Reports from the 2010 Clinical and Translational Cancer Research Think Tank Meeting: Design Strategies for Personalized Therapy Trials

Donald A. Berry¹, Roy S. Herbst², and Eric H. Rubin³

Abstract

It has long been evident that cancer is a heterogeneous disease, but only relatively recently have we come to realize the extent of this heterogeneity. No single therapy is effective for every patient with tumors having the same histology. A clinical strategy based on a single-therapy approach results in overtreatment for the majority of patients. Biomarkers can be considered as knives that dissect the disease ever more finely. The future of clinical research will be based on learning whether certain therapies are more appropriate than others for biomarker-defined subsets of patients. Therapies will eventually be tailored to narrow biomarker subsets. The ability to determine which therapies are appropriate for which patients requires information from biological science as well as empirical evidence from clinical trials. Neither is easy to achieve. Here we describe some nascent approaches for designing clinical trials that are biomarker-based and adaptive. Our focus is on adaptive trials that address many questions at once. In a way, these clinical experiments are themselves part of a much larger experiment: learning how (or whether it is possible) to design experiments that match patients in small subsets of disease with therapies that are especially effective and possibly even curative for them. Clin Cancer Res; 18(3); 638–44. ©2012 AACR.

Introduction

Despite the burgeoning number of cancer drugs currently in development, fewer new cancer drugs are being submitted for market approval to the U.S. Food and Drug Administration (FDA). Moreover, the proportion of successful phase III clinical trials in oncology is the lowest among all therapeutic areas (1).

Recognizing the need to build a better foundation for drug development, the FDA initiated its Critical Path Initiative (CPI) in 2004. Its goal was to accelerate the translation of biomedical discoveries into therapies. Among other recommendations, the CPI encouraged the use of innovative trial designs, including adaptive designs. Updating the CPI in 2006, the FDA indicated that it had "uncovered a consensus that the two most important areas for improving medical product development are biomarker development and streamlining clinical trials" (2). These 2 areas are the principal focuses of this article.

Other groups have also encouraged the improvement of clinical trial design strategies. In concert with the FDA’s CPI, in 2005 statisticians at pharmaceutical companies formed an Adaptive Design Working Group under the auspices of the Pharmaceutical Research and Manufacturers of America. The objectives of the group were "to foster and facilitate wider usage and regulatory acceptance of adaptive designs to enhance clinical development" (3). In 2007 the European Medicines Agency issued a "Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design" (4). In 2010 the FDA released a draft guidance entitled "Adaptive Design Clinical Trials for Drugs and Biologics" (5).

In a similar vein, in 2010 a committee of the Institute of Medicine responded to a request from the National Cancer Institute (NCI) by publishing a review of the NCI’s Cooperative Group Program of clinical trials (6). The committee concluded that "[b]etter phase 2 trial designs are needed to more accurately assess which patients benefit from a particular therapy, and thus guide the decisions about whether to move into Phase 3 trials. Improved designs for Phase 3 trials...could lead to faster more accurate conclusions about new therapeutics and in the process reduce costs and conserve resources."

These initiatives reflect wide recognition that traditional approaches to drug development too often fail, and they too often fail in late phases, leading to excessive development costs and duration. Some of these failures are due to ineffective drugs whose ineffectiveness should have been discovered sooner in more-informative, early-phase clinical trials. Another reason for failure is that an effective drug was poorly developed. Moreover, some drugs that eventually prove to be successful spend too much time (and resources)
in clinical trials. Strategies to improve drug development should consider adaptive designs and the use of biomarkers to help guide trials with such designs in personalized therapy trials. The goal of such trials is to identify which therapies are best for which patients while preserving the scientific integrity of the trial. The statistical issues are substantial, as are the logistics and timeliness of biomarker assessment and data flow (3, 7).

In this article we describe 2 personalized therapy trials, known as I-SPY 2 (Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2) and BATTLE (Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination), as well as lessons learned in designing and running these trials. Both of these trials are based on prospective Bayesian designs. The Bayesian approach is ideally suited for building adaptive trials because its basic inferential measures (posterior probabilities of unknown parameters, and predictive probabilities of future observations) can be updated (using Bayes’ rule) as information is accrued in the trial (8–11).

