MAIN PAPER

Pharmaceutical Statistics

Received 7 February 2009, Revised 22 December 2009, Accepted 11 January 2010 Published online 22 February 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/pst.418

An alternative phase II/III design for continuous endpoints

Wong-Shian Huang,[a,b†] Jen-pei Liu,[b,c†] and Chin-Fu Hsiao[b*]

The success rate of drug development has been declined dramatically in recent years and the current paradigm of drug development is no longer functioning. It requires a major undertaking on breakthrough strategies and methodology for designs to minimize sample sizes and to shorten duration of the development. We propose an alternative phase II/III design based on continuous efficacy endpoints, which consists of two stages: a selection stage and a confirmation stage. For the selection stage, a randomized parallel design with several doses with a placebo group is employed for selection of doses. After the best dose is chosen, the patients of the selected dose group and placebo group continue to enter the confirmation stage. New patients will also be recruited and randomized to receive the selected dose or placebo group. The final analysis is performed with the cumulative data of patients from both stages. With the pre-specified probabilities of rejecting the drug at each stage, sample sizes and critical values for both stages can be determined. As it is a single trial with controlling overall type I and II error rates, the proposed phase II/III adaptive design may not only reduce the sample size but also improve the success rate. An example illustrates the applications of the proposed phase II/III adaptive design. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: type I error rate; power; sample size

1. INTRODUCTION

Despite of a better understanding of disease etiology, a rapid increase in resources, technological advance, and seemingly large amount of potential candidates, the performance of drug development is disappointing. The success rate of drug development has been declined drastically in recent years [1]. One of the many possible reasons for the dissatisfying performance is that the current paradigm for drug development is no longer functioning for the 21st century. Therefore, new concepts, strategies, and methodologies are urgently needed to reduce the development cost and to shorten the development duration.

In recent years, the use of adaptive design methods in clinical research and development based on accrued data has become very popular due to its flexibility and efficiency. One of the most commonly considered adaptive designs in clinical research and development is probably a two-stage seamless adaptive trial design. In general, the objective of the traditional paradigm for the phase II and III drug development is to confirm the efficacy of the doses selected from the phase II development in the pivotal phase III trials. Issues concerning the current phase II and III paradigm include different patient populations recruited for phase II and phase III trials, inability to establish the dose–response relationship for the test drug, possible different primary efficacy endpoints used in the phase II and III trials, and large sample sizes because of multiple phase II and III trials with individualized type I and II error rates. Therefore, interest in combining two separate studies into a single study has developed. An adaptive seamless phase II/III trial design is a trial design that combines two separate trials (i.e. a phase IIb and a phase III trial) into one trial and would use data from patients enrolled before and after the adaptation in the final analysis. Such designs can reduce the lead time that would have occurred between the trials that they had been conducted separately, and thus possibly shorten the duration of the trials and reduce the development cost.

For the cytotoxic agents for cancer treatment, a combined phase II/III design has been suggested for time-to-event endpoint [2,3]. Follmann *et al*. [4] consider comparison of several treatments and a control as a testing problem of the global null hypothesis that all treatments are equal in the framework of group sequential tests with Pocock and O'Brien-Fleming-type boundaries chosen to maintain the type I error rate. On the other hand, Todd and Stallard [5] then consider same study objectives but different primary endpoints for selection and confirmation stages. More specifically, they propose a group sequential design that incorporates treatment selection based on a short-term endpoint, followed by a confirmation stage comparing the selected treatment with control using a longer-term primary endpoint.

[†]*These two authors contributed equally to this work.*

[a]*Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan*

[b]*Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli County, Taiwan*

[c]*Division of Biometry, Department of Agronomy, National Taiwan University, Taipei, Taiwan*

**Correspondence to: Chin-Fu Hsiao, Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, No. 35, Keyan Road, Zhunan Town, Miaoli County 350, Taiwan.*
E-mail: chinfu@nhri.org.tw

One of the most frequently types of the variables for evaluation of efficacy is the continuous endpoints such as total cholesterol level, sitting diastolic blood pressure, HbA1c, and many others. Based on continuous endpoints, Tsou *et al*. [6] propose a two-stage screening design that minimizes the expected sample size if the new candidate has no or low efficacy activity subject to the constraints of the type I and type II error rates. Bischoff and Miller [7] also develop a competing two-stage group-sequential adaptive design with a minimal expected number of patients that controls the type I error rate, achieves a desired power to detect a given clinically relevant difference in means, and controls the probability of wrong selection. In their approach, they test the hypothesis that each of the treatment effects is smaller than or equal to the control effect against the hypothesis that the effect of one of the treatments is larger than the control effect for the whole trial.

In seamless phase II/III design, the drop-the-loser mechanism is used very often at the phase II clinical stage especially when there are uncertainties regarding the dose levels [8]. A drop-the-losers design is a design that allows dropping the inferior treatment groups. Sill and Sampson [9] explored an inferential technique for drop-the-losers designs for the binomial distribution setting. Chang [10] proposed an adaptive seamless phase II/III design that uses the weak α-control method for the drop-the-losers design based on a contrast test at the phase II stage. Similarly, in this paper, we consider the phase II and III development as a single trial with controlling overall type I and II error rates. Our proposed alternative phase II/III design consists of two stages: a selection stage and a confirmation stage. Most importantly, in this paper, for the selection stage, a randomized parallel design with several doses and a concurrent placebo group is employed for characterization of the dose–response relationship and selection of doses for the confirmation stage. In our phase II/III design, the drop-the-losers mechanism is also used at the phase II stage. However, different from Chang [10], our approach for selection of doses during selection phase is based on the magnitude of the slope of regression line and is not based on the contrasts of group means. After the best dose is chosen, the patients of the selected dose group and placebo group will be continued to enter the confirmation stage. In addition, new patients will be recruited and randomized to receive the selected dose or the placebo group for the confirmation stage. At completion of the confirmation stage, the final analysis is performed with the cumulative data of patients from both stages. With the pre-specified probabilities of rejecting the drug at each stage, the sample sizes and critical values for both stages can be determined. Special features of our proposed phase II/III design are: (1) the same targeted patient population is evaluated in both stages with the same primary continuous endpoints, evaluation criteria, schedules under the same experimental conditions specified in the same protocol, (2) an empirical dose–response relationship is used for selection of the dose for the confirmation stage, and (3) a portion of the patients in the selection stage is also evaluated in the confirmation stage. In the next section, the current approach to the phase II and III development is reviewed. The proposed alternative phase II/III design is introduced in Section 3. Determination of sample sizes and critical values are also provided in this section. An example for illustration of the applications of our proposed design is presented in Section 4. Discussion and final remarks are provided in Section 5.

