Serum tumor markers: Part II
Practical considerations and limitations of testing

By Daniel M. Hoefner, MT, PhD, DABCC, FACB

Part I of this review [MLO, December 2005 Clinical Issues, p. 20] summarized the utility of serum tumor markers in highest usage; the analytes included α-fetoprotein (AFP), CA 125, CA 15-3, CA 27.29, CA 19-9, carcinoembryonic antigen (CEA), choriogonadotropin (hCG), PSA, and thyroglobulin. The primary use of these markers is to assess the efficacy of therapy in patients being treated for cancer, as well as monitoring patients for recurrence following curative measures. Practical considerations and caveats associated with the testing of these analytes will be addressed here in Part II. To explore these issues, attributes of an ideal tumor marker will be illustrated and contrasted with the tests that are presently in common use; such features are listed in Table 1. If such an idealized marker existed, it could function for all of the uses that were mentioned in the first part of this manuscript, viz., aiding cancer diagnosis; monitoring for recurrence, prognosis, and staging; detection of residual disease; screening; and for monitoring treatment. While great efforts are being put forth to find markers that fulfill these criteria, currently, no analyte comes close to meeting them.

Sensitivity and specificity
An ideal marker should be highly specific so that samples from patients without cancer would not give false positive results. This is one of the greatest weaknesses of most tumor mark-
Table 1. Attributes of an ideal tumor marker

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor-specific (no overlap between health and disease)</td>
<td>Released exclusively from organ-specific tumor tissue, but absent or below a threshold in health.</td>
</tr>
<tr>
<td>Type-specific</td>
<td>Marker is only elevated in response to one cancer type.</td>
</tr>
<tr>
<td>Detection of early cancer and early prediction of recurrence</td>
<td>High sensitivity—marker is produced at high levels, elevated in all patients with cancer, and detectable before clinical symptoms become apparent.</td>
</tr>
<tr>
<td>Concentration change relative to total tumor burden</td>
<td>If a tumor grows or gets smaller, this should be reflected by a corresponding increase or decrease in tumor marker levels.</td>
</tr>
<tr>
<td>Easily measured and cost-effective</td>
<td>Relatively easy to measure with common instrumentation from minimally invasive sampling (i.e., blood or other body fluids).</td>
</tr>
</tbody>
</table>

ers in contemporary use. In order to be sensitive enough to detect a large proportion of patients who have the disease, the upper reference level (URL) would need to be set so low that a significant number of healthy patients would be flagged as positive for the test. The clinical (diagnostic) sensitivity of a tumor marker assay refers to the percentage of individuals with cancer who have an elevated tumor marker result. While a test may be elevated in the majority of patients with malignancy, if it is elevated in even a small proportion of healthy patients (non-specificity), there may be an unacceptably high number of false positives (due to the relatively low incidence of cancer in the population) if it is used for screening purposes.

Another drawback is that most markers in use are rarely elevated in just one type of cancer; therefore, they do not function to definitively identify the source of the malignancy. Even PSA (or prostate-specific antigen) is not limited to the prostate gland, as was once thought. Numerous studies have shown that other tissues also express the protein, including the perirethral gland and breast in women and pancreas and salivary gland, independent of gender. The key to sensitivity and specificity is to find an analyte in which no overlap exists at a given concentration between patients in health and disease — no such marker exists for cancer.

Reference intervals

PSA has been frequently described as the most ideal tumor marker available. Therefore, PSA will be used to exemplify some of the problem issues encountered with tumor markers. Traditionally, the reference range for PSA has been 4.0 ng/mL. Partitioning factors are important, however, and analyte levels are significantly affected by age and race. For tumor markers, the 95th percentile determined (non-parametrically) from an apparently healthy patient cohort is frequently used to set the URL. The concentration of PSA at the 95th percentile for white, apparently healthy 40-, 50-, 60-, 70-, and 80-year-old men is about 2.0, 2.7, 3.8, 5.4, and 7.4 ng/mL, respectively. African-Americans have significantly higher PSA levels than similarly aged white males and there are differences in Latinos and Asians as well. There are arguments for and against the use of age-specific ranges and controversy exists over where to set the upper limit. For the most part, however, two URLs are currently promoted, namely 2.5 ng/mL and 4.0 ng/mL, regardless of age and race. The dilemma is that up to two-thirds of cancers may be missed at the 4.0 ng/mL level; and, based on recent data from a different cohort using a URL of 2.5 ng/mL, only 40.5% of cancer cases were detected. Conversely, the lower the URL is set, the greater the number of men in whom false positive results occur, leading to anxiety, additional laboratory testing, and unnecessary biopsies. While it is a well-known phenomenon that PSA concentrations increase as individuals get older, age appears to also impact other tumor marker results as well. For example, Lopez, et al, showed that CEA, CA 19-9, CA 15-3, and PSA were all increased in healthy older individuals and concluded that the increase was related to the aging process rather than the presence of cancer.

