December 16, 2008
Oncologic Drugs Advisory Committee Meeting

FDA Briefing Document

EXECUTIVE SUMMARY

The selection of a drug based on genomic biomarker profile is desirable because it limits drug exposure to patients who will benefit/are most likely to benefit from drug treatment, avoids drug use in patients who will be/are likely to be harmed by drug treatment, or enhances safe use by optimizing drug dosing. In the ideal case, the development of the assay methodology for the genomic biomarker should be an integral part of the clinical drug development program, such that the clinical studies required to establish the efficacy of the drug and those needed to establish the prognostic and predictive value of the genomic biomarker as measured by a well-characterized assay occur in tandem. FDA strongly endorses such scientifically guided drug development as part of FDA's Critical Path Initiative. The principles of this Initiative include integration of the scientific knowledge of the drug effects in determining patient and dose selection. This ideal approach to drug development continues to be underutilized.

In contrast, there are multiple examples of "retrospective" or post-hoc genomic biomarker assessment or clinical disease characteristics. In the worst examples, this involves a retrospective *re-analysis* of a "failed" clinical trial in which efficacy is purported to be established in a subset defined by a genomic biomarker/patient characteristics without consideration of multiplicity (i.e., data dredging), substantial missing data, and poorly characterized assays. FDA discourages such practices and should not be considered during this advisory meeting discussion. However, FDA also recognizes that there may be legitimate reasons for failure to prospectively consider early in drug development the impact of genomic biomarkers, primarily due to advances in the scientific knowledge of a drug or disease that occur while drug development is ongoing. In this latter situation, FDA seeks guidance regarding how to incorporate new scientific information without compromising the legal mandate to ensure that marketed drugs show substantial evidence of efficacy and are reasonably safe. The levels of evidence needed may differ depending on the claim being sought. For example, restriction of drug use to patient subsets to improve safe use of the drug might not require the same level of scientific rigor as claims for specific drug benefits.

FDA will present a recent example of retrospective biomarker analyses intended to support changes to product labeling and support device approval, to provide a context for the questions posed to the Committee. The following example provides a "real-world" context faced by FDA in which considerations of the type and extent of data needed to support labeling claims must be made. Additional context is provided in the following sections regarding the considerations for an optimal drug-device co-development program and the statistical and clinical study design principles that underlie FDA's statutory requirements for establishing efficacy.

CASE STUDY OF "RETROSPECTIVE" DEVICE-DRUG CO-DEVELOPMENT

ImClone, the license holder for cetuximab (Erbitux) and Amgen, the license holder for panitumumab (Vectibix), described the results of retrospective analyses assessing efficacy outcomes as a function of KRAS genomic status. Based on these data, both companies have proposed inclusion of information on drug use in the subset of patients with metastatic colorectal cancer whose tumors express wild type KRAS.

ImClone submitted summary results obtained in subgroups defined by retrospective KRAS testing of available tumor tissue from four randomized studies in patients with metastatic colorectal cancer. The clinical trial outcomes data for patients with available KRAS tumor test results were re-analyzed by treatment arm and by KRAS genomic status. Information regarding these studies is presented in the following table.

Title	Population	% tested for KRAS	Assay
CRYSTAL (EMR 62202-013)	1st line EGFR positive mCRC, FOLFIRI ± cetuximab	45% (540/1198)	LNA-mediated qPCR based assay
OPUS (EMR 62 202-047)	1st line, EGFR positive mCRC FOLFOX-4 ± cetuximab	69% (233/337)	LNA-mediated qPCR based assay
EPIC CA225006	2nd line, EGFR positive mCRC irinotecan ± cetuximab	23% (300/1298)	Direct sequencing
NCIC 017 CA225025	3rd- line EGFR positive mCRC Best support care ± cetuximab	65% (366/572)	Direct sequencing

Amgen submitted summary results obtained in subgroups defined by retrospective KRAS testing of available tumor tissue in a single randomized study in patients with EGFR-positive, third-line therapy of metastatic colorectal cancer. Information on this study is presented in the table below.

Title	Population	% subjects tested for KRAS status	Assay
20020408	3rd line EGFR positive mCRC, best supportive care ± panitumumab	92% (427/463)	PCR-based assay

In discussions with these manufacturers, FDA stated that the optimal approach would be to conduct an adequate and well-controlled trial, prospectively designed to assess efficacy in subgroups based on KRAS testing by a validated assay. However, the widespread

publication and presentation of the retrospective KRAS analyses have resulted in practice changes in the community, thus, a prospectively designed trial may no longer be feasible.

