

## Adaptive Clinical Trial Designs for Simultaneous Testing of Matched Diagnostics and Therapeutics

Howard I. Scher<sup>1</sup>, Shelley Fuld Nasso<sup>2</sup>, Eric H. Rubin<sup>3</sup>, and Richard Simon<sup>4</sup>

### Abstract

A critical challenge in the development of new molecularly targeted anticancer drugs is the identification of predictive biomarkers and the concurrent development of diagnostics for these biomarkers. Developing matched diagnostics and therapeutics will require new clinical trial designs and methods of data analysis. The use of adaptive design in phase III trials may offer new opportunities for matched diagnosis and treatment because the size of the trial can allow for subpopulation analysis. We present an adaptive phase III trial design that can identify a suitable target population during the early course of the trial, enabling the efficacy of an experimental therapeutic to be evaluated within the target population as a later part of the same trial. The use of such an adaptive approach to clinical trial design has the potential to greatly improve the field of oncology and facilitate the development of personalized medicine. *Clin Cancer Res*; 17(21); 6634–40. ©2011 AACR.

### Introductory Note

At the 2010 Conference on Clinical Cancer Research, convened by Friends of Cancer Research and the Engelberg Center for Health Care Reform at the Brookings Institution, participants explored 4 pressing challenges in the field. Articles summarizing the panel's recommendations on each of these topics are featured in this issue of *Clinical Cancer Research* (1–4).

### Key Role of Companion Diagnostics in Oncology Drug Development

Nearly all cancer drugs being developed today are designed to inhibit molecular targets that have been identified as being dysregulated in human tumors. Genomics has established that the dysregulated pathways and mutated genes in tumors originating in a particular primary site are highly variable. To optimally evaluate and utilize a targeted approach requires the concurrent development of diagnostics that enable the identification of those tumors that are most likely to be sensitive to the anticancer effects of a particular drug or drug combination. The reality of code-

veloping a matched diagnostic and therapeutic has profound implications for the clinical trial designs used in drug development. Trials of cytotoxic drugs typically enroll unselected patients at a particular point in the continuum of a disease in the hope that the response of tumors that are sensitive to the treatment will be sufficient to show benefit for the population as a whole. Although this approach may lead to broad labeling indications, it also exposes patients with nonsensitive tumors to unnecessary toxicities and increases the possibility of discarding a drug that may dramatically benefit a subset of patients. Consequently, this strategy is not viable for molecularly targeted agents, in which the activity is likely to be restricted and determined more by the genomic alteration(s) within a tumor at the time treatment is being considered than by the primary site in which the tumor originated. The use of anatomically based (i.e., primary site of disease), "all comers" approaches to develop targeted approaches has typically led to failure in phase III studies, or demonstration of "success" based on statistically significant but clinically questionable benefits (5).

Although developing the right drug for a specific patient has great value to the individual and is critical for controlling the costs of health care, it dramatically increases the complexity of the drug development process. For many drugs, the complexities of identifying a predictive biomarker and the practical complexities of developing analytically valid diagnostic tests for the biomarker are grossly underestimated. Knowing when to start the development of the diagnostic is also an issue, particularly when the effectiveness of the drug in any population is uncertain. Developing the right drug for the right subset of patients requires new clinical trial designs and new paradigms of data analysis.

Efforts to codevelop a matched diagnostic and therapeutic face other challenges as well. Even with extensive

**Authors' Affiliations:** <sup>1</sup>Department of Medicine, Genitourinary Oncology Service, Memorial Sloan-Kettering Cancer Center, and Weill Cornell Medical College, New York, New York; <sup>2</sup>Susan G. Komen for the Cure Advocacy Alliance, Washington, District of Columbia; <sup>3</sup>Oncology Clinical Research, Merck Research Laboratories, Whitehouse Station, New Jersey; and <sup>4</sup>Biometric Research Branch, National Cancer Institute, Bethesda, Maryland

**Corresponding Author:** Howard I. Scher, Department of Medicine, Sidney Kimmel Center for Prostate and Urologic Cancers, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY. Phone: 646-422-4323; Fax: 212-988-0851; E-mail: scherh@mskcc.org

doi: 10.1158/1078-0432.CCR-11-1105

©2011 American Association for Cancer Research.

