From Personalized Medicine

US FDA and Personalized Medicine: *In vitro* Diagnostic Regulatory Perspective

Živana Težak; Marina V Kondratovich; Elizabeth Mansfield

**Authors and Disclosures**


**Abstract and Introduction**

**Abstract**

Personalized medicine has captured the attention of the public, including patients, healthcare providers, scientists, medical product manufacturers and many others. The US FDA will evaluate many of the products that will allow personalized medicine to be successfully implemented in the USA. This article addresses the FDA's approach to regulation of one component of personalized medicine, *in vitro* diagnostic devices. It also describes the FDA's efforts to integrate the various medical product regulatory authorities provided by Congress in the Federal Food, Drug and Cosmetic Act to develop effective mechanisms for oversight of medical products used to personalize treatment. Finally, it presents some of the current challenges in *in vitro* diagnostics oversight that may be of interest for personalized medicine.

**Diagnostics in Personalized Medicine**

The definition and scope of the term personalized medicine varies widely, depending on the stakeholder providing the definition, and ranges from the extremely broad to the very narrow. Personalized medicine is perhaps most often referred to as providing the right intervention or therapy, at the right dose, for the right person, at the right time, by understanding the individual's biology. Alternatively, the US National Cancer Institute, defines personalized medicine 'as a form of medicine that uses information about a person's genes, proteins and environment to prevent, diagnose and treat disease'. Here, we will define personalized medicine as a model for taking into account a patient's particular genetic, genomic or proteomic constitution, together with environmental and other factors, to deliver treatments that are as safe and as effective for that patient as possible. This definition accounts for both constitutional variations and environmental exposures that are particular to an individual, as well as variations and exposures that are shared within a group of individuals. It also encompasses various uses of the model, including predicting therapeutic response, nonresponse, likelihood of adverse reactions to therapy and implementation or modification of a therapeutic approach to alter a patient's predisposition to disease.

The success of personalized medicine depends on safe and effective diagnostics. If the diagnostic performance is poor, variable or not properly validated in the specific therapeutic context of use, the performance of the therapeutic that depends on a diagnostic result likely will be unpredictable. For
example, with an incorrect diagnostic result, an unsuitable drug may be given to a patient who will as a result, be harmed or will not benefit, because the drug will cause an otherwise avoidable adverse event, will be ineffective for that patient, or both. Therefore, proper design and validation of the diagnostic is essential in order to gain the best risk–benefit profile for the accompanying therapeutic.

Those diagnostics that are necessary to successfully select patients for therapy, distinguish likely responders from nonresponders, identify patients at high risk for adverse events, or select an appropriate dose for safe and efficacious use of the therapy, are referred to here collectively as 'companion diagnostics'. They may be distinguished from other types of diagnostics by their intended uses and by the importance placed upon them in therapeutic product labeling. See the section entitled 'Companion diagnostics' for a more detailed description of this product type.

Regulatory Challenges in Personalized Medicine

The US FDA's oversight activities for personalized medicine products primarily include the three medical product review centers: the Center for Devices and Radiological Health (CDRH), the Center for Drug Evaluation and Research (CDER), and the Center for Biologics Evaluation and Research (CBER). Each of these centers applies specific sets of regulations translating different statutory authorities, most of which have been in place for many years. Thus, not surprisingly, existing regulations for drugs, biologics and medical devices do not address the current situations in personalized medicine in which different types of medical products are dependent upon one another to achieve safety and effectiveness. This has led to some inconsistencies in regulating products used in personalized medicine. The FDA is now identifying these areas, and establishing regulatory processes and implementing policies that will clearly delineate the activities and responsibilities of the different centers in the oversight of personalized medicine products within each center's existing regulatory framework. These activities are intended to help coordinate premarket reviews for the different products (therapeutic and diagnostic) to provide consistency and timeliness in regulatory decision-making for these products.

Diagnostic devices include both in vitro tests such as assays used in measurement of genetic factors and in vivo tests, such as electroencephalography, electrocardiography or diagnostic imaging equipment. For this article, we will focus only on in vitro diagnostic (IVD) tests that are regulated by the FDA's CDRH, Office of IVD Device Evaluation and Safety (OIVD; MD, USA), although at least some of the principles described can be applied to other types of diagnostics. The OIVD has three IVD divisions, each addressing a different diagnostic focus: Immunology and Hematology; Chemistry and Toxicology; and Microbiology. In addition to these, the OIVD has recently added a cross-cutting personalized medicine staff as an integral part of the office. This group is charged with addressing the important and unique issues for diagnostics used in personalized medicine, including policy and process related issues. The personalized medicine staff is developing regulatory approaches for co-development of a diagnostic with a therapeutic, recommendations for FDA internal intercenter coordination for more efficient co-reviews, and is helping to guide the balance of regulation between centers for oversight of personalized medicine products.
What is a Diagnostic Device?

