Evolution of Quality Assurance for Clinical Immunohistochemistry in the Era of Precision Medicine – Part 2: Immunohistochemistry Test Performance Characteristics

Emina E. Torlakovic, MD, PhD,*†‡ Carol C. Cheung, MD, PhD, JD,*§
Corrado D'Arrigo, MB, ChB, PhD, FRCPath,‖¶# Manfred Dietel, MD, PhD,**
Glenn D. Francis, MBBS, FRCPA, MBA, FFSc (RCPA),††‡‡§§ C. Blake Gilks, MD,‖‖
Jacqueline A. Hall, PhD,¶¶ Jason L. Hornick, MD, PhD,## Merdol Ibrahim, PhD,***
Antonio Marchetti, MD, PhD,††† Keith Miller, FIBMS,*** J. Han van Krieken, MD, PhD,‡‡‡
Soren Nielsen, BMS,§§§‖‖‖ Paul E. Swanson, MD,¶¶¶ Mogens Vyberg, MD,§§§‖‖‖
Xiaoge Zhou, MD,###**** Clive R. Taylor, MD,†††† and
From the International Society for Immunohistochemistry and Molecular Morphology (ISIMM)
and International Quality Network for Pathology (IQN Path)

Abstract: All laboratory tests have test performance characteristics (TPCs), whether or not they are explicitly known to the laboratorian or the pathologist. TPCs are thus also an integral characteristic of immunohistochemistry (IHC) tests and other in situ, cell-based molecular assays such as DNA or RNA in situ hybridization or aptamer-based testing. Because of their descriptive, in situ, cell-based nature, IHC tests have a limited repertoire of appropriate TPCs. Although only a few TPCs are relevant to IHC, proper selection of informative TPCs is nonetheless essential for the development of and adherence to appropriate quality assurance measures in the IHC laboratory. This paper describes the TPCs that are relevant to IHC testing and emphasizes the role of TPCs in the validation of IHC tests.

This is part 2 of the 4-part series "Evolution of Quality Assurance for Clinical Immunohistochemistry in the Era of Precision Medicine."

Historically, immunohistochemistry (IHC) has for all practical purposes been considered a "special stain" similar to traditional histochemical preparations; how-

ever, it is more accurately described as a "cell-based in situ immunoassay" and resides firmly within a class of on-slide, in situ molecular tests that use one of a number of different probes (antibodies, RNAs, DNAs, aptamers or other synthetic molecules) and enable the identification of specific complementary molecular targets.[1] Unlike histochemical tests, "on-slide molecular tests" use reagents that have high (or very high) sensitivity and specificity (in the nanomolar range) for individual targets.[2–5] The great majority of in situ tests, including IHC tests, are developed in the individual laboratory and as such, they are best designated as "laboratory developed tests" (LDTs).[6] Application of appropriate quality assurance measures and the proper validation of "in situ cell-based molecular tests" require a full understanding of which test performance characteristics (TPCs) apply and how validation of these tests is defined. In addition, it is necessary to adapt specific terminology relating to both quality assurance and validation that was initially developed for liquid-based assays. TPCs are essential for the characterization of LDTs.[7–11] As we will show, TPCs as a group are an inherent component of LDT validation.

In this paper, we will define the TPCs that are relevant to in situ cell-based assays that have descriptive readouts primarily generated by human readers (ie, pathologists) ± assistive image analysis and will describe the role of TPCs in the validation of the IHC test. This work builds upon recently published recommendations and guidelines for the validation of antibodies, and molecular and IHC assays from the College of American Pathologists, recommendations for the standardization of controls from the Ad Hoc International Expert Committee, as well as research applications of biomarkers in preclinical and clinical trials for novel and emerging anticancer drugs.[12–19]

TPCs FOR IN SITU CELL-BASED ASSAYS

Here, we adopt and define the following TPCs for IHC: (1) sensitivity and specificity, (2) reproducibility and robustness, (3) accuracy and precision, and (4) reportable range (Table 1).

