Pharmaceutical Research, Vol. 22, No. 4, April 2005 (© 2005)
10.1007/s11095-005-2495-9

Conference Report

Method Validation and Measurement of Biomarkers in Nonclinical and Clinical Samples in Drug Development: A Conference Report

Jean W. Lee,[1,17] Russ S. Weiner,[2] Jeff M. Sailstad,[3] Ronald R. Bowsher,[4] Dean W. Knuth,[5] Peter J. O'Brien,[6] Jean L. Fourcroy,[7] Rakesh Dixit,[8] Lini Pandite,[9] Robert G. Pietrusko,[10] Holly D. Soares,[11] Valerie Quarmby,[12] Ole L. Vesterqvist,[2] David M. Potter,[11] James L. Witliff,[13] Herbert A. Fritche,[14] Timothy O'Leary,[15] Lorah Perlee,[16] Sunil Kadam,[6] and John A. Wagner[8]

Received November 15, 2004; accepted January 14, 2005

Biomarkers are increasingly used in drug development to aid scientific and clinical decisions regarding the progress of candidate and marketed therapeutics. Biomarkers can improve the understanding of diseases as well as therapeutic and off-target effects of drugs. Early implementation of biomarker strategies thus promises to reduce costs and time-to-market as drugs proceed through increasingly costly and complex clinical development programs. The 2003 American Association of Pharmaceutical Sciences/Clinical Ligand Assay Society Biomarkers Workshop (Salt Lake City, UT, USA, October 24–25, 2003) addressed key issues in biomarker research, with an emphasis on the validation and implementation of biochemical biomarker assays, covering from preclinical discovery of efficacy and toxicity biomarkers through clinical and postmarketing implementation. This summary report of the workshop focuses on the major issues discussed during presentations and open forums and noted consensus achieved among the participants on topics from nomenclature to best practices. For example, it was agreed that because reliable and accurate data provide the basis for sound decision making, biomarker assays must be validated in a manner that enables the creation of such data. The nature of biomarker measurements often precludes direct application of regulatory guidelines established for clinical diagnostics or drug bioanalysis, and future guidance on biomarker assay validation should therefore be adaptable enough that validation criteria do not stifle creative biomarker solutions.

KEY WORDS: biomarkers; nonclinical and clinical drug development; quantitative method development and validation.

[1] MDS Pharma Services, Lincoln, Nebraska, USA.
[2] Bristol-Myers Squibb, Princeton, New Jersey, USA.
[3] Trimeris Inc., Durham, North Carolina, USA.
[4] LINCO Diagnostic Services, St. Charles, Missouri, USA.
[5] Jasper Clinic, Kalamazoo, Michigan, USA.
[6] Eli Lilly and Company, Indianapolis, Indiana, USA.
[7] Walter Reed Army Hospital, Bethesda, Maryland, USA.
[8] Merck and Company, Inc., West Point, Pennsylvania, USA.
[9] GlaxoSmithKline, Research Triangle Park, North Carolina, USA.
[10] Millenium Pharmaceuticals, Cambridge, Massachusetts, USA.
[11] Pfizer Global Research, Groton, Connecticut, USA.
[12] Genetech, Inc., South San Francisco, California, USA.
[13] University of Louisville, Louisville, Kentucky, USA.
[14] MD Anderson Hospital, University of Texas, Austin, Texas, USA.
[15] Food and Drug Administration, Rockville, Maryland, USA.
[16] Molecular Staging Inc., New Haven, Connecticut, USA.
[17] To whom correspondence should be addressed. (e-mail: jean.lee@mdsinc.com)

ABBREVIATIONS: AAPS, American Association of Pharmaceutical Sciences; BQL, below quantifiable limit; CDER, Center for Drug Evaluation and Research; CMS, Centers for Medicare & Medicaid Services; CLAS, Clinical Ligand Assay Society; CLIA, Clinical Lab Improvement Amendments; GLP, good laboratory practices; LBABFG, Ligand Binding Assay Bioanalytical Focus Group; LOD, lower limit of detection; LLOQ, lower limit of quantification; NCCLS, National Committee for Clinical Laboratory Standards; PD, pharmacodynamic; PK, pharmacokinetic; OIVD, Office of *in Vitro* Diagnostics Device Evaluation and Safety; QCs, quality controls; ULOQ, upper limit of quantification.

INTRODUCTION

Biomarkers[0] are widely used in drug development, contributing to increased productivity and improved patient care across a spectrum of drug development activities. The recent explosion in biomarker research is largely driven by the widespread belief that appropriate application of biomarkers to preclinical and clinical drug development will accelerate the process (i.e., speeding time to market), increase efficiency (by providing early indications of efficacy or toxicity), and facilitate dose selection prior to expensive phase III clinical trials (1–5). Studies of efficacy and toxicity biomarkers can aid risk assessment and candidate selection in animal models and can improve clinical trials outcomes through improved patient stratification and dose selection. Biomarkers thus seem likely to hasten the attrition of undesirable candidate compounds and to increase productivity in drug development, reducing the time and costs associated with bringing new therapies to patients.

Given the intense interest in biomarkers in the pharmaceutical and diagnostics industries, it is not surprising that biomarkers have become a focus of major new governmental initiatives. Biomarkers are included in the NIH Road Map (6), which aims to facilitate a more efficient and productive system of medical research. Also, the FDA Critical Path Initiative, designed in part to identify the sources of major delays and other problems in drug development (7), cites the need

0724-8741/05/0400-0499/0 © 2005 Springer Science+Business Media, Inc.

for biomarkers in "toolkits" for new product development. The use of biomarkers extends to clinical practice as well, with recent examples of drug development biomarkers that have evolved into diagnostics suitable for use in patient care.

In October 2003, a workshop was held in Salt Lake City, Utah, to address the analytical validation of biomarker assays in support of drug development. Reflective of the close relationship between drug development biomarkers and clinical diagnostic assays, the workshop was organized by the American Association of Pharmaceutical Sciences (AAPS) Ligand Binding Assay Bioanalytical Focus Group (LBABFG) and cosponsored by Clinical Ligand Assay Society (CLAS). Approximately 150 scientists participated in the workshop, which included presentations on biomarker research in drug development and diagnostic applications. These presentations were grouped by topics, and sessions were followed by open-forum discussions to identify commonly encountered issues in biomarker research and development and to obtain consensus on approaches to this research wherever possible. Workshop presentations and subsequent discussions provided the context for the assay validation recommendations and discussions of "best practices" below.

This summary report of the workshop is organized around the larger themes identified at the workshop and focuses on the essence of the bioanalytical issues and relevant discussions held there. Because a comprehensive treatment of all the materials presented at the workshop is beyond the scope of this report, the authors aimed to capture the spirit of "state-of-the-art" biomarker method development and validation.

Understanding the process of biomarker discovery, implementation, and qualification in the pharmaceutical industry provides the context under which biomarker assay validation occurs. Common practices presented and discussed at the workshop are reviewed below to help provide that context. Given that the majority of workshop participants had backgrounds in clinical and preclinical bioanalytical assay development in the pharmaceutical industry and that only a few of the participants were from the diagnostics industry, both the workshop and this report emphasize the use of biomarkers in the clinical phase of drug development.

NOMENCLATURE AND DEFINITIONS

Achieving consensus on clearly defined terminologies is an essential prerequisite to understanding biomarker data in a proper clinical context. Yet, as indicated by several speakers and discussion participants, overlapping and ambiguous terminologies surrounding biomarker research has led to confusion among scientists from different disciplines. Several important government and trade organization initiatives have addressed the issue of nomenclature in drug development (8,9). There was a consensus among workshop participants to adopt the NIH Working Group definition for a biomarker (9). Additional nomenclature was recommended to avoid future confusion in the biomarker literature.

By definition, a biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological or pathologic processes or of pharmacologic responses to a therapeutic intervention. Pharmacodynamic (PD) biomarkers specifically refer to time-associated measures of a pharmacologic response. A clinical end-point differs from a surrogate end-point in that the former quantifies a characteristic of the patient's condition (e.g., how they feel or function, or the survival rate of a population), whereas the latter is meant to substitute for a clinical end-point. Surrogate end-points are thus based on scientific observations that are expected to predict a clinical benefit or lack thereof in a drug trial. Biomarker biology often fits into the cascade of pathologic events that underlie a disease and can, with adequate clinical evidence, eventually serve as surrogate end-points for disease progression. Biomarkers that fall short of surrogacy can nonetheless provide measures of safety, tolerability, and drug efficacy or may be used to monitor disease progression. Two categories of this type of biomarker were defined.

