Clinical Chemistry 54:8
e1–e10 (2008)

National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines for Use of Tumor Markers in Clinical Practice: Quality Requirements*

Catharine M. Sturgeon,[1†] Barry R. Hoffman,[2] Daniel W. Chan,[3] Soo-Ling Ch'ng,[4] Elizabeth Hammond,[5] Daniel F. Hayes,[6] Lance A. Liotta,[7] Emmanuel F. Petricoin,[7] Manfred Schmitt,[8] O. John Semmes,[9] György Söletormos,[10] Elena van der Merwe,[2] and Eleftherios P. Diamandis[2]

BACKGROUND: This report presents updated National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines summarizing quality requirements for the use of tumor markers.

METHODS: One subcommittee developed guidelines for analytical quality relevant to serum and tissue-based tumor markers in current clinical practice. Two other subcommittees formulated recommendations particularly relevant to the developing technologies of microarrays and mass spectrometry.

RESULTS: Prerequisites for optimal use of tumor markers in routine practice include formulation of the correct clinical questions to ensure selection of the appropriate test, adherence to good clinical and laboratory practices (e.g., minimization of the risk of incorrect patient and/or specimen identification, tube type, or timing), use of internationally standardized and well-characterized methods, careful adherence to manufacturer instructions, and proactive and timely reactions to information derived from both internal QC and proficiency-testing specimens. Highly desirable procedures include those designed to minimize the risk of the reporting of erroneous results attributable to interferences such as heterophilic antibodies or hook effects, to facilitate the provision of informative clinical reports (e.g., cumulative and/or graphical reports, appropriately derived reference intervals, and interpretative comments), and when possible to integrate these reports with other patient information through electronic health records. Also mandatory is extensive validation encompassing all stages of analysis before introduction of new technologies such as microarrays and mass spectrometry. Provision of high-quality tumor marker services is facilitated by dialogue involving researchers, diagnostic companies, clinical and laboratory users, and regulatory agencies.

CONCLUSIONS: Implementation of these recommendations, adapted to local practice, should encourage optimization of the clinical use of tumor markers.

We report National Academy of Clinical Biochemistry (NACB)[11] guidelines and recommendations regarding general laboratory testing for tumor markers, (unpublished data and *(1)*), microarrays for cancer diagnosis, and MALDI and related mass spectrometry (MS) techniques *(2)* for profiling of tumors and patient sera. This report presents a condensed version of the full NACB document, which cannot be published here in

* This report is a condensed version of the full NACB document, which because of constraints of space cannot be published here in its entirety. However, the whole manuscript is accessible in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol54/issue8.

[1] Department of Clinical Biochemistry, Royal Infirmary of Edinburgh, Edinburgh, UK; [2] Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, and Department of Laboratory Medicine and Pathobiology, University of Toronto, Ontario, Canada; [3] Department of Pathology, Center for Biomarker Discovery, Johns Hopkins Medical Institutions, Baltimore, MD, USA; [4] Tanjung Bungah, Penang, Malaysia (previously Nuclear Medicine Section, Division of Human Health, International Atomic Energy Agency, Vienna, Austria); [5] Department of Pathology, LDS Hospital, Salt Lake City, UT, USA; [6] Breast Oncology Program, University of Michigan Comprehensive Cancer Center, Ann Arbor, MI, USA; [7] Center for Applied Proteomics and Molecular Medicine, College of Sciences, George Mason University, Manassas, VA, USA; [8] Clinical Research Unit, Department of Obstetrics and Gynecology, Technical University of Munich, Munich, Germany; [9] Department of Microbiology and Molecular Cell Biology, Center for Biomedical Proteomics, Eastern Virginia Medical School, Norfolk, VA, USA; [10] Department of Clinical Biochemistry, Hillerød Hospital, Hillerød, Denmark.

† Address correspondence to this author at: Department of Clinical Biochemistry, Royal Infirmary of Edinburgh, 51 Little France Crescent, Edinburgh EH16 4SA, UK. Fax 44-131-242 6882; e-mail c.sturgeon@ed.ac.uk.

Previously published online at DOI: 10.1373/clinchem.2007.094144

[11] Nonstandard abbreviations: NACB, National Academy of Clinical Biochemistry; MS, mass spectrometry; IQC, internal QC; PT, proficiency testing; PSA, prostate-specific antigen; IS, international standard; IRR, international reference reagent; hCG, human chorionic gonadotropin; LOE, level of evidence.

its entirety, owing to space limitations. Much additional information is available in online supplemental tables, and the full original manuscript is also accessible as an online supplemental file (see the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol54/issue8).

An explanation of the methods used for developing these guidelines will be provided (unpublished data). The disciplines of all authors and statements of conflicts of interest, declared according to NACB requirements, are provided in the Supplemental Data Disclosures Table in the online Data Supplement. The latter are also listed at the end of this manuscript. All comments received about these guidelines are also recorded in the online Data Supplement (see the Supplemental Data Comments Received Table), together with responses to the comments.

The preparation of these guidelines included review of the literature relevant to the use of tumor markers. Particular attention was given to review articles, including the few relevant systematic reviews, and to guidelines issued by expert panels. When possible, the consensus recommendations of the NACB panels were based on available evidence, i.e., were evidence based.