Sensitivity and Specificity

An Ad Hoc Committee recently published a proposal for a new classification of sensitivity and specificity in IHC testing.[17] The new concept, for the first time, allowed for observation of analytical sensitivity and specificity in a descriptive, nonlinear method such as IHC. The introduction of the concept of analytical sensitivity and specificity for descriptive tests, and its distinction from diagnostic sensitivity and specificity, and clinical sensitivity and specificity, resulted in a 3-tier classification of sensitivity and specificity, which sets the stage for building an appropriate validation framework for IHC. As will be shown technical validation is not possible in the absence of analytical sensitivity and analytical specificity.[20,21] Diagnostic and clinical sensitivity and specificity do not apply to technical validation of IHC tests (even though a valid methodology based on proper technical validation is

TABLE 1. Test Performance Characteristic for Technical Validation of Immunohistochemistry (IHC) Assays

Test Performance Characteristic	Definition
Analytical sensitivity	Determines LOD* of an IHC test
Analytical specificity	Ability of IHC assay to detect one particular substance, rather than others
Preanalytical reproducibility/ robustness	Ability of IHC assay to produce the same results with the same tissue sample for various preanalytical conditions
Analytical reproducibility/ robustness	Ability of IHC assay to produce the same results with the same tissue sample for various analytical conditions
Readout accuracy	Readout agreement with a designated gold standard or designated true value for the readout
Readout precision	Readout reproducibility (intraobserver and interobserver reproducibility of the readout)
Reportable range	Clinically relevant range of values that can be described (or measured by image analysis) for a certain antigen that can be detected by IHC in specimens known to express the antigen with a provision that the specimens had undergone ideal/standard tissue processing
EQA accuracy for reportable range	Ability of IHC test to produce a correct result in EQA (PT or interlaboratory), where external IHC results act as the agreed upon actual (true) value
NIM accuracy for reportable range	Ability of IHC test to produce a correct result in NIM (PT, interlaboratory, or intralaboratory), where results of another (previously validated) method (eg, FISH, PCR, other) act as the agreed upon actual (true) value

*Limit of detection (LOD); the smallest amount of substance in a sample that can accurately be detected by an assay.

EQA, external quality assurance; FISH, fluorescent in situ hybridization; NIM, non-IHC method; PCR, polymerase chain reaction; PT, proficiency testing.

a fundamental requirement for clinical validation studies), but should be demonstrated/known before an IHC test is introduced into clinical practice.[22,23] Therefore, they are also briefly discussed below.

Analytical Sensitivity and Specificity

Generally speaking, analytical sensitivity represents the smallest amount of substance in a sample that can accurately be measured (for quantitative assays) or detected (for qualitative assays) by an assay. In other words, analytical sensitivity determines the limit of detection (LOD).[24,25] Analytical specificity refers to the ability of an assay to uniquely measure or detect one particular substance in a sample.[26] It is critical to recognize that for a descriptive method such as IHC, analytical sensitivity, and analytical specificity are also descriptive. A pertinent example is the demonstration of low molecular weight cytokeratin (LMW-CK) in hepatocytes (a known weak expressor of CK, thus suitable as a descriptive LOD cell type). If an IHC test for LMW-CK shows positivity in hepatocytes, then adequate analytical sensitivity has been demonstrated.[17] LMW-CK is not expressed in benign

lymphocytes, skeletal muscle, and many other tissues. If an IHC test for LMW-CK is reproducibly negative in known negative tissues, then adequate analytical specificity has been demonstrated.[27] In the clinical laboratory analytical sensitivity and specificity are traditionally determined against an analytical reference standard test (ie, a criterion standard or colloquially, a "gold standard"), but this is currently not the norm in IHC testing because for most IHC tests, gold standards are not yet defined. An Ad Hoc Expert Committee recently initiated development of descriptive gold standards for descriptive analytical sensitivity and specificity in IHC testing by establishing the concept of standardized controls for IHC tests, also referred to as IHC critical assay performance controls (iCAPCs).[17] Although iCAPCs include relevant LOD tissue(s), analytical specificity is only narrowly addressed. To establish a more extended analytical specificity it is desirable to include a spectrum of benign human tissues in the evaluation of analytical specificity. Such tissues may be combined into a so-called "specificity tissue microarray (TMA)" (see part 4 of the Evolution Series).

