Development and Use of Integral Assays in Clinical Trials

Richard L. Schilsky1,2, James H. Doroshow3, Michael LeBlanc4, and Barbara A. Conley3

Abstract

Clinical trials that include integral biomarkers to determine eligibility, assign treatment, or assess outcome must employ robust assays to measure the molecular analyte of interest. The decision to develop a biomarker assay into a test suitable for use in humans should be driven by clinical need, that is, there should be a clear clinical purpose for undertaking the test development. Supporting in vitro or in vivo research on the ability of the marker to distinguish subgroups of patients with a given characteristic is necessary. The magnitude of the difference in treatment effect expected with use of the marker should be sufficient to support differential treatment prescription for marker-positive and -negative patients. Analytical and clinical validation of the marker assay should be completed before the clinical trial is initiated to ensure that the assay is stable enough for clinical use throughout the trial. Clinical use of the assay requires that it be performed in a Clinical Laboratory Improvement Amendments–accredited laboratory, and the need to apply for an Investigational Device Exemption from the U.S. Food and Drug Administration should be considered. In this article we elaborate on the steps required to get a biomarker assay ready for use as an integral component of a clinical trial and give an example of the use of an integral assay in a phase III trial.

Clin Cancer Res; 18(6); 1540–6. ©2012 AACR.

Introduction

The elucidation of molecular signaling pathways that contribute to cancer initiation, progression, and metastasis, and the development of drugs targeted to interrupt these pathways, present the opportunity to conduct clinical trials that use these molecular features to select patients, assign treatment, or assess treatment outcomes. This will potentially allow the design of clinical trials with larger effect sizes, smaller sample sizes, and faster completion times. Current research efforts are focused on the characteristics of tumors and the patient host that can provide insights into which treatment, even with standard therapy, is likely to benefit or harm a particular patient the most. Integral assays are assays that must be performed before the clinical trial can proceed. The results of these assays may inform about eligibility, stratification, or treatment assignment (see Box 1). They are distinct from "proof of mechanism" assays, which are research assays that assess whether a drug has affected its target. They are also distinct from integrated assays, which may be performed on all or most patients in a clinical trial. The results of such assays will not influence the treatment of patients in the trial; rather, they will be used to inform further research or clinical trials. Other types of markers may also be the subject of clinical research (1). In this article we discuss the analytical, clinical, and regulatory issues that must be considered in the development of integral assays that will be used in a clinical trial to select patients, assign treatment, stratify patients, or assess outcomes. Relevant definitions of terms can be found in Table 1.

Clinical Need

The development of any biomarker into an assay for use in humans, particularly an integral assay in a clinical trial, should be driven by the clinical need. For example, will use of the assay result in better treatment outcomes than could be achieved without it? Our goal is to maximize the chance that a patient will benefit from the treatment and minimize the chance that he or she will not benefit. The choice of an integral assay in a clinical trial will vary depending on the objectives of the trial, the types of biospecimens to be studied, the analyte of interest, and the complexity and analytical performance of the assay technology. Selection of an integral marker often relies on previous research findings, which have rarely been obtained with the goal of developing an assay that meets the requirements for clinical utility. Thus, when the use of the assay to direct patient therapy is contemplated, a clinical assay must be developed, often under substantial time constraints, so that it is ready when the clinical trial begins accruing patients.

Before committing to the development of an assay for use in directing therapy, investigators should ensure that...
previous research has demonstrated several qualities of the molecular marker. An example of a marker/assay that might not be useful is one that predicts that marker-positive patients will have a survival of 6 months, whereas marker-negative patients will have a survival of 5 months. In addition, if both marker-positive and marker-negative patients benefit from a new treatment but the magnitude of benefit is smaller in the marker-negative patients, a clinician would not be likely to withhold the new treatment from the marker-negative patients. Ideally, before development of a clinical-grade assay is undertaken, a research-grade assay that works in the types of human tissues to be used in the trial should be available.