There are other confounding issues. Since PSA levels may be increased in benign prostatic hyperplasia (BPH) and prostatitis, elevated levels are not specific for malignancy (~50% of men with BPH have levels >4.0 ng/mL). Furthermore, PSA levels do not correlate with the aggressiveness of the cancer. Relative to the detection of early recurrence, many of the tumor markers serve adequately in this capacity, but only if the tumor has been shown to produce the particular marker of interest. Yet, even in some cases of advanced cancer, PSA is never elevated, as is the case for other malignancies and their respective archetypal markers. Finally, there are small differences between manufacturers of assay kits that may also influence reference interval determinations.

Tumor markers serve adequately in this capacity, but only if the tumor has been shown to produce the particular marker of interest. Yet, even in some cases of advanced cancer, PSA is never elevated, as is the case for other malignancies and their respective archetypal markers. Finally, there are small differences between manufacturers of assay kits that may also influence reference interval determinations.

PSA has been frequently described as the most ideal tumor marker available.

Kinetics and other factors

Tumor marker levels are frequently used to monitor the effectiveness of treatment. If the treatment is successful, levels usually drop; if the patient is not responding favorably to the therapy, levels may stay flat or actually increase. Due to the heterogeneity of cancer, however, it is possible for the tumor burden to increase while a particular tumor marker is decreasing. The caution here is that concentrations can sometimes be misleading and low and/or decreasing levels do not always indicate patient improvement. As a whole, most analyte levels are reflective of tumor burden, but they suffer from not fulfilling the criteria in all situations. As with other diagnostic tools, the expense should allow testing to be accessible to the entire patient population, the procedure should be minimally invasive, and testing should be analyzable on routine testing platforms. For the

Continue on page 14
most part, immunoassays fulfill these latter criteria of cost effectiveness and ease of testing.

High-dose hook effect (prozone phenomenon or antigen excess)

Many assay interferences are not specific to tumor markers but are a consequence of the type of assay. In single-step “sandwich” immunoassays, very high levels of antigen (tumor marker) may exist in excess of the amount of capture antibody, so that the excess (free) antigen competes with bound antigen for the signal-labeled antibody (conjugate) binding sites. This “uncaptured” conjugate-analyte complex gets removed in the wash step prior to quantification, which leads to falsely decreased results. This is known as the “high-dose hook effect,” which is illustrated in Figure 1. When this phenomenon occurs, the apparent results are still elevated but may represent only a fraction of the true levels. It is most commonly seen with assays in which the analyte has a very large pathophysiological range, which is common among tumor markers. While two-step assays appear to overcome this difficulty, they may not be completely impervious to the problem.11

Heterophile antibodies

Antibodies that are produced against poorly defined antigens are known as heterophile antibodies. This term is also used to identify human antibodies that are directed at animal antibodies (human anti-animal antibodies or HAAA). They may be produced following passive exposure from handling and working with pets and animals or arise following direct exposure from injected medical therapeutics that were made from animal-derived immunoglobulins. Some of the more common (and problematic) are antibodies that develop against mouse antibodies, commonly referred to as HAMA. When two-site (sandwich) immunoassays are performed, HAMA in the patient sample may interact with the animal-derived reagent antibodies and act to either sterically hinder (block) or cross-link them, resulting in interference that may either falsely decrease or increase the true result.

The product literature for nearly all immunoassays includes comments regarding the potential for interference from heterophile antibodies. In addition, most package inserts included in clinical laboratory kits for tumor marker assessment include verbiage resembling, “the results of this test should be interpreted in combination with the entire clinical picture (patient history, clinical exam, other testing, and diagnostic procedures)” and warn that elevated levels may occur in non-neoplastic conditions. These admonitions are especially true for tumor markers in which so many caveats exist, and relying solely on laboratory data to make a clinical decision is unwise. Even in light of manufacturer’s best efforts to prevent this occurrence through the use of blocking agents, and given their product warnings, reports of mistakes are not infrequent. In a recent paper,12 12 women were incorrectly diagnosed with choriocarcinomas due to falsely elevated hCG results caused by heterophile antibodies. More tragic than the adverse cancer diagnosis was that most of the women had already undergone pointless surgical and/or chemotherapeutic treatments before the patient’s interfering heterophile antibodies were identified.