Our description of personalized therapy trials is not comprehensive with respect to either adaptive or Bayesian approaches or the use of biomarkers. Generic descriptions have been published elsewhere (3, 11–13). The 2 examples given here represent a very special kind of adaptive design. It is special in several ways, but perhaps most noticeably in that many therapies, including those involving experimental drugs from different pharmaceutical companies, are compared within the same trial. Quite generally, the adaptive Bayesian approach is most useful for trials that address many issues, including identifying which of many possible therapies are better for which patients.

I-SPY 2

I-SPY 2 involves a randomized phase II screening process in which experimental agents are evaluated in combination with standard neoadjuvant chemotherapy for patients with high-risk primary breast cancer (i.e., those with tumors at least 2.5 cm in size; refs. 14–16). Pharmaceutical companies submit drugs to the trial’s Agent Selection Committee. Experimental arms can be added to the trial at any time, assuming that adequate phase I safety information has been obtained and the overall trial’s accrual rate is sufficient to accommodate additional treatment arms.

The primary endpoint is pathologic complete response (pCR) at the time of surgery, which is a potential path to accelerated marketing approval (17). An agent is evaluated for its effect on pCR and can be graduated from the trial at any time for lack of effect on pCR in every subset of the disease (i.e., a ≤10% predictive probability of phase III success in all 10 biomarker signatures).

Patient screening includes an MRI to establish tumor size at baseline and a biopsy to identify the tumor’s hormone-receptor status (HR), HER-2 status (HER2), and NKI 70-gene profile (NKI; ref. 18). Patients within the HR/HER2/NKI strata are assigned to therapy in an adaptively randomized fashion. The randomization probabilities depend on the performance of the various therapies within the trial in comparison with control (which has a fixed randomization probability of 20%), and in particular for patients in the same stratum as the patient being randomized. Therapies that show a high rate of pCR for such patients have greater randomization probabilities, and thus better-performing therapies can be moved through the process more rapidly.

Figure 1 illustrates the design of I-SPY 2. The patient population in panel A is shown as being heterogeneous. Five experimental arms are shown. (The ongoing trial has 3 experimental arms, with the 3 drugs from different companies and with additional drugs under consideration.) Adaptive randomization is performed within the patient subsets, as indicated above. Panel B shows the hypothetical possibility that experimental arm 2 will graduate with a particular biomarker signature, indicated schematically by the subset of patient population symbols from panel A. Panel C shows the setting in which experimental arms 2, 3, and 5 have moved on from the trial and have been replaced by experimental arms 6 and 7.

Figure 1D shows a configuration of arms that is possible but has not yet been used in I-SPY 2. The panel also suggests the use of other settings and diseases by replacing standard therapy with standard of care (SOC) and indicating progression-free survival, overall survival, and pCR as possible endpoints. The bottom 4 arms constitute a factorial design in which agents C and D plus SOC are compared with SOC alone and in combination. The trial could proceed just as it did with independent arms, but the analysis would exploit the benefits of the factorial design as a subtrial within the bigger trial. The randomization probabilities for the single-arm arms would be down-weighted within subsets of the disease if the combination C+D were shown to be better than both alone within those subsets. This approach could be used in an effort to increase the efficiency of early studies of new therapeutics, which often use separate trials to explore the efficacy of monotherapy and combinations with SOC, sometimes including separate biomarker-defined cohorts.

The standard therapy in I-SPY 2 consists of 12 weekly cycles of paclitaxel followed by 4 cycles of doxorubicin/cyclophosphamide. In the experimental arms, experimental agents are added to standard therapy during the paclitaxel phase of treatment. MRI to assess changes in tumor volume from baseline is conducted at weeks 3 and 12 of the paclitaxel phase. Consistent with the Bayesian approach, randomization and phase III success probabilities are based on all available data, including MRI volume measurements.
for all patients. Week 3 and week 12 measurements for patients undergoing surgery are used to inform a longitudinal statistical model for predicting pCR. This model is used to multiply impute pCR results for those patients who have not yet had surgery but have undergone at least one postrandomization MRI. Longitudinal MRI volume measurements are predictive of outcomes at surgery (19, 20). The predictions are not perfect, but interim MRI measurements are informative and improve the performance of the adaptive design algorithm.