**Analytical Validation**

Analytical and clinical validations should be completed before patient accrual begins, to finalize and standardize the performance of the assay (i.e., "lock down" the assay) so that its use in the proposed clinical trial will provide reliable results. Analytical validation involves documentation that the performance characteristics of the assay are reliable and suitable for the intended clinical use in the specimens of interest. The desired characteristics of the assay must be defined. For example, what are the clinical implications of "wrong" results, such as missing the abnormality when it is present (false negative), or detecting the abnormality when it is absent (false positive)? The desired sensitivity and specificity of the assay must be defined. The desired limit of detection must also be defined. If only 1% of cells possess the analyte, will the assay be able to detect this level? Is this level of detection necessary for the intended use? Specific acceptance criteria must be set in advance. One must also define preanalytic variables that might affect the performance of the assay, such as how much tumor tissue (or other biospecimen) is needed, how the sample should be collected (e.g., needle aspirate vs. core vs. open biopsy), how the tissue should be handled and shipped, and how the specimen should be preserved. Clinical trials are often conducted at multiple institutions, possibly internationally, and over several years of patient accrual; therefore, during the validation procedure, the effects of storage time and storage conditions must be determined if these factors are relevant to the trial. Please see the article by Hewitt and colleagues (3) in this CCR Focus section for further discussion about the role of preanalytic factors.

The procedures to be followed for analytical validation depend on the technology platform. However, the assay must meet specifications for precision, accuracy, selectivity (i.e., quantification of the analyte in the presence of other substances), elucidation of any interfering substances, and stability of the analytes. Accuracy is defined as the closeness of the mean test result to the true result. Precision refers to how often the assay gives the same result when samples are repeatedly analyzed (e.g., within run, within day, between run, or between days). Precision estimates may involve different operators, equipment, reagents, or laboratories. If

---

**Box 1. Characteristics of integral markers**

- Integral markers are required for the trial to proceed and are used for medical decision making.
- Integral markers are used
  - to determine patient eligibility (e.g., somatic mutation for a kinase inhibitor)
  - to assign patients to treatment [e.g., FLT3-ITD for risk assessment in pediatric trials with consequent assignment to specific treatment (17)]
  - for risk stratification if such stratification leads to different treatments (as in the FLT3-ITD example)
  - for risk classification (as in the FLT3-ITD example)
- Because the assay is performed for medical decision making, and the patient or physician is informed of the result, a Clinical Laboratory Improvement Amendments–certified laboratory is generally required for assay.
- Integrated markers are performed on all or a statistical subset of patients, but are not used for medical decision making.
- Research markers include all other assays; often referred to as correlative research.

*Integral Assays*
Clinical Validation

Clinical validation refers to how well the assay result relates to the clinical outcome of interest. In initial attempts to achieve clinical validation, investigators may use a research assay or "samples of convenience" (e.g., a collection of samples from the tumor of interest that may or may not have come from a clinical trial). Such samples may have biases that affect interpretation of the clinical utility of the assay (e.g., methods of specimen collection, differences in treatment, drug dosing, comorbidities, performance status, and methods and timing of tumor assessment). When use in a clinical trial is contemplated, a marker clinical validation study requires a defined validation protocol with a specified analysis plan. As part of that plan, the assessment of performance should be focused on a completely defined specified biomarker, and the assay should be locked down. If the assay involves combinations of features, all parameters should be specified prior to the evaluation. It is possible to clinically validate an assay in a prospective-retrospective fashion if an adequate number of samples can be obtained from relevant clinical trials that have already been completed and for which the outcome data are available (12). The
assay must be performed blinded with respect to patient outcome. Samples from phase II trials of an agent could be useful for defining the correlation of the assay result with the outcome of interest. However, if the phase II study does not include an appropriate control group, it may not be possible to assess whether the marker is predictive of treatment efficacy or is only prognostic. Even a randomized phase II study typically has only modest power to determine whether a marker is predictive (i.e., tests an interaction) or to identify the marker-defined subgroup associated with greatest treatment efficacy.

An evaluation of clinical validity will show the variability of the biomarker in the intended population and in a population without the disease of interest.

