Phase II trials powered to detect tumor subtypes

John D. Roberts
Massey Cancer Center and the Department of Internal Medicine,
Virginia Commonwealth University

V. Ramakrishnan
Division of Biostatistics and Epidemiology,
Medical University of South Carolina

Running Title: Phase II trials powered to detect tumor subtypes

Acknowledgements: Supported in part by NIH P30 CA016059. The authors wish to acknowledge the encouragement of Dr Daniel Sullivan.

Corresponding Author:

V. Ramakrishnan
Professor, Division of Biostatistics and Epidemiology
Medical University of South Carolina,
135 Cannon St. Suite 303S
Charleston, SC 29425
843-876-1153
843-876-1127 (fax)
www.musc.edu
Abstract

Classical phase II trial designs, including “adaptive” designs, require the prospective characterization of tumors. We propose a two stage phase II design that allows for characterization of tumors and selection of a tumor subtype of interest at the conclusion of stage one. The stage two objective is either a classical estimate of the response rate for either the tumor or a subtype or a formal test of the hypothesis that the response rate for a subtype is greater than the overall response rate. Considering likely scenarios, stage one sample sizes approximately range from 20 to 100 with a usual size of 50. This compares with typical classical stage one sample sizes of 12 to 30. Total sample sizes range from identical to classical designs (tens to scores) to large sizes typical of phase III trials in metastatic disease (hundreds). Our design is more efficient than previous adaptive designs because it allows for the selection of a tumor subtype of interest on the basis of results from stage one. It complements classical phase II and phase III designs, in which different treatments are compared in “similar” patients and tumors, by positioning a treatment as fixed (control) and utilizing tumor subtype as the variable of interest.
Introduction

Classical oncology phase II trial designs provide minimum sample sizes necessary to estimate the activity of a treatment against a tumor type with a certain precision [1][2]. These designs typically involve two stages in which the first stage is an early stopping rule for futility. Whereas these designs consider a specified tumor type as homogenous with regard to the probability of response, for most tumors it is likely that there is incompletely understood, predictable heterogeneity in the probability of response. “Adaptive” modifications of classical phase II trial designs to address this have proposed [3][4][5][6][7]; To date, however, all these designs require prospective tumor characterization, which is cumbersome and limits adaptation. We propose a phase II design that does not require prospective tumor characterization in stage one and therefore is less cumbersome in stage one and more adaptable in stage two.

Methods

Definitions. Consider “tumor type” to refer to the set of tumors conforming to the tumor-related eligibility criteria of a phase II trial. Consider tumor “subtype” to refer to a subset of those tumors identifiable by further characterization. For simplicity, the design is discussed as if a subtype is identifiable by the presence of a biomarker for which testing is error free.

Design. At the outset one must calculate a stage one sample size and create a discrete, prioritized list of candidate biomarkers predictive of response.
During the trial one must bank a tumor sample from each enrolled patient that is amenable to testing in terms of all candidate biomarkers.

**Stage one sample size calculation.** In classical phase II trial designs the parameter of interest is the response rate, or probability of response, $p_R$ (Table 1); in order to formulate the early stopping rule, one must specify a minimum overall response rate of interest. Once one presumes the existence of a subtype, in order to calculate the stage one sample size one must specify something about the subtype in terms of prevalence and response rate. One approach is to specify a minimum subtype prevalence of interest as defined by the presence of a biomarker, $p_M$, and a minimum subtype response rate of interest, $p_{R|M}$, which is the conditional probability of response within the subset of biomarker positive tumors. Another approach is to specify a minimum subtype prevalence within responding tumors of interest, $p_{M|R}$, which is the conditional probability of the subtype within the subset of responding tumors, and a minimum response rate of interest, $p_R$. We consider the former approach to be more intuitive. In either case, the product of prevalence and conditional probability yields the probability that a tumor is both a responder and biomarker positive, $p_{M,R}$, a joint probability:

$$p_{M,R} = p_{R,M} = p_M \times p_{R|M} = p_R \times p_{R|M}$$  \hspace{1cm} (1)$$

Of course, a third approach is to specify directly the minimum prevalence of tumors that are both responders and biomarker positive of interest.