NACB QUALITY REQUIREMENTS FOR THE USE OF TUMOR MARKERS

The quality requirements presented here, further developments of previous recommendations of the NACB and EGTM *(1)*, are relevant to the clinical use of the tumor markers most frequently measured in routine clinical laboratories (see Supplemental Table 1 in the online Data Supplement for a list of these markers and their general characteristics), under the following broad headings:

- Preanalytical requirements: choice of tumor marker, specimen type, specimen timing, sample handling.
- Analytical requirements: assay standardization, internal and external QC, interferences.
- Postanalytical requirements: reference intervals and interpretation and reporting of tumor marker results.

This report also briefly addresses some of the issues relevant to enhancing the clinical utility of tumor marker testing, both now and in the future.

PREANALYTICAL QUALITY REQUIREMENTS

Reporting of erroneous tumor marker results is more likely to cause undue alarm to patients than is reporting of erroneous results for many other laboratory tests. In addition to adhering to general preanalytical recommendations applicable to all diagnostic tests *(2)* and encouraging appropriate test requesting *(1, 3, 4)*, the laboratory must exercise extra vigilance in ensuring that correct results are reported *(5)*. Errors reportedly occur more often in the preanalytical than the analytical phase [30%–75% and 13%–31%, respectively, as reported in one review *(6)*, and 10 times as often in a transfusion medicine study *(7)*]. The majority of preanalytical errors for tumor markers are attributable to simple specimen-handling errors, such as inappropriate timing and incorrect specimen identification, and data-entry errors, the occurrence of which should be minimized by good laboratory practice and effective auditing procedures. A number of additional circumstances may lead to misleading results (see Supplemental Table 2 in the online Data Supplement). Implementing the NACB Guidelines for Tumor Markers *(8)*, particularly by discouraging inappropriate test requesting *(9, 10)*, ensuring appropriate specimen timing, and encouraging requests for confirmatory specimens when required, should decrease the risk of causing patients to suffer unwarranted distress or undergo unnecessary clinical investigations.

With the advent of electronic health records, the process of tumor-marker ordering should be linked with preanalytical precautions available through databases designed to provide support for clinical decision-making *(11)*, such as the database recently made available through the AACC *(12)*.

ANALYTICAL QUALITY REQUIREMENTS

With the increasingly widespread use of automated immunoassay analyzers, responsibility for analytical quality has come to rest largely with the diagnostics industry, which must meet quality requirements defined by national and international regulatory authorities. For satisfactory measurement of any analyte, it is crucial that laboratories independently monitor their own performance carefully to ensure that analyzers are being used appropriately and to confirm that methods are being performed according to specifications. This process is best achieved by implementation of rigorous internal QC (IQC) procedures and participation in well-designed proficiency testing (PT) (external quality assessment) programs *(1)*. It is of course crucial that laboratories not only participate in such programs but also take appropriate remedial action, immediately investigating the cause of unsatisfactory results.

NACB recommendations for both IQC and PT are presented in Supplemental Table 3 in the online Data Supplement. Although most of these recommendations are applicable to all analytes, several have particular relevance to tumor markers. Specimens for both IQC and PT should always resemble clinical sera as closely as is feasible. When clinical decision points are employed, stable and consistent performance is essential, including the use of IQC specimens at concentrations close to decision-point concentrations. Such protocols are critical for screening tests performed on

asymptomatic individuals, such as screening for prostate cancer by measuring prostate-specific antigen (PSA) *(13)*, or for cases in which chemotherapy may be instituted on the basis of an increasing tumor-marker concentration in the absence of other scan evidence, such as monitoring of testicular cancer patients *(14)*. Functional sensitivity (i.e., the lowest result that can be reliably reported, best defined as the concentration at which the day-to-day CV is <20%) is also very important for certain tumor marker applications, e.g., monitoring PSA in prostate cancer patients after radical prostatectomy. By repeatedly issuing specimens of the same low concentration pool, PT schemes can provide valuable information about the stability of results over time *(15)*. Because tumor markers are often used to monitor cancer patients for long time periods, assessment of long-term assay stability at other analyte concentrations is also advisable.

Long-term monitoring presents major challenges, because patients may change hospitals or laboratories may change methods of tumor-marker measurement. Ideally results obtained with different methods would be fully interchangeable, but unfortunately this is not the case, with between-method CVs in excess of 20% observed in PT schemes for some tumor markers *(16)*. Major causes of between-method variation for these complex analytes include poor calibration, differences in antibody specificity, and differences in method design *(17)*.

Reasonably standardized and accurate calibration should be achievable, but only for those analytes for which a recognized international standard (IS) or international reference reagent (IRR) is available (see Supplemental Table 4 in the online Data Supplement) and universally adopted by manufacturers for primary calibration of their methods. There are no IS for any of the important CA series of tumor markers, a major gap that should be addressed urgently. When relevant IS or IRR are available, recovery experiments undertaken through PT schemes, together with linearity and stability studies, provide the independent validation of consensus target values that is essential for a well-designed PT scheme. Conveniently, because PT providers should be working toward improving between-method agreement, these experiments also permit assessment of the accuracy of calibration of individual methods, helping to identify methods requiring improvement (e.g., methods over- or underrecovering the relevant IS by more than 10%). Long-term PT data can also confirm the effect of successful introduction of a new IS. Data from one scheme, for example, suggests that mean CVs decreased from 21.9% in 1995, before the first PSA IRR was introduced, to 9.5% in 2004 *(18)*.