Diagnostic Sensitivity and Specificity

Diagnostic sensitivity is the probability that a test outcome is positive in a diseased person (eg, for IHC tests this usually relates to a particular type of tumor), and diagnostic specificity is the probability that a test outcome is negative in a nondiseased person (eg, in a different type of tumor by IHC). In a clinical setting, 2 useful measures are positive predictive value and negative predictive value. Positive predictive value is defined as the probability that a person randomly chosen from the population who has tested positive actually has the disease; negative predictive value is the probability that a person who has tested negative does not have the disease. Clearly, these definitions assume that the values are calculated and are relevant in a defined diagnostic setting.[28] Therefore, if sensitivity and specificity are defined in these terms, we refer to "diagnostic sensitivity and specificity"; this concept is applicable to IHC in the setting of diagnosing a disease or a condition. Diagnostic sensitivity and specificity of IHC tests are often published in the pathology literature (eg, sensitivity and specificity of DOG1 IHC assay in the diagnosis of gastrointestinal stromal tumor).[29–31]

Clinical Sensitivity and Specificity

To establish clinical sensitivity and specificity (eg, patient response to hormonal therapy for estrogen receptor-positive breast cancer), clinical trials are necessary; this requirement generally goes beyond the purview of an individual laboratory.[32,33]

Clinical utility is a concept that incorporates both diagnostic and clinical sensitivity and specificity. Diagnostic and clinical sensitivity and specificity may be known either from studies performed by the laboratory itself, or those reported in published literature or other reliable sources. For example, studies that document diagnostic, prognostic, and predictive relevance of IHC tests are frequently provided in the peer-reviewed literature.[34–39] IHC tests that have not been diagnostically and clinically validated in published literature should not be introduced in a clinical IHC laboratory without that laboratory itself performing a diagnostic validation and/or clinical validation of its own. Tissue tools for demonstration of IHC analytical sensitivity and specificity are described in part 4 of the Evolution Series.

In summary, "clinical sensitivity and specificity" relate to the ability of an IHC test to predict clinical outcomes and clinical responses to therapeutic agents (prognostic and predictive); "diagnostic sensitivity and specificity" relate to the ability of an IHC test to diagnose a disease or biological state; "analytical sensitivity and specificity" relate to the ability of an IHC test to clearly demonstrate signal in an agreed upon "LOD" cell type, along with optimal signal-to-noise ratio in a fit-for-purpose manner.

Reproducibility and Robustness

Reproducibility and robustness were not specifically addressed by the recent College of American Pathologists guidelines for IHC validation.[15] However, we would argue that because reproducibility is defined as the ability of a test to produce the same result for a given sample when repeatedly tested, and robustness relates to test reproducibility given some changes in test conditions, both are applicable to IHC validation.

Reproducibility Versus Precision

Reproducibility is the ability of a test to produce the same results in repeat testing within a laboratory (it may also be assessed across different laboratories as discussed below). It has been argued that reproducibility is to a qualitative test what "precision" is to a quantitative test.[7,10,11,19,21,40] Therefore, for qualitative IHC assays, the term "reproducibility" is preferred and the term "precision" should not be used.

How to Measure Reproducibility?

Reproducibility is measured by the percentage of times the same results are obtained when a test is performed repeatedly on the same sample (specimen). Although this approach to measuring reproducibility is applicable to IHC testing, the phrase "the same results" needs to be defined for each test because IHC tests are descriptive in nature (eg, for CD31 testing, demonstration of strong staining of endothelial cells plus weak staining of mantle zone lymphocytes in 9 of 10 slides from the same paraffin block demonstrates 90% reproducibility). Intralaboratory reproducibility for qualitative testing was also recently referred to as "accordance."[41] The difference between the usual understanding of intralaboratory reproducibility and accordance is that reproducibility is expressed as a percentage of the same set of results (compared with a predefined standard, such as control materials), while accordance is measured by using paired samples without a predefined standard; it is a measure (also in percentage) of how often the paired samples produce the same results (positive vs. negative). Although

reproducibility can be easily accessed by monitoring results of "on-slide standardized (or well defined) controls" daily, to assess accordance, we would need to prepare at least 10 paired samples and compare their results for agreement in a separate evaluation.

