

2

Analytical Aspects of Biomarker Immunoassays in Cancer Research

*Fred C.G.J. Sweep, PhD,
Chris M.G. Thomas, PhD,
and Manfred Schmitt, PhD*

CONTENTS

INTRODUCTION
IMMUNOASSAY DESIGNS
VARIABILITY IN TEST RESULTS
ASPECTS OF BIOMARKER ASSESSMENT
QUALITY ASSURANCE
CONCLUSIONS
REFERENCES

SUMMARY

Many difficulties associated with immuno(metric) assay kits designed for quantification of a particular biomarker arise from their variation in specificity and binding affinity of the employed antibodies. Other important sources causing varying assay results are the use of different standard preparations in these kits and the nonuniform preanalytical specimen processing procedures employed, each of which should be subjected to standardization. To improve the performance and comparability of assays, continuous interlaboratory external quality control procedures are needed. Such quality assurance

*Cancer Drug Discovery and Development: Biomarkers in Breast Cancer:
Molecular Diagnostics for Predicting and Monitoring Therapeutic Effect*

Edited by: G. Gasparini and D. F. Hayes © Humana Press Inc., Totowa, NJ

programs provide a forum for expert laboratory investigators to discuss technical details and to exchange laboratory issues and related practical information. This chapter addresses some of these issues and presents initial analytical validation procedures of newly developed biomarker assays, the validation of already established assay procedures for routine use on a day-to-day basis, and finally discusses some aspects on adequate (external) quality control proficiency testing.

Key Words: Biomarkers; immunoassay; cancer; tumor markers.

1. INTRODUCTION

In many cases progression of cancer growth is rather slow and often it may take years for a malignancy to manifest clinically. Because early cancer detection is required to significantly reduce cancer mortality, screening procedures are needed that are highly specific (i.e., providing almost a 100% proportion of negative test results for a tumour marker in nondiseased individuals) and sensitive enough to detect malignancies at an early stage of development. Thus, the screening procedure should give assay results above a defined cutoff value in a reasonable proportion of early stage diseased persons. As yet, there are no assay procedures available that meet such a specification, although there is a growing public interest in improving early cancer detection. Ideally, determination in biological specimens of cancer-derived analytes for a particular type of cancer not only should provide valuable information for initial diagnosis, but also should have prognostic value to guide the choice of treatment, and such a test should provide a reflection of the tumor burden of the patient, being predictive for recurrent disease after initial treatment, and of help in monitoring the course of the disease throughout time of follow-up. Each of these properties should contribute to more effective treatment of an individual patient and thus provide indispensable information for improving the quality of life and outcome of the disease by increasing disease-free and overall survival. Despite extensive research efforts in the last decades and numerous papers dealing with development and clinical testing of potentially promising biochemical markers, no assays are as yet available that are sensitive enough to convincingly detect any of the major types of cancer at the most early stage. Although an impressive number of biochemical markers with the capacity to predict disease recurrence and/or early death have been introduced, a comprehensive understanding of the tumor biological processes involved is still lacking.

1.1. Guidelines for Evaluating Clinical Value of Biomarkers

At present, cancer diagnosis is based mainly on clinical symptoms and confirmed by histomorphological findings. Application of biochemical markers in this process may have additional value but still, depending on the

marker test applied, the reliability criteria may become less important. In case of screening and diagnosis (“rule-in-disease”) specificity is of utmost importance (to avoid false-positive assay results leading to unfavorable and unnecessary medical examination and treatment), although it should be realized that increasing specificity of a test goes at the cost of decreasing its sensitivity (that should remain high enough to detect early-stage diseased individuals). If the purpose of the test is disease monitoring to detect recurrence during follow-up, precision should be high; providing a prognosis for treatment, the test should put emphasis on specificity and accuracy. All of these criteria are not well established and should become standard criteria for evaluation of biomarker assays and their clinical application. In line with this, Hayes et al. (1) proposed certain criteria to standardize the available biomarker information for clinical use in a biomarker utility grading system.