IRR for isoforms of human chorionic gonadotropin (hCG) *(19)* and PSA *(20)* (developed under the auspices of the IFCC) provide additional tools for elucidating method-related differences associated with the second major cause of method-related differences, antibody specificity. Carefully designed experiments with the IRR for PSA and free PSA enable assessment of the calibration and equimolarity of assays for PSA, which are particularly critical in the context of prostate cancer screening. Similarly, experiments with the 6 recently established IRR for hCG isoforms should elucidate what is actually measured by available methods for hCG analysis *(19, 21)*, an issue of major importance for oncology applications in which recognition of a broad spectrum of hCG-related molecules is recommended *(22)*. Epitope-mapping projects such as those carried out under the auspices of the International Society for Oncology and Biomarkers may enable the formulation of broad recommendations regarding the most clinically appropriate antibody specificities for some tumor markers, as has already been achieved for hCG *(22)*.

Such studies may lead to better understanding of optimal method design for complex tumor markers, thereby addressing the third major cause of method-related differences. Differences in method design are likely to contribute to numerical differences in results observed and to differences in method robustness attributable to clinically relevant interferences (see Supplemental Table 5 in the online Data Supplement). Maintaining vigilant awareness of the potential for the latter is essential. Ultimately, the most effective way to minimize the risk of unrecognized interference leading to serious clinical error is to promote regular dialogue between laboratory and clinical staff, thereby encouraging early discussion and investigation of any results not in accord with the clinical picture *(14)*.

POSTANALYTICAL QUALITY REQUIREMENTS

Provision of helpful reports prepared according to the NACB recommendations in Supplemental Data Table 6 encourages communication between laboratory and clinical staff, a highly desirable component of the process of achieving the best use of tumor marker tests. Clinical biochemistry laboratories should be prepared to engage more actively in the interpretation of tumor marker results, ensuring that appropriately validated reference intervals are provided (taking into account age and/or sex when relevant), and incorporating estimates of analytical and biological variation as well as considering other variables specific to particular tumor markers and malignancies, such as tumor marker half-life *(23)* and kinetics *(24)*. Tumor marker results, together with all other observations made during patient care, can be contained in electronic health records *(11)* and used to determine the baseline tumor marker concentration for individual cancer patients during periods of remission, thereby facilitating earlier diagnosis

of progression. The electronic health record definition has matured nationally and internationally to a degree that enables such attributes to be documented in a common fashion *(25)*.

CLINICAL ISSUES THAT ENHANCE THE RELIABILITY AND UTILITY OF TUMOR MARKERS

Tumor markers are surrogate indicators that increase or decrease the clinician's suspicion that future clinically important events, such as cancer onset, recurrence, or progression or patient death, will or will not happen, and/or that a specific treatment will decrease the risk of such events. Markers can be used to determine risk, screen for early cancers, establish diagnosis, estimate prognosis, predict that a specific therapy will work, and/or monitor for disease recurrence or progression *(26)*. Tumor markers are valuable because they permit more efficient application of therapies, enabling their use to be tailored to those patients most likely to benefit while preventing their use, with its concomitant exposure to toxicities, in patients who would not benefit *(27)*.

Tumor markers are only useful if 3 circumstances pertain:

- The marker results are appropriate precisely for the required application, i.e., risk assessment, screening, diagnosis, prognosis, prediction, or posttreatment monitoring.
- The marker results separate patients into 2 or more populations whose outcomes differ so strikingly that healthcare providers would treat one group differently from another. This consideration depends on several factors, including the endpoint in question (patients might be more willing to accept therapy for very small mortality reductions but not for similar reductions in occurrence of a new cancer), the toxicity of the therapy (patients are more likely to accept a therapy with small benefits if the toxicities are few), and the cost of the therapy.
- The estimate of the separation in outcomes for marker positivity and negativity is reliable.

These issues are interrelated. For example, studies of the prognostic value of a marker that do not address the manner in which the study populations were treated are not helpful to the clinician trying to decide whether to apply treatment.

Furthermore, although statistical analysis is important for estimating the reliability of marker-identified differences between groups, the *P*-value alone does not indicate clinical utility. If a study is sufficiently powered, a small difference in outcomes of 2 groups separated by marker results (positive vs negative) might be statistically significant. Too often an investigator will conclude that a marker is clinically useful because a derived *P*-value is <0.05. More important for clinical utility, however, is that one population (marker positive or negative) does extremely well while the other does very poorly, so that one group might accept the therapy of interest while the other would elect not to. In such cases it is imperative that the *P*-value suggests statistical significance, but this criterion is not the determining factor for clinical utility. Finally, a single study does not establish a scientific fact. Rather, secondary validation of the results of an interesting study in a subsequent data set is imperative, and the validation study should use the same assay and the same cut point(s). In addition, the patient population must be very similar to that of the preceding study. These requirements are among those highlighted in the excellent reporting recommendations for tumor marker prognostic studies recently developed and published as the REMARK guidelines *(28)*, complementing previous broader statements on the Standards for Reporting of Diagnostic Accuracy *(29)*.

Studies leading to the acceptance of a tumor marker for clinical utility must be carefully and thoughtfully designed to ensure results that are meaningful in the clinical setting. Unfortunately, most tumor marker investigations have been studies of convenience, using archived samples that happened to be available *(26)*. Such studies, considered to be at level of evidence (LOE) III, are useful for generating hypotheses, but without follow-up investigations involving careful investigational planning and design the results cannot be accepted as fact. Indeed, LOE II studies, in which the marker is considered prospectively as a secondary objective in a clinical trial, or better yet, LOE I studies, in which the marker in question is the primary objective, are much more likely to yield clinically useful results. In other words, it is better to ask the question and get an answer than to get an answer and then ponder the question. Such evidence-based considerations are particularly important when patient lives are at stake and should be remembered whenever a tumor marker test is requested.