Which Samples to Use to Demonstrate Reproducibility?

Proper selection of quality control samples is critical for demonstration of reproducibility in a descriptive test such as IHC, and such samples should optimally include LOD tissues/cell types. Therefore, iCAPCs or similar samples are well suited to be used for demonstration of reproducibility of qualitative IHC assays (see part 4 of the Evolution Series).

How Many Repeats of a Given Test are Necessary to Demonstrate Reproducibility?

The number of repeat tests that is sufficient to demonstrate reproducibility may be determined by regulatory agencies, professional bodies, or by a laboratory medical director. If standardized on-slide controls are used for IHC testing, then the reproducibility of the test is continually demonstrated and monitored with each IHC slide, because the obtained results are regularly inspected at the intra-laboratory QC stage of IHC testing. This approach is endorsed because in clinical practice, slides that are deemed less than optimal must be recognized, investigated (with laboratory-initiated repeat testing if necessary), and properly documented. Whether a laboratory-identified suboptimal IHC slide is released to the pathologist who ordered the test is usually the decision of the medical director of the laboratory. A recent publication reported approximately 98% overall reproducibility of the IHC tests in a single large clinical laboratory.[42]

Robustness

The robustness of an IHC test (or other in situ cell-based assays) is defined as test reproducibility in the face of changes in various test conditions.[43,44] These changes are often in preanalytical conditions, such as fixation time, decalcification, warm and cold ischemic time, among others. Reproducibility of IHC results for a clinically relevant range of preanalytical conditions defines preanalytical robustness. Similarly, reproducibility of IHC results for a clinically relevant range of analytical conditions (eg, different equipment, different operators, etc.) defines analytical robustness of the IHC test. Demonstration of preanalytical robustness is under the purview of the IHC laboratory, while analytical robustness is the responsibility of both the laboratory and industry (eg, manufacturers/suppliers of IHC reagents and equipment). Many IHC laboratories do not establish robustness because robustness has not yet been mandated for IHC validation and the number of tissue samples for establishing preanalytical or analytical robustness in IHC validation has not yet been defined. Tissue tools for demonstration of IHC reproducibility and robustness are described in part 4 of the Evolution Series.

Accuracy and Precision

Accuracy is usually defined as the degree of closeness of a test measurement to the actual (true) or expected (previously measured) value; this approach to accuracy is not applicable to most IHC testing.[43–45] Similar to polymerase chain reaction tests, IHC tests use amplification and are generally qualitative, and therefore cannot be analyzed based on exact quantitative measurements; this is in contrast to other, more quantitative methods that do not use variable amplification, where the output of the test more closely reflects actual protein content/expression (eg, mass spectrometry, or even immunoassays that use directly labelled primary antibody, such as multicolor flow cytometry and direct immunofluorescence).[10] However, accuracy may be applicable to IHC testing in defined circumstances as follows: (1) accuracy of the IHC protocol and (2) accuracy of the IHC readout.

(1) IHC protocol accuracy is sometimes inappropriately referred to as "concordance." "Concordance" has also been considered as the equivalent of reproducibility when measured across different laboratories; namely the percentage chance that 2 identical test samples sent to different laboratories will both yield the same result (ie, both found positive or both found negative result).[41] However, in this context most of the time one of the results is considered as the designated gold standard result (designated "true value") and therefore, accuracy is a more appropriate term because concordance reflects agreement of the paired samples irrespective of the correctness of the results. In clinical IHC, the output ("stained" slide) is compared with a designated "gold standard," where the designated gold standard result (designated "true value") is provided by a reference laboratory (either through interlaboratory exchange or through a proficiency testing or PT program). In this context, the IHC protocol accuracy is best referred to as external quality assurance (EQA) accuracy.