A diagnostic device is a type of medical device. Medical devices are defined in statute as 'an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including any component, part, or accessory, which is … intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals … and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes.'[1] By regulation, the FDA has further defined, IVDs as a specific subset of medical devices that include ‘…those reagents, instruments and systems intended for use in the diagnosis of disease or other conditions, including a determination of the state of health, in order to cure, mitigate, treat or prevent disease or its sequelae’. [2] Many companion diagnostics fit into this category.

As can be observed from the regulatory definition of IVDs, the regulatory term 'diagnostic' in context of an IVD (and the colloquial short-hand phrases 'diagnostic', or 'diagnostic test', are often used in place of IVD tests) makes use of the word 'diagnostic' in a broad sense, covering any use that can be construed as a medical use. Indeed, IVDs are products that can be used not only for diagnosis, but also for screening, health classification, risk assessment, monitoring, staging, prediction, prognosis and more.

IVD Regulation: Management of Risk

The FDA regulates IVDs through pathways corresponding to the classification of the device according to its stated ‘intended use’. [3] The intended use for a medical device, including an IVD device, is established according to the claims the manufacturer or sponsor intends to make for the device, and includes the target population and the clinical setting for the use of an IVD. The classification of an IVD into one of three classes (class I, II or III,[4] in ascending order of risk) determines the premarket review pathway,[5] as well as the types of regulatory controls to which the IVD will be subject.

For most IVDs, the intended use and indications for use are often folded together under the umbrella of ‘intended use’, and should generally include the following elements:

- What is measured, identified or detected (names of one or more specific analytes, such as specific gene, protein, polymorphisms or signatures);
- The measurement principle of the test (including whether the test is qualitative, semiquantitative or quantitative);
- The specimen type, source (e.g., whole blood, serum, CSF, nasal aspirates and so on);
- The setting in which the device is meant to be used (clinical laboratory, point-of-care and so on);
- A specification of the instrumentation required in order to perform the test;
- The target condition (a particular disease, disease stage, health status, or any other identifiable condition or event, or a health condition that should prompt clinical action such as the initiation, modification or termination of treatment);
- The clinical purpose of the measurement (such as aid in diagnosis, prognosis and monitoring of disease);
- The target population for whom the test is intended to be used (such as individuals with particular genotypes or phenotypes).
Classification through risk determination for an IVD is typically based on the harm to a patient that might be incurred based on an incorrect test result when the test is used as intended, although it can include other types of risks. For example, if a consequence of a false-positive test result is an invasive medical procedure or a therapy with toxic side effects, this type of test would generally be considered high risk, since a false-positive test result will likely lead to unnecessary harm for the patient. Similarly, a false-negative test result might alter medical management and the appropriate intervention for the patient may be unnecessarily delayed, or not pursued at all.

The FDA expects that many companion diagnostics will be classified into the highest risk class (class III), owing to the likelihood of harm to the patient if the diagnostic result is incorrect. Class III devices require the more rigorous premarket approval application (PMA), pre-approval inspection of manufacturing facilities, yearly reports summarizing minor product changes, and other controls to assure that the device continues to perform in the same way over time. One can easily see why this level of control would be critical for most companion diagnostics – changes in diagnostic performance owing to product changes or inconsistent manufacture could be disastrous when applied to use of a therapeutic that depends on diagnostic results in order to be used properly.

Another area of risk that must be addressed, but is not directly related to the classification of the diagnostic, is the risk to patients when a diagnostic is used in an investigational manner. For all types of diagnostics, performance data is generally gathered under an investigational plan, subject to CDRH's investigational device regulations. These regulations direct the CDRH to apply oversight to the investigational process at a level that is commensurate with the risk to patients and users of the device as it is to be used in the investigation. Risk in investigational settings for devices is parsed as 'significant' and 'nonsignificant'. Many IVDs do not present serious risks to patients or users in the investigational phase of the device (and in fact, are exempt from most of the investigational requirements), primarily because the patient under study is not affected by the diagnostic result. However, if results from an IVD under investigation are returned for patient selection or care, such as in certain therapeutic clinical trials, the investigation is often considered to present significant risk, owing to the possibility that the patient could be harmed by consequences of the IVD result (assignment to a specific therapeutic trial arm and so on). In these cases, or if the device under investigation requires an invasive sampling procedure such as a biopsy, approval of a regulatory submission, termed an 'investigational device exemption', may be required prior to using the IVD in the investigational setting. Where the investigational setting is a therapeutic clinical trial, investigational device requirements can be addressed either as a part of the therapeutic investigational new drug submitted for review by CDER or CBER and/or as part of an investigational device exemption submitted for review by CBER or CDRH. In either case, all review centers will interact to assure that both the therapeutic and the IVD are used under the appropriate investigational controls in the trial.

Finally, a premarket submission for approval of the IVD should include, in most cases, all the required information regarding performance of the IVD in the context of therapeutic use. The clinical performance for a companion diagnostic will generally be provided by data from the therapeutic trial(s) indicating that the IVD properly identifies patients for specific treatment choices, whether to select, avoid or determine appropriate dose for the therapy.