Mechanism-specific, "proximal" biomarkers are a subset of PD biomarkers that reflect drug action. Qualification of proximal biomarkers for safety and efficacy requires extensive knowledge of both the underlying pathophysiology and the on-target and off-target mechanisms of action of the investigative drug. "Distal" biomarkers reflect disease progression. Additional nomenclature included routine biomarkers, which were defined as those that are analyzed in laboratories with well-established methods, such as in the field of diagnostic clinical chemistry. More recently, novel biomarkers have emerged as an important research tool in drug development. Measures of such novel biomarkers are usually conducted in research laboratories and require specialized reagents or technologies that are not available routinely in a clinical laboratory setting. Because of the perceived value of novel biomarkers in facilitating the discovery and development of new and innovative therapies, extensive discussions focused on this type of analysis.

An ideal novel biomarker supporting drug development would be biologically and clinically relevant, analytically sound, operationally practical, timely, and interpretable and cost effective (10). The definitions described above may apply to a number of different biomarkers, including those acquired using physiologic, psychological, and imaging techniques. The biomarkers discussed at the 2003 AAPS/CLAS workshop and described in this report, however, are biochemical measures typically measured in a laboratory setting. The consensus and recommendations below are limited to such assays.

INDUSTRY TRENDS IN BIOMARKER DEVELOPMENT

Preclinical Phases

Given the evolving and dynamic nature of biomarker research projects, where internal and external influences on decision-making can rapidly change, what strategies can be used to see biomarkers through validation and implementation? Representatives from innovator companies described a common pattern of preclinical biomarker implementation. Ideally, biomarker research contributes to regulatory approval. However, considering the high attrition rate of candidate compounds, a minimal goal of this exercise is often to provide feedback to discovery scientists regarding the performance of predictive biomarkers, thus allowing improvements in subsequent generations of compounds and improvements in future study designs.

In order for novel biomarkers to be used to their fullest extent, they must be identified early in the drug discovery process. This allows the bioanalytical methods to be developed and validated in time for early phase clinical trials. Because study objectives can often include disparate factors, an interdisciplinary approach is often adopted for effective development of biomarkers. Some companies have established biomarker working groups that are composed of scientists and other relevant personnel from discovery, pharmacology, absorption/distribution/metabolism/elimination (ADME), medical, regulatory, and analytical divisions. Team discussions and effective communication are vital to the timely, successful development, validation, and application of biomarker methods to sample analysis. The timing of assay development and validation relative to drug development milestones, operational and logistical issues including preanalytical factors, and limitations in the interpretation of data were identified as crucial topics of discussion for biomarker teams. This integrated approach is a well-accepted model, and most innovator companies are using the following approach or a close variant for biomarker development.

Following target validation and drug candidate selection, typically using *in vitro* models, animal models help to predict the *in vivo* activity of the candidate compound. If the compound exhibits suitable activity with minimal if any toxicities, attempts are made to identify efficacy biomarkers that can serve as early predictors of clinical outcome. These can include novel biochemical or pharmacogenomic markers that may be predictive of disease susceptibility, drug metabolism, or other factors influencing efficacy and other markers for patient stratification. During preclinical development, discovery organizations examine disease-causing and interventional mechanisms to support the selection and use of each biomarker. The biomarker may then be evaluated in pharmacology and toxicology studies, often including animal models of human disease. PD biomarkers may provide early mechanistic or efficacy information about an investigative compound, aiding internal “go–no-go decisions” thereby allowing a team to move forward with added confidence (1–4). Panels of these biomarkers are sometimes developed, including a subset of markers that serve as early indicators of PD efficacy; that is, if the drug has hit its intended target.

Another important area is the identification of biomarkers that can predict toxicity, also termed safety or toxicity biomarkers. The initial development of these markers often occurs during preclinical development. Historically, nonclinical toxicity biomarkers have focused on those that could be measured by well-established clinical biochemical methods. Novel toxicity biomarkers are typically linked to the mechanism(s) of action of a drug and off-target side effects that can sometimes be predicted in light of that pharmacology. Such novel toxicity biomarkers can provide important, compound-specific information regarding toxic drug side-effects that might otherwise not be detected using standard assays. Such information can aid candidate selection and the refinement of next-generation drugs through rational drug design. Once identified, a mechanism-related toxicity biomarker can be treated like an efficacy biomarker and ideally complements standard clinical measures. Such a marker must therefore undergo both bioanalytical method validation and nonclinical qualification in experimental toxicology studies.

Preapproval Clinical Phases

In the process of validating the biomarker assay, samples from healthy individuals and patients are used to establish expected ranges and biologic variations for a given analyte. Knowledge of these ranges and assay performance in these matrices aids the design of studies, thereby allowing researchers to understand the sample size necessary to confidently detect relevant changes in a biomarker after therapeutic intervention. Successful application of biomarkers in early-phase clinical trials requires the choice of a suitable biomarker and an accurate definition of the role that marker will play in subsequent decision-making. A suitable biomarker may reflect a dose-proportional response to therapy but may be too far downstream to adequately model the pharmacokinetic/pharmacodynamic (PK/PD) relationships that underlie the measured response. For example, biomarkers can be used to facilitate dose selection in pivotal trials (5). In ideal situations, a biomarker will also provide predictive evidence that disease processes have been beneficially impacted. This is particularly desirable in chronic diseases, where clinical outcomes may take years to present. As with preclinical biomarkers, workshop participants shared experiences surrounding the implementation of biomarkers in clinical trials, including trends surrounding successes and difficulties.

The makeup and focus of biomarker teams will likely need to evolve in response to clinical and other influences that may necessitate the modification or refinement of a biomarker assay during clinical development. In light of new developments, careful consideration of the continued relevance of a biomarker or its measurement to the clinical situation at hand is paramount. Reasonable “turnaround times” should be planned for bioanalytical method development, data flow, and interpretation of clinical biomarker data. Team access to real-time data is often beneficial at this stage. It is crucial to consider the intended purpose of the biomarker measurement in this assessment (i.e., if the biomarker is exploratory or confirmatory, proximal or distal). It is beneficial for the bioanalytical scientists performing the assays to be included in data interpretation and to regularly communicate with the team on biomarker bioanalytical performance.

Clinical qualification, necessary for the use of a biomarker as a surrogate end-point, was described in the workshop. This is a graded process where scientific evidence is acquired in order to link a biomarker with a clinical end-point. A cycle of qualification, the nature of which is based on the evidentiary status of the biomarker relative to knowledge of the disease and compound pharmacology, precedes biomarker qualification and demonstration of surrogacy. This cycle roughly approximates the drug development process, where basic science and preclinical discovery characterize an exploratory phase of research.

Exploratory findings can lead to a demonstration of their linkage with the clinical end-point and further clinical characterization. These steps lead to the assembly of a biomarker database to demonstrate that a biomarker can substitute for a clinical end-point, otherwise known as surrogacy. Examples of identifying type-2 diabetes biomarkers were given in the case of PPARγ agonists. Free fatty acid was linked to insulin resistance as mechanism-specific, proximal demonstration biomarker; adiponectin was identified by cDNA microarrays, and serum adiponectin concentrations were correlated with

insulin sensitivity to be a proximal demonstration biomarker (11–13); and glycosylated hemoglobin (Hb_{a1c}) was qualified as a surrogate marker for PPARγ in type 1 and type 2 diabetes (14).

Postapproval Clinical Phase

Phase IV monitoring of safety and toxicity biomarkers can be mandated by the FDA prior to approval of a new drug. Prominent examples include the monitoring the immunogenicity of protein therapeutics and surveillance for idiosyncratic adverse events. The most common example of the latter is postmarket pharmacovigilance for rare hepatotoxicity (see www.fda.gov/cder/livertox/postmarket.pdf).