Where the designated gold standard result (designated "true value") is determined using a non-IHC based testing method, the IHC protocol accuracy is best referred to as non-IHC methodology (NIM) accuracy. In NIM accuracy, the ability of an IHC protocol to produce a correct result is assessed in comparison with another (non-IHC) test methodology for the same biological condition (eg, comparison of IHC test results with results obtained using fluorescent in situ hybridization or polymerase chain reaction testing).[46] EQA accuracy may involve a reference laboratory that provides both IHC and fluorescent in situ hybridization results for a set of PT samples while testing for NIM accuracy may be performed by the same laboratory performing the validation (intra-laboratory NIM accuracy) or by an external laboratory.

To assess IHC protocol accuracy, samples that contain no analyte, as well as samples with low, moderate, and high levels of analyte are usually included. Professional organizations are likely to

continue the practice of issuing recommendations regarding the number of samples for intralaboratory NIM accuracy, as well as the number of samples for interlaboratory IHC exchanges and/or proficiency testing that need to be included to enable meaningful calculations.[14,46–50] Although IHC protocol accuracy is also often referred to as "concordance," when comparison is conducted with the agreed upon "true value" of the IHC test, or with a designated "true value" using another methodology, then "EQA accuracy" and "NIM accuracy" are the preferred terms. Concordance could be used when there is no reference to "correct results," but only interlaboratory reproducibility is being assessed.

(2) There are many types of nonquantitative measurements that are applicable to the IHC readout, which produce different types of data (eg, binary data, categorical data, ordinal data, etc).[10] Where accuracy and precision apply only to the validation of the readout as a specific segment of the IHC protocol, they should be designated as "readout accuracy" and "readout precision." Readout accuracy refers to readout agreement with a designated gold standard or designated true value, and readout precision refers to reproducibility of the readout per se (intraobserver and interobserver reproducibility of the readout). Readout accuracy and readout precision should be determined for type 2-IHC tests (class II and class III U.S. Food and Drug Administration IHC tests), and for type 1-IHC tests (class I) where the read out includes an element of quantification (eg, % cells with a defined qualitative cutoff point that distinguishes between a "negative" and "positive" result). Note that readout accuracy and readout precision are not the same (eg, image analysis is expected to have very high readout precision, but it may have very low readout accuracy if an inappropriately calibrated algorithm is used for the readout). The level of acceptability for readout accuracy and readout precision of various IHC tests may be set by regulatory bodies, professional organizations, or expert/subspecialty groups. Readout accuracy and readout precision are particularly relevant for biomarkers that are used for targeted therapy.[32]

Reportable Range

For liquid-based quantitative laboratory tests, "analytical measurement range" is the term used by the College of American Pathologists to refer to the range of values that a quantitative method can measure for a sample, without dilution, concentration, or other pretreatment of the specimen before testing. Analytical measurement range is the same as "reportable range" in The Clinical Laboratory Improvement Amendments terminology.[51] When applied to IHC testing, which is a descriptive qualitative technique, the reportable range is usually considered to be the clinically relevant range of results that can be described (or measured by image analysis) for a given IHC target that can be detected in specimens known to express the antigen, with the provision that the specimens had undergone ideal tissue processing (eg, ischemic time <1 h, overnight fixation in formalin, and embedding in paraffin). For example, a given range may extend from rare cells positive, to all cells positive, etc. Depending on the IHC protocol conditions, the intensity of staining, above a threshold defined in the controls, may or may not be relevant. Examples include range of expression of estrogen receptor in breast cancer, range of expression of CD10 in follicular lymphoma, etc. Depending on the objective of the test, 3 to 10 samples may be well representative of the reportable range. Reportable range samples may be combined in the form of a TMA paraffin block (also known as "Index TMA"; see part 4 of the Evolution Series). Depending on the type and the number of cases included, accuracy TMA may also be suitable for demonstrating reportable range for most IHC assays (see part 4 of the Evolution Series).