Given these practical considerations, FDA indicated that retrospective analyses from clinical trials could be submitted provided that

- The trial was adequate, well-conducted and well-controlled;
- The sample size was sufficiently large to be likely to ensure random allocation to each of the study arms for factors (i.e., KRAS status) that were not used as stratification variables for randomization;
- Tumor tissue was obtained in ≥ 95% of the registered and randomized study subjects and an evaluable result (wild type or mutant KRAS) is available for ≥ 90% of the registered and randomized study subjects;
- Prior to analysis, FDA has reviewed the assay methodology and determined that it has acceptable analytical performance characteristics [e.g., sensitivity, specificity, accuracy, precision] under the proposed conditions for clinical use;
- Genetic analysis is performed according to the qualified assay method by individuals who are masked to treatment assignment and clinical outcome results; and
- Prior to analysis of clinical outcomes based on the genetic testing, agreement with FDA has been reached on the analytic plan for hypothesis testing for proposed labeling and promotional claims.

In response, both companies have proposed retrospective testing of KRAS status from large randomized trials that have completed accrual (Amgen) or are actively accruing patients (Imclone). The ongoing studies have been modified to enroll only patients with KRAS wild type (WT) tumors through the completion of the studies.

CETUXIMAB			
Protocol	**Population**	**Status**	**Assay**
CALGB 80405	1st line mCRC, 3-arm, 2 x 3, RCT FOLFIRI or FOLFOX with • bevacizumab (Arm A) • cetuximab (Arm B) • bevacizumab +cetuximab (Arm C)	Ongoing > 1400/2289 subjects enrolled	DxS
N0147	2-arm RCT of FOLFOX ± cetuximab for adjuvant treatment of Stage III colon cancer	Ongoing 2344/2650 enrolled; ↑ sample size to 3768	DxS
PANITUMUMAB			
Protocol	**Population**	**Status**	**Assay**
20050203	1st line, mCRC FOLFOX ± panitumumab	Accrual complete 1150 patients	DxS
20050181	2nd line mCRC FOLFIRI ± panitumumab	Accrual complete 1100 patients	DxS

As discussed below, retrospective analysis of clinical trials are frequently limited by deficiencies in study design elements relating to the biomarker analysis. For example, retention of tumor samples was required for all patients enrolled in the four studies in the table above, however compliance with this requirement has not been determined. In addition, details regarding tumor acquisition and sample handling were not pre-specified in the protocol to ensure that accurate test results could be obtained. Although results of genomic marker assay methodology and read-out can be impacted by sample handling (see recent ASCO/CAP guidelines on HER2 testing), there is no evidence to date that this is the case with KRAS testing methods applied to routinely processed formalin fixed, paraffin embedded tissue.

CONSIDERATIONS FOR OPTIMAL DRUG-DEVICE CO-DEVELOPMENT

The use of laboratory testing to find, explain and treat disease grows in parallel with our rapidly increasing scientific and technical capability to define and measure biomarkers. Through substantial research and development programs, biomarkers become the basis for in vitro diagnostic devices (IVDs) used to help diagnose and treat patients. The informed use of IVDs is at the heart of increasingly personalized medicine.

When the indications for use of an FDA-approved or licensed therapeutic agent are tied to results from an IVD, FDA clearance or approval of the IVD is also needed. A prominent example is the link between results from HER2 testing and the indications for use of the drug Herceptin for the treatment of breast cancer. FDA's regulation of the marketed IVD (a "companion diagnostic") aims to ensure that the claims and

performance characteristics of the test support the informed use of the therapeutic agent throughout the commercial life cycles of both products.

An ideal scenario is one in which the relationship of the biomarker to potential action of the drug is recognized very early – indeed, such a relationship might be the motivation for starting the drug's development. In this setting, many milestones for development of the IVD might be reached in an orderly way. The identity of the biomarker should be established early, along with reliable means for its measurement. If the biomarker has an impact on the natural course of disease (prognosis), such a relationship might be elucidated. Through pre-clinical studies and early clinical trials, support might grow for applicability of the biomarker as an indicator of drug effect. This is the delicate circumstance in which formulation of a specific intended use for the biomarker might emerge, and resources are committed to complete the analytical validation of a fully specified IVD. When a definitive efficacy trial for an investigational drug is undertaken, its design should incorporate a test of the IVD, so that firm conclusions can be drawn concerning both the safety and efficacy of the drug and the safety and effectiveness of the IVD for informing use of the drug. With a trial that is successful from all perspectives, the drug will be approved and the test will be clinically validated and approved for prediction of drug effect.

