

Marie A. Vodicka, PhD
Assistant Vice President
Biologics & Biotechnology
Scientific & Regulatory Affairs



2605 5 JUL 20 110 :00

July 19, 2005

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

Re: Draft Drug-Diagnostic Co-Development Concept Paper [Docket No. 2004N-0279]

Dear Sir or Madam:

The Pharmaceutical Research and Manufacturers of America (PhRMA) is pleased to submit its comments to the Food and Drug Administration's (FDA's) *Draft Drug-Diagnostic Co-Development Concept Paper (draft concept paper)*. 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 about \$40 billion annually in discovering and developing new medicines, PhRMA companies are leading the way in the search for cures.

PhRMA welcomes and supports FDA's initiative to provide guidance on this complex and important topic. PhRMA member companies have worked closely with FDA to clarify how the regulatory processes for interdependent therapeutics and diagnostics can be managed most efficiently. We were pleased to have the opportunity to be a co-sponsor with FDA of the joint workshop organized by the Drug Information Association to discuss this topic, which was held in Washington, D.C. on July 29, 2004. We appreciate FDA's action in issuing a concept paper to begin to address the complexity of the co-development of drugs and diagnostics, and we look forward to the issuance of formal guidance in due course.

Before commenting on the details of the draft paper, we would like to address the issues that concern us most.

1. The Presented Model for Co-development is Generally Unrepresentative of Reality

The paper states (Section 4.1, para. 2, p. 10), "*Ideally, a new diagnostic intended to inform the use of a new drug will be studied in parallel with early drug development (phase 1 or 2 trials) and diagnostic development will then have led to prespecification of all key analytical validation aspects for the subsequent (late phase 2 and phase 3) clinical studies.*" We agree that this may be the ideal situation but it is, in the reality of drug development, rarely the case. The

Pharmaceutical Research and Manufacturers of America

1100 Fifteenth Street, NW, Washington, DC 20005 • Tel: 202-835-3561 • FAX: 202-835-3597 • E-Mail: mvodicka@phrma.org

2004N-0279

C6

scenario given in Figure 1 is unrepresentative because it does not take into account the significant differences in normal development timelines between therapeutic and diagnostic products. We strongly suggest that the eventual guidance address the regulatory processes for the more common situation in which adequate qualification of a biomarker or the need for a diagnostic is only identified late in Phase 2 or in Phase 3 of the drug/biologic development. The guidance should also provide recommendations on the sequence of regulatory interactions and the potential for cross-referencing data for the regulatory development of a diagnostic that is only commenced when the related therapeutic is in Phase 3.

Further, we are concerned that the draft concept paper limits the definition of “co-development” to the development of a single test with a single drug. We recommend that any future guidance on co-development recognize the realistic possibility of having multiple tests for the same biomarker or multiple biomarkers identified. Further, there should be some discussion of the development of follow-up tests (to be based on predicate) to make clear whether “clinical utility” as defined in the concept paper needs to be established for each new diagnostic or only the first. We recommend that the scope be carefully considered so that the regulatory pathway does not burden the industry by creating hurdles to adding an improved diagnostic test to a drug label.

2. Requirements for Establishing Clinical Utility Are a High Hurdle for Diagnostics

We are concerned by the concept paper’s definition and use of the term “Clinical Utility”. Regulatory requirements for demonstrating clinical utility of a diagnostic agent currently require a pre-market approval application (PMA). We recommend that the concept paper or any subsequent guidance does not in any way limit the pathway to approval of the diagnostic product to the pre-market approval application (PMA) process alone, thereby allowing use of the least burdensome pathway (510(k), or *de novo* 510(k)) where appropriate on a case-by-case basis.

The discussion of the need to establish clinical utility (Section 6, page 22) suggests that two confirmatory clinical trials might be needed to support the approval of the diagnostic under this co-development process. Although this is a normal expectation for establishing clinical efficacy of a new therapeutic agent, this is not a requirement for a PMA and this is an excessive request in terms of the Least Burdensome Provisions of §205 of the FDA Modernization Act of 1997 (FDAMA).