MICROARRAYS IN CANCER DIAGNOSTICS

BACKGROUND

Genomic microarrays, first introduced in 1996 by Affymetrix *(30)*, were developed through successful exploitation of principles first described in the early 1980s *(31)* and further developed by others [see e.g., *(32)*]. Briefly, Ekins' Ambient Analyte Theory recognizes that a minute amount of binding material (e.g., antibody) does not significantly change the sample concentration and can give much higher sensitivity than assay formats using 100 or 1000 times more binding material *(31, 33)*. Use of microscopic spots of a binding agent

located at high surface density on a solid support (together with very high specific activity labels) can yield higher sensitivity and shorter incubation times than conventional ligand assay methods, especially so-called noncompetitive assay methods *(34)*.

Although still generally restricted to research use, microarrays—biochips, DNA-chips, protein-chips, or cell-chips—have major potential clinical applications in oncology, enabling parallel and simultaneous analysis of complex systems and pathways and providing a complex snapshot of biological properties of the cell, tissue, organ, or fluid under consideration. Microarrays relevant to cancer diagnostics have been commercially introduced and/or are being developed (see Supplemental Table 7 in the online Data Supplement).

PRINCIPLES OF MICROARRAYS

A microarray is a compact device containing a large number of well-defined immobilized capture molecules (e.g., synthetic oligos, PCR products, proteins, or antibodies). The best-known microarrays, DNA-biochips, are miniature arrays of oligonucleotides attached to glass or plastic surfaces. These microarrays are used to examine gene activity (expression profiling) and identify gene mutations or single-nucleotide polymorphisms, by hybridization between the microarray sequences and a labeled probe (the sample of interest). There are 2 major methods for microarray fabrication: photolithography and mechanical deposition on materials such as glass slides or membranes *(35)*.

Major potential advantages of microarray-based assays include high sensitivity, requirement for only small amounts of binding reagents, independence of sample volume, decreased incubation times, minimal reagent wastage, simultaneous access to many genes or proteins, and potentially, quantification. More detailed information on the subject is readily available in many specialized books and reviews.

TISSUE MICROARRAYS

High-throughput analysis of tissues is facilitated by new technologies such as multitissue northern blots, protein arrays, and real-time PCR *(36)*. However, tissues are disintegrated before analysis, preventing identification of the cell types expressing the gene of interest. This limitation can be overcome by tissue microarrays, which consist of up to 1000 tiny cylindrical tissue samples assembled on a routine histology paraffin block, enabling simultaneous cost-effective analysis of up to 1000 tissue samples in a single experiment.

APPLICATIONS OF MICROARRAYS

Microarrays have been successfully applied in a variety of settings, including gene expression profiling, detection of single-nucleotide polymorphisms, sequencing by hybridization, protein expression profiling, protein-protein interaction studies, and whole-genome biology experiments.

Although the potential of microarrays has yet to be fully realized, they have shown great promise as tools for deciphering complex diseases, including cancer *(37)* (see Supplemental Table 7 in the online Data Supplement).

LIMITATIONS OF MICROARRAYS

Microarray technologies are still evolving, presenting difficulties for standardization because no gold standards are yet available to facilitate comparison of data between laboratories and platforms. Recent reports suggest that microarray data are not reproducible *(38)* and may be biased *(39)*.

KEY POINTS: MICROARRAYS IN CANCER DIAGNOSTICS

Microarrays will undoubtedly become routine diagnostic tools (see Supplemental Table 8 in the online Data Supplement), but first several variables must be optimized (e.g., capture molecules, hybridization protocols, data collection), QC systems must be established, and the appropriate level of analytical and clinical validation must be determined, as has been proposed *(40, 41)*.

Evaluation of microarray expression data requires careful consideration of the validity and accuracy of results for the biological system and whether the data fundamentally describe the phenomenon being investigated, together with the possibility of the presence of artifacts, which can occur at any time. Validation requires experimental QC, independent confirmation of data, and universality of results. Automation is also essential to minimize variability and increase robustness. Currently, for most studies on the clinical application of microarrays *(26)*, the LOE is only V. An exception to this is the 70-gene signature for predicting outcome in breast cancer *(42)*, a microarray application that has undergone both internal *(43)* and external validation *(44)* and was cleared by the FDA in 2007.

Based on the information above, the NACB Panel has formulated the recommendations outlined in Table 1.

MS IN CANCER DIAGNOSTICS

The principles of MS as applied to cancer diagnostics are summarized here, together with recommendations for the use of MS in clinical practice, focusing on MALDI and related MS techniques such as SELDI for proteomic analysis.

Table 1. NACB recommendations for use of microarrays in cancer diagnostics.