Currently, many biochemical markers of potentially prognostic value are intensively tested in multicenter clinical trials. Only a few of these show a benefit for predicting prognosis of node-negative breast cancer patients. To conclude that these newly developed biochemical markers have independent prognostic value over already known factors, McGuire and Clark (2) some years ago proposed strict guidelines for evaluating newly developed prognostic markers, addressing the biological role of the new factor as well as the extent of the sample size, the risk of sample bias, the appropriate testing system, the establishment of cutoff values in a training data set, and confirmation of these observations in an independent validation data set.

2. IMMUNOASSAY DESIGNS

Immunochemical assay procedures can be classified according to the kind of analysis (qualitative, semiquantitative, or quantitative), type of assay format (manual or automated) and assay system (liquid phase, solid–liquid phase, [non-]equilibrium), making use of (radioisotopic or nonisotopic) labeled markers (to detect the antigen–antibody complex) or nonlabeled markers (in which the antigen–antibody complex is detected without labeled markers). The term immunoassay refers to competitive methods while immunometric assays refer to noncompetitive, sandwich-type assay formats. As early as 1969 the first generation of binding assays emerged with development of a radioimmunoassay (RIA) for quantification of insulin antibody formation. Later, the evolution of technical developments led to nonisotopic labels (enzyme-, fluorescence, time-resolved fluoro-, (chemo-) luminescence immunoassays, etc.), monoclonal antibodies, phase matrices, and two-site immunometric sandwich-type assay formats employing two or more antibodies that will bind the analyte at repetitive or different epitope binding sites. The enzyme-linked immunosorbent assay (ELISA) format is

a commonly used type of two-site sandwich type assay in clinical routine work. The analyte is allowed to react noncompetitively with an excess of immobilized (“capture”) antibody (coupled to a solid phase) and after addition and washing off the excess amount of sample specimen, an excess amount of marker labeled (“signal”) antibody is added to bind to another epitope of the analyte. The sandwich thus formed is provided with marker label proportional to the amount of analyte present in the sample. These quantitative assay formats have several advantages: large numbers of specimens can be processed in parallel, providing reproducible results with reasonable precision, sensitivity, specificity, and accuracy. A major advantage of immunometric assay formats over semiquantitative or qualitative techniques is provided by the quantitative endpoint as measured against a defined standard, although there are limitations. Often, the analyte standard is not well defined, the assay procedure is not fully validated prior to use in patient studies, or the possibility to make comparisons to a reference method is lacking. Strict measures of quality assurance or good manufacturing practice protocols are needed to ensure proper assay performance before they should be applied.

Because early detection of small breast lesions is becoming common practice, there is an increasing demand to measure correctly biomarkers in smaller pieces of tumor specimens obtained through fine needle aspiration, core biopsies, or cryostat sections. This implies that there is an immediate need for more sensitive techniques than the standard immunoassays available to date. Alternative approaches to ELISA are proteomic methods such as MALDI and SELDI TOF mass spectrometry (MS), tandem MS, plasma resonance techniques, and antibody chip technologies. Of course the same rigorous principles of quality assurance should be applied for these new methodologies as for the more conventional immunoassays.

3. VARIABILITY IN TEST RESULTS

Assay results are often heterogeneous because of variations in specimen composition, tissue processing, design and specificity of the employed assay, as well as the statistics used for analyzing the collected data. In each of these stages, intrinsic differences in molecular forms (isoforms) of the biomarker present in the tumor tissue are augmented by external causes. The sampling procedure (e.g., fine-needle aspirate, core biopsy, or large biopsy obtained during surgery), the source of tissue (fresh or frozen), storage conditions (time, temperature, freeze–thawing cycles, etc.), and tissue processing (cytosol fraction, membrane extracts) may severely influence the final assay results (3). Likewise, this also holds true for the quantification of biological markers in serum or plasma (4).

