Two-stage design for phase II oncology trials with relaxed futility stopping

Anastasia Ivanova* and Allison M. Deal

Many oncology phase II trials are single arm studies designed to screen novel treatments based on efficacy outcome. Efficacy is often assessed as an ordinal variable based on a level of response of solid tumors with four categories: complete response, partial response, stable disease and progression. We describe a two-stage design for a single-arm phase II trial where the primary objective is to test the rate of tumor response defined as complete plus partial response, and the secondary objective is to estimate the rate of disease control defined as tumor response plus stable disease. Since the goal is to estimate the disease control rate, the trial is not stopped for futility after the first stage if the disease control rate is promising. The new design can be generated using easy-to-use software that is available at http://cancer.unc.edu/biostatistics/program/ivanova/.

Keywords and phrases: Simon's design, Two-stage design, Ordinal outcome, Disease control, Tumor response.

1. INTRODUCTION

Phase two trials in oncology are usually single arm studies to screen oncology treatments. According to the International Harmonization Project for Response Criteria [1], responses to treatment fall into one of four categories: complete response (CR), partial response (PR), stable disease (SD) and progressive disease. In most trials, these categories are combined into a dichotomous primary outcome, tumor response (TR), which is defined as complete or partial response. Disease control (DC), defined as complete or partial response or stable disease, is often of interest as well. Alternatively, clinical benefit, defined as disease control observed for at least a certain period of time, can be used. Often, one outcome (such as tumor response) is assessed as the primary endpoint, and another (such as disease control) is assessed as a secondary endpoint. These types of trial designs have been used in various disease groups. Abrey et al. [2] described a phase II trial in patients with recurrent or progressive brain metastases with tumor response as a primary outcome and disease control as a secondary outcome. Similarly, in a large phase II trial in non-small cell lung can-

cer [3], both response rate and disease control were reported as important endpoints of the study. Studies of advanced breast cancer often include clinical benefit as a secondary endpoint [4].

Our motivating example is a University of North Carolina Lineberger Comprehensive Cancer Center single arm phase II trial to investigate the efficacy of the daily use of a novel mTOR inhibitor in combination with other agents in the setting of HER2-positive breast cancer brain metastases. The primary endpoint was tumor response defined as complete or partial response, but also of particular interest was the clinical benefit endpoint. Clinical benefit was defined as complete or partial response or stable disease lasting for at least 3 months. The tumor response rate of standard of care regimen without the addition of the novel mTOR inhibitor was thought to be approximately 0.05, and the TR rate for the alternative hypothesis was set to 0.20. Initially the investigators were going to test TR rate by using Simon's optimal two-stage design [5]. The total of $n = 29$ provide a power of 80% with type I error rate of at most $\alpha = 0.05$. The trial is stopped for futility if no TRs are observed in the first 10 patients. The null hypothesis is to be rejected if 4 or more TRs are observed among 29 patients. This design yields an actual type I error rate of 0.0468. Since the clinical benefit rate was also of interest, it was decided to stop the trial for futility after the first stage only if none of the patients in stage 1 had clinical benefit; that is, to apply the stopping rule for futility to the outcome of clinical benefit, defined as TR or lasting SD, rather than TR alone. If the probability of lasting SD is high, the likelihood of stopping for futility is low and the type I error rate of the trial will be, in fact, higher than 0.0468 and possibly higher than the nominal level of 0.05. If the probability of stopping the trial for futility is close to 0, the type I error rate is close to 0.0548, the type I error rate of a single-stage design. Therefore, one cannot simply replace TR with clinical benefit in the futility stopping rule in a two-stage design. Instead, our new two-stage design that yields the desired type I and II errors under certain assumptions on lasting SD rate should be used.

In Section 2 of this paper we describe how to construct two-stage designs with relaxed stopping for futility that yield the specified type I error rate. In Section 3 we describe how to estimate rates of interest after the trial. We give examples in Section 4.

*Corresponding author.

