2007

Phase II Trials Powered to Detect Activity in Tumor Subsets with Retrospective (or Prospective) Use of Predictive Markers

Grishma S. Sheth
Virginia Commonwealth University

PHASE II TRIALS POWERED TO DETECT ACTIVITY IN TUMOR SUBSETS

WITH RETROSPECTIVE (OR PROSPECTIVE) USE OF PREDICTIVE MARKERS

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

by

GRISHMA SEEMIT SHETH
Masters of Science in Mathematics, Gujarat University, India, May 2003

Director: V. RAMAKRISHNAN, PH.D.
ASSISTANT PROFESSOR, DEPARTMENT OF BIOSTATISTICS

Virginia Commonwealth University
Richmond, Virginia
May, 2007

Acknowledgement

First of all, I would like to thank the Department of Biostatistics, who gave me the opportunity to pursue masters in Biostatistics. I would like to thank Dr. Ramesh for spending his time from his busy schedule in editing my thesis constantly. I would also like to thank Jessi for giving me guidance in writing the Thesis and Dr. Roberts for being in my committee. I also thank Dr. Johnson, Dr. Archer, Russ, Dr. McClish and Dr. Elswick and Dr.Best for their quick support on any question or situation.

I don't have words to thank Seemit, who has been my husband and my best friend for more than two year. I can't imagine past three years of life without him. I have come up to this stage just because of his caring and loving support. He has always appreciated me and my dedication towards study and work, which includes night and weekend work also. I am a better person only because of him.

Next, I want to thank my parents, who have been a great source of motivation to me. They have taught me to become positive and yet realistic in life. I also thank my younger brother Manan, who has shared a wonderful childhood with me and talking to him and sharing thoughts with him is always fun.

I also want to thank my parents in-laws, who has known the value of education and have always motivated me.

Last but not the least; I would like to thank God to give me such wonderful people in life and for making my entire life so joyous.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Abstract

PHASE II TRIALS POWERED TO DETECT ACTIVITY IN TUMOR SUBSETS WITH RETROSPECTIVE (OR PROSPECTIVE) USE OF PREDICTIVE MARKERS

By Grishma Seemit Sheth, M.S.

A Thesis submitted in partial fulfillment of the requirements for the degree of Masters at Virginia Commonwealth University.

Virginia Commonwealth University, 2007

Major Director: V. RAMAKRISHNAN, PH.D.
ASSISTANT PROFESSOR, DEPARTMENT OF BIOSTATISTICS

Classical phase II trial designs assume a patient population with a homogeneous tumor type and yield an estimate of a stochastic probability of tumor response. Clinically, however, oncology is moving towards identifying patients who are likely to respond to therapy using tumor subtyping based upon predictive markers. Such designs are called targeted designs (Simon, 2004). For a given phase II trial predictive markers may be defined prospectively (on the basis of previous results) or identified retrospectively on the basis of analysis of responding and non-responding tumors. For the prospective case we propose two Phase II targeted designs in which a) the trial is powered to detect the presence of responding subtype(s) as identified either prospectively or retrospectively by

predictive markers or b) the trial is powered to achieve a desired precision in the smallest subtype. Relevant parameters in such a design include the prevalence of the smallest subtype of interest, the hypothesized response rate within that subtype, the expected total response rate, and the targeted probabilities of type I and II errors (α and β). (The expected total response rate is needed for design a) but not for b)). Extensions of this design to simultaneous or sequential multiple subtyping and imperfect assays for predictive markers will also be considered. The Phase II targeted design could be formulated as a single stage or Simon two-stage design. For multiple subtyping corrections to the significance level will be considered. Sample size calculations for different scenarios will be presented. An implication of this approach is that phase II trials based upon classical designs are too small. On the other hand, trials involving "reasonable" numbers of patients must target relatively high threshold response rates within tumor subtypes. For the retrospective case we will provide the power to detect desired rates in the subtypes and provide the sample sizes required to achieve desired power. Retrospective analysis has the advantages that the analysis can be "supervised" by grouping responding and non-responding tumors; and multiple hypotheses, including hypotheses not formulated at the time of trial design, can be tested

CHAPTER 1 Background

1.1 Clinical Trials

Clinical trials are prospective studies comparing the effect and value of an intervention (drug or therapy) against a control (placebo or standard therapy). A properly planned and executed clinical trial is a powerful experimental tool for assessing the effectiveness of the intervention.

While the ultimate goal of the clinical trials is the design and analysis for the purpose of comparing the effectiveness of one or more interventions, several steps or phases of clinical research must occur before reaching that goal. These steps include, in addition to the pre-clinical trials four clinical segments called Phases. The pre-clinical trial focuses on creation of a new drug and testing it on lab animals for toxicity and efficacy. This includes Pharmacokinetics (PK) and pharmacodynamic (PD) modeling. They are the key tools in proper dose selection as an early stage of the phase I trial. PK attempts to characterize the fate of the drug in the body following dosing, primarily by sampling its concentration-time profile in the circulation. PD investigates the relationship between the response induced by the drug and its circulating concentration. Once the drug passes the pre-clinical stage and the appropriate dose is found the drug is then tested on humans. The human phase of the clinical trial follows the pre-clinical trials. These phases are briefly summarized next.

The first phase, called Phase I or sometimes referred as the dose finding phase, deals with determining a dose of a drug or a regimen to understand how well it is tolerated in a small number of individuals. The main aim here is to estimate how large a dose can be given before unacceptable toxicity is experienced by patients. This dose is referred to as the maximally tolerated dose (MTD). Several designs have been published to find MTD, among which more commonly used designs are provided by Fleming (1982), O'Brien (1979) and Piantadosi (1997).

The second phase, called Phase II trial or Safety and Efficacy Design, evaluates the drug's biological activity or effect on humans and may also be used to estimate the rate of adverse events. This trial is conducted at the maximally tolerated dose is established in the Phase I trial. Thus, the phase II design depends on the quality and adequacy of the phase I study. Phase II studies are sometimes divided into Phase IIA and Phase IIB. Phase IIA is specifically designed to assess dosing requirements, whereas Phase IIB is specifically designed to study efficacy. The most commonly used sample size calculations for the phase II designs in cancer includes Gehan's (1960) approach or the optimal two-stage design proposed by Simon (1988). A detailed summary of these papers will be presented in section 1.2 and 1.3 respectively.

The third phase, called the Phase III trial is designed to assess the effectiveness of the new drug or therapy and thereby its role in clinical practice by comparing it with standard therapy or drug or in the case of new drugs with placebo. The comparison treatment and the new treatment are referred to as arms of the trial.

The fourth phase, called Phase IV is in fact post clinical research mainly concerned with marketability of the new treatment or drug. This is a long term surveillance of an intervention and does not involve control groups.

The focus of this thesis is on the phase II part of the clinical trials applied in cancer research, in which the trials need to be powered to estimate the response rate in tumor 'subtypes' through the application of 'predictive markers'. A formal definition of a subtype and an explanation of a predictive marker are provided in section 1.5. A review of sample size estimation and a review of use of subtype analysis in cancer trials are provided in the next section.

1.2. Sample Size for Phase II and Follow-Up Clinical Trials

Gehan (1960) describes a Phase II design and provides the number of patients to be used in the phase II trials under two scenarios. The first scenario is a simple case in which the trial is designed to estimate a specified response rate given a rejection error. The response rate specified to be of interest may arise from the researchers' experience or from past trials. The rejection error corresponds to incorrectly rejecting the response rate to be less than the specified rate. This rejection error is usually set at 5%. This scenario is primarily to decide whether or not a drug shows any biological activity to a particular drug leading to a minimum effectiveness so as to proceed to a larger trial. The second scenario involves, in addition to the parameters above, a precision within which a researcher wishes to estimate the response rate. In this scenario, depending on the precision specified, the sample size needed may increase considerably.

As an example, a researcher may specify that a drug must have some minimal level of activity, say in 20% of the patients. If the estimated activity level from a trial involving a small number of patients is less than 20%, she/he may choose not to consider this drug any further. If the estimated activity level exceeds 20%, she/he may want to study more patients to get a better estimate of the response rate (the number of patients required depends on the number of successes in the first stage). A typical study for ruling out a 20% or lower response rate with 5% rejection error enters 14 patients. If no response is observed in these 14 patients the drug is considered not likely to have a 20% or higher activity level and therefore the trial ends. This is because, failure in 14 consecutive patients would happen 5% or less times if the drug were truly effective 20% or more of the time. This can be established by considering response as a success and a non-response as a failure in a binomial trial with success probability equal to the specified response rate of interest. The number of patients required in the first stage (n) according to the specified response rate (p) can be found by using the exact binomial probability. That is, one could find n with 5% rejection error and a response rate (p) using

$$P(x < r) = \sum_{x=0}^{r-1} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} < 0.5, \qquad (1.1)$$

where r is number of successes.