In order to calculate the stage one sample size, let $N$ denote the number of subjects with tumors evaluable for response. Each subject has a tumor that either is a responding tumor or a non-responding tumor. Let $R$ denote the number
of responding tumors. Each tumor is either biomarker positive or biomarker negative. But we intend to characterize only responding tumors. Let \( J \) be the number of biomarker positive, responding tumors. Now, the entire set of subjects can be allocated among three mutually exclusive categories: subjects with biomarker positive, responding tumors; subjects with biomarker negative, responding tumors; and subjects with non-responding tumors. The numbers of subjects in each of these three categories are \( J \), \( R - J \), and \( N - R \), respectively, and correspond to a multinomial distribution with cell probabilities \( p_{M,R} \), \( p_R - p_{M,R} \), and \((1 - p_R)\). The proposed design is a straightforward extension of Gehan’s design using this multinomial model [1].

As in classical designs, stage one is an early stopping rule for futility. Define \( n_J \) as the minimum number of biomarker positive, responding tumors required to proceed to stage two and \( n_R \) as the minimum number of responses required to proceed to stage two. For a specified false negative rate, \( \beta \), the stage one sample size will be determined by requiring the joint probability of the event \( J < n_J \) and the event \( R < n_R \) to be smaller than \( \beta \). Under the multinomial model this translates to numerically solving the equation

\[
P(J < n_J, R < n_R \mid N) = \sum_{i=0}^{n_J} \sum_{j=0}^{\min(r,n_J)} \frac{N!}{i!(r-i)!(N-r)!} (p_{M,R})^i (p_R - p_{M,R})^{r-i} (1 - p_R)^{N-r} < \beta \tag{2}
\]

for \( N \), \( n_R \) and \( n_J \). For a fixed \( n_J \) several combinations of \( n_R \) and \( N \) may satisfy the above inequality. The recommended combination is the case that leads to the smallest \( N \) such that \( n_R \) is the minimum needed to observe at least \( n_J \), where \( n_J \geq 1 \).
List of candidate biomarkers. A list of candidate biomarkers predictive of response is created from any available knowledge. In general a candidate biomarker might be any feature that is differentially expressed within a tumor type. Such a biomarker might or might not have known prognostic or predictive significance. For an agent “targeted” to, for example, a specific mutation, the presence of that mutation would be an obvious candidate.

Analysis of stage one. Count the number of responding tumors ($R$). Test responding tumors for as many candidate biomarkers as is feasible or desired. Count the number of responding tumors ($J$) for each candidate biomarker. Compare numbers of responding ($R$) and biomarker positive, responding tumors ($J$) with minimum numbers needed to proceed to stage two ($n_R, n_J$) with four possible outcomes (Figure 1): (1) the overall response rate is not potentially of interest, and the prevalence of biomarker positive, responding tumors for any candidate biomarker is not potentially of interest ($R < n_R$ and $J < n_J$); (2) the overall response rate is not promising but there is a promising biomarker(s) ($R < n_R$ and $J > n_J$); (3) the overall response rate is promising but there is no promising biomarker ($R \geq n_R$ and $J < n_J$); and (4) the overall response rate is promising and there is a promising biomarker(s) ($R \geq n_R$ and $J \geq n_J$). In scenarios (1) and (3), one has demonstrated that none of the candidate biomarkers is of interest. In scenarios (2), (3), and (4) one has demonstrated that the treatment is potentially of interest, either when matched to a promising subtype (2) or in general (3), or both (4), respectively. For the purposes of simplicity, we consider only the case in which a single candidate marker is identified as promising.
**Design of stage two.** If the overall response rate is not promising and there is no promising biomarker, then the study has been completed. It might be appropriate to archive responding tumors for future testing should new candidate biomarkers become known. Indeed, the observation of a number of responding tumors that expressed none of the candidate biomarkers might prompt a search for biological similarities among them. But classical statistical considerations would not allow one to attach any significance to a "promising" biomarker identified in such a manner.