2. THE CURRENT APPROACH

For the purpose of illustration, during the phase II development, we only consider on a phase II trial using the randomized parallel-group design for comparing several doses of a test product, $d_1, d_2, \ldots, d_k$ with a concurrent placebo control, d_0. In addition, we also assume that one of the k doses is selected for a single phase III confirmation trial if it is statistically significant. Let Y_{ij} be some observed continuous endpoint for patient j assigned to dose d_i, $i=0,\ldots, k$; $j=1,\ldots, n_2$; where n_2 is the sample size per group for the phase II trial. We also assume that Y_{ij} follows a normal distribution with mean μ_i and known variance σ^2, $i=0,1,\ldots, k$. The current approach to selection of the doses for the phase III confirmation trials is based on the following hypothesis:

$$H_0 : \mu_i - \mu_0 \leqslant 0 \quad \text{vs} \quad H_A : \mu_i - \mu_0 > 0, \quad i = 1, \ldots, k. \tag{1}$$

Let $\bar{Y}_{\text{II}i}$ be the sample mean of group i, $i = 0,1,\ldots, k$, and s_{II}^2 the pooled sample variance obtained from the phase II trial. A dose is declared to have a statistically significantly superior efficacy over placebo if

$$t_{\text{II}} = (\bar{Y}_{\text{II}i} - \bar{Y}_{\text{II}0})/s_{\text{II}}\sqrt{2/n_2} > z_{1-\alpha},$$

for no adjustment

or

$$t_{\text{IIB}} = (\bar{Y}_{\text{II}i} - \bar{Y}_{\text{II}0})/s_{\text{II}}\sqrt{2/n_2} > z_{1-\alpha/k}, \tag{2}$$

with Bonferroni correction,

where $z_{1-\alpha}$ is the upper αth quantile of the standard normal distribution.

Let δ be the required minimal clinically meaningful improvement on efficacy for a dose to be selected for the phase III trial. Without considering adjustment for multiple comparisons, the sample size required for each dose group to provide a power of $(1-\beta)$ at the α significance level is given by

$$n' = 2(z_{1-\alpha}+z_{1-\beta})^2/(\delta/\sigma)^2 \tag{3}$$

If the Bonferroni correction is employed to adjust p-values for multiple comparisons, the required sample size is given by

$$n'^{\text{B}} = 2(z_{1-\alpha/k}+z_{1-\beta})^2/(\delta/\sigma)^2 \tag{4}$$

On the other hand, instead of conducting pairwise comparison, we may want to test whether the dose–response relationship can be described by the simple linear regression as follows:

$$E(Y_{ij}) = \xi + \eta d_i$$

where ξ is the intercept and η is the slope. Let n'^{L} be the required sample size per dose level in traditional phase II trial for dose–response to test the null hypothesis H_0: $\eta \leqslant c$ against H_A: $\eta > c$. In this case, the sample size required for each dose group can be calculated by

$$n'^{L} = \left(\frac{\sigma}{c'-c}\right)^2 \frac{(Z_{1-\alpha}+Z_{1-\beta})^2}{\sum_{i=0}^{k}(d_i-d_0)^2} \tag{5}$$

with a specified value c' under the alternative hypothesis.

Suppose dose r is selected from the phase II trial, confirmation of the efficacy in the phase III trial is based on the following hypothesis:

$$H_0 : \mu_r - \mu_0 = 0 \quad \text{vs} \quad H_A : \mu_r - \mu_0 \neq 0 \tag{6}$$

Let Δ' be the minimal clinically meaningful requirement for confirmation of a superior efficacy over the placebo in the phase III trial. The sample size required to provide a power of $(1-\beta)$ at the α significance level is given as

$$n'' = 2(z_{1-\alpha/2}+z_{1-\beta})^2/(\Delta'/\sigma)^2 \quad (7)$$

As mentioned before, the phase II trials of the current approach are conducted in much restricted experimental conditions with much stringent inclusion/exclusion criteria. On the other hand, the phase III trials are usually conducted in a much heterogeneous patient population under an environment much close to the clinical practice. Therefore, the patient populations and experimental conditions are different between phase II and phase III trials. Although it is assumed that the variance is same for both phase II and III trials, in fact, the variability observed from the phase III trials is much larger than that of the phase II trials. On the other hand, the restricted patient population and the tight experiment environment of the phase II trials are often to over-estimate the efficacy of the test drug in the clinical practice. Consequently, the efficacy of the test drug may not be confirmed during the phase III trials.

For the current approach, the phase II and phase III trials are conducted as separate trials in a sequential manner. Therefore, the duration of the current phase II and III development is longer than necessary. In addition, phase II and III trials are conducted independently with individual type I and II error rates. As a result, the overall sample size is also larger than it should be. Suppose that we only consider one phase II trial and one phase III trial. Chow and Chang [11] point out that in the current approach, the actual overall type I error rate (α) is equal to $\alpha_{II}\alpha_{III}$, where α_{II} and α_{III} are the type I error rates controlled at phase II and phase III respectively. That is, α for the current design is too conservative. On the other hand, if power refers to the probability of correctly detecting a true but not hypothetical treatment difference, then in the current design, the actual power is given by

$$\text{power} = \text{power}_{II} * \text{power}_{III}$$

Consequently, when power_{II} and power_{III} are not large enough, the actual power for the current approach might be small [11]. In other words, the current approach has a very low type I error rate and may not provide sufficient power either. New strategies and methodology are urgently needed to improve the current situation.

3. THE PROSPOSED ALTERNATIVE PHASE II/III DESIGN

Our proposed alternative phase II/III design is a single trial consisting of two stages: a selection stage and a confirmation stage. Because it is a single trial, the inclusion/exclusion criteria, evaluation methods and schedules, primary and secondary efficacy endpoints, and other experimental conditions are the same and are pre-specified in the same protocol. The first stage is the selection stage in which the patients are randomly assigned to receive either one of the k doses of the test drug or to the placebo group. Although there are many different forms of dose–response relationship, for the purpose of illustration, during the selection stage, the dose–response relationship is assumed linear and can be investigated by the simple linear regression method. If the slope is not greater than some pre-specified threshold, say c, then the trial stops and the test drug is concluded no efficacy. On the other hand, if the slope is greater than the specified threshold, the lowest dose meeting the minimal clinically meaningful requirement, say δ, is selected for the confirmation stage of the trial. In addition, the patients in the selected dose and placebo groups will be continued to the confirmation stage. In addition, new patients will be recruited and randomized to receive either the selected dose of the test drug or to the placebo group. The final analysis includes the data of the selected dose and placebo groups from both selection and confirmation stages.

Suppose that the dose–response relationship can be described by the simple linear regression as follows:

$$E(Y_{ij}) = \xi+\eta d_l$$

where ξ is the intercept and η is the slope. It follows that the corresponding hypothesis for the selection stage is given as

$$H_0^{II} : \eta \leqslant c \text{ vs } H_A^{II} : \eta > c \quad (8)$$

where $c \geqslant 0$ is some pre-specified threshold.

If we fail to reject the null hypothesis of Equation (8) at a pre-specified significance level, then the trial stops and the drug is eliminated from consideration of further development. On the other hand, dose r is selected for the confirmation stage if the null hypothesis of Equation (8) is rejected at the pre-specified significance level and it is the lowest among the doses such that the efficacy is better than the placebo by a magnitude greater than δ, where δ is the required minimal clinically meaningful improvement on efficacy for a dose to be selected for the confirmation stage. The sample size can then be determined to ensure that if some doses are superior to placebo by the pre-specified amount δ, a dose will be selected with high probability, say 0.8.