Usually r is set at 1, which yields a simplified version of the above equation,

$$P(x < 1) = p \ (1-p)^{n-1} < 0.5.$$

The number of additional patients required for the follow-up trial depends on the degree of precision required, which could be specified in terms of the standard error. Since some information on the response rate is available from the first scenario, Gehan suggests using the upper 75% confidence limit as the specified response rate. This is an improvement over the response rate specified by the researcher at the beginning of the trial, because it incorporates the information regarding the number of successes in the first stage. Now, using this estimated response rate ($\hat{p}$) and a desired precision (p_e), the sample size for the second scenario is calculated by the following formulae

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = p_e . \tag{1.2}$$

To illustrate this, considering 20% response rate with 10% rejection error, 11 patients are needed for the phase I. Suppose 1 success is observed out of 11 trials. The upper 75% confidence limit is found for the precision found from the above formulae using the response rate $\hat{p}$ =1/11, which turns out to be 0.1906. Now, using this upper limit as the response rate $\hat{p}$ in the above equation (1.2) again with specified precision of 5% gives the required number for the follow-up trial to be 71. Thus, one can conclude that one may need 60 additional patients for the follow-up trial to satisfy the precision required. On the other hand, if there were 2 out of 11 successes, it can be shown that 78 additional patients are required for the follow-up trial to satisfy the required precision. Therefore the additional information from the first stage is essential in the calculation of sample size needed in the second stage. Sample size increases up to 50% of observed success and it starts decreasing beyond that. This happens because of the nature of binomial distribution.

This design could be extended as an optimal design in the sense that the expected size is minimized if the regimen has low activity, subject to constraints upon the size of type I and type II errors. This design due to Simon (1988) is next summarized.

1.3 Optimal II stage design for Phase II Trials

Simon (1988) develops the design to estimate the sample size of phase II trials based on testing a null hypothesis $H_0: P \leq P_0$ that the true response probability is less than an uninteresting probability level P_0. This hypothesis is tested at significance level of α to conclude that the drug is accepted for further study. If H_0 is rejected, the new alternative hypothesis $H_1: P > P_1$ for some P_1 is considered. If H_1 is true, then the probability of rejecting the drug for further trial is set to be less than β. Essentially, α and β are type I and type II errors of hypothesis testing. Unlike Gehan, Simon uses hypothesis tests rather than confidence interval for the precision and applies the calculations of exact binomial probability.

The two-stage design considers the expected sample size (EN) to be n_1 + (1-PET) n_2, where n_1 and n_2 are the numbers of patients studied in stages I and II, respectively. Here, PET is probability of early termination after stage I, which is based on the number of responses observed over the first n_1 patients. The EN and PET depend on true probability of response *p*. The trial is considered ineffective at the end of stage I and the drug is rejected, if r_1 or fewer responses are observed, where PET is equated to $B(r_1; p, n_1)$, and B denotes the cumulative binomial distribution. That is,

$$\text{PET} = \text{B}(r_1; p, n_1) = \Pr(\text{X} \le r_1) = \sum_{i=0}^{r_1} \frac{n_1!}{i!(n_1 - i)!} p^i (1-p)^{n_1 - i} \quad . \qquad (1.3)$$

If r or fewer responses are observed, the trial is considered ineffective and the drug is rejected at the end of stage II. The probability of rejecting a drug given success probability, p, is

$$\text{Probability of rejecting a drug} = \text{B}(r_1; p, n_1) + \sum_{x=r_1+1}^{\min\{n_1, r\}} b(x; p; n_1)\ \text{B}(r - x; p, n_2). \qquad (1.4)$$

where b denotes the binomial probability mass function.

For specified parameters p_0 , p_1 , α and β, the two stage design is defined as the one that satisfies the error probability constraints α, β and minimizes the EN, given the response probability is p_0 . Here, the optimization is performed over all values of n_1 , n_2, r_1 and r.

First, n is found, starting from the lowest value, which is given by the formula,

$$p(1-p)\left[\frac{(z_{1-\alpha} + z_{1-\beta})}{(p_1 - p_0)}\right]^2 \quad . \qquad (1.5)$$

where $p = (p_0 + p_1)/2$. Now keeping n fixed, n_1 can be searched within the range (1, n-1). For each value of total sample size n and each value of n_1 in the range (1, n-1), the integer values of r_1 and r are determined, that satisfy two constraints α, β and minimize the expected sample size when $p = p_0$. r_1 can be ranged between (0, n_1). For each value of r_1 the maximum value of r is determined that satisfies type II constraints. That is, value of

equation (1.4) is at least 1-β. Then the set of parameters (n_1, n_2, r_1, r) are examined to check the whether they satisfy the type I constraint. That is, the value of equation (1.4) is at the most α. If it does, then the EN is compared and the minimum value can be determined.

The optimization criterion chosen here is not unique. One could minimize the expected sample size averaged with regard to a prior distribution for the true response probability p.

1.4 Evaluating the Efficiency of Targeted Designs

Simon and Maitournam (2004) propose a design for evaluating the efficiency of targeted designs for randomized clinical trials. They also study the efficiency of targeted designs in comparison with traditional randomized design. For targeted and untargeted design, the comparison of a control versus a new experimental treatment with the same number of randomized patients in two groups is considered. The efficiency is evaluated with regard to number of patients required for randomization and number required for screening in the context of a binary outcome such as survival or time-to-progression end points.

According to their notation specifications, $R+$ is the population of patients who are predicted to be responsive to the new treatment and $R-$ is the remaining population. The p_C is the response probability in control group, $p_C + \delta_0$ is the response probability in treatment group for patients in $R-$ and $p_C + \delta_1$ is the response probability in treatment group for $R+$ patients. Let p_e be the overall response probability for experimental

treatment group and $p_e^T = p_e + \delta_1$ be the response probability for experimental treatment group in targeted design. The hypothesis of interest is $H_0 : p_C = p_e$. Also, let n and n^T denote the number of patients required to randomize in untargeted and targeted designs, respectively, to achieve the same statistical power for testing H_0. The relative efficiency of untargeted and targeted designs can be expressed in the form:

$$\mathrm{n/n^T} = \left[\frac{\delta_1}{R\_\delta_0 + (1-R\_)\delta_1}\right]^2 f \qquad (1.6)$$

They consider the two scenarios when R- patients do not benefit from the new treatment ($\delta_0 = 0$: case i) and R- patients benefit half as much as the $R+$ patients ($\delta_0 = \delta_1/2$: case ii). For both the scenarios, considering the response probability for control group, p_C to be equal to 0.1 or 0.5 and the improvement in the response probability for $R+$ patients, δ_1 to be equal to 0.2 or 0.4 with $\alpha = 0.025$, $\beta = 0.2$, they conclude that when the new treatment benefits only a subset of patients and those patients can be accurately identified, the targeted design required fewer randomized patients than the untargeted design (ratio>1 in the equation 1.6). However, the advantage of the targeted design is much greater for case i compared to case ii. The degree of reduction depends heavily on the availability of an assay for identifying all patients who will benefit from the new treatment and the prevalence of such patients. They do not provide methods for calculating sample sizes for designing such targeted designs.

1.5 Introduction to Subtype and Targeted Designs

Cancer therapy often is found to be effective only for a subset of treated patients. If the tumors of patients enrolled are homogeneous with respect to a specific characteristic, then almost all of the tumors of that type are likely to respond in a similar manner to the drug of interest, therapy or regimen. However, if the patients are heterogeneous (say, have different mutations of a gene or different types of melanoma, etc.), the treatment response may vary with respect to the patients characteristics. If the patients' characteristics, which are called 'predictive markers', could be determined and the patients could be categorized into known 'subtypes' it may be possible to target the trials to specifically estimate the response rates in the subtypes. The Phase II clinical trials that are designed to study the responses in such homogeneous subtype populations are called targeted designs (Simon, 2004). In the next section, examples of subtype are provided.

1.6 Mutation in EGFR and responsiveness of lung cancer to Gefitinib

Lynch et al. (2004) present a study designed to evaluate the effectiveness of gefitinib for treating the tumor of non-small-cell lung cancer. A total of 275 patients with advanced, chemotherapy-refractory non-small-cell lung cancer were treated with gefinitib as a single agent since 2000 at Massachusetts General Hospital. Out of these, 25 patients were identified by physicians as having clinically significant responses to the drug. Thus most of the patients with non-small-cell lung cancer showed no response to the gefitinib. It

was also found that those who responded 10 percent patients (25 out of 275) showed a rapid and dramatic clinical response to gefitinib. If one seeks the typical response rate of interest in phase II trials of 20%, and designed this trial using the Gehan's approach described in section 1.2 the trial may have not proceeded after the first 14 patients. This is because; the probability of observing no successes out of 14, when the response rate is 10% is 23%, which is larger than usually accepted 5% rejection error.