preclinical investigations, it is often difficult to identify a predictive biomarker and evaluate it in the phase I trial of a given drug. As a result, it becomes necessary to have a test for, and develop preliminary data showing, the predictive power of the candidate marker in the context of phase II investigations, so that properly focused phase III trials can be designed, conducted, and completed. In the unusual case in which a single predictive biomarker has been identified and a validated assay has been developed prior to the start of the phase III trial, targeted enrichment designs and stratification designs can be used (6, 7). For example, in the targeted enrichment design used in the development of trastuzumab, only "marker-positive" patients were included. In the stratification design, patients were not excluded on the basis of marker status, but the size of the trial was adequately powered for the anticipated frequency of marker-positive patients and the overall 5% type I error allocated between the comparison of treatments overall and the comparison within marker-positive patients. Adaptive phase II designs, such as the design recently used in the Biomarker-Integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE) clinical trial in non-small cell lung cancer (8) and the Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2 (I-SPY2) trial in breast cancer (9), are useful for identifying the most promising predictive biomarker in phase II development, but they require large sample sizes. The outcome-adaptive randomization weights used in the BATTLE study design complicate the interpretation of results, and the determination of whether they improve efficiency has not been established (10). The BATTLE study did, however, show the feasibility of a biopsy-based, hypothesis-driven biomarker trial, and the follow-up phase II trial, BATTLE 2, will test the predictive value of the gene signatures prospectively. In 2010, the U.S. Food and Drug Administration (FDA) issued a draft Guidance to Industry on Adaptive Design Clinical Trials for Drugs and Biologics (11).

Due to the complexity of cancer biology, it is often not possible to firmly establish the biomarker(s) most likely to predict sensitivity to a particular drug or class of drug by the time pivotal phase III trials are set to begin. Recently, however, several adaptive clinical trial designs have been published that show how to design the trial(s) so that the most suitable target population of patients is adaptively identified during the trial and the effectiveness of the drug is evaluated in that population in a rigorously defined and statistically valid manner (12–14). For example, when the biomarker assay has been validated and standardized, and performance characteristics are known, the adaptive signature design (12) and cross-validated adaptive signature design (14) are carefully crafted adaptive phase III non-Bayesian approaches that preserve the desired type I error rate while identifying an optimal target population. Neither design results in a change in randomization weights or in eligibility criteria (both of which could require statistical adjustments to avoid introduction of bias), which makes them better suited for licensing registration trials than the

Bayesian methods used in the phase II BATTLE trial. Interestingly, although the FDA Guidance on adaptive trial designs acknowledges that a Bayesian framework can be useful for planning purposes to evaluate model assumptions and decision criteria, they recommend that the study design be planned in a framework to control the overall study type I error rate (11).

These are, however, complex designs that have not been tested in practice. Challenges to the use of these designs are that the treatment comparisons can only be conducted after completion of the study, that the developed predictive signature may be based on a combination of factors with unclear biological meaning, and that it may be difficult to interpret the results if there are imbalances in other baseline prognostic factors between treatment arms in the marker-positive subgroup. Although these designs are in some ways conservative, they are nevertheless dramatically different from the kinds of designs used for the vast majority of clinical trials being conducted today.

Here, we describe how adaptive methods can be used for indication determination in a manner that provides the level of confidence in conclusions that we expect from phase III registration trials and in a manner consistent with the FDA Guidance on adaptive design. The current draft of the Guidance defines a clinical study using an adaptive design as one that "includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study. Analyses of the accumulating study data are carried out at prospectively planned time points within the study, can be performed in a fully blinded manner or in an unblinded manner, and can occur with or without formal statistical hypothesis testing" (11). In some cases, adaptive designs require fewer patients but much more upfront planning.

