

Effective Incorporation of Biomarkers into Phase II Trials

Lisa M. McShane,¹ Sally Hunsberger,¹ and Alex A. Adjei²

Abstract The incorporation of biomarkers into the drug development process will improve understanding of how new therapeutics work and allow for more accurate identification of patients who will benefit from those therapies. Strategically planned biomarker evaluations in phase II studies may allow for the design of more efficient phase III trials and better screening of therapeutics for entry into phase III development, hopefully leading to increased chances of positive phase III trial results. Some examples of roles that a biomarker can play in a phase II trial include predictor of response or resistance to specific therapies, patient enrichment, correlative endpoint, or surrogate endpoint. Considerations for using biomarkers most effectively in these roles are discussed in the context of several examples. The substantial technical, logistic, and ethical challenges that can be faced when trying to incorporate biomarkers into phase II trials are also addressed. A rational and coordinated approach to the inclusion of biomarker studies throughout the drug development process will be the key to attaining the goal of personalized medicine.

It is widely believed that incorporation of biomarkers into the drug development process will improve understanding of how new therapeutics work and allow for more accurate identification of patients who will benefit from those therapies. Many aspects of the conduct of phase II trials might be considered when evaluating how trials might be made more efficient and successful (1), but this article specifically discusses the benefits and challenges of incorporating biomarkers into phase II cancer clinical studies. The term biomarker will be understood to mean “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (2). Biomarkers may be measured by laboratory assays on a variety of specimens, including, for example, tumor tissue, whole blood, plasma, serum, bone marrow, or bodily fluids such as urine. The biomarkers may be tumor-based or may measure host characteristics such as germ-line DNA mutations or polymorphisms. In addition, biomarkers may be assessed using molecular imaging techniques *in vivo* (3). There are many different roles that a biomarker can play in a phase II trial. Some roles discussed in this article include predictor of response or resistance to specific therapies, patient enrichment, correlative endpoint, or surrogate endpoint.

The use of biomarkers is particularly appealing for molecularly targeted therapies, as it seems likely that obtaining

biomarker measurements associated with the target may be helpful in evaluating those treatments. For example, the target of interest might be a protein, the therapy could be a monoclonal antibody directed at that protein, and the biomarker measurement might be the expression level of that protein. Striking examples of biomarkers that had a pivotal role in development of new therapies over the last decade include HER-2 protein overexpression or gene amplification for trastuzumab therapy in breast cancer and BCR-ABL fusion product for imatinib mesylate therapy in chronic myelogenous leukemia (4). Although inclusion of biomarkers into phase II trials appears highly attractive from a scientific perspective, inclusion of biomarker studies can also present substantial technical, logistic, and ethical challenges. Ultimately, the hope is that rational inclusion of biomarkers into phase II trials will lead to a higher success rate and more efficient design of phase III trials while avoiding premature abandonment of useful therapies at the phase II development stage.

Predictive and Enrichment Biomarkers

A predictive biomarker is a measurement associated with response or lack of response (e.g., resistance) to a particular therapy (5). Biomarkers of toxicity could also be viewed as a type of predictive biomarker for which the prediction is for harm that is to be avoided. Perhaps the best-known predictive biomarker is estrogen receptor status for prediction of response to endocrine therapy for breast cancer. Estrogen receptor-negative breast tumors are unlikely to respond to endocrine therapy, whereas a substantial percentage of estrogen receptor-positive breast tumors will respond to endocrine therapy. For molecularly targeted therapies, biomarkers related to the target are natural candidates for predictive biomarkers. Ideally, one would like to have some knowledge of potential predictive biomarkers before testing a new therapy in phase II trials, but often a predictive biomarker will not be clearly identified or

Authors' Affiliations: ¹Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, Maryland and ²Katherine Anne Gioia Chair in Cancer Medicine, Roswell Park Cancer Institute, Buffalo, New York
Received 11/13/08; revised 1/12/09; accepted 1/12/09; published OnlineFirst 3/10/09.

Requests for reprints: Lisa M. McShane, Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Room 8126, Executive Plaza North, MSC 7434, 6130 Executive Boulevard, Bethesda, MD 20892. Phone: 301-402-0636; Fax: 301-402-0560; E-mail: lm5h@nih.gov.

© 2009 American Association for Cancer Research.
doi:10.1158/1078-0432.CCR-08-2033

there will not be a suitably well-developed assay available for measuring the biomarker at the start of a phase II trial.

If information is available to suggest subgroups of patients who are more likely to benefit from a therapy, it may be reasonable to conduct the phase II trial only in those patients. Factors used to limit the study population to patients believed more likely to benefit from the experimental therapy are termed enrichment factors. Enrichment factors may be predictive biomarkers, or they may be biomarkers or clinicopathologic characteristics (such as squamous cell lung cancer for pemetrexed) or demographic characteristics (such as females and never-smokers with lung adenocarcinoma for epidermal growth factor receptor inhibitor therapy) associated with a predictive biomarker or with the target of a therapeutic agent. The enrichment factors considered in this article are assumed to be biomarker-based. The lower the proportion of truly benefiting patients, the more advantageous it is to consider studying an enriched population (6) even if the enrichment biomarker can only approximately identify the benefiting patient group. Biomarkers that could be useful as enrichment biomarkers during the drug development process might still need further refinement before they are ready for clinical use as predictive factors. This is because many enrichment biomarkers used in drug development do not have sufficiently high positive or negative predictive value to justify clinical use or the assay used to measure the biomarker during the drug development process might not be sufficiently robust and reproducible for routine clinical use. The main purpose of using an enrichment biomarker in drug development is to improve the chances that the drug will show benefit in the tested subgroup of patients to more quickly establish that the drug is worth pursuing further. Once it has been shown that there is some group of patients who benefit, the enrichment biomarker and its assay can be further developed into a clinically useful predictive biomarker test.