I-SPY 2 is sponsored by the Biomarkers Consortium of the Foundation for the National Institutes of Health (21), a public-private partnership that includes the FDA, the National Institutes of Health, major pharmaceutical companies, and QuantumLeap Healthcare (ref. 22; Figs. 1 and 2).

**BATTLE-1**

Multiple signaling pathways have been implicated in the development and progression of non–small cell lung cancer (NSCLC). Important differences in signaling pathway alterations between chemo-naive and -resistant tumor tissues in patients with advanced NSCLC necessitate molecular examination of the tumor at the time of therapy selection for these patients, rather than the use of data from the original diagnostic biopsy. In the Department of Defense–sponsored BATTLE-1 trial (23), the Thoracic Medical Oncology team at The University of Texas MD Anderson Cancer Center developed a program to obtain fresh tissue biopsies from patients with refractory NSCLC. We employed real-time molecular analyses of those biopsies to guide treatment decisions while continuing to discover new pathways and markers relevant to this disease. These molecular assessments were performed in the thoracic research laboratory and used to guide patient assignments, via a Bayesian adaptive randomization algorithm, to 4 corresponding targeted therapies: erlotinib (EGFR inhibitor), vandetanib (dual EGFR/VEGFR inhibitor), bexarotene + erlotinib (targeting cyclin D1/RXR pathways and EGFR, respectively), and sorafenib (RAF/VEGFR2/PDGFR inhibitor). Based on the most up-to-date patient-derived data elucidating the relationship between biomarker status and treatment outcomes, it was concluded that the adaptive randomization model assigned patients to more-effective treatments with a higher probability depending on the current results in each patient’s biomarker profile. After the patients were enrolled in the trial and provided consent, they underwent a core biopsy of their lung tumor or metastasis for biomarker analysis of 11 prespecified...
biomarkers/marker groups: EGFR, KRAS, and BRAF gene mutation (PCR); EGFR and cyclin D1 copy number (FISH); and 6 proteins by IHC (VEGFR and RXR receptors/cycD1).

Of critical importance, each patient's identification and consent, tissue collection, biomarker analysis, and randomization were all achieved within 14 days of enrollment. The primary endpoint was the 8-week disease control rate, with patients being treated until disease progression or unacceptable toxicity occurred.

From November 30, 2006, to October 28, 2009, the BATTLE-1 trial enrolled 341 patients, 255 of which were randomized to one of the 4 treatments previously listed. The patients were heavily pretreated and many had received multiple therapies for metastatic disease, including prior erlotinib therapy (116 patients, 45%). The mandated biopsies were shown to be feasible and safe, with an 11.5% pneumothorax incidence in patients receiving lung biopsies and, of these, only 1 patient had a grade 3 pneumothorax (no grade 4 or 5). We equally randomized the first 97 patients (24%) into the 4 treatments to acquire sufficient data to inform the statistical model, and then switched to the adaptive randomization phase for the remaining 60% of patients. We calculated the associations between tumor molecular profiles and treatment efficacy and continually updated the results during the trial, which allowed us to increasingly randomize new patients to the most effective treatments for that profile. The overall 8-week disease control rate was 46%.

The major outcomes of this trial were as follows:

- More than 250 patients were biopsied and randomized to one of the 4 treatments in <3 years (an accrual rate of >8 patients/month), and biomarker analyses were completed in our thoracic molecular pathology research laboratory within 2 weeks. The study achieved its primary endpoint (assessment of disease control rate) and showed our ability to successfully complete a large, complex, biopsy-driven clinical trial with mandated fresh tumor biopsies in poor-prognosis NSCLC patients.
- Findings that EGFR mutations were predictive for erlotinib benefit confirmed information emerging at the time of BATTLE-1’s development (2005), and showed the potential of biomarkers to predict patient outcomes after treatment with a targeted agent.
- An unexpected level of benefit was observed in sorafenib-treated patients with both wild-type and mutant KRAS; however, the biologic underpinnings of this activity are unknown (23). Given the lack of response in patients with KRAS-mutated tumors to any targeted agent to date, including sorafenib, these results
warrant further study of sorafenib’s clinical activity and potential markers to identify patients who are most likely to benefit in a more durable fashion.