According to the device regulations, two types of evidence must be evaluated by the FDA for all devices. Safety assessment includes evaluation of reasonable assurances, based on valid scientific evidence, that the probable benefits to health from use of the device outweigh any probable risks.
Effectiveness assessment evaluates whether there is reasonable assurance, based on valid scientific evidence, that the use of the device in the target population will provide clinically significant results. Safety and effectiveness for companion diagnostics will generally be determined based on satisfactory analytical (measurement) performance, clinical performance in the context of use, that is, a demonstration that the diagnostic measurement correlates significantly with the specified therapeutic action and outcome, and other factors such as ability to repeatedly manufacture the device to specifications, appropriate ethical conduct within the trials, and labeling that is compliant with the labeling regulations for IVDs.

When an IVD is being considered for use in selecting patients to receive or avoid a particular therapy, or to select a safe and efficacious dose, the same performance measures should be addressed. Analytical performance demonstrates the ability of the IVD to accurately and reproducibly select patients whose samples contain (or lack) the analyte(s) of interest, either in an absolute sense (e.g., present/absent or above/below a threshold value), a semiquantitative sense (e.g., low/medium/high) or a quantitative sense (e.g., amount of analyte in measurement units as related to certain diagnostic outcome). Clinical performance demonstrates the ability of the IVD to select patients in a way that will improve the risk–benefit balance of the drug. This is often dependent on the selection of the appropriate cutoff value that will differentiate patients into the desired outcome classifications, for example, responder's versus nonresponders.

The HER-2/Herceptin therapeutic–diagnostic pair provides a good example of the difference between analytical and clinical performance, and the importance of clinical outcome in evaluating the IVD. HER-2 testing is not used for the purpose of detecting the presence of HER-2 per se (analytical performance) but to identify patients with greater likelihood to respond to treatment (have a good outcome) with Herceptin (clinical performance). If analytical performance is inadequate to measure HER-2 relative amounts, clinical performance will be compromised in predicting ability to respond to the drug. In the context of a therapeutic trial, any instability of the IVD, or selection of a noninformative marker, has clear potential to jeopardize therapeutic approval. In the context of real-world use, instability of the IVD will put patients at risk.

Important Issues in IVD Selection & Performance

A definitive clinical study for a joint therapeutic–diagnostic would be one that allows for assessment of the therapeutic's safety and efficacy, as well as for validation of the clinical validity (and usually utility) of the marker in guiding the therapeutic's use. Ideally, an appropriate IVD for use in therapeutic selection will be identified at least by the time the pivotal therapeutic clinical study is performed, and will be completely analytically validated with preliminary cutoff points defined. This IVD would be used during the therapeutic clinical trial(s), resulting in the assessment of the clinical performance of the IVD together with that of the therapeutic. For different approaches for selection of the cutoff, see. The efficiency of the combined validation is obvious. However, it is critical for this strategy that there are no changes made to an IVD during the clinical trial process. Any changes with the potential to result in modified analytical performance may compromise the use of the IVD data in the trial. This would imperil not only the IVD approval, but also that of the therapeutic.
Development & Validation of Diagnostic Marker

To ensure objective evaluation, the initial development of a novel IVD should use a clinical dataset that is completely separate from the prospective or retrospective set of samples on which it is to be clinically validated. This is of particular importance when attempting to retrospectively establish a diagnostic marker to guide a therapeutic use. Retrospective characterization of IVD performance on the clinical validation dataset can be particularly misleading. For example, a diagnostic marker developed and evaluated using a single clinical validation dataset for both development and validation can often identify a subgroup that appears to be associated with drug response or drug toxicity, but is actually an artifact due to chance or sample bias within the selected study population. Identification of spurious associations is more likely when the number of markers being studied is large since the chance of a false association increases with the number of comparisons performed. While prospective studies are ideal for addressing the problem of false associations, alternative techniques using robust retrospective validation may be considered, particularly in cases where the marker-treatment mechanism is well understood, the strength of association is high, and replicate testing of an independent collection of samples is possible.

Sample Size

Depending on the primary end point(s) selected, the sample size needed in drug-diagnostic studies must consider factors such as the marker prevalence in specific subpopulations and the magnitude of expected drug effects in subsets defined by marker status or other stratification factors. Sample size should adequately address the diagnostic effect if the therapeutic response (efficacy or adverse events) is used to establish the performance of a diagnostic marker. The sample size needed in therapeutic–diagnostic studies may also need to account for additional factors, based on preliminary feasibility or early therapeutic studies. In practice, the number of patients and patient samples required to characterize IVD clinical performance will vary depending on the nature of the claim and particular circumstances, for example in rare diseases, and so on, for the therapeutic, IVD and population being studied.

