The FDA's Perspective on the Evaluation of Tumor Marker Tests

Kaiser J. Aziz1,2 and Peter E. Maxim2

Cancer marker tests are often proposed for three intended uses: screening, diagnosis, and monitoring. For each intended use, performance characteristics need to be well defined. The utility of a marker in a given setting depends heavily on two predominant performance characteristics—sensitivity and specificity. These parameters must be established with respect to the intended clinical use of the marker. The value of the marker in a particular situation also depends on the effectiveness of therapy for the malignancy. In reviewing a cancer marker test, the US Food and Drug Administration focuses on both the proposed intended use statement and the clinical utility of the marker. The sponsor is expected to provide specific claims data in support of the safety and effectiveness of the device through well-designed and executed clinical studies. Several cancer markers are already available. In the future, new markers are anticipated that may greatly expand the range of usefulness in cancer diagnosis screening and monitoring.

Indexing Term: government regulation of medical devices

The Medical Device Amendments to the Food, Drug, and Cosmetic Act were enacted in 1976. That law directed the US Food and Drug Administration (FDA) to regulate medical devices under those control levels that are necessary to ensure safety and effectiveness.4 To achieve this goal, the law required the agency to issue regulations placing all devices on the market at that time into one of three regulatory classes (Table 1). To initiate the classification process, the agency was directed to obtain a classification recommendation from an advisory panel.

In 1978, the Immunology Advisory Panel recommended that new cancer markers be placed in class III because these devices had certain potential risks associated with their intended medical uses.

The Safe Medical Devices Act of 1990, the first major revision of the 1976 amendments, extended the previous amendments, which had allowed the use of two major mechanisms to bring a medical device to market: premarket notification [510(k)] and premarket approval. Premarket notification is reserved for devices that can be classified as class I or II and can be compared to a device that is legally, commercially marketed (substantial equivalence). Class III devices are associated with greater risk and are subject to premarket approval to ensure their safety and effectiveness (3). To illustrate the premarket review and approval process, we will use cancer markers for two reasons: these substances (analytes) and their respective technologies belong to the class III category, and the markers are currently the subject of much interest and concern.

The manufacturer of a device placed in class III must submit an application for premarket approval that contains safety and effectiveness data from clinical investigations (Table 2). The application must also contain a description of the manufacturing facility and process, including statements about the quality-assurance and quality-control procedures and materials. Through administrative review, the FDA determines whether the premarket approval application includes the required information and whether it is suitable for filing. If the application is complete and suitable for scientific review, the filing date is the date the FDA receives the complete application, and the 180-day review period provided by the Food, Drug, and Cosmetic Act begins on this filing date. The application is reviewed by the FDA scientific staff, whose evaluations include scientific, medical, statistical, and manufacturing reviews. After this scrutiny, the submission may then be referred to an advisory panel. After receiving recommendations from the advisory panel, the FDA approves or denies a premarket approval application within the 180-day review period, unless the FDA and the applicant have agreed to a longer period.

This evaluation of a product by the premarket approval process takes into consideration many criteria, ranging from test requirements to medical usefulness of the product in diagnosis, screening, prognosis, treatment, or monitoring. For example, the FDA determines the appropriate performance requirements for each class III analyte. To do so, the agency's scientists consider factors such as consequences of a false-positive or false-negative finding in a qualitative test or the importance of an absolute value vs a change in value for a quantitative test. The performance parameters of a new device may be compared with those of previously approved class III products or, for new class III analytes, the device's performance in well-designed medical studies must stand alone for evaluation.

The establishment of class III performance parameters for a product is a complicated process because accuracy and precision are unique for each type of analysis performed in the clinical laboratory. Establishing methodological limits for accuracy, precision, sensitivity, and

---

1 Clinical Chemistry and Toxicology Branches, and Immunology Branch, Food and Drug Administration Center for Devices and Radiological Health, Office of Device Evaluation, Division of Clinical Laboratory Devices, 1090 Piccard Dr., Rockville, MD 20850.
2 Author for correspondence.
3 Nonstandard abbreviations: FDA, Food and Drug Administration; IDE, investigational device exemption; IVD, in vitro diagnostic product; CFR, Code of Federal Regulations; CLIA '88, Clinical Laboratory Improvement Amendments; and HCFA, Health Care Financing Administration.
4 Received March 8, 1993; accepted July 8, 1993.
**Table 1. Classification of Devices**

<table>
<thead>
<tr>
<th>Categories</th>
<th>Level of controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>General: devices whose safety and effectiveness can be ensured by the application of the general regulatory controls that are applicable to all devices, e.g., labeling or manufacturing requirements.</td>
</tr>
<tr>
<td>Class II</td>
<td>Special: devices for which general controls alone are insufficient to provide reasonable assurances of safety and effectiveness and for which sufficient information exists to establish special controls to provide this assurance. Until a special control is established by regulation, only general controls apply. One possible special control is a mandatory performance standard.</td>
</tr>
<tr>
<td>Class III</td>
<td>Premarket approval: devices for which the general and special controls are insufficient to provide reasonable assurance of safety and effectiveness.</td>
</tr>
</tbody>
</table>

Details of regulatory classifications are listed in references 1 and 2.