Assessment of Clinical Utility

Clinical utility refers to how useful the assay is for directing treatment. This can be determined with a prospective clinical trial or with more than one prospective-retrospective trial (12, 13). For this assessment, the assay should be analytically and clinically validated (locked down). Analysis of more than one retrospective dataset will allow refinement of the cut points, if they are necessary. A prospective clinical trial (see example below) will also allow refinement of the cut points for the assay. Clinical utility is generally the goal of trials using integral assays to guide treatment or to select patients for a given treatment.

Although typically investigators will conduct null hypothesis significance tests, evidence of a sufficiently strong effect for the assay to be of clinical utility is more relevant for the future use of the methodology. If there is substantial uncertainty from prospective-retrospective analyses of the magnitude of the marker’s predictiveness with respect to treatment efficacy, a prospective randomized trial that concurrently assesses the activity of a new drug and the biomarker should be considered (12, 14). The clinical protocol for the prospective evaluation must include the important aspects related to use of the integral assay, including the impact on patient eligibility, the statistical design (including implications for power and type I error control if subgroup analyses are involved), trial futility monitoring, and logistical details such as sample preparation, submission details or timing, and transmission of assay results. Various trial designs have been proposed for this purpose, including marker-stratified designs (14) and Bayesian adaptive designs (15). In marker-stratified designs, the marker is measured in all of the patients. The patients are then stratified into marker-negative or -positive groups, and each group is randomized to receive either the experimental or the standard therapy. A particular strength of this design is the possibility for retrospective assessment of a different cut point or potentially different biomarkers. Bayesian designs begin with an initial model of how the assay will perform relative to the primary outcome, and the model is subsequently updated with accruing data. These designs are being used particularly in earlier therapeutic development clinical trials, and their efficiency relative to other designs is under study. Although enrichment designs that include only a defined marker group can be an efficient strategy for assessing a new drug or therapy in a biomarker-positive subgroup, such designs do not allow a full assessment of the performance of the marker (i.e., the drug’s effect on biomarker-negative patients).

Legal, ethical, and scientific standards

An integral assay or test must be performed in a Clinical Laboratory Improvement Amendments (CLIA)–accredited laboratory. In addition, integral markers in clinical trials are often considered to pose a significant risk to the health of patients. The U.S. Food and Drug Administration (FDA) recently increased its surveillance of the risk posed by integral markers in trials, with risk defined in the Federal Code of Regulations as a marker and its assay “for a use of substantial importance in diagnosing, curing, mitigating, or treating disease, or otherwise preventing impairment of human health and presents a potential for serious risk to the health, safety, or welfare of a subject” (16). Thus, many integral markers are considered to pose a significant risk, and, as discussed by Meshinchi and colleagues (17) elsewhere in this Focus section, it is wise to contact the FDA early in the protocol development process and submit an application for a pre-Investigational Device Exemption (pre-IDE). If the FDA indicates that a full IDE is needed, there are various ways to facilitate the IDE application (17).