Changes to an IVD

In some cases, the diagnostic modality utilized during therapeutic trials may not be ideal or practical for use in the real-world clinical setting, or additional IVDs are developed with the same intended use. If changes to an IVD are contemplated, either in late stages of therapeutic development or upon marketing of the approved therapeutic, the 'new' IVD (measuring the same marker) must either have very similar analytical performance characteristics to the 'old' one, or must, by testing banked specimens from the pivotal therapeutic trials, provide assurance that the diagnostically predicted clinical outcome remains the same with the 'new' IVD. However, in order to provide convincing assurance of bridged clinical performance of the diagnostic, evaluation of banked sample stability (especially with regard to the analyte needed to determine marker status, but also including stability or introduction upon storage of possible interfering and crossreacting substances) is critical. Careful consideration of possible bias in sample availability cannot be overemphasized when retrospective bridging of IVD is undertaken. Bias at this juncture has proven to be a serious problem, and often very difficult, if not nearly impossible, to overcome. Therefore, it is worth considering designing a sample collection and storage element into every therapeutic-diagnostic trial as a fail-safe, even when the need for such samples is not
predicted *a priori*. It is also generally a good idea to consult with the FDA review division on the validation protocol for the new or modified assay prior to beginning bridging the studies.

**Uses of IVDs in the Context of Therapeutic Selection & Management**

As discussed briefly above, the use of IVD assays is not limited to diagnostic claims. In general, IVDs can be used for the evaluation of the current state (e.g., diagnosis or residual disease detection), change in state (e.g., monitoring or recurrence), or a future state (e.g., risk of disease, prognosis or prediction). Besides diagnosis, IVD assays include those that are used for:

- Risk assessment, leading to preventive interventions for those at sufficient risk;
- Early detection, enabling intervention at an earlier and potentially more curable stage than under usual clinical conditions;
- Prognosis, indicating disease aggressiveness and patients with poorer or better prognosis;\(^{[108]}\)
- Prediction of response to a therapy, providing guidance in choice of therapy;\(^{[105-107]}\)
- Monitoring of disease response during therapy,\(^{[109]}\) with potential for adjusting the level of intervention (e.g., dose) on a dynamic and personal basis;
- Early detection of recurrence.

**Prognostic & Predictive Markers**

Although all of the above-mentioned uses of IVDs in the context of therapeutic use are important, we will discuss prognostic and predictive markers here in more detail, owing to their relatively large contribution to date in the FDA's therapeutic-related IVD reviews, as well as the critical importance of understanding the distinction between these two types of markers and what information they provide (Figure 1).
Figure 1.
**Predictive versus prognostic biomarkers.**
Marker-positive population is marked in red, and marker-negative population is marked in blue. The figures only illustrate a few simple ways in which biomarker–therapy–outcome interactions might occur. Other factors (such as risk:benefit ratio, safety concerns, availability of other treatment and so on) that may affect assessment of the biomarker and therapeutic effect are not taken into account. (A) No biomarker effect tested. The effect of T versus S is assessed. T shows improved outcome (green arrow) compared with the S in all comers. (B) Prognostic biomarker. Only S is used to assess the effect of biomarker; the effect of therapy is not assessed. When the same type of care is used (regardless of whether there is treatment or no treatment), marker-positive population (dashed red line) shows better outcome than the marker-negative population (dashed blue line). Biomarker shows prognostic effect (yellow arrow). (C) Prognostic biomarker. The effect of S versus T is assessed in both biomarker-positive (red) and biomarker-negative population (blue). Similar therapy versus standard-of-care effect size is observed (green arrows), regardless of biomarker status. For the purposes of the point described, the therapeutic effect is the same, for example, in an ‘absolute’ survival sense (the green arrows are the same length). Biomarker-positive population has better outcome than biomarker-negative population (yellow arrows) regardless of whether the S or T.
is used. The biomarker shows prognostic effect, and there is no predictive biomarker effect (i.e., treatment effect is independent of marker status). (D) Predictive biomarker. The effect of S versus T is assessed, in both biomarker-positive (red) and biomarker-negative population (blue). T does not appear to improve patient outcomes over S in the marker-negative population (circled green arrow between blue lines T and S). T shows large improvement in patient outcomes when compared with S in marker-positive population (green arrow between T and S red lines). Biomarker shows predictive effect. (E) No biomarker effect. The effect of S versus T is assessed, in both biomarker-positive (red) and biomarker-negative population (blue). Similar therapy versus standard-of-care effect size is observed (green arrow), regardless of biomarker status, and T shows improved patient outcomes when compared with S. There appears to be no biomarker effect on patient outcomes in either S or T arm (marked by yellow circles). There is no predictive or prognostic biomarker effect.

Figures are simplified illustrations of the relevant points, and not depictions of biological data.

S: Standard of care; T: New therapy.

Prognostic markers are useful to assess the risk of disease recurrence, by comparing the outcome for marker-positive and marker-negative patients, regardless of the treatment (Figure 1B), where intervention (e.g., drug therapy) is not a variable. A prognostic marker can be defined as either a single trait or signature of traits that separates different populations with respect to the risk of an outcome of interest in absence of treatment, or despite nontargeted 'standard' treatment. Conversely, predictive markers compare intervention effect (i.e., treatment vs control) for marker-positive versus marker-negative patients, and predict differential effect of treatment on the outcome (Figure 1D). Predictive markers can be defined as a single trait or signature of traits that separate different populations with respect to the outcome of interest in response to a particular targeted treatment. Figure 1 illustrates the effects of predictive and prognostic markers on therapeutic trial outcomes. In Figure 1D there is predictive biomarker effect; in Figure 1C there is no predictive biomarker effect, but there is prognostic effect; and in Figure 1E there is neither a predictive nor a prognostic biomarker effect. A predictive marker implies relative sensitivity or resistance to specific treatments (or adverse events), and can ultimately be useful for selecting or avoiding specific therapy.

