patients with SD also achieve clinical benefit (Shepherd et al., 2005). Take one kinase inhibitor indicated for the treatment of unresectable hepatocellular carcinoma and advanced renal cell carcinoma, Sorafenib (NEXAVAR), for example. Clinical studies show that Sorafenib extends

progression-free survival (PFS) but the response rate is only 2% (Llovet et al., 2008). So the overall response rate (CR+PR) as used in Simon's two-stage design may not be appropriate to assess the anti-tumor activity of cytostatic drug such as Sorafenib. Therefore, when evaluating the anti-tumor activity of a new treatment to a specific type of tumor, clinicians are interested not only in CR+PR (in terms of tumor shrinkage), but also in other types of measurements indicating clinical benefit (such as PFS). If the new treatment can improve on either endpoint(s), it may be promising for further evaluation. Such relationship between endpoints belongs to the type of "alternative primary endpoints" (Offen et al., 2007). Very few studies, if any, have discussed phase II trial designs with "alternative primary endpoints", although several authors (Bryant and Day, 1995; Thall and Cheng, 2001; Conaway and Petroni, 1995) have proposed two-stage phase II clinical trial designs considering both efficacy and toxicity, which belongs to the category of "multiple co-primary endpoints". In this article we discuss the development of a single-arm, two-stage design for phase II cancer clinical trials to answer the research question of determining whether the new treatment has sufficient anti-tumor activity for further evaluation when there are two alternative dichotomous primary endpoints of efficacy.

2. Method

The research question of an initial rough estimate of the degree of antitumor activity of the treatment or drug can be answered under the framework of hypothesis testing. In the context of phase II cancer clinical trials, denote the probabilities of success for each of the binary efficacy endpoints as π_1 and π_2, respectively. The research question then can be translated into the hypotheses as follows:

$$H_0 : \pi_1 \leq p_1^{(0)} \text{ and } \pi_2 \leq p_2^{(0)}$$
$$H_A : \pi_1 \geq p_1^{(A)} \text{ or } \pi_2 \geq p_2^{(A)}$$

where $p_1^{(0)}$ and $p_2^{(0)}$ are specified values that are believed to be uninteresting or comparable to the current standard of care, and $p_1^{(A)}$ and $p_2^{(A)}$ are the targeted response rate, and $p_1^{(0)} < p_1^{(A)}$ and $p_2^{(0)} < p_2^{(A)}$. In two-stage design settings, let X_1 and X denote the total number of responses at the end of stage 1 and stage 2 for endpoint 1, respectively, while letting Y_1 and Y denote the corresponding quantities for endpoint 2. And (s_1, s) and (t_1, t) are critical values associated with the occurrence of endpoint 1 and endpoint 2, respectively.

The trial proceeds as follows:

- Accrue n_1 patients in stage 1. If $X_1 \leq s_1$ and $Y_1 \leq t_1$, terminate the trial due to futility;
- Otherwise, accrue additional $(n - n_1)$ patients into the second stage.

Recommend the treatment only if $\{(X_1 > s_1 \text{ or } Y_1 > t_1) \text{ and } (X > s \text{ or } Y > t)\}$. Then a two-stage design can be specified by a vector of six parameters $Q = (n, n_1, s_1, t_1, s, t)$. The goal of this study is, given pre-specified (α, β_1, β_2, β, $p_1^{(0)}$, $p_2^{(0)}$, $p_1^{(A)}$, $p_2^{(A)}$), to search for feasible solutions of Q that satisfy type I/II error constraints, and then use optimality criteria to find the "optimal" design.