If the overall response rate is promising but there is no promising biomarker, then the situation is analogous to a promising stage one outcome of a classical phase II trial, and a stage two sample size can be calculated according to one of these designs. Similarly, if the overall response rate is not promising but there is a promising biomarker, a stage two sample size can be calculated according to a classical design but powered to estimate the prevalence of tumors that are both biomarker positive and responding. An alternative would be to proceed to a new trial limited to tumors prospectively characterized as promising biomarker positive. This is analogous to the approach of previously proposed adaptive designs.

If both the overall response rate is promising and there is a promising biomarker, then it is recommended that stage two be formulated as a test of the hypothesis that the response rate among biomarker positive tumors ($p_{R|M}$) is significantly greater than the overall response rate ($p_R$). This approach requires
an estimate of the prevalence of the promising biomarker in the overall tumor population, $p_M$.

One might come by an estimate of the probability of biomarker positivity in the overall tumor population in one of three ways: (1) Characterize the non-responding tumors; this resource intensive approach yields the most precise unbiased estimate possible and, therefore, the greatest power. (2) Characterize a sample of the non-responding tumors; this yields a less precise unbiased estimate and lesser power. (3) Characterize an unrelated sample of “similar” tumors; this might allow one to use existing data or to increase the precision of the estimate, but any such estimate would be subject to whatever bias was introduced by the different methods of selection of the different tumor samples.

In order to design a hypothesis-testing stage two, the investigator also must provide a minimum difference of interest between the response rate among biomarker positive tumors and the overall response rate. The number of responses needed to detect this difference is calculated using the marginal binomial distribution of $J$, the number of biomarker positive, responding tumors. Then, for a specified significance level $\alpha$ and power $(1 - \beta)$, using the marginal binomial $[(N, p_{M,R})]$ distribution of $J$, exact one-sample methods [8] can be used to determine the required sample size. Examples were calculated using PROC POWER in SAS® with the TEST=EXACT option for testing the null and alternate hypotheses stated in terms of the minimum biomarker prevalence at 5% significance level and 80% power.
**The issue of multiple testing.** As we consider only cases in which a single biomarker is designated as promising, no statistical adjustments for multiple testing are required in stage two. As in classical designs, stage one is an early stopping rule for futility. The efficiency of this stopping rule is a function of the scope of the candidate markers. For example, if virtually all tumors of the tumor type mark positive for at least one candidate biomarker, then, if there are any responses in stage one, it is a virtual certainty that the trial will proceed to stage two.

**Results**

Stage one sample sizes are largely an inverse function of the minimum biomarker positive, responding tumor prevalence of interest with little dependence upon the minimum overall response rate of interest (Figure 2). We think it would be unusual for a stage one sample size to be less than 20, which is, for example, the number necessary to exclude a biomarker with a response rate of 50% and prevalence less than 34%. For a common tumor a stage one sample size of approximately 50 might be common. Such a stage one sample size would routinely detect a biomarker with a prevalence of 10% and a response rate of 50%.

Stage two, incremental sample sizes \((N^+)\) vary considerably according to the goal of stage two. In a situation in which the overall response rate is promising but there is no promising biomarker \((R \geq n_R \text{ and } J < n_j)\), the stage two, incremental sample size would be less and the total \((N + N^+)\) sample size would
be identical to classical phase II designs, that is, tens to scores of patients. In a situation in which the overall response rate is not promising but there is a promising biomarker \((R < n_R \text{ and } J \geq n_J)\), the stage two, incremental sample size and the total sample size would be larger than for classical phase II designs. We anticipate that such trials might involve one hundred patients or more. Such a trial would expose additional patients to a treatment that is not likely to be effective for most (see Discussion). In a situation in which the overall response rate is promising and there is a promising biomarker \((R \geq n_R \text{ and } J \geq n_J)\), the stage two, incremental sample size would be much larger than for classical phase II designs and the total sample size would approach that of classical phase III trials for metastatic disease, that is, hundreds of patients (Figure 3).