Trastuzumab is an example of a targeted therapeutic for which enrichment strategies were used in the clinical development process. Trastuzumab is a monoclonal antibody that targets the HER-2/*neu* receptor. Preclinical studies (7, 8) provided evidence that trastuzumab was most likely to be effective against tumor cells that overexpressed the HER-2/*neu* receptor. Pivotal phase II studies of trastuzumab as monotherapy in patients with metastatic breast cancer (9, 10) required evidence of HER-2 overexpression by immunohistochemistry for eligibility and this enrichment strategy was maintained in subsequent clinical trials evaluating trastuzumab in other settings. Further studies suggested that HER-2 gene amplification as measured by fluorescence *in situ* hybridization may be a more reliable indicator of benefit from trastuzumab therapy (11). Given that the positivity rate for HER-2 (immunohistochemistry 3 positive or fluorescence *in situ* hybridization positive) is ~25% to 30%, the benefit of trastuzumab might not have been detected in metastatic patients had the drug development occurred in a nonenriched (all comers) setting. When trastuzumab was tested in the adjuvant setting, the patient population was also enriched for patients whose tumors were immunohistochemistry 3 positive or fluorescence *in situ* hybridization positive. Interestingly, a

current controversy is whether the benefit of trastuzumab delivered as adjuvant therapy is limited to this enriched group of patients or whether patients with tumors that are immunohistochemistry 1 or 2 positive without amplification (HER-2-low) might also derive some benefit from trastuzumab (12). Should further studies confirm that HER-2-low patients benefit from trastuzumab in the adjuvant setting, possible explanations would include variations in assay methodology or alternative mechanisms of action of trastuzumab in early-stage disease.

The story of epidermal growth factor receptor targeting agents is far more confusing than the story of HER-2 and trastuzumab (4). Small-molecule inhibitors and the monoclonal antibodies targeting epidermal growth factor receptor have been studied in several cancers, with a wide range of results both for the effectiveness of the treatment and for biomarkers that predict treatment benefit. In colorectal cancer, for example, the monoclonal antibody cetuximab has been shown to have clinical benefit, but there is no clear association between epidermal growth factor receptor overexpression (as measured by immunohistochemistry) and benefit. In contrast, the presence of certain activating KRAS mutations may confer resistance to cetuximab through dysregulation of downstream signaling pathways (13).

In some instances, the putative target of the agent in early clinical testing is found to be wrong. Thus, using this target as a biomarker would lead to major errors in the development of that agent. A recent example is sorafenib, which was originally developed and tested as an inhibitor of the kinase activity of c-Raf but was later found to be an inhibitor of the kinase activity of angiogenic receptors, particularly, vascular endothelial growth factor receptors (14). Misspecification of a target biomarker can have significant consequences. If an agent truly benefits all patients equally without regard to target biomarker status, then studying only those patients whose tumors are positive for the biomarker will only slow accrual to trials and increase expense while producing no improvement in the chance of detecting a benefit of the new therapy and unnecessarily limiting the size of the patient population offered the agent. If an agent benefits a certain subset of patients but the wrong subgroup of patient is studied because of a faulty biomarker, then a good agent could mistakenly be abandoned in the drug development process. These examples show the potential pitfalls in pursuing enrichment strategies when the biological pathways and mechanism of action of the therapeutic agent are not well understood.

Economic considerations will likely play a role in determining how biomarkers will be used in the drug development process. Assay development, biomarker screening for trial entry, and eventual market size for the drug all have cost implications. If the proportion of patients identified by an enrichment biomarker as having an increased likelihood of benefiting from the therapy is large, 85% or 90% of the general patient population, it might not be cost-effective to spend the time or resources to develop a biomarker assay for enrichment in phase II trials. In contrast, if the proportion of patients judged likely to benefit is modest, it may be essential to have in place a reasonably well-developed assay at the phase II trial stage even if the magnitude of benefit for that minority group of patients is fairly large or a beneficial therapeutic may be overlooked in

a trial that includes all patients. The total sample size required for an enriched trial (number of subjects screened for eligibility using the biomarker) compared with the sample size required for a similar trial design without enrichment will depend on the proportion of patients in the enriched subgroup and the magnitude of treatment benefit in that enriched subset. The goal at the end of the drug development process is to have a drug that works in some group of patients and to be able to reasonably accurately identify that group of benefiting patients. Decisions that have to be made during the development process include whether an enrichment or predictive biomarker is needed at all and, if needed, at what point resources should be committed to refining an enrichment biomarker assay into a clinically usable predictive biomarker test. There exists a tension between the goal of rapid therapeutic development and the goal of developing a reasonably robust and accurate assay that will be useful for identifying individual patients who will or will not benefit from a particular therapy.