To illustrate the careful planning necessary for proper use of adaptive methods in this context, a detailed illustration of the use of the adaptive signature design of Freidlin and Simon (12) is provided. The approach includes 3 components: (i) a statistically valid identification, based on the first stage of the trial, of the subset of patients who are most likely to benefit from the new agent; (ii) a properly powered test of overall treatment effect at the end of the trial with all randomized patients; and (iii) a test of treatment effect for the subset identified in the first stage but only with patients randomized in the remainder of the trial. The design is adaptive in the sense of the FDA Guidance because the primary plan for the final analysis is influenced by the results of the trial. The adaptive signature design (12) and the more recently published cross-validated adaptive signature design (14) were developed for use in gene expression profiling settings when there are enormous numbers of candidate measurements that can be combined to provide a classifier of which patients are likely (or unlikely) to benefit from a new treatment relative to a control regimen. These designs can be used much more broadly, however, regardless of the candidate predictors, and are discussed in greater generality by Simon (15).

## Background

Application of the original adaptive signature paradigm to a real clinical development setting has many complexities, which will be illustrated here for castration-resistant prostate cancer (CRPC), the advanced, lethal form of the disease. Molecular profiling studies show that reactivation of androgen receptor (AR) function is a consistent feature of CRPC (16), in part, through AR overexpression and overexpression of androgen-synthetic enzymes leading to increased intratumoral androgens (17, 18). The clinical significance of these findings has been validated in trials of abiraterone acetate, an inhibitor of androgen synthesis in the testis, adrenal gland, and tumor (19, 20), and MDV3100, a novel AR antagonist selected for activity in prostate cancer model systems with overexpressed AR (21). Abiraterone was recently shown to confer a survival benefit after chemotherapy in patients with CRPC and is now approved by the FDA for this indication (22). MDV3100 has shown activity comparable with that of abiraterone in postchemotherapy CRPC (23), and a phase III registration trial for this population has been fully accrued. Noteworthy in the trials of both agents was the similarity of response in matched patient populations, which ranged from dramatic prostate-specific antigen declines with durable radiographic control in some to intrinsic resistance in others, suggesting the presence of predictive biomarkers in tumors. A number of other agents targeting different points in the AR signaling pathway are currently in development (24), and although predictive biomarkers of sensitivity have been postulated, none has warranted the development of a validated assay or begun the formal process of clinical qualification (11). As biotechnology continues to provide the tools to characterize tumors at the genomic scale and basepair resolution, it is likely that relevant predictive markers will be identified.

Further adding to the complexity of developing drugs for CRPC is the recent demonstration that 3 additional agents, with different mechanisms of action—Provenge (sipuleucel-T; Dendreon), Jevtana (cabazitaxel; Sanofi), and Alpharadin (radium-223; Bayer)—also confer a survival benefit (25–28). With the expanded armamentarium, it will become increasingly more difficult to show the survival benefit of an agent without enriching for patients most likely to respond. Identification and development of validated assays for predictive biomarkers of sensitivity are likely to play a significant role in the ultimate approval of these and future therapies.

## The Adaptive Signature Approach

The adaptive signature approach provides for a final analysis consisting of 2 parts: first, outcomes for all patients randomized to receive the new drug will be compared with outcomes for all patients randomized to receive the control. If this comparison is significant at a more stringent than usual 2-sided significance level of  $\alpha_0$ , then the new drug is considered broadly effective. Otherwise, a single subset analysis is conducted. The patients in the trial are randomly

partitioned into a training set and a validation set. The training set is used to develop a "classifier" that identifies the subset of patients who seem to benefit from the new treatment compared with the control. This classifier can be based on a combination of all the clinical and biomarker candidate variables measured before treatment. When this single classifier is completely specified using only the training set, it is applied to classify patients in the validation set with respect to whether they are predicted to benefit from the new treatment. Outcomes for patients in this subset of the validation set who were randomized to receive the new treatment are compared with outcomes for patients in this subset who were randomized to receive the control regimen. Only the patients in the validation set are used for this comparison. Because the training set was used to develop the classifier, it cannot be used to evaluate it. If this difference is significant at the reduced 2-sided significance level of  $0.05-\alpha_0$ , then the new treatment is considered effective for the subset of patients defined by the classifier developed in the training set. The cross-validated adaptive signature design is a more statistically powerful version of this approach (14). However, even application of the original adaptive signature paradigm to a real clinical development setting has many complexities, an example of which will be illustrated here for CRPC.