Variable design of immuno(metric) assays results in the generation of different test results because different kits incorporate a broad spectrum of

antibodies, sometimes with different antibody specificities and/or affinities. Also, the use of different standards and reference materials provided with the kits are a source of variations in test results. Furthermore, different data reduction processes and statistical techniques are used to analyze tumor marker data and this may lead to a variety of conclusions regarding the clinical interpretation. The computational data processing of laboratory results must be appropriate, uniform, and evaluated extensively (5). McGuire and Clark stated that the design of confirmatory clinical studies should be identical to that of the definite study (2). It is of most importance to note that this also applies to all laboratory steps including tissue storage and processing, the analytical procedures, and the subsequent data processing.

The number and diversity of biomarkers for assessment of cancer prognosis is expanding rapidly, as is the variety of analytical formats and procedures used for quantification. A substantial proportion of assays is based on immunochemical principles and there is a widespread use of nonvalidated assay formats in clinical research settings. Because many assays are poorly standardized and (external) quality control is lacking in most cases, nonvalidated assay results without provided certified guidelines for interpretation become available at a too preliminary stage of assay development. Thus, biomarker testing procedures in laboratories participating in clinical trials should be standardized and externally quality assessed. This requires settlement of quality standards of all assay reagents included in assay kits, provision of guidelines for standardized assay protocols, standardized algorithms for calculation of assay results, and statistical procedures to allow unequivocal interpretation of clinical effect measures. Finally, to ascertain continuity of reliable biomarker data generation, there is a need for guidelines toward uniform internal and external quality assessment procedures. The next sections will discuss preanalytical, analytical, and postanalytical aspects of assay performance.

4. ASPECTS OF BIOMARKER ASSESSMENT

4.1. Preanalytical Criteria

4.1.1. SAMPLING BIAS OF TISSUE SPECIMENS AND TISSUE PROCESSING

Because many tumors are heterogeneous the size of a tumor tissue specimen is important to avoid sampling bias. This bias may lead to different assay results if different areas of a tumor are analyzed (different content of tumor cells, nonmalignant cells, extracellular matrix, fat, and necrotic spots). Thus, fine needle biopsy results may differ from those obtained from a tumor tissue biopsy specimen. Selection bias may occur if frozen tissue specimens from large tumor banks are used in retrospective studies as generally in tumor banks relatively larger samples of frozen tumor tissues are overrepresented (6).

The use of blood specimens requires standardization of blood collection conditions (fasting, fixed time of day, supine position), type of specimen (whole blood, serum, or plasma) and type of anticoagulant. Care should be taken to immediately transport tissue specimens or blood directly after surgery or blood collection to the laboratory in a standardized manner (time, temperature). Disintegration or extraction procedures of tissue samples should be performed according to the consensus protocols written by internationally acknowledged experts. Errors in this preanalytical phase of biomarker level quantification will affect the reliability of the final experimental data.

4.2. Analytical and Reliability Criteria

Prior to producing and subsequent reporting of test results it is the task of the laboratory to verify or establish performance specifications for each analytical procedure, irrespective whether the assay of interest has been developed in an academic institution or by a commercial company. In their instructions for use, kit manufacturers have often included disclaimers for misusing or overinterpreting the information included in their product information. It is common practice of diagnostic kit manufacturers to advise their clients that each laboratory should establish its own reference values in particular for specified populations or applications, irrespective of already available data provided by the manufacturer. The next sections deal with reliability criteria of analytic testing systems.