Biomarker-Based Correlative Endpoints

It has proven very difficult to establish robust clinical trial endpoints based on biomarkers. The terms surrogate and surrogate endpoint have been intentionally avoided up to this point in this article because these terms are widely misunderstood and misused (15). A perfect trial-level surrogate endpoint would be one for which the surrogate (e.g., biomarker) could be substituted for a definitive trial endpoint in a new trial and that trial would reach the same conclusion regarding treatment effect (16). To make this assessment, usually a meta-analytic approach is needed where data are analyzed from a series of trials in which both the putative surrogate endpoint and the definitive trial endpoint were measured. The series of trials allows for proper inference about whether the surrogate endpoint could be used reliably in a new trial conducted in a similar patient group, with therapies having mechanisms of action similar to the therapies used in the previous trials. Although substitution of endpoints rarely will be perfect, several investigators have proposed methods for quantifying the reliability of the substitution such as R^2 -type measures (17) or plots of within-trial treatment arm differences of the definitive trial endpoint versus within-trial treatment arm differences of the surrogate variables (16–19). In many situations, there would not be sufficient data from previous trials to even attempt to formally verify trial-level surrogacy using such a meta-analytic approach. Moreover, the results may be very specific to a particular class of therapeutic agents and to a specific patient population. An example is the diminution of the standard uptake value on fluorodeoxyglucose-positron emission tomography scanning of patients with gastrointestinal stromal tumors after therapy with imatinib (3). Response to imatinib therapy and clinical benefit appear highly associated with early, and often dramatic, changes in tumor metabolism as measured by fluorodeoxyglucose-positron emission tomography, but traditional tumor response measures based on computed tomography scans may lag far behind the fluorodeoxyglucose-positron emission tomography indicators. In contrast, computed tomography-based measurements may be

more relevant to assessing response to conventional cytotoxic therapies.

It is important to distinguish the notion of trial-level surrogacy from the notion of individual-level surrogacy in a specific trial. For individual-level surrogacy in a specific trial, it needs to be established that the surrogate can be reasonably substituted for the definitive endpoint for the patients in that specific trial without changing the conclusion regarding treatment effect. For example, the well-known Prentice criteria (20) are often applied to the situation of establishing individual-level surrogacy in the context of a single trial. This does not provide direct evidence, however, that the surrogacy would hold in a broader population of trials, such as for a trial one might be planning that involves a new therapy. Only if the biology of the surrogate biomarker and the mechanism of action of the drug of interest are so well understood that it can be asserted that the action of the new drug on the definitive endpoint is entirely mediated through the surrogate endpoint can one confidently substitute the surrogate endpoint for the definitive endpoint in the absence of empirical evidence from a series of trials. For these reasons, evaluating a first-in-class therapeutic using a surrogate endpoint previously established in a different setting can be especially risky.

Despite the difficulties in establishing that a biomarker is a reliable surrogate endpoint, biomarker measurements made during and after therapy may still be helpful in understanding how a therapy is interacting with its target or may give earlier indication of the likely effectiveness of a therapy than more traditional clinically based outcome measures, particularly in the setting of cytostatic agents. Even if the biomarker endpoints do not replace more conventional clinical endpoints in phase II trials, they might, for example, be useful as early indicators of treatment efficacy that could be used in the conduct of phase II screening trials (21, 22). Biomarker-based endpoints that are useful despite a lack of sufficient data to formally establish surrogacy will be termed correlative endpoints.

The use of prostate-specific antigen (PSA) in advanced prostate cancer illustrates some of the difficulties in establishing and using biomarker-based correlative endpoints. For metastatic castrate-resistant prostate cancer, PSA measures have been recommended and widely used in phase II trials (23), but substantial skepticism remains (24). For example, different studies have suggested different cutoffs for percent decline in PSA to best correlate with overall survival benefit, and the situation has become even more confusing with the increasing numbers of targeted agents being evaluated in phase II clinical trials. Targeted agents may affect different cancer cell subpopulations. Some targeted agents modulate PSA and others do not, and the change in PSA may or may not correlate with degree of clinical benefit as measured by a definitive endpoint such as overall survival. For example, finasteride is well known to profoundly lower PSA, but it is not an effective treatment for prostate cancer (25). Therefore, PSA should not be considered a broadly validated surrogate endpoint.

In response to continued questions about the use of PSA endpoints, a working group was formed to review issues of

Table 1. Effect of enrichment for biomarker-positive cases on the probability of falsely concluding that a new therapy increases the response rate when a single-arm phase II study is designed using assumptions from an unselected population

Historical response rate in the unselected population (%)	Historical response rate in the subpopulation enriched for biomarker-positive cases (%)	Sample size	No. observed responses required to conclude that the response rate is increased by the new therapy	Actual probability of falsely concluding that the new therapy increases the response rate in the enriched subpopulation
10	15	30	6	0.29
10	20	30	6	0.57
20	25	36	11	0.27
20	30	36	11	0.53
30	35	39	16	0.26
30	40	39	16	0.51

NOTE: Single-arm phase II studies are typically designed to test against benchmark historical response rates in the range of the example rates provided in column 1. When the historical response rate in the subpopulation enriched for biomarker-positive cases is unknown, it is often assumed to be the same as in the unselected population. Column 2 presents the (unknown) historical response rate in the subset of patients who are positive for the biomarker. Sample size (column 3) and required observed response rate (column 4) are calculated to satisfy two conditions: (a) control the probability of falsely concluding that the true response rate has been increased from the historical rate in the unselected population to be ≤ 0.10 and (b) provide at least 0.90 probability of concluding that the new therapy has increased the response rate when the new therapy has a true response rate that represents an absolute increase of 20% compared with the historical response rate in the unselected population.

design and endpoints for clinical trials for patients with progressive castrate-resistant prostate cancer (26). The working group report issued further cautions in using PSA-based endpoints and recommended shifting emphasis to time-to-event endpoints (failure to progress), particularly for non-cytotoxic therapies. These concerns do not imply that PSA measurements have no value when conducting prostate cancer clinical trials. Elevated PSA levels are an established indicator for worse prognosis and rapid PSA doubling time may be an early indicator of treatment ineffectiveness. It will be important to continue to record PSA measurements in prostate cancer in a consistent way to amass data that might eventually provide the information needed to better evaluate the usefulness of PSA-based endpoints in a broad range of treatment contexts.