Let n_2 be the sample size per group for the selection stage. The least squares estimator of η is given as

$$\hat{\eta} = \frac{\sum_{i=0}^{k}\sum_{j=1}^{n_2}(d_i-\bar{d})(Y_{ij}-\bar{Y})}{\sum_{i=0}^{k} n_2(d_i-\bar{d})^2} \quad (9)$$

where

$$\bar{d} = \frac{\sum_{i=0}^{k} d_i}{k+1}$$

and

$$\bar{Y} = \frac{\sum_{i=0}^{k}\sum_{j=1}^{n_2} Y_{ij}}{n_2(k+1)}$$

In addition, $\hat{\eta}$ follows a normal distribution with mean η and variance

$$\frac{\sigma^2}{\sum_{i=0}^{k} n_2(d_i-\bar{d})^2}$$

Assume that we will reject H_0^{II} if $\hat{\eta} \geqslant C_2$ and dose d_r and the placebo group (d_0) are chosen into the confirmation stage. Denote the population means of dose r and placebo by μ_r and μ_0, respectively, and $\Delta = \mu_r - \mu_0$. The hypothesis for the confirmation stage is given as

$$H_0^{III} : \Delta = 0 \text{ vs } H_A^{III} : \Delta \neq 0. \quad (10)$$

The estimator of Δ is $\hat{\Delta} = \bar{Y}_r^* - \bar{Y}_0^*$ where

$$\bar{Y}_r^* = \frac{\sum_{j=1}^{n_2} Y_{rj} + \sum_{j=n_2+1}^{n_2+n_3} Y_{rj}}{n_2+n_3}$$

$$\bar{Y}_0^* = \frac{\sum_{j=1}^{n_2} Y_{0j} + \sum_{j=n_2+1}^{n_2+n_3} Y_{0j}}{n_2+n_3}$$

and n_3 is the sample size of the new patients per group required for the confirmation stage. Consequently, $\hat{\Delta}$ follows a normal distribution with mean $\mu_r-\mu_0$ and variance $2\sigma^2/(n_2+n_3)$.

As our proposed phase II/III design is a single trial, it follows that the overall failure rate is α and the overall power is $1-\beta$. To control the overall type I error rate, our approach is similar to Simon's two-stage design [12]. We will terminate the experiment at the end of the phase II stage and reject the drug if $\hat{\eta} < C_2$. This will occur with probability $P_c(\hat{\eta} < C_2)$. We will reject the drug at the end of the phase III stage if $|\hat{\Delta}| \geqslant C_3$. Consequently, the overall probability of rejecting the new drug in both stages with the true parameters η and Δ is a function of $\eta, \Delta, n_2, n_3, C_2, C_3$, and σ and it is given as

$$\begin{aligned}&\varphi(\eta, \Delta, n_2, n_3, C_2, C_3, \sigma)\\ &= P_\eta(\hat{\eta} < C_2) + \int_{C_2}^{\infty} f_{\hat{\eta}}(b) P_\Delta(-C_3 < \hat{\Delta} < C_3)\,db,\end{aligned} \tag{11}$$

where P_η and P_Δ denote the probability measure with respect to η and Δ respectively, and $f_{\hat{\eta}}(\cdot)$ represents the probability density function of $\hat{\eta}$ with respect to η. Then the overall type I error rate evaluated at $\eta = c$ and $\Delta = 0$ is given as

$$\alpha = 1 - [P_c(\hat{\eta} < C_2) + \int_{C_2}^{\infty} f_{\hat{\eta}}(b) P_0(-C_3 < \hat{\Delta} < C_3)\,db] \tag{12}$$

Equation (12) can be re-expressed as

$$1 - \alpha = P_c(\hat{\eta} < C_2) + \int_{C_2}^{\infty} f_{\hat{\eta}}(b) P_0(-C_3 < \hat{\Delta} < C_3)\,db$$

As our proposed phase II/III design is to eliminate ineffective drugs and doses as early as possible, unlike the traditional sequential design, a weighing factor γ_1 is used to determine how much the overall failure rate is spent in the two stages such that

$$P_c(\hat{\eta} < C_2) = \gamma_1(1-\alpha) \tag{13}$$

and

$$\int_{C_2}^{\infty} f_{\hat{\eta}}(t) P_0(-C_3 < \hat{\Delta} < C_3)\,dt = (1-\gamma_1)(1-\alpha) \tag{14}$$

where $0 < \gamma_1 < 1$. Note that the larger the γ_1 is, the larger the C_2 is. Also larger γ_1 indicates that we spend fewer type I error rate for phase II stage. In addition, the objectives of the proposed design are to eliminate the inefficacious drugs or doses as early as possible and to ensure a better success rate for the selected doses in the confirmation stage, we suggest that γ_1 be greater than 0.5, say 0.6 or above, be spent at the selection stage under $c = 0$.

The overall type II error with a specified value c' and Δ' under the alternative hypotheses Equations (8) and (10) is given as

$$\begin{aligned}\beta &= \varphi(c', \Delta', n_2, n_3, C_2, C_3, \sigma)\\ &= P_{c'}(\hat{\eta} < C_2) + \int_{C_2}^{\infty} f_{\hat{\eta}}(b) P_{\Delta'}(-C_3 < \hat{\Delta} < C_3)\,dt\end{aligned}$$

Again we need to determine how much the type II error probability is spent at each stage. Consequently we introduce another weighting factor γ_2 such that

$$P_{c'}(\hat{\eta} < C_2) = \gamma_2 \beta \tag{15}$$

and

$$\int_{C_2}^{\infty} f_{\hat{\eta}}(t) P_{\Delta'}(-C_3 < \hat{\Delta} < C_3)\,dt = (1-\gamma_2)\beta \tag{16}$$

where $0 < \gamma_2 < 1$. As seen, the larger the γ_2 is, the smaller the n_2 is. Considering $\Delta = \eta(d_r - d_0)$ under the linear trend, Equation (11) can be re-expressed as

$$\begin{aligned}&\varphi(\eta, \Delta, n_2, n_3, C_2, C_3, \sigma)\\ &= P_\eta(\hat{\eta} < C_2) + \int_{C_2}^{\infty} f_{\hat{\eta}}(b)\\ &\times P_\Lambda\left(\frac{n_2+n_3}{n_3}\left(-C_3 - \frac{n_2}{n_2+n_3} b(d_r - d_0)\right) < \hat{\Delta}_3\right.\\ &\left. < \frac{n_2+n_3}{n_3}\left(C_3 - \frac{n_2}{n_2+n_3} b(d_r - d_0)\right)\right) db\end{aligned}$$

where $\hat{\Delta}_3 = \frac{\sum_{j=n_2+1}^{n_2+n_3} Y_{rj}}{n_3} - \frac{\sum_{j=n_2+1}^{n_2+n_3} Y_{0j}}{n_3}$.

Under the specifications of the values for design parameters $c, c', \Delta', \gamma_1, \gamma_2, \alpha$, and β, the proposed alternative phase II/III design is to determine n_2, n_3, C_2, and C_3 numerically based on constraints of the overall type I and II error rates given in Equations (13)–(16). In conjunction with R function 'Integrate' for numerical integration, with respect to specified values of c, c', σ, α, β, γ_1, and γ_2, R function 'NLM' is used to solve nonlinear functions in conjunction with a Newton-type algorithm to find the values of n_2 and C_2, under constraints Equations (13) and (15). With specification of Δ and Δ', we again use R function 'NLM' to find values of n_3 and C_3 satisfying constraints Equations (14) and (16). An R program is available from the authors upon request.