4.2.1. STANDARD CALIBRATION PREPARATIONS

Standards are used to prepare a standard dose response curve that relates the response reading as the independent variable to the quantity of the standard as the dependent variable. This allows calculation of the quantity of analyte from the response reading obtained for the unknown sample. It is not always possible to obtain sufficient quantities of a reasonably pure biomarker for characterization, which is the reason why in many cases arbitrary nonpurified or semipurified preparations of biomarkers are used to produce standard curves. Protein analytes may be present in different molecular forms (“isoforms”) which may cause differences in affinity or other binding characteristics with antibodies. In case there are differences in affinity of the antibody for the calibrator standard and the analyte present in the unknown sample, different assay results will be obtained at different sample dilutions. For this reason we propose to analyze biological markers in at least two or three different dilutions to detect this phenomenon. This means that the suitability of a biomarker assay has to be validated for each biological specimen of concern because the procedure for the measurement of an analyte in tissue extracts is not always suitable for assaying the same analyte in plasma or serum. Stability of the standard can best be followed by longitudinal monitoring of the consecutively produced slopes of the

standard dose–response curves. Thus, assays should use well-defined, well-characterized standard calibrator material with known sequence and degree of purity. Also, different kit manufacturers should adhere to internationally accepted standards and preferably use identical standards in their diagnostic kits. An important source of providing biological reference materials to the scientific community covering many areas of clinical medicine is the WHO International Laboratory for Biological Standards (National Institute for Biological Standards and Control [NIBSC], Potters Bar, UK). Finally, as an example of advancements that contribute to standardization of widely used biomarker assays, we mention the introduction of an assay procedure for prostate-specific antigen (PSA) that determines several molecular forms of PSA on an equimolar basis, and the calibration of this assay with the Stanford 90:10 Reference Material, composed of 90% PSA-ACT and 10% free-PSA (7).

4.2.2. ACCURACY

Definition of accuracy of an assay by the International Federation of Clinical Chemistry (IFCC) is the agreement between the best estimate of a quantity and its true value. As this quantity has no numerical value, the term inaccuracy is used. Thus, inaccuracy is the difference between the mean of a set of replicate measurements and the true value. Although the concept is clear, it has realistic value for those analytes for which a reference method is available. As no such reference values are available or even feasible for many biomarkers, the concept of (in)accuracy has limited significance, emphasizing even more the necessity of standardization of assays.

4.2.2.1. Linearity

As outlined earlier, linearity of an assay in fact refers to identity between affinity of the antibody for the calibrator standard and the analyte present in the unknown sample. This should be the case and these tests of parallelism between standard and unknown analyte can be conducted by measuring samples at different sample dilutions and multiplying the amount of analyte measured with the dilution factor. Linearity studies are used to assess and establish the working range of an assay that is in between the lowest and highest concentration that can reliably be measured with that assay. This can easily be realized by mixing two different samples in several proportions (e.g., 1:3, 1:1, 3:1). See also the National Committee for Clinical Laboratory Standards (NCCLS) evaluation protocol (EP6).

4.2.2.2. Recovery

Recovery experiments are conducted to test whether the standard of the assay and analyte in the unknown sample behave chemically identical, or to exclude whether disturbing interactions of the analyte with the matrix or other compounds of the assay will lead to different assay results. Thus, in order to obtain insight into the identity of the analyte vs the standard, or

to study matrix interactions with the standard, known amounts of standard are added to samples with an already known amount of endogenous biomarker and the recovery of the added amount is calculated.

4.2.2.3. High-Dose Hook Effect

This phenomenon is a source of error specifically occurring in double determinant one-step sandwich-type assays and comprises the saturation of capture and/or signal antibodies resulting from extremely high concentrations of biomarker analyte present in the incubation medium. This leads to a falsely low concentration calculated for the analyte. High-dose hook effects can be avoided by conducting a two-step assay protocol in which the immobilized capture antibody is incubated with an appropriately diluted unknown sample and excess of unbound analyte is washed off. The assay is completed by addition of signal antibody in the second incubation step. It is also advised to analyze samples at different dilutions to check whether the assay is vulnerable for the high-dose hook effect.