## A Phase III Adaptive Trial Design

The design we describe for the clinical trial is an application of the adaptive signature approach of Freidlin and Simon (12) and could be used with many more candidate predictive markers. This design is appropriate for settings in which (unlike the case of HER2 overexpression and trastuzumab development) there is not yet a single predictive biomarker candidate, in which there is high confidence by the time of initiation of the phase III clinical trial of the drug.

Eligible patients are individuals with progressive CRPC for whom a targeted therapeutic approach is being developed, and for whom tumor material is available. The requirement of sufficient tumor for analysis at entry ensures near-complete ascertainment of the biomarker or biomarker panel. Formalin-fixed, paraffin-embedded (FFPE) samples that were obtained either as part of the routine testing to establish diagnosis or during radical prostatectomy are typically the most readily available; therefore, for practical reasons, assays that can be conducted on FFPE specimens are preferred. For biomarkers present at a higher frequency in progressive metastatic CRPC (relative to primary tumors that are noncastrate), a repeat biopsy of the metastatic lesion will be required, whereas for those assays that can be conducted reliably only in frozen tumor, a repeat biopsy of either metastatic or primary tumor immediately before trial entry will be necessary. Tumor specimens are stored for future assay. After confirmation that sufficient tumor is available for analysis, a patient is randomized to treatment with compound X or control. A key aspect of this design is that neither the predictive biomarker nor the

analytically validated tests are needed until the time of the final analysis of the trial.

The primary endpoint for the study is overall survival, which is the primary regulatory endpoint for new drug approval for CRPC. A total of 935 patients will be accrued, and the final analysis will be conducted when there are 700 total deaths. This will provide approximately 90% statistical power for detecting a 25% reduction in hazard of death for compound *X* relative to control at a 2-sided statistical significance level of 1%. The remaining 4% of type I error will be used for evaluating the statistical significance of treatment effect on survival in the adaptively defined biomarker subset that is anticipated to derive greater benefit than the population as a whole. This will provide approximately 80% statistical power for detecting a 37% reduction in the hazard of death in the adaptively defined subset of the validation set, which consists of only 33% of the validation set, as described in more detail below. By splitting the traditional 5% significance threshold into a portion to be used for the overall comparison and a portion to be used for the comparison within the subset, the type I error rate of the trial is preserved at 5%.

The type I error rate of 5% can be partitioned into a part for the overall analysis and a part for the subset analysis in a variety of ways. One could attempt to optimize the split to minimize the total sample size subject to constraints on the statistical power for both the overall analysis and the subset analysis. We have not attempted such an optimization. We have allocated most of the 5% to the subset analysis because the power of the subset analysis drives the overall sample size, particularly when a minority of patients benefit from the new treatment. By taking into account the correlation between the 2 analyses, less stringent significance levels could be used (29).

The final analysis will be conducted in the following manner. A log-rank test will be used to compare survival times in the 2 treatment arms for all randomized patients. If the 2-sided significance level is less than 0.01 and favors compound *X*, then compound *X* will be considered effective for the randomized population as a whole. If not, then the following analysis will be conducted with the fallback design of the adaptive signature approach developed by Freidlin and Simon (12).

A predictive classifier  $P(B_1, B_2, B_3, B_4)$  will be developed that identifies whether a patient with biomarker values  $B_1, B_2, B_3,$  and  $B_4$  (each representing the result of a specific validated assay) is likely to benefit from drug *X* compared with control *C*. For the purpose of illustration, we have arbitrarily specified 4 individual markers that can be used for building the classifier. The number is arbitrary as long as the markers and the algorithm for building the classifier with the candidate markers are specified before the data are examined and as long as an analytically validated assay is available for measuring each marker. The value of the adaptive signature design is greatest when the number of candidate markers is large. The  $P$  classifier will be developed using a randomly selected training set of patients consisting of 33% of the cases. The split proportion of 33% of the

patients for development and training of the classifier and 67% for evaluation of the classifier is somewhat arbitrary but influences the ability to develop a good classifier and to adequately compare the treatment in the subset of the validation set determined by the classifier. Dobbin and Simon (30) have studied the optimal splitting of data sets into a training set and a validation set for prognostic classifiers, but similar studies have not been reported for predictive classifiers as used in the adaptive signature design. We believe that a training set consisting of approximately 233 events should be adequate for developing a predictive classifier in which accuracy is close to that of the optimal classifier that could be developed with an infinite-sized training set, but a quantitative evaluation of this along the lines described by Dobbin and Simon should be pursued. Reducing the size of the validation set further constrains the statistical power of the subset analysis, as shown below in the paragraph describing how the power for the subset analysis in the validation set drives the total size of the study. The advantage of the more recently developed cross-validated adaptive signature design is that a fixed training-validation split is not required. (14)