**Table 2. Premarket Approval Review Statistical Checklist**

I. Organizational and administrative checklist included
   A. Summary of safety and effectiveness
      1. Indications for use
      2. Claims for the device
      3. Summary of studies
   B. Clinical investigations
      1. Protocol
      2. Patient accountability
      3. Description of safety and effectiveness parameters
      4. Documentation of statistical analysis and results

This abbreviated outline is based on the checklist by MR Hanna: Checklist accelerates the review of PMA applications. Med Dev Diag Ind, December 1991, p 31.

Specificity often requires comparative studies, standard reference materials, quality-control materials, and actual clinical specimens. Accuracy and precision must be measured over the dynamic range for which the class III device is intended to be used. Sensitivity and specificity must be considered with respect to the intended clinical use of the product.

Using widely accepted criteria reported in the scientific literature, the FDA examines each class III product and evaluates it against the manufacturer's own labeling claims as to how well it performs and compares it with other class III products marketed for that purpose. Each claim (e.g., screening vs monitoring) must be supported by statistically significant scientific data. Changing the labeling of a device for a particular use can mean that the device may have to meet different sets of supporting data requirements. For example, a cancer marker proposed for longitudinal monitoring will require clinical data accumulated over time to support that claim. A claim for screening or diagnosis would require different types of data in support of these claims.

Investigational and Research In Vitro Diagnostic Products (IVDs)

The Code of Federal Regulations (CFR) describes the appropriate IVD labeling requirements at 21 CFR 809.10(c).6 Regulation 21 CFR 812.2(c)(3) may apply if the manufacturer elects to conduct the investigation under an investigational device exemption (IDE). This type of application is required to show that the IVD is being proposed for investigational use instead of in vitro diagnostic use. An IDE is proposed when the product is being evaluated for use in conjunction with confirmation of the diagnosis by another approved procedure. Manufacturers and distributors of IVDs intended for research or investigational use must ensure that these products are not diverted for unintended uses. Regulation 21 CFR 809.10(c) provides specific guidance relative to labeling for such products and the conditions under which research and investigational IVDs should be represented or used.

**Research use:**

(a) The labeling, including the label on the box of each device, includes the statement "For Research Use Only." The device is not for use in diagnostic procedures.

(b) A written document may be obtained by the manufacturer in which the researcher agrees that such device will not be used for investigations involving clinical use, including diagnosis, prognosis, and monitoring, and that results of tests will not be used in conjunction with patient records or treatment.

**Investigational use:**

(a) The labeling, including the label on the box of each device, includes the statement "For Investigational Use Only." The performance characteristics of this product have not been established.

(b) A written document may be obtained by the manufacturer in which the investigator agrees that the tests are not to be used as an independent diagnostic procedure without confirmation of the diagnosis by another medically established diagnostic product or procedure.

(c) For an IVD, generally, the sponsor submits to the FDA an exemption request with a study protocol under IDE regulation and states that the investigation meets the five exemption criteria listed in 21 CFR 812.2(c)(3).

**Evaluation of Cancer Markers**

The FDA review and evaluation of an IVD cancer marker test focuses on the proposed intended use statement and the clinical utility of the device. The success of a premarket application depends on the sponsor to establish testing criteria, to prove all specific clinical claims, and to confirm the safety and effectiveness of the device through well-designed clinical studies. The nonclinical laboratory studies required for a premarket approval application are those generally associated with the development of an assay system. The characterization of both the antigen and antibody(ies) are very important and critically evaluated. Testing of performance characteristics of the device includes determining interassay, intraassay, and lot-to-lot precision. Linearity, recovery studies, and matrix (e.g., serum, plasma, urine) effects as

---

6 The *Federal Register* provides all the official notices and proposed and final regulations that are issued by federal agencies, including the FDA.
well as the effects of interfacing substances also must be documented. The possible interferences by drugs and therapeutics in the circulation (or in urine) must also be established. A "hook effect," if one is possible in the test system, must also be shown. Finally, a series of stability studies must be performed (in real time) to establish final dating of the product and components (4). Generally, the clinical studies acceptable to the FDA's review process require at least three clinical investigators chosen at separate and geographically diverse sites. A suitable patient accounting system should be adopted. Either patients should be included in all appropriate studies or their exclusion should be explained.