Non-linearity of dilutions

Infrequently, an apparent over-recovery of a tumor marker occurs when serum samples are diluted. In the absence of dissociating conditions, highly mucinous tumor markers, such as CA 19-9, have an opportunity to aggregate together — a phenomenon that occurs more commonly with increasing serum tumor marker concentrations. Many factors may contribute to this, including the extensive heterogeneity in carbohydrate composition and quantity, and the in vivo level of naturally occurring weak-affinity anti-carbohydrate antibodies.13 Laboratorians and clinicians should be aware of this potential occurrence, since results achieved on diluted samples may be significantly higher than from the neat sample.

Addressing reagent changes

Unlike many other analytes measured in the laboratory, it is the course of tumor marker concentrations over time that is most useful. Single values should not drive decisions — a solitary result is much more meaningful when considered in light of previous values. Consequently, it is critical to thoroughly evaluate new reagent lots to ensure that patient results are returning comparable results with the current and new lots, prior to moving to a new kit. In addition, if moving from one manufacturer or platform to another, it is important to notify the clinician and offer him the opportunity to rebaseline his existing patient cohort for the particular analyte. Even if the evaluation of a new method does not indicate significant differences between the two, occasional patient samples may still give anomalous results with two different methods.
Parallel rebaselining seems to be the most appropriate means of performing this type of study. Using this approach, a new baseline is established for each patient by parallel testing each sample with the current and new lot of reagents, as tests are routinely ordered. Both results should be reported to the clinician and he should be given a window of time that the rebaseline testing will continue to be performed. The difficult aspect of parallel rebaselining is that it must be offered for an extended period of time to allow all patients to be cycled through.

An alternate approach is to use samples that are saved in anticipation of a test change. Here, routinely ordered samples for the marker in question are analyzed with the current method and then stored frozen for several weeks to months, prior to testing with the new method. After enough samples have been collected to comprise the entire patient cohort, they are tested with the new reagents and results are compared to the originals. With this approach, one must ensure that analyte degradation does not occur during storage, which can be done by retesting some of the frozen samples by the original method. The least satisfactory means is to simply compare the values from the two kits during the method evaluation and alert the clinician that all results could be expected to shift by a given amount, based on linear regression of the data. This approach should only be used in the unfortunate event that there is not enough of the current lot of reagent (and it is no longer available) to perform the appropriate studies.

**Variability of test results**

In addition to the above comments regarding reagent changes, one must also be mindful of the biological, analytical, and pre-analytical variables that can affect the results. For illustration purposes, let us consider total PSA again. Suppose that a patient had an initial PSA result of 2.9 ng/mL and on a subsequent visit, the result had climbed to 5.1 ng/mL. By most criteria, the first result would be considered well inside the reference interval of ≤4.0, but the second would fall into the gray zone (4.1 ng/mL to 10 ng/mL) and, depending on the percent free PSA, could trigger a biopsy. At the very least, it would likely cause patient anxiety. To determine whether the two results are actually statistically different from one another, however, the variability that is involved must be considered. A recent review of biological variation for PSA suggests that the SD at this concentration would be about 0.8 ng/mL. In addition, preanalytical and analytical variability must also be factored in. For the sake of simplicity, let us assume that the preanalytical factor is negligible and that the analytical component has an SD of 0.16 ng/mL (i.e., 4% CV at 4 ng/mL), which is in optimistic agreement with most product claims. In order to be considered different (P<0.05), two results would need to vary by more than indicated in the following formula:

\[
1.96 \times \sqrt{2} \times \sqrt{\text{Total SD}} \quad \text{or} \quad 1.96 \times \sqrt{2} \times \sqrt{0.8^2 + 0.16^2} = 2.26
\]

The value 1.96 is the bidirectional Z-score (P<0.05), and the square root of two is the factor used when comparing two results. The product of these two numbers can be considered a constant (~2.77) that is multiplied by the overall SD (the square root of \([0.8^2 + 0.16^2]\)) when two results are compared. Therefore, in this example, the value of 2.9 ng/mL would not be considered statistically different from 5.1 ng/mL. This approach may be used to compare two results for any analyte in which the variability is known or can be estimated.