The algorithm for developing the classifier is described in the Appendix, which follows the Discussion section. The value of the classifier function  $C(B_1, B_2, B_3, B_4)$  equals 1 if the patient with those biomarker values is likely to benefit from *X*, and equals 0 otherwise. The set IND of combinations of biomarker values  $(B_1, B_2, B_3, B_4)$  for which the classifier equals 1 is the indication for treatment *X* should the subset analysis be statistically significant. As part of the final analysis, this indication will be described graphically, analytically, by decision tree, and as a classification function.

The training set data are extensively analyzed to develop a single completely specified classifier. Predictive classifier development is different from traditional subset analysis. Although the development algorithm may involve evaluation of subsets determined by single variables, a classifier must be developed that integrates all such information into a single function of all the baseline variables to predict whether a patient will benefit from receiving the new treatment relative to the control. Although a large body of literature exists on prognostic signatures, very little literature is available on predictive 2-treatment classifiers. A single completely specified classifier should be developed with the training data. If multiple classifiers were developed, they would have to be evaluated in the validation set and that would require additional portions of the type I error to be allocated to evaluate them.

The estimated improvement in survival for *X* versus *C* in the indicated population IND will be estimated by classifying each patient in the trial who was not included in the training set used to develop the classifier. Let  $S$  denote the set of patients in this "test set" classified as likely to benefit from *X* using  $C(B_1, B_2, B_3, B_4)$ . Kaplan-Meier survival curves will be computed for the patients in  $S$  who received *X* and for the patients in  $S$  who received *C*. The difference between these 2 survival curves will be summarized with a log-rank statistic (LR) and a log hazard ratio (LHR) and a 95%

confidence interval for LHR. If the log-rank statistic LR is significant at the 4% level of the  $\chi^2$  distribution with one degree of freedom and the HR of  $X$  versus  $C$  is less than 1, then the treatment  $X$  will be considered effective in improving survival of patients with an indication specified by the set IND defined based on the classifier  $C$  ( $B_1, B_2, B_3, B_4$ ) as described above.

The statistical power of the biomarker-specified subset analysis depends on the proportion of patients who are included in the adaptively defined subset  $S$ . To have 80% power for detecting a 37% reduction in the hazard of death for  $X$  versus  $C$  (a reasonable target effect size given historical results with predictive biomarker-based treatments such as trastuzumab), approximately 157 deaths are required in the classifier-positive subset of the test set of patients (i.e., patients not used for developing the classifier). If one third of patients are classifier positive, then 471 total deaths are required in the test set. The test set will contain about two thirds of the patients and events. The total number of deaths at the time of final analysis will be 700, and hence this power target should be achievable.

## Discussion

For the goal of developing the right drug for the right patient to become more than a cliché, sponsors, investigators, and regulators must recognize that some of the conventional wisdom used to guide clinical trial design and analysis in the era of broadly targeted cytotoxic agents is no longer appropriate. Indeed, the continued use of traditional clinical trial designs is likely to hamper the development of new drugs that are highly effective for molecularly well-defined subsets of patients.

The use of conventional, primary site-based approaches to develop targeted cancer therapeutics is in many cases not consistent with our knowledge of the underlying biology of a tumor, exposes patients to toxic drugs from which they are not expected to benefit, and may result in long delays for the approval and ultimately the availability of drugs that offer substantial benefit to molecularly characterized subsets of patients. Clearly, in this new era, issues previously considered to be standard, such as the role of subset analysis, the role of stratification, the need to have broad eligibility criteria, and the use of adaptive methods, must be critically reexamined. However, new methods for clinical trial design and analysis must be no less rigorous than conventional designs in their use of randomized controls, clinically meaningful endpoints, and protection against type I error.