Reportable range is also relevant for the proper selection of quality control samples.[17] Selection of quality control samples for daily controls, and other quality control samples must satisfy the minimum requirement of a negative control and 2 positive controls containing different concentrations of the target antigen, to be included in every run. Note that for modern automated IHC testing, each slide in fact represents a separate run and, hence, the on-slide control rule applies.[17,42] The expression levels should be chosen to verify assay performance for relevant analytic and clinical decision points. This usually means including a high-expressor positive control, near the upper limit of the reportable range, as well as a low-expressor positive control, near or at the lower LOD.[52]

TPCs IN TEST VALIDATION

Demonstration of relevant TPCs is required for technical IHC assay validation. In this discussion it is assumed that primary antibodies and other reagents, as well as equipment that will be used in clinical IHC testing, have already been validated either by industry, or in published literature. Therefore, clinical IHC laboratories must perform additional technical validation of IHC protocols. Furthermore, the medical director (or designate) of the clinical IHC laboratory is responsible for validation of IHC readouts where applicable.

Technical validation of IHC testing is described in detail in part 3 of Evolution Series.

REFERENCES

1. Taylor CR. Predictive biomarkers and companion diagnostics. The future of immunohistochemistry: "in situ proteomics," or just a "stain"? *Appl Immunohistochem Mol Morphol.* 2014;22:555–561.
2. Kominami G. Sensitivity of immunoenzymometric assay and detection method of enzyme. *J Immunoassay.* 1994;15:79–92.
3. Anonymous. KD value: a quantitative measurement of antibody affinity. Abcam. Available at: www.abcam.com/primary-antibodies/kd-value-a-quantitive-measurement-of-antibody-affinity. Accessed August 7, 2016.

4. Pasch A, Farese S, Gräber S, et al. Nanoparticle-based test measures overall propensity for calcification in serum. *J Am Soc Nephrol.* 2012;23:1744–1752.

5. Palliyil S, Downham C, Broadbent I, et al. High-sensitivity monoclonal antibodies specific for homoserine lactones protect mice from lethal *Pseudomonas aeruginosa* infections. *Appl Environ Microbiol.* 2014;80:462–469.

6. U.S. Food and Drug Administration. Draft guidance for industry, food and drug administration staff, and clinical laboratories— ucm416685. 2014. Available at: www.fda.gov/downloads/medical-services/deviceregulationand%20guidance/guida ncedocuments/ucm416685.pdf. Accessed August 7, 2016.

7. Lee JW, Devanarayan V, Barrett YC, et al. Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res.* 2006;23:312–328.

8. Sweep FCGJ, Fritsche HA, Gion M, et al. Considerations on development, validation, application, and quality control of immuno(metric) biomarker assays in clinical cancer research: an EORTC-NCI working group report. *Int J Oncol.* 2003;23: 1715–1726.

9. United Nations Office on Drugs and Crime, Laboratory and Scientific Section. *Guidance for the Validation of Analytical Methodology and Calibration of Equipment Used for Testing of Illicit Drugs in Seized Materials and Biological Specimens: A Commitment to Quality and Continuous Improvement.* New York, NY: United Nations; 2009.

10. Lee JW, Weiner RS, Sailstad JM, et al. Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. *Pharm Res.* 2005;22: 499–511.

11. Marchiò C, Dowsett M, Reis-Filho JS. Revisiting the technical validation of tumour biomarker assays: how to open a Pandora's box. *BMC Med.* 2011;9:1.

12. Bordeaux J, Welsh A, Agarwal S, et al. Antibody validation. *BioTechniques.* 2010;48:197–209.

13. Howat WJ, Lewis A, Jones P, et al. Antibody validation of immunohistochemistry for biomarker discovery: recommendations of a consortium of academic and pharmaceutical based histopathology researchers. *Methods.* 2014;70:34–38.

14. Fitzgibbons PL, Hammond MEH, Valenstein PN. Recommendations for validating estrogen and progesterone receptor immunohistochemistry assays. *Arch Pathol Lab Med.* 2010;134:930–935.

15. Fitzgibbons PL, Bradley LA, Fatheree LA, et al. Principles of analytic validation of immunohistochemical assays: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med.* 2014;138:1432–1443.

16. Torlakovic EE, Francis G, Garratt J, et al. Standardization of negative controls in diagnostic immunohistochemistry: recommendations from the international ad hoc expert panel. *Appl Immunohistochem Mol Morphol.* 2014;22:241–252.

17. Torlakovic EE, Nielsen S, Francis G, et al. Standardization of positive controls in diagnostic immunohistochemistry: recommendations from the International Ad Hoc Expert Committee. *Appl Immunohistochem Mol Morphol.* 2015;23:1–18.

18. Khleif SN, Doroshow JH, Hait WN, et al. AACR-FDA-NCI cancer biomarkers collaborative consensus report: advancing the use of biomarkers in cancer drug development. *Clin Cancer Res.* 2010;16: 3299–3318.

19. Cummings J, Ward TH, Greystoke A, et al. Biomarker method validation in anticancer drug development. *Br J Pharmacol.* 2008;153:646–656.

20. U.S. National Archives and Records Administration. Code of Federal Regulations. Title 42. Public health, vol. 4, chapter V. Health Care Financing Administration, Department of Health and Human Services, part 493. Laboratory requirements, section 493.1253. Standard: establishment and verification of performance specifications. US Government Printing Office, Washington, DC. 2009. Available at: www.gpo.gov/fdsys/pkg/CFR-2015-title42-vol5/pdf/CFR-2015-title42-vol5-sec493-1253.pdf. Accessed August 7, 2016.

21. Burd EM. Validation of laboratory-developed molecular assays for infectious diseases. *Clin Microbiol Rev.* 2010;23:550–576.

22. McGuire WL. Breast cancer prognostic factors: evaluation guidelines. *J Natl Cancer Inst.* 1991;83:154–155.

23. Clark GM, Wenger CR, Beardslee S, et al. How to integrate steroid hormone receptor, flow cytometric, and other prognostic information in regard to primary breast cancer. *Cancer.* 1993;71:2157–2157.

24. Armbruster DA, Pry T. Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev.* 2008;29:S49–S52.

25. Clinical and Laboratory Standards Institute. *Protocols for Determination of Limits of Detection and Limits of Quantitation, Approved Guideline.* Wayne, PA: CLSI; CLSI document EP17; 2004.

26. Saah AJ, Hoover DR. "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. *Ann Intern Med.* 1997;126:91–94.

27. Moll R, Franke WW, Schiller DL, et al. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *cell.* 1982;31:11–24.

28. Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. *BMJ.* 1994;308:1552.

29. Liegl B, Hornick JL, Corless CL, et al. Monoclonal antibody DOG1.1 shows higher sensitivity than KIT in the diagnosis of gastrointestinal stromal tumors, including unusual subtypes. *Am J Surg Pathol.* 2009;33:437–446.

30. Espinosa I, Lee C-H, Kim MK, et al. A novel monoclonal antibody against DOG1 is a sensitive and specific marker for gastrointestinal stromal tumors. *Am J Surg Pathol.* 2008;32:210–218.

31. Miettinen M, Wang Z-F, Lasota J. DOG1 antibody in the differential diagnosis of gastrointestinal stromal tumors: a study of 1840 cases. *Am J Surg Pathol.* 2009;33:1401–1408.

32. Roach C, Zhang N, Corigliano E, et al. Development of a companion diagnostic PD-L1 immunohistochemistry assay for pembrolizumab therapy in non–small-cell lung cancer. *Appl Immunohistochem Mol Morphol.* 2016;24:392–397.

33. Mohsin SK, Weiss H, Havighurst T, et al. Progesterone receptor by immunohistochemistry and clinical outcome in breast cancer: a validation study. *Mod Pathol.* 2004;17:1545–1554.

34. Baltayan A, Naritoku WY, Chaiwun B, et al. LEA.135 expression: identifies low-risk patients with breast ductal carcinoma in situ. *Anticancer Res.* 2002;22:2933–2937.

35. Bilalovic N, Blystad AK, Golouh R, et al. Expression of bcl-6 and CD10 protein is associated with longer overall survival and time to treatment failure in follicular lymphoma. *Am J Clin Pathol.* 2004;121:34–42.

36. Bodey B, Bodey B, Gröger AM, et al. Clinical and prognostic significance of the expression of the c-erbB-2 and c-erbB-3 oncoproteins in primary and metastatic malignant melanomas and breast carcinomas. *Anticancer Res.* 1997;17:1319–1330.

37. Han L, Ming TM, Xu X, et al. LTBP2 is a prognostic marker in head and neck squamous cell carcinoma. *Oncotarget.* 2016. [Epub ahead of print]. Available at: www.oncotarget.com/abstract/8855. Accessed August 7, 2016.

38. Kim S, Nam SJ, Kwon D, et al. MYC and BCL2 overexpression is associated with a higher class of Memorial Sloan-Kettering Cancer Center prognostic model and poor clinical outcome in primary diffuse large B-cell lymphoma of the central nervous system. *BMC Cancer.* 2016;16:363.

39. Torlakovic EE, Bilalovic N, Golouh R, et al. Prognostic significance of PU.1 in follicular lymphoma. *J Pathol.* 2006;209:352–359.

40. Pennello GA. Analytical and clinical evaluation of biomarkers assays: When are biomarkers ready for prime time? *Clin Trials.* 2013;10:666–676.

41. Langton SD, Chevennement R, Nagelkerke N, et al. Analysing collaborative trials for qualitative microbiological methods: accordance and concordance. *Int J Food Microbiol.* 2002;79:175–181.

42. Cheung CC, Taylor CR, Torlakovic EE. An audit of failed immunohistochemical slides in a clinical laboratory: the role of on-slide controls. *Appl Immunohistochem Mol Morphol.* 2015. [Epub ahead of print].

43. Clinical Laboratory Standards Institute (CLSI). Harmonized Terminology Database. Available at: http://htd.clsi.org/listterms.asp?searchd term = reproducibility&button = Submit. Accessed August 7, 2016.

44. European Medicines Agency (EMEA) > ICH Topic Q2 (R1). Validation of Analytical Procedures: Text and Methodology.

CPMP/IHC/381/95. June 1995. Available at: http://www.ema.europa. eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC 500002662.pdf. Accessed on September 6, 2016.

45. Food and Drug Administration. Draft Guidance for Industry, Food and Drug Administration Staff, and Clinical Laboratories Framework for Regulatory Oversight of Laboratory Developed Tests (LDTs). 2014. Available at: www.fda.gov/downloads/medicalservices/deviceregulation and%20guidance/guidancedocuments/ucm416685.pdf. Accessed August 7, 2016.

46. Wolff AC, Hammond MEH, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer. *J Clin Oncol*. 2006;25:118–145.

47. Hammond MEH, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch Pathol Lab Med*. 2010;134:e48–e72.

48. Rüschoff J, Dietel M, Baretton G, et al. HER2 diagnostics in gastric cancer—guideline validation and development of standardized immunohistochemical testing. *Virchows Arch*. 2010;457:299–307.

49. Pazhoomand R, Keyhan E, Banan M, et al. Detection of HER2 status in breast cancer: comparison of current methods with MLPA and real-time RT-PCR. *Asian Pac J Cancer Prev*. 2013;14:7621–7628.

50. Hanna WM, Hammond E, Taylor CR, et al. High concordance between immunohistochemistry and fluorescence in situ hybridization testing for HER2 status in breast cancer requires a normalized IHC scoring system. *Mod Pathol*. 2008;21:1278–1280-1281.

51. Killeen AA, Long T, Souers R, et al. Verifying performance characteristics of quantitative analytical systems: calibration verification, linearity, and analytical measurement range. *Arch Pathol Lab Med*. 2014;138:1173–1181.

52. Centers for Medicare & Medicaid Services, Department of Health and Human Services. Medicare, Medicaid, and CLIA programs; laboratory requirements relating to quality systems and certain personnel qualifications; final rule [published correction appears in *Fed Regist*. 2003;68(163):50722–50725]. *Fed Regist*. 2003; 68(16):3707–3714.