3. Interaction with the Voluntary Genomic Data Submission (VGDS) Process Should Be Clarified

PhRMA supports FDA’s recognition (page 3) that “*optional or exploratory tests that are not intended for further development or those that do not affect the results of clinical trials*” are not within the scope of the paper. This distinction between the use of pharmacogenetic data

for regulatory decision-making and for drug development is important, since over-regulation of the latter could stifle innovation.

In the recently finalized guidance on Pharmacogenomic Data Submissions (March 2005), FDA has established clear guidelines for early development situations where the interactions between the sponsor and the Interdisciplinary Pharmacogenomics Review Group (IPRG) should not influence the Agency's subsequent regulatory decisions. Point 2, page 6, of the draft concept paper states, "*What additional information is needed for information previously submitted under a VGDS if a VGDS becomes a required submission?*" FDA's intent here, and the operation of the VGDS process in conjunction with the pre-IDE process, are not at all clear. We strongly recommend that FDA add a specific discussion of the impact of using the VGDS process for the initial submission of data on the development pathway for a drug-diagnostic co-development product. It would be important that the potential for obligations to submit data in the future under particular contingencies be highlighted.

4. Guidance Must Represent Views of All Involved Centers

The draft concept paper seems to represent disproportionately the views or concerns of the Center for Devices and Radiological Health (CDRH). It does not seem to be equally reflective of the views of the Center for Drug Evaluation and Research (CDER) or the Center for Biologics Evaluation and Research (CBER), which would be crucial in a true co-development paradigm. We recommend that FDA ensure the resulting guidance adequately reflects the views and concerns of all of the review Centers.

5. Some Guidance Should be Given on Labeling Principles

While we are committed to the concept that labeling is highly product specific and should be left to individual company/FDA negotiations on a case-by-case basis, we suggest that more guidance should be given on when a specific diagnostic test should be included in the drug label vs. a general comment on use of a test. For example, will there be situations when a therapeutic and all of the diagnostics that have been approved for use with it will be cross-labeled (i.e., the therapeutic's label will list all of the approved diagnostics by name and the diagnostics' labels will all mention their use with the therapeutic)? If so, what are they? We believe that labeling of separately marketed drugs and diagnostics that are used together does not need to be identical, but should not be contradictory. The regulations (21 CFR 3.2(e)(3)) and the Intercenter Agreement ("ICA") between the CDER and CDRH support the concept of flexibility with respect to mutually conforming labeling such that the intended use, indications and effect should be consistent with, but not necessarily identical to, the approved drug labeling. The review procedures section should be expanded to include greater detail on inter-center review considerations and potential FDA-industry interactions.

In our view, the impact of device co-development on the clinical trial program is poorly addressed throughout the document. The recommendation of having a test analytically and clinically validated early during the drug development, as proposed in Figure 1, would require large size early phase clinical trials. This would significantly increase development time and cost. We find many of the procedures discussed in the document more applicable to diagnostics that are initially developed independently, with no specific tie-in the a co-development program.

6. Clinical Study Design and Analysis Considerations

The role and acceptability of retrospective analysis in the development of a co-development product should be specifically considered in the guidance. The requirements for clinical evaluation in a test-negative population (assessment, study design etc.) should be addressed on a case-by-case basis, being dependent upon a number of factors including whether safety or efficacy is the consideration, the risk/benefit profile of the drug and the role of the test in the established standard of care.

7. Special Protocol Assessment (SPA) Should be Available for Co-Developed Products

We propose that SPA should be available for co-development of drugs and diagnostics. It may be appropriate for members of CDRH to consult with the CDER/CBER review division in the SPA process. The agreement made with the sponsor on Phase 3 study design under an SPA agreement should be binding on both the therapeutic and diagnostic review teams.

SPECIFIC COMMENTS

Section 1: Introduction, Background, and Scope

1.3 Scope

The draft concept paper does not limit its scope specifically to pharmacogenomic tests, and it would be helpful if a few examples of the different types of tests covered by the guidance could be given.