1. Gene expression microarrays are new and promising devices used for cancer diagnosis, prognosis, prediction of therapeutic response, and monitoring and selection of therapy. The level of evidence from most published studies, according to Hayes et al. (27) is Level V (lowest category). Consequently, microarrays should continue to be used as research devices, but not as tools for making clinical decisions.
2. Standardization and clinical validation of expression microarrays is warranted.
3. Quality control and quality assurance programs for expression microarrays need to be further developed.
4. Microarray automation is encouraged for improving reproducibility, throughput, and robustness.
5. Tissue microarrays are devices suitable for high-throughput analysis of large numbers of samples and are recommended for use in clinical trials and retrospective studies for evaluating and validating new tumor markers by immunohistochemical analysis.
6. Use of microarrays for single nucleotide polymorphism analysis is recommended for establishing haplotypes and for correlating these haplotypes to disease predisposition.
7. Use of microarrays is recommended for high-throughput genotyping and mutation/sequence variation detection for cancer diagnostics and pharmacogenomics. More validation is necessary to ensure equivalent results between standard technologies (such as DNA sequencing) and microarray analysis.
8. Protein microarrays and other similar technologies are recommended as research tools for multiparametric analysis of large numbers of proteins. The level of evidence is not yet high enough for clinical applications.
9. Standardized protocols should be developed for sample collection, handling, and processing.

PRINCIPLES OF DIAGNOSTIC MS

The typical mass spectrometer consists of an ion source, a mass analyzer that measures the mass-to-charge ratio (*m/z*) of the ionized analytes, and a detector that registers the number of ions at each *m/z* value. MS measurements are carried out in the gas phase of ionized species, often using electrospray ionization and MALDI to volatize and ionize the proteins or peptides. A variant of the latter is SELDI. The mass analyzer separates ionic species according to their *m/z* ratios. Four basic types of mass analyzers are commonly used in proteomic research: the ion trap, TOF, quadrupole, and Fourier transform ion cyclotron resonance; a potential fifth variant is the new Orbitrap™ mass spectrometer. These analyzers may be variously combined.

There are 2 approaches for biomarker discovery using MALDI/SELDI-TOF MS. One approach uses the differences between MS profiles of the disease and control specimens to generate a diagnostic model. A variation of this approach is to select several discriminate protein/peptide peaks and identify their nature. Diagnostics are based on multiplex immuno-MS or ELISA. The other approach is to degrade enzymatically the proteins to peptides, separate the peptides by techniques such as HPLC, and direct the eluted fractions into an ion source (electrospray ionization or MALDI), where they are converted into ionized species that enter the mass spectrometer, followed by algorithmic identification of the protein fragments comprising the mass spectra.

Protein identification is achieved through either peptide mass fingerprinting or peptide sequencing. In the former, peptide masses are compared with mass spectra of proteins listed in databases using appropriate software. Peptide sequencing is based on inducing random cleavage of peptide bonds between adjacent amino acid residues, e.g., by collision-induced dissociation, and determining the amino acid sequence of the resulting ion series.

APPLICATION OF MS IN CANCER DIAGNOSTICS

MS is particularly well suited to serve as a diagnostic or biomarker discovery tool in cancer, because during cancer development, cancer cells and/or the surrounding microenvironment generate proteins and peptides of different types and in different concentrations than normal cells. These abnormal tissue distributions can be analyzed by imaging-based MS and patterns compared with controls to identify cancer-specific changes that may be clinically useful. Should leakage into the circulation occur from the tumor-host microenvironment, cancer-specific analytes may also be detectable in the blood (Fig. 1). MS has been used to demonstrate many cancer-specific protein patterns, most often in blood and urine (see Supplemental Table 9 in the online Data Supplement).

Almost every published report of the profiles generated by MALDI-TOF MS has suggested that this method yields better diagnostic sensitivities and specificities than cancer biomarkers in current use, resulting in extensive publicity. However, initial enthusiasm has been tempered by other reports identifying potential problems with this approach and its clinical reliability *(45)*, highlighting the importance of rigorous validation.

If MALDI-TOF profiling is to transition successfully from research technique to clinical laboratory, all sources of variation—preanalytical, analytical, and postanalytical—must be understood and controlled *(45)*. The effects of sample storage and processing, sample type, patient selection, and demographic variables on test outcome must be clearly established, and analytical performance must be improved such that sensitivity, specificity, and dynamic range are comparable to those of established immunoassay techniques.

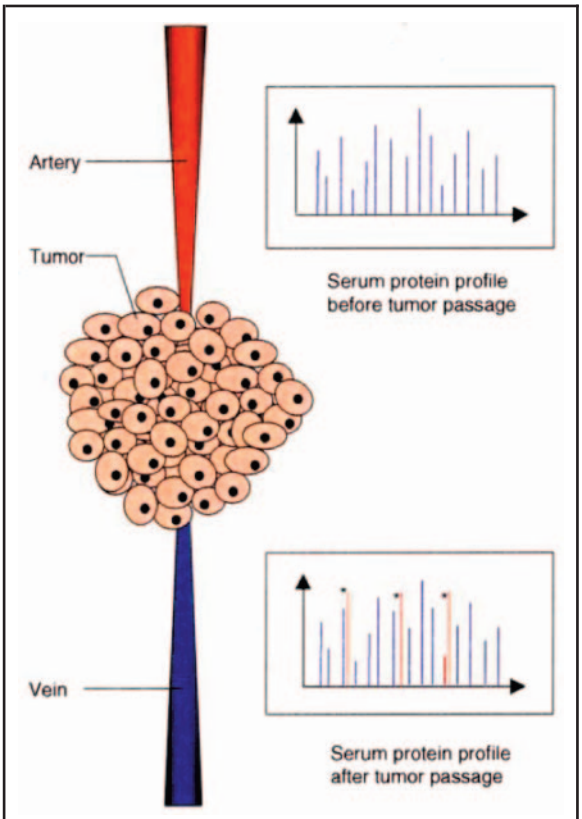

Fig. 1. Secretion of specific biomarkers into the blood circulation by tumors.