Although the study led to only a small number of successes, the authors study the responders further with the intention to test the hypothesis that gefitinib might target the epidermal growth factor receptor (EGFR). In other words, their hypothesis is that the patients with non-small-cell lung cancer who have striking responses to gefitinib had somatic mutations in EGFR gene, which would indicate the essential role of the EGFR signaling pathway in the tumor response. Therefore, they searched for mutations in the EGFR gene in primary tumors from a subset of patients with non-small-cell lung cancer who had responded to gefitinib and a sample of those who did not respond. Also, to estimate the prevalence of EGFR mutations, they tested tumor patients who have not been exposed to gefitinib.

Out of the 25 responders tumor specimens were available only for nine patients. These nine patients responded substantially to gefitinib therapy. When tested, somatic mutations were identified in the EGFR gene in eight of the nine patients, while no mutations were observed in a sample of seven patients with non-small-cell lung cancer who did not respond to gefinitib (Table 1). They use the Fisher's exact test to establish

statistical significance at a p-value of 0.001. Next they studied 25 patients with primary non-small-cell lung cancer, who have not been exposed to gefitinib, to estimate the prevalence. The EGFR mutations were detected in two tumors out of 25. Therefore the prevalence is estimated as 8% (with a standard error of ± 5.4%). They also mentioned that these two patients had heterozygous mutations. They also reported that there were no mutations among 95 primary tumors and 108 cancer-derived cell lines, which represented diverse tumor types. This suggests that only a subgroup of cancers patients, namely non-small-cell lung cancer patients, showed EGFR mutations and these may be variability in the mutations.

Table 1: EGFR mutations and Response to Gefitinib

	R+	*R*-	Total
EGFR Mutation (+)	8	0	8
EGFR Mutation (-)	1	7	8
Total	9	7	16

R+= Response to gefitinib and *R*-= No Response to gefitinib

In summary, they conclude, only a subgroup of patients with non-small-cell lung cancer who have specific mutations in the EGFR gene, which correlate with clinical responsiveness to gefitinib, would respond to gefitinib. Therefore they propose screening for such mutations in lung cancers to identify patients who would respond to gefitinib.

The trial was not designed as a prospective study to examine the response to gefitinib in the EGFR subtype. Instead, the authors perform this retrospectively. Since a large number of patients were included in the trial and since they identified somatic mutations in the EGFR gene in a preponderantly many responders (eight out of nine or 89 %) the question of power did not arise. In addition they found no EGFR mutations in the sample of non-responders (zero out of 7). Adding further to the power of the test, in general, one may not be so serendipitous and therefore if such subtypes are anticipated one should design the trials to adequately study the response rate among them.

1.7 Somatic activation of *KIT* distinct subtypes of melanoma

This article has been provided by Curtin et al. (2006), which studies melanoma and presented here as an example of other cancers, where subtypes are present. In this article there is no treatment and no phase II study is proposed. In general, the melanoma is categorized into four groups based on sun exposure and anatomic site namely Chronic sun exposure with damage (CSD) (e.g., face), intermittent sun exposure without damage (acral) (e.g., trunks, arms, legs), Minimal sun exposure (Non CSD) (e.g., soles, palms) and protected from sun (mucosal) (e.g., mucosal membranes).

They hypothesize that mutations or multiple copies of *KIT* may affect a protein called MAP. Also, mutations occur in genes NRAS or BRAF or NRAS with *KIT* or BRAF with *KIT*. They notice that mutations in NRAS and BRAF do not occur simultaneously and therefore do not consider *KIT* in combination with both. Mutations in *KIT* are more likely to occur in acral, mucosal melanomas and CSD but do not occur in Non CSD. Mutations

in BRAF gene are highly prevalent (59%) in Non CSD melanomas, while mutations in BRAF occur significantly less frequently on mucosal melanomas and on CSD melanomas.

They analyzed DNA specific to melanoma subtypes where mutations in *BRAF* and *NRAS* are infrequent among 102 primary melanomas (38 from mucosa, 28 from acral skin, and 18 from CSD and 18 from Non CSD). They use two tailed Mann – Whitney U test to compare *KIT* antibody expression levels between melanomas with *KIT* mutations or melanomas without such mutations in which P–values less than 0.05 were regarded as significant. Mutations in *KIT* were found in three of seven tumors. Examination of all 102 primary melanomas found mutations and/or copy number increases of *KIT* in 39% of mucosal, 36% of acral, and 28% of melanomas on CSD, but not in any (0%) melanomas on non CSD. Eleven out of 14 (79%) of mutations with *KIT* mutations and eight of 15 (53%) with multiple copies of *KIT* demonstrated increased *KIT* protein levels.

They conclude that *KIT* is an important gene in melanoma. Thus there are different subtypes of melanoma. A treatment may work on some subtypes but may not on all subtypes. Because the majority of the *KIT* mutations were found in melanoma also occurs in imatinib-responsive cancers of other types, imatinib may offer an immediate therapeutic benefit for a melanoma. In order to design a trial to study this, a phase II trial that has enough sample size is desirable. This will be addressed in chapter 2.

1.8 Other related study designs:

There are numerous articles on the existence of subtypes and statistical methods for designing targeted trials. One of them is provided by Freidlin and Simon (2005). They propose an adaptive design that prospectively combines the development of a classifier based on gene expression to select sensitive patients to test the overall effect with a proper power. To address a question whether addition of a new targeted agent to the standard treatment is beneficial, a phase III clinical trial is conducted where patients are randomly assigned to experimental group (which consists of combination of new and standard treatment) and control group (which consists of standard treatment alone). Performance of the adaptive design, relative to the more traditional design, is evaluated in a simulation study. It is shown that when the proportion of patients sensitive to the new drug is low, the adaptive design is shown to substantially reduce the chance of false rejection of effective new treatments. When the new treatment is broadly effective, the adaptive design seems to have adequate power to detect the overall effect similar to the traditional design. They conclude that development of a gene expression–based classifier to identify the subset of sensitive patients can be prospectively incorporated into a randomized phase III design without compromising the ability to detect an overall effect. Pusztai et al., (2007) provided the limitations and their alternatives of pharmacogenomic predictors in phase II clinical trials.

CHAPTER 2 Sample Size for Phase II Design

2.1 Introduction to targeted design within classical phase II trial

In cancer research, a classical phase II trial design assumes a patient population with a homogeneous tumor type and aims to estimate the probability that a patient's tumor will respond to a particular drug. If the homogeneity of tumor holds, the measurement of the outcome (response or non-response) is an exercise involving only a few patients. The phase II trial design focuses on optimizing cost by minimizing the numbers of patients treated.

In Chapter 1 several examples of subtype were presented. In those examples, the number of tumors with the subtype form a sample design constructed for a traditional phase II trial may not yield adequate power or precision. Therefore, phase II clinical trials that are targeted to prospectively study the effectiveness of the treatments in the subtypes, extensions of the traditional designs should be considered. To design such trials, ideally the subtype information would be desirable from all the treated individuals. But due to cost, often the subtype information is obtained only for the responders. However, the prevalence of the subtype in the affected population may be available. A method for computing sample size under these circumstances is proposed below for this purpose. Here estimation of a specified response rate in subtypes is a desirable quantity.

2.2 Methods

In this section, the methods for obtaining sample sizes for different scenarios that might arise in targeted Phase II trials are described. In these scenarios it is assumed that the status of the subtype will be determined only for the responders. (The case in which the entire sample of treated patients is examined for the subtype will be discussed later in section four)

The primary goal here is to determine the number of responders needed to estimate the proportion of responders in any given subtype of interest and subsequently to determine the total number of patients needed given an expected total response rate. In standard phase II trials (one-stage or two-stage) the total sample size needed is determined only as a function of the expected total response rate, say p_r along with an acceptable rejection error. Sometimes the required precision (width of the confidence interval) is given along with rejection error, within which one wants to estimate p_r. Sometimes it might be of interest to test the hypothesis on the response rate on the subtype with required precision and from which to derive the total sample size. For all of these scenarios, according to the given specifications and the sample size, a certain number of responses are expected. If the trial produces fewer than the expected number of responses the treatment is considered ineffective. The method proposed here first determines the number of responses needed to estimate the subtype response rate with the specified precision and adequately test hypotheses on the subtype from which the total sample size is derived.