2.1 Derivation of the power function

The original individual data in a single-stage design with n patients can be collapsed into 4 response patterns, which can be represented by a 2×2 table as follows:

Table 1: Response Pattern

Pattern	*Endpoint* 1	*Endpoint* 2	*total obs.*
1	Yes	Yes	n_{11}
2	Yes	No	n_{10}
3	No	Yes	n_{01}
4	No	No	n_{00}

Table 2: Observed Counts

		Endpoint 1		
		Yes	No	
Endpoint 2	Yes	$C_{11} = n_{11}$	$C_{01} = n_{01}$	$Y = n_{+1}$
	No	$C_{10} = n_{10}$	$C_{00} = n_{00}$	
		$X = n_{1+}$		n

Table 3: Probability

		Endpoint 1		
		Yes	No	
Endpoint 2	Yes	π_{11}	π_{01}	π_2
	No	π_{10}	π_{00}	
		π_1		

The random quantities in each cell of the 2×2 table (Table 2), $(C_{11}, C_{10}, C_{01}, C_{00})$, are distributed as:

$$(C_{11,}\ C_{10,}\ C_{01,}\ C_{00}) \sim Multinomial(n, (\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00}))$$

The joint distribution of (X, Y) can be calculated via multinomial probability mass function as:

$$\begin{aligned} & P(X = x, Y = y) = p(x, y; n, \pi_1, \pi_2, \pi_{11}) \\ = & \sum_{\max(0,\ x+y-n) \le n_{11} \le \min(x,\ y)} \binom{n}{n_{11}, x - n_{11}, y - n_{11}, n - x - y + n_{11}} \times \pi_{11}^{n_{11}} \\ & \times (\pi_1 - \pi_{11})^{x - n_{11}} \times (\pi_2 - \pi_{11})^{y - n_{11}} \times (1 - \pi_1 - \pi_2 + \pi_{11})^{n - x - y + n_{11}} \end{aligned}$$

where

$$\max(0, \pi_1 + \pi_2 - 1) \le \pi_{11} \le \min(\pi_1, \pi_2),$$

and π_{11} is used to describe the correlation between endpoint 1 and endpoint 2 within the same individual.

Let

$$D(s,t;n,\pi_1,\pi_2,\pi_{11}) = Pr(X \le s, Y \le t \mid X \sim Bin(n,\pi_1), Y \sim Bin(n,\pi_2),\pi_{11}) = \sum_{x=0}^{s}\sum_{y=0}^{t} p(x,y;n,\pi_1,\pi_2,\pi_{11}),$$

then the power function for two-stage designs can be written as:

$$\begin{aligned}
&G_t(\mathbf{Q},\pi_{11}|H) = Pr(\text{Recommend Treatment}|\mathbf{Q},\pi_{11},H)\\
=&G_t(n_1,n,s_1,t_1,s,t,\pi_{11},\pi_1,\pi_2)\\
=&Pr(X_1 > s_1 \text{ or } Y_1 > t_1),(X > s \text{ or } Y > t))\\
=&Pr((X > s \text{ or } Y > t)) - Pr(X_1 \le s_1, Y_1 \le t_1, (X > s \text{ or } Y > t))\\
=&1 - Pr(X \le s, Y \le t) - Pr((X > s \text{ or } Y > t)|X_1 \le s_1, Y_1 \le t_1)Pr(X_1 \le s_1, Y_1 \le t_1)\\
=&1 - D(s,t;n,\pi_1,\pi_2,\pi_{11}) - D(s_1,t_1;n_1,\pi_1,\pi_2,\pi_{11})\\
&+ \sum_{i=0}^{s_1}\sum_{j=0}^{t_1}\{D(s-i,t-j;n_2,\pi_1,\pi_2,\pi_{11}) \times p(i,j;n_1,\pi_1,\pi_2,\pi_{11})\}
\end{aligned}$$

2.2 Properties of the power function and the expected total sample size

We have used the method of mathematical induction (MI) to prove the following properties of the power function $G_t(\mathbf{Q},\pi_{11}|H)$ and the expected total sample size $E(N|H_0,\mathbf{Q},\pi_{11})$:

1. The power function is non-decreasing in π_1 given $\mathbf{Q}$, π_{11}, and π_2. The same applies to π_2.
2. The power function is non-increasing in t given (n_1,n,s_1,t_1,s).
3. The expected sample size under the null hypothesis, $E(N|H_0,\mathbf{Q},\pi_{11})$, is non-increasing in π_{11} given $\mathbf{Q}$, π_1 and π_2 .