There are several other examples of biomarkers that have shown promise for use either as correlative endpoints or as part of composite endpoints combining biomarker measurements with a clinical endpoint. A series of studies have suggested that CA-125 may be useful as a response indicator in clinical trials in advanced ovarian cancer (27–30). Several types of biomarkers have shown promise as response indicators for antiangiogenic agents. These include growth factors or soluble growth factor receptors in blood or urine, circulating endothelial cells, circulating endothelial progenitor cells, and noninvasive imaging measurements (31). Such endpoints are especially important for antiangiogenic agents because many antiangiogenic agents are cytostatic, making traditional measures of tumor shrinkage unreliable indicators of agent activity or clinical benefit. Validation of these biomarkers, however, with respect to both the analytic properties of the assays used to measure them and the clinical value of those assay results, needs to be done before they can be widely used in clinical decision-making.

Biomarker-Based Phase II Trial Designs

A variety of trial designs are used for phase II investigations. Historically, single-arm trials using objective response rate as the endpoint have been widely used for cytotoxic agents, but randomized phase II study designs are increasingly being proposed (32) and used. For evaluation of molecularly targeted agents, a variety of novel designs have been suggested (33). A main goal of incorporating biomarkers into phase II studies of molecularly targeted agents is to determine if the new drug should be developed for all patients without patient selection or whether it should be developed for a biomarker-defined patient subset only. Definitive comparison of the new drug with existing therapies (with or without enrichment of the patient population using the biomarker) is accomplished in phase III trials and will not be addressed in this article. Presented here are brief remarks on a few phase II designs that have been proposed that specifically incorporate biomarker-based subgroups for testing for response from a new therapeutic. More extensive discussions of specific phase II design strategies and endpoints are discussed elsewhere (21, 22).

One must be cautious in using traditional single-arm trial designs, for example, the commonly used Simon two-stage design (34), when examining biomarker-defined subgroups of patients. Single-arm trials with tumor response endpoints rely on the ability to specify a benchmark response rate that a new agent must exceed to make it of sufficient interest to warrant further study. These benchmark response rates are generally based on prior experience in unselected patients receiving standard therapies. If an enrichment strategy is used to determine eligibility for the trial, it is important to recognize the possibility that the enrichment characteristics could be

prognostic. Patients in the enriched group might be more or less likely to respond to any type of therapy than the general population of patients. This implies that the benchmark response rate for an enriched trial might require adjustment accordingly. Unfortunately, the biomarkers that define the enriched subgroup may never have been measured in historical patient cohorts. To obtain these revised benchmark estimates of response rate might require retrospectively performing the new biomarker assays on archived specimens from patients who received standard therapies. The calculations presented in Table 1 show how misleading the results of a single-arm phase II trial based on testing against a benchmark response rate can be when the benchmark response rate is inappropriate for the enriched study population. (A single-stage design is considered for simplicity.) For example, consider a situation in which the historical response rate in an unselected population is 20% and a single-arm phase II study is designed to test whether a new therapy yields an improved response rate. The sample size is calculated so that there will be at least 0.90 probability of concluding that the new therapy has a response rate >20% if the true response rate is actually 40%. The statistical test that will be done at the end of the study will have probability no more than 0.1 of concluding that the new therapy response rate is >20% when, in truth, it is not. Now, if an enriched subpopulation is studied and the historical response rate with standard therapy

(or no therapy) in this subpopulation is 30% rather than the assumed 20%, then there will be 0.52 probability of concluding that the new therapy yields an improved response rate when it really does not offer an improvement for that enriched subpopulation. In situations where there are no active agents currently available for the patient group under study, a single-arm phase II study with response endpoint might still be very appropriate because even modest activity of the agent in the enriched subgroup might warrant further investigation of the agent.

A biomarker-adaptive parallel Simon two-stage design has recently been proposed for the evaluation of a targeted agent that it is thought might have different activity (response rate) in subgroups defined by biomarker positive versus negative (35). The design assumes that the biomarker is prespecified. The investigators discuss running two parallel two-stage designs, one in each of the biomarker-positive and biomarker-negative groups, and then propose an adaptive parallel design. The adaptive parallel two-stage design initially begins with two parallel studies, one conducted in N_1^- biomarker-negative subjects and the other conducted in N_1^+ biomarker-positive subjects. The design continues enrolling N^{un} unselected subjects during the second stage if the number of responses to the drug in the biomarker-negative group in the first stage, X_1^- , meets or exceeds a cutoff of k_1^- . The design enrolls