Methods for adaptive characterization and validation of the patients most likely to benefit from a new treatment in phase III oncology trials have been developed in recent years (12–14). The specific designs are adaptive in distinct ways, but most have focused on intratrial modification of the number of patients to be included (sample size reestimation) or modification of the randomization weights (response-adaptive). Controversies with these designs include the question of whether adaptive sample size reestimation is more effective than traditional

sequential analysis methods, whether response-adaptive methods provide statistical analyses that are robust to time trends in unmeasured prognostic factors, and whether response-adaptive methods improve efficiency (10). As a result, response-adaptive designs are rarely used in phase III clinical trials.

A more promising area is the adaptive characterization of patients enrolled in phase III trials who are most likely (or least likely) to benefit from a new treatment. This adaptive determination of the treatment indication represents a paradigm shift in phase III clinical trial design with the potential for a major impact on oncology drug development and a major benefit to patients. However, there is a need for dialogue among academic investigators, government and industry sponsors, and regulators on how best to use this methodology. It was for this reason that this area of adaptive clinical trial design was chosen for focus by the members of the adaptive design panel of the 2010 Conference on Clinical Cancer Research.

Use of adaptive methods to identify the patients who are most likely or least likely to benefit from a new regimen requires substantial prospective planning. The methods cannot be used reliably in an exploratory *post hoc* manner. In fact, if done improperly they can introduce bias and risk "disqualifying" a trial as adaptive in the view of the FDA.

The following are some of the key features of the clinical trial we designed:

1. Use of an acceptable regulatory endpoint such as overall survival as the primary endpoint for final analysis.
2. Use of a randomized design with an appropriate control arm.
3. Obtaining tumor specimens prior to randomization for all patients registered on the trial. Tumor assays may be conducted at a later time, but prior to data analysis, if the analytically validated tests are not available when the clinical trial is initiated.
4. Use of an intermediate endpoint for interim futility analysis is considered necessary but not sufficient to ensure a treatment effect on the primary endpoint, even though it is not a validated surrogate of the primary endpoint.
5. Use of analytically validated tests for measuring all candidate predictive biomarkers.
6. Prespecification of the algorithm to be used for developing the classifier in the training set, and prespecification of how the validation analysis will be conducted.
7. Adequately powering the clinical trial for validation of a substantial treatment effect in the adaptively identified subset.
8. At the time of final analysis the patients are randomly partitioned into a portion (e.g., one third) for training a classifier that identifies which patients are most and least likely to benefit from the new treatment, and a portion (e.g., two thirds) for validating that classifier. The full set of patients is used for the overall

comparison of the new treatment with the control, but only the validation set is used for the comparison of new treatment versus control in the adaptively determined subset. The type I error is allocated among those 2 comparisons.

Oncology therapeutics development is in an era of fundamental change. The discoveries of both the molecular basis for human cancers and the heterogeneity of cancer provide a great opportunity to develop more effective treatments and to properly deliver them to the right patients. Many challenges remain to be addressed, and some familiar paradigms require reevaluation. What does not change is the need for clinical trials that are fundamentally science based, statistically sound, and responsive to the urgency for reducing mortality and morbidity from cancer.

## Appendix

A wide variety of classifier development algorithms are possible. The algorithm should be described completely in the protocol. For the study being illustrated here, the classifier will be developed by the following algorithm:

A proportional hazards model will be fit to the data for the combined treatment  $X$  and control group. Denote this model by

$$\log[\lambda(t, B1, B2, B3, B4, v)/\lambda_0(t)] = \delta v + \beta_1 B1 + \beta_2 B2 + \beta_3 B3 + \beta_4 B4 + v(\gamma_1 B1 + \gamma_2 B2 + \gamma_3 B3 + \gamma_4 B4)$$

where  $v$  is a binary treatment indicator ( $v = 1$  for  $X$ ,  $v = 0$  for  $C$ ),  $\delta$  is the regression coefficient that represents the main effect of treatment on survival, the  $\beta$ s reflect the prognostic effects of the biomarkers, and the  $\gamma$ s are the interaction effects that represent the predictive effects of the biomarkers. The left-hand side of the equation represents the log hazard relative to the baseline hazard. The markers will only be binary if a cut point is predefined based on preliminary data. Otherwise, no cut point will be imposed on the modeled values.