2. TWO-STAGE DESIGN WITH RELAXED STOPPING FOR FUTILITY

Let p_T, p_S and p_D denote the true unknown tumor response, stable disease and disease control rates respectively, $p_D = p_T + p_S$. Without loss of generality we will use these three outcomes to describe our method with DC defined as TR or SD. Other outcomes can be used as well, for example, clinical benefit defined as TR or lasting SD, or tumor response defined as complete response or partial response. Let p_{0T} be an uninteresting TR rate, and p_{AT} be a rate which would be considered promising for future study. Consider the one-sided hypothesis about TR rate

$$H_0 : p_T \leq p_{0T} \quad \text{versus} \quad H_1 : p_T > p_{0T}.$$

Consider a two-stage design, where n is the total number of patients in the trial, n_1 is the number of patients in stage 1, $n_1 < n$. Let $X_T^{(1)}$ and $X_S^{(1)}$ be random variables describing the number of patients with tumor response and stable disease in stage 1, out of n_1 patients, and $X_T^{(2)}$ and $X_S^{(2)}$ the number of patients with TR and SD at the end of the trial out of n patients. $X_T^{(2)} \geq X_T^{(1)}$ and $X_S^{(2)} \geq X_S^{(1)}$. Let $x_T^{(1)}$, $x_S^{(1)}$, $x_T^{(2)}$ and $x_S^{(2)}$ be the corresponding observed quantities. Simon's, or other two-stage, design testing H_0 will stop for futility if the observed number of tumor responses in stage 1, is less than or equal to the r_1, ($x_T^{(1)} \leq r_1$), otherwise $n - n_1$ patients are enrolled. H_0 is rejected if $x_T^{(2)} > r_2$. For a given type I error rate α and type II error rate β, parameters r_1 and r_2 in a two-stage design are such that

$$\Pr\{X_T^{(1)} > r_1 \text{ and } X_T^{(2)} > r_2 | p_{AT}\} \geq 1 - \beta,$$
$$\Pr\{X_T^{(1)} > r_1 \text{ and } X_T^{(2)} > r_2 | p_{0T}\} \leq \alpha.$$

In our situation, the primary endpoint is TR, but the investigators are also interested in estimating the DC rate, if the DC rate is promising. Therefore, if a low TR rate is observed in stage 1, but the DC rate is promising (because many patients had stable disease), we do not want to stop for futility at the interim. Consider the following two-stage design: stop after stage 1 and declare treatment uninteresting only if the number of TRs plus the number of patients with SD is low with $x_T^{(1)} + x_S^{(1)} \leq r_1$. We will also stop after stage 1 if the number of patients with TR is so low that it is impossible to reject H_0 at the end of the trial even all $n - n_1$ patients in stage 2 have TR, that is, we stop if $x_T^{(1)} \leq r_2 - (n - n_1) - 1$. If $p_S = 0$, the two designs are the same since $\Pr\{X_S^{(1)} > 0\} = 0$. However, when $p_S > 0$, changing the futility stopping rule from $x_T^{(1)} \leq r_1$ to $x_T^{(1)} + x_S^{(1)} \leq r_1$ will affect type I and type II error rates in a two-stage design. The probability of stopping for futility will decrease, and, therefore, both type I error rate and power (1 minus type II error rate) will increase.

To compute type I and type II error rates in the modified design, we need to make assumptions on the rate of stable disease. Note that both the type I and II error should be computed under the same assumptions on p_S. If assumptions on p_S are not the same under the null and alternative hypotheses, the type I and II error rates are computed for two different designs. For example, if we assume that, under H_0, the SD rate is equal to the smallest possible value, $p_S = 0$, the type I error rate is computed for the two-stage design that stops for futility when $X_T^{(1)} \leq r_1$. If we assume that, under H_1, the SD rate is equal to the largest possible value, $p_S = 1 - p_{AT}$, the type II error rate is computed for a single stage design since the probability of stopping for futility is 0. Since we need to have the same design under both H_0 and H_1, the assumptions on p_S should be the same under both hypotheses.