Building on lessons learned from the BATTLE-1 program, we are currently studying the effects of KRAS on response to inhibitors of its downstream signaling pathways in a prospective, multi-arm, adaptively randomized trial entitled “BATTLE-2 Program: A Biomarker-Integrated Targeted Therapy Study in Previously Treated Patients with Advanced Non–Small Cell Lung Cancer” (BATTLE-2). This is an NCI-funded program (1R01CA155196-01A1) which will be conducted at both MD Anderson and the Yale Cancer Center. In particular, BATTLE-2 will identify new biomarkers that can effectively predict disease control for EGFR-wild-type patients treated with targeted agents. In this trial (see Fig. 3), 400 patients with refractory NSCLC will undergo a mandated fresh biopsy prior to therapy and receive one of 4 treatments [(i) erlotinib (EGFR inhibitor), (ii) erlotinib plus an AKT inhibitor (MK-2206), (ii) MK-2206 plus a MEK inhibitor (AZD6244), and (iv) sorafenib], including combinations that target downstream markers of KRAS-activated pathways. Discovery of new markers and mutations in patients with no known dominant pathway will be guided by Clinical Laboratory Improvement Amendments–certified molecular analyses of their tumor tissue.

The laboratory component of this clinical trial will enable us to identify novel biomarkers to more effectively select patients who might benefit from these therapies. We will use high-throughput sequencing technologies to identify gene mutations in BATTLE-1 and BATTLE-2 tumor tissues. These high-throughput technologies will include analyses of hotspot mutations in 20 known NSCLC-related oncogenes (via Sequenom, Inc.), and the newly developed next-generation (nex-gen) sequencing platforms encompassing whole-genome sequencing (DNA), full transcriptome sequencing (mRNA), and miRNA analysis (24).

Evidence from studies using a panel of molecularly characterized NSCLC cell lines indicates that different KRAS amino acid substitutions may have variable effects on KRAS-activated signaling. We have also identified new compounds that selectively inhibit proliferation of cells with mutant but not wild-type KRAS, and we will explore their mechanism of action (25). Fully annotated clinical data and biopsy samples from our BATTLE-1 NSCLC trial fully support the feasibility and scientific strength of this approach, and can also be used to validate our discoveries. Thus, we will have available >400 tissue (core needle biopsy) and cytology (fine-needle aspiration) specimens collected prospectively from BATTLE-2, as well as clinical and molecular data from both of these unique, biopsy-driven adaptive clinical trial programs (BATTLE-1 and BATTLE-2) to further explore the efficacy of a personalized medicine approach to the treatment of NSCLC, and to better understand and target this critical oncogenic pathway.

Limitations of I-SPY 2 and BATTLE, and Alternatives

If a predictive biomarker is expected to identify a population of patients who will respond to a new therapy with high confidence, the simplest path to development of both the drug and the companion diagnostic test (based on the predictive biomarker) would be to restrict study enrollment...
to the selected population early in development. Notable recent examples of the successful use of this strategy include the development of vemurafenib (BRAF inhibitor) and crizotinib (ALK inhibitor), along with their companion diagnostic tests, in melanoma and NSCLC, respectively. In both of these cases, enrollment was restricted beginning in late phase I studies.

However, a major risk with this strategy is the selection of a predictive biomarker that is predominantly based on preclinical studies. Preclinical models often do not fully recapitulate the clinical setting and can suggest incorrect predictive biomarkers, as in the case of EGFR and IGFR1 protein expression for EGFR- and IGFR-targeting antibodies, respectively (26, 27). With the increasing demands for efficiency in drug development, selection of the wrong biomarker in early studies in which enrollment is restricted can lead to an incorrect "no go" due to an apparent lack of efficacy. Another issue with the restricted-enrollment development strategy is that studies to determine the effect of a new drug in a biomarker-negative population may be delayed or never done, leaving open the question of whether the drug would have shown efficacy in that population. This is exemplified by a study of trastuzumab in combination with chemotherapy in patients with HER2-low breast cancer, which was initiated in 2011, >10 years after the drug was initially approved (ClinicalTrials.gov identifier NCT01275677).