Discussion

**Comparison with previous designs.** The problem of response heterogeneity has been approached with adaptive or flexible, including Bayesian, phase II designs \([4][5][6][7]\). All of these designs incorporate inefficiencies that render them unpopular. All require stratified patient enrollment based upon prospective characterization of tumor biomarkers. This is cumbersome for a variety of reasons: (1) Prospective testing frequently causes treatment delays of days or weeks that are frustrating for both patients and clinicians and may discourage enrollment. (2) Although in most cases batch processing of tumor samples would be more efficient and more accurate, usually tumors are tested one at a time in order to avoid longer treatment delays. (3) Although experience
with a new candidate biomarker often is limited to a research laboratory, if treatment decisions (for example, whether or not a patient is eligible) are based upon testing results, then testing must be performed within a CLIA certified laboratory. (4) In the case of multiple candidate biomarkers, it is cumbersome to enroll sufficient numbers of patients into each stratum. Our design, which does not require prospective tumor testing, addresses all of these problems. In addition, in the case of negative studies, our design uses fewer resources, as initial testing to identify a promising biomarker is performed only on responding tumors; testing of non-responding tumors is necessary only if a promising biomarker is tentatively identified.

Our design fills an unmet need in the spectrum of oncology clinical trials. Currently, tumor subtypes are matched with treatments largely on the basis of preclinical findings and retrospective analysis of phase III trials. Experience to date indicates that preclinical findings are poorly predictive of tumor response type (see Historical examples). Retrospective analyses of phase III trials also are a poor platform for accomplishing this objective. Phase III trial sample sizes may be unnecessarily large or too small to answer clinically relevant questions concerning tumor subtypes. Phase III trials typically do not include prospective tumor banking; or bank tumor samples from only a subset of patients enrolled, which only begs the question of appropriate sample size; or include optional tumor banking, which renders sample size unknown and may introduce unknown biases. Phase III trials generally are limited to treatments that have shown efficacy in phase II; a subtype-specific treatment that is never appropriately
matched to its subtype in phase II is unlikely to be studied in phase III and therefore unlikely to be appropriately matched in any phase. Finally, once the sample sizes are estimated using the proposed method, it is important to consider the corresponding operating characteristics. Although we have not discussed this in this manuscript, we are exploring this through simulations and we will publish these subsequently.

**Historical examples.** Consider how our design could be both more effective and more efficient than current approaches:

**Example 1:** CALGB 500104 was a classical phase II trial of the farnesyltransferase inhibitor, tipifarnib, in cutaneous melanoma [NCT00060125 at ClinicalTrialsFeeds.org updated March 17, 2010]. The trial demonstrated with sufficient confidence that tipifarnib is not active in the tumor type cutaneous melanoma when no responses were observed in 14 patients. At the time the trial was initiated, however, it was known that approximately 15% of cutaneous melanomas harbor activating mutations in N-Ras, a farnesylated signaling protein [9]. It would have been quite rational, then, to propose N-Ras mutation as defining a responding subtype. If mutated N-Ras is associated with a response rate of approximately 50%, the false negative rate for this classically designed trial – that is, the probability that tipifarnib is active in mutated N-Ras cutaneous melanoma – is approximately 33%. If interest in tipifarnib in melanoma was stimulated by knowledge of N-Ras mutations, then this study was too small to effectively rule out tipifarnib activity in a relevant subtype of melanoma.
Example 2: In 2004 it was reported that within non-small cell lung cancer there is a subtype defined by a family of activating mutations of the tyrosine kinase domain of the epidermal growth factor receptor (EGFR) that exhibits unusual responsiveness to the EGFR inhibitor gefitinib [10]. The circumstances that led to this discovery were that despite an overall response rate of only 8%, which generally would signify an “inactive” drug, the investigators treated the unusually large number of 275 patients. This allowed them to identify 25 patients with responses, among whom they were fortunate enough to have archived tumor specimens from 9 patients. EGFR gene mutation analysis revealed tyrosine kinase domain mutations in 8 of these 9 tumors. Analysis of samples of non-gefitinib-treated non-small cell tumors and non-small cell tumor cell lines suggested that the frequency of these mutations in the overall tumor population was 8% or less. From this they concluded that activating tyrosine kinase domain EGFR mutation was a likely biomarker for gefitinib responsiveness. The two unusual circumstances that facilitated this discovery – treatment of an unusually large number of patients and availability of archived tumor samples for retrospective analysis – are explicit features of our proposed design. But the approach taken by the investigators was inefficient. Assuming the observed parameters, namely an overall response rate of 9%, a subtype prevalence of 8%, a subtype response rate of 89%, and accepting a 5% false positive rate, discovery of the EGFR mutation would have required a stage one minimum sample size of 36 patients. The expected number of responses would have been two, one of which would have had the EGFR mutation. This would correspond to
our scenario (2) in which the overall response rate is not promising but there is a
promising biomarker; in our notation, \( R < n_R \) and \( J \geq n_J \). A non-hypothesis testing
stage two would estimate the prevalence of biomarker positive, responding
tumors \( p_{M,R} \). Although in theory one might object that pursuing stage two in the
face of a less than promising overall response rate is medically or ethically
suspect, this is an example in which investigators were willing to do this. In order
to complete stage two an additional 62 patients \( N^+ \), for a total of 98 patients
\( (N^+ + N^+) \), would be required. This is about one third the number of patients
treated in the absence of our trial design.

Limitations and issues for future research. As our design relies upon
testing of a single tumor specimen, it incorporates the simplifying assumptions
that tumor testing is error free and individual patient's tumors are biologically
homogeneous. Biological tumor heterogeneity both within a tumor mass and
between and among a tumor primary and metastases is a well-described,
potentially complex phenomenon and a problem for all study designs including
ours. We anticipate that an appropriate adjustment for less than perfect test
sensitivity would be to increase the sample size. If the sensitivity is known, this
adjustment could be calculated; if unknown, the adjustment could be arbitrary.
Less than perfect test specificity might result in false positive trial results. This
might lead to further, futile investigations, but eventually the facts would become
known.

Example 3: A limitation of our design is that it allows only a single candidate
biomarker to proceed to stage two. A third example identifies the need for
extensions of our trial design to accommodate multiple biomarkers. Current clinical practice guidelines for treatment of metastatic colon cancer include the combination of the EGFR-targeted monoclonal antibody cetuximab with the cytotoxic agent irinotecan [NCCN Clinical Oncology Practice Guideline v2.2011]. This combination is recommended regardless of the presence or absence of expression or overexpression of the putative cetuximab target, EGFR, but only in the absence of activating mutations of the K-Ras oncogene. Neither of these recommendations reflects a result from a prospective, stratified clinical trial. The presumption that cetuximab would be active only in EGFR-expressing tumors was so strong that patients with EGFR non-expressing tumors were excluded from the studies that led to approval [11][12] Retrospective analysis of these studies, however, could demonstrate no relationship between degree of EGFR expression and clinical benefit. The recommendation not to test for or consider results of EGFR testing is a non-evidence-based extrapolation of this analysis to patients with tumors that do not express EGFR. As K-Ras is “downstream” of EGFR, a logical inference would be that cetuximab is inactive in tumors harboring activating K-Ras mutations. This inference has been confirmed and therefore incorporated into guidelines. This confirmation, however, is based exclusively upon retrospective analysis, as prospective typing for K-Ras mutation status was not incorporated into these trials.