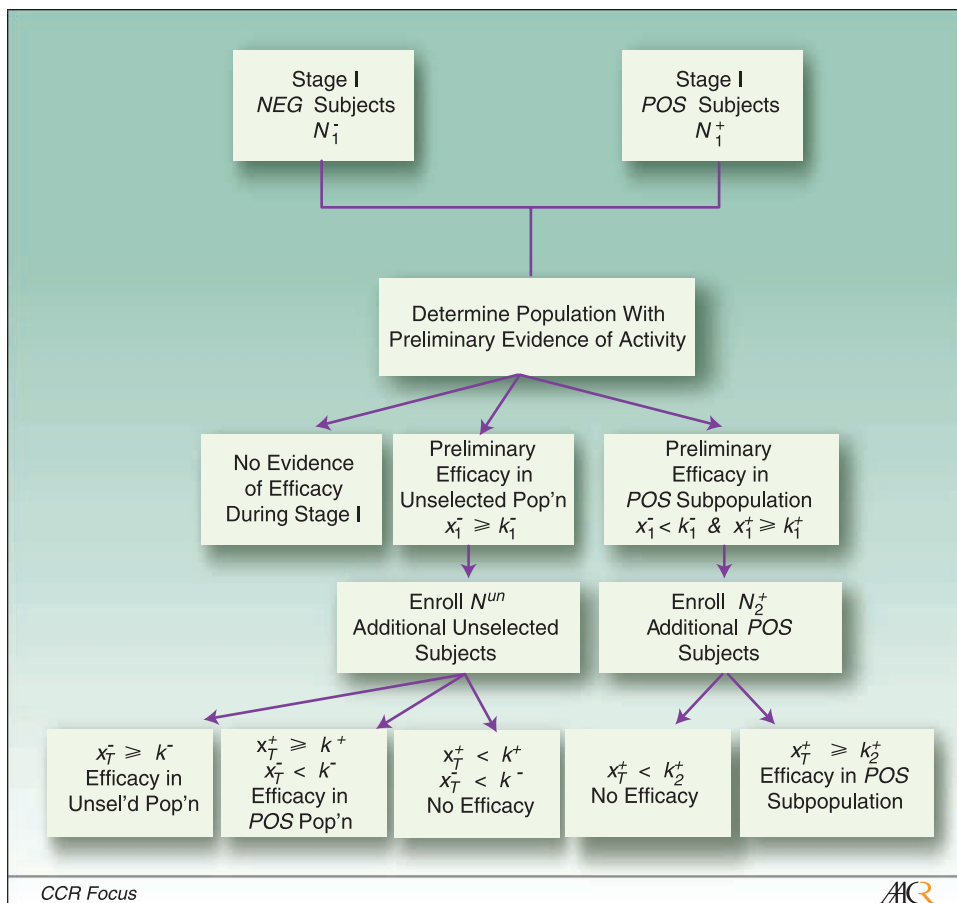
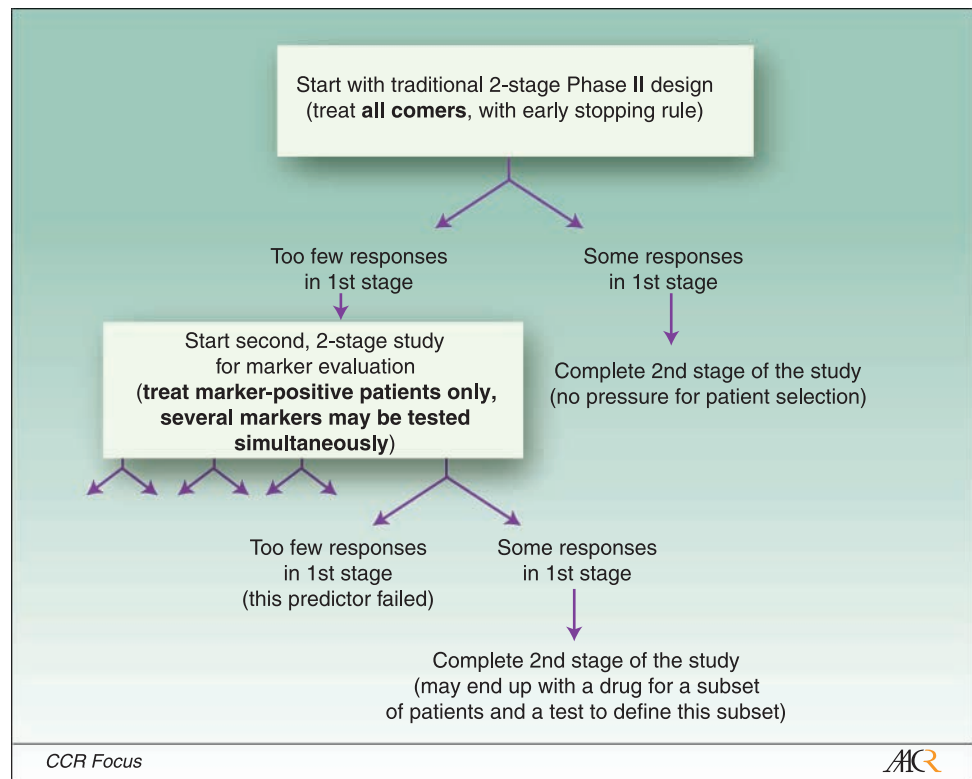