Clinicians have used biomarkers in patient diagnosis for a long time, though the term has only recently come into use. Previously, physicians depended on a standard set of laboratory markers to gauge patient progress. Today, data related to novel biomarkers are also correlated to clinical data and biology to assess their predictive power for clinical validation or qualification of the biomarkers. Given the uses described above for biomarkers in clinical trials, the integration of biomarkers into standard clinical practice is enabling the evolution of so-called personalized medicine. This term describes the process of selecting optimal treatments for an individual patient to ensure maximum therapeutic response with minimal side effects and toxicity. Additional major uses of biomarkers have arisen in the postapproval phase, which relate to personalized medicine. These include 1) the evolution of a biomarker into an approved clinical diagnostic test that enables postapproval surveillance of drug safety and clinical benefits, and 2) the correlation of safety and efficacy biomarkers that allows pharmaceutical companies and regulatory agencies to get a comprehensive picture of both the negative and positive clinical outcomes of a given drug when it is used in large patient populations. Increasing knowledge of diseases has led to such personalized therapies, where drugs or dosing regimens are targeted at specific disease mechanisms, and patient populations are stratified to maximize the utility of the targeted therapies. Clinical decision-making in breast cancer treatment, for example, can benefit from the use of biomarkers in disease staging and predicting patient prognoses (15,16).

METHOD DEVELOPMENT AND VALIDATION

Contextual Differences of Biomarker vs. Drug Bioanalysis

The speed with which biomarkers have been adopted by industry has left the field in something of a regulatory vacuum that can only partly be filled using established best practices and guidelines. The consequence of this lack of regulatory guidance is that confusion exists concerning what experiments should be performed and what data are necessary and appropriate for biomarker assay validation (17). This has led to inconsistent application of validation procedures.

A major concern expressed at the workshop was that forthcoming recommendations for biomarker assay validation might adhere too strictly to existing bioanalytical guidelines, particularly those established for drug and safety monitoring. A widely voiced opinion was that such an outcome would ultimately hinder novel biomarker research. The consensus and recommendations that arose from the workshop align with the "spirit" of previous guidelines, though great care was urged in the application of established bioanalytical principals to this young and emerging field. Workshop participants agreed that reliable and accurate data are required for making sound decisions during drug development and that there should be common, minimal expectations for biomarker data. It was also recognized that analytical requirements could be stage specific, based on the intended use of the biomarker data. Bioanalytical method validation issues were identified for subsequent actions to be addressed by the AAPS LBABFG Biomarker Committee.

Almost universally, regulatory guidance for bioanalytical validation indicates that the paramount objective of a bioanalytical method validation is to ensure that assay is reliable for its intended use (18,19). This philosophical viewpoint is central to guidance documents, conference reports, and "white" papers for bioanalytical validation of assays for conventional small molecule drugs and macromolecular therapeutics (20–25). Moreover, this has led to prescribed procedures and acceptance criteria for bioanalytical method validation. Although biomarker laboratory analyses can have many similarities to those used in toxicology and ADME studies, the variety of novel biomarkers and the nature of their applications often preclude the use of previously established bioanalytical validation guidelines in biomarker research. The same method can be used for various purposes (e.g., either exploratory or confirmatory research) or in different therapeutic areas, including those where significant sample matrix differences are encountered. For example, method validation requirements for inflammatory biomarker assays may be different in infectious disease, allergy, cancer, and cardiovascular disease samples or during different drug development phases. These considerations are pivotal for designing biomarker methods, defining the scope of validation and the degree of associated document control.

Categories of Assays and Data Defining Bioanalytical Method Validation

Workshop speakers echoed common themes, particularly with regard to "fit-for-purpose" validation guidelines. Still, there was no consensus on the specifics for biomarker assay development and validation. There was an agreement, however, that there needs to be some foundation for the validation efforts including all of the procedures required to demonstrate that a particular method is "reliable for the intended use" (18,19). Furthermore, regardless of the process used, the goal of assay development and validation should be to "develop a valid method" rather than simply to "validate a developed method" (26). Assay development and assay validation are intertwined and often occur within a process as shown in Fig. 1. The following sections review the presentations from the workshop, focusing on the common topics and themes that arose. Below we describe various influences on biomarker assay development and validation, including how the stage of development, sample and data type, and intended use of the assay influence decisions surrounding assay validation.

The standardization of validation processes for biomarkers is complicated because the assays are inherently diverse (27). Unlike bioanalytical assays for conventional xenobiotic

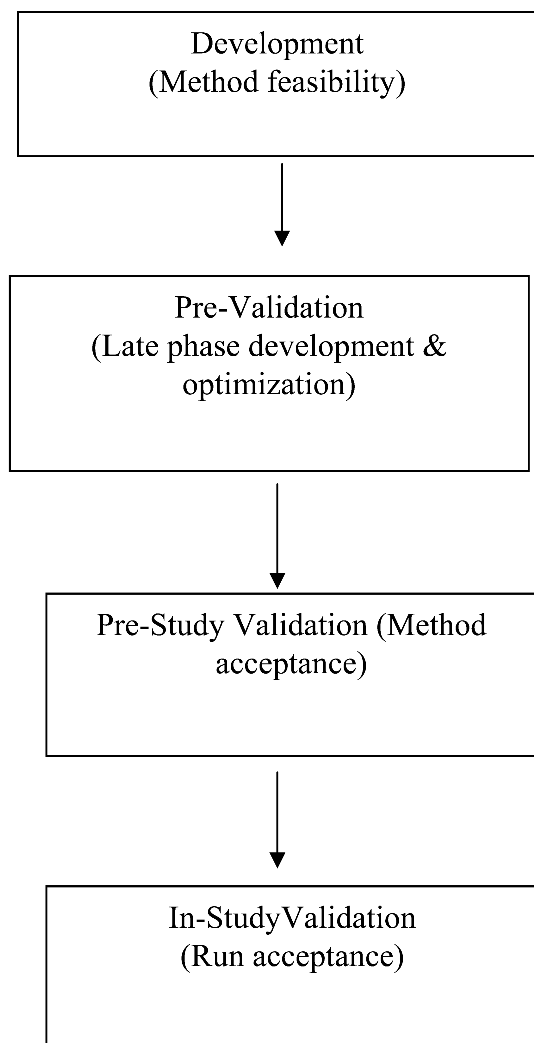

Fig. 1. Process of assay development and validation for bioanalysis.

drugs, where quantitative results are obtained by calibration against a highly purified and well-defined reference standard, biomarker assays can differ considerably depending on the type of analytical measurement, the type of analytical data that arises from the assay, and the intended use of the reported result. As described above, study objectives vary with the stages of drug development. Accordingly, if the intended use is exploratory in nature (i.e., the biomarker is not clinically validated), the extent of bioanalytical method validation is usually less than that for a confirmatory biomarker. A useful starting point in determining the direction of a biomarker assay validation is to consider the assay to be used and the data type that the assay will generate.

From a statistical perspective, biomarker data can be categorized as continuous, using numerical values, or categorical (with discrete or descriptive values). Categorical data may be classified further into either one of two types: ordinal, with discrete numerical or nonnumeric values that are spaced in a logical order but where a spacing interval is not implied; or nominal, characterized by nonnumeric values that classify results but are not of an implicit order. Lee and colleagues have provided a framework for categorizing bioanalytical methods that is useful in decision-making surrounding biomarker assay validation (27).

The first category is the definitive quantitative assays. For this category, calibration is performed using a reference standard that is well defined or characterized and is representative of the endogenous biomarker. Bioanalytical results are expressed in continuous numeric units of the definitive reference standard (e.g., a human insulin or a steroid assay). Definitive quantitative assays represent the ideal situation for biomarker measurement. The second category includes *relative quantitative assays,* where calibration is performed with a reference standard that is not well characterized, not available in a purified form, or not fully representative of the endogenous biomarker. Results from these assays are expressed in continuous numeric units of the relative reference standard (e.g., cytokine ELISAs). Standard clinical phase laboratory–based biomarker assays are typically quantitative in nature. However, only a small segment of novel biomarker assays are "definitive" quantitative measurements. More biomarker assays generate "relative" results, due to the nature of the reference materials or sample matrix. The latter complication often results from the presence of detectable, endogenous levels of the analyte in the sample. For relative quantitative assays, it is therefore appropriate to place greater emphasis on temporal changes in biomarker concentrations rather than the absolute concentrations. This is an example where input from bioanalytical scientists can be of value for interpretation of biomarker data.