4.2.2.4. Interferences

Heterophilic antibodies are an often underestimated source of error in immunometric assays. In particular the treatment of patients with monoclonal mouse antibodies for immune-imaging and immune-targeting purposes has emerged occurrence of human antimouse antibodies (HAMAs), that is, the generation of human immune globulins G and M (IgG, IgM) in the blood of these patients. These antimouse IgG or IgM may also originate from other iatrogenic animal sources, all of these interfering to variable extents with the antibodies incorporated in biomarker sandwich-type assays. For a review of HAMA occurrence and its consequences for assay methodology see Kricka (8).

4.2.3. SPECIFICITY

In epidemiological terms specificity refers to the proportion of true-negative test results of a control population and in fact is similar to its definition in analytical terms where it is defined as (absence of) interference (cross-reaction) in an assay system of compounds more or less related to the analyte to be measured in that assay. Thus, specificity of immunoassays refers to the degree of interference by compounds that may resemble but differ from the analyte to be quantified. One established manner to express cross-reaction is comparison of the amount of analyte homologous for the assay with the amount of another compound tested for interference with the assay. This is performed at half the maximum response level (often referred to as $B/B_0 = 0.5$) of the linearized standard dose-response curve. The specificity of immunometric assays strongly depends on antibody characteristics because polyclonal or monoclonal antibodies or mixtures of both are applied in different testing kits. Specificity will be highest with monoclonal anti-

bodies because these are directed against one epitope on the analyte molecule. Many tumor-associated antigens have epitopes also common to other proteins present in a variety of many other tissues. Because epitope mapping data of antibodies is not often documented, investigators have to check cross reactivity of a number of compounds related structurally or biologically to the assay's analyte.

4.2.4. SENSITIVITY

In epidemiological terms, sensitivity refers to the proportion of true-positive test results of a diseased population. Analytically, sensitivity may be defined as the limit of detection of the analyte in the assay, that is, the lowest concentration of analyte significantly different from zero, also called the analytical sensitivity. The limit of quantification at which a test can be reliably measured with a coefficient of variation of less than 20% is called the functional sensitivity. It is recommended to report clinical assay results not below the functional sensitivity limit that can easily be retrieved from the precision profile of an assay that is constructed by plotting the coefficients of variation of replicate measurements of all the samples assayed against the concentrations of the obtained results. Data on sensitivity should be provided by the kit manufacturer and checked by the investigator on first use of a kit. One of the goals of immunoassay methodology is to optimize continuously the lower detection limits of assays in order to settle clinically relevant cutoff points. Defining such low thresholds requires a high degree of reproducibility of assay results, that is, precision.

4.2.5. PRECISION

According to the IFCC, precision is defined by the agreement between replicate measurements. As is also the case with accuracy, precision has no numerical value, the reason why the use of imprecision is more practical, although not commonly used. The imprecision is the standard deviation or coefficient of variation of the results of a set of replicate measurements.

The precision of a biomarker determination varies depending on whether duplicate determinations are performed in one sample, different samples in the same batch, or in different batches, and so forth. Obviously, the estimate of the precision used to assess the validity of experimental results must be related to the assay conditions in the definite study. For instance, if the concentration of a biomarker in malignant vs nonmalignant tissue of the same patient is determined in one assay run, the statistical significance of relevant difference is referred to the intraassay precision of the method. On the other hand, when a marker is monitored over a long time of observation (follow-up), samples will be assayed in different batches of test kits and the interassay precision is the more relevant parameter. For validation of an assay, at least the intrasample, intraassay precision performance should

be investigated. The precision profile is an ideal tool to assess this (see below). The NCCLS offers a practical evaluation protocol (EP5) for evaluating the precision performance of an assay.

4.2.6. MINIMAL CONSISTENCY CRITERIA

Apart from the aforementioned assay characteristics that should be assessed by the investigator once a new kit is introduced into the laboratory, assay performance may be hampered by day-to-day, performer-to-performer, and batch-to-batch variability. Run-to-run performance errors may be reduced by daily consistency testing of the calibration curve, the precision profile, and data on quality control specimens. Charting of standard dose–response curve characteristics comprises at least the calibrated slope; y-intercept, correlation coefficient, analyte concentration at 50% response (ED_{50}), and minimum detectable analyte concentration. The shape of this curve defines the quality of the performed assay and offers a basis for selection of the working range of the assay, while it also quite easily allows to detect unreliably scattering duplicates.