Tumor-specific proteins may be actively secreted by tumor cells or released into the circulatory system by necrosis and apoptosis of these cells. Either of these conditions leads to an alteration of the serum protein profile. When sera from normal and disease samples are compared, this alteration may result in differences detectable in relative and/or unique signal intensities. Reproduced from Pharmacogenomics 2003;4(4):463–76 *(49)* with the permission of Future Medicine Ltd.

Further investigations must address reproducibility of protein patterns across different batches of chips, different analysts, different sites, and different instrumentation. Robustness of the methodology is of concern, as are issues related to bioinformatic artifacts, data overfitting, and bias arising from experimental design. Some of these issues may relate to inappropriate use of publicly available mass spectral data sets, a consortium of investigators having recently succeeded in obtaining reproducible mass spectral signatures at multiple sites across time and instruments. This development is encouraging for those attempting to employ MALDI-TOF–type approaches for protein fingerprinting based diagnostics *(46)*.

Current limitations and promises of MALDI-TOF, particularly as applied to clinical practice and cancer diagnostics, are addressed more fully in several recently published reviews (for references see the original unabridged manuscript in the online Data Supplement).

KEY POINTS: MASS SPECTROMETRY PROFILING IN CANCER DIAGNOSTICS

Despite numerous publications describing diagnostic use of MALDI-TOF MS (see Supplemental Table 9 in the online Data Supplement), most are only LOE IV-V studies (i.e., either retrospective or small pilot studies) *(26)*. According to published criteria *(47)*, the stage of development of this technology is phase 1 (preclinical exploratory studies). NACB recommendations have been formulated on the basis of this information (Table 2). MALDI-TOF MS approaches are promising for biomarker discovery and validation, but before clinical application can be instituted the issues discussed here must be addressed. Advantages of proteomic profiling include analysis without the need for a labeling mole-

Table 2. NACB Recommendations for use of MALDI-TOF MS in Cancer Diagnostics

1.	MALDI-TOF MS profiling in the realm of cancer diagnostics should currently be considered an investigational and research tool that at this time, like all unvalidated methods, is insufficiently reliable to be the basis of clinical decisions.
2.	For MALDI-TOF MS profiling to become a clinically reliable tool, it must undergo validation according to principles such as those described by Pepe et al. *(47)* and avoid biases, as described by Ransohoff (39).
3.	For MALDI-TOF MS profiling testing, validation of discriminatory peaks in the mass spectra should include statistically powered independent testing and validation sets that include large numbers of inflammatory controls and samples from patients with benign disorders and from healthy controls as well as samples from patients with other cancers. The degree of statistical powering of the validation studies should be carried out under methods such as those described by Pepe et al *(47)* along with taking into consideration the intended clinical use of the test itself.
4.	Stability of bioinformatic algorithms should be evaluated using large numbers of samples, preferably from several institutions and countries.
5.	Standardized protocols should be developed for sample collection, handling, and processing.
6.	QC and reference materials for MALDI-TOF MS must be developed and used more widely to monitor and improve method reliability.
7.	Protein/peptide sequence identification or specific immune recognition of the analytes facilitates reproducibility, robustness, and overall biomarker validation.

cule, potentially high specificity, multiparametric analysis, high throughput, very low sample volume requirements, and direct interface with computer algorithms. Important limitations include variable cross-platform reliability of signatures generated, major vulnerability to minor changes in sample handling and processing, and poor analytical sensitivity, particularly when the analyte is present in minute amounts in a highly complex mixture containing high-abundance molecules. However, there are inherent advantages to certain MALDI-TOF approaches. For instance, combining immune isolation with MALDI-TOF analysis allows for elimination of secondary antibodies and detection of multiple derivative analytes such as protein isoforms. In addition, exciting new research suggests that many low-abundance proteins and low molecular weight analytes exist in bound states in serum and are effectively amplified by carrier-protein–based sequestration. These low molecular weight analytes appear to have underpinned many past spectral fingerprints, indicating that many of these ions may be generated from low-abundance analytes. A list of these low molecular weight carrier protein–bound analytes has recently been provided for early stage ovarian cancer patients *(48)*.

Until extensive validation studies are performed, MALDI-TOF MS fingerprinting approaches should not be used for cancer diagnosis. When MS fingerprinting is employed, independent validation sets including appropriate numbers of inflammatory, benign, and unaffected controls are essential, because specificity will determine success in the clinic. Encouraging developments include better appreciation of the influence of bias and variance and of the need for carefully defined operating procedures essential for validation of this technology and its introduction into clinical practice.