Before describing the methodology, some notations are introduced. Let the total sample size (e.g., the total number of melanoma patients receiving the treatment irrespective of subtype status) be denoted by N. Let n_r denote the number of patients responding to the treatment (e.g., responding to Gefinitib). Let r be the number of patients with the subtype among the responders (e.g., patients with EGFR mutation among those responding to Gefitinib). Let the prevalence of the subtype in the population of affected individuals be denoted by P_S, which will be assumed to be known or the minimum acceptable is specified. Let the conditional probability of response to a treatment given that the individual is in the subtype be denoted by θ_r. That is,

$$\theta_r = \text{P (Response | Subtype)}.$$

Let p_S be the conditional probability of subtype given the number of responders. That is,

$$p_S = \text{P (subtype | responders)}.$$

In targeted phase II trials, the subtype information is only obtained among the responders, as mentioned earlier, and often the expected response rate θ_r is specified. Using the definition of conditional probability, the joint probability that a patient with tumor will respond and have the subtype is,

$$P(\text{subtype and responder}) = P(\text{subtype}|\text{responder})\,P(\text{responder}) = p_S p_r.$$

Since the number of responders with the subtype and without the subtype and the non responders form a mutually exclusive group and since the total number of patients, N

can fall into either one of these three categories, $r, n_r - r$ or $N - n_r$ follow a multinomial distribution with cell probabilities $p_S p_r$, $p_r(1-p_S)$ and 1- p_r. That is,

$$(r, n_r - r, N - n_r) \sim \text{Multinomial}(N, p_S p_r, p_r(1-p_S), 1-p_r)$$

Note that $r + n_r - r + N - n_r = N$ and $p_S p_r + p_r(1-p_S) + 1 - p_r = 1$, as required by the multinomial distribution. Although the distributions are defined in terms of p_S and p_r, the investigator often specifies θ_r, as mentioned above. Therefore a derivation of θ_r, as shown below, in terms of the parameters in the multinomial distribution is necessary.

$$\begin{aligned}
\theta_r &= \text{Probability of responders in a subtype} \\
&= P\,(\text{Response} \mid \text{Subtype}) \\
&= \frac{P\,(\text{Subtype and Response})}{P\,(\text{Subtype})} \\
&= \frac{p_S p_r}{P_S}. \qquad (2.1)
\end{aligned}$$

Now, the specifications of the design, such as the precision or effect sizes in the tests of hypotheses could be restated in terms of the parameters of the multinomial distribution. This can be achieved by rewriting for given θ_r, P_S *and* p_r

$$p_S = \frac{\theta_r P_S}{p_r} \qquad (2.2)$$

2.3 Five Scenarios of interest

2.3.1 The response rate in the total population is specified

In this scenario detecting a sample size for specified response rate in the total sample size is of interest. This case is exactly the same as the traditional case but included here for completeness. The subtype information is not necessary. The number of patients responding to the treatment n_r can be shown to follow a marginal binomial distribution. That is, $n_r \sim binomial(N, p_r)$. The number of patients required for a specified trial of a new agent for therapeutic effectiveness in the total population, for a given rejection error, can be found using exact binomial probability as prescribed by Gehan (1961, Table 1). That is, given the response rate in the total population (p_r) and rejection error (α) the required number of patients can be found by solving for N using

$$P(x < r) = \sum_{x=0}^{r-1} \frac{N!}{x!(N-x)!} {p_r}^x (1-p_r)^{N-x} < \alpha , \tag{2.3}$$

where r is number of successes, which is usually set at 1.

2.3.2 The response rate in the total population is specified with precision

In this scenario detecting a sample size for specified response rate with required precision in the total sample size is of interest. The subtype information is not necessary. The number of patients responding to the treatment n_r can be shown to follow a marginal binomial distribution. That is, $n_r \sim binomial(N, p_r)$. The number of patients required for a specified trial of a new agent for therapeutic effectiveness in the total population, for a

given rejection error and with required precision can be found using exact binomial probability as prescribed by Gehan (1961, Table 2), which is explained in detail in section 1.2.

2.3.3 The response rate in the Subtype is specified

In this scenario detecting a sample size for specified response rate among the subtype is of interest. The prevalence of the subtype is assumed to be known. To design this Phase II study the required sample size is of interest. The number of patients in the subtype r can be shown to follow marginally a binomial distribution by summing the joint multinomial distribution of $r, n_r - r, N - n_r$ over n_r.

That is,

$$P(r) = \sum_{n_r=r}^{N} \frac{N!}{r!(n_r - r)!(N - n_r)!} (p_S p_r)^r (p_r(1 - p_S))^{n_r - r} p_r{}^{N - n_r} \tag{2.4}$$

$$= \frac{N!}{r!(N-r)!} (p_r p_S)^r (1 - p_r p_S)^{N-r} \tag{2.5}$$

This is denoted as $r \sim binomial(N, p_r p_S)$. The number of patients required for a specified trial of a new agent for therapeutic effectiveness in the subtype for given an expected total response rate of p_r and for a given rejection error, can be found by using exact binomial probability as prescribed by Gehan (1961, Table 1). That is, given the response rate in the total population (p_r) and rejection error (α) the required number of patients can be found by solving for N using

$$\sum_{x=0}^{r-1} \frac{N!}{x!(N-x)!}(p_r p_S)^x (1-p_r p_S)^{N-x} < \alpha\,, \tag{2.6}$$

where r is number of successes, which is usually set at 1.

When only θ_r and P_S are specified, one can calculate $p_r p_S$ using

$$p_r p_S = p_r \frac{\theta_r P_S}{p_r} = \theta_r P_S\,. \tag{2.7}$$

2.3.4 The response rate in subtype is specified with desired precision

In this scenario detecting a sample size for specified response rate with desired precision among the subtype is of interest. The prevalence of the subtype is assumed to be known. To design this Phase II study the required sample size needs to be provided. As mentioned in scenario 3 the number of patients in the subtype *r* can be shown to follow marginally a binomial distribution by summing the joint multinomial distribution of $r, n_r - r, N - n_r$ over n_r.

This is denoted as $r \sim binomial(N, p_r p_S)$. The number of patients required for a specified trial of a new agent for therapeutic effectiveness in the subtype for given an expected total response rate of p_r and for a given rejection error, can be found by using exact binomial probability as prescribed by Gehan (1961, Table 2), which is explained in detail in section 1.2. When only θ_r and P_S are specified, we use

$$p_r p_S = p_r \frac{\theta_r P_S}{p_r} = \theta_r P_S\,.$$

2.3.5 Test the hypothesis on the specified subtype response rate

This scenario deals with the case when the expected response rate in subtype (θ_r) and in the entire population (p_r) are expected to be different. In other words the hypothesis of interest is to examine if the expected response rate in subtype is greater than that of the total population. This could be also stated in terms of null and alternative hypotheses regarding the subtype of interest, respectively as follows:

$$H_{01}: \theta_r \leq p_r,$$

$$H_{a1}: \theta_r > p_r.$$

In order to test this hypothesis the subtype status of the entire sample is desirable. However, it is often possible to test the subtype status of only the n_r responders and not to test all the N individuals. Consequently, the estimation of θ_r is not feasible. Therefore, the determination of the sample sizes for this scenario will be formulated in terms of hypotheses on the conditional probability of an individual being in the subtype given that the individual is a responder, denoted by p_S. Also, the prevalence P_S of the subtype among the tumor population will be assumed to be known or minimum P_S of interest will be specified. Then, if the alternate hypothesis H_{a1} above is true, p_S would be greater than the prevalence P_S. So, the hypotheses regarding θ_r and p_r could be formulated as

$$H_{02}: p_S \leq P_S,$$

$$H_{a2}: p_S > P_S.$$

Under this setup the sample size problem reduces to translating the desired magnitude of the effect specified in terms of difference between θ_r and p_r into a magnitude specified in terms of the difference between p_S and P_S.

Consider the following 2 × 2 table in which the rows represent whether or not an individual is in the subtype and the columns represent whether or not an individual responds to treatment. The corresponding cell probabilities and frequencies are shown in the table.

Table 2: 2×2 table of response and subtype

Subtype	Response: Yes	Response: No	Total
Yes	R p_S, θ_r	$n_S - r$ 1 - θ_r	P_S n_S
No	$n_r - r$ 1 - p_S	$N - n_S - n_r + r$	1- P_S N- n_S
Total	p_r n_r	1 - p_r N - n_r	1 N

Suppose, under H_{a1}, a difference $\Delta_1 = \theta_r - p_r$ is specified for the power (sample size) calculations. As discussed earlier, since θ_r is inestimable the problem will be reformulated in terms of $\Delta_2 = p_S - P_S$.

Notice,

$$\Delta_1 = \theta_r - p_r$$

$$= \frac{\pi_{11}}{P_S} - p_r$$

$$= (\pi_{11} - p_r P_S) / P_S ,$$

where π_{11} is the joint probability of responders and subtype, namely $p_r p_S$.