Thus, optimal tumor subtyping may require multiple biomarkers, which raises problems of multiple testing and potential increases in sample size. On the other hand, there may be interactions among biomarkers (for example,
biomarkers may be mutually exclusive or nested) that, if recognized, reduce the real number of tests. Future work that incorporates multiple, possibly interacting biomarkers in the trial design is desirable.

Although our presentation and examples focus upon tumor response heterogeneity, genetic or environmentally-determined differences among patients that affect treatments, for example in terms of pharmacokinetics or pharmacodynamics, may contribute to response rates. Our trial design is easily adapted to discover patient biomarkers that predict for response to a particular therapy.

Although originally described with a binary tumor response endpoint, classical designs have been adapted to other endpoints including various time-to-event endpoints such as time-to-progression \[13][14][15][16]. An advantage of tumor response is that, as most tumors do not spontaneously shrink, it uniquely reflects a tumor-treatment interaction; a disadvantage is that tumor shrinkage may signify neither a symptomatic or time-to-event benefit. An advantage of time-to-event endpoints is that they measure something that is almost always relevant to patients; a disadvantage is that they reflect both biomarker “effects” and tumor-treatment interactions. In a single arm study utilizing exclusively a time-to-event endpoint it is impossible to distinguish prognostic effects from tumor-treatment interactions for which a biomarker is predictive. We anticipate that extension of our design to time-to-event endpoints would be feasible through randomization of patients among two or more study arms. Additional study arms could involve observation, a placebo treatment, a “standard” treatment, and/or a
different investigational treatment. An observation or placebo treatment arm would permit an analysis to distinguish prognostic from predictive effects, but might be challenging from a medical or ethical standpoint. A “standard” treatment or alternative investigational treatment might yield clear signals about how to proceed clinically, but differentiation of predictive versus prognostic effects might be confounded by tumor-treatment interactions.

**Our design in context.** In oncology one of the purposes of clinical trials is to match tumors and treatments. The question underlying multi-arm, randomized clinical trials is: given this group of “similar” patients, which treatment is best? Pursuit of this question has led to trials involving increasing numbers of patients in order to detect small treatment benefits. An alternative approach is to ask the question: which patients benefit most from this treatment? Pursuit of this question also will involve increasing numbers of patients. Our design offers a rational approach to placing limits on these increases. In multi-arm trials the control is the patient population; in our design the treatment is controlled and tumor subtype is the variable of interest. As an alternative approach to the problem of matching tumors and treatments, our design complements classical phase II and phase III designs.

When numbers of agents and tumor types were small, empiric exploration of the tumor-treatment matrix was tractable and largely accepted. Increasing understanding of cancer biology and advances in technology have increased both the number of recognizable tumor types and the number of available treatments. These increases have rendered empiric exploration of the tumor-
treatment matrix intractable. As a “targeted” agent emerges from the laboratory to the clinic, the hypothesis implicit in the label “targeted” is that the agent has been appropriately matched to its target. This has encouraged many to consider testing of a targeted agent in non-target-marked tumors as unnecessary and “empiricist”. We take a different view: Testing of a targeted agent in both target and non-target marked tumors, that is, hypothesis testing, is both the essence of the scientific method and necessary given the imperfection of current preclinical models and limitations of retrospective analysis. Our design can be considered either or both an efficient approach to this hypothesis testing and/or a rational solution to the intractability of purely empiric exploration of the tumor-treatment matrix.
Figure 1. Stage One Flow Diagram

Figure 2. Stage one sample size contours \((N)\) as a function of minimum overall response rate \((p_R)\) of interest and minimum prevalence of biomarker positive, responding tumors \((p_{M,R})\) of interest, at a 5% false negative rate. The lines interrupted by numbers represent sample size. Example: When \(p_R \sim 0.3\) and \(p_{M,R} \sim 0.045\) the sample size is 60. Inset: Smoothed relationship between the minimum prevalence of biomarker positive, responding tumors of interest and stage one sample size.