For many reasons, the ideal scenario is a rarity to date. When a definitive efficacy trial has been conducted and completed without reference to the biomarker, then the urge to test for the biomarker retrospectively in available clinical trial specimens seems inescapably attractive. One justification is that the follow-up for patients accrued to a well-executed efficacy trial is already in hand. A second is that the patients who accrued to the completed trial included both patients who were "positive" and patients who were "negative" for the biomarker of interest – a likely requirement for gaining insight on a predictive claim for the IVD.

Several considerations or caveats should be kept in mind for such retrospective analyses. Retrospective analyses include the potential for bias in "selecting" the biomarker and failure to adjust for multiplicity, since retrospective analyses might be carried out for many biomarkers with only "significant" results carried forward. Another consideration is that (through selective retention or exhaustion) the specimens available for retrospective analysis might not be representative of specimens from the intended use population. When specimens were collected during the trial without reference to any special requirements for measuring the biomarker, some or all of those specimens might be analytically unfit for testing. Another caveat is that early testing for the biomarker might use a device that has not yet been analytically validated, and needs significant modification for reduction to a commercially distributable product with adequate performance characteristics. Beyond these practical caveats related to specimens and testing, there may be concerns about absence of biomarker-oriented stratification, and inadequate power or effect size for both the device and the drug in retrospective re-analyses of clinical trials. One more caveat, if the completed trial is old enough, is that

the clinical and therapeutic context that existed when it was designed and executed might no longer be relevant.

CONSIDERATIONS FOR ESTABLISHING EFFICACY IN SUPPORT OF REGULATORY MARKETING AND PROMOTIONAL CLAIMS

The general approach to examining treatment responses and treatment effects as it is reflected in different subject subpopulations in clinical trials, usually determined by phenotypic characteristics, has long been considered an important part of clinical trial design as well as analysis. When done retrospectively, after results are known, or when planned for examination but not in the set of pre-specified primary or secondary study hypotheses, the approach has been considered by most clinical trialists, as well as by FDA, as an exploratory exercise whose results should be descriptively interpreted.

A biomarker is defined as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Biomarkers Definitions Working Group, 2001). Current technologies now allow for a subject level genomic biomarker to be used to classify a subject into a distinct subpopulation and this strategy is now being incorporated into clinical trial design for drug development and drug registration. Unlike a surrogate biomarker which is an intermediate outcome, a genomic biomarker classifier in a randomized controlled trial is used to classify patients into refined subsets prior to and independently of treatment administration, and is often expected to predict treatment response.

Current practice in some randomized controlled trials is to determine a patient's genomic biomarker status based on his/her baseline genomic materials, e.g., blood, tissues, which generally should be collected prior to any treatment intervention. The genomic material is assayed, often but not necessarily with some diagnostic test, for biomarker status. For a binary classifier, the result of the genomic diagnostic assay test is either considered positive (present) or negative (absent). Whether or not an FDA cleared diagnostic assay is used to classify the patients' genomic biomarker status, there is often interest in demonstrating a favorable benefit/risk in genomic marker (positive or negative) patient subsets or in minimizing a harmful/risk outcome by excluding treatment in a selected genomic patient subset

There are many issues to be addressed with this strategy, a substantial number of which are matters of statistical design, analysis and interpretation.

1. The chance of erroneously concluding that there is a real treatment effect when in fact it is not true, or the chance of concluding there is no treatment effect when in fact one actually exists, are two critical concerns for the design and interpretation of study results of any clinical trial. There are many examples in the medical literature and in FDA files of subpopulation findings that are spurious and false. To address this problem at the study design stage, it is necessary to control the

chances of making these false conclusions, usually by pre-specifying the hypotheses and the number of subgroups for which a treatment effect in the subpopulation is sought as a primary objective of the trial.

2. The likelihood that the statistical evidence for reproducing a treatment effect identified in a subpopulation in a single clinical trial can be demonstrated in another independent study. It is surprising how many treatment effects identified in one study cannot be demonstrated in a second independent study, even for the overall population studied, not to say for a subpopulation group itself.