We assume that $p_S \in [p_S^L, p_S^U]$, $p_S^U \leq 1 - p_{AT}$. We are interested in finding two-stage designs such that

(1)
$$\Pr\{X_T^{(1)} + X_S^{(1)} > r_1 \text{ and } X_T^{(1)} > r_2 - (n - n_1) - 1$$
$$\text{and } X_T^{(2)} > r_2 | p_{AT}, p_S\} \geq 1 - \beta,$$
$$\Pr\{X_T^{(1)} + X_S^{(1)} > r_1 \text{ and } X_T^{(1)} > r_2 - (n - n_1) - 1$$
$$\text{and } X_T^{(2)} > r_2 | p_{0T}, p_S\} \leq \alpha.$$

When the treatment response rate is fixed at p_T, $\Pr\{X_T^{(1)} + X_S^{(1)} > r_1 | p_T, p_S\}$ is increasing in p_S. Therefore, the power is minimized when $p_S = p_S^L$ and the type I error rate is maximized when $p_S = p_S^U$. Hence conditions (1) can be written as

(2)
$$\min_{p_S} \Pr\{X_T^{(1)} + X_S^{(1)} > r_1 \text{and } X_T^{(1)} > r_2 - (n - n_1) - 1$$
$$\text{and } X_T^{(2)} > r_2 | p_{AT}, p_S\}$$
$$= \Pr\{X_T^{(1)} + X_S^{(1)} > r_1 \text{ and } X_T^{(1)} > r_2 - (n - n_1) - 1$$
$$\text{and } X_T^{(2)} > r_2 | p_{AT}, p_S^L\} \geq 1 - \beta,$$
$$\max_{p_S} \Pr\{X_T^{(1)} + X_S^{(1)} > r_1 \text{ and } X_T^{(1)} > r_2 - (n - n_1) - 1$$
$$\text{and } X_T^{(2)} > r_2 | p_{0T}, p_S\}$$
$$= \Pr\{X_T^{(1)} + X_S^{(1)} > r_1 \text{ and } X_T^{(1)} > r_2 - (n - n_1) - 1$$
$$\text{and } X_T^{(2)} > r_2 | p_{0T}, p_S^U\} \leq \alpha.$$

That is, to control type I error rate for all possible values of p_S we need to control it for p_S^U and to ensure required power for all values of p_S we need to ensure it for p_S^L.

As there are many designs satisfying criteria (2), Simon proposed to consider the minimax and the optimal designs. The minimax design minimizes the maximum total sample size n, while the optimal design minimizes the expected sample size under the null hypothesis EN_0. When there is more than one minimax or optimal design we choose the design with the smallest value of $EN_0 + n$. Such defined minimax and optimal designs are unique. To compute the expected sample size under H_0 for our design, we need to know the distribution of p_S. Assume that random variable P_S is distributed with density function $f(x)$ defined for $x \in [p_S^L, p_S^U]$, $p_{0T} + p_S^U \leq 1$, then probability of stable disease is a ran-

Table 1. Two-stage designs with relaxed stopping for futility assuming that SD rate $p_S \sim Unif(0, p_S^U)$. For each value of p_S^U we report design parameters, the expected sample size under H_0 (EN_0), the probability of stopping the trial early after stage 1 (PES) and the weight w in $wn + (1-w)EN_0$. When $p_S^U = 0$ the designs are based on TR only. The power is 0.8 and the type I error rate is one-sided 0.05

p_S^U	N	n_1	r_1	r_2	EN_0	PES	w	Design
$p_{0T} = 0.05$, $p_{AT} = 0.2$								
0	27	13	0	3	19.8	0.51	[0.598,1]	Minimax
0	28	11	0	3	18.3	0.57	[0.414,0.597]	Admissible
0	29	10	0	3	17.6	0.60	[0,0.413]	Optimal
0.1	27	13	0	3	23.1	0.28	[0.443,1]	Minimax
0.1	28	11	0	3	22.3	0.34	[0,0.442]	Optimal
0.2	27	13	0	3	24.6	0.17	[0.208,1]	Minimax
0.2	28	11	0	3	24.3	0.22	[0,0.209]	Optimal
$p_{0T} = 0.5$, $p_{AT} = 0.7$								
0	37	23	12	23	27.7	0.66	[0.556,1]	Minimax
0	39	16	8	24	25.2	0.60	[0.304,0.555]	Admissible
0	43	15	8	26	23.5	0.70	[0,0.303]	Optimal
0.1	37	11	4	23	32.3	0.18	[0.267,1]	Minimax
0.1	46	15	8	28	29.1	0.55	[0,0.266]	Optimal
0.2	37	29	15*	23	32.9	0.51	[0,1]	Minimax, Optimal
$p_{0T} = 0.4$, $p_{AT} = 0.6$								
0	39	34	17	20	34.4	0.91	[0.815,1]	Minimax
0	41	17	7	21	25.6	0.64	[0.182,0.814]	Admissible
0	46	16	7	23	24.5	0.72	[0,0.181]	Optimal
0.1	42	20	0	22	36.3	0.26	[0.857,1]	Minimax
0.1	43	15	6	22	30.3	0.45	[0,0.856]	Optimal
0.2	42	33	15*	22	37.7	0.48	[0.455,1]	Minimax
0.2	45	13	5	23	35.2	0.31	[0,0.454]	Optimal
0.3	42	35	16	22	37.8	0.60	[0,1]	Minimax, Optimal