We have used both simulation and method of mathematical induction to find that the power function is not monotone in π_{11} given $\mathbf{Q}$, π_1 and π_2 .

The above property 1 makes it possible to use hypothesis testing on simple null and alternative hypotheses as a reasonable substitute for hypothesis testing on composite null and alternative hypotheses. Therefore, the design parameters $Q = (n,n_1,s_1,t_1,s,t)$ may be determined by solving:

$$\min_{Q}\max_{\pi_{11}} E(N|Q,\pi_{11},H_0:\pi_1 = p_1^{(0)},\pi_2 = p_2^{(0)}), \tag{1}$$

subject to

$$\max_{\pi_{11}} G(Q,\pi_{11}|H_0:\pi_1 = p_1^{(0)},\pi_2 = p_2^{(0)}) \le \alpha, \tag{2a}$$

$$\min_{\pi_{11}} G(Q,\pi_{11}|H_{A_1}:\pi_1 = p_1^{(A)},\pi_2 = p_2^{(0)}) \geqslant 1-\beta_1, \tag{2b}$$

$$\min_{\pi_{11}} G(Q,\pi_{11}|H_{A_2}:\pi_1 = p_1^{(0)},\pi_2 = p_2^{(A)}) \geqslant 1-\beta_2, \tag{2c}$$

$$\min_{\pi_{11}} G(Q,\pi_{11}|H_{A_3}:\pi_1 = p_1^{(A)},\pi_2 = p_2^{(A)}) \geqslant 1-\beta. \tag{2d}$$

If $\beta_1 = \beta_2 = \beta$, then (2d) is included in (2b) or (2c).

We found that more than one, actually many, feasible solutions share the same minimum expected sample size of N under the null hypothesis because of discreteness of the underlying bivariate binomial distribution and the small difference in the value of E(N| H_0) between feasible solutions sharing the same (n, n_1, s_1, t_1). So the optimality criteria for the optimal design now is:

- Minimum $E(N|H_0)$;
- Maximum type I error(closer to nominal level) since there are three type II errors and the directions of the magnitude of them are not the same in most time.

2.3 Searching Algorithm

Due to the introduction of the correlation parameter π_{11} into the bivariate joint distribution, the time cost in optimization and exhaustive searching in the nested loops of $n \to n_1 \to s_1 \to t_1 \to s \to t$ has increased dramatically. Theorems in Bryant and Day (1995) have inspired us to adopt a pre-screening strategy via starting searching assuming the two alternative primary efficacy endpoints are independent. Find those feasible solutions satisfying type I/II error constraints under this independence assumption, and sort them by the optimality criteria (which is, minimum expected sample size under H_0 and maximum type I error rate closer to the nominal level) under the independence assumption. And then, among the top 5% of the sorted feasible solutions, relax the independence assumption, do computation-intensive calculations of real maximized type I/II error rates allowing π_{11} to assume any values in its defined range, and search and locate the optimal design after applying the optimality criteria.