3. The required sample size for the biomarker negative subpopulation in order to have sufficient statistical chances of detecting a treatment effect, if it exists, considering that the effect may not be of the same magnitude as in the biomarker positive subpopulation

4. The minimum performance characteristics (e.g., sensitivity, specificity, reproducibility) of the assay used to define patient subgroups and the consequences of that performance for correct decision making and inferences from the study.

5. Clarity and consensus on the definitions of biomarker classifier prognostic and prediction properties and the consequences of these definitions on study design planning, sample size, and ability to draw valid inferential conclusions.

6. The role of randomization to assure comparable comparisons, especially in small sample size subpopulation identified after completion of a clinical study.

7. The choice of alternative clinical trials designs, and their efficiency and feasibility, to demonstrate prospectively that there are biomarker subpopulation differential treatment effects.

In a prospectively planned randomized controlled trial, the biomarker status can be readily available prior to trial initiation and to subject randomization, thus, it can be used as a stratification factor for randomization. However, as in the case study discussed above, in some randomized controlled trials randomization did not incorporate baseline biomarker status, and in fact, the biomarker status was not determined until after the trial began, but before unblinding the clinical trial data, based on information from external trial sources. .

The clinical utility of a genomic biomarker (g+ vs. g-) in controlled trials can be judged in terms of its properties as prognostic, predictive, or prognostic-predictive of a treatment effect measured by the primary efficacy outcome. To illustrate these concepts, the efficacy outcome of tumor response rate is used in the following example. Table 1 presents three general scenarios, where the “control” group can be an active control or placebo.

Table 1. True response rates by genomic biomarker status and by treatment intervention

Genomic Status*	Scenario A		Scenario B		Scenario C	
	Control	Drug A	Control	Drug B	Control	Drug C
g–	10%	10%	10%	10%	10%	20%
g+	10%	25%	20%	20%	20%	40%

* *g*+ or *g*– is patient's genomic biomarker status determined from a diagnostic assay

When a treatment effect exists only in the g+ patient subset and no effect in the g- patient subset, such as Scenario A, the genomic biomarker is said to be predictive of treatment effect. As shown in Table 1, there is no impact of response to control treatment whether a patient is classified as g+ or g-. However, it is also possible to have responses that are different between g+ and g- patient subsets in the control arm, that are separate and independent from the relationship to the treatment group. The primary focus of clinical trial is on the comparative effect that is the difference in the patient response on treatment relative to the patient response on the control, so that Scenario A describes a superior treatment effect relative to its comparator in the g+ patient subset and no treatment effect in the g- patient subset.

When there is no treatment effect in any patient (sub)set, the genomic biomarker may only be prognostic of the underlying disease mechanism, or be prognostic for a higher or lower event rate as shown in Scenario B. The practice in many large outcome trials is to recruit patient with prognostic risk factors as there is then the expectation that more events will be observed, thus, increasing the statistical power of the clinical trial.

The most interesting situation is when the genomic biomarker can be both prognostic of disease response and prognostic of a therapeutic effect, as shown in Scenario C, resulting in a treatment effect that is present in both biomarker subgroups but is also quantitatively different in each genomic biomarker subset. In this case, the genomic biomarker is predictive of differential treatment effect. We note that there is some confusion with this terminology as the commonly known term prognostic (of therapeutic effect) is often used for Scenario C, a scenario often observed in clinical trials. For Scenario C, there is a need for more sophisticated and relevant statistical analysis methods to adjust for the genomic biomarker status to improve the power of detecting an overall treatment effect in the study and to formally test for differential treatment effect (qualitative, quantitative statistical interactions). Table 1 data can easily be translated into a time to event endpoint data, such as, overall survival or progression free survival.

The conventional clinical trial design and analysis addressing a single study objective that new treatment is superior to its comparator on the prespecified primary efficacy endpoint is unlikely to be able to formally detect a treatment effect limited to the *g*+ patient subset. When the biomarker classifier is prospectively used to define the clinical objective and statistical hypotheses, for instance, an all-favorable treatment or a favorable treatment benefit/risk profile limited to a biomarker-classifier defined patient subset, statistical literatures on the study designs (fixed or adaptive) and analysis methods that account for multiplicity adjustments of multiple patient hypotheses are available.