*Additional stopping for TR: the trial is stopped for futility if the number of TRs in stage 1 is such that it is impossible to reject H_0 after stage 2, that is, the number of TRs is less than $r_2 - (n - n_1)$.

dom variable defined as $P_D = p_{0T} + P_S$. The probability of early stopping under H_0, PES_0, is computed as $PES_0 = E_{P_D}[\Pr(Y \leq r_1 | P_D)]$, where Y is a number of patients with disease control out of n_1 patients. The expected sample size under the null hypothesis is $EN_0 = PES_0 n_1 + (1 - PES_0)n$.

Apart from the minimax and optimal designs, it is helpful to calculate other admissible designs, designs that minimize $wEN_0 + (1-w)n$ for some w, such that $0 \leq w \leq 1$ [6] since admissible designs often provide different sample size breakdowns over the two stages. These other admissible designs are often preferred by investigators because they allow a design with a more balanced first and second stage sample size to be chosen. We developed easy-to-use software that generates all admissible designs with relaxed stopping for futility; it can be found at http://cancer.unc.edu/biostatistics/program/ivanova/. We also give an example write-up that can be used in a clinical trial protocol.

Table 1 presents several examples of input parameters and all admissible designs for these parameters. The designs are generated by our online software. The software assumes that P_S has a uniform distribution. Since the power of the test is minimized when $p_S = p_S^L$ and the type I error rate is maximized when $p_S = p_S^U$, power and type I error rate depend only on the range of P_S, $[p_S^L, p_S^U]$, and do not depend on the distribution of P_S. Thus, all designs presented in Table 1 guarantee (1). The only quantity that depends on the distribution of P_S is EN_0. When there is more than one design that minimizes $wEN_0 + (1-w)n$ for a given w, we choose the design with the smallest value of $EN_0 + n$, hence our designs are unique.

3. ESTIMATION OF TUMOR RESPONSE AND STABLE DISEASE RATES AFTER THE TRIAL

The goals of a trial with relaxed stopping for futility are to test the null hypothesis based on the primary endpoint and provide point and interval estimates for the primary and secondary endpoints at the end of the trial. The testing and estimation procedures should take into account the interim analysis after stage 1, otherwise the estimates can be biased. Jovic and Whitehead [7] adopted the methodology of Fiarbanks [8] to estimate the rate of the primay outcome after Fleming's two-stage design [9]. In this section we show how to obtain p-values and estimates with confidence in-

tervals for primary and secondary endpoints in a two-stage trial with relaxed futility stopping, taking design into account. For that, we enumerate all of the outcomes that are as extreme or more extreme than the one we observed. Let $r_1' = r_2 - (n - n_1) - 1$. Define two functions

$$P(p_T) = \begin{cases} \Pr\{X_T^{(1)} > r_1 - x_S^{(1)} \text{ and } X_T^{(2)} \geq x_T^{(2)}|p_T\} \\ \quad \text{if } x_T^{(1)} + x_S^{(1)} > r_1 \text{ and } x_T^{(1)} > r_1', \\ \Pr\{x_T^{(1)} \leq X_T^{(1)} \leq n_1|p_T\} \\ \quad \text{if } x_T^{(1)} + x_S^{(1)} \leq r_1 \text{ or } x_T^{(1)} \leq r_1', \end{cases}$$

and

$$Q(p_T) = \begin{cases} \Pr\{X_T^{(1)} > r_1 - x_S^{(1)} \text{ and } X_T^{(2)} > x_T^{(2)}|p_T\} \\ \quad \text{if } x_T^{(1)} + x_S^{(1)} > r_1 \text{ and } x_T^{(1)} > r_1', \\ \Pr\{x_T^{(1)} < X_T^{(1)} \leq n_1|p_T\} \\ \quad \text{if } x_T^{(1)} + x_S^{(1)} \leq r_1 \text{ or } x_T^{(1)} \leq r_1'. \end{cases}$$