Under that the nominal dose levels are 0, 10, 20, and 30, Tables I–IV provide the sample sizes per group (n_2 and n_3) and the critical values (C_2 and C_3) for the selection and confirmation stages of the proposed phase II/III adaptive designs for different combinations of design parameters with $\gamma_1 = 0.6$ and 0.8, and $\Delta' = 1$ and 2. In addition, the sample sizes of the phase II trial with and without the Bonferroni adjustment (n' and n'^B) and phase III trial (n'') of the current approach are also presented in Tables I–IV with the relative efficiency (r_s and r_c), defined as the ratio of the total sample of our proposed phase II/III design to that of the current approach.

Tables I–IV illustrate that the sample size at the selection stage increases as γ_1 increases. This fact is because that the larger the γ_1 is, the larger failure rate spent in selection stage and more difficult for a dose will be selected for the confirmation stage. On the other hand, when γ_2 decreases, less type II error rate or more power is spent in the selection stage and large sample sizes are required for the selection stage. The sample sizes for our proposed phase II/III designs given in Tables I–IV confirm this and show that as γ_2 decreases, the required sample size per group for the confirmation stage decreases but the required sample size per group for the selection stage increases. In addition, the critical value at the final analysis also increases as γ_2 decreases. On the other hand, a large value of γ_2 indicates that more power will be spent at the confirmation stage. Therefore, when γ_2 is sufficiently large, the sample size required for the confirmation stage might be greater than that required for the traditional phase III trial.

A simulation study was conducted to compare the proposed alternative phase II/III design with the current approach in terms of success rate which is defined as the overall power for both methods. The overall power for the current approach is the probability of rejection of both null hypotheses of Equations (1)

Table I. Designs with $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 1$, $\gamma_1 = 0.6$, $k = 3$, $(\alpha, \beta) = (0.05, 0.2)$, and $(d_0, d_1, d_2, d_3) = (0, 10, 20, 30)$.

γ_2	n_2	n_3	C_2	C_3	n'	n'^B	n''	r_s*	r_c†
0.1	100	1022	0.0079	0.6369	1237	1764	1570	0.3022	0.2397
0.2	75	1137	0.0092	0.6209	1237	1764	1570	0.3182	0.2525
0.3	60	1239	0.0102	0.6036	1237	1764	1570	0.3361	0.2666
0.4	51	1346	0.0112	0.5851	1237	1764	1570	0.3581	0.2840
0.5	43	1465	0.0121	0.5650	1237	1764	1570	0.3835	0.3042
0.6	37	1606	0.0131	0.5427	1237	1764	1570	0.4154	0.3295
0.7	32	1785	0.0140	0.5172	1237	1764	1570	0.4572	0.3627
0.8	28	2032	0.0151	0.4866	1237	1764	1570	0.5163	0.4096
0.9	24	2448	0.0162	0.4449	1237	1764	1570	0.6172	0.4896

*$r_s = (4n_2 + 2n_3)/(4n' + 2n'')$.
†$r_c = (4n_2 + 2n_3)/(4n'^B + 2n'')$.

Table II. Designs with $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 1$, $\gamma_1 = 0.8$, $k = 3$, $(\alpha, \beta) = (0.05, 0.2)$, and $(d_0, d_1, d_2, d_3) = (0, 10, 20, 30)$.

γ_2	n_2	n_3	C_2	C_3	n'	n'^B	n''	r_s*	r_c†
0.3	103	845	0.0312	0.5524	1237	1764	1570	0.2599	0.2062
0.4	90	948	0.0335	0.5331	1237	1764	1570	0.2789	0.2213
0.5	80	1060	0.0355	0.5122	1237	1764	1570	0.3017	0.2393
0.6	71	1191	0.0375	0.4894	1237	1764	1570	0.3296	0.2615
0.7	64	1356	0.0395	0.4637	1237	1764	1570	0.3670	0.2911
0.8	58	1582	0.0415	0.4331	1237	1764	1570	0.4199	0.3331
0.9	53	1964	0.0436	0.3921	1237	1764	1570	0.5119	0.4060

*$r_s = (4n_2 + 2n_3)/(4n' + 2n'')$.
†$r_c = (4n_2 + 2n_3)/(4n'^B + 2n'')$.

Table III. Designs with $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 2$, $\gamma_1 = 0.6$, $k = 3$, $(\alpha, \beta) = (0.05, 0.2)$, and $(d_0, d_1, d_2, d_3) = (0, 10, 20, 30)$.

γ_2	n_2	n_3	C_2	C_3	n'	n'^B	n''	r_s*	r_c†
0.2	100	124	0.0079	1.2974	310	441	393	0.3198	0.2541
0.3	75	188	0.0092	1.2646	310	441	393	0.3337	0.2651
0.4	60	231	0.0102	1.2296	310	441	393	0.3465	0.2753
0.5	51	269	0.0112	1.1915	310	441	393	0.3662	0.2910
0.6	43	307	0.0121	1.1498	310	441	393	0.3880	0.3082
0.7	37	349	0.0131	1.1033	310	441	393	0.4176	0.3318
0.8	32	399	0.0140	1.0502	310	441	393	0.4571	0.3631
0.9	28	465	0.0151	0.9864	310	441	393	0.5143	0.4086

*$r_s = (4n_2 + 2n_3)/(4n' + 2n'')$.
†$r_c = (4n_2 + 2n_3)/(4n'^B + 2n'')$.

Table IV. Designs with $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 2$, $\gamma_1 = 0.8$, $k = 3$, $(\alpha, \beta) = (0.05, 0.2)$, and $(d_0, d_1, d_2, d_3) = (0, 10, 20, 30)$.

γ_2	n_2	n_3	C_2	C_3	n'	n'^B	n''	r_s*	r_c†
0.6	90	130	0.0335	1.1547	310	441	393	0.3060	0.2431
0.7	71	207	0.0375	1.0398	310	441	393	0.3445	0.2737
0.8	64	254	0.0395	0.9772	310	441	393	0.3771	0.2996
0.9	58	316	0.0415	0.9055	310	441	393	0.4265	0.3388

*$r_s = (4n_2 + 2n_3)/(4n' + 2n'')$.
†$r_c = (4n_2 + 2n_3)/(4n'^B + 2n'')$.

and (5). For the proposed phase II/III design, alternatively the overall power is the probability of rejecting both null hypothesis of Equations (8) and (10). Suppose the test drug has dose levels of 10, 20, and 30, respectively. Also assume that the placebo group has dose level of 0. Figure 1 displays simulation results for the case of $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 1$, and