Similarly,

$$\Delta_2 = p_S - P_S$$

$$= (\pi_{11} - p_r P_S) / p_r .$$

Solving for Δ_2 using the above two equations yields,

$$\Delta_2 = p_S - P_S = \Delta_1 \frac{P_S}{p_r} . \qquad (2.8)$$

Thus, given P_S, p_r and $\Delta_1 = \theta_r - p_r$ one can easily determine Δ_2 using equation (2.7). Subsequently the sample size n_r can be determined using the conditional distribution of r given n_r for testing H_{02} for specified significance level α and power, 1 - β. That is using,

$$P(r \mid n_r) = \frac{P(r, n_r; N)}{P(n_r)} \sim \text{binomial}\ (p_S, n_r). \qquad (2.9)$$

Once n_r is determined the total sample size N can be estimated by using marginal distribution of n_r, which is a binomial (p_r , N). Given an expected total response rate, p_r of interest, the number of responses needed above (i.e., n_r) and a rejection error, the total sample size N could be calculated using Gehan (1961). Expanding Gehan's approach for r successes, the equation can be written as,

$$\sum_{x=0}^{n_r-1} \frac{N!}{x!(N-x)!} {p_r}^x (1-p_r)^{N-x} < \alpha \,. \tag{2.10}$$

CHAPTER 3 Sample Size Tables for the Targeted Phase II trials

In this chapter tables of sample sizes for various combinations of prevalence, response rate in total population as well as in subtype are provided for all possible scenarios.

3.1 Sample sizes for scenarios 1 and 2

In scenario 1, the expected response rate in the total population of tumor patients (p_r) is specified. The sample sizes for the typical response rates, namely, p_r = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3 and 0.35 assuming 5% and 10% rejection errors are calculated. Since this situation is identical to the usual Phase II trials, these values are identical to the tables provided by Gehan (1961, Table 1). To obtain this for any response rate and any rejection errors a SAS program is provided.

In scenario 2, in addition to the expected response rate a required precision must also be specified. The required sample sizes for the values of p_r mentioned above in scenario 1, for 5% and 10% precisions and 5% and 10% rejection error are provided. For scenario 2 these values are recalculated from tables provided by Gehan (1961, Table 2). SAS program is provided for this scenario to find sample size for any required precision and any response rate with any rejection errors.

In scenario 1 if there are no responses (that is, at least one response is not observed) in the sample, the trial will be terminated. This is true also in scenario 2. That is, if no responses in the number of trials mentioned under scenario 1 are observed the trial will terminate and will not accrue the additional patients required under scenario 2. Therefore, whether one is in scenario 1 or 2 the sample sizes needed under scenario 1 is essential. Therefore the sample sizes for scenario 1 and 2 are provided together in Table 3 below.

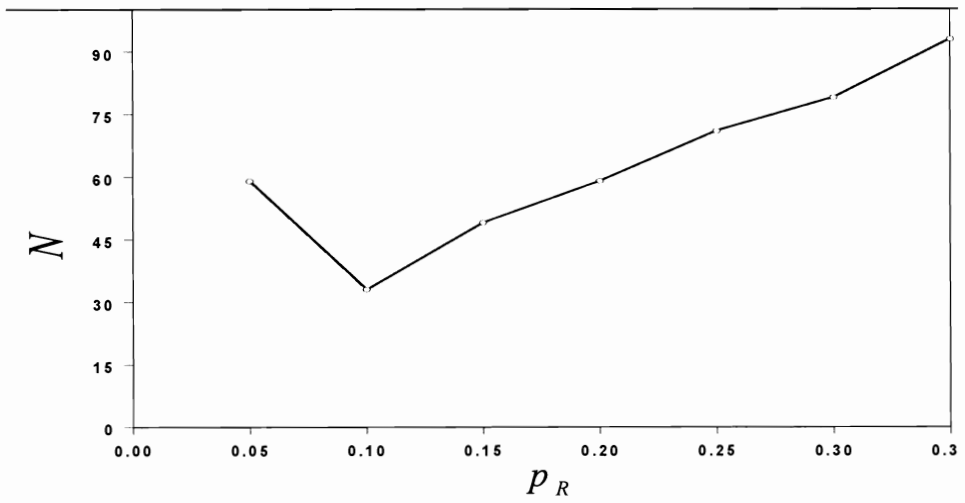
Table 3: Sample size for scenario 1 and 2

p_r	Scenario 1	Scenario 2 (5% precion)	Scenario 2 (10% precion)
Rejection Error=5%			
0.05	59	59	59
0.1	29	33	29
0.15	19	48	19
0.2	14	61	16
0.25	11	72	18
0.3	9	82	21
0.35	7	93	24
Rejection Error=10%			
0.05	45	45	45
0.1	22	42	22
0.15	15	58	15
0.2	11	72	18
0.25	9	82	21
0.3	7	93	24
0.35	6	98	25

For example, if the expected response rate is 20% and the rejection error is 5% the trial will stop if no responses are observed in the first 14 patients. If the trial proceeds and one

also requires 5% precision the trial will require a total of 59 patients. That is, it will require an additional 45 patients.

Figure 1: Sample size as a function of p_r for Scenario 2



The figure 3.1 represents the total sample size as a function of p_r. The figure indicates that the sample size increases linearly for response rates larger than 10% while it is quadratic on the whole if the 5% case is included. This is because; the standard errors of 5%, 10% and 20% at rejection error of 0.05 for example, are 0.0002, 0.0015 and 0.0008, which is quadratic.

Program 3.1: SAS codes for scenario 1

```
%macro table1(pr,re);
data x;
Re=&re;
Pr = &pr;
do i = 1 to 100 until (flag);
n = (1-pr)**i;
if n < &re then do;
flag=1;
```

```
flag=1;
N1 = i;
output;
end;
end;
run;
proc print data=x noons;
var Re Pr N1;
title ' The Required Sample Size for Scenario 1';
run;
%mend;
%table1(.35,.05);

run;
```

In the above program the inputs pr and re correspond response rate in total population and rejection error respectively. If one specifies the response rate in the total population (p_r) and rejection error, the program will provide the required sample size. The output is provided below as an example, where p_r =0.35 and rejection error 95% are specified. The required sample size comes out to be 7 from the above program.

Table 4: Sample size for scenario 1

The Required Sample Size for Scenario 1

Re	Pr	N1
0.05	0.35	7

Program 3.2: SAS codes for scenario 2

```
%macro table2(pr,re,os,prec);
data x;
Pr = &pr;
Re=&re;
Observed_Success=&os;
Precision=&prec;
do i = 1 to 100 until (flag);
n = (1-Pr)**i;
if n < &re then do;
flag=1;
N1 = i;
output;
end;
end;
      p1 = &os/n1;
            se1 = sqrt(p1*(1-p1)/n1);
            phat1 = p1 + 1.64*se1;
            N2 = int((phat1*(1-phat1)/&prec**2)) + 1;
            output;
run;
data x;
set x;
if N2 = . then delete;
run;
proc print data=x noobs;
var Re Pr Observed_Success Precision N1 N2;
title ' The Required Sample Size for Scenario 2';
run;
%mend;
%table2(.05,.05,1,.05);run;
```

In the above program the inputs pr, re, os and prec correspond response rate in the total population, rejection error, observed success and precision respectively. If one specifies the response rate in the total population (p_r), rejection error, required precision and the observed success the program will provide the required sample size. This program also provides the sample size for more than 1 success. The output is provided below as an

example, where p_r =0.05 and rejection error 95%, with 5% precision and 1 observed success are specified. The required sample size comes out to be 59 for scenario 1 and 18 for scenario 2 (if at-least 1 success is observed in scenario 1) from the above program.

Table 5: Sample size for scenario 2

The Required Sample Size for Scenario 2

Re	Pr	Observed_Success	Precision	N1	N2
0.05	0.05	1	0.05	59	18

3.2 Sample Sizes for Scenarios 3 and 4

In scenario 3, the expected response rate in the subtype of tumor patients (θ_r) and the subtype prevalence (P_S) are specified. The sample size for the typical response rates, namely θ_r =0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65 and 0.7 assuming 5% and 10% rejection error with subtype prevalence P_S =0.2. Some of these values are identical to the tables provided by Gehan (1961, Table 1) using $p_r p_S$ instead of p_r (unlike in scenario 1), from the equation 2.1. That is,

$$p_r p_S = \theta_r P_S$$

The SAS program (3.3) provided below for this scenario finds sample size for any response rate in the subtype, for any subtype prevalence and with any rejection errors.

In scenario 4, in addition to the expected response rate in the subtype and the subtype prevalence a required precision is also specified. The required sample size for the values of θ_r and P_S mentioned above in scenario 3, sample size has been constructed for scenario 4 using P_S =0.2 and the same θ_r as in scenario: 3, for 5% and 10% precisions and 5% and 10% rejection error were calculated.

These values are recalculated using the technique provided by Gehan (1961, Table 2) using $p_r p_S$ instead of p_r (unlike in scenario 1), form the equation 2.1. That is,

$$p_r p_S = \theta_r P_S$$

The SAS program (3.4) provided below for this scenario finds sample size for any response rate in the subtype, for any subtype prevalence including any required precision and with any rejection errors.