4.3. Postanalytical Criteria

Once an assay has been performed the results of unknown samples must be derived from the obtained response parameters by calculating the analyte concentrations from the standard dose–response curve. Numerous computerized algorithms are available, but irrespective of the choices made, it is highly advisable to use the same statistical approach to process assay data, especially if one participates in or conducts a multicenter study. Each laboratory should establish its own reference values to circumvent population sampling errors and biological variation.

5. QUALITY ASSURANCE (QA)

Defined protocols for (internal and external) quality control (QC) should be part of routine practice in the laboratory. QA not only comprises the analytical process as such (QC), but it also regards the total of the managerial, technical, and interpretative aspects and is intended to prevent, monitor, and correct mistakes in the laboratory chain process. Reasonable quality management requires knowledge about the level of quality that is needed. It is useless to implement and adhere to too strict control rules because this may cause unnecessary, false rejection of assay runs. Ideally, an adequate control procedure should be based on a definition of quality requirements weighing acceptable error against needed clinical decision levels.

5.1. Internal and External QC

Every biomarker assay should include control sample procedures to check the validity of the unknown sample results. Control samples and compari-

son of their results against control limits should always be integral part of a complete assay procedure.

5.1.1. QC SAMPLES

QC samples are stabilized specimens and available in liquid or lyophilized form because freshly collected sample materials are not always available and unstable for long-term QC use. Important requirements of QC preparations are that they should be time and temperature resistant with little or no vial-to-vial variation, homogeneous, similar in matrix structure to the test material, available at concentrations that cover the physiological range expected in the experimental material, and available in sufficient quantities. Unfortunately these requirements are not always achieved. For serum assays, large pools of serum can be established, aliquoted, and made available to laboratories. However, many manufacturers nowadays supply reference samples on a non-serum-based matrix, and in some cases this yields assay results different from those of true native serum samples. Thus, control samples should resemble as close as possible the analyte fractions representative of those routinely encountered in patient specimens.

5.1.2. MONITORING OF DAILY PERFORMANCE

At least two samples of different concentrations of control material should be included in each assay run to make multirule/decision control procedures possible, for example, by applying Westgard evaluation rules for internal QC (IQC) (9). Thus, repeated measurement of control samples allows to determine imprecision of the assay system. In addition to the use of IQC for day-to-day assay monitoring, the long-term trend in assay performance should be regularly checked in order to detect any shift or drift. Obviously, there should be agreed criteria for batch rejection. Levey–Jennings charts (10) are practical tools to evaluate the controls simply by plotting the individual values on a chart and compare these with a predefined mean with signaling limits (e.g., ± 2 SD). The chart patterns bring different kinds of technical problems (random error, systematic error, etc.) to light, and are also useful, simple tools for investigators or supervisors to decide whether or not assay results are within (or beyond) acceptable ranges and whether the data can be reported. Lot-to-lot variation errors of commercial reagents can be reduced by prescreening of critical reagents and be rejected before use if not consistent.

For external QC (EQC) purposes, preparations distributed by a reference laboratory should be included in assay runs if available. In proper EQC programs, the obtained data of control samples should be submitted to an external organization for statistical evaluation. These programs serve to monitor long-term assay performance within each participating laboratory. Moreover, they provide comparison of assay results between laboratories

and between different assay designs or brands, if available. This enables the organization to assess systematic errors between laboratories just by comparing the reported mean values of the individual laboratories with the mean of the total or reference group (all laboratory trimmed mean).