3. Results

The following table is part of the found optimal designs for a variety of design parameters with $p_1^{(0)} \leq p_2^{(0)}$. The operating characteristics for each of the optimal two-stage sequential design (including the maximized type I error rate, minimized powers and the minimized value of maximum possible expected sample size under the null hypothesis in the defined range of π_{11}) are presented as well. The following notations are used in these tables

$$
\begin{aligned}
G_0(Q, \pi_{11}) &= G(Q, \pi_{11}|H_0 : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(0)}),\\
G_1(Q, \pi_{11}) &= G(Q, \pi_{11}|H_{A_1} : \pi_1 = p_1^{(A)}, \pi_2 = p_2^{(0)}),\\
G_2(Q, \pi_{11}) &= G(Q, \pi_{11}|H_{A_2} : \pi_1 = p_1^{(0)}, \pi_2 = p_2^{(A)}),\\
G_3(Q, \pi_{11}) &= G(Q, \pi_{11}|H_{A_3} : \pi_1 = p_1^{(A)}, \pi_2 = p_2^{(A)}).
\end{aligned}
$$

Table 4: Given $(\alpha, \beta_1, \beta_2) = (0.05, 0.20, 0.20)$. $\delta = 0.20$.
Reject H_0 if $(X_1 > s_1$ or $Y_1 > t_1)$ and $(X > s$ or $Y > t)$.

$p_1^{(0)}$	$p_1^{(A)}$	$p_2^{(0)}$	$p_2^{(A)}$	(n, n_1, s_1, t_1, s, t)	$\max_{\pi_{11}} G_0(Q, \pi_{11})$	$\min_{\pi_{11}} G_1(Q, \pi_{11})$	$\min_{\pi_{11}} G_2(Q, \pi_{11})$	$\min_{\pi_{11}} G_3(Q, \pi_{11})$	$\max_{\pi_{11}} E(N\|H_0, Q, \pi_{11})$
0.05	0.25	0.05	0.25	(25, 12, 1, 1, 3, 3)	0.0497	0.8061	0.8061	0.8062	15.0
		0.1	0.3	(27, 15, 1, 2, 4, 5)	0.0491	0.8161	0.8101	0.8688	19.0
		0.2	0.4	(37, 17, 2, 4, 5, 11)	0.0490	0.8130	0.8040	0.9010	22.6
		0.3	0.5	(39, 19, 2, 6, 6, 16)	0.0483	0.8308	0.8041	0.9236	26.8
		0.4	0.6	(53, 21, 3, 10, 7, 26)	0.0470	0.8045	0.8010	0.9223	27.1
		0.5	0.7	(39, 26, 3, 15, 6, 24)	0.0496	0.8623	0.8018	0.9491	28.6
		0.6	0.8	(38, 18, 2, 12, 6, 27)	0.0477	0.8066	0.8119	0.9455	23.2
		0.7	0.9	(28, 13, 1, 10, 4, 23)	0.0494	0.8082	0.8045	0.9513	17.9
0.1	0.3	0.1	0.3	(38, 14, 2, 2, 7, 7)	0.0460	0.8053	0.8053	0.8054	21.2
		0.2	0.4	(41, 22, 4, 6, 8, 12)	0.0499	0.8053	0.8042	0.8760	25.6
		0.3	0.5	(47, 22, 4, 8, 9, 19)	0.0462	0.8142	0.8013	0.9061	28.0
		0.4	0.6	(45, 22, 4, 10, 9, 23)	0.0491	0.8030	0.8035	0.9195	28.7
		0.5	0.7	(49, 18, 3, 10, 9, 30)	0.0482	0.8227	0.8004	0.9360	28.2
		0.6	0.8	(43, 18, 3, 12, 8, 31)	0.0456	0.8175	0.8005	0.9437	25.5
		0.7	0.9	(38, 14, 2, 11, 7, 31)	0.0489	0.8115	0.8215	0.9515	21.7
0.2	0.4	0.2	0.4	(50, 22, 6, 6, 15, 15)	0.0446	0.8025	0.8025	0.8027	29.3
		0.3	0.5	(54, 22, 6, 8, 16, 22)	0.0480	0.8110	0.8038	0.8777	31.9
		0.4	0.6	(55, 25, 7, 12, 17, 28)	0.0418	0.8008	0.8011	0.8969	32.8
		0.5	0.7	(59, 22, 6, 13, 18, 35)	0.0498	0.8035	0.8021	0.9157	32.4