Figure 3. Hypothesis-testing stage two, incremental sample sizes \((N^+)\) (at 5% significance level and 80% power) for the stage one outcome of \(n_J = 1\) and \(R \geq n_R\) at a false negative rate of 5%. The total sample size for a study is the sum of the stage one sample size \((N)\) and the stage two sample size \((N^+)\). The lines interrupted by numbers represent stage two sample size \((N^+)\). Example: When \(p_{M,R} \sim 0.045\) and \((p_{M,R} - p_R) \sim 0.04\) the incremental stage two sample size is 246. If the stage one sample size was 60, then the total sample size needed is 306.
References


Table 1. Summary of Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_R$</td>
<td>Overall response rate</td>
</tr>
<tr>
<td>$p_M$</td>
<td>Prevalence of biomarker positive tumors</td>
</tr>
<tr>
<td>$p_{M</td>
<td>R}$</td>
</tr>
<tr>
<td>$p_{R</td>
<td>M}$</td>
</tr>
<tr>
<td>$p_{M,R}$</td>
<td>Prevalence of tumors that are both responding and biomarker positive</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of subjects/tumors in stage one</td>
</tr>
<tr>
<td>$N^+$</td>
<td>Additional number of subjects/tumors in stage two (stage two, incremental sample size)</td>
</tr>
<tr>
<td>$R$</td>
<td>Number of responding tumors</td>
</tr>
<tr>
<td>$J$</td>
<td>Number of biomarker positive, responding tumors (Joint event)</td>
</tr>
<tr>
<td>$n_R$</td>
<td>Minimum number of responding tumors required to proceed to stage two</td>
</tr>
<tr>
<td>$n_J$</td>
<td>Minimum number of biomarker positive, responding tumors required to proceed to stage two</td>
</tr>
</tbody>
</table>
Specify:
A minimum overall response rate of interest
AND
A list of candidate predictive biomarkers (or subtypes)
AND
A minimum subtype response rate of interest and a minimum subtype prevalence of interest
OR
A minimum prevalence of responding tumors with the subtype of interest

- Treat patients
- Count $R$, the number of responding tumors
- Identify a promising biomarker by subtyping responding tumors
- Count $J$, the number of responding tumors of the subtype of interest

Classical stage two: estimate $p_{M,R}$

Hypothesis-testing stage two: Is $p_{RM} > p_M$?
Minimum prevalence of biomarker positive, responding tumors ($p_{M,R}$)

Minimum overall response rate ($p_R$)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$p_{M,R}$</th>
</tr>
</thead>
</table>
| 0.02 | 0.10  
| 0.04 | 0.12  
| 0.06 | 0.14  
| 0.08 | 0.16  
| 0.10 | 0.18  
| 0.12 | 0.20  
| 0.14 | 0.22  
| 0.15 | 0.24  

Statistics in CCR
Minimum difference of interest between the prevalence of responding tumors among biomarker positive tumors ($p_{M|R}$) and the overall response rate ($p_R$).

Minimum prevalence of biomarker positive responding tumors ($p_{M,R}$) of interest:

- 0.01
- 0.02
- 0.03
- 0.04
- 0.05
- 0.06
- 0.07
- 0.08
- 0.09
- 0.10
- 0.11
- 0.12
- 0.13
- 0.14
- 0.15

© 2011 American Association for Cancer Research.
Clinical Cancer Research

Phase II trials powered to detect tumor subtypes

John D. Roberts and Viswanathan Ramakrishnan

Clin Cancer Res  Published OnlineFirst July 7, 2011.

Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-10-2466

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.