Figure 1 does not highlight the large differences in the development life cycle of therapeutics and diagnostics. The latter are commonly brought to market within 2-4 years of the start of development, depending on the scope of the test system and its intended use. The development of the diagnostic test includes (1) prototype assay (2) final assay for analytical validation, (3) clinical validation. Changing the platform after analytical validation (as shown in Fig. 1) can present problems, depending on the scope of the changes in the test system.

Figure 1 does not depict the common scenario whereby emerging data from Phase 2 or Phase 3 studies constitute the impetus for consideration and potential development of a biomarker to select patients.

The third bullet on page 4 under clinical test validation discusses predicting an “associated disorder”. We suggest that this should be termed a “clinical state” instead.

Section 2: Review Procedure Issues

2.1 Co-development and Intercenter Review Considerations

We appreciate FDA’s recognition that co-development products may, or may not, be combination products. As mentioned in the general comments, however, this section should address the point that there is a significant difference in timelines for therapeutic and diagnostic development decisions and regulatory interactions and decision-making.

2.2 Procedures

Figure 2: Drug Device Co-development Process:

The guidance should also consider a co-development pathway that starts during the Phase 3 of drug development. It may help to adapt Figure 2 to show how to time events to allow the preparation, filing, review and approval of the PMA or 510(k) for the diagnostic test during the same timeframe in which the NDA or BLA is reviewed and approved. One possible timeline in such situation would be that the pre-IDE meeting takes place in mid-to-late Phase 3 drug development.

There is no discussion of how the VGDS discussion process will advance to discussions aimed at regulatory decision-making for a co-development program. This will be an important pathway for many products and needs to be clearly described.

It would be helpful if FDA would expand on the option for ‘sequential’ approval as noted in 2.2 (7).

Section 3: Analytical Test Validation

Many of the processes and procedures described in this section are common to all IVD systems and not unique to diagnostic tests being co-developed with a drug. It is suggested that the guidance refer to specific CDRH guidances and not be repetitive.

3.1. General Recommendations to Support Premarket Review

The paper states, “*Study design should take into account statistical considerations for both the drug and the diagnostic.*” There should be recognition that clinical validation of the diagnostic product may come from clinical trials that did not take into account statistical considerations for the diagnostic. It would be helpful if FDA provided specific examples in the

document pertaining to analytic validity, clinical validity and clinical utility to assist in the definitions and further illustrate the concepts.

It is often the case that a sponsor is not in a position to design a Phase 3 study (for clinical validation and utility assessment) based on biomarkers discovered in Phases 1 or 2, since larger trials may be required to determine the markers associated with a given response. We request that FDA consider this scenario in its recommendations.

We suggest that the following sentence should be changed to “*Clinical trial specimens should be banked in storage conditions **adequate** to enable subsequent test development and/or retrospective hypothesis generation or confirmation of test performance.*” The word “optimal” should be removed, otherwise it would be necessary to define what “*optimal conditions*” would be.

We would urge FDA to ensure that a globally consistent recommendation be made for sample storage/banking, especially given the current activity of the European Medicines Agency (EMA) with regard to this aspect.

3.3 Analytical Studies

Adequate bridging between the analytical test and clinical diagnostic will be required. If the data are blinded, the correlation with efficacy/safety assessment in the clinical program needs to be established. If the analytical validity of an IVD is proven, but the clinical utility is not evident in the clinical trials, could the IVD still be marketed under a 510(k), with claims of analytical validity only, or should the IVD intended use include a statement that clinical utility has not been fully established?

3.5 Analytical Validation of Changes to a Device in Late Stages of Development

The paper states, “*The stability and validity of using banked samples should be documented by demonstrating that the original assay results can be repeated at the time when the new assay results are obtained from the specimens.*” We believe this statement is unreasonably prescriptive. We recommend changing it to, “*The stability and validity of using banked samples should be documented and information supporting sample integrity should be provided.*” This recommendation is consistent with CDRH’s Guidance, “Drug Metabolizing Enzyme Genotyping System.”