5.1.3. EXTERNAL QUALITY ASSESSMENT (EQA)

DOES NOT COVER ALL PROCESS STEPS

EQA based on lyophilized tumor tissue extracts or blood specimens does not allow any conclusion with regard to preanalytical, methodological issues such as variation in tissue collection, transport from operating theater to the laboratory, sample storage conditions, homogenization of tissue, and extraction procedures, as the use of external controls covers only reproducibility of the analytical assay procedure and subsequent computation of data. Providing proper instructions and careful observation of the results obtained is the only feasible way to monitor (between-hospital) variations in sample treatment conditions. Because most clinical trials are carried out on a multicenter basis, the interlaboratory QC is very important but the obtained deviations are most probably underestimations of true differences. Therefore, all steps in the procedure from taking biopsies to reporting assay results to the clinician including the preanalytical items should be conducted according to strict protocol guidelines.

5.2. Normalization of Assay Results

Long-term QA trials on steroid hormone receptors (estrogen receptors [ERs] and progesterin receptors [PgRs]) assays by the Receptor and Biomarker Group of the European Organisation for Research and Treatment of Cancer (RBG EORTC) have shown that even highly experienced laboratories, with excellent intralaboratory between-run performance, can have difficulties in directly comparing their results with those of another institution. As variation among laboratories in general appeared to be not random (11), a high interlaboratory coefficient of variation (CV) does not necessarily mean inconsistencies in performance of all individual laboratories. These systematic differences in ER and PgR test results pave the way for calibration (*see* Fig. 1). However, normalization can be achieved only when a marker is homogeneous with only one molecular form present. The presence of more molecular forms of the analyte will yield a broad range of data, especially when different immuno(metric) assays (with varying sets of antibodies, each with other affinities to these molecular forms of the analytes) are used.

6. CONCLUSIONS

An important issue in applying (pre-)clinical immuno(metric) testing kits is that different kits in many cases generate different assay results in the same tumor specimen owing to variation in test design, antibody specificity

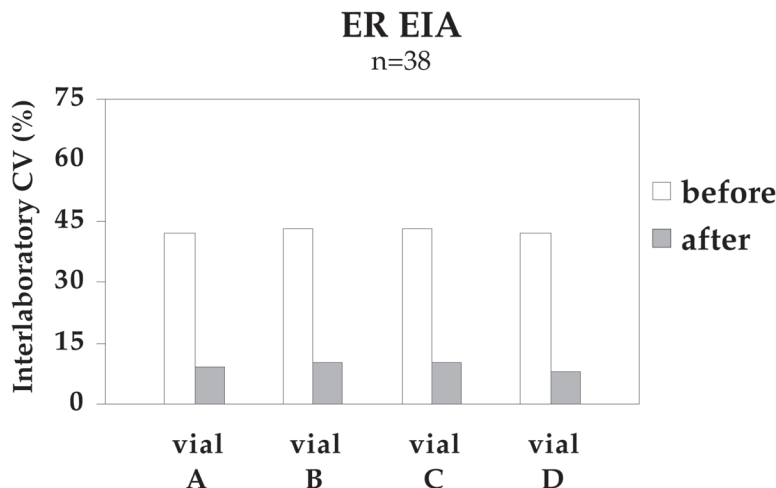


Fig. 1. Interlaboratory CV before (*open bars*) and after normalization of the observed values using a common fifth QA vial (Vial E, gray bars) in estrogen receptor enzyme immunoassay (EIA). Normalization substantially reduces the interlaboratory CVs from 45% to <15%.

and affinity between different kits (or even within a kit from one manufacturer between lots or batches), and use of different calibrators. Although of potential interest, newly explored biomarkers in our view should therefore not be included in large clinical studies unless the assay procedures are carefully evaluated, and common assay protocols, common standards, and QC preparations allowing proper EQA established. At first, such a parameter should be examined in a single expert laboratory. In addition, we strongly advice that in multicenter studies the laboratory performance to be scrutinized prior to generating results from patients in clinical trials.