Example

The increased expression of eicosanoids in NSCLC has been associated with adverse prognosis (18). The combination of cyclooxygenase 2 (COX-2) inhibitors with lipoxigenase (LOX) inhibitors or with chemotherapy was shown to enhance cytotoxicity in vitro (19). Cancer and Leukemia Group B (CALGB) conducted a randomized phase II study for patients with stage IIIb (pleural effusion) or stage IV chemotherapy-naïve NSCLC in which patients received standard cytotoxic chemotherapy (carboplatin and gemcitabine) with either celecoxib (a COX-2 inhibitor), zileuton (an antiasthmatic drug that also inhibits 5-LOX), or both. Although none of the 3 arms met the primary study endpoint of 50% failure-free survival at 9 months, an exploratory analysis showed that administration of celecoxib plus zileuton with chemotherapy resulted in longer failure-free survival than the other 2 arms. The difference was of borderline significance ($P = 0.054$) when adjusted for known prognostic factors (e.g., stage and performance status). Because tumor tissue and serum samples were collected in this study, a subsequent retrospective-prospective analysis was able to show that patients whose tumor overexpressed COX-2 had a worse prognosis, and patients whose tumor overexpressed COX-2 and who received celecoxib (with or without zileuton) had superior survival compared with patients who did not receive celecoxib. The test for interaction between treatment (celecoxib) and marker (COX-2 index ≥ 4) was significant (20). This
provocative finding led to the design of a phase III trial, CALGB-30801 (A Randomized Phase III Double Blind Trial Evaluating Selective Cox-2 Inhibition in Cox-2 Expressing Advanced Non–Small Cell Lung Cancer), in which COX-2 expression will be used as an integral biomarker to select patients for enrollment. This prospective trial will begin with an enrichment design (COX-2 ≥ 2) and then randomize patients to standard treatment with or without a COX-2 inhibitor. This trial will serve to confirm or refute an interaction between the expression of COX-2 in tumors and benefit from the addition of a COX-2 inhibitor to standard treatment, and will allow exploration of the cut point for the assay (Fig. 1). The primary objective of CALGB-30801 is to evaluate the benefit of COX-2 inhibition combined with chemotherapy (arm A) in comparison with chemotherapy only (arm B) in patients with advanced NSCLC whose tumors overexpress COX-2 (defined as COX-2 index ≥ 4). The trial investigators also plan to evaluate the survival benefit of celecoxib in patients with a lower level of COX-2 expression (COX-2 index < 2).

All potentially eligible patients are preregistered. The CALGB Molecular Pathology Reference Laboratory will perform all COX-2 assays for the trial. The laboratory is CLIA-certified and accredited by the College of American Pathologists. At the time the trial was initiated, the FDA was not requiring a pre-IDE review for such National Cancer Institute–supported studies.

COX-2 is measured by immunohistochemistry [IHC (20)]. Paraffin-embedded tissue sections (4 mm) are heated at 60°C, cooled, deparaffinized, and rehydrated through xylene and graded alcohol solutions to water. Endogenous peroxide is blocked with 3% hydrogen peroxide in water, and antigen retrieval is performed with the use of Target Retrieval Solution (Dako) and steaming. Automated staining with antibody to COX-2, clone SP21 (Labvision Corp.) diluted 1:50 for 1 hour, is performed at room temperature. The detection system is a labeled streptavidin-biotin complex. After IHC is performed, images of the slides and the pathology reports are transferred to the reading pathologist by means of a virtual microscopy system. All slides are reviewed without knowledge of the patient’s history. Scoring for COX-2 expression, established in the previous phase II trial, is semiquantitative/ordered categorical. The neoplastic cells are scored for intensity (range of scores, 0–3) and percentage of cells stained [0 (0%), 1 (1%–9%), 2 (10%–49%), 3 (50%–100%)]. An IHC index (range of scores, 0–9) is defined as the product of the intensity and percentage of cells staining. Two cut points (COX-2 index ≥ 2 and ≥ 4) will be evaluated in this trial. Three expert pathologist members of the CALGB Pathology Committee will score each case using a digital slide scanning system. Results will be sent within 72 hours of receipt of the specimen to the CALGB Statistical Center, and the institution will be notified of the patient’s eligibility status within 72 hours. Once COX-2 testing is completed, patients with an index of ≥ 2 will be registered and randomized with equal allocation to chemotherapy with celecoxib or placebo. Patients with a score of < 2 will not be registered and will be treated at the discretion of their physician.

Assay Validation

As mentioned above, an assay must be locked down before it can be used as an integral assay in a clinical trial. Although the College of American Pathologists has established some proficiency standards for IHC assays, no such standards, or any standard or certified reference material, are yet available for the COX-2 assay (4). Therefore, CALGB set up proficiency testing with a small set of cases that were
Integral Assays

tested by 2 different laboratories: the CALGB Molecular Pathology Reference Laboratory and the Ohio State University clinical IHC laboratory.