The one-sided p-value for testing H_0 is computed as $P(p_{0T})$, the 95% confidence interval for p_T is given by (p_T^*, p_T^{**}) where $Q(p_T^*) = 0.025$ and $P(p_T^{**}) = 0.975$, and the median unbiased estimate of p_T is given by $(p_T^- + p_T^+)/2$, where $P(p_T^-) = Q(p_T^+) = 0.5$.

To estimate the rate of the secondary endpoint, disease control, p_D, define

$$P'(p_D) = \begin{cases} \Pr\{X_T^{(1)} + X_S^{(1)} > r_1 \\ \quad \text{and } X_T^{(2)} + X_S^{(2)} \geq x_T^{(1)} + x_S^{(1)}|p_D\} \\ \quad \text{if } x_T^{(1)} + x_S^{(1)} > r_1 \text{ and } x_T^{(1)} \leq r_1' \\ \Pr\{x_T^{(1)} + x_S^{(1)} \leq X_T^{(1)} + X_S^{(1)} \leq n_1|p_D\} \\ \quad \text{if } x_T^{(1)} + x_S^{(1)} \leq r_1 \text{ or } x_T^{(1)} \leq r_1', \end{cases}$$

and

$$Q'(p_D) = \begin{cases} \Pr\{X_T^{(1)} + X_S^{(1)} > r_1 \\ \quad \text{and } X_T^{(2)} + X_S^{(2)} > x_T^{(1)} + x_S^{(1)}|p_D\} \\ \quad \text{if } x_T^{(1)} + x_S^{(1)} > r_1 \text{ and } x_T^{(1)} \leq r_1' \\ \Pr\{x_T^{(1)} + x_S^{(1)} < X_T^{(1)} + X_S^{(1)} \leq n_1|p_D\} \\ \quad \text{if } x_T^{(1)} + x_S^{(1)} \leq r_1 \text{ or } x_T^{(1)} \leq r_1'. \end{cases}$$

As before, the 95% confidence interval for p_D is given by (p_D^*, p_D^{**}) where $Q'(p_D^*) = 0.025$ and $P'(p_D^{**}) = 0.975$, and the median unbiased estimate of p_D is given by $(p_D^- + p_D^+)/2$, where $P'(p_D^-) = Q'(p_D^+) = 0.5$.

4. EXAMPLES

In the introduction we described the Lineberger Comprehensive Cancer Center Phase II trial in HER2-positive breast cancer brain metastases with the null TR rate of $p_{0T} = 0.05$ and alternative rate of $p_{AT} = 0.20$. The alternative TR rate of 0.2 was set based on previous studies [10] that have shown a tumor response rate of about 0.2 in a similar patient population. Desired power was 0.8 and the

type I error rate was 0.05. The initial plan was to use Simon's optimal design with $n = 29$, $n_1 = 10$, $r_1 = 0$, $r_2 = 3$. However, the investigators were also interested in the rate of clinical benefit and wanted to have full enrollment if the rate of clinical benefit was promising even though the rate of tumor response was small. Therefore, a stopping rule for futility after stage 1 of the form $x_T^{(1)} + x_S^{(1)} \leq r_1$ was used. Calculations show that for any stable disease rate higher than 0.048, type I error rate of the design defined above exceeds 0.05. Therefore taking Simon's optimal design and modifying its stopping rule from $x_T^{(1)} \leq r_1$ to $x_T^{(1)} + x_S^{(1)} \leq r_1$ is not going to yield a design with type I error rate below or at the nominal level of 0.05. It was assumed that $p_S \sim Unif(0, 0.2)$ in the HER2-positive breast cancer brain metastases trial, and the optimal design with $n = 28$, $n_1 = 11$, $r_1 = 0$, $r_2 = 3$ was chosen for the study. In fact, in stage 1 there was at least one patient with lasting stable disease, and therefore the trial continued to stage 2 without waiting for responses from the remaining stage 1 patients still in follow-up. Table 1 presents all admissible designs satisfying condition (2) for this trial under various assumptions on p_S.