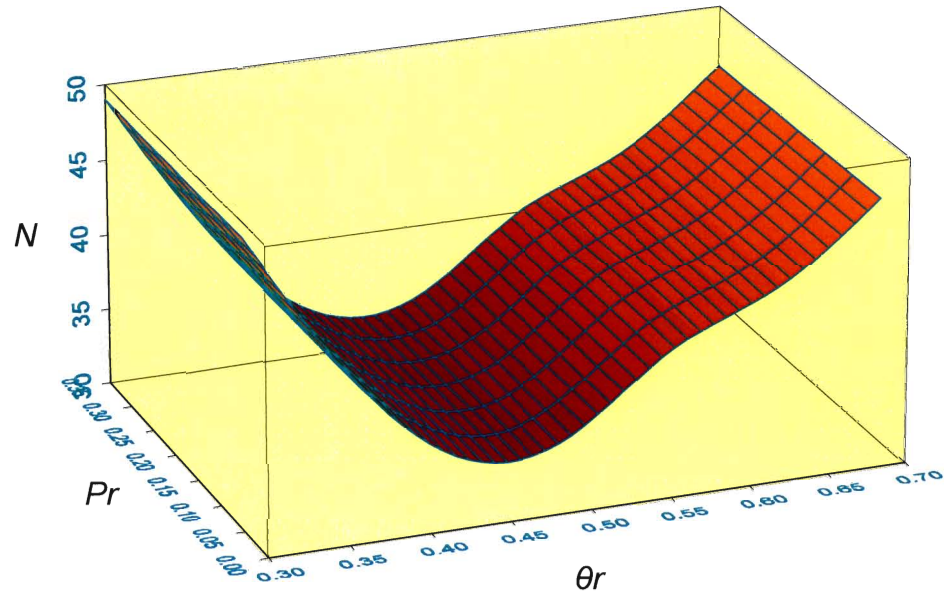
In scenario 3 if there are no responses (that is, at least one response is not observed) in the subtype, the trial will be terminated. This is true also in scenario 4. That is, if no responses in the number of trials mentioned under scenario 3 are observed the trial will terminate and will not accrue the additional patients required under scenario 4. Therefore, whether one is in scenario 3 or 4 the sample sizes needed under scenario 3 is essential. Therefore the sample sizes for scenario 3 and 4 are provided together in Table 4 below.

Table 6: Sample size for scenario 3 and 4 for P_S =0.2

θ_r	Scenario 3	Scenario 4	Scenario 4
		(5% Precision)	(10% Precision)
Rejection Error=5%			
0.30	49	49	49
0.35	41	41	42
0.40	36	36	36
0.45	32	32	32
0.50	29	33	29
0.55	26	37	26
0.60	24	39	24
0.65	22	42	22
0.70	20	46	20
Rejection Error=10%			
0.30	38	38	38
0.35	32	32	32
0.40	28	34	28
0.45	25	38	25
0.50	22	42	22
0.55	20	46	20
0.60	19	48	19
0.65	17	52	17
0.70	16	55	16

For example, if the expected response rate is 50% in the subtype and the rejection error is 5% the trial will stop if no responses are observed in the first 29 patients. If the trial proceeds and one also requires 5% precision the trial will require a total of 33 patients. That is, it will require an additional 4 patients.

Figure 2: Sample size as a function of θ_r and p_r ; P_S =0.2 for Scenario 4



The figure above indicates that the required sample size increases linearly for subtype response rate (θ_r) larger than 45% while it is quadratic if the 35% case is included. This is because the standard errors of 35%, 45% and 55% at 5% rejection error for example are 0.0005, 0.0009 and 0.001. So, the sample size for the specified response rate in the subtype is proportional to the standard error. The sample size remains constant for different values of p_r , which makes sense since the calculation of sample size involve only $p_r p_S$ as $p_r p_S = \theta_r P_S$ in the scenario 3 and 4.

for different values of p_r , which makes sense since the calculation of sample size involve only $p_r p_S$ as $p_r p_S = \theta_r P_S$ in the scenario 3 and 4.

Program 3.3: SAS codes for scenario 3

```
%macro table3(Theta_r,prev,re);
data x;
Prevalence=&prev;
Theta_r=&Theta_r;
Re=&re;
p = &Theta_r*&prev;
do i = 1 to 100 until (flag);
n = (1-p)**i;
if n < &re then do;
flag=1;
N1 = i;
output;
end;
end;
run;
proc print data=x noobs;
var Theta_r Prevalence Re N1;
title ' The Required Sample Size for Scenario 3';
run;
%mend;
%table3(0.35,.2,.05);run;
```

In the above program, if one specifies the response rate in the total population (p_r), rejection error and the subtype prevalence, the program will provide the required sample size. The output is provided below as an example, where p_r =0.35 and rejection error 95% with subtype prevalence=0.2. The sample size comes out to be 42 for scenario 3 from the above program.

Table 7: Sample size for scenario 3

The Required Sample Size for Scenario 3

Theta_r	Prevalence	Re	N1
0.35	0.2	0.05	42

Program 3.4: SAS codes for scenario 4

```
%macro table4(thetar,prev,re,os,prec);
data x;
Theta_r=&thetar;
Prevalence=&prev;
Re=&re;
Observed_Success=&os;
Precision=&prec;
p = &thetar*&prev;
do i = 1 to 100 until (flag);
n = (1-p)**i;
if n < &re then do;
flag=1;
N1 = i;
output;
end;
end;
      p1 = &os/N1;
            se1 = sqrt(p1*(1-p1)/n1);
            phat1 = p1 + 1.64*se1;
            N2 = int((phat1*(1-phat1)/&prec**2))+1;
            if N2<N1 then N2=N1;
output;
run;
```

```
data x;
set x;
if N2 = . then delete;
run;
proc print data=x noobs;
var Theta_r Prevalence Re Observed_Success Precision N1 N2;
title ' The Required Sample Size for Scenario 4';
run;
%mend;
%table4(.3,.2,.05,1,.05);run;
```

In the above program, if one specifies the rejection error, observed success, prevalence, precision and θ_r, the program will provide the required sample size. The output is provided below as an example, where θ_r=0.3, rejection error 95% with 1 observed success, subtype prevalence=0.2 and 10% precision. The required sample size comes out to be 42 for scenario 3 and no more patient is needed for scenario 4 (if the study proceed further. That is, if at-least 1 success is observed in scenario 3) from the above program.

Table 8: Sample size for scenario 4

The Required Sample Size for Scenario 4

Theta_r	Prevalence	Re	Observed_Success	Precision	N1	N2
0.3	0.2	0.05	1	0.05	49	49

3.3 Sample size for scenario 5

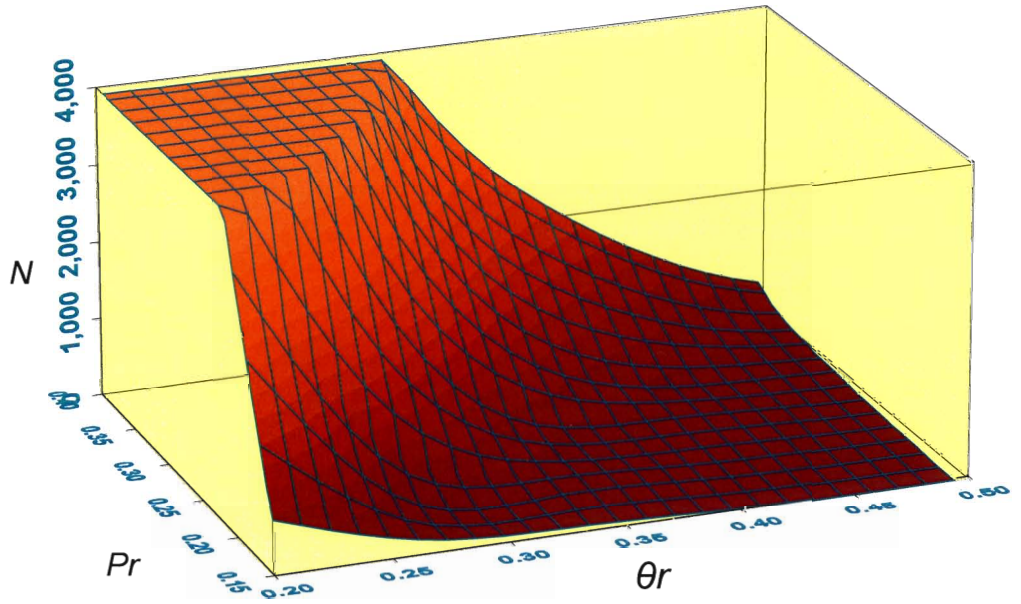
In this scenario, subtype prevalence (P_S), response rate in the subtype (θ_r) as well as in the total population (p_r) are specified. The number of patients responding a treatment (n_r), total sample size (N) and number of patients with subtype among responders (r) for the typical cases when P_S =.2, θ_r =0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and p_r =0.1, 0.15, 0.2 with α_r =5% and β=80% are calculated using the method defined in this paper in chapter 2. To obtain this for any subtype prevalence and any response rate in the subtype as well as in the total population for any significance level and any power, a SAS program is provided.