Fig. 1. Schema of the adaptive parallel two-stage design. In the first stage of the design, N_1^- biomarker-negative (NEG) subjects and N_1^+ biomarker-positive (POS) subjects are studied. If the number of responses to the drug in the biomarker-negative group in the first stage, X_1^- , meets or exceeds a cutoff of k_1^- , then N^{un} unselected subjects are accrued during the second stage. If X_1^- is less than k_1^- but the number of responses in the biomarker-positive group in the first stage, X_1^+ , meets or exceeds a cutoff of k_1^+ , then the design enrolls N_2^+ additional biomarker-positive subjects during the second stage and no further biomarker-negative subjects are accrued. A total of N^+ and N^- biomarker-positive and biomarker-negative subjects, respectively, will have been enrolled by the end of the second stage. A total of X_T^+ (biomarker-positive group) and X_T^- (biomarker-negative group) responders will have been observed. To determine drug benefit, total responses X_T^+ and X_T^- are compared against cutoffs k^+ and k^- if unselected patients continued to be enrolled during the second stage or X_T^+ is compared against the cutoff k_2^+ if only biomarker-positive subjects were enrolled in the second stage. The trial stage- and group-specific samples sizes N_1^- , N_1^+ , N^{un} , N_2^+ and cutoffs k_1^- , k_1^+ , k^- , k^+ , k_2^+ are determined so that they control the probability of correct conclusions in the biomarker-positive and unselected patient groups. Adapted from Jones CL and Holmgren E. An adaptive Simon two-stage design for phase 2 studies of targeted therapies. *Contemp Clin Trials* 2007;28:654–61, with permission from Elsevier.

Fig. 2. Schema of the tandem two-step phase II predictor biomarker evaluation trial design. In the first stage, a group of patients unselected with regard to biomarker status (all comers) is studied. If sufficient numbers of objective responses are observed in the first stage, then the study continues into the second stage, still accruing unselected patients. If the number of responses observed in the first stage is not sufficiently high, then the study continues accruing only patients in the subgroup predicted by the pharmacogenomic classifier to be responders; study termination is governed by a standard optimal two-stage phase II trial design in that subgroup of patients predicted to be responders. Adapted with permission from Pusztai et al. (Fig. 3). Pharmacogenomic predictor discovery in phase II clinical trials for breast cancer. Clin Cancer Res 2007;13:6080–6.



N_2^+ additional biomarker-positive subjects during the second stage and no further biomarker-negative subjects if the number of responses in the biomarker-negative group fails to attain the cutoff k_1^- , whereas the number of responses in the biomarker-positive group in the first stage, X_1^+ , meets or exceeds a cutoff of k_1^+ . A total of N^+ and N^- , biomarker-positive and biomarker-negative subjects, respectively, will have been enrolled by the end of the second stage. A total of X_2^+ (biomarker-positive group) and X_2^- (biomarker-negative group) responders will have been observed. To make final conclusions regarding efficacy, total responses X_2^+ and X_2^- are then compared against cutoffs k^+ and k^- if unselected patients continued to be enrolled during the second stage or X_2^+ is compared against the cutoff k_2^+ if only biomarker-positive subjects were enrolled in the second stage.

The trial stage- and group-specific samples sizes N_1 , N_1^+ , N_1^{un} , N_2^+ and cutoffs k_1^- , k_1^+ , k^- , k^+ , k_2^+ are determined so that they control the probability of correct conclusions in the biomarker-positive and unselected patient groups. An example trial schema is presented in Fig. 1. The adaptive design can result in a reduction of expected sample size compared with the nonadaptive two parallel designs. Both adaptive and nonadaptive designs require specification of appropriate benchmark response rates in each of the biomarker-defined subgroups, which may be difficult to specify for reasons already discussed for the standard single-arm trial in an enriched patient population.

Pusztai et al. (36) proposed a tandem two-step phase II trial design that incorporates a prespecified pharmacogenomic

Table 2. Principles to guide the inclusion of biomarkers in the drug development process

Stage in drug development process	Recommendation
Preclinical	Conduct comprehensive preclinical studies aimed at reliably identifying the drug target and developing a robust biomarker coupled with a sufficiently reliable and reproducible assay. At this point, efforts should be directed at developing an analytically validated assay for the biomarker.
Phase I	Incorporate the biomarker in phase I testing with the goal of better characterizing the biomarker and assay performance in human samples.
Phase II	The biomarker should be incorporated into trials for hypothesis testing. Questions of whether use of the biomarker is likely to inform treatment decisions by identifying populations likely to respond or whether it may be useful as a surrogate or correlative endpoint can be preliminarily assessed. To achieve this, larger and more complex phase II designs, including randomized designs, may be needed.
Phase III	Once the biomarker assay is analytically validated for use on clinical specimens, an indication of efficacy in phase II testing will allow for the design of a phase III trial that will lead to clinical validation and eventually incorporation of the biomarker into clinical use.