Section 4: Preclinical Pilot Feasibility Studies

4.1 Introduction

We have already expressed our concern that the scenario envisaged here is hardly representative of reality, and other scenarios must be considered in the guidance.

4.2 & 4.3

The premise appears to be that the basic observation upon which a test result will be based varies continuously. However, for pharmacogenetics specifically, the test result will be based on genotype and it would be helpful if the document contained specific recommendations for this scenario.

Specification of subpopulations by a biomarker device is the most problematic development aspect, as illustrated by these two sub-sections. Use of a cutoff defined *a priori* is one method of providing statistical rigor, but it is not free from problems. If the analysis is based on small studies with a sparse sampling near the identified cutoff, the imprecision of the estimate will greatly reduce the power of subsequent studies.

The example of IHC staining scores is a poor illustration of this problem, because the indeterminate nature of a 2+ score is more reflective of the imprecision of the test than of the actual association between this level of expression and response. (See the distinction between sections 4.5.1 and 4.5.2). Considering only the extreme values of 1+ and 3+ would increase the association and the consequent test performance. In the case of more continuous variables, the variable may be precisely measured, avoiding the problems in the example. However there may still be a continuous rather than threshold association between expression and clinical response, resulting in strong predictive ability at either end of the range and poor predictive ability in the middle, leading to an apparently imprecisely determined threshold.

A more rigorous discussion of the statistical and design issues inherent in specifying an expression level cutoff should be developed.

4.4 Use of Receiver-Operating Characteristic (ROC) Curves to Aid in Setting the Cutoff Values for Diagnostic Tests

We feel that this is a curious discussion coming after Section 4.2, where use of a statistically identified cutoff is prescribed. Use of an ROC curve to identify a cutoff is merely the application of some economic weights to positive and negative prediction errors. These prediction rates are still sample specific and just as prone to over-specification as the maximally predictive cutoff is. The ROC analysis discussed pertains only to a continuous variable (e.g. expression level), not for a test based on genotype or on a set of genotypes. We request that FDA clarifies the purpose of this section.

Section 5: General Approaches To Define Clinical Test Validation

It is stated in the paper "*Clinical test validation of a new diagnostic for use in selecting drug therapy or avoiding drug therapy should be characterized by studying the test in relation to the intended clinical outcome in patient subgroups with and without the analyte of interest.*" Clinical test validation of a pharmacogenetic test may not be done in patient subgroups without the analyte of interest when that analyte defines the disease (e.g., chronic

myelogenous leukemia). This possibility also needs to be accounted for in the fourth bullet point on page 18.

The concept paper focuses on a test in which there are positive and negative results with a single cutoff value (e.g., responder/non-responder). While this simplification is useful to present some concepts, the guidance also needs to account for tests that have more than two categories, return continuous values that place an individual in a specific portion of a benefit/risk spectrum, or that provide a range of probable outcomes for individuals based on their genotype. To illustrate, the following scenarios are discussed.

- A possible example of a relevant efficacy biomarker is one that identifies three groups of asthmatics who can be expected on average to have a 5, 12 or 20% increase of FEV1 after 2 weeks on drug (or to put it another way, have a 20, 50 or 85% probability of attaining a clinically meaningful response after 2 weeks on drug).
- A possible example of a relevant safety biomarker is one that identifies three groups of cancer patients who can be expected to have different ranges of metabolic changes on drug.
- Positive predictive value (PPV) and negative predictive value (NPV) may not be the main metrics if the outcome is continuous (see examples above). Hence, specification of cutoff values may not be so important for many pharmacogenetic tests.

It is not clear whether FDA is suggesting that if a study features an ‘enriched’ sample, meaningful estimates of NPV and PPV cannot be derived. We would propose that an estimate of PPV and NPV can be made for ‘generalizability’ to clinical practice as a function of the sensitivity and specificity observed in the study sample and by assuming a specific prevalence of cases in an unselected population (also applicable to Addendum C).