Conversely, the use of an "all-comers" approach early in clinical drug development can also be risky in settings where an early efficacy signal must be obtained before investment is made in large phase II or III studies, especially when the prevalence of the responsive population is low in a historically defined cancer type. For example, ~5% of patients with NSCLC have cancers with an ALK translocation that is associated with responsiveness to crizotinib. Assuming that a 20-patient lung cancer phase 1b cohort is used as a screen for efficacy, the probability of enrolling 3 patients with an ALK translocation is 8%, and the probability of enrolling 2 such patients is 26%. Thus, assuming the drug has little activity in patients who do not have ALK-translocated tumors, the most likely outcome using an all-comers approach in this scenario is that there would be no or 1 response among these 20 patients, which could result in a "no go" for the development of the drug in lung cancer.

The BATTLE and I-SPY approaches have several advantages over the all-comers and restricted-enrollment clinical trial strategies: (i) a single control arm is used for multiple experimental drugs, (ii) there are no screen failures for enrollment based on a specific diagnostic assay, and (iii) each drug is evaluated for efficacy in multiple biomarker-defined subgroups. Thus, the BATTLE and I-SPY approaches have the potential to greatly improve the efficiency of codeveloping a new therapeutic with a matching diagnostic. However, there are limitations to these approaches. First, because patients are assigned to treatments according to the results of biomarker analyses, the biomarker assays must be chosen carefully. Similarly to the restricted-enrollment strategy, selection of the wrong biomarker may lead to an incorrect no-go decision for a given compound.

A related limitation is the selection of the cutpoint for biomarkers that are measured on a continuous scale, which is used to classify patients as positive or negative with regard to the biomarker, and must be determined before the study starts. Considering the limited clinical information that is typically available regarding the relationship between the biomarker and the efficacy of the investigational drug, it may be difficult to select this cutpoint. Setting the cutpoint too high could reduce the ability of adaptive randomization to discriminate among potential treatments for this subpopulation. Conversely, setting the cutpoint too low could dilute the effect size of an investigational treatment in the biomarker-positive group and also decrease the discriminatory ability of the approach.

An alternative adaptive approach that does not require selection of biomarkers or cutpoints before initiation of a study has been proposed but has not yet been implemented in the clinic. Referred to as the adaptive signature design (or the related cross-validated adaptive signature design [28, 29]), this is an adaptive but frequentist approach to testing multiple biomarkers in a relatively large study. The key aspect of this design is that a potential predictive biomarker is identified in a randomly selected training set of enrolled patients, and the remaining patients are used to validate the predictive biomarker. An attractive aspect of this approach is that whole-genome, agnostic methods to identify a predictive biomarker can be used in the test set, which arguably decreases the risk of selecting the wrong biomarker. An additional important aspect of this approach is that analytically validated tests for potential predictive biomarkers are not needed until the time of the final analysis. If it is embraced by regulatory authorities and drug developers, similarly to I-SPY and BATTLE, this approach has the potential to improve the efficiency of codeveloping new therapeutics with matching diagnostics.

**Conclusions**

Adaptive clinical trials that use prospectively assessed biomarkers to assign therapy are feasible. They promise to shorten overall drug development times and help define which patients will benefit from which therapies. We have described some example trials. For reasons indicated above, none of these trials are the final answer to personalizing cancer medicine. They are themselves part of an experiment. It is an essential experiment to help us understand how, or whether, we can build complicated and yet informative and efficient clinical trials that match biomarker subsets with therapies.

**Disclosure of Potential Conflicts of Interest**

D.A. Berry is part owner of Berry Consultants, LLC, a company that designs adaptive clinical trials for pharmaceutical and medical device companies. E.H. Rubin is an employee of Merck & Co., Inc. R.S. Herbst disclosed no potential conflicts of interest.

Received October 26, 2011; revised December 14, 2011; accepted December 14, 2011; published online February 1, 2012.
References


5. Food and Drug Administration [homepage on the Internet]. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceR-


Reports from the 2010 Clinical and Translational Cancer Research Think Tank Meeting: Design Strategies for Personalized Therapy Trials

Donald A. Berry, Roy S. Herbst and Eric H. Rubin


Updated version  Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/18/3/638

Cited articles  This article cites 14 articles, 4 of which you can access for free at: http://clincancerres.aacrjournals.org/content/18/3/638.full#ref-list-1

Citing articles  This article has been cited by 9 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/18/3/638.full#related-urls

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, use this link http://clincancerres.aacrjournals.org/content/18/3/638.
   Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.