Design of clinical studies for the management of cancer patients should include validation of a decrease in the cancer marker after resection (or other therapy) for the target tumor. Maintenance of a decreased amount of cancer marker in the absence of recurrence or increase of the cancer marker due to recurrence of the tumor needs to be demonstrated in real-time longitudinal studies. The time interval between increase of the cancer marker and the diagnosis (by other means) of recurrence of tumor is the predictive utility of the cancer marker for recurrence. To be useful, the earlier detection of recurrence should be of value for therapy and increased change for therapeutic benefit.

Sample sizes should be established through use of acceptable statistical procedures before the initiation of the clinical study. Selection of sample types to substantiate proposed claims should be adequately described [e.g., serum, plasma (EDTA, citrate, heparin)]. Sample sizes and types need to reflect age, sex, race, and ethnic differences if these exist for the marker being evaluated. Generally, the population for a cancer marker study would include healthy subjects, controls with nonmalignant (benign) diseases, patients with tumors or malignant diseases other than the one under study, and the study population with the malignant disease. The patients in the study group need to be stratified on the basis of stage of disease or other parameters, and care should be taken to include a sufficient number of patients in each subpopulation to achieve meaningful statistical analysis. With regard to clinical parameters, the effect of differentiation of the tumor on the production of cancer marker can be demonstrated by the stratified clinical data, which should include histopathological grade of the tumor as well as the clinical stage of the tumor. Therefore, cancer marker data may need to be analyzed against both clinical factors. If the claim is made to monitor the effect of treatment, the clinical testing should reflect, e.g., the effects of anti-androgens on decreased production of prostate-specific antigen in benign conditions and prostate cancer.

The distribution of the marker concentrations in the various populations studied can be based on the values obtained in single-sample studies. Additionally, the clinical sensitivity and specificity of the test should be calculated. Patients who are being monitored serially for recurrence or response to therapy must be included in numbers sufficient to establish the monitoring claim. The various stages of disease may need to be represented, along with sufficient numbers of patients, to establish the utility of the test during different treatment regimens (surgery, radiation therapy, chemotherapy, and hormonal therapy). Particular attention should be given to show that the patients' clinical course and response to treatment are being monitored by the values of the serum cancer marker values. Analysis of the data by acceptable statistical procedures is required, as is the careful presentation of the results, so that the correlation of marker values with clinical changes is obvious (Table 3).

### Analysis of Cancer Marker Data

It is virtually impossible to generalize what analyses are appropriate for these studies. Various aspects of the data lend themselves to specific statistical treatments. Many references and recommendations detail how the data obtained in the laboratory studies can be analyzed statistically. The most difficulty is encountered in analyzing the serial samples obtained in longitudinal monitoring studies. Often, the most valid analysis is determining the correlation with the patients' clinical status.

The following represent some general parameters that are often included in cancer marker analyses:

The predictive value is calculated by the formula frequently referred to as Bayes' formula. Before a cancer marker test result is evaluated for its intended clinical use, an understanding of the epidemiological sensitivity, specificity, and prevalence as well as analytical interpretation of sensitivity and specificity are required to establish clinical utility of cancer markers. The predictive value of a marker changes significantly when there is a change in the prevalence of the type of cancer in the population under study. For example, in a hypothetical situation, we might know from the value for sensitivity and specificity that the marker test was positive in 95% of cancer patients and negative in 95% of patients without cancer; i.e., it has a sensitivity of 95% and a specificity of 95%. The relationship between these two parameters and the prevalence of the type of cancer under

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Indication for use</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Fetoprotein</td>
<td>Neural tube defects, and tumor marker for testicular cancer (nonseminoma)</td>
</tr>
<tr>
<td>Carcinoembryonic antigen</td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td>Estrogen receptor</td>
<td>Aid in prediction of patient response to hormonal therapy, and prognosis and management of breast cancer patients</td>
</tr>
<tr>
<td>Ovarian CA (CA 125)</td>
<td>Ovarian cancer (second-look)</td>
</tr>
<tr>
<td>Progesterone receptor</td>
<td>Aid in prediction of patient response to hormonal therapy, and prognosis and management of breast cancer patients</td>
</tr>
<tr>
<td>Prostate-specific antigen (PSA)</td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>Soluble IL-2 receptor</td>
<td>Hairy cell leukemia</td>
</tr>
</tbody>
</table>
study is such that the predictive value will vary according to how many individuals in the population being studied have that type of cancer. In other words, the number of false positives and false negatives will vary between populations that have different prevalences of the type of cancer under study. A particular marker test will have higher predictive value when it is applied to a population with a higher prevalence of the type of cancer being studied. Frequently, this is why a diagnostic marker test fails a screening test when applied to a population in which the prevalence of that type of cancer is very low. Usually, cancer marker tests are applied on the basis of previous clinical information about cancer patients; however, this has the effect of selecting a given population for testing that has a higher prevalence than the general population for the proposed cancer test. Therefore, prevalence is probably the most important factor affecting the usefulness of a cancer marker test result, particularly the intended clinical use of the test. As mentioned above, the accuracy of the test will depend on the patient population chosen. The most important determinants of accuracy are sensitivity, specificity, and disease prevalence in the population under study.