As far as estimation, if, for example, the following counts were observed in a HER2-positive breast cancer trial $x_T^{(1)} = 0$, $x_S^{(1)} = 2$, $x_T^{(2)} = 1$, $x_S^{(2)} = 6$. The trial is not stopped for futility since $x_T^{(1)} + x_S^{(1)} = 2 > r_1 = 0$. The maximum likelihood estimate of disease control rate is $(1+6)/29 = 0.24$ with corresponding 95% naïve exact confidence interval $(0.10, 0.44)$. Adjusted median unbiased estimate with corresponding 95% exact confidence interval using methodology from Section 3 is $0.23(0.10, 0.40)$. The adjusted estimate is slightly smaller becuase the trial did not stop for futility after stage 1.

Our second example is a Lineberger Comprehensive Cancer Center Phase II trial in older patients with previously untreated diffuse large B-cell lymphoma. The primary endpoint in this trial was complete response. The null hypothesis that the CR rate is 0.5 was tested against the one-sided alternative. Required power is 80% when CR rate is 0.7 and the type I error rate is 0.05. Partial response was also of interest; therefore, stopping after stage 1 was based on the number of CRs plus PRs. With partial response $p_P \sim Unif(0, 0.2)$, the only design that minimizes $wEN_0 + (1 - w)n$ is the design with $n = 37$, $n_1 = 29$, $r_1 = 15$, $r_2 = 23$ and additional futility rule to stop after stage 1 if the number of CRs in stage 1 is less than or equal to 14. This design was used for the study. The study was stopped early for futility because the number of CRs observed in stage 1 was not high enough to reject H_0 at the end of the trial. The number of PRs was, actually higher than expected.

Two-stage designs without and with relaxed stopping for futility for these examples are displayed in Table 1. The original Simon's minimax and optimal designs [5], that is, designs without relaxed stopping for futility are presented in the lines where $p_S^U = 0$. Relaxing the stopping rule for futility often leads to less efficient designs compared to designs

in [5]. For example, in the case of testing $p_{0T} = 0.4$ versus $p_{AT} = 0.6$ (last example in Table 1), there are three admissible designs without relaxed futility stopping, minimax with $n = 39$, an admissible design with $n = 41$ and the optimal design with $n = 46$. If we would like to relax stopping for futility and assume $p_S^U = 0.3$, there is only one available design, and it requires $n = 42$. Its maximum total sample size, $n = 42$, is larger compared to designs in [5], $n = 39$ for the minimax design. If we are interested in minimizing the expected sample size under the null hypothesis it yields $EN_0 = 37.8$ compared to the best design with $p_S^U = 0$, the optimal design with $EN_0 = 24.5$.

5. DISCUSSION

This work was motivated separately by the two investigators leading the trials described in Section 4. They were each unsatisfied with the lack of flexibility in a standard two-stage design which does not allow for a trial to continue if important secondary endpoint shows promise in stage 1. We have developed easy to use software to generate designs we describe here. From the user's input values of α, $1 - \beta$, $p_{0T}, p_{AT}, p_S^L, p_S^U$, the software calculates the sample sizes (n and n_1), and decision rules (r_1 and r_2), probability of early stopping under the null hypothesis, EN_0 and weight w for each admissible design. The website also provides a recommended write-up to help with interpretation and to be used in the protocol.

Another solution to designing a trial with two outcomes CR and TR is to test the intersection hypothesis about CR and TR [11, 12] instead of testing CR alone. In this case the trial is stopped for futility if both CR and TR are low and the drug is considered promising if either CR or TR is good. This approach is also a part of our software together with various Phase II methods that are frequently used at the Lineberger Comprehensive Cancer Center. Among methods available at http://cancer.unc.edu/biostatistics/program/ivanova/ are Simon's and Fleming's two-stage designs and the method to generate stopping boundary for continuous toxicity monitoring in a Phase II trial [13].