Hayes proposed criteria for implementing biomarkers in clinical practice, and he defined levels of evidence (LOE) and levels of utility (1). For the highest level (LOE-1) large consistent meta-analysis and validation in a prospective clinical trial should be conducted. Recently, in case of urokinase-type plasminogen activator (uPA) and its inhibitor PAI-1 the level of evidence type-1 was reached, based on the results of a prospective randomized node-negative breast cancer therapy trial (12) and a meta-analysis combining most of the published data sets (13). The therapy trial was under strict external QC by the Receptor and Biomarker Group of the EORTC. The participating laboratories received meticulous instructions on how to run the assays, participated in workshops, used common assays, and were subject of EQC (14). They thus can be regarded

as experienced and qualified, which most likely contributed to the success of this trial as well.

Although considerable progress has been made for some analytes as exemplified above for uPA and PAI-1, standardization of biomarker assay protocols, and development of proficiency testing programs for biomarkers, should be an ongoing process. Only the stringent application of QC systems enables a consistent assessment of the prognostic and/or predictive power of biomarkers.

REFERENCES

1. Hayes DF, Bast RC, Desc CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996;88:1456–1466.
2. McGuire WL, Clark GM. Prognostic factors and treatment decisions in axillary-node-negative breast cancer. *N Engl J Med* 1992;326:1756–1761.
3. Benraad ThJ, Geurts-Moespot J, Grondahl-Hansen J, et al. Immunoassays (ELISA) of urokinase-type plasminogen activator (uPA): report of an EORTC/BIOMED-1 workshop. *Eur J Cancer* 1996;32A:1371–1381.
4. Dittadi R, Meo S, Fabris F, et al. Validation of blood collection procedures for the determination of circulating vascular endothelial growth factor (VEGF) in different blood compartments. *Int J Biol Markers* 2001;16:87–96.
5. Biganzoli E, Boracchi P, Daidone MG, Gion M, Marubini E. Flexible modelling in survival analysis. Structuring biological complexity from the information provided by tumor markers. *Int J Biol Markers* 1998;13:107–123.
6. McGuire WL. Breast cancer prognostic factors: evaluation guidelines. *J Natl Cancer Inst* 1991;83:154–155.
7. Stamey TA, Chen Z, Prestigiacomo AF. Reference material for PSA: the IFCC standardization study. International Federation of Clinical Chemistry. *Clin Biochem* 1998;31:475–481.
8. Kricka LJ. Human anti-animal antibody interferences in immunological assays. *Clin Chem* 1999;45:942–956.
9. Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewart chart for quality control in clinical chemistry. *Clin Chem* 1981;27:493–501.
10. Levey S, Jennings ER. The use of control charts in the clinical laboratory. *Am J Clin Pathol* 1950;20:1059–1066.
11. Geurts-Moespot J, Leake R, Benraad ThJ, Sweep CGJ. Twenty years of experience with the steroid receptor External Quality Assessment program—the paradigm for tumour biomarker EQA studies (review). *Int J Oncol* 2000;17:13–22.
12. Jänicke F, Prechtel A, Thomssen C, et al. Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. *J Natl Cancer Inst* 2001;93:913–920.
13. Look MP, van Putten WLJ, Duffy MJ, et al. Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J Natl Cancer Inst* 2002;94:116–128.
14. Sweep CGJ, Geurts-Moespot J, Grebenschikov N, et al. External quality assessment of trans-European multicentre antigen determinations (ELISA) of urokinase-type plasminogen activator (uPA) and its type-1 inhibitor (PAI-1) in human breast cancer tissue extracts. *Br J Cancer* 1998;78:1434–1441.



<http://www.springer.com/978-1-58829-227-8>

Biomarkers in Breast Cancer

Gasparini, G.; Hayes, D.F. (Eds.)

2006, XII, 342 p. 22 illus., Hardcover

ISBN: 978-1-58829-227-8

A product of Humana Press