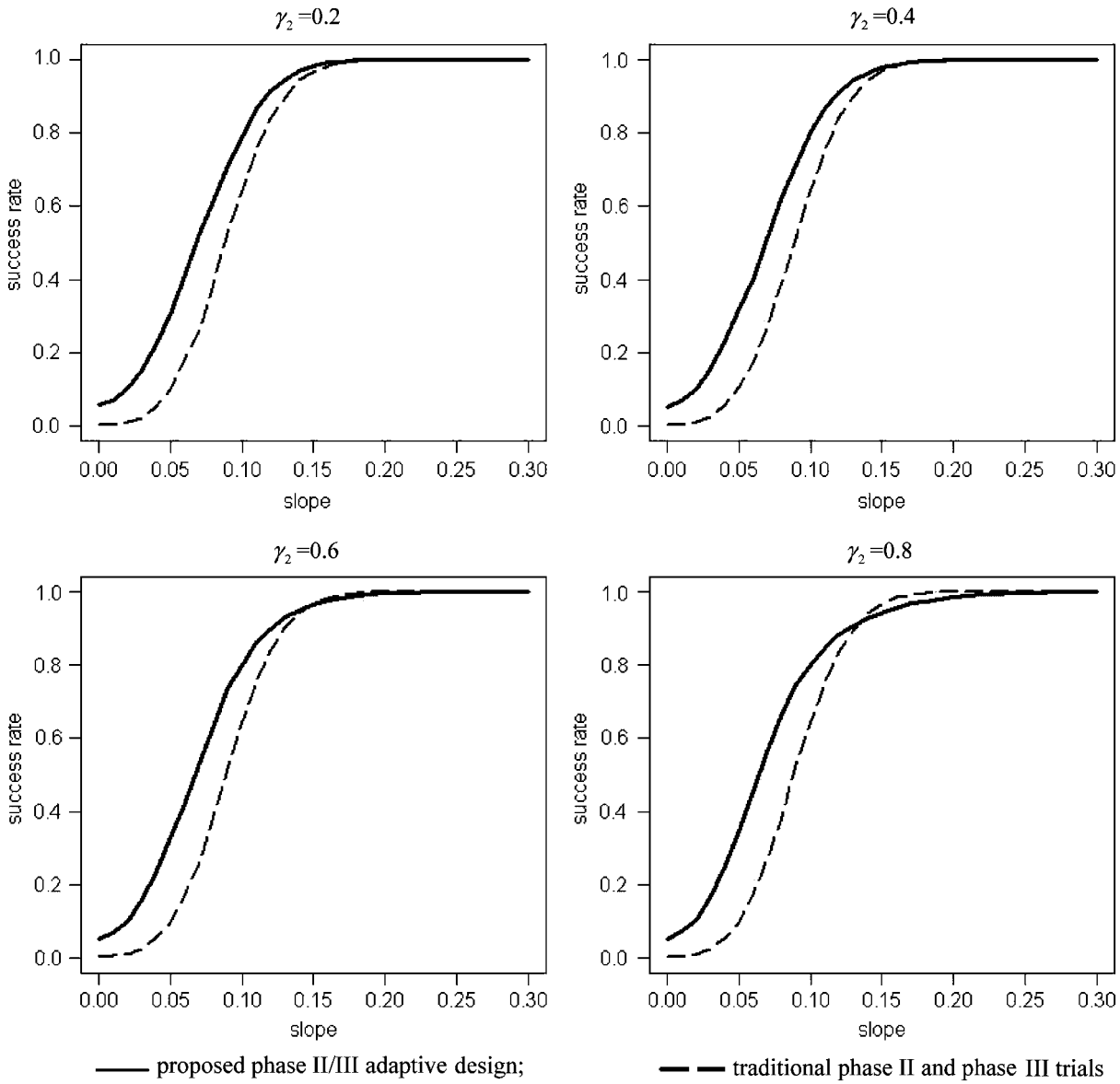

Figure 1. Simulated success rates for the case of $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 1$, $\gamma_1 = 0.6$, $k = 3$, and $(d_0, d_1, d_2, d_3) = (0, 10, 20, 30)$.

$(\alpha, \beta) = (0.05, 0.2)$, $\gamma_1 = 0.6$ with various values of γ_2. For instance, given $\gamma_2 = 0.2$, we can derive that $n_2 = 75$, $C_2 = 0.0092$, $n_3 = 1137$, $C_3 = 0.6209$, $n' = 1237$, and $n'' = 1570$. The power is evaluated for η being from 0 to 0.30 by 0.02. Under the assumption of a linear trend such that $\Delta = 10\eta$, the success rate at each η was obtained as the proportion of successes from 10,000 replicates. From Figure 1, the proposed phase II/III design is uniformly more powerful than the current method for the range of η between 0 and 0.3 under the above specifications of parameters for the design. Figures 2–4 compare the success rates of the proposed phase II/III design with the current approach under other specifications of the parameters of the design. Similar conclusion on the success rate is reached from Figures 2–4.

4. EXAMPLES

Suppose that a new test drug is being developed for the indication of the patients with scleroderma lung disease [13]. Three doses of 10, 20, and 30 mg and placebo are selected for evaluation in phase II and III development. One of the primary efficacy endpoints is the change from baseline of the forced vital capacity (FVC, % of predicted) at month 12. We apply the alternative proposed phase II/III design to the development of this new test drug with the assumed standard deviation of 10. The pre-specified requirements for the slope for the selection stage and for the treatment effect for the confirmation stage are 0.1 and 1.0, respectively. In other words, $c' = 0.1$ and $\Delta = 1$. For the type I error rate of 0.05 and type II error rate of 0.2, Table I indicates that if $\gamma_1 = 0.6$ and $\gamma_2 = 0.5$, 43 patients per group for a total of 172 patients are required for the selection stage. At the completion of the selection stage, if the observed value of slope $\hat{\eta}$ does not exceed 0.0121, the development is terminated and the new test drug is concluded as lack of efficacy for further development. On the other hand, if the observed value of the estimator of slope $\hat{\eta}$ is greater than 0.0121, the trial continues to the confirmation stage with selection of the lowest dose level of 10 (i.e. $\Delta' = 1$). An additional 1465 patients per group are