predictor of response. The trial schema is presented in Fig. 2. The initial stage of the study is carried out in an unselected group of patients. If sufficient numbers of objective responses are observed in the first stage, then the study continues into the second stage to more fully characterize the response rate in unselected patients. If the number of responses observed in the first stage is not sufficiently high, then the study continues accruing only patients in the subgroup predicted by the pharmacogenomic classifier to be responders; study termination is governed by a standard optimal two-stage phase II trial design in that subgroup of patients predicted to be responders. The investigators stress the importance of having the pharmacogenomic predictor completed specified (fully defined, including cutoff values for defining positive or negative) before the initiation of the trial. It is assumed that this predictor could be derived using cell line models or other preclinical data or perhaps using archived specimens obtained from patients who had received the same or similar treatment. It may be challenging to develop a predictor from these other types of studies that transfers readily and performs well in the phase II trial. Moreover, based on some simulations studies, Puzstai et al. (36) expressed skepticism that pharmacogenomic response biomarkers would be successfully discovered using high-throughput gene expression profiling embedded within a typical phase II study due to small sample size with few responses and the problems of multiplicity and noise associated with high-throughput technologies. Despite the skepticism about discovery of tumor-based biomarkers for predicting response in phase II trials, phase II trials remain a viable setting for early testing of prespecified candidate tumor-based biomarkers for predicting response or for evaluation of germ-line DNA-based markers of toxicity or tumor response suggested by known biological pathways or mechanisms.

Specimen Collection for Biomarker Analyses

Enthusiasm for the incorporation of biomarkers into phase II trials must be tempered by recognition of the many challenges that can be encountered in obtaining appropriate specimens on which to conduct the biomarker assays. Many phase II trials are conducted in patients with advanced or recurrent disease who may have received prior systemic treatment. From a biological perspective, it is not clear whether the most relevant tissue specimen to evaluate is the primary tumor or the recurrent or metastatic lesion. The biomarker characteristics of the primary tumor may differ from the characteristics of a metastatic or recurrent lesion in the same patient, and the biological characteristics of a recurrent tumor could be altered by any systemic therapies the patient received for the original (primary) tumor. It may be difficult to obtain suitable archived tissue from a primary lesion that was collected years earlier and with unknown acquisition and handling procedures. Alternatively, obtaining a new sample of the metastasis or recurrence may require an additional biopsy that could pose some risk and discomfort to the patient. Image-based biomarkers and blood or serum-based biomarkers would have clear advantages in this respect, but much work remains to be done to further develop biomarkers of this type. One should be cautious about mixing

both primary and recurrent or metastatic lesions within the same biomarker study.

Given the difficulty in obtaining specimens, it is imperative that specimens not be squandered on unfocused exploratory studies or be evaluated using assays that are seriously lacking in robustness and reproducibility. If there is substantial uncertainty about the best biomarkers to examine or the most appropriate assay methodology to use, it may be better to bank the specimens for later study. If banking will be pursued, it is desirable to collect, process, and store the specimens in a prespecified and consistent way to increase the chances that the specimens will perform well when assayed, perhaps years later. Recommendations for collection and handling of breast cancer specimens from clinical trials have recently been published (37), but guidelines are lacking for many other cancer types and guidelines may evolve as new assay technologies are introduced. If consistency in specimen collection cannot be achieved, then every effort should be made to document whatever collection, processing, and storage methods were used. It may be unethical to ask patients to undergo risky or uncomfortable procedures to obtain specimens if the specimen collection and handling procedures are not carefully controlled and the studies using those specimens have not been designed to answer scientifically meaningful questions.