Quasi-quantitative assays (quasi: "possesses certain attributes") are a third category. Unlike the previous quantitative assay categories, this does not use a reference standard (or calibration curve). However, the analytical response is continuous (numeric), and the bioanalytical results are expressed in terms of a characteristic of the test sample. Some examples of quasi-quantitative assays include anti-drug antibody assays (where the readout is a titer or % bound), enzymatic assays (where activity might be expressed per unit volume), and flow cytometric assays (28).

In contrast to the "quantitative" categories, qualitative analytical results of biomarkers are discrete (discontinuous) and reported in either ordinal or nominal formats. For example, low, medium, and high, or +, ++, and +++, are used for ordinal data, and yes/no or positive/negative for nominal data. These qualitative assays, like quasi-quantitative assays, use no calibration standards, and the bioanalytical results are expressed in terms of a characteristic of the test sample. The major difference is that the analytical response is categorical. The readout data may be ordinal (ordered rather than continuous responses) or nominal (nonordered and noncontinuous responses). An example of a qualitative assay would be a method to detect the presence of a single nucleotide polymorphism or gene mutation in a sample of DNA. Table I shows diverse examples of biomarkers, from gene expression to various clinical observations, and provides typical data types associated with these markers and examples of the technologies used to generate that data.

Regulatory Guidance on Biomarker Method Validation and Assay Acceptance Criteria

Discussions at the workshop highlighted a significant lack of consensus surrounding the application of method validation guidelines and acceptance criteria for biomarkers in clinical trials. For example, different criteria are used within

Table I. Method and Data Type Diversity: Varieties of Biomarkers from Genes to Clinical End-Points and the Plethora of Technologies for Their Evaluation

Biomarkers	Method/data type	Technologies
Genotyping	Descriptive, qualitative or quasi-quantitative	Genomics, microscopy, polymerase chain reaction (PCR)
Gene expression	Qualitative or quasi-quantitative	Quantitative-PCR, flow cytometry
Proteins	Quasi-, relative, or definitive quantitative	MALDI- and SELDI-ToF MS, ligand binding assays, cell-based assays
Biometabolites	Quasi-, relative, or definitive quantitative	Small molecule LCMS, ligand binding assays
Clinical markers and end points	Descriptive, qualitative, or quasi-quantitative	Physiological measurements, imaging

the pharmaceutical and diagnostic industries for method and assay run acceptance. FDA guidelines for bioanalytical drug assays (GLP, or good laboratory practices regulations), NCCLS (National Committee for Clinical Lab Standards), and CLIA (Clinical Lab Improvement Amendments) guidelines for diagnostic assays may not be suitable for some biomarker assays. Nevertheless, the lessons learned and principles applied in these previously established assay validation guidelines will no doubt significantly influence recommendations for biomarker assay validation. Widespread support was voiced for the use of statistics-based approaches for the establishment of acceptance criteria in the context of drug study objectives. It is illustrative to review the key differences between NCCLS, CLIA, GLP, and other relevant regulatory guidance on biochemical assay validation to better understand how biomarker validation decisions are made.

Given the diversity of biomarker assays and their intended applications, the 1990 Crystal City Bioanalytical Method Validation report statement that "all pharmacodynamic measures for definitive bioequivalency or related studies must be fully validated" (19) would only be applicable to very rare or specific situations. Also, the 2001 FDA Bioanalytical Method Validation guidance does not specifically address biomarker assay validation (21). Currently, there are two basic bioanalytical approaches that are being applied to biomarker validation. One is based on the FDA bioanalytical drug assay guidance and is generally being referred to as "GLP-like" (21,22). The second is based on NCCLS and CLIA (29,30). Table II highlights the analytical performance characteristics examined under these two approaches. The relative advantages of GLP and CLIA approaches were discussed during the open forum. For novel biomarkers, a GLP-like performance standard would be recommended for use in clinical trials.

Although the NCCLS/CLIA approach is well defined and is often more flexible in that it aligns the acceptance criteria to the proven performance of the method and adjusts them over time using a confidence limit–based approach, it is generally more appropriate for routine clinical assays. Several factors however, preclude its use in the validation of novel biomarker assays. This is particularly true for early clinical phase development where attrition of the clinical candidate and associated biomarker assays is high. In other words, the assay may only be used for a short period of time precluding the establishment of appropriate confidence limits.

Table II provides a comparison of various components of CLIA-type and GLP guidance on analytical validation. Accuracy assessments within NCCLS are based on comparison to "gold standard" methods, which rely on measured concentrations in an official reference or involve measuring a concentration in comparison to an official standard. In contrast, GLP accuracy assessments reflect the mean bias determined in spiked recovery experiments. For novel biomarkers, gold standard or comparator assays seldom exist, especially when considering macromolecule measurements. Consequently, spiked recovery is generally the most practical and feasible approach for accuracy evaluation. Sensitivity (i.e., the lowest reportable result) is defined quite differently between GLP and NCCLS. NCCLS/CLIA defines sensitivity as any response above the limit of detection (LOD), whereas GLP defines the LLOQ (lower limit of quantification) as the lowest concentration that can be measured with acceptable accuracy and precision (e.g., ±20%). There is a profound difference between linearity definitions. NCCLS defines linearity as the measure of the degree to which a calibration curve approximates a straight line, whereas the GLP and ICH guidelines define it as the condition in which the test results are directly proportional to the concentration of that analyte of interest. Additional differences are noted on how prestudy method acceptance criteria for accuracy and precision are assigned and used. For NCCLS, criteria are established depending on the intended clinical use of an assay. For GLP, and based on the 2000 Crystal City Meeting Report, mean bias (% relative error) and % coefficient of variation should be less than 15% (20% at the LLOQ), though newer publications call for more lenient criteria (25). There are also important differences in the approaches used by NCCLS and GLP concerning in-study run/batch acceptance criteria. NCCLS recommends use of control charts, such as Westgard rules (31,32), whereas GLP uses "4-6-X rule" (4 out of 6 QCs must fall within acceptance criteria X).

As mentioned above, it was recognized that, unlike drug assays, novel biomarker assays are accompanied by unique analytical issues, in many cases ruling out the use of universal, strict validation guidelines. These issues include the common absence of suitable reference standards, the employment of unique analytical reagents and assay platforms, the presence of endogenous biomarkers in a sample, analyte heterogeneity, and a variety of disease-specific effects. Therefore, at the conference there was an agreement that assays for novel biomarkers should not be defined as GLP assays. There was general consensus that these assays should be viewed as "GLP-similar" assays, meaning that the process of GLP validation should serve as a framework for biomarker assay validation. In the spirit of GLP, assay validation could be outlined ahead of time in a validation plan. This plan would include the scope and purpose of the assay, relevant background information, a description of issues that are expected

Table II. Comparison of CLIA- and GLP-Type Assays

	CLIA/NCCLS	GLP
Purpose	Human diagnostics and prognostics	Nonclinical drug safety studies and drug assays
FDA regulated	OIVD	CDER
CMS regulated	Yes	No
Requirements		
Written SOPs	Yes	Yes
Lab accreditation	Yes	No
Lab license	Yes	No
Proficiency testing	Yes	No
Compliance audits	Annually	Each study
Compliance inspections	By CMS and state authorities	By FDA
Validation characteristic		
Accuracy	Based on official reference gold standard	No gold standard; based on mean bias from spiked recovery experiments
Precision	√	√
Sensitivity	LOD	LLOQ
Linearity	Degree to which a calibration curve approximates a straight line	Test results directly proportional to analyte concentration
Specificity/selectivity	√	Against matrix and comedication
Range	√	Covering the expected concentration of study samples
Parallelism	√	On incurred samples
Dilutional linearity	–	√
Analyte stability	–	√
Standard stability	–	√
Reagent stability	√	√
Run acceptance criteria	Use control charts such as Westgard rules	4-6-X rule

√, validation required.

to affect the design of the validation experiments, a description of the planned validation experiments, and most importantly (when feasible) *a priori* criteria for method acceptance.

What constitutes a full validation is highly dependent on the type of assay being validated. Table III shows some of the prestudy performance validation characteristics that would be useful to explore in the validation of each the four categories of assays. For prestudy validation of definitive and relative quantitative methods, it is recommended that validation samples be prepared at five concentration levels: at the anticipated LLOQ, about 2–4 times the LLOQ, near the mid-range (on a log scale), about the 70–80% anticipated ULOQ, and at the anticipated ULOQ. The validation samples should be analyzed in duplicate or more in at least six runs. The default method acceptance criteria for a novel quantitative biomarker assay is suggested to be as follows. Method acceptance criteria: both accuracy (mean bias) and imprecision (%CV) should be within 25% (30% for the LLOQ). More or

Table III. Prestudy Validation Characteristics of Various Assay Performance Categories of Assays

Assay performance category	Definitive quantitative	Relative quantitative	Quasi-quantitative	Qualitative
Accuracy	√	√	–	–
Precision	√	√	√	–
Sensitivity	√ LLOQ	√ LLOQ	√	√
Specificity	√	√	√	√
Dilutional linearity	√	√	–	–
Parallelism	√	√	–	–
Assay range	√ LLOQ-ULOQ	√ LLOQ-ULOQ	√	–
Standard and key reagent stability	√	√	–	–
Matrix stability	√	√	√	√

√, validation required.

less lenient criteria can be required on a case-by-case basis depending on analyte properties, the type of assay and its limitations, and the intended use of the data. Batch (run) acceptance criteria: in contrast to conventional drug assays (e.g., 4-6-30 rule), more lenient run acceptance criteria can be justified in most cases for novel biomarkers. Criteria can be more or less lenient depending on the criteria used for prestudy method acceptance. In contrast to novel biomarkers, validation procedures for well-defined routine clinical assays may be closer to the CLIA/NCCLS guidelines, which set the quality standards for clinical lab testing on human specimens for the purpose of providing information for the diagnosis, prevention, treatment of disease, or impairment of or assessment of health (30).

When should biomarker assays follow the GLP or CLIA validation approaches? These decisions are often made by the biomarker team (see above), and it is thus necessary to disseminate information to that team and others who will use the data. The team should be consulted on the study objectives and timeline expectations, assay classification, and data type. These considerations can be considered in light of the intended use of the data to determine the course of action for assay validation. In general, biomarker assays should be GLP/GCP-like, unless the intended use is exploratory. In this case, the limited scope of the method validation should be defined in advance, and the future use of the data should be accordingly restricted. As stated previously, a limited or abbreviated method validation may be acceptable for certain stages of biomarker assay development, but the extent of the validation effort has to depend on the gravity of the decision to be made based on the biomarker data. For exploratory work, one should not avoid an assay just because tight controls are not in place. If a go–no-go decision will be made on the compound or program, then biomarker assay may deserve full validation. It is impractical and not cost-effective to require full method validation for exploratory biomarker research. It is also impossible to apply one set of rules to all assay platforms. The guidelines need to allow assay- and technology-specific flexibility. They should be very general to include both common sense and good science. For example, documentation of change and version control of biomarker methods is important, especially to those who produce elements of the assay/kit.

One of the issues discussed at the workshop concerned the level of method validation when transferring or outsourcing biomarker assays. Another issue was the basic requirement for full disclosure of data and validation efforts to reviewers and downstream users. The next goal of the LBABFG Biomarker Committee would be to drive a minimum acceptable standard and provide a spectrum and list of method validation processes to reflect staged progression of method validation.

Sample Integrity During Collection and Analyte Stability

As a biomarker progresses from preclinical to clinical application, researchers often encounter unexpected developments that negatively impact the performance of a biomarker assay. Preclinical experimentation is often quite difficult to extrapolate to the human clinical condition, as many of the advantages of animal models and *in vitro* human tissue models are absent in patients or normal volunteers. Animals used in discovery research are generally healthy and genetically homogeneous. Moreover, specimens are collected under highly controlled conditions with extrinsic variants kept to a minimum. This is rarely the case in clinical trials, where genetic and environmentally induced variables (notably preanalytical issues) often confound interpretation of early clinical biomarker data. Extrinsic factors are often uncontrollable and can include diet, coadministered medications and substances including tobacco and alcohol, and non-disease-related physiologic parameters including age-related differences in metabolism. Such factors, as does genetics, often account for as much of the observed inter-individual variability of biomarker responses.

An initial understanding of sample collection and stability is required by both CDER and CLIA regulations governing assay validation and should include investigation of preanalytical factors, such as analyte integrity during sample collection and processing. Preanalytical issues can be significant for novel biomarkers and represent an operational challenge in some cases. Control and standardization of sample collection is necessary to understand and minimize the variability from multiple clinical sites.

There is usually some delay between biomarker sample processing at a clinical site and its arrival at a bioanalytical laboratory. This delay can have significant effects on assay performance. Prominent examples of assays where this is problematic were discussed at the workshop and include *ex vivo* blood cell stimulation assays and tissue estrogen and progestin receptor assays. In the latter case, thermal lability of the receptors in the biopsy and cell-free extracts could compromise the assay performance. The collection and processing of human tissue biopsies has historically focused on obtaining clinical data (e.g., cancer diagnosis and staging). Until recently, little emphasis has been put on the development of appropriate sampling handling techniques (e.g., cryopreservation), which are essential for tissue-based biomarker analyses. Changes in sample collection procedures were described, beginning in the operating room and the pathology suite. For tumor marker analysis, it is necessary to freeze tissue specimens on dry ice in the pathology suite within 20–30 min of collection or immediately immersed in ice and rapidly transported to a site for cryopreservation and storage at −80°C.

Standardization of tissue collection and storage could benefit from the coordination and sample integrity control of a central sample repository. Explicitly written protocols for sample collection by a well-trained clinical staff should be in place before the clinical study begins. For both tissue collection and *ex vivo* stimulation assays, it is best to validate sample collection procedures using conditions that closely mimic the planned sampling protocol and to examine sample stability as soon as possible. In some cases, sample handling is so complicated that an effective strategy would be to establish the limitations surrounding sample collection and processing, then incorporate those limitations in the planning of sampling times.

Reference Standards, Calibrator Matrices, and Preparation of Quality Control Sample Matrices

For novel biomarkers, a widespread problem is the lack of characterization and standardization of the reference materials, as the cutting-edge and proprietary nature of many

novel biomarkers confounds development of an "official" standard. In contrast, when novel biomarkers become well established in the research community, reagent standardization can be addressed more effectively.

Standard Curve Range and Calibrator Matrix

The levels of the biomarker in disease and healthy populations should help to define the initial standard range. Samples from disease populations should be acquired and tested in biomarker assays to obtain the concentration range prior to the clinical study if possible. It is also important to understand how diurnal variability can affect the data. In such cases, it is prudent to pool samples or to collect them at the same time of the day. The initial surveys of healthy and patient samples also provide a rough idea of the assay variability. The clinical question is to compare the treatment vs. placebo. Appropriate clinical and assay controls should be included to produce unbiased clinical answers. For cancer studies, there may not be placebo or baseline samples available to provide data for better understanding of the true drug effect vs. the nonspecific variability of the biomarker *in vivo* expression or measurement variability.

Most biomarkers are endogenous compounds with measurable baseline levels in the biological matrix of interest. The endogenous and heterogeneous natures of biomarkers post a challenge to find analyte-free biological matrix to prepare calibrator standards. There is no limitation to what can be an appropriate substitute matrix for standard preparation. Often, an altered substitute matrix is used after being depleted of the analyte of interest. It may involve a stripped matrix, other species matrix, buffer, or healthy normal with minimal endogenous levels. Stripping by charcoal or more specific affinity molecules such as antibodies or molecular imprint polymers could be used.

The preparation of calibrators not in the intended sample matrix is one major difference of biomarker assays from that of drug compounds. In consequence, validation is required to demonstrate that the concentration-responses relationships are similar in the sample matrix and the substituted matrix. If the biomarker reference standard is a recombinant product, which may not have the same immunochemical behavior as the endogenous counterpart, the laboratory should perform parallelism studies on the recombinant vs. the endogenous materials.

Quality Controls

Method variability in buffer should be checked first during method development. Serum or other binding proteins can present bioanalytical challenges that can affect biomarker results. Spike recovery experiments should be performed on multiple individual lots of matrix to assess accuracy, matrix effects, and interference. The background levels of the endogenous analyte and other interferences should be taken in consideration, because they can skew the results and affect accuracy and the application of the acceptance criteria. If there is an interference or matrix effect, the sample could be diluted to reduce the background. One suggestion was to set the lowest spike concentration at ≥50% of the background level in order to quantify the spike recovery.

Quality controls (QCs) prepared from a healthy population may not reflect the true assay performance in patient samples. When patient samples can be made available, comparison between healthy population and patients should be made. If patient samples are not available during method validation, it is necessary to test randomly selected patient samples retroactively to characterize the variability in the diseased population.

The use of QCs can pose a problem for clinical labs and CLIA tests, which may not have any QCs. QCs are handled very differently in the diagnostics arena. Diagnostics assays run QCs prior to assaying samples in order to assess the condition of the instrument. If the QCs are not within acceptable limits, then instrument parameters are adjusted and QCs rerun until acceptable. Typically non-CLIA, the bioanalytical laboratories (GLP/GLP-like) use five concentration levels of QCs during prestudy validation and three levels during sample analysis. This is an unfamiliar approach to CLIA laboratories. No consensuses was reached at the workshop open forum regarding how many QC levels and how many replicates per QC should be used.

Multiple dilutions would be necessary for samples with high endogenous levels. It may take multiple dilutions to bring the sample responses into the calibration range. In many instances, the technical range of a biomarker standard range is not the same as the study samples. In such case, rather than preparing QCs at the high, mid, and low levels of the standard curve range, the levels should mimic the intended study samples.

Although an appropriate substitute matrix is used for standard preparation, QCs must be in the same matrix as the authentic samples. However, one has to consider possible exceptions with rare matrices, such as tear, CSF, or synovial fluid, where in these cases a surrogate matrix may be the only practical option.

Data Handling: Curve Fitting and Data Assessment by Statistical Methods

Contributing variance factors come from both method variability and biological variability; possible artifacts could rise from *in vivo* and *in vitro* sources. Statistical assessment should be made on biological variation of markers between disease and nondisease, method variability, and whether single or multiple biomarkers were involved. Determination of outliers using confidence interval should be made.

Components of method variability (e.g., reagents, operator, incubation conditions, intra- and inter-assays) and biological variability (e.g., % differences of one sample vs. the reference time point, individuals, placebo and various doses, clinical vs. discovery settings) can be identified and the data compared. Random effects (variance components) models provide estimates and confidence intervals. The data will help study design to provide power for the study and determine assay expectations.

For biomarker studies, often the baseline value could be below the LLOQ. However, often this value is required for comparison of the drug effect. In a PK study, the LLOQ would be reported as an alphanumeric value "BQL" (below quantifiable limit), as LLOQ concentration, or as zero. None of these three options would be appropriate for biomarker assessment. For example, if the data were to be expressed in fold increases or % of basal, reporting as zero skews the

variability to nothing, whereas reporting as the LLOQ value would underestimate the drug effect, and reporting as BQL would not allow calculation. It is possible and useful to compute a value between 0 and LLOQ, which can provide a better understanding of the true value in these patients. In some instances, it may be appropriate to report "not detected" levels in addition to BLQ. We should think about broad applications, including situations that use no calibration curves. Not all technologies generate data of the same type—imaging is one example. The issue of LLOQ required being lower than endogenous levels should not affect method validation, as the purpose of validation is to show what the assay is capable of, not the clinical validity of the result.

For binding assays, the 4- or 5-parameter logistic model is the gold standard model for curve fitting (23). Sometimes, standards in commercial kits are provided only at the low portions of the 4-parameter logistic curve function. One can add standards and anchor points to extend the range to better define the curve function parameters or use a 5-parameter logistic curve fitting.

Document Control Considerations

An issue common to all biomarkers is to what extent should the assays and experimental results be documented. With respect to regulatory submission of biomarker data, it is preferred that a team have a well-thought-out biomarker plan that is project-specific and focuses on the anticipated use of the biomarker data. Should biomarker data be used in support of claims as to the mechanism of action of a compound, to predict efficacy, or support a safety profile, it is best that these issues be identified during the preclinical stages of development. Without question, if the biomarker assays are used to support GLP studies, then they should approach the validation requirements of a GLP assay as closely as possible. This includes document control.

Just as the bioanalytical assay evolves over time, so should the level of documentation control. For example, the depth of detail in a validation report is expected to increase as the assay evolves and data are collected. Also, it may not be appropriate or necessary to establish standard operating procedures for exploratory or early clinical biomarker assays, but later stages may demand it. Because most biomarkers identified in discovery will not progress into clinical studies, it makes practical sense to concentrate the documentation efforts only on those biomarkers and associated method validations that have a reasonable chance of progressing into clinical development. This eliminates a significant amount of unnecessary effort that can be better focused on those biomarkers with the best chance of providing useful clinical data.

Currently, individual companies are struggling with defining and developing a process for what nonclinical biomarker data to include in regulatory submissions. In terms of nonclinical data, there currently exist regulatory requirements on what needs to be submitted to the agency. Investigational new drug (IND) applications require pharmacology and toxicology information "on the basis of which the sponsor has concluded that it is reasonably safe to conduct the proposed clinical investigations." In the case of a new drug application (NDA), nonclinical studies that should be submitted include those "studies that are pertinent to possible adverse effects." Unfortunately, biomarker data is often not fully "biologically" validated or well understood, so questions often arise as to what really needs to be submitted. For example, genomic data might indicate that a toxicity finding is likely to be unique to a specific species of laboratory animal. If the rationale for determining the level of patient safety includes information derived from that genomic data, it would need to be part of the regulatory submission. In other words, those results intended to influence the course of the clinical development process would be considered part of the safety and efficacy evaluation. In contrast, some nonclinical biomarker data that does not have such a regulatory impact would thus not need to be part of the regulatory submission. Data from microarray toxicity screening, which is only used for hypothesis generation or biomarker discovery, is one example of the latter data type.

Regulations on *in Vitro* Diagnostic Tests

The premarket review of medical devices or new diagnostic tests (510K) evaluates the performance characteristics of a new laboratory test on accuracy and precision of the laboratory test, analytical sensitivity and specificity of the method, and clinical or diagnostic sensitivity and specificity. The evaluations include the intended use; indication(s) for use, special condition for use statement(s), and special instrument requirements, and the device description. A Statement for Reporting Studies of Diagnostic Accuracy was prepared with the objective toward complete and accurate reporting of studies on diagnostic accuracy (33). In 1997, the FDA published the Analyte Specific Reagent Rule (ASR Rule). Analyte specific reagents include antibodies, specific receptor proteins, nucleic acid sequences, and similar biological reagents, which through chemical binding or reaction with substances in specimen are intended for identification and quantification of an individual chemical substance or ligand in biological specimens.

The Office of *in Vitro* Diagnostics Device Evaluation and Safety evaluates emerging technologies for infectious diseases detection, microarrays, proof-of-concept testing, over-the-counter tests, expert systems, and diagnostic/drug pairs. A good example of the codevelopment of a diagnostic assay and therapeutic is the Herceptest/Herceptin combination, where the diagnostic test was applied to pivotal clinical trials, and the simultaneous submission of applications for both components aided in the approval of both the drug and the test (34). The development of any new test/service requires that the new test must have clinical importance in its impact on medical action and affect patient outcome in overall survival, quality of life, and reduced cost of care. The new test must meet financial objectives on the minimal setup cost of capital equipment and personnel and provide adequate third party reimbursement on income generation and meet physician demand.

Emerging Novel Assays

The discovery of efficacy and toxicity biomarkers is often based on pattern differentiation in experimental animal models and human populations. A presentation by Molecular Staging on high-throughput protein microarrays for disease sample analysis was given in the workshop. The platform and assay validation, along with some details of data analysis strat-

egies, have been recently described (35–37). Selectivity, sensitivity, broad dynamic range, and reproducibility were the components emphasized in the method validation. Multiple assay quality control spots were included in each run to test assay variability such as array printing, reagents, individual spots and slides, background signal, matrix effects, and entire assay runs. Statistical design of experiments was used to minimize well-to-well and slide-to-slide imprecision. The Molecular Staging chip system is currently in use to survey a broad panel of analytes and is thus less amenable to large clinical trials than other, robust and validatable immunoassay platforms. Complementary methods will be necessary to advance subsets of analytes discovered in early stages into later stage, higher throughput clinical trials.

The workshop open forum also focused on several potentially difficult issues that need to be addressed in multiplex assays in general, and particularly in multiplex ligand binding. These include the availability of well-characterized reference standards, understanding reference ranges for different analytes, assay reproducibility and generalizability, coverage of other species, correction for biases in the analyte set, and bridging various phases of drug and diagnostics development.

The design of standard curve ranges can be complicated during method development for multiple analyte assays. In order to accommodate assay capabilities, sometimes compromises lead to the truncation of the assay ranges for a subset of analytes. The limit of quantitation and specificity in multiplex assays depends on the antibody for each analyte. To evaluate specificity and cross-reactivity, it is advisable to check each analyte once with all the bead sets and determine whether there is interference (i.e., each analyte of the multiplex should be treated as a unit during method evaluation). When the data generated is to support a study and repeat analysis is required for one of the analytes, the assay would be performed with all of the analytes, but only report the re-assay results for the one that failed and ignore the results of the others.

Microarray-based analysis of mRNA expression profiles provides is another technology that generates enormous amounts of information. Such analysis holds the potential to aid in clinical biomarker discovery and patient stratification for responsiveness to a given drug, as well as the identification of novel drug targets and markers of drug action. Current understanding of the huge data sets that arise from complex cDNA microarray experiments falls far short of what is necessary for their implementation in the regulation-intensive environment of clinical trials. Scientists at the Lilly Research Laboratories have been working on developing a validation plan to determine bioanalytical- and instrument-level variation in Affymetrix gene chip experiments. An experimental scheme was designed to estimate variance attributable to analyst, reaction conditions, chips and fluidics stations as shown in Fig. 2. Statistical analysis of data from 55 Affymetrix U95Av.2 chips in this experiment produced a %CV of 21 over a p value up to 0.06 for approximately 4600 genes that were called either present or marginal in 75–100% of chips analyzed. About 12% of that variation was attributed to the chip and less than 2% was due to the analyst. When the method was applied to the U133A array, it produced similar %CVs for the same input RNA. The key to development of a successful validation strategy for microarray-based analysis in clinical trials is to initiate a dialogue with the FDA and seek their input in the earliest phases of these experiments.

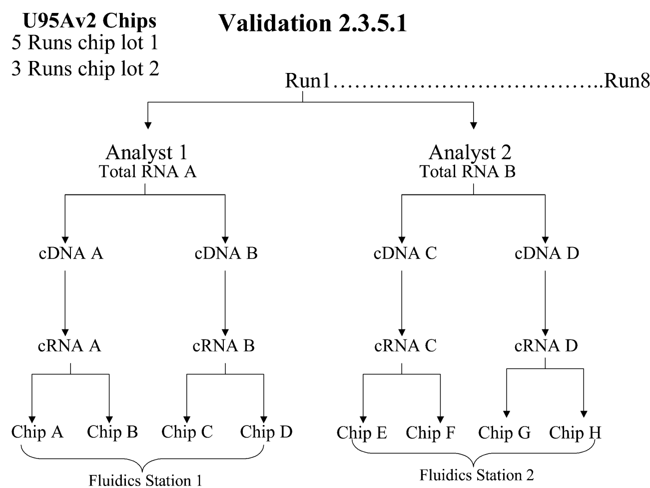

Fig. 2. An experimental scheme to estimate variance attributable to analyst, reaction conditions, chips, and fluidics stations. Statistical analysis of data from 55 Affymetrix U95Av.2 chips in this experiment produced a %CV of 21 over a p value up to 0.06 for approximately 4600 genes.

CONCLUSIONS, FUTURE PERSPECTIVES, AND FOLLOW-UP ACTIONS

A shared conclusion from the workshop was that, because the quality of clinical decisions can depend on the quality of biomarker data, appropriate analytical validation of biomarker assays is essential to ensure high-quality data to maximize the value of such decisions. The extent of assay validation is often determined by the study objectives, as these data should support the desired level of biomarker qualification. A properly validated biomarker assay will help ensure that performance characteristics of the measurement (i.e., sensitivity, specificity, and reproducibility) would meet predefined acceptance criteria, but do not necessarily link the biomarker with a clinical end-point. Stage-specific validation can be a challenging concept to explain to clinical "end-users" or internal "customers" in order to justify undertaking a given validation effort. Given that the use of an appropriately validated biomarker maximizes investments in time and money while ensuring adequate regulatory compliance, these discussions are undoubtedly worthwhile.

This workshop occurred at the beginning of a scientific collaborative effort to address the issues of method development and validation of quantitative biomarker assays. There are many other important aspects of biomarkers that were not covered in the workshop. These include an in-depth analysis of clinical biomarker validation (qualification), the use of combinations of biomarkers, including pharmacogenomic and other population-based methods. One particularly relevant example to be covered is a description of the validation efforts required for clinical applications of single nucleotide polymorphism (SNP) genotyping assays in drug development and therapy. The concept of using pharmacogenomic markers to provide "individualized therapy" has been widely promoted in both the lay media and scientific literature, via approaches such as the use of SNP genotyping to define meta-

bolic phenotypes (CYP2D6, etc.) or identification of target-specific polymorphisms that may affect the pharmacodynamic effects of a drug (i.e., EGFR mutations and Iressa). As a step toward increasing the application of pharmacogenomic information in drug development, the FDA is preparing a guidance document to industry to aid in the submission of pharmacogenomic data (draft version available at http://www.fda.gov/cder/guidance/5900dft.pdf) and has encouraged voluntary submission of pharmacogenomic data in an effort to increase the knowledge base for therapeutic candidates (see http://www.fda.gov/oc/initiatives/criticalpath/Lesko/Lesko.html). As this field moves closer to clinical reality, issues regarding how such assays will be applied and validated will need to be addressed. In addition, bottlenecks in biomarker development, including the translation of biomarker targets and assays from preclinical to clinical use, notably how species specificity affects the soundness of data and limits its extrapolation to human models, need to be explored. Future biomarker method validation workshops will likely expand beyond the focus of the 2003 workshop and are likely to include nonconventional technologies for discovery and implementation of biomarkers, including imaging, proteomics, and synergistic combinations of these and newer emerging technologies. Characterization and standardization of biomarker reagents also needs to be addressed and will likely remain an issue for some time.

The costs associated with newer technologies are currently greater than for more standard immunoassays. In the near term, immunoassays and related ligand binding assays will likely continue to be the analytical method of choice for biomarker measurements. Continued clinical application of novel biomarkers will therefore depend on the development of new antibodies for ligand binding assays. Ironically, despite the rapid advances in technologies and assay validation and the promise that biomarkers will ultimately save money in both drug development and patient care, more mundane issues must also be addressed in order to maximize the utility of biomarkers. For example, the establishment of billing codes and affordable charges for novel biomarker assays will have to be worked out well before these markers are widely adopted beyond the world of drug development. In the absence of government reimbursement, the cost issues associated with biomarker assays may ultimately have to be folded into the cost of new drug development.

As in most arenas, education is key to the advancement of biomarkers in clinical practice. Physicians need education on the nature and use of biomarker assays and associated technologies and the interpretations of biomarker data to better convey their benefits to patients. For example, it was noted that breast cancer patients typically wish they were better informed. This will no doubt be aided as examples of successful biomarker implementation emerge, though concerns surrounding the disclosure of proprietary information inhibits and delays information dissemination and discussion.

The LBABFG Biomarker Committee has established work teams to address the bioanalytical issues, seeking input from a wide network of participating scientists. A position white paper addressing many of the important issues outlined in this report is currently being prepared and will be available in the not too distant future.

ACKNOWLEDGMENTS

The authors are grateful to the following contributors to the workshop and this report. Workshop Discussion Facilitators: Larry Duan, John Allinson, Bob Millham, and Daisy McCann. Scribes who recorded the discussions throughout the sessions: Binodh DeSilva, Keith Joho, Marion Kelley, Marie Green, and Micheala Golob. The authors are indebted to Marie Green and Robert Masse for providing critical review of the manuscript.

REFERENCES

1. J. A. DiMasi, R. W. Hansen, and H. G. Grabowski. The price of innovation: new estimates of drug development costs. *J. Health Econ.* **22**:151–185 (2003).
2. G. Levy. Mechanism-based pharmacodynamics modeling. *Clin. Pharmacol. Ther.* **56**:356–358 (1994).
3. C. C. Peck, W. H. Barr, L. Z. Benet, J. Collino, R. E. Desjardins, D. E. Furst, J. G. Harter, G. Levy, T. Ludden, and J. H. Rodman. Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. *Pharm. Sci.* **81**:600–610 (1992).
4. W. A. Colburn. Selecting and validating biologic markers for drug development. *J. Clin. Pharmacol.* **37**:355–362 (1997).
5. P. R. Jadhav, M. U. Mehta, and J. V. S. Gobburu. How biomarkers can improve clinical drug development. *Am. Pharm. Rev.* **7**:62–64 (2004).
6. E. Zerhouni. Medicine. The NIH Roadmap. *Science* **302**:63–72 (2003). Available at http://nihroadmap.nih.gov.
7. F. D. A. March 2004 report. Innovation or stagnation: challenge and opportunity on the critical path. Available at http://www.fda.gov/oc/initiatives/criticalpath/.
8. G. J. Downing. *Biomarkers and Surrogate Endpoints: Clinical Research and Applications*. Proceedings of the NIH-FDA Conference held on 15–16 April 1999. Elsevier, New York, 2000.
9. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69**:89–95 (2001).
10. R. R. Bowsher. Analytical validation of assays for novel biomarkers. Presentation at AAPS Workshop, Method Validation and Measurement of Biomarkers in Nonclinical and Clinical Samples in Drug Development, Salt Lake City, Utah, 2003.
11. J. A. Wagner. Early clinical development of pharmaceuticals for type 2 diabetes mellitus: from pre-clinical models to human investigation. *J. Clin. Endocrinol. Metab.* **87**:5362–5366 (2002).
12. J. A. Wagner. Overview of biomarkers and surrogate endpoints in drug development. *Dis. Markers* **18**:41–46 (2002).
13. J. Berger and J. A. Wagner. Physiological and therapeutic roles of peroxisome proliferator-activated receptors. *Diabetes Technol. Ther.* **4**:163–174 (2002).
14. U. Krishnamurti and M. W. Steffes. Glycohemoglobin: a primary predictor of the development or reversal of complications of diabetes mellitus. *Clin. Chem.* **47**:1157–1165 (2001).
15. D. A. Dillon. Molecular markers in the diagnosis and staging of breast cancer. *Semin. Radiat. Oncol.* **12**:305–318 (2002).
16. R. Dates, M. Schmitt, and N. Harbeck. Advanced statistical methods for the definition of new staging models. *Recent Results Cancer Res.* **162**:101–113 (2003).
17. B. N. Swanson. Delivery of high-quality biomarker assays. *Dis. Markers* **18**:47–56 (2002).
18. I. C. H. Guidelines. Text on validation of analytical procedures, Q2A. International Conference on Harmonization, Geneva, Switzerland, 1994.
19. V. P. Shah, K. K. Midha, S. Dighe, I. J. McGilveray, J. P. Skelly, A. Yacobi, T. Layloff, C. T. Viswanathan, C. E. Cook, R. D. McDowall, K. A. Pittman, and S. Spector. Analytical methods validation: bioavailability, bioequivalence, and pharmacokinetic studies. *Pharm. Res.* **9**:588–592 (1992).
20. V. P. Shah, K. K. Midha, J. W. A. Findlay, H. M. Hill, J. D. Hulse, I. J. McGilvary, G. McKay, K. J. Miller, R. N. Patnaik, M. L. Powell, A. Tonnelli, C. T. Viswanathan, and A. Yacobi. Bioanalytical method validation. A revisit with a decade of progress. *Pharm. Res.* **17**:1551–1557 (2000).

21. Guidance for industry on bioanalytical method validation: availability. *Federal Register* **66**:28526–28527 (2001).
22. FDA government document. Code of Federal Regulations. Title 21, Vol. 1. Good Laboratory Practice for Nonclinical Laboratory Studies. Revised April 1, 2001.
23. J. W. A. Findlay, W. C. Smith, J. W. Lee, G. D. Nordblom, I. Das, B. S. DeSilva, M. N. Khan, and R. R. Bowsher. Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J. Pharm. Biomed. Anal.* **21**:1249–1273 (2000).
24. K. J. Miller, R. R. Bowsher, A. Celniker, J. Gibbons, S. Gupta, J. W. Lee, S. J. Swanson, W. C. Smith, and R. S. Weiner. Workshop on Bioanalytical Methods Validation for Macromolecules: summary report. *Pharm. Res.* **18**:1373–1383 (2001).
25. B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, D. Tracey, H. Hill, and A. Celniker. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm. Res.* **20**:1885–1900 (2003).
26. W. C. Smith and G. S. Sittampalam. Conceptual and statistical issues in the validation of analytic dilution assays for pharmaceutical applications. *J. Biopharm. Stat.* **8**:509–532 (1998).
27. J. W. Lee, W. C. Smith, G. D. Nordblom, and R. R. Bowsher. Validation of Assays for the Bioanalysis of Novel Biomarkers. In J. C. Bloom and R.A. Dean (eds.), *Biomarkers in Clinical Drug Development*. Marcel Dekker, New York, 2003, pp. 119–149.
28. A. R. Mire-Sluis, Y. C. Barrett, V. Devanarayan, E. Koren, H. Liu, M. Maia, T. Parish, G. Scott, G. Shankar, E. Shores, S. J. Swanson, G. Taniguchi, D. Wierda, and L. A. Zuckerman. Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J. Immunol. Methods* **289**:1–16 (2004).
29. National Committee for Clinical Laboratory Standards (NCCLS). Document EP5-A: Evaluation of Precision Performance of Clinical Chemistry Devices: Approved Guideline (1999); Document EP6-P: Evaluation of the Linearity of Quantitative Analytical Method: Proposed Guideline (1986); Document EP7-P: Interference Testing in Clinical Chemistry: Proposed Guideline (1986); Document EP9-A: Method Comparison and Bias Estimation Using Patient Samples: Approved Guideline (1995).
30. FDA government document. Code of Federal Regulations. Title 42, Vol. 3. Clinical Laboratory Improvement Amendment. Revised October 1, 2001.
31. J. O. Westgard, P. L. Barry, M. R. Hunt, and T. Grove. A multirule Shewhart chart for quality control in clinical chemistry. *Clin. Chem.* **27**:493–501 (1981).
32. J. O. Westgard and G. G. Klee. Quality management. In C. Burtis (ed.), *Fundamentals of Clinical Chemistry,* 4th ed. WB Saunders, Philadelphia, 1996, pp. 211–223.
33. P. M. Bossuyt, J. B. Reitsma, D. E. Burns, C. A. Gatsonis, P. P. Glasziou, L. M. Irwig, D. Moher, D. Rennie, H. C. de Vet, and J. G. Lijmer. Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin. Chem.* **49**:7–18 (2003).
34. DAKO HercepTest™ Facts, DAKO, Fort Collins, CO, USA, 2000.
35. B. Schweitzer, S. Roberts, B. Grimwade, W. Shao, M. Wang, Q. Fu, Q. Shu, I. Laroche, Z. Zhou, V. T. Tchernev, J. Christiansen, M. Velleca, and S. F. Kingsmore. Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat. Biotechnol.* **20**:359–365 (2002).
36. W. Shao, Z. Zhou, I. Laroche, H. Lu, Q. Zong, D. D. Patel, S. Kingsmor, and S. P. Piccoli. Optimization of rolling-circle amplified protein microarrays for multiplexed protein profiling. *J. Biomed. Biotechnol.* **5**:299–307 (2003).
37. S. F. Kingsmore and D. D. Patel. Multiplexed protein profiling on antibody-based microarrays by rolling circle amplification. *Curr. Opin. Biotechnol.* **14**:74–81 (2003).