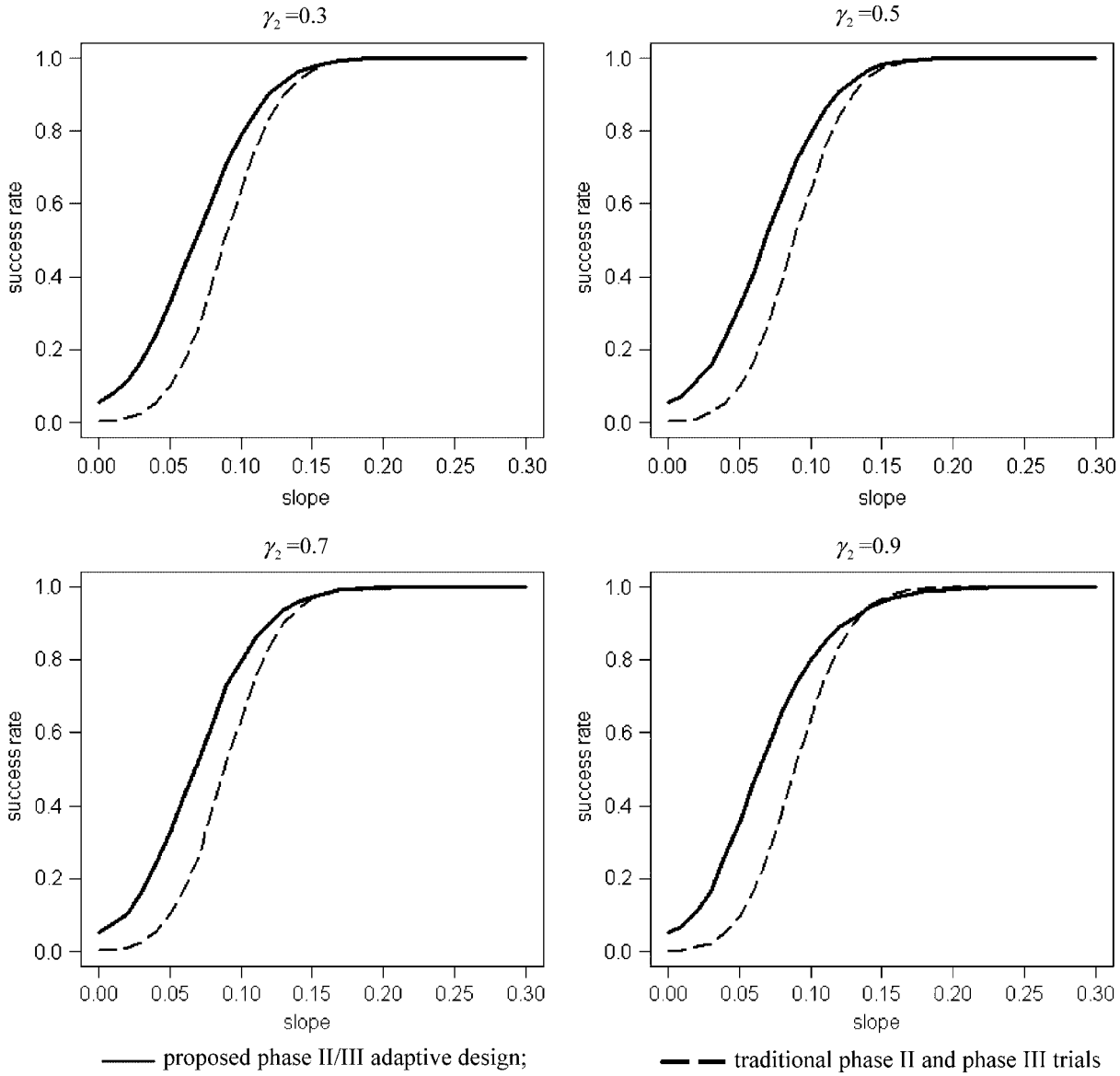

Figure 2. Simulated success rates for the case of $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 1$, $\gamma_1 = 0.8$, $k = 3$, and $(d_0, d_1, d_2, d_3) = (0, 10, 20, 30)$.

required for the selected dose and placebo. At the conclusion of the confirmation stage, if the overall observed absolute value of mean difference, $\hat{\Delta}$, based on the cumulative data of a total 3016 patients ($2 \times 43 + 2 \times 1465$) obtained at the end of the confirmation stage of the trial does not exceed 0.5650, the new test drug is concluded no efficacy. On the other hand, if the observed absolute value of $\hat{\Delta}$ is greater than 0.5650, efficacy of the new test drug is superior to the placebo group with respect to FVC (% of predicted) at the 0.05 significance level. In addition, the number of required sample sizes for the current phase II trial and phase III trials are 1237 (and 1764 for Bonferroni *p*-value adjustment) and 1570 per group, respectively. It leads to reduction of 61.65% and 69.58% of the total sample size for the proposed phase II/III design as compared with the current approach without and with the Bonferroni adjustment, respectively.

Note that in this example, it may not make intuitive sense to differentiate between three doses with 43 patients per arm, and then require 1465 patients per arm to confirm an effect against placebo in confirmation stage. As mentioned, larger γ_1 indicates that we spend fewer type I error rate for phase II stage. If we choose $\gamma_1 = 0.8$, then the number of required sample sizes for the phase II trial and phase III trials are 80 and 1060 per group, respectively (cf. Table II). Therefore, the determination how we want to spend the type I error rate at each stage is very important.

5. DISCUSSION

In this paper, we propose an alternative phase II/III design for evaluation of drugs efficacy based on continuous endpoints. Under this design structure, a single trial with the selection and confirmation phases is conducted using the same protocol with the same inclusion/exclusion criteria, the same concurrent control, the same methods for evaluation, and the same efficacy/safety endpoints. In other words, the data from both the dose selection and confirmation of efficacy are generated within the same study. Another attractive feature is that our phase II/III design would in fact use the data from patients

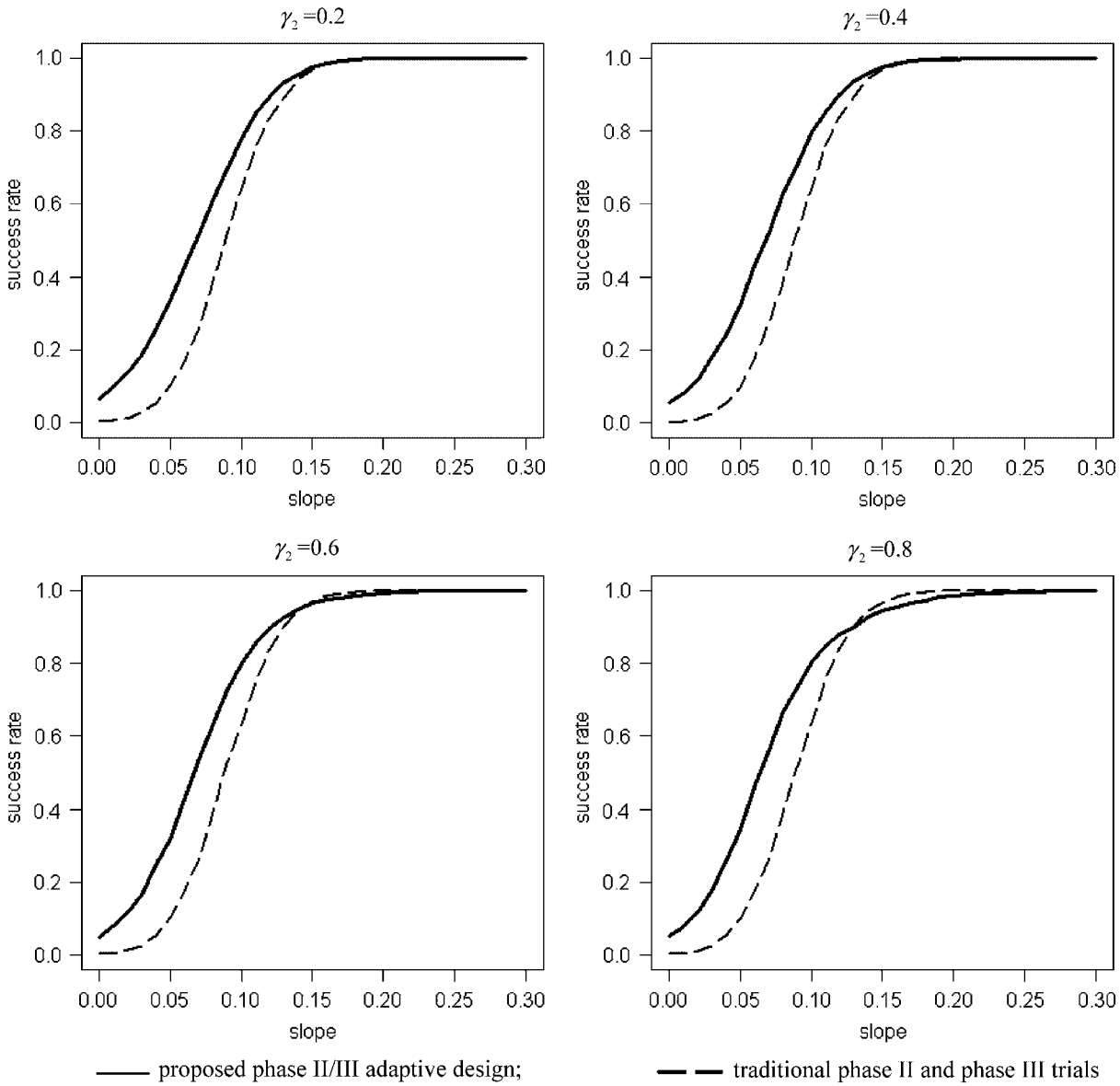

Figure 3. Simulated success rates for the case of $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 2$, $\gamma_1 = 0.6$, $k = 3$, and $(d_0, d_1, d_2, d_3) = (0, 10, 20, 30)$.

enrolled from the selection stage and from the confirmation stage in the final analysis. With this approach, reduction of the total sample size might be possible. This in term may shorten the total duration of drug development and hence can save considerably valuable resource and cost.