For a patient with biomarker values ( $B1, B2, B3, B4$ ), the LHR if the patient receives treatment  $X$  minus the LHR if the patient receives the control  $C$  is

$$\Delta(B1, B2, B3, B4) = \delta + \gamma_1 B1 + \gamma_2 B2 + \gamma_3 B3 + \gamma_4 B4$$

By fitting the model to the data, we obtain estimates of the regression coefficients and a covariance matrix for these estimates. Hence, for any vector of biomarker values, we can compute  $\hat{\Delta}(B1, B2, B3, B4)$  in which the regression coefficients are replaced by their estimates, and we can compute the variance  $V[\hat{\Delta}(B1, B2, B3, B4)]$ . A

binary classifier will be defined by

$$C(B1, B2, B3, B4) = 1$$

$$\text{if } \hat{\Delta}(B1, B2, B3, B4) / \sqrt{V[\hat{\Delta}(B1, B2, B3, B4)]} \leq c$$

The patient is classified as likely to benefit from  $X$  if the standardized LHR of  $X$  relative to  $C$  is less than or equal to constant  $c$ . The constant will be determined by 10-fold cross-validation within the training set to maximize the log-rank statistic for treatment effect within the training set patients classified as likely to benefit from  $X$ . Application of this algorithm to the training data provides a completely specified classifier that can be used to classify each of the patients in the validation set. Each patient in the validation set has biomarker values  $B1, B2, B3$ , and  $B4$ . By plugging in these values to the classifier, the patient is classified as likely or unlikely to benefit from  $X$  relative to  $C$ . The patients in the validation set who are classified as likely to benefit from  $X$  are the subset to be analyzed. In that subset, outcomes for patients who received  $X$  are compared with outcomes for those who received the control  $C$ .

The classifier illustrated here is based on a proportional hazards regression analysis of 4 biomarker values. Alternatively, variable selection strategies could be used to include only variables that seem informative for distinguishing outcome on  $X$  from outcome on  $C$ . When the number of candidate variables is large, variable selection is essential. It should be recognized, however, that the objective is to accurately classify patients as to whether they will benefit from  $X$ , not to document with statistical significance the importance of individual variables.

## Disclosure of Potential Conflicts of Interest

Dr. Scher has received research funding from Ortho Biotech Oncology Research and Development and Medivation. He has served as a consultant or in an advisory role for Ortho Biotech Oncology Research and Development and Medivation. No potential conflicts of interest were disclosed by the other authors.

## Acknowledgments

Rajeshwari Sridhara contributed on behalf of the U.S. FDA. The authors thank Samantha Roberts for a critical review of the manuscript and Margaret McPartland for editorial assistance.

## Grant Support

This work was supported in part by the MSKCC SPORE in Prostate Cancer (P50 CA92629), the Department of Defense Prostate Cancer Research Program (PC051382), The Research and Therapeutics Program for Prostate Cancer, and The Prostate Cancer Foundation.

Received May 6, 2011; revised September 22, 2011; accepted September 22, 2011; published online November 1, 2011.

## References

1. Roberts SA, Karnes EK, Allen JD, Benner JS, Sigal EV, McClellan M. Achieving the goals of effective, safe, and individualized cancer care. *Clin Cancer Res* 2011;17:6632-3.
2. Turteltaub KW, Davis MA, Burns-Naas LA, Lawton MP, Clark AM, Reynolds JA. Identification and elucidation of the biology of adverse events: the challenges of safety assessment