Grant/Funding Support: None declared.
Financial Disclosures: C. Sturgeon has over the past ten years received speaker expenses and/or honoraria from Abbott Diagnostics, Bayer Diagnostics, Beckman Coulter, Becton Dickinson, Dako, DPC, Roche, Tosoh Biosciences, and Wallac Oy for lectures on quality aspects of tumor markers (with no product endorsements). D. Hayes has received expenses and/or honoraria as a consultant, advisory panel participant, or lecturer/speaker during the last 12 months from Abraxis, American Biosciences, AviaraDx, Cytogen, Monogram Bioscience, Pfizer, Precision Therapeutics, Predictive Biosciences, Siemens Medical Solutions Diagnostics, and StemCapter and during the last twelve months has been principal or coinvestigator of clinical research sponsored by GlaxoSmithKline, Pfizer, Novartis, and Wyeth Ayerst-Genetics Institute. D. Chan has received research funding as an investigator for Abbott Labs and Beckman Coulter, fees as a consultant for Tosoh Biosciences, and fees as a scientific advisory board member for Ciphergen and Roche Diagnostics. L. Liotta is a board member for and holds stocks in Theranostics but receives no honoraria or salary. E. Petricoin is an inventor and owner of US Government and university-assigned patent applications that cover aspects of the technologies discussed in this manuscript. M. Schmitt is a research collaborator with American Diagnostica and Epigenomics AG.
Acknowledgments: We thank the numerous scientists and clinicians who have contributed to this undertaking, especially the invited reviewers listed in the Appendix, Mrs. Hassima Omar Ali for her excellent assistance, David Bruns for agreeing to consider publishing these guidelines in *Clinical Chemistry* and for his highly valued advice, and of course the NACB and the AACC for their much appreciated support and encouragement.

References

1. Fleisher M, Dnistrian A, Sturgeon C, Lamerz R, Witliff J. Practice guidelines and recommendations for use of tumor markers in the clinic. Tumor markers: physiology, pathobiology, technology and clinical applications. Washington (DC): AACC Press; 2002:33–63.
2. WHO. Use of anticoagulants in diagnostic laboratory investigations. Available from: WHO/DIL/LAB/99.1 Rev. 2. http://www.eoc.ch/allegati/doc_preanalitica_eolab.pdf.
3. Pariente A, Prevost J, Montaut N, Quesnel-Tueux N, Boulet JM, Gasnier Y, et al. [Improvement of prescriptions for serum tumor markers in a general hospital]. Presse Med 1998;27:153–6.
4. Durieux P, Ravaud P, Porcher R, Fulla Y, Manet CS, Chaussade S. Long-term impact of a restrictive laboratory test ordering form on tumor marker prescriptions. Int J Technol Assess Health Care 2003;19:106–13.
5. Basuyau JP, Leroy M, Brunelle P. Determination of tumor markers in serum: pitfalls and good practice. Clin Chem Lab Med 2001;39:1227–33.
6. Bonini P, Plebani M, Ceriotti F, Rubboli F. Errors in laboratory medicine. Clin Chem 2002;48:691–8.
7. Boone DJ, Steindel SD, Herron R, Howanitz PJ, Bachner P, Meier F, et al. Transfusion medicine monitoring practices: a study of the College of American Pathologists/Centers for Disease Control and Prevention Outcomes Working Group. Arch Pathol Lab Med 1995;119:999–1006.
8. NACB: Tumor markers 2006. Draft guidelines: second posting. Available from: http://www.aacc.org/AACC/members/nacb/LMPG/OnlineGuide/DraftGuidelines/TumorMarkers/.
9. Loi S, Haydon AM, Shapiro J, Schwarz MA, Schneider HG. Towards evidence-based use of serum tumour marker requests: an audit of use in a tertiary hospital. Intern Med J 2004;34:545–50.
10. McGinley PJ, Kilpatrick ES. Tumour markers: their use and misuse by clinicians. Ann Clin Biochem 2003;40:643–7.
11. Bates DW, Gawande AA. Improving safety with information technology. N Engl J Med 2003;348:2526–34.
12. Young DS, Friedman RB, Narayanan S. Young's effects online. Available from: http://www.fxol.org/aaccweb/.
13. Roddam AW, Price CP, Allen NE, Ward AM. Assessing the clinical impact of prostate-specific antigen assay variability and nonequimolarity: a simulation study based on the population of the United Kingdom. Clin Chem 2004;50:1012–6.
14. Sturgeon CM,. Limitations of assay techniques for tumor markers. In: Tumor markers: physiology, pathobiology, technology, and clinical applications. Washington (DC): AACC Press; 2002:65–81.
15. Seth J, Sturgeon CM, Al-Sadie R, Hanning I, Ellis AR. External quality assessment of immunoassays of peptide hormones and tumour markers: principles and practice. Ann Ist Super Sanita 1991;27:443–52.