The limitation of discussion to categorical endpoints is unnecessary and too restrictive. Other sections deal with the potential of a device to identify patients with different potential to respond. (e.g., Section 6.2, paragraph immediately following bulleted list). Continuous measures of response would be just as valid and potentially more powerful outcomes.

Clinical trials will generally consist of enriched populations. When results can be generalized to the broad target population, measures such as odds ratios are robust to differences in positive test rates. Moreover, case/control frequencies can be variable across subpopulations and time further reducing the utility of NPV/PPV for assessing a device.

Section 6: Clinical Utility

We suggest that this section should further address the scenario where a definitively identified predictive pharmacogenetics marker (or set of markers) will not have occurred to the point that a Phase 3 clinical program can be specifically designed as a consequence.

- For section 6.2, this section should be expanded to include discussion of additional aspects that need to be considered such as what comprises ‘clinical utility’ and potential differences for safety vs. efficacy considerations.
- FDA is requested to clarify the scenarios for potential evaluation in a test-negative population. We propose that this be addressed on a case-by-case basis, being dependent upon potentially a number of factors including whether safety or efficacy is the consideration, the risk:benefit profile of the drug and the role of the test in the established standard of care.
- FDA is requested to clarify the recommendations for the role and acceptability of retrospective analyses for the validation of markers. This is a critical component of the regulatory decision-making process.
- As with all research, consents should cover the intended scope; however, we consider that requirements for overly specific details in consents may hamper innovation. We urge FDA to recognize the rigorous coding and handling mechanisms already used and further, given that these are outlined by the EMEA as well, that this also will facilitate development of a global framework.

A detailed discussion of clinical utility from a test standpoint is needed. FDA should work with all of its stakeholders on this effort. The definition of clinical utility should also be such that the requirements of other HHS departments would accept the concept as well.

It is stated *“To confirm clinical performance, including clinical utility, additional clinical studies may be called for to avoid post-hoc specification of the diagnostic cut-off points.”* The paper should recognize that a prospectively defined analysis of drug clinical trial data could be used to clinically validate the performance characteristics of the diagnostic test, negating the need to conduct additional clinical studies.

6.1 Coordinating Drug and Diagnostic Studies

6.1.1 Study Objective and Timing

The concept that there will be a prospective study simultaneously assessing both drug response and the quality of the diagnostic is ideal, but it must be acknowledged as often unobtainable.

Figure 3 and its accompanying texts should be modified to allow for the possibility that the diagnostic statistical analysis may be conceived and conducted after the drug clinical trial is completed.

6.1.2 Clinical Trial Design Considerations

Because test outcome rates are unknown, there is limited ability to power such a study or control for confounding factors. This design should be classified as a preliminary trial that would need to be followed by a more robust confirmatory trial. We recommend that FDA address the limitations around powering and analyzing such a study.

6.2 Issues to Consider in Selecting Study Populations

The paper states, *“In some cases, sponsors may wish to use enriched study populations to evaluate the likelihood of response to a drug treatment, such as in a proof of concept trial in early phase 2 of drug development... Consideration should be given to how enrichment will relate to the ultimate claims made for the drug being evaluated.”* The use of a pharmacogenetic test for a proof-of-concept trial is not a registration issue. Justification of the enrichment technique is an internal business decision for the sponsor, as long as there is no intent to also enrich the pivotal Phase 3 studies.

Further, the paper states, *“Optimally, further confirmatory testing would be performed in prospective trials.”* It should be recognized that this will be the exception rather than the rule in development programs for regulatory co-approval of drugs and tests. We recommend that the guidance, when issued, address the “usual” situation instead of describing only scenarios considered “optimal.”

It is stated, *“The approach to these associations and analysis should be prespecified in advance and not after the study is completed.”* It must be made clear that the intent to perform the genetic analysis should be specified in advance, but that the definite analysis plan may only be decided upon after the clinical analysis has been completed (in fact, in many situations this will be preferred).