After a linear dose response is established in the selection stage, selection of dose level for the confirmation stage is also critical. First of all, we need to choose the dose level with the pre-specified requirement for efficacy. However, the choice of dose level should be determined not only on the efficacy but also on safety. In general, the toxicity might also increase as the dose level increases. In this case, the lowest dose level which meets the efficacy requirement with the best safety profile is selected for the confirmation stage. On the other hand, even if the linear trend of the dose–response for the selection stage is statistically significant, the dose–response might increase first and then reach the plateau at the higher dose levels. In this case, we may also select the lowest dose level reaching the plateau and meeting the pre-specified requirement. As a result, our proposed alternative design selects the lowest dose level whose slope meets the minimal clinically meaningful requirement, δ.

For the current approach, the phase II and phase III trials are conducted sequentially but independently and the individual type I error rates at phase II and phase III are separately controlled both 0.05. It follows that the actual overall type I error rate is in fact equal to $0.05 \times 0.05 = 0.0025$, which is unnecessarily stringent. However, in our proposed phase II/III design, the actual type I error rate is only equal to 0.05. In other words, the type I error rate of our proposed phase II/III design is 20 times larger than the traditional approaches. In other words, the traditional approach is more conservative than our proposed phase II/III design. Similarly, in traditional approaches, if the powers for both phase II and phase III trials are 0.8, then the overall power is 0.64. On the other hand, in our proposed phase II/III design, the actual power is equal to 0.8 which is 1.25 times larger than the traditional approaches. That is, our proposed alternative phase II/III design can gain more power than the traditional approach. This phenomenon is also demonstrated in Figures 1–4.

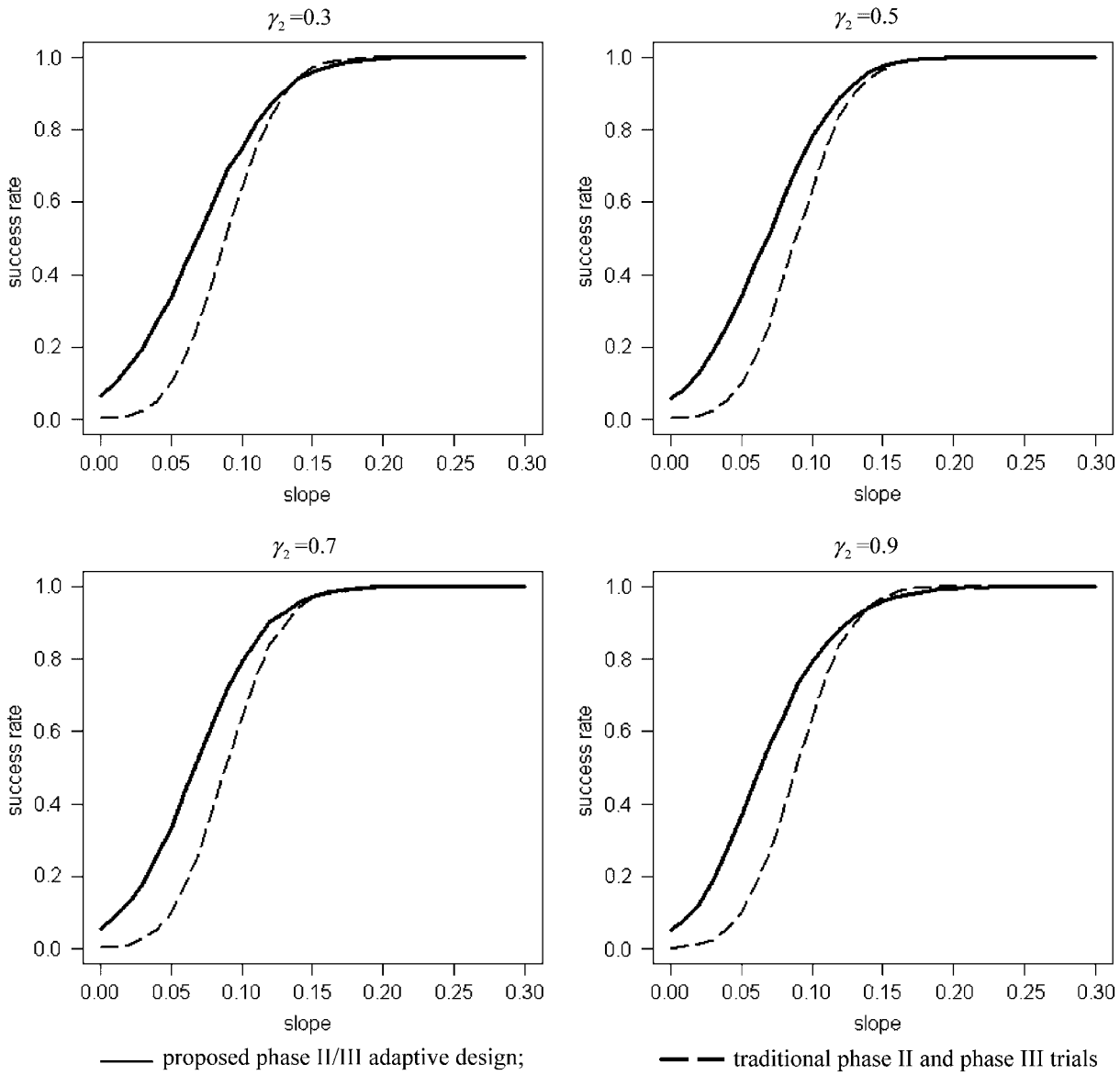

Figure 4. Simulated success rates for the case of $\sigma = 10$, $(c, c') = (0, 0.1)$, $\Delta' = 1$, $\gamma_1 = 0.8$, $k = 3$, and $(d_0, d_1, d_2, d_3) = (0, 10, 20, 30)$.

It should be noted that in our proposed phase II/III design, the probability of success in phase III trial does not depend so much on statistical significance of the dose–response test from phase II stage. However, the expected treatment effect for the confirmation stage (phase III) can be estimated from the data obtained in the selection stage (phase II). Consequently, the proposed alternative phase II/III design can be extended as follows. First, given γ_1 and γ_2, we can first determine sample size n_2 and the critical C_2 for the selection stage based on the pre-specified values of undesirable and desirable slopes for the dose response, c and c′ and σ. If the conclusion of the selection stage is successful and one dose is chosen for the confirmation stage, the estimates of Δ' and σ can be obtained from the selection stage. With the updated estimates of Δ' and σ, we can therefore calculate the required total sample size and the critical value for the confirmation stage. This may increase the accuracy of the estimate of the required sample size for the confirmation stage and consequently improve the overall success rate of development. Doing so may also increase the probability of success in phase III trial.

