How to Validate Predictive Immunohistochemistry Testing in Pathology?

To the Editor.—I read with interest a recent editorial by Erik Thunnissen entitled "How to Validate Predictive Immunohistochemistry Testing in Pathology? A Practical Approach Exploiting the Heterogeneity of Programmed Death Ligand-1 Present in Non-Small Cell Lung Cancer."[1] It would be significant and desirable if laboratories could simplify different spheres of validation of predictive immunohistochemistry (IHC) biomarkers. The introduction of programmed death ligand-1 (PD-L1) testing for immunotherapy started a new, more complex era for IHC assay development and validation. The challenges that laboratories face forced us to rethink what type of laboratory test IHC is, what "fit-for-purpose" validation means, and a few other parameters that were first embraced by drug development and pharmaceutical research rather than by anatomic pathologists.[2] In following their steps, we have discovered how to distinguish analytic/technical sensitivity and specificity from diagnostic sensitivity and specificity, and that this distinction is essential in the validation of predictive biomarkers.[3,4] Thunnissen explored the role of what he termed "critical samples, which have an epitope concentration close to the threshold of the validated assay," the type of samples that were traditionally used by proficiency testing programs to assess calibration/analytic sensitivity of the IHC assays, and were also previously termed "descriptive limit of detection" and incorporated in IHC Critical Assay Performance Controls (iCAPCs) in 2015.[3] These samples are derived from human tissues or other sources (cell lines, xenografts, etc) and can be used to demonstrate basic analytic sensitivity, specificity, and reproducibility as described for iCAPCs.[4] However, testing of 20 positive and 20 negative samples still applies for technical validation because their purpose is not to show analytic sensitivity, but that the assay protocol performs as it should in a representative set of clinical samples (eg, specific tumor type) by demonstrating reportable range, cellular localization, tissue distribution, results with the representative range of preanalytic conditions, etc.[5] Furthermore, any clinical validation, including "indirect clinical validation" that is used by the author in this editorial, requires at least demonstration of assay diagnostic accuracy in comparison with an established reference standard (or "diagnostic accuracy criteria") and, for the purpose of comparison of methods, at least 50 positive and 50 negative samples are recommended by the CLSI DP12-A2 User Protocol for Evaluation of Qualitative Test Performance.[6] The purpose of the indirect clinical validation is to ensure that the candidate test has the same or nearly the same diagnostic accuracy as the comparator assay (reference test/diagnostic accuracy criteria), and therefore that it is safe for patient selection for a specific therapy. Although analytic sensitivity and specificity are related to diagnostic sensitivity and specificity, no studies have yet demonstrated that we can make direct assumptions from one to the other. Although calibration and optimization of the assay are greatly helped with "critical samples" (aka iCAPCs), to benchmark analytic sensitivity, unfortunately these samples do not tell us too much about diagnostic accuracy and are not sufficient for indirect clinical validation. I am concerned that taking the shortcut approach for indirect clinical validation for predictive biomarkers could compromise patient safety.

Emina Emilia Torlakovic, MD, PhD

Department of Pathology and Laboratory Medicine, Saskatchewan Health Authority and University of Saskatchewan, Saskatoon, Saskatchewan, Canada

1. Thunnissen E. How to validate predictive immunohistochemistry testing in pathology?: a practical approach exploiting the heterogeneity of programmed death ligand-1 present in non–small cell lung cancer. *Arch Pathol Lab Med*. 2019; 143(1):11–12.

2. 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(4): 499–511.

3. 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 AIMM*. 2015; 23(1):1–18.

4. Torlakovic EE, D'Arrigo C, Francis GD, et al. Evolution of quality assurance for clinical immunohistochemistry in the era of precision medicine–part 2: immunohistochemistry test performance characteristics. *Appl Immunohistochem Mol Morphol*. 2017;25(2):7.

5. 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.

6. Garrett PE, Lasky FD, Meier KL, Clark LW; Clinical and Laboratory Standards Institute. *User Protocol for Evaluation of Qualitative Test Performance: Approved Guideline*. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.

Accepted for publication April 5, 2019.

The author has no relevant financial interest in the products or companies described in this article.

doi: 10.5858/arpa.2019-0056-LE

Predictive Markers Require Thorough Analytic Validation

To the Editor.—The recent editorial by Erik Thunnissen, "How to Validate Predictive Immunohistochemistry Testing in Pathology? A Practical Approach Exploiting the Heterogeneity of Programmed Death Ligand-1 Present in Non–Small Cell Lung Cancer,"[1] propagates misunderstandings regarding immunohistochemistry (IHC) assay validation that could be deleterious to accurate predictive marker IHC test development, especially given the emerging impact of programmed death ligand-1 (PD-L1) testing.

Thunnissen considers whether an immunohistochemical slide with heterogeneous staining might be considered a "composite of several hundreds or thousands of analytes." We disagree with this premise. First, if 2 assays were to be compared using serial sections cut from the same block, it would be impossible to directly compare cell-to-cell results, even with sophisticated image analysis capabilities. We consider one *slide* or *case* to be one *analyte* for validation purposes. Furthermore, one cannot assume that heterogeneity is due only to variable protein expression in a given area of tissue, particularly in "critical samples, which have an epitope concentration close to the threshold of the validated assay."[1,2] Besides the stated differences in protein expression, these "critical samples" are more likely to yield heterogeneous results for multiple reasons, including (but not limited to) uneven chemical