The term "prospective/retrospective" study design is used here to characterize a current practice in clinical trials. In this situation, a patient's biomarker status is determined using genomic samples collected and banked at the baseline pre-treatment stage of a completed clinical trial that is then assayed at some other time during or after the trial has been completed. Since the collection of the genomic materials requires a genomic consent form, which is different from the usual clinical trial consent, the genomic samples for each randomized subject are often collected only on those subjects who consent, and could be considered a convenience sample. This could be problematic if the reasons are different between treatment groups for the consent and if the patient risk factors for the study outcome are differentially distributed between the patients consenting in each treatment group. Because of the quality of the banked genomic samples or of the analytical performance of the diagnostic assay (e.g., unvalidated, unregulated diagnostic test), some of the genomic samples cannot be used to clearly determine the biomarker status. The combination of the amount of missing biomarker status on all randomized subjects, if unduly influenced by poor quality of genomic sample ascertainment methods or genomic diagnostic assay validation issues, can raise further concerns of the genomic study quality overall and the interpretation of the study results in the retrospective pursuit.

A trial that is designed prospectively to collect biomarker data and to test a *de novo* hypothesis that a subject's genotype is indicative of a particular treatment response, based upon a hypothesis generated from clinical data from previously completed drug trials or cohorts is called a 'prospective' study. Here, the genotyping laboratory is blinded to the clinical data, eliminating any bias in the interpretation of the study results after they are unknown. This would then be a prospective genetic analysis study carried out using unblinded available clinical data.

There are several alternative clinical trial designs that can be considered, ranging from incorporation of all subjects regardless of marker status, to exclusion of subjects with a particular marker status, to adaptive study designs, that may modify the second stage entrance criteria of a trial based upon the first stage results of the trial.

CONCLUDING COMMENTS

FDA is seeking ODAC's deliberations on all of these issues raised above. In particular, the committee should discuss the conditions, if any, where a prospective/retrospective clinical study design may provide evidence for treatment effects that are limited (or restricted) to biomarker classified subpopulation, thereby being judged as evidence of a predictive biomarker. In addition, if a retrospective analysis can be performed to show benefit in a genomic subset and it is considered acceptable that randomization on biomarker status was not done, what level of evidence should be considered for reproducibility of the finding? That is, can one single large prospective/retrospective trial serve as the basis for label consideration or is an independent prospectively randomized controlled trial needed to replicate the finding?

REFERENCES:

1. DeMets, DL.Statistical Issues in interpreting clinical trials. J Int Med 2004: 255: 529-537.

2. Huang Y, Pepe MS, Feng Z. Evaluating the predictiveness of a continuous marker. Biometrics 2007: 63: 1181-1188.

3. Li, Z., Chuang-Stein, C., Hoseyni, C. The probability of observing negative subgroup results when the treatment effect is positive and homogeneous across all subgroups. Drug Information Journal, 2007:41:47-56.

4. Mandrekar SJ, Grothey A, Goets MP, Sargent DJ. Clinical Trial Designs for prospective validation of biomarkers. Am J Pharmacogenomics. 2005:5(5): 317-25.

5. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thomson IM, Zheng Y. Integrating the predictiveness of a marker with its performance as a classifier. Amercican Journal of Epidemiology 2008 167:362-368.

6. Sargent DJ, Conley BA, Allegra C, Collette, L, Clinical Trial Designs for predictive marker validation in cancer treatment trials. JCO 2005 Mar 20:23(9):2020-7.

7. Simon R, Maitourman, A. Evaluating the efficiency of targeted design for randomized clinical trials. Clin Cancer Res, 2005:10(20):6759-63.

8. Simon R, Wang SJ. Use of genomic signatures in therapeutic development. The Pharmacogenomics Journal 2006:6:1667-73.

9. Wang SJ, O'Neill, RT, Hung HMJ, Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharmaceutical Statistics, 2007:6:227-44.

10. Pennello GA and Vishnuvajjala L. Statistical design and analysis issues with pharmacogenomic drug-diagnostic co-development. Statistical Design and Analysis Issues with Pharmacogenomic Drug-Diagnostic Co-Development, In American Stat. Assoc. 2005 Proc. of the Biopharm. Section, Joint Statistical Meetings, Minneapolis, MN, August, 2005.

11. Hayes DF, Bast RC, Desch CE, et al Tumor marker utility grading system: A framework to evaluate clinical utility of tumor markers. JNCI, 1996. 88:1456-66.

12. Bossuyt PM, Reltsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003. 138: W1-W12.

13. McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK). JNCI 2005. 97:1180-1184.

14. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. JNCI 2007. 99:1036-43.

15. Wolff AC, Hammond EH, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch Pathol Lab Med. 2007: 131: 18-43.