Table 9: Sample size for scenario 5; for P_S =0.2

P_S =.2

p_r	θ_r	r	n_r	N
0.1	0.2	11	36	469
0.1	0.3	4	10	166
0.1	0.4	2	4	88
0.15	0.2	61	254	1864
0.15	0.3	11	36	310
0.15	0.4	7	18	173
0.15	0.5	3	7	84
0.15	0.6	2	4	58
0.2	0.3	30	116	666
0.2	0.4	11	36	231
0.2	0.5	6	17	123
0.2	0.6	4	10	81

Figure 3: Sample size as a function of θ_r and p_r ; P_S =0.2 for scenario 5



The figure indicates that as θ_r decreases the sample size increases exponentially. Also, as p_r increases the sample size increases.

Program: 3.5 SAS codes for scenario 5

```
%macro table3(thetar,prev,pr,alpha,power);
data a;
Theta_r=&thetar;
```

Program: 3.5 SAS codes for scenario 5

```
%macro table3(thetar,prev,pr,alpha,power);
data a;
Theta_r=&thetar;
Prevalence=&prev;
Pr=&pr;
Alpha=&alpha;
Beta=&power;
l=100;
ps=&prev;
/*do thetar=.2 to .9 by .05;
do pr=.1 to .5 by .05;*/
del1=(&thetar-&pr);
del2=del1*ps/&pr;
thetas=ps+del2;

do r=0 to l;
N1=14+r*10;
do nr=r to N1;
b1=1-cdf('binom',r,ps,nr);
b2=1-cdf('binom',r,thetas,nr);
retain;
output;
end;
end;
/*end;
end;
*/
run;
data b;
set a;
tol1=.04;
tol2=.04;
dif1=abs(b1-&alpha);
dif2=abs(b2-&power);
difsum=dif1+dif2;
if dif1 > tol1 or dif2 > tol2 then delete;
if difsum > 0.1 then delete;
if b1 < .01 or b2 <&power then delete;
if b1 > &alpha then delete;
data c;
set b;
do N=nr to 3000;
b3=cdf('binom',nr,&pr,N);
output;
retain;
end;
data d;
set c;
p=&pr;
theta=&thetar;
```

```
tol3=.04;
dif3=abs(b3-&alpha);
if dif3 > tol3 then delete;
if b3 < &alpha then delete;
run;
proc sort data=d out=ds;by p theta descending b1 b2 b3 ;run;
data e;
set ds;
by p theta;
fb3=first.p;
fb1=first.theta;
run;
data f;
set e;
if fb3 or fb1;
proc print data=f noobs;
var Theta_r Prevalence Pr Alpha Beta r nr N ;
title' The Sample Size for Scenario 5';run;*/
run;
%mend;
%table3(.3,.2,.2,.05,.8);run;
```

In the above program, if one specifies θ_r , p_r , α , β and subtype prevalence, the program will provide the number of patients responding treatment (n_r) and the required total sample size (*N*). The output is provided below as an example, where θ_r =0.3, p_r =0.2, α =95%, β =80% and subtype prevalence=0.2. n_r and *N* come out to be 116 and 666 respectively from the above program.

Table 10: Sample size for scenario 5

The Sample Size for Scenario 5

Theta_r	Prevalence	Pr	Alpha	Beta	r	nr	N
0.3	0.2	0.2	0.05	0.8	30	116	666

The more interactive and user friendly SAS program for scenario 1 and 3 are provided in the Appendix A using the same technique prescribed in this chapter. This program will give the same result as the SAS program provided in this chapter.

3.4 Examples of typical cases on 5 scenarios

Below is the table indicating the required sample size for the specific cases for all five scenarios.

Table 11: Examples on 5 scenarios

	I	II	III	IV	V	VI	VII	VIII
P_S	0.1	0.1	0.1	0.2	0	0	0.3	0.1
θ_r	0.5	0.5	0.5	0.5	0	0	0.6	0.6
p_r	0.05	0.1	0.25	0.1	0.05	0.1	0.3	0.3
Scenario1	59	59	59	29	†	†	16	49
Scenario2	59	29	11	29	59	29	9	9
Scenario3	59	59	59	33	†	†	55	49
Scenario4	59	29	71	33	59	29	79	79
Scenario5	†	166	372	†	59	28	80	†

Here, † represents the cases where sample sizes could not be determined.

CHAPTER 4 Extension as future work

All our calculations in scenarios 1 to 4 are based on a Phase II design proposed by Gehan. Recall that when the precision is specified the sample size calculations are based on a normal approximation. This was improved by Simon (1988), who suggests two-stage designs in which designs are optimal in the sense that the expected size is minimized if the regimen has low activity subject to constraints upon the size of type I and type II errors. This design has been explained in detail in chapter 1. The scenarios 2,3 and 5 could also be extended to the Simon two stage optimal design. This would follow the same arguments as in the case described in chapter 2. However the scenarios have to be reformed using the conditions set by the optimal design.

All the calculations in scenarios 1, 3 and 5 are based on the assumption that the tumors tested for the subtype will find the subtype 100% of the time. In reality diagnostic tests are not 100% accurate. Implementing information such as sensitivity and specificity of the tests in the sample size calculation may be necessary.

The prevalence of the subtype is assumed to be known, which is estimated from the sample. To estimate this with a certain precision in addition to treatment response is also of interest.

Study designs including multiple subtypes such as the melanoma example with mutations on *KIT* are also of interest. The calculations and program need to be repeated for various scenarios. These designs may be achieved by simply correcting for the overall rejection error or by considering more complicated designs such as sequential designs. Calculations and programs can be repeated for various scenarios.

List of References

List of References

John A. Curtin, Klausn Busam, Daniel Pinkel, and Boris C. Bastian (2006), "Somatic Activation Of KIT In Distinct Subtypes Of Melanoma," *Journal of Clinical Oncology*, volume 24, Number 26

Gehan EA (1961), " The Determination Of The Number Of Patients Required In A Preliminary And A Follow-Up Trial Of A New Chemotherapeutic Agent," J Chron Dis 13:346-353,

Richard Simon, Ph.D., (1988), "Optimal Two-Stage Designs For Phase II Clinical Trials," From Biometric Research Branch, National Cancer Institute, Bethesda, Maryland

Boris Freidlin and Richard Simon (2005), "Adaptive Signature Design: An Adaptive Clinical Trial Design For Generating Prospectively Testing Patients," Clin Cancer Res 11(21) November 1, 2005

Richard Simon and Aboubakar Maitournam (2004), " Evaluating the efficiency of Targeted Designs for Randomized Clinical Trials", Clin Cancer Res; 10, 6759-6763, October 15, 2004

Thomas J. Lynch et al. (2004), "Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib", The NEJ of Medicine, vol.350 No.21, May 20, 2004

Steven Piantadosi, "Clinical Trials – A Methodologic Perspective"

Lajos Pusztai et al., "Discovery of pharmacogenomic predictors in phase II clinical trails: limitations and alternatives", AACR Annual Meeting, 2007

Fleming TR: "One sample multiple testing procedure for phase II clinical trials". Biometrics 38:143-151, 1982

O'Brien PC, Fleming TR: "A multiple testing procedure for clinical trials" Biometrics 35:549-555, 1979