The analyses referred to can just as easily be done using the designs presented in Figures 3 and 4 using subsample analyses of patients with positive and negative tests, along with an analysis ignoring test result. The statistical validity of subset analyses by test result status is no different from any other secondary outcome and should be treated the same. Because test outcome rates are unknown, there is limited ability to power such a study or control for confounding factors. This design should be classified as a preliminary trial that would need to be followed by a more robust confirmatory trial.

In the first and third bullet points, page 18, these considerations apply only to enriched pivotal Phase 3 studies.

It is stated that *“In cases where the testing is done as an ancillary part of the trial (i.e., not incorporated into the trial design or primary outcomes), resulting associations between test results and clinical outcomes would usually be considered exploratory and therefore these results would be more appropriate for assessing clinical test performance or generating hypothesis about clinical utility rather than confirming clinical performance or utility.”* The

paper, as written, appears to recommend that additional prospectively designed confirmatory studies are necessary for confirmation of observations obtained from an ancillary part of a clinical trial. In reality, FDA's Least Burdensome Approach, as required by statute, may permit use of such data without confirmation.

6.4 Verification of Clinical Test Utility – Statistical Considerations

It is stated that "... *the analytical characterization of a diagnostic test should be based on a dataset that is independent from and prior to the prospective or retrospective samples on which it is to be clinically verified.*" Clarification of what constitutes an independent dataset for analytical characterization would help. That is, the same samples should not be used for both analytical and clinical validation, but different subsets of the same therapeutic clinical trial should be acceptable. A more complete discussion of datasets and references to specific statistical papers in the topic of validation sets would be helpful.

The paragraph on "*post-hoc characterization of a test*" is misleading because it does not highlight the prospective (genetic)-retrospective (clinical) approach. Again more discussion of the statistical considerations with references is needed.

It is not clear what is meant by "[a] *dataset that is... prior to the prospective or retrospective samples on which it is to be clinically verified*". Does FDA mean to suggest the parameterization of the test, including measures and cutoffs for them, be defined prior to the analysis of the datasets?

6.5 Comments on Drug Efficacy and Safety Studies

Proofing comment: *Figure 3* in the text (end of second paragraph) should be *Figure 5*.

Addendum A: DEVICE DESCRIPTION– Examples of Elements to be Described

We propose an additional aspect: Evidence should be provided that the ruggedness of the device has been studied in a systematic fashion. Ruggedness refers to the ability of the device to give reproducible results even when slight deviations from recommended conditions are used in operating the device. In the event that deviations cannot be tolerated for some factors, then this should be clearly defined in the operating instructions, and allowable tolerances should be specified.

Addendum B: STUDY DESIGN –Examples of Issues to be Considered

3. Analyte concentration specifications (page 28).

A corollary for these considerations should be that no extra (array) elements should be included in an IVD.

4. Cut-off (page 29)

Note that cut-off values are applicable only to tests with categorical outcomes.

Addendum C: DETERMINING IF A DIAGNOSTIC TEST IS INFORMATIVE

The paper states: *"The first step in interpreting diagnostic test results is determining if a test is informative. A test is clinically useful only if it provides information to discriminate between patients with and without the condition or interest (e.g., response or adverse event). Examples of standard diagnostic test performance metrics are clinical sensitivity and specificity".* This is an example of "informational utility". It should not be predicated on response or outcome. This is further reinforced in Addendum C: *"A test is informative only if its sensitivity plus its specificity is greater than 100%. For tests with a combined sum of more than 100%, the strength of the test should be considered in terms of both numerical and clinical impact of the combined numbers. Obviously, the closer the sum comes to 200% (sensitivity and specificity each of 100%), the better the test performs. However, values between 100% and 200% that are considered clinically meaningful would depend on clinical rather than mathematical considerations."*

We appreciate the opportunity to comment on this draft concept paper and trust our comments will be useful to FDA in evolving the guidance.

Best regards,



Marie A. Vodicka, PhD