Clinical laboratorians are interested in having new tests available that are suitable for screening, diagnosis, monitoring, and prognosis of malignant disease. Several tests are currently under clinical investigation, but it remains to be seen whether any of them offer any new advantages over current tests. An ideal screening test would detect cancer in asymptomatic individuals before the disease reaches a symptomatic stage. However, given that the prevalence of any one cancer at a specific point in time is low, the cancer markers available thus far have limitations as screening tests in asymptomatic individuals. Cancer marker tests usually do not offer sufficient sensitivity to detect early disease or small tumors. They also lack specificity, in that they are present, or increased, in subjects with benign disease or even in normal individuals. This lack of sensitivity and specificity diminishes the usefulness of markers for screening or diagnosis of carcinomas. An extreme increase in a marker often indicates a poor prognosis and in some malignancies can indicate the need for more aggressive treatment. The value of most cancer markers to date has been in the monitoring of cancer patients for response to therapy or for recurrence of disease. The clinical utility of cancer markers is therefore most often expressed as an aid in the management of cancer patients and in assessing prognosis (4-6).

To be useful, the change in the concentration of a cancer marker should reflect the change in tumor burden and allow for therapeutic intervention. If medical treatment for a malignancy is ineffective, increasing concentrations of a marker may correlate strongly with prognosis based on increasing tumor burden. Cancer markers are useful in therapies in which the marker can assist in predicting response to therapy, in predicting prognosis, and, perhaps most importantly, in following the course of the disease.

Labeling

The FDA reviews product labels for conformance with regulations in section 502(a) and section 502(0)(1) of the Food, Drug, and Cosmetic Act and 21 CFR 509.10, resulting in a fairly standardized package insert familiar to laboratorians. The package insert for cancer marker kits should advise the clinical laboratory personnel not to use kits interchangeably and to identify the assay being used on the report form, because the source of the test kit reagent components (e.g., monoclonal vs polyclonal antibodies) for a given cancer marker can influence the results. Therefore, accurate and precise serial testing of markers requires that same-test methodology be used throughout the study.

Known limitations may also be described in the labeling, including interferences such as human anti-mouse antibody in patients receiving monoclonal-antibody therapy or hormonal therapy.

Integration of CLIA '88 with FDA Programs

The FDA's new role under the Clinical Laboratory Improvement Amendments of 1988 (CLIA '88; 42 CFR 493(k)) (which includes both premarket as well as postmarket activities), in conjunction with the Centers for Disease Control and Prevention and the Health Care Financing Administration (HCFA) interagency activities, will require that IVDs used by clinical laboratories be properly validated either by the manufacturer through FDA's regulatory process or by laboratories through HCFA's responsibilities under CLIA. For example, HCFA should have access to information from the FDA on IVDs labeled for research and investigational use for use by laboratory surveyors. Such information could help surveyors better assess the scope of the validation data needed to comply with CLIA standards.

Special consideration is also needed for the quality-control section of the product labeling as a result of CLIA '88 implementation. According to the FDA's role in clinical laboratory device approval, as defined in 42 CFR 493(k) of the CLIA '88 regulation, the FDA has developed a guidance document for manufacturers that describes the type of validation studies for premarket submissions relating to quality control.

In this article, we have presented the basic concepts and applications of premarket review and approval process for cancer markers. The relationship between a particular cancer-related analyte and the clinical status of the patient must be emphasized, and tests designed to measure cancer-related analyte concentrations must be accurate and precise. Continued technological advances are expected to provide not only new analyte markers but also new technologies to detect them. These technological advances may provide an opportunity to extend the use of markers to early diagnosis and detection. It is commonly accepted that early intervention offers the most promise for better prognosis in cancer patients. In addition, the use of panels of two or more approved markers may contribute
to the improved sensitivity and perhaps the diagnostic
efficacy for a particular disease.

We thank members of the Division of Clinical Laboratory De-
vices for discussions that contributed to the preparation of this
manuscript.

References
1. Aziz KJ, Kennedy R. Medical Device Amendments of 1976 and
classification of clinical chemistry in vitro diagnostics devices.