APPENDIX A

```
********** Interactive codes for all Scenarios **********;

%window ssize
   rows=25 irow=5 color=gray
group=scenario
      #1 @5 'Specify which scenario you want sample size for' +2 scenario
1
      #2 @5 ' '
      #3 @5 '-------------------------------------------------------'
group=askforit1
     #4 @5 'This program calculates the sample size for scenario 1.'
     #5 @5 'You need to provide p_r and rejection error(re)'
       #6 @5 'Execute the program by pressing the ENTER key.'
     #7 @5 'To end this program, press HOME and enter STOP on the command
line.'
       #8 @5 '-------------------------------------------------------'
     #9 @2 'Enter the response rate p_r' +2 pr 3
     #10 @2 'Enter the rejection error (re)' +2 re 3
group=askforit2
     #6 @5 'This program calculates the sample size for scenario 2.'
     #7 @5 'You need to provide p_r, rejection error(re), observed
success and precision'
       #8 @5 'Execute the program by pressing the ENTER key.'
     #9 @5 'To end this program, press HOME and enter STOP on the command
line.'
       #10 @5 '-------------------------------------------------------'
     #11 @2 'Enter the response rate p_r' +2 pr 3
     #12 @2 'Enter the rejection error (re)' +2 re 3
       #13 @2 'Enter the precision' +2 prec 3
     #14 @2 'Enter the observed number of success' +2 os 3
group=askforit3
     #6 @5 'This program calculates the sample size for scenario 3.'
     #7 @5 'You need to provide Theta_r, Prevalence(P) and rejection
error(re)'
       #8 @5 'Execute the program by pressing the ENTER key.'
     #9 @5 'To end this program, press HOME and enter STOP on the command
line.'
 #10 @5 '-------------------------------------------------------'
     #11 @2 'Enter the response rate among subtype (Theta_r)' +2 thetar 3
     #12 @2 'Enter the Prevalence of the sugtype (P)' +2 prev 3
   #13 @2 'Enter the rejection error (re)' +2 re 3
group=askforit4
     #6 @5 'This program calculates the sample size for scenario 4.'
```

```
     #7 @5 'You need to provide Theta_r, Prevalence(P) and rejection
error(re)'
       #8 @5 'Observed number of successes and precision'
       #9 @5 'Execute the program by pressing the ENTER key.'
     #10 @5 'To end this program, press HOME and enter STOP on the
command line.'
 #11 @5 '-------------------------------------------------------'
     #12 @2 'Enter the response rate among subtype (Theta_r)' +2 thetar 3
     #13 @2 'Enter the Prevalence of the sugtype (P)' +2 prev 3
   #14 @2 'Enter the rejection error (re)' +2 re 3
 #15 @2 'Enter the precision' +2 prec 3
     #16 @2 'Enter the observed number of success' +2 os 3
group=askforit5
     #6 @5 'This program calculates the sample size for scenario 5.'
     #7 @5 'You need to provide Theta_r, Prevalence(P), response rate,
alpha and power'
       #8 @5 'Observed number of successes and precision'
       #9 @5 'Execute the program by pressing the ENTER key.'
     #10 @5 'To end this program, press HOME and enter STOP on the
command line.'
 #11 @5 '-------------------------------------------------------'
     #12 @2 'Enter the response rate among subtype (Theta_r)' +2 thetar 3
     #13 @2 'Enter the Prevalence of the sugtype (P)' +2 prev 3
  #14 @2 'Enter the response rate (re)' +2 pr 3
 #15 @2 'Enter the significance level' +2 alpha 3
     #16 @2 'Enter the power' +2 power 3
     group=showtable1
     #20 @2 "You need at least 1 success out of "   n1 +1  "to proceed"
       #21 @10 "You may change the values and try again..."
 group=showtable4
     #20 @2 "You need at least 1 success out of "   n1 +1  "to proceed"
     #21 @2 "If you observed" os1 +1 "successes you will need" n2 +1
"total sample"
       #22 @10 "You may change the values and try again..."
  group=showtable2
     #20 @2 "You need at least"  r1 +1 "success out of "   nr1 +1
"responses in" n1 +1  "patients"
       #21 @10 "You may change the values and try again..."
       group=schoscen
#23 @10 "You may change the scenario..."
     group=byebye
     #24 @2  " ";
  ;
%macro table3;*(thetar,prev,re);%do %while(%upcase(&syscmd) ne STOP);

%askhere:
   %display ssize.askforit3;
   %if &thetar<1 %then %do;
   data x;
p = &thetar*&prev;
do i = 1 to 100 until (flag);
n = (1-p)**i;
if n < &re then do;
```

```
flag=1;
n1 = put(i,3.0);
call symput("n1",n1);
output;
end;
end;
run;
%end;
%display ssize.showtable1 noinput;
%display ssize.schoscen;
%choscen;
%end;
%display ssize.byebye blank noinput;

%mend table3;
%macro table1;
%do %while(%upcase(&syscmd) ne STOP);
%askhere:
   %display ssize.askforit1;
  data x;
p = &pr;
do i = 1 to 100 until (flag);
n = (1-p)**i;
if n < &re then do;
flag=1;
n1 = put(i,3.0);
call symput("n1",n1);
output;
end;
end;
run;
%display ssize.showtable1 noinput;
%display ssize.schoscen;
%choscen;
%end;%display ssize.byebye blank noinput;

%mend table1;
%macro table2;
%do %while(%upcase(&syscmd) ne STOP);
%askhere:
   %display ssize.askforit2;
data x;
p = &pr;
do i = 1 to 100 until (flag);
n = (1-p)**i;    * chance of consecutive treatment failure by number of
patients;
if n < &re then do;
flag=1;
n1 = put(i,3.0);
call symput("n1",n1);
output;
end;
end;
```

```
      p1 = &os/n1;
            se1 = sqrt(p1*(1-p1)/n1);
            phat1 = p1 + 1.64*se1;
            new_se1 = (phat1*(1-phat1)/&prec**2);
            if new_se1<n1 then new_se1=n1;
            output;
run;
data b;
set x;
os1 = put(&os,3.0);
call symput("os1",os1);
n2 = put(new_se1,3.0);
call symput("n2",n2);
run;
%display ssize.showtable4 noinput;
%display ssize.schoscen;
%choscen;
%end;%display ssize.byebye blank noinput;

%mend table2;
%macro table4;
%do %while(%upcase(&syscmd) ne STOP);
%askhere:
   %display ssize.askforit4;
data x;
p = &thetar*&prev;
do i = 1 to 100 until (flag);
n = (1-p)**i;    * chance of consecutive treatment failure by number of
patients;
if n < &re then do;
flag=1;
n1 = put(i,3.0);
call symput("n1",n1);
output;
end;
end;
      p1 = &os/n1;
            se1 = sqrt(p1*(1-p1)/n1);
            phat1 = p1 + 1.64*se1;
            new_se1 = (phat1*(1-phat1)/&prec**2);
            if new_se1<n1 then new_se1=n1;
output;
run;

data b;
set x;
os1 = put(&os,3.0);
call symput("os1",os1);
n2 = put(new_se1,3.0);
call symput("n2",n2);
run;
%display ssize.showtable4 noinput;
%display ssize.schoscen;
```

```
%choscen;
%end;%display ssize.byebye blank noinput;
%mend table4;
%macro table5;
%do %while(%upcase(&syscmd) ne STOP);
%askhere:
   %display ssize.askforit5;
data a;
l=100;
ps=&prev;
/*do thetar=.2 to .9 by .05;
do pr=.1 to .5 by .05;*/
del1=(&thetar-&pr);
del2=del1*ps/&pr;
thetas=ps+del2;

do r=0 to l;
N1=14+r*10;
do nr=r to N1;
b1=1-cdf('binom',r,ps,nr);
b2=1-cdf('binom',r,thetas,nr);
retain;
output;
end;
end;
/*end;
end;
*/
run;
data b;
set a;
tol1=.04;
tol2=.04;
dif1=abs(b1-&alpha);
dif2=abs(b2-&power);
difsum=dif1+dif2;
if dif1 > tol1 or dif2 > tol2 then delete;
if difsum > 0.1 then delete;
if b1 < .01 or b2 <&power then delete;
if b1 > &alpha then delete;
data c;
set b;

do N=nr to 3000;
b3=cdf('binom',nr,&pr,N);
output;
retain;
end;
data d;
set c;
p=&pr;
theta=&thetar;
tol3=.04;
```

```
dif3=abs(b3-&alpha);
if dif3 > tol3 then delete;
if b3 < &alpha then delete;
run;
proc sort data=d out=ds;by p theta descending b1 b2 b3 ;run;
data e;
set ds;
by p theta;
fb3=first.p;
fb1=first.theta;
run;
data f;
set e;
if fb3 or fb1;
run;
data k;
set f;
N1 = put(N,3.0);
nr1=put(nr,3.0);
r1=put(r,3.0);
call symput("n1",n1);
call symput("nr1",nr1);
call symput("r1",r1);
run;
%display ssize.showtable2 noinput;
%display ssize.schoscen;
%choscen;
%end;%display ssize.byebye blank noinput;
%mend table5;
%macro choscen;
%do %while(%upcase(&syscmd) ne STOP);
%display ssize.scenario blank;
%if &scenario = 1 %then
%table1;
%else
%if &scenario = 2 %then
%table2;
%else
%if &scenario = 3 %then
%table3;
%else
%if &scenario = 4 %then
%table4;
%else
%table5;
%end;

%mend choscen;
%choscen;
```

VITA

Grishma Sheth was born on December 14, 1980 in Ahmedabad, India. She went to Navrang High School in Ahmedabad, India. She graduated with M.S. in Mathematics from Gujarat University, India in 2003. Grishma will graduate from Virginia Commonwealth University with M.S. in Biostatistics in May, 2007. She has served as a biostatistician at Department of surgery, MCV-VCU.