ACKNOWLEDGEMENTS

REFERENCES

[1] CHESON, B. D., PFISTNER, B., JUWEID, M. E., GASCOYNE, R. D., SPECHT, L., HORNING, S. J., COIFFIER, B., FISHER, R. I., HAGENBEEK, A., ZUCCA, E., ROSEN, S. T., STROOBANTS, S., LISTER, T. A., HOPPE, R. T., DREYLING, M., TOBINAI, K., VOSE, J. M., CONNORS, J. M., FEDERICO, M., and DIEHL, V. (2007. International harmonization project on lymphoma. Revised response criteria for malignant lymphoma. *Journal of Clinical Oncology* **25**(5) 579–86.

[2] ABREY, L. E., OLSON, J. D., RAIZER, J. J., MACK, M., RODAVITCH, A., BOUTROS, D. Y., and MALKIN, M. G. (2001). A phase II trial of temozolomide for patients with recurrent or progressive brain metastases. *Journal of Neuro-Oncology* **53**(3) 259–65.

[3] FUKUOKA, M., YANO, S., GIACCONE, G., TAMURA, T., NAKAGAWA, K., DOUILLARD, J. Y., NISHIWAKI, Y., VANSTEENKISTE, J., KUDOH, S., RISCHIN, D., EEK, R., HORAI, T., NODA, K., TAKATA, I., SMIT, E., AVERBUCH, S., MACLEOD, A., FEYEREISLOVA, A., DONG, R. P., and BASELGA, J. (2003). Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. *Journal of Clinical Oncology* **21**(12) 2237–46. Erratum in: *Journal of Clinical Oncology* 2004, **22**(23) 4863.

[4] ROBERTSON, J. F., LLOMBART-CUSSAC, A., ROLSKI, J., FELTL, D., DEWAR, J., MACPHERSON, E., LINDEMANN, J., and ELLIS, M. J. (2009). Activity of fulvestrant 500 mg versus anastrozole 1 mg as first-line treatment for advanced breast cancer: results from the FIRST study. *Journal of Clinical Oncology* **27**(27) 4530–5.

[5] SIMON, R. (1989). Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials* **10**(1) 1–10.

[6] JUNG, S. H., LEE, T., KIM, K., and GEORGE, S. L. (2004). Admissible two-stage designs for phase II cancer clinical trials. *Statistics in Medicine* **23**(4) 561–9.

[7] JOVIC, G. and WHITEHEAD, J. (2010). An exact method for analysis following a two-stage phase II cancer clinical trial. *Statistics in Medicine* **29**(30) 3118–25. MR2758706

[8] FAIRBANKS, K. and MADSEN, R. (1982). P-values for tests using a repeated significance test design. *Biometrika* **69**(1) 69–74. MR0655671

[9] FLEMING, T. R. (1982). One-sample multiple testing procedure for phase II clinical trials. *Biometrics* **38**(1) 143–51.

[10] CARDOSO, F., DIERAS, V., CAMPONE, M., MASSACESI, C., MANLIUS, C., THEVENAZ, P., ZHANG, Y., JERUSALEM, G., SAHMOUD, T., ANDRE, F., and GIANNI, L. (2009). Everolimus, trastuzumab and paclitaxel or vinorelbine in the treatment of HER2+ metastatic breast cancer pre-treated with lapatinib-containing therapy: pooled analysis of 2 phase 1 studies. *Cancer Research* **69**(24 supplement) 5064.

[11] LU, Y., JIN, H., and LAMBORN, K. R. (2005). A design of phase II cancer trials using total and complete response endpoints. *Statistics in Medicine* **24** 3155–70. MR2209049

[12] IVANOVA, A., MONACO, J., and STINCHCOMBE, T. (2012). Efficient designs for phase II oncology trials with ordinal outcome. *Statistics and Its Interface* **5**(4) 463–469. MR3010666

[13] IVANOVA, A., QAQISH, B. F., and SCHELL, M. J. (2005). Continuous toxicity monitoring in phase II trials in oncology. *Biometrics* **61**(2) 540–5. MR2140926

Anastasia Ivanova
Department of Biostatistics
CB #7420
The University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-7420
USA

Lineberger Cancer Center at the University
of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-7295
USA
E-mail address: aivanova@bios.unc.edu

Allison M. Deal
Biostatistics Core
Lineberger Comprehensive Cancer Center at the University
of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-7295
USA