If a predictive marker is used to guide treatment decisions, some of the issues to be considered include the following:

- Differences between assays and test systems may yield inconsistent or nongeneralizable results across different platforms. Therefore, even if the marker has clear clinical significance, the specific test used for measuring the biomarker may make a (sometimes large) difference in test results for the same patient;
- Although some markers may have well-accepted or plausible mechanistic significance, the clinical significance of the marker will need to be demonstrated in a prospective manner as applied to a specific therapeutic;
- The findings from a specific study often cannot be extrapolated to different study populations (e.g., various human subpopulations with different genetic backgrounds).

**Design Considerations for Predictive Biomarker Tests**

One objective of joint therapeutic–diagnostic studies is to ensure that the results of an IVD used in the target population have been properly linked to the expected therapeutic response, in order to assess safety and/or efficacy of treatment. Useful applications of predictive markers include identifying patients who are good candidates for a therapy, and therefore, will show favorable efficacy (i.e., responders), identifying patients who are likely to develop adverse outcomes with a therapy (i.e.,
experience toxicity) and are not good candidates for a treatment, and so on. In studies using predictive diagnostics, the analytical and clinical end points for diagnostics should be carefully chosen to reflect useful outcome measures, and selection of appropriate intended use populations and patients should be considered, based on marker prevalence, disease stage, prior therapeutic history and so on. The results obtained from well-controlled therapeutic–diagnostic trials will provide information on the predictive results of the IVD as they relate to therapeutic response (safety and/or efficacy), as well as the differential therapeutic effect in marker positive and negative patients.

Numerous designs can be considered in a joint therapeutic–diagnostic study. In an ideal study, there would be at least some descriptive data to allow evaluation of the therapeutic action independent of marker, the marker action independent of therapeutic, and the intersection between the two. This type of study would allow assessment of the fraction of therapeutic responders who are marker positive (i.e., sensitivity of the IVD), fraction of therapeutic nonresponders who are marker negative (i.e., specificity of the IVD), fraction of marker positives who respond to the therapy (positive predictive value of the IVD), and fraction of marker negatives who do not respond to the therapeutic (negative predictive value of the IVD). This can be achieved in several ways, including:

- Identify patients by marker status; randomize therapy across all patients;
- Identify patients by marker status; randomize therapy in subsets defined by marker status;
- Randomize by treatment; look back at markers.

The second design results in assessment of the effect of an IVD within a therapeutic trial by employing randomization by results (i.e., marker-positive and marker-negative arm), where each arm would be randomized to either new therapy or standard of care (or a treatment and placebo) (Figure 2B). This results in stratification by the results of a marker result. Alternatively, the third bullet point lists the usual simple two-arm randomization (Figure 2A); comparing a treatment and a control may be employed, with the results from the diagnostic marker that is being investigated used as a stratification factor in the statistical analysis. One reason such a design may be adopted would be if the results of the IVD are not expected to be readily available at the clinical sites for informing randomization.
Clinical trial design considerations when using biomarker tests.

(A) Nontargeted design. Standard randomized clinical trial design. Randomization is not stratified by biomarker result. This type of trial design evaluates the effect of therapeutic in the entire treatment-eligible population. All subjects are randomized to either a new therapy or a standard-of-care treatment, regardless of biomarker status. Therefore, overall therapeutic effect is a mixture of marker-positive and marker-negative therapeutic effects. This type of initial trial design is often seen when studies on banked/retrospective samples are performed to post hoc evaluate marker effect.

(B) ‘Marker-by-treatment’ interaction design. Test everybody for biomarker status. Randomize biomarker-positive and biomarker-negative populations separately to either new treatment or a standard-of-care. Information will be obtained on whether the therapeutic works in the marker-positive subset, the marker-negative subset of patients or both. However, accrual rate may not be equal in marker-positive and marker-negative populations. Sensitivity, specificity, positive predictive value and negative predictive value of the biomarker-based test used in the study can be calculated.

(C) Targeted design. Test everybody for biomarker status. If the therapeutic is thought to work in the marker-positive subset of patients only, then enroll only these patients. Therapeutic effect in marker-negative population is not assessed. This type of design cannot establish whether the marker is informative.

However, in many cases and for many different reasons, studies are designed where only one of the subsets defined by the IVD is studied in the clinical trial (Figure 2C). For example, for ethical reasons, a marker selected patient population expected to have an adverse reaction or to not respond to the
therapeutic may be excluded from the study. However, the results of such a study would yield only a positive predictive value for the IVD (that is, the fraction of IVD marker positives who respond to the therapeutic), without providing any information on whether the therapeutic may also work in a marker-negative population. In addition, no information would be available regarding sensitivity (the fraction of therapeutic responders who are marker positive), or specificity of the IVD (the fraction of therapeutic nonresponders who are marker negative). Nevertheless, this is the type of study conducted in many therapeutic trials.

