Alice E. Till, Ph.D.
VICE PRESIDENT
SCIENCE POLICY AND TECHNICAL AFFAIRS

PhRMA

July 11, 2003

Dockets Management Branch (HFA-305),
Food and Drug Administration,
5630 Fishers Lane, Rm. 1061,
Rockville, MD 20852

Re: Draft Guidance for Industry and FDA Reviewers; Multiplex Tests for Heritable DNA Markers, Mutations, and Expression Patterns [Docket No. 03D-0120, 68 *Federal Register*, 19549-19550, April 21, 2003]

Dear Sir or Madam,

The Pharmaceutical Research and Manufacturers of America (PhRMA) represents the country's leading research-based pharmaceutical and biotechnology companies, which are devoted to inventing medicines that allow patients to lead longer and more productive lives. Investing more than $30 billion annually in discovering and developing new medicines, PhRMA companies are leading the way in the search for cures.

We are pleased to send our comments on the subject guidance above. Generally, we welcome FDA's initiative to provide guidance on this complex and important topic and hope that our comments here will be useful to the agency in refining the final version.

General Comments on the Draft Guidance

The guidance should include a glossary of terms to ensure consistency and clarity.

Involvement of multiple Centers in shared guidance

We hope that future guidance documents on the use of microarray data in submissions to the FDA reflect the range of requirements of different FDA centers, and that future revisions of this document would also involve all of the Centers that review data from microarray analyses.

The guidance document was issued by the Center for Devices and Radiological Health (CDRH) for use in preparing Pre-Market Approval (PMA) and 510(k) applications. PhRMA members are using, and plan to use, multiplex assays including microarrays for a variety of purposes in drug discovery and development. Exploratory studies as early as discovery will likely have an impact on the design of later development studies in preparation for submission to the Agency. Some of those uses should lead to applications for marketable devices to CDRH, but other uses will be in clinical development of drugs that will be reviewed by other Centers. We believe it is important that a common regulatory framework be available for these uses, which requires the involvement and guidance of multiple FDA Centers.

03D-0120 C7

Given the likelihood that studies in clinical drug development using microarrays will be coordinated with development of clinical test applications based on microarrays, a shared regulatory framework seems not only desirable but necessary. Questions about comparison of clinical samples in submissions to CDRH (Section III of the guidance) might be addressed in part by leveraging microarray data from clinical development studies, but that will be of limited value if guidance for analytical validation differs between FDA Centers.

Introduction – Purpose

Scope of "Multiplex Tests"

The scope of "multiplex tests" covered by the guidance is not always clear, but is apparently very broad. It should be clarified to address both technical issues and any related guidance. Technical issues include:

- one can imagine many different ways of assessing DNA markers without using arrays, but using procedures that measure genotype at one locus or many loci. The simplest case to consider is the assessment of genotype at one marker. This could be technically construed as a multiplex test, because two alleles are being assayed. More complex tests include assessing multiple markers (or mutations) within a gene, assessing the haplotypes of a gene directly, estimating a haplotype from genotype data, and combining genotype data or haplotype data across many genes. The complexity of multiplex tests goes beyond these statements made for genetics. The same type of queries could be made for multiple TaqMan markers or ELISAs
- the use of multiple features in a single test where the measurements are not strictly simultaneous. Platforms that measure multiple analytes in rapid succession should be considered multiplex platforms if the multiple analyte measurements are combined into a single test.
- The guidance refers to "oligonucleotide, cDNA, protein and tissue arrays" (Purpose) but most of the examples of the guidance are restricted to the DNA arrays assaying for DNA sequence or RNA expression level variation. If the guidance retains its broad scope, this lack of correspondence between scope and example should be rectified, at least allowing for amendments to the guidance as new technologies develop.

The Least Burdensome Approach

The discussion of achieving the Least Burdensome Approach (LBA) for compliance is brief. The LBA in practice may be dependent on resolution of the other issues presented in the guidance.

The least burdensome regulatory pathway may vary by assay and intended use. The FDA should consider which path to recommend on a case by case basis for each application. Points to consider when making this decision include the following:

- Intended use
    - Predisposition

- Aid to diagnosis
- Prediction of risk for recurrence or disease severity
- Therapy choice/dose

- Safety Issues
  - Treatment availability
  - Risk from misclassification
  - Benefit to the patient

Intended Use / Regulatory Pathway/ Risk:Benefit Analysis

We suggest that the guidance document be written more broadly so that it will apply to most or all classifications of multiplex and array assays. We believe that the most appropriate regulatory pathway will vary for various types of intended use. This approach suggests a more stringent regulatory pathway for those products which have higher risk and lower relative benefit to the patient and a less burdensome approach for those in which many benefits can be delivered to the patient with relatively few risks. Specifically, we suggest the wording in section 1.8 of the FDA document: *Guidance for Submission of Immunohistochemistry Applications to the FDA* (June 3, 1998) offers a useful precedent.

The discussion above is restricted to the scope of the CDRH guidance. We believe that this framework can be extended beyond the use of multiplex assays in a PMA or 510(k) submission, and may be useful when developing a broader framework in collaborative guidance from CDRH, the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER).

Genetics versus expression

Different types of multiplex assays report qualitatively different types of information. In particular, assays of genetic variation at the DNA level are fundamentally distinct from assays of expression variation at the RNA or protein level. This distinction is not consistent throughout the guidance, and should be clarified in Section III of the introductory material.

The guidance may be easier to follow if the detailed recommendations are also broken out along these lines.

What follows is a suggested wording for this section.

III. Different types of information: DNA tests v. RNA tests

DNA, RNA and protein all have the potential to provide information regarding the genetic state of germinal cells, somatic cells and tumor cells. However, information is qualitatively and quantitatively different for tests measuring DNA versus tests measuring RNA or protein. Measurement of DNA provides information about allelic state, gene copy or chromosome number. Measurement of mRNA or protein levels provides information about the relative level of a gene product only at the given time point of tissue isolation and processing. It is important to note that mRNA and protein levels are dynamic.

Because of DNA polymorphism, epistasis, response to environmental stimuli and varying efficiencies in biological processing, relative transcription levels of messenger RNAs and protein production vary within and among individuals. Also, levels of messenger RNA or protein are measured as continuous variables and the level of any one or a few transcripts may or may not provide enough power to be used for a certain type of test. Consequently, for many analyses, gene expression pattern information is more informative and reliable than the level of any given transcript or protein.

DNA information including genotype, gene dosage and karyotype information is more absolute or digital in nature than RNA or protein expression information. Informative DNA analysis can identify specific alleles or mutations, specific whole numbers of gene copies and/or whole numbers of chromosomes.. Results from these tests can be characterized as dichotomous, e.g. present or absent, trichotomous, e.g. homozygous A, homozygous B or heterozygous A, B, or categorized, e.g. haplotypes. For DNA analysis measurements of single base changes as well as patterns of SNPs can be highly informative.

DNA based tests: Genotype, gene dosage, karyotype

DNA array tests should be carefully designed, highly reproducible and have well established performance characteristics. Process for interpretation of results from tests identifying allele or mutation state, gene copy number etc., should be relatively straightforward. Clinical studies should account for allele frequencies in unaffected populations and disease prevalence in the populations being studied.

mRNA based tests: Gene expression patterns

Expression patterns represent complex interactions between genes, pathways, and networks. Expression pattern tests should be highly reproducible, both technically and biologically. The primary correlation of the test should be to the outcome being measured, not to any procedural step. Applications can include disease predisposition, disease class/subclass, prognosis, treatment response/monitoring information.

... Suggested wording ends.

Comments on Recommendations

I. Intended use of a test or device

The recommendation of *"a separate application for each intended use"* when a test has multiple intended uses should be reconsidered to allow amendments to existing applications for a new use when appropriate.

II. Analytical Validation

In general, we suggest rewriting the analytical validation section to reflect a general structure in which guidance for assay components, including arrays, reagents, samples, and equipment, and software is dealt with in Part A, and the actual validation of the specific assay and platform at a site (Pharma or CRO) is described in Part B. This section

would include assay validation experiments and specific documentation which are not generic to the components and reflect the utilization of all the components described in Part A to address a specific assay used to derive data associated with a clinical study, test plan, or operating laboratory.

Additional comments are as follows:

- Certain sections of the draft guidance are written for single analyte tests specifically and are not applicable to multiplex tests. These sections should be removed.
- The assay validation section should be expanded to specifically recommend assay validation activities such as documentation, site specific validation, tracking of process controls, and assay performance over time.
- This section should be expanded to include reporting of sample collection and processing methods.
- In general, the language needs to be clarified for multiplex tests with continuous measurements per analyte. Specifically, the section referring to assay sensitivity and reproducibility should reflect the use of these terms in a multiplex context, i.e., sensitivity for all genes or a subset.
- This section may have to be written differently for multiplex genetic tests vs. gene expression assays.

Pre-analytical issues such as mRNA stabilization or tissue microdissection for cancer applications are not addressed in the guidance document. Some clarification of CDRH expectations would be helpful.

II.A. Validation of assay components/systems

We recommend that submissions include analytical data that demonstrate that the device performs accurately and reliably under given conditions. The following elements of arrays and multiplex platforms should be well-characterized: design, internal controls used, oligonucleotides, primers, probes, or other capture elements, conditions for producing arrays, including washing procedure and drying conditions (e.g., temperature, length of time), methods used to attach the target material to the matrix, composition and spatial layout of arrays or other spatially fixed platforms, specificity for markers or targets, and stability of the platform. In addition, annotation (gene ID, etc.) associated with each nucleic sequence should be described in its derivation and source.

Section II.A.1 briefly discusses sample collection, storage, and handling conditions. The guidance should elaborate on this since it is a crucial issue in generating comparable samples across the industry and across clinical centers.

Additional comments.

- The terminology is somewhat ambiguous. As a single example, "target" and "probe" are used in various ways. We advocate a definition list or glossary to accompany a revised guidance.

- We recommend that software and LIMS must conform to applicable parts of 21 CFR Part 11 electronic records requirements – this should be mentioned in the guidance.
- The guidance should refer to existing quality system regulations (QSR) and other existing software guidance documents for software used for data interpretation and validation.
- The derivations of annotation of array elements (e.g., with associated gene identifiers) should be documented
- Reagent Assay components should be well characterized – including buffers, enzymes, signal detection systems such as fluorescent dyes, chemiluminescent reagents, other signaling reagents, controls and/or calibrators: negative and positive controls, characterized as internal or external.
- We recommend that documentation and reported performance of all assay validation experiments be included in assay validation.

II.B. Validation of specific performance characteristics

Much of the terminology here is ambiguous, and should be revised for consistency with an accompanying glossary. In particular, the following terms require clarification -

- Sensitivity: should be clarified when used for classification.
- Positive sample: It is not clear what "positive sample" means here for expression measuring devices. For genetic tests, a positive sample means a disease allele is present. For expression tests, it is unclear if this means that "absolute expression level" is above some lower bound. Two-channel cDNA microarrays are more accurate on relative expression level than absolute expression level. FDA should clarify if this means that the relative expression ratio to a "designed" reference sample is above a lower bound.
- Assay range: This should be described as the dynamic range of the asssay.

II.C. Array and data processing

Paragraph *3. Computational methods for data processing* - The documentation of computational methods should be applicable to transformations of data, e.g., from raw images to spot-level estimates to gene-level expression estimates in the case of RNA expression microarrays. A number of other criteria should be documented separately if not documented as part of the data transformations listed above. These include: use of background measurements, assessment and treatment of poor quality features, and normalization.

The least burdensome approach to data handling may be aided by the support and adoption by the Agency and the industry of emerging standards for microarray data interchange such as the MIAME standard. If it is too early to support such a standard at this stage, CDRH should be open to revision or amendment of the guidance as standards develop.

III. Comparison studies using clinical samples

This section should be expanded and should reference other standards or guidance where appropriate. It should be noted that comparison studies may only rarely be appropriate for these types of novel devices.

- Comparison to another well-characterized device should not use percent agreement if a high level of agreement means both devices are making the same mistakes.
- Sensitivity and specificity are a tradeoff. We recommend a Receiver Operating Characteristic (ROC) curve, rather than a report of individual specificity and sensitivity numbers, if possible.
- We recommend the use of appropriate reference standards for comparing RNA expression profiling results across multiple sites.

IV. Clinical evaluation studies comparing test performance to accepted diagnostic procedures

The reference to "clinical truth" is of concern. Even if "clinical truth" is eventually known, the diagnosis or classification based on the multiplex test in an application may be unique when compared with existing diagnoses, and the benefit to patients of an improved diagnostic may outweigh the waiting time required for patient outcomes (and thus 'clinical truth') to be revealed. In practice, comparisons against existing diagnostic procedures may rarely be available.

A possible solution would be a stratification of product types depending on intended use, which may be valuable in determining the most appropriate path for evaluation of new diagnostics based on multiplex tests.

The guidance should provide more clarity on the validation process. This is a critical issue for genomics, proteomics and metabonomics if a set of changes associated with simultaneous measurements of multiple analytes are defined as the test.

Paragraph *IV.B.3.* states, *"[v]erify with a second detection system"*. This may not be appropriate or possible in all cases. If a second detection system is not available, an alternative verification procedure should be described. Even when a second detection system is available, it is not clear a second method can be used.

An example may be helpful. How can a second detection system such as quantitative RT-PCR be used to verify the results of an expression array? qRT-PCR can be used to verify expression changes for a few individual genes, and an array of qRT-PCR reactions may be used to generate independent predictive expression patterns. (It should be noted that the second method under consideration is now *itself* a multiplex method, and thus falls under the same guidance.) But the correlation between the quantitative expression measurements of qRT-PCR for individual genes and less sensitive and semi-quantitative hybridization results varies between genes, and may not be an appropriate validation criterion in any case. The goal is the analytical and clinical validation of the test method. A test method using an expression pattern may be intentionally developed to withstand variability in expression levels of individual genes. The test may be valid even when

results of individual genes are not well correlated between platforms. qRT-PCR verification for a subset genes would not be very helpful for the verification of the clinical validity of the microarray assay.

Appendix I

Paragraph *4. "Determine sample size prior to beginning the clinical study".* This may not always be possible. The guidance should describe acceptable alternative data sets.

Paragraph *7.* This is an example of singling out specific data types (in this case, genotyping) for a general problem (data summarization). See comments on "Genetics versus Expression" above.

The appendix is missing paragraphs 10-12.

Paragraph 15. This should be clarified for expression and genotyping data.

Appendix II

This appendix seems to be a compilation of helpful hints for analysis, but it only scratches the surface of methods that could be applied to multiplex data. For example, there is no description of haplotyping methods. If this appendix originally comes from other material, the reference to the original document should be given and this section should stress how multiplex tests may be different.

The method for comparison between studies is not clear. If there is no measure of truth, why does the measure of a new method have to be in agreement with the measure of a control method? In the paragraph on *Use of null hypothesis testing*, the use the slope of linear regression to test agreement (being close to 1) is suggested. Even if the slope equals 1, it still could mean both methods are not giving the correct value. If the new method measures the expression level lower than the control method in a consistent way, such that the classification with a proper cutoff value is as good as the control method, does this mean the two methods are in agreement?

Without a "true" measurement, how can reproducibility be evaluated? If we knew the true expression level of a gene, for example, between 1 and 1.2, we would like to have the mean of the expression level between 1 and 1.2 at every site (for site-to-site reproducibility) with a small standard deviation and small number (or none) of individual measurements outside of (1, 1.2). Similar comments apply to array-to-array reproducibility, day-to-day reproducibility, and sample-to-sample reproducibility, and so on.

For a diagnostic device, single sensitivity/specificity or positive or negative predicted value is not sufficient, because a device is usually not used under a single point of ROC. ROC should be provided for the full range, which the device is intended to use. Without a ROC curve, bias may occur since investigators can choose a best point of sensitivity and specificity for their device for the comparison. Sampling plan and acceptance criteria for performance characteristics for analytical validation should be provided. The draft guidance recommends NCCLS guidelines be consulted. This, however, seems to be

inconsistent with the recommendations given in other guidances for analytical validation, which requires the performance characteristics such as accuracy, precision, linearity, range, LOD, LOQ, and ruggedness as described in the USP/NF be considered.

It is suggested that qualitative and quantitative analytical validation be distinguished in the draft guidance. Statistical methods for assessment of qualitative and quantitative analytical validation could be very different under different study designs.

The leave-one-out method is known to be biased, and should not be recommended. A better approach in the absence of a completely independent dataset would be multiply repeated X-fold cross-validation (where x = 5 or 10).

The draft guidance does not provide information regarding what sample sizes are required for specific studies.

As indicated in the guidance, FDA may request different types of data and statistical analyses in pre-market applications for *in vitro* diagnostic tests. However, little information detailing this possibility is provided.

Variation associated with multiplex tests is always a concern, which may have an impact on the accuracy and reliability of the test results. The draft guidance does not address the issue of variability that may come from various resources.

We thank you for this opportunity to comment, and trust that our remarks will be helpful to you in elaborating the next version of the guidance.

Please feel free to contact me if you have any questions.

Sincerely,

Alice E. Till

Alice E. Till, Ph.D.

Cc E. Mansfield
M. Schoonmaker