**Poor judgment**

In a 2003 report of data collected at a large teaching hospital in the United Kingdom that evaluated 27,323 tumor marker requests, many tests were apparently ordered erroneously. It was discovered that about 17% of all CA 125 and ~26% of all CA 15-3 tests were performed on samples from males. In addition, tumor marker testing was frequently ordered before patients were even diagnosed as having cancer. The authors concluded that “tumor markers are frequently and inappropriately requested…” Hopelessly, the clinical practice guidelines that were discussed in the first part of this report [MLO, December 2005 Clinical Issues, p. 20] will begin to have a greater impact in clinical practice and continued educational efforts by organizations such as the American Association for Clinical Chemistry, the National Academy of Clinical Biochemistry, and European Group on Tumour Markers will impart the effectiveness and limitations of such testing upon the minds of the caregivers.

In a recent paper, 12 women were incorrectly diagnosed for choriocarcinomas due to falsely elevated hCG results caused by heterophile antibodies.

**Summary**

The attributes of an ideal tumor marker are well beyond the capacity of the tests that are currently available. While some markers do an adequate job at the early detection of recurrence and as a means to monitor the efficacy of treatment, there are serious deficits regarding sensitivity and specificity. In addition, although they are easily measured and cost-effective, the limitations of immunoassays are also a detriment. PSA was used to illustrate many of the limitations and difficulties associated with tumor marker testing and cancer assessment. For the most part, however, it is a better marker than many of the other analytes examined for their corresponding malignancies. Nevertheless, PSA still has major limitations — as is the case for all tumor markers in current use. Although the use of tumor markers is widespread and they do have much to offer, clinicians need to be made aware of their limitations, which can be achieved through good communication between laboratorians and the caregivers who are ordering the tests.

Daniel M. Hoefner, MT, PhD, DABCC, FACB, is corporate director of Clinical Science Assessment for Laboratory Corporation of America, Burlington, NC.

**References**


Continues on page 16
Join us at the nation’s biggest lab and pathology management gathering. Our 50 speakers bring you the best and the latest to help you and your lab prosper in tough times. Hot topics this year are “Lab Readiness with a Capital ‘R’!” and “Building the Automated Histology Lab of the 21st Century.” Get information you can’t find anywhere else from our esteemed faculty of 50 top-notch lab leaders. Discover for yourself why more than 50 percent of the nation’s savviest lab directors and pathologists make the Executive War College their “must-attend” event of the year. Register today!

**Essential Insights! 50 Top Speakers!**

- Meet America’s newest lab competitor: Sonic Healthcare, Ltd. of Australia in a compelling strategic case study.
- First look at the paperless, fully-automated laboratory at the newly-opened Century City Doctors Hospital.
- Consumer-Directed Health Plans: essential knowledge for labs and path groups on how to collect from consumers.
- Beating back the 88305 threat to pathology.
- “Big Bang” lab makeover: new space, new instruments, new workflow trigger gains in TAT and lab productivity.
- Microarrays and other “hot” molecular assays.
- Integrated informatics and middleware solutions: lessons from first mover labs and pathology group practices.
- Laboratory automation round-up: evaluating all options for pre-analytical, analytical, and post-analytical processes.
- Real-time pathology generates AP reports in 2 hours at the newly-opened Century City Doctors Hospital.
- Meet America’s newest lab competitor: Sonic Healthcare, Ltd. of Australia in a compelling strategic case study.
- First look at the paperless, fully-automated laboratory at the newly-opened Century City Doctors Hospital.
- Consumer-Directed Health Plans: essential knowledge for labs and path groups on how to collect from consumers.
- Beating back the 88305 threat to pathology.
- “Big Bang” lab makeover: new space, new instruments, new workflow trigger gains in TAT and lab productivity.
- Microarrays and other “hot” molecular assays.
- Integrated informatics and middleware solutions: lessons from first mover labs and pathology group practices.
- Laboratory automation round-up: evaluating all options for pre-analytical, analytical, and post-analytical processes.
- Real-time pathology generates AP reports in 2 hours at the newly-opened Century City Doctors Hospital.

For Details and to Register, go to www.darkreport.com

800-560-6363

Visit www.rsleads.com/602ml-007