In the seamless phase II/III design, controlling the experiment-wise error is an issue. Chang [10,14] proposed a method where the test statistic at each stage is a linear combination of the *p*-values calculated using subsamples from the phase II and phase III stages. This method offers great flexibility in the selection of stopping boundaries and no numerical integration is needed for the two-stage designs. On the other hand, in our design, the overall type I error and power spent at each stage are controlled by the weighting factors γ_1 and γ_2 respectively. At the design stage, once γ_1 and γ_2 are specified, the corresponding stopping boundary and sample size required for each stage can be derived. Although our design cannot avoid the calculation of numerical integration, it can allow flexibility on how we want to spend the type I error rate and power at each stage. A proper balance for selection of weighting factors γ_1 and γ_2 between the selection and confirmation stages is important for a successful implementation of the proposed alternative phase II/III design. The value of γ_1 should be stringent (large) enough that ineffective doses should be quickly eliminated during the

selection stage and the probability of confirming the efficacy of the selected doses can exceed the pre-specified level. On the other hand, γ_2 should be chosen to provide sufficient sample sizes for both selection and confirmation stages. For small c, we recommend that the value of γ_1 be between 0.6 and 0.8 and the value of γ_2 be 0.3 and 0.7. In addition, we also suggest that the number of patients in the confirmation stage be larger than that of the selection stage. Within these ranges for γ_1 and γ_2, as shown in Tables I–IV, the required sample size of our proposed alternative phase II/III design can reduce to 30% and 45% of the sample size required by the current approach.

Note that in our method, we assume that the dose–response relationship can be described by the simple linear regression. This assumption may not be true. Alternatively, some other dose–response relationships can be examined at the phase II stage, and the dose–response patterns can be expressed by various linear contrasts of μ_i, $i=0, 1,\ldots, k$, say $\sum_{i=1}^{k} c_i\mu_i$, where $c_1+\cdots+c_k=0$. In this case, the objective of the phase II trial will in stead test the following hypothesis:

$$H_0 : \sum_{i=1}^{k} c_i\mu_i = 0 \text{ vs } H_A : \sum_{i=1}^{k} c_i\mu_i \neq 0$$

Another point we wish to make is that while assuming linear dose–response, the test of slope may easily be significant in situation when there is very weak dose–response. Therefore, we need a threshold for slope. In other words, the determination of the threshold of the slope c in Equation (8) is rather critical. It should adequately reflect the minimal clinically meaningful dose–response.

One intriguing feature about the use of an adaptive phase II/III design is probably the possibility of shortening the time of development of a new medication. As indicated earlier, such a design is not only flexible but also efficient as compared with separate phase II and phase III studies. However, in practice, not all clinical development may be suitable for such a design. Maca *et al*. [15] propose a list of criteria for determining the feasibility of the use of an adaptive design in clinical development plan. As the use of an adaptive phase II/III design is to get effective medications to patients sooner, whether such a design would achieve a reduction in development time would be an important factor for feasibility consideration. When the adaptive phase II/III trial is the only pivotal trial required for regulatory submission, the reduction in clinical development time is clear. On the other hand, if the phase II/III trial is one of two required pivotal trials, then the second pivotal trial should be completed within a reduced time frame that shortens the overall development time. Maca *et al*. [15] suggest the second pivotal trial, which is more traditionally designed, could begin immediately after the interim analysis so that it is possibly completed close to the time the adaptive phase II/III study is completed. Doing so may need more time for planning, development, and health authority review for such a design. Consequently, this extra time must also be incorporated into the evaluation of the overall development time.

In our adaptive phase II/III design, prior to the interim analysis at which the dose to be continued will be chosen, there will be a period during which some patients have been randomized but have not yet been followed long enough to reach the endpoint for evaluation. During this 'transition' period, Maca *et al*. [15] suggest that when the time needed to reach this endpoint is short relative to the total enrollment time of the study, enrollment can still continue uninterrupted with relatively few patients enrolled. Even though those patients enrolled during this period and randomized to doses that will not be continued will not be providing direct evidence for the comparison of the selected dose vs the control at the confirmation stage, they can be used to understand better the dose response and safety profile. On the other hand, when the endpoint duration is too long, many patients will have been randomized during this period, which could cause undesirable inefficiencies. In this case, enrollment may need to be temporarily hauled during the transition period, but doing so can result in disruption to the trial and erode the benefits in time saving for the adaptive phase II/III design. Thus, Maca *et al*. [11] suggested that well-established and understood endpoints (or surrogate markers) be considered when implementing an adaptive phase II/III design in clinical development.

Acknowledgements

Thanks are due to two referees for their detailed, constructive and thoughtful comments and suggestions which we believe have led to a significant improvement to this paper. The views expressed in this article are personal opinions of the authors and may not necessarily represent the position of the National Health Research Institutes and National Taiwan University, Taiwan.

REFERENCES

[1] The Economist. *Mercky Prospects; Pharmaceuticals*, Vol. 364, The Economist, London, U.K., 2002; p. 60.
[2] Scher HI, Heller G. Picking the winners in a sea of plenty. *Clinical Cancer Research* 2002; **8**:400–404.
[3] Schaid DJ, Wieand S, Therneau TM. Optimal two-stage screening designs for survival comparisons. *Biometrika* 1990; **77**:507–513.
[4] Follman DA, Proschan MA, Geller NL. Monitoring pairwise comparisons in multi-armed clinical trials. *Biometrics* 1994; **50**:325–336.
[5] Todd S, Stallard N. A new clinical trial design combining phases II and III: sequential designs with treatment selection and a change of endpoint. *Drug Information Journal* 2005; **39**:109–118.
[6] Tsou HH, Hsiao CF, Chow SC, Liu JP. A two-stage design for drug screening trials based on continuous endpoints. *Drug Information Journal* 2008; **42**:253–262.
[7] Bischoff W, Miller F. Adaptive two-stages test procedures to find the best treatment in clinical trials. *Biometrika* 2005; **92**:197–212.
[8] Chow SC, Chang M. Adaptive design methods in clinical trials – a review. *Orphanet Journal of Rare Diseases* 2008; **3**:11, published online 2 May 2008.
[9] Sill MW, Sampson AR. Drop-the-losers design: binomial. *Computational Statistics and Data Analysis* 2009; **53**: 586–595.
[10] Chang M. *Adaptive Design Theory and Implementation using SAS and R*. CRC Press: Boca Raton, 2008.
[11] Chow SC, Chang M . *Adaptive Design Methods in Clinical Trials*. Chapman & Hall: New York, 2006.
[12] Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials* 1989; **10**:1–10.
[13] Tashkin DP, Elashoff R, Clements PJ *et al*. Cyclophosphamide versus placebo in scleroderma lung disease. *New England Journal of Medicine* 2006; **354**:2655–2666.
[14] Chang M. Adaptive design method based on sum of *p*-values. *Statistics in Medicine* 2007; **26**:2772–2784.
[15] Maca J, Bhattacharya S, Dragalin V, Gallo P, Krams M. Adaptive seamless phase II/III designs – background, operational aspects, and examples. *Drug Information Journal* 2006; **40**:463–474.