It is important to note that study descriptions listed above are simplistic and lack critical details. There is, however, an evolving body of literature on clinical study designs that might be appropriate to adequately evaluate performance of both IVD and therapeutic for use in personalized medicine.[15–19]

Some considerations when designing clinical programs in which a marker-defined subset of patients will be studied include the clinical performance of the IVD (i.e., the strength of the association between the IVD result and a particular treatment response, whether beneficial, neutral or toxic), and whether there is an expectation that the therapeutic will only be used in an enriched population in practice (i.e., that patients will be universally tested and the therapeutic used only in patients with a certain marker status). The amount and extent of clinical trials data needed to establish the clinical validity (and usually utility) of an IVD will differ, depending on the prior knowledge of the pathophysiology involved, the level of mechanistic understanding of the therapy and its effect in relation to the IVD, and the amount of previously collected relevant clinical data. If previous clinical performance data is unclear on whether there is reasonable support for the hypothesis that an IVD would predict the enhancement of efficacy or safety when compared with the unselected population, it would likely be more useful to define the therapeutic response in both patients with known marker status (both marker positive and marker negative) and in unselected patients. In cases where there is significant uncertainty about the marker-response correlation, banking samples for the purposes of retrospective exploration and testing should be strongly considered. See previous section on 'Changes to an IVD' for important issues about banked samples and bias.

**Future Perspective: Evolving Regulatory Challenges**

Although the FDA's authority to regulate all tests that are within the definition of a medical device includes 'laboratory developed tests' (LDTs) developed and used in the same laboratory, the FDA has historically exercised enforcement discretion towards most LDTs. However, in recent years there has been considerable widening of the types of tests offered as LDTs, resulting in calls for FDA to re-evaluate its informal policy of enforcement discretion towards these tests. Currently, some companion diagnostic tests for personalized treatment are also offered as LDTs. In this section, we provide background, and list evolving regulatory challenges related to LDTs and companion diagnostics, as well as describe some steps the FDA has taken in response to growing concerns about certain types of LDTs.

**Laboratory Developed Tests**

The expansion of molecular testing was largely aided by completion of the draft human genome, which was greatly facilitated by development of new, rapid and accurate molecular technologies. These technologies range from the simplest nucleic acid amplification methods, to various complex sequencing
technologies, microarrays, multiplex tests, bioinformatics and so on. To date, the vast majority of molecular-based diagnostic tests are being offered as LDTs without direct FDA oversight. The decision to generally apply enforcement discretion and not enforce regulations that apply to LDTs may be traced to the manner in which these tests were offered at the time the US Congress passed the Medical Device Amendments Act and the device regulations were implemented. At that time, most LDTs were simple, developed for use by specific laboratories for their local populations, with ordering and test-result reporting performed with direct interaction between physicians and laboratory personnel, often within the institution providing clinical care to patient being tested.

As molecular technology evolved, many products previously only used in scientific research made the transition into the clinical setting, often without formal control of manufacturing. Concerned over use of research-grade materials, especially for molecular diagnostics, the FDA developed regulatory requirements for commercially distributed reagents that are the active ingredients of clinical diagnostic tests, called analyte specific reagents (ASRs). The result was the set of regulations addressing ASRs and commonly termed the ‘ASR rule’. This set of three inter-related regulations defined ASRs; classified ASRs; applied FDA's general controls, including manufacturing under parts of the FDA's Quality System Regulation (‘good manufacturing practice’) but exempted most ASRs from premarket notification or approval; specified labeling and distribution limitations; and called for explicit notation of laboratory results generated using ASRs, warning that the test itself had not been cleared or approved by the FDA. \[20\]

Most molecular tests offered today are LDTs that make use of ASRs and general purpose reagents.

In recent years, the US healthcare system has witnessed considerable expansion of the types of test offered as LDTs alongside profound changes in laboratory business models. LDT offerings have broadened since 1976 from local laboratories performing relatively simple, well-characterized tests in close interaction with the ordering provider, to today's laboratory networks and systems that offer novel, complex and high-risk tests to healthcare providers across the country. Starting from 1997, several prominent groups have questioned the appropriateness of the FDA's policy of enforcement discretion toward LDTs, including the NIH's Department of Energy, Joint Task Force on Genetic Testing, \[110\] the Secretary's Advisory Committee on Genetic Testing, and the Secretary's Advisory Committee on Genetics, Health and Society. \[111\] Each of these groups published recommendations suggesting that the FDA should revisit its LDT oversight policy, in particular for genetic tests. The reports stressed the need for objective review of tests offered to the public.

In the last several years, other groups have pointed out the lack of effective oversight and made specific recommendations regarding the regulation of LDTs:

- A petition to the FDA from the pharmaceutical manufacturer Genentech (CA, USA) requested that the FDA 'exercise its regulatory authority over all IVD tests pursuant to the risked-based classification system it uses for medical devices', particularly those used to direct treatment with an approved therapeutic. \[112\]
- The Advanced Medical Technology Association (DC, USA) has proposed a modified regulatory paradigm in which all tests are subject to active oversight by the FDA, according to the risk of the test result. \[113\]
- The College of American Pathologists (IL, USA) developed a proposal to include all LDTs in a three-tier 'risk based' oversight model, recommending strengthening of LDT oversight. \[114\]
Many meetings and much time has been spent discussing possible ways to assure that LDTs are safe and effective. Indeed, on 16th June, 2010, the FDA announced plans to hold a public meeting on July 19–20, 2010 to discuss how the agency will oversee LDTs. The meeting discussions are divided into four sessions: Patient Considerations; Challenges for Laboratories; Direct-to-Consumer Marketing of Testing; and Education and Outreach.

**Companion Diagnostics**

Diagnostics used in selecting patients for treatment with a particular therapeutic or in determining what and/or how treatment will be administered have been termed companion diagnostics. These tests hold great promise for personalizing medicine. Companion diagnostics determine a condition of use for a therapeutic drug or biological product, and assure that the effectiveness of the associated therapeutic or biologic when used with the approved companion diagnostic is in accordance with the therapeutic label.

The term companion diagnostic applies to several different scenarios – concurrent development of diagnostic and therapeutic, development of a diagnostic test intended to optimize treatment with a therapeutic that has already been approved, or optimizing a treatment with newly developed therapeutic with previously approved diagnostic test. Companion diagnostics can provide information such as the following:

- Test results that identify a population in which the therapeutic product will achieve greater (or little) effectiveness;
- Test results that identify a patient population that should not receive a particular therapeutic product due to the possibility for therapy-related serious adverse events;
- Test results that identify the characteristics of a disease, condition, or disorder to specifically determine what type of treatment is appropriate;
- Test results that are the basis for selecting a safe and efficacious therapeutic dose.

Companion diagnostics can use any methodology, such as genetic test platforms, immunohistochemistry, enzyme-linked immunosorbent assays, proteomic assays, and so on. Companion diagnostic tests encompass both commercially distributed test kits and LDTs.

The FDA has attempted to clarify the regulatory framework for companion diagnostics and their related therapies. A concept paper was issued in 2005, and the FDA is currently working towards issuing a guidance in this area. There are ongoing efforts towards greater synchronization and harmonization of the FDA centers primarily involved in co-development – CDRH, CDER and CBER. Some of the challenging areas to be addressed include differing laws and regulations governing the review process of different centers, different review timelines imposed on different types of medical products, and postmarket controls that would affect both therapeutic and diagnostic products. Some of the specific considerations that will need to be taken into account include whether a single diagnostic test can have broad applicability to multiple therapeutics, whether multiple IVDs can be used in order to guide the administration of a single drug, whether the IVD and the therapeutic are co-developed simultaneously, or whether either the therapeutic or IVD were developed and marketed first.
A companion diagnostic impacts the ability of a specific therapeutic product to achieve its established safety and effectiveness. Therefore, the risk of the companion diagnostic reflects the risks of use of the associated therapeutic product because it is used to identify patients who respond to particular therapies or regimens, or those who are at higher risk for certain adverse events from a specific therapeutic drug or biologic product. Because of these risks, the FDA believes that diagnostics that fit the definition of companion diagnostics should be subject to oversight with appropriate controls. In general, it is expected that the risk profile of most companion diagnostics will be high, and that many of these products will be regulated as class III devices.

Premarket review of therapeutic products and their associated companion diagnostic applications occurs through an intercenter collaborative process that is structured to provide appropriate review expertise to each product; for example CDER or CBER review the therapeutic product; CDRH or CBER review the companion diagnostic, and a collective review between centers is made of the combination as a whole. Review of post-approval modifications to the labeling of an approved companion diagnostic/therapeutic combination product that will change the conditions of the currently approved therapeutic product or diagnostic, for example, to add new therapeutic indications, new conditions of use, or labeling modifications to the companion diagnostic, may be similarly structured.

Future FDA efforts will target further development of the regulatory structure for companion diagnostics in an effort to provide appropriately scaled scientific review and rigor to products, and clarity on co-development issues described in this article.

**IVD Multivariate Index Assay**

In response to growing concern over certain tests being offered commercially as LDTs, in 2006, the FDA published a guidance draft establishing a particular type of test called an IVD Multivariate Index Assay (IVDMIA). The FDA deduced, through examination of scientific literature and product advertisements made by IVDMIA manufacturers that these tests were beyond the level of complexity envisioned at the time that enforcement discretion was initiated, and presented new and serious risks to patients if improperly designed and validated. The guidance draft announced, therefore, that at a defined period, the exercise of enforcement discretion for tests meeting the definition of an IVDMIA would be terminated.[117]

Among the most common public comments to the draft guidance was a suggestion that the FDA should create a new classification of regulation for this test type, which should be established by a process of rulemaking, and that enforcement should not proceed on the basis of issuing a guidance document. However, this recommendation overlooks an important aspect of IVD oversight: IVDs are regulated according to their intended uses, not according to the methodology used to develop, validate and calculate results.[22] Thus, it is not clear how a rule would adequately define or provide oversight of the universe of existing and possible IVDMIAs. For example, the FDA has cleared assays that fit the IVDMIA description that fall under a number of different regulations:

- 21CFR §866.6040 (gene-expression profiling test system for breast cancer prognosis),[118] classification product code NYI;[119]
A second draft of the guidance document, published for comment in 2007, provided additional clarity to the definition of 'IVDMIA', and described in detail a number of standard regulatory processes that would be applied to oversight of such tests. The IVDMIA definition describes test algorithms that are derived from complex correlations between often large numbers of analytes and patient status. An IVDMIA is defined as a test that: 'combines the values of multiple variables using an interpretation function to yield a single, patient-specific result (e.g., a 'classification,' 'score,' 'index', and so on), that is intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment or prevention of disease, and provides a result whose derivation is nontransparent and cannot be independently derived or verified by the end user.'

It is also important to understand that, by definition, an IVDMIA must include at least one IVD, since a combination of patient characteristics without laboratory tests is not an IVD, according to FDA regulations.

Publishing of the two drafts of the IVDMIA guidance contributed to a broader discussion on risks and benefits of a blanket enforcement discretion policy based solely on the place of manufacture of a test. The FDA's effort to withdraw enforcement discretion from a small category of LDT tests such as IVDMIA has attracted both intense criticism and strong support from various stakeholders, including LDT manufacturers and patient advocacy groups. To date, the finalized guidance has not been published.

**Conclusion**

The success of personalized medicine depends on safe and effective diagnostics. IVDs fall under the FDA's medical device authority and are classified and regulated in a risk-based manner. General regulatory principles that apply to other types of IVDs are highly instructive for IVDs used to personalize treatment, with similar classification, evaluation and review issues. However, personalized medicine introduces new regulatory challenges in developing effective mechanisms to synchronize reviews of therapeutics with IVDs used to personalize treatment. When developing and validating IVDs for use in guiding treatment, there are distinct study design considerations for the evaluation of biomarker-based assays for prognostic and predictive marker-based assays. Today, however, many tests used to guide treatment are being offered as LDTs, without the FDA's review or comprehensive regulatory control. Clarifying the regulatory framework for companion diagnostics and their related therapies, and resolving evolving oversight challenges in this area will likely enable development of successful IVDs for use in providing optimized treatment for patients.
References

   • Provides statutory definition of a medical device.
2. US FDA: Code of Federal Regulations Title 21 §809.3.
   • Provides a definition of in vitro diagnostic product (IVD), defined by regulation.
   • Provides a good overview of IVD regulatory framework.
   • Provides a good overview of IVD regulatory framework with emphasis on genetic testing.
   • Provides updated look on ongoing clinical trial design thinking for predictive and prognostic markers with emphasis on oncology applications.
20. US FDA: Code of Federal Regulations Title 21 § 809.10(e), 809.30 and 864.4020.
   • Guidance for industry and US FDA staff – commercially distributed analyte specific reagents: frequently asked questions.


Websites
103. FAQs about Investigational Device Exemption www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/InvestigationalDeviceExemptionIDE/ucm051480.htm#invitro
108. MammaPrint decision summary www.accessdata.fda.gov/cdrh_docs/reviews/K062694.pdf
112. FDA docket number FDA2008-P-0638–0001 www.regulations.gov/search/Regs/home.html#docketDetail?R=FDA-2008-P-0638
113. FDA docket number FDA-2008-P-0638–0009 www.regulations.gov/search/Regs/home.html#documentDetail?R=0900006480933b6c
114. College of American Pathology; Proposal for oversight of laboratory-developed test www.cap.org/apps/cap.portal?_nfpb=true&cntwrPttl_actionOverride=%2Fportlets%2FcontentViewer%2Fshow&_windowLabel=cntwrPttlt&cntwrPttlt%7BactionForm.contentReference%7D=advocacy%2FIdt%2Findex.html&_state=maximized&_pageLabel=cntwr
115. FDA/CDRH public meeting: oversight of laboratory-developed tests, 19–20 July (2010) www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm212830.htm
•• Provides a good background on laboratory-developed test oversight.
117. FDA docket number FDA-2006-D-0233–0002 (initial guidance published September 2006) and FDA-2006-D-0233–0012 (second draft of guidance, published August 2007)
www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079148.htm
118. Guidance for industry and FDA staff – class II special controls guidance document: gene-expression profiling test system for breast cancer prognosis
www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079163.htm
• Provides guidance on studies and performance requirements for a specific type of IVD Multivariate Index Assay.

119. FDA Product Classification Database
120. Guidance for industry and FDA staff – class II special controls guidance document: cardiac allograft gene-expression profiling test systems
www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187084.htm
121. FDA Product Classification Database: Product classification
www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPCD/classification.cfm?ID=625
122. FDA Product Classification Database: Cytochrome P450 2C9 (CYP450 2C9) drug metabolizing enzyme genotyping system
123. FDA Product Classification Database: Quality control material, genetics, DNA
www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPCD/classification.cfm?ID=2700

Papers of special note have been highlighted as:
• of interest
•• of considerable interest