16. Sturgeon C, Ellis A, Al-Sadie R. Annual review for 2007. Available from: www.ukneqas.org.uk.
17. Sturgeon CM, Seth J. Why do immunoassays for tumour markers give differing results? A view from the UK National External Quality Assessment Schemes. Eur J Clin Chem Clin Biochem 1996;34:755–9.
18. White P. Annual review for PSA. Available from: http://www.immqas.org.uk/.
19. Bristow A, Berger P, Bidart JM, Birken S, Norman R, Stenman UH, Sturgeon C. Establishment, value assignment, and characterization of new WHO reference reagents for six molecular forms of human chorionic gonadotropin. Clin Chem 2005; 51:177–82.
20. Rafferty B, Rigsby P, Rose M, Stamey T, Gaines DR. Reference reagents for prostate-specific antigen (PSA): establishment of the first international standards for free PSA and PSA (90:10). Clin Chem 2000;46:1310–7.
21. Sturgeon CM, Ellis AR. Standardization of FSH, LH and hCG-Current position and future prospects. Mol Cell Endocrinol 2007;260:301–9.
22. Berger P, Sturgeon C, Bidart JM, Paus E, Gerth R, Niang M, et al. The ISOBM TD-7 Workshop on hCG and related molecules: towards user-oriented standardization of pregnancy and tumor diagnosis: assignment of epitopes to the three-dimensional structure of diagnostically and commercially relevant monoclonal antibodies directed against human chorionic gonadotropin and derivatives. Tumour Biol 2002;23:1–38.
23. Vogl M, Muller MM. Tumor markers: review and clinical application. Milan (Italy): IFCC; 2002.
24. Riedinger JM, Wafflart J, Ricolleau G, Eche N, Larbre H, Basuyau JP, et al. CA 125 half-life and CA 125 nadir during induction chemotherapy are independent predictors of epithelial ovarian cancer outcome: results of a French multicentric study. Ann Oncol 2006;17:1234–8.
25. Greenes RA. Clinical decision support: the road ahead. Burlington (MA): Elsevier Academic Press; 2007.
26. Hayes DF, Bast RC, Desch CE, Fritsche H Jr, Kemeny NE, Jessup JM, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst 1996;88:1456–66.
27. Hayes DF. Prognostic and predictive factors for breast cancer: translating technology to oncology. J Clin Oncol 2005;23:1596–7.
28. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 2005;97:1180–4.
29. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Clin Chem 2003; 49:1–6.
30. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 1996;14:1675–80.
31. Ekins RP. Multi-analyte immunoassay. J Pharm Biomed Anal 1989;7:155–68.
32. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (Wash DC) 1995;270:467–70.
33. Joos TO, Berger H. The long and difficult road to the diagnostic market: protein microarrays. Drug Discov Today 2006;11:959–61.
34. Ekins R, Chu F. Ultrasensitive microarray-based ligand assay technology. In: Albala JS, Humphery-Smith I, eds. Protein arrays, biochips, and proteomics: the next phase of genomic discovery. New York (NY): Marcel Dekker; 2003:81–125.
35. Finckh P, Berger H, Karl J, Eichenlaub U, Weindel K, Hornauer H. Microspot, an ultrasensitive microarray-based ligand assay system. A practical application of ambient analyte assay theory. In: Sturgeon CM, Seth J, Middle JG, Halloran SP, eds. Proc UK NEQAS Endocrinology Meeting (1998), vol. 3. Edinburgh: Association of Clinical Biochemists; 1998:155–65.
36. Kallioniemi OP. Biochip technologies in cancer research. Ann Med 2001;33:142–7.
37. Chung CH, Bernard PS, Perou CM. Molecular portraits and the family tree of cancer. Nat Genet 2002;32 Suppl:533–40.
38. Marshall E. Getting the noise out of gene arrays. Science (Wash DC) 2004;306:630–1.
39. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer 2005;5:142–9.
40. Burgoon LD. The need for standards, not guidelines, in biological data reporting and sharing. Nat Biotechnol 2006;24:1369–73.
41. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 2007;99:147–57.
42. van't Veer LJ, Dai H, van dV, M.J., He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature (Lond) 2002;415:530–6.
43. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
44. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 2006;98:1183–92.
45. Diamandis EP. Point: proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? Clin Chem 2003;49:1272–5.
46. Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer; I, assessment of platform reproducibility. Clin Chem 2005;51: 102–12.
47. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–61.
48. Zhou M, Lucas DA, Chan KC, Issaq HJ, Petricoin EF, 3rd, Liotta LA, et al. An investigation into the human serum "interactome". Electrophoresis 2004;25:1289–98.
49. Pusch W, Flocco MT, Leung SM, Thiele H, Kostrzewa M. Mass spectrometry-based clinical proteomics. Pharmacogenomics 2003;4:463–76.

Appendix

NACB Subcommittee members: Quality Requirements: Catharine M. Sturgeon, Chair; Soo-Ling Ch'ng, Elizabeth Hammond, Daniel F. Hayes, and György Sölétormos; Microarrays: Eleftherios P. Diamandis, Chair; Manfred Schmitt and Elena van der Merwe; Mass spectrometry: Daniel W. Chan, Chair; Eleftherios P. Diamandis, Lance A. Liotta, Emmanuel F. Petricoin; O. John Semmes, and Elena van der Merwe.

NACB Quality Requirements Subcommittee Members: Catharine M Sturgeon, Chair; Soo-Ling Ch'ng, Elizabeth Hammond, Daniel F. Hayes, and György Sölétormos. All comments received about the NACB Recommendations for Quality Requirements are included in the online Data Supplement. Dr Jean-Pierre Basayau, Professor Per Hyltoft Petersen, Professor Mathias Müller, and Professor Hans Schneider were invited expert reviewers.

NACB Microarray Subcommittee members: Eleftherios P. Diamandis, Chair; Manfred Schmitt and Da-elene van der Merwe. All comments received about the NACB Recommendations for Microarrays in Cancer Diagnostics are included in the online Data Supplement. Professor Roger Ekins was an invited expert reviewer.

NACB Mass Spectrometry Subcommittee members: Daniel W. Chan, Chair; O. John Semmes, Emmanuel F. Petricoin, Lance A. Liotta, Da-elene van der Merwe, and Eleftherios P. Diamandis. All comments received about the NACB Recommendations for Mass Spectrometry in Cancer Diagnostics are included in the on-line supplement. Professor William T. Morgan and Professor Roz Banks were invited expert reviewers.