Accuracy is generally determined by comparing the measurement of the analyte against the true or accepted value. Because there is currently no certified quality control standard for measuring COX-2 protein in tissue, investigators determined the accuracy of the COX-2 IHC assay by showing that the monoclonal rabbit anti-human COX-2 antibody that was used for the assay was specific for the antigen as it was used to identify COX-2 in both Western blotting and precipitation of COX-2 as determined by high-performance liquid chromatography.

CALGB performed 2 forms of precision testing. First, a set of samples with a range of IHC indices were stained on consecutive days and analyzed by the study pathologists to determine the day-to-day variability and the consistency of performance. Second, a set of 26 specimens were reread by 2 pathologists. A linear model was fitted with the measure from one pathologist as the outcome variable and that of the other pathologist as the explanatory variable. The 95% confidence interval (CI) for the intercept estimate contains 0, and that of the slope estimate contains 1. The joint test for the intercept = 0 and the slope = 1 was not statistically significant (P = 0.5136). Similar findings held for COX-2 intensity and COX-2% positive cells. The intraclass correlation was 0.81260 for COX-2 index, 0.81308 for intensity, and 0.67726 for percentage. After the COX-2 index was dichotomized at 2, the agreement between the pathologists as measured by the k coefficient (0.7797; 95% CI, 0.3666–1.000) was good. Despite a wide confidence interval, these findings suggest that the reading of COX-2 expression is reproducible across different pathologists. COX-2 expression readings of 2 distinct runs on a tumor microarray of samples from 24 patients with NSCLC and 36 patients with colon cancer also revealed similar reproducibility and accuracy values.

Controls (assayed with each study specimen) include lung tissue typed as negative or positive, and a nonspecific matched isotype antibody used in place of the primary antibody.

If an assay yields uninterpretable results due to failed quality control, tissue staining, or loss or distortion of the tissue from the slide, the assay is repeated. If the assay remains uninterpretable after a second attempt, the pathology readout will indicate this, and the patient will not be enrolled.

Conclusions

Integral biomarker assays will likely become more common in future phase II and III clinical trials. In clinical practice, tests are already being performed fairly commonly for molecular abnormalities in single genes or for copy number alterations in breast, lung, colon, melanoma, and other tumors [e.g., estrogen and progesterone receptor expression; expression or amplification of Her2/neu, KRAS mutations, epidermal growth factor receptor (EGFR) mutations, BRAF V600E mutation; and expression of c-KIT and BCR/ABL, PML/RAR, and EML4-ALK translocations]. Several cancer centers and some companies are sequencing tumor DNA to identify actionable targets. The development of a powerful technology that can perform massively parallel (next-generation) sequencing in less time has brought the prospect of whole-genome and whole-exome sequencing to the clinical realm, along with the possibility of identifying mutated genes that may be targets for therapy, or polymorphisms that may portend a greater risk for toxicity. The pace of technology development promises additional capabilities to molecularly characterize tumors in the near future. Thus, it will continue to be essential to assess the clinical utility of new putative biomarkers.

When the use of an integral assay is contemplated, there should be a strong biologic rationale and/or a correlation with tumor biology or clinical outcome. The assay should be analytically validated in a CLIA-accredited laboratory, in a manner adequate to support its use in a clinical trial. Serious consideration must be given to obtaining an IDE and consulting with the FDA (for trials conducted in the United States). Currently, not all trials with integral assays have met these criteria. However, if we are to attain the benefits of molecularly driven cancer treatment, we must invest in the development of assays that are robust and validated and provide clinically useful information.

Disclosure of Potential Conflicts of Interest

R.L. Schilsky has an ownership interest (including patents) in Universal Oncology and Foundation Medicine, is a consultant to and serves on the advisory board of Foundation Medicine, and receives compensation from Roche and GlaxoSmithKline. No other potential conflicts of interest were disclosed.

Received November 22, 2011; revised January 21, 2012; accepted January 27, 2012; published online March 15, 2012.
16. Medical Devices: Title 21 Code of Federal Regulations Pt 812.3 [21CFR812.3 (revised April 1, 2011)].