Principles of Analytic Validation of Clinical Immunohistochemistry Assays

Jeffrey D. Goldsmith, MD, Patrick L. Fitzgibbons, MD,† and Paul E. Swanson, MD‡*

Abstract: All assays performed in anatomic and clinical pathology laboratories must be validated before they are placed into clinical service. This review summarizes strategies for validation of clinical immunohistochemistry assays, and is chiefly based on the recently released guideline released by The College of American Pathologists.

Key Words: clinical immunohistochemistry, analytic validation, practice guideline

(*Adv Anat Pathol* 2015;22:384–387)

In the current practice of anatomic pathology, immunohistochemistry (IHC) is a critical ancillary test that aids in the accurate diagnosis of a host of neoplastic and non-neoplastic conditions. In addition, IHC is being increasingly used to predict response to therapy and screen for inherited diseases. In the last decades of the 20th century, IHC assays were being developed that could be reproducibly performed on paraffin-embedded, formalin-fixed tissues; these methods were developed as "home-brew" assays, more appropriately termed "laboratory developed tests." As such, assay conditions often varied significantly between laboratories. As IHC became more widespread and its use expanded to industry, detection methods became more standardized. However, as many preanalytic factors may affect the results of IHC tests, assay conditions still may vary significantly between laboratories.

Many IHC laboratories continue to use laboratory developed tests; as preanalytic factors may significantly affect assay results, robust and standardized analytic validation before use on patient samples is required, particularly for those assays with quantitative results or for IHC tests that predict responsiveness to specific therapies. Indeed, analytic validation of all clinical laboratory tests, including IHC, is required by the Clinical Laboratory Improvement Amendments of 1988.[1] Despite both this regulatory mandate and the common sense notion that quality testing is predicated on carefully validated methodology, up to 28% of surveyed IHC laboratories did not have a written procedure for initial assay validation at the time a recent interlaboratory practice survey.[2] The same survey noted that laboratories in compliance with CLIA'88 validation requirements nonetheless followed strikingly variable IHC assay analytic validation practices. To address these challenges to the uniformity and quality of diagnostic IHC, the College of American Pathologists (CAP) convened a panel of experts in 2012 with the charge of creating an evidence-based guideline that would serve as a standard for analytic validation of IHC assays. The resulting recommendations were published in 2014.[3]

With these introductory comments in mind, we herein review the relevant concepts behind analytic validation with particular focus on analytic validation of IHC assays. The authors of this review served on the expert panel that created above-mentioned guidelines; however, this article has been created without input from the CAP.

GENERAL CONSIDERATIONS

The United States Food and Drug Administration defines "validation" as "confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use can be consistently fulfilled."[4] In other words, analytic validation is a process that confirms that a test has the expected level of sensitivity, specificity, and reproducibility for its intended use. In the context of the clinical pathology laboratory, validation is achieved by comparing the test's result with a known gold standard. However, a vast majority of IHC assays do not have a gold standard referent test that can be feasibly obtained by most laboratories. As such, most laboratories must compare their results to comparators that are not considered gold standards in the strict sense. Such comparators tend to fall in the following 4 categories.

(1) Morphology and expected results according to the medical literature: This comparator is frequently used when new assays are being initiated. Typically, the medical director of the laboratory performs a review of the literature pertinent to the new assay. From those data, a set of validation cases is chosen, typically from the laboratory archives from cases fixed and processed in the same manner as those that will be run on patient samples.

(2) Previous results from a previously validated assay from the same laboratory: This method is often used if the assay conditions change to such an extