Multiplex Tests for Heritable DNA Markers, Mutations and Expression Patterns; Draft Guidance for Industry and FDA Reviewers

Draft Guidance
27 Feb. 2003 Draft

This guidance document is being distributed for comment purposes only.

Document issued on: April 21, 2003

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to Dockets Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact Elizabeth Mansfield at 301-594-1293.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Office of In Vitro Diagnostic Device Evaluation and Safety
Division of Immunology and Hematology Devices

TABLE OF CONTENTS

Preface

Public Comment

For 90 days following the date of publication in the Federal Register of the notice announcing the availability of this guidance, comments and suggestions regarding this document should be submitted to the Docket No. assigned to that notice, Dockets Management Branch, Division of Management Systems and Policy, Office of Human Resources and Management Services, Food and Drug Administration, 5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD 20852.

Additional Copies

Additional copies are available from the Internet at:
http://www.fda.gov/cdrh/oivd/guidance/1210.pdf or to receive this document via your fax machine, call the CDRH Facts-On-Demand system at 800-899-0381 or 301-827-0111 from a touch-tone telephone. Press 1 to enter the system. At the second voice prompt, press 1 to order a document. Enter the document number 1210 followed by the pound sign (#). Follow the remaining voice prompts to complete your request.

Multiplex Tests for Heritable DNA Markers, Mutations and Expression Patterns; Draft Guidance for Industry and FDA Reviewers

INTRODUCTION

I. Purpose

This document is intended to provide guidance on preparing and reviewing premarket approval (PMA) submissions for multiplex tests, or tests that assay multiple analytes simultaneously. Array-based tests, such as oligonucleotide, cDNA, protein and tissue arrays, are a subset of multiplex tests. The following recommendations for elements of a multiplex test submission apply to array-based tests as well as other types of multiplex tests. This guidance primarily considers nucleic acid based analysis, but many of the principles apply to protein and tissue arrays as well.

FDA is committed to working with manufacturers to facilitate the transfer of multiplex and microarray technology into the marketplace. We recognize that new tests based on this technology have the potential to enhance medical care by refining patient diagnosis for disease, disease susceptibility, and drug selection among other potential applications. While the full impact of this new technology is uncertain, the likelihood for improvement in the quality of medical care seems intuitive. FDA is anxious to provide clear guidance to assist sponsors in developing submissions that will support marketing of safe and effective products using this technology.

To date, FDA has little experience with multiplex or array device submissions. This document is intended to initiate dialogue with stakeholders regarding the basic framework for the types of data and regulatory issues that should be addressed in a multiplex device submission. FDA's goal is to establish a set of recommendations that will both define the levels of data needed to establish a reasonable assurance of safety and effectiveness of a device, and suggest the least burdensome path to market for manufacturers of multiplex and array devices.

FDA is requesting that interested parties to comment on this draft guidance within the next 60 days. Following review of the comments we receive, the agency intends to issue a new draft guidance for additional discussion. We are taking this approach because we believe that the public health will benefit from dialogue with the industry about appropriate ways to review this new and important technology. Because we have not had significant experience in this area and because no industry drafts of appropriate review criteria have been made available, we are hoping that this initial overview of the areas we believe should be addressed in submissions will stimulate the kind of interchange that can lead to a more refined draft for further review. All comments will be available on an open electronic docket that will give each stakeholder an opportunity to view and respond to other points of view.

This guidance document provides recommendations for the preparation and review of a multiplex test submission, thereby providing a common baseline from which both the manufacturers and scientific reviewers can operate. Although it is intended for PMA submissions, many of the scientific issues may be relevant to 510(k) multiplex or microarray submissions as well. Depending on claims and information available, multiplex and array submissions are expected to be processed as PMAs, de novo 510(k)s and traditional 510(k)s. We recommend that the sponsor or manufacturer consult with FDA to determine the appropriate type of submission. We also encourage sponsors to consider submitting protocols ("pre-IDEs") before carrying out studies to ensure review issues are addressed and resolved prior to submission of new tests.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. The Least Burdensome Approach

The issues identified in this guidance document represent those that we believe should be addressed before your device can be marketed. In developing the guidance document, we carefully considered the relevant statutory criteria for Agency decision-making. We also considered the burden that may be incurred in your attempt to comply with the guidance and address the issues we have identified. We believe that we have considered the least burdensome approach to resolving the issues

presented in the guidance document. If, however, you believe there is a less burdensome way to address the issues, you should follow the procedures outlined in **A Suggested Approach to Resolving Least Burdensome Issues**. It is available on our Center web page at: http://www.fda.gov/cdrh/modact/leastburdensome.html.

III. Genetics vs. Expression

The measurement of expression changes, whether RNA or protein, will raise different validation and safety and effectiveness questions than the measurement of DNA changes or variations.

"Genetic" tests: DNA differences are fixed, whether germinal or somatic. Results from these tests can generally be described as dichotomous (either present or not present), trichotomous (homozygous A, heterozygous, homozygous B), or categorized (e.g., haplotypes). Interpretation of tests designed to measure these types of differences will be, in most cases, straightforward. DNA array tests nevertheless should be carefully designed and highly reproducible, and have well-established performance. Clinical studies should account for disease prevalences in the populations studied.

Tests measuring expression changes: Expression changes, in contrast to DNA changes, can be responses to a variety of factors. These may include simple individual-to-individual differences, time of day, and specific effect of a therapeutic treatment on a tissue. Results can vary markedly as a result of these factors. Tests to diagnose, predict, or select based on expression patterns may consequently be difficult to interpret. Sponsors of these tests should consider array physical design strategies, quality control (QC), reproducibility and readout/interpretation.

RECOMMENDATIONS FOR THE PREPARATION OF THE MULTIPLEX TEST APPLICATION

The following are areas that we believe should be addressed in the preparation of a submission for a device incorporating multiplex or array technology, whether the device measures genetic or expression differences.

FDA may request different types of data and statistical analyses in premarket applications for *in vitro* diagnostic tests. The information requested depends on the (1) intended use (for example, to detect cytochrome P450 enzyme alleles), (2) indications for use (for example, predictive or prognostic for disease, response, or sensitivity), (3) methodology (for example, polymerase chain reaction), (4) technical interpretation of results (for example, positive for variant alleles), (5) performance characteristics (for example, analytical validity, quality control and assay limitations), (6) clinical validity (for example, false positives and negatives), (7) clinical interpretation (for

example, benefits and risks) and (8) claims made by the manufacturer (for example, effectiveness). Recommendations for addressing these issues follow.

I. Intended Use of a Test or Device

The intended use should specify what the test is intended to measure, why it is measured, and should specify populations to which the test is targeted, where appropriate.

Some tests may have multiple intended uses. FDA recommends a separate application for each intended use that requires unique and separate supporting studies. You should consult the appropriate review divisions in FDA for advice on submitting tests with multiple intended uses.

II. Analytical Validation

A. Design and Manufacturing

Product design, manufacturing, and controls must conform with applicable parts of the Quality System regulation (QSR) as set forth in 21 CFR §820, see also 21 CFR § 814.20(b)(4)(v).

Specifically, 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.

We recommend that submissions include analytical data that demonstrate that the device performs accurately and reliably under given conditions; this may include:

1. Specimen/sample (for each claimed matrix): identity, preparation, acceptance criteria where applicable, and methods for determining label incorporation, probe length, and so forth, for samples that will be hybridized to the array. Also, include specimen collection, storage, and handling conditions.

2. Assay components: including buffers, enzymes, signal detection systems such as fluorescent dyes, chemiluminescent reagents, other signaling reagents, instruments, and software.

3. Controls and/or calibrators: negative and positive controls, characterized as internal or external.

B. Validation of Specific Performance Characteristics: Analytical Laboratory Studies

We recommend that you describe the following performance characteristics for each target, pattern, marker or mutation claimed in the intended use statement:

1. Assay sensitivity: ability to accurately identify positive samples.

2. Reproducibility: Consult NCCLS EP-5A and EP-12A for information on reproducibility studies, http://www.nccls.org/.

3. Validation of cut-off, reference range, or medical decision point.

4. Assay range.

5. Effect of excess sample and limiting sample. Investigate the sample concentrations and conditions that reproducibly yield acceptable results

6. Assay specificity and interfering substances (endogenous and exogenous).

C. Array and data processing

We recommend that you describe the:

1. Optimization of multiple simultaneous target detection/differentiation, for example, hybridization conditions, concentration of reactants, control of specificity.

2. Potential for sample carryover.

3. Computational methods for data processing. We recommend that you develop computational methods using the CDRH software development and validation guidance documents that are available at http://www.fda.gov/search/databases.html.

4. Limiting factors of the array, including saturation level of hybridization.

D. Validation of instrumentation

We recommend that the submission include the following:

1. Characterization: Characterize instruments used in the assay, including how the instrument assigns values to or interprets assay variables such as feature location, size, concentration, volume, drying of small samples, effect on small volume reactions and its impact on test results.

2. Calibration: Describe instrument calibration.

3. Uncertainties: Describe sources and estimates of uncertainties in results introduced by hardware components.

III. Comparison studies using clinical samples

Where comparison studies are appropriate to establish performance of a device, the following items could be used to support submissions:

A. Comparison to another device: Results of comparison studies with another well-characterized or predicate device; usually reported as percent agreement.

B. Comparison to a Reference Method: Results of comparison studies to a validated reference method or clinical diagnosis; usually reported as sensitivity and specificity.

C. Resolution of Comparison Discrepancies: Results of discrepancy testing should be reported; resolution should be performed only using unbiased statistical techniques

D. Identification of analytical/technical false positive or false negative results, estimates of expected assay failure rates.

E. Evaluation of tests employing quantitative measurement techniques: evaluation of random and systematic error in comparison to the predicate or reference method.

IV. Clinical Evaluation Studies Comparing Test Performance to Accepted Diagnostic Procedure(s)

Where clinical studies are necessary to establish safety and effectiveness of a multiplex or array device, you should address the following points:

A. Clinical Data to Support Intended Use

You should provide appropriate clinical data to support each intended use. In some cases, it may be appropriate to include a direct reference to a professional statement or guideline in the intended use statement. We encourage sponsors to consult with FDA to determine the suitability of reference to such statements or guidelines.

B. Clinical Validation

FDA recommends that the following items be addressed in the submission to support clinical validation:

1. Informed Consent and Investigational Research Board (IRB) Requirements: Samples that are used in the clinical portion of the device validation must be obtained in conformance with FDA requirements, see 21 CFR Parts 50 and 812.

2. "Clinical truth": Define clinical truth as it will be used in evaluating the clinical performance of the device.

3. Clinical data: Validate expression patterns, genotype/phenotype correlations, and so on, on a statistically adequate number of specimens for each intended use, including clinical samples for all matrices claimed in the intended use statement; verify with a second detection system (e.g. quantitative RT-PCR) if applicable. When defining the populations used, submissions should include the following information::

    - Number of samples from the normal population with samples summarized according to appropriate demographic characteristics.
    - Number of specimens included in each disease, condition, pathogen, genotype, or group summarized according to appropriate demographic characteristics.

4. Reference ranges: Calculate reference ranges when appropriate. To establish references ranges, FDA recommends that sponsors follow NCCLS C28 "How to Define and Determine Reference Intervals in the Clinical Laboratory", (http://www.nccls.org/).

5. Statistical method: Describe statistical methods used for calculations. Measures of precision, e.g., confidence intervals, should be described and presented.

6. Literature: For some markers, mutations, or patterns in an array-based test system, there may be a sufficient literature base to establish clinical validity with the new test. If a sponsor intends to use literature to support clinical validity, include a summary of available published and unpublished information and/or published clinical data pertinent to the device. When literature is intended to support bridging from analytical to clinical performance, the literature should identify the same technology as the new test and a similar patient population. We recommend that you consult FDA to determine the suitability of literature to supplement or substitute for clinical performance studies.

VI. Clinical Effectiveness of the Device

A. New markers

Evaluations of new markers, mutations, patterns, or other outputs of multiplex tests should meet the FDA standard for clinical effectiveness for their intended use, as outlined in 21 CFR § 860.7.

B. Established Markers

When analytical performance is validated in the specimen matrix claimed, the sponsor may use the medical literature as evidence of the effectiveness of the marker or mutation. If a sponsor wants to use peer-reviewed literature to support effectiveness, you should furnish copies of all relevant articles and provide a justification for the use of the literature in place of clinical studies. The sponsor should establish comparability between the new device and the device used in the published literature in order to ensure that the data can be confidently extrapolated.

Appendix I: General considerations for planning and evaluating clinical studies.

The following are general recommendations that may be used when planning and evaluating clinical studies. We recommend that you consult with FDA review divisions to determine the most appropriate strategies for clinical studies. You should:

1. Describe all protocols for external evaluation studies. Clearly define the study population and inclusion and exclusion criteria and the chosen clinical endpoint. If literature is to be used, the study population, inclusion/exclusion criteria, and endpoints should be clearly explained in the publication and be reflective of how the device will be used in practice. The study populations and endpoints should correspond to the intended use and claims of the manufacturer.

2. Use investigational sites appropriate to the intended use and claims being sought. Efforts to define population sampling bias should be clearly outlined when this issue may affect performance.

3. Establish uniform protocols for all external evaluation sites prior to study and follow them consistently throughout the course of data collection.

4. Determine sample size prior to beginning the clinical study. The sample size should have sufficient statistical power to detect differences of clinical importance for each marker, mutation, or pattern. FDA will consider alternate data sets in cases with a small available sample size, for example, a disease or condition having a low prevalence or with markers or mutations of very low frequency.

5. Describe the sampling method used in the selection and exclusion of patients. If it is necessary to use archived specimens or a retrospective design, provide adequate justification for why the sampled population is relevant to your patient population.

6. Analyze test data both by separate investigator/site and pooled over investigators, if statistically and clinically justified. For heritable markers and mutations, gender and race or ethnicity demographics should be similar between sites if data pooling is otherwise appropriate.

7. Display genotype data in the appropriate N x N table (e.g., 3 X 3 for homozygous wildtype, heterozygous and homozygous mutation) where applicable.

8. Support the intended use claim for the device with data that are representative of the population for whom the device is intended. Include a diversity of ethnic groups if the marker/mutation varies according to ethnicity.

9. Include samples from individuals with diseases or conditions that may cause false positive or false negative results with the device (i.e., within the differential diagnosis), if appropriate.

13. Account for all patients and samples. Perform appropriate data audits and verification before submitting to FDA. Give specific reasons for excluding any patient or result after enrollment.

14. Perform studies using appropriate methods for quality control. Describe the materials and methods used to assess quality control.

15. Describe how the cut-off point (often the distinction between positive and negative, or the medical decision limit) was determined, if appropriate. Describe the performance characteristics the cutoff identifies for each marker/mutation. The description of how each cut-off was determined should include the statistical method used (e.g., receiver operating characteristic curve).

16. The "Minimum Information About a Microarray Experiment" (MIAME) guidelines (see www.mged.org/miame for more information) describe many of the sources and types of data and information that should be available for most types of microarray studies, whether they are used to support drug development or diagnostic device submissions.

Appendix II: Statistical Considerations for Analyzing Array Data

Expression arrays: Appropriate statistical analyses for discriminating subjects into groups (e.g., normal, diseased) include supervised analyses (e.g., discriminant analysis, multinomial regression, support vector machines). Unsupervised analyses that allow for discovery of new groups would also be considered (e.g., cluster analysis, pattern recognition, self-organizing maps, factor analysis) as a basis for building diagnostic categories. While such analyses can be complex, they are generally used only to establish the cut-offs for discriminating between pre-specified groups; actual performance of tests could be evaluated with simple statistical analyses of sensitivity, specificity, percent agreement, positive and negative predicted values, among others. Evaluation of the performance of cutoffs should be based on a dataset that is independent of the dataset used to establish the cut-offs, otherwise the performance will tend to be overstated. Alternatively, statistical methods can be used to correct for this bias (e.g., the leave-one-out method, the jackknife, or the bootstrap). Receiver operating characteristic curves are useful for evaluating performance, but cut-offs still need to be chosen to apply the test in practice.

The statistical analysis should account for lack of independence due to, among others, correlation of replicates within chips, samples within runs, runs within days. For example, a variance component analysis could be used to account for correlation. Multivariate measurements by methods such as multiple SNP analysis or multiplex DNA-based tests, complicate comparison of tests. For multivariate measurements, a summary of the measurements could be helpful in making comparisons.

Method Comparisons: Comparisons of tests without a measure of truth are comparisons of agreement and have the following limitations:

- Because the true value (the diagnosis or the quantity being measured) is unknown, comparison of the methods can only establish equivalence, not superiority of one method over another.
- Agreement is not a measure of correctness because both methods could agree on an incorrect value.
- For diagnosis, level of agreement usually depends on prevalence because it depends on whether the true diagnosis is positive or negative. For quantitative measurement, level of agreement often depends on the magnitude of the measurement. When agreement is heterogeneous over a variable, its statistical analysis should be stratified by that variable.
- For diagnosis, relative sensitivity and specificity and discrepant resolution can be very misleading and are not appropriate for primary evaluation of approvability. FDA recommends reporting of positive and negative percent agreement.

Recently, statistical methods have been developed that allow comparison of methods with the unknown true value being measured. These methods might be useful for obtaining estimates of performance with respect to the unknown true value and for establishing superiority of one method over another. However, these methods often make strong assumptions about the correlation between test results, the distribution of the unknown true values (e.g., the prevalence), and the performance of the reference or predicate test (e.g., its sensitivity and specificity). Such assumptions need to be justified.

For quantitative method comparison data without a truth standard, plots are very useful for decomposing the error of agreement into systematic and random error. In a scatter plot of the experimental method measurement (y) versus the corresponding control (reference or predicate) method measurement (x), systematic error can be evaluated by comparing the scatter with the identity line (y=x), which indicates no systematic error on average. Random error is assessed by the spread of the scatter, i.e., the variability between the methods. A Bland-Altman scatter plot of the difference between paired measurements from the two methods versus their average is especially useful for detecting trends in systematic and random errors over the measurement range.

Use of null hypothesis testing: Formal statistical analyses test equivalence of the experimental method with the control method. For example, two methods could be defined as equivalent based on the slope b, from a linear regression of paired measurements, being close to one. For this definition, a valid approach tests the null hypothesis that the b is more than d units away from one, where d is pre-specified to be the smallest clinically meaningful difference. Rejection of this null hypothesis then implies that equivalence has been demonstrated. A common, but invalid, approach tests the null hypothesis that the b equals one. This approach is invalid because insufficient evidence to reject the null hypothesis does not imply sufficient evidence to accept it. In fact, a sufficiently small sample size can be chosen to guarantee that the null hypothesis will not be rejected.

Other considerations: Poor agreement between the experimental and control methods might be simply due to poor repeatability of the control method, even if the experimental method measures the true value perfectly. Duplication of measurements under the same conditions may be needed to identify this problem. This variability or error in the measurements also biases downward the estimate of the slope in a standard linear regression comparing two methods. Alternative regression methods that account for measurement error include Deming regression and Passing-Bablok regression.