Discussion

Over the past decade, biomarkers and targeted therapies have been important to many of the major success stories in cancer treatment (e.g., imatinib mesylate therapy for BCR-ABL-positive chronic myelogenous leukemia and c-KIT-positive gastrointestinal stromal tumors and trastuzumab for HER-2-positive breast cancer). For each of these success stories, however, hundreds of biomarker studies have been done without bearing fruit. A common feature of the success stories is the investment that was made early in the drug development process, including in the preclinical phase, to develop biomarkers that would improve the understanding of the biology of the disease and mechanism of action of the therapeutic agent. Haphazard inclusion of biomarkers into phase II trials is likely to be too late and not very informative. The drug development community will also have to accept that phase II trials may need to be somewhat larger and more complex and more randomized phase II trials may be needed to fully evaluate the potential of biomarkers for their usefulness in the conduct of phase III trials and ultimately for clinical decision-making. For example, to get an early indication of whether a kinase inhibitor may be effective only in patients with a mutated target, one may want to perform a phase II study both in patients with the wild-type target and the mutated target to assess whether there is evidence for differential efficacy. Table 2 summarizes some suggested principles to guide the use of biomarkers in the drug development process. With greater investment and more rational approaches to biomarker research in earlier stages of drug development, greater rewards await at the end.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Adjei AA, Christian M, Ivy P. Novel designs and endpoints for phase II clinical trials. *Clin Cancer Res* 2009;15:1866–72.
2. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89–95.
3. Shankar LK, Van den Abbeele A, Yap J, Benjamin R, Schuetz S, Fitzgerald TJ. Considerations for the use of imaging tools for phase II treatment trials in oncology. *Clin Cancer Res* 2009;15. In press.
4. Park JW, Kerbel RS, Kelloff GJ, et al. Rationale for biomarkers and surrogate endpoints in mechanism-driven oncology drug development. *Clin Cancer Res* 2004;10:3885–96.
5. Clark GM, Zborowski DM, Culbertson JL, et al. Clinical utility of epidermal growth factor receptor expression for selecting patients with advanced non-small cell lung cancer for treatment with erlotinib. *J Thorac Oncol* 2006;1:837–46.
6. Betensky RA, Louis DN, Cairncross JG. Influence of unrecognized molecular heterogeneity on randomized clinical trials. *J Clin Oncol* 2002;20:2495–9.
7. Park JW, Stagg R, Lewis GD, et al. Anti-p185HER2 monoclonal antibodies: biological properties and potential for immunotherapy. In: Dickson RB, Lippman ME, editors. *Genes, oncogenes, and hormones: advances in cellular and molecular biology of breast cancer*. Boston: Kluwer; 1992. p. 193–211.
8. Lewis GD, Figari I, Fendly B, et al. Differential responses of human tumor cell lines to anti-p185HER2 monoclonal antibodies. *Cancer Immunol Immunother* 1993;37:255–7.
9. Baselga J, Tripathy D, Mendelsohn J, et al. Phase II study of weekly intravenous recombinant humanized anti-p185HER2 monoclonal antibody in patients with HER2/*neu* overexpressing metastatic breast cancer. *J Clin Oncol* 1996;14:737–44.
10. Cobleigh MA, Vogel CL, Tripathy D, et al. Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *J Clin Oncol* 1999;17:2639–48.
11. Smith IE. Efficacy and safety of Herceptin in women with metastatic breast cancer: results from pivotal clinical studies. *Anticancer Drugs* 2001;12:S3–10.
12. Paik S, Kim C, Wolmark N. HER2 status and benefit from adjuvant trastuzumab in breast cancer. *N Engl J Med* 2008;358:1409–11.
13. Spano JP, Milano G, Vignot S, et al. Potential predictive markers of response to EGFR-targeted therapies in colorectal cancer. *Crit Rev Oncol Hematol* 2008;66:21–30.
14. Wilhelm SM, Carter C, Tang L, et al. BAY 43-9006 exhibits broad spectrum oral anti-tumor activity and targets the Raf/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res* 2004;64:7099–109.
15. Ratain MJ, Glassman RH. Biomarkers in phase I oncology trials: signal, noise, or expensive distraction? Commentary on Goulart et al. *Clin Cancer Res* 2007;13:6545–8.
16. Korn EL, Albert PS, McShane LM. Assessing surrogates as trial endpoints using mixed models. *Stat Med* 2005;24:163–82.
17. Buyse M, Molensberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000;1:49–67.
18. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med* 1997;16:1965–82.
19. Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998;17:2815–34.
20. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–40.
21. Rubinstein L, Crowley J, Ivy P, LeBlanc M, Sargent D. Randomized phase II designs. *Clin Cancer Res* 2009;15. In press.
22. Dhani N, Tu D, Sargent DJ, Seymour L, Moore MJ. Alternate endpoints for screening phase II studies. *Clin Cancer Res* 2009;15. In press.
23. Bublely GJ, Carducci M, Dahut W, et al. Eligibility and response guidelines for phase II clinical trials in androgen-independent prostate cancer: recommendations from the Prostate-Specific Antigen Working Group. *J Clin Oncol* 1999;17:3461–7.
24. Ramiah R, George DJ, Armstrong AJ. Clinical endpoints for drug development in prostate cancer. *Curr Opin Urol* 2008;18:303–8.
25. Thompson IM, Tangen CM, Kristal AR. Prostate-specific antigen: a misused and maligned prostate cancer biomarker. *J Natl Cancer Inst* 2006;100:1487–8.
26. Scher HI, Halabi S, Tannock I, et al. Design and endpoints of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. *J Clin Oncol* 2008;26:1148–59.
27. Bridgewater JA, Nelstrop AE, Rustin GJ, et al. Comparison of standard and CA-125 response criteria in patients with epithelial ovarian cancer treated with platinum or paclitaxel. *J Clin Oncol* 1999;17:501–8.
28. Rustin GJ, Nelstrop AE, Bentzen SM, Bond SJ, McClean P. Selection of active drugs for ovarian cancer based on CA-125 and standard response rates in phase II trials. *J Clin Oncol* 2000;18:1733–9.
29. Stevenson A, Gore M. Specific keynote: clinical trial design problems in ovarian cancer. *Gynecol Oncol* 2003;88:S117–21.
30. Rustin GJS, Bast RC, Jr., Kelloff GJ, et al. Use of CA-125 in clinical trial evaluation of new therapeutic drugs for ovarian cancer. *Clin Cancer Res* 2004;10:3919–26.
31. DePrimo SE, Bello C. Surrogate biomarkers in evaluating response to anti-angiogenic agents: focus on sunitinib. *Ann Oncol* 2007;18:S11–9.
32. Rubinstein LV, Korn EL, Freidlin B, et al. Design issues of randomized phase II trials and a proposal for phase II screening trials. *J Clin Oncol* 2005;23:7199–206.
33. Booth CM, Calvert AH, Giaccone G, et al. Design and conduct of phase II studies of targeted anticancer therapy: recommendations from the task force on Methodology for the Development of Innovative Cancer Therapies (MDICT). *Eur J Cancer* 2008;44:25–9.
34. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 1989;10:1–10.
35. Jones CL, Holmgren E. An adaptive Simon two-stage design for phase 2 studies of targeted therapies. *Contemp Clin Trials* 2007;28:654–61.
36. Puzstai L, Anderson K, Hess KR. Pharmacogenomic predictor discovery in phase II clinical trials for breast cancer. *Clin Cancer Res* 2007;13:6080–6.
37. Leyland-Jones BR, Ambrosone CB, Bartlett J, et al. Recommendations for collection and handling of specimens from group breast cancer clinical trials. *J Clin Oncol* 2008;26:5638–44.