4. Discussion

In this paper, we propose a two-stage optimal design for a single-arm phase II cancer clinical trial with two alternative binary primary efficacy endpoints under a variety of parameter settings. Since the two alternative primary efficacy endpoints within a patient are correlated, the inclusion of the nuisance correlation parameter has made the joint distribution and the power function more complicated for the study design. Sill et al. (2012) has mentioned the necessity of considering the case with two alternative primary efficacy endpoints although they used different terminology. They only considered three relatively extreme cases for the correlation parameter: independent, partially and fully dependent. This paper, however, has considered all possible values of this correlation parameter since we may not have much information about this nuisance correlation parameter at the design stage and we want to be conservative. The searching results show that the correlation parameter π_{11} may assume different values to achieve the maximized type I error rate, minimized powers, the minimum of the maximized value among all possible expected sample sizes under the null hypothesis. Due to the time cost of thorough searching in the defined range of π_{11}, we only did the computation-intensive calculations of real maximized type I and II error rates among the top 5% of sorted feasible solutions from independence assumption, so the resulting designs we got may not be global optimal, but close to as shown in Bryant and Day (1995). The optimal two-stage designs and the corresponding operating characteristics in this study can be referenced when planning a phase II cancer clinical trial with two binary alternative primary efficacy endpoints. Statistical inference procedures for this two-stage optimal design are under development.

References

Richard Simon. Optimal two-stage designs for phase ii clinical trials. *Controlled Clinical Trials*, (10):1–10, 1989.

Edmund Gehan. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *Journal of Chronic Diseases*, 13(4):346–353, 1961.

Thomas R. Fleming. One-sample multiple testing procedure for phase ii clinical trials. *Biometrics*, 38(1):143–151, 1982.

S. J. Mandrekar, Yingwei Qi, S. L. Hillman, KL Allen Ziegler, NF Reuter, KM Rowland, SA Kuross, RS Marks, SE Schild, and AA Adjei. Endpoints in phase ii trials for advanced non-small cell lung cancer. *Journal of Thoracic Oncology*, 5(1):3–9, 2010.

FA Shepherd, PJ Rodrigues, T Ciuleanu, and et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med*, 353:123–32, 2005.

Josep M. Llovet, et al., and the SHARP Investigators Study Group. Sorafenib in advanced hepatocellular carcinoma. *New England Journal of Medicine*, 359(4): 378–90, 2008.

Walter Offen, Christy Chuang-Stein, Alex Dmitrienko, Gary Littman, Jeff Maca, Laura Meyerson, Robb Muirhead, Paul Stryszak, Alex Boddy, Kun Chen, Kati Copley-Merriman, Willard Dere, Sam Givens, David Hall, David Henry,

Joseph D. Jackson, Alok Krishen, Thomas Liu, Steve Ryder, A. J. Sankoh, Julia Wang, and Chyon-Hwa Yeh. Multiple co-primary endpoints: Medical and statistical solutions - a report from the multiple endpoints expert team of the pharmaceutical research and manufacturers of america. *Drug Information Journal*, 41(1):31–46, 2007.

John Bryant and Roger Day. Incorporating toxicity considerations into the design of two-stage phase ii clinical trials. *Biometrics*, 51(4):1372–1383, 1995.

Peter F. Thall and Su-Chun Cheng. Optimal two-stage designs for clinical trials based on safety and efficacy. *Statistics in Medicine*, 20(7):1023–1032, 2001.

Mark R. Conaway and Gina R. Petroni. Bivariate sequential designs for phase-ii trials. *Biometrics*, 51(2):656–664, 1995.

Michael W Sill, Larry Rubinstein, Samuel Litwin, and Greg Yothers. A method for utilizing co-primary efficacy outcome measures to screen regimens for activity in two-stage phase ii clinical trials. *Clinical Trials*, 9(4):385–395, 2012.