- and translational medicine. *Clin Cancer Res* 2011;17:6641–5.
3. Cleeland CS, O'Mara A, Zagari M, Baas C. Integrating pain metrics into oncology clinical trials. *Clin Cancer Res* 2011;17:6646–50.
  4. Ambrose LF, Freedman J, Buetow K, Friend S, Schilsky RL. Using patient-initiated study participation in the development of evidence for personalized cancer therapy. *Clin Cancer Res* 2011;17:6651–7.
  5. Rubin EH, Anderson KM, Gause CK. The BATTLE trial: a bold step toward improving the efficiency of biomarker-based drug development. *Cancer Discov* 2011;1:17–20.
  6. Simon R. The use of genomics in clinical trial design. *Clin Cancer Res* 2008;14:5984–93.
  7. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J Clin Oncol* 2009;27:4027–34.
  8. Printz C. BATTLE to personalize lung cancer treatment. Novel clinical trial design and tissue gathering procedures drive biomarker discovery. *Cancer* 2010;116:3307–8.
  9. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther* 2009;86:97–100.
  10. Korn EL, Freidlin B. Outcomeadaptive randomization: is it useful? *J Clin Oncol* 2011;29:771–6.
  11. United States Food and Drug Administration: 2010 draft guidance for industry on adaptive design clinical trials for drugs and biologics. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm201790.pdf>.
  12. Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res* 2005;11:7872–8.
  13. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J Natl Cancer Inst* 2007;99:1036–43.
  14. Freidlin B, Jiang W, Simon R. The cross-validated adaptive signature design. *Clin Cancer Res* 2010;16:691–8.
  15. Simon R. Clinical trials for predictive medicine: new challenges and paradigms. *Clin Trials* 2010;7:516–24.
  16. Scher HI, Sawyers C. Biology of progressive castration resistant prostate cancer: directed therapies targeting the androgen receptor signaling axis. *J Clin Oncol* 2005;23:8253–61.
  17. Holzbeierlein J, Lal P, LaTulippe E, Smith A, Satagopan J, Zhang L, Ryan C, et al. Gene expression analysis of human prostate carcinoma during hormonal therapy identifies androgen-responsive genes and mechanisms of therapy resistance. *Am J Pathol* 2004;164:217–27.
  18. Stanbrough M, Bubley GJ, Ross K, Golub TR, Rubin MA, Penning TM, et al. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res* 2006;66:2815–25.
  19. Danila DC, Morris MJ, de Bono JS, Ryan CJ, Denmeade SR, Smith MR, et al. Phase II multicenter study of abiraterone acetate plus prednisone therapy in patients with docetaxel-treated castration-resistant prostate cancer. *J Clin Oncol* 2010;28:1496–501.
  20. Reid AH, Attard G, Danila DC, Oommen NB, Olmos D, Fong PC, et al. Significant and sustained antitumor activity in post-docetaxel, castration-resistant prostate cancer with the cyp17 inhibitor abiraterone acetate. *J Clin Oncol* 2010;28:1489–95.
  21. Tran C, Ouk S, Clegg NJ, Chen Y, Watson PA, Arora V, et al. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 2009;324:787–90.
  22. de Bono JS, Logothetis CJ, Molina A, Fizazi K, North S, Chu L, et al. Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med* 2011;364:1995–2005.
  23. Scher HI, Beer TM, Higano CS, Anand A, Taplin ME, Efstathiou E, et al. Antitumor activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet* 2010;375:1437–46.
  24. Knudsen KE, Scher HI. Starving the addiction: new opportunities for durable suppression of AR signaling in prostate cancer. *Clin Cancer Res* 2009;15:4792–8.
  25. Kantoff PW, Higano CS, Shore ND, Berger ER, Small EJ, Penson DF, et al. Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N Engl J Med* 2010;363:411–22.
  26. de Bono JS, Oudard S, Ozguroglu M, Hansen S, Machiels JP, Kocak I, et al. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. *Lancet* 2010;376:1147–54.
  27. Nilsson S, O'Bryan-Tear CG, Bolstad B, Lokna A, Parker CC. Alkaline phosphatase (ALP) normalization and overall survival in patients with bone metastases from castration-resistant prostate cancer (CRPC) treated with radium-223. [abstract]. In: Proceedings of the 2011 ASCO Annual Meeting; 2011 Jun 3–7; Chicago, IL: ASCO; 2011. Abstract nr 4620.
  28. Positive outcome of interim analysis of pivotal Alpharadin study: primary endpoint met in phase III ALSYMPCA study [press release]. Oslo, Norway: Algeta ASA; 2011.
  29. Wang SJ, O'Neill RT, Hung HM. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm Stat* 2007;6:227–44.
  30. Dobbin KK, Simon RM. Sample size planning for developing classifiers using high-dimensional DNA microarray data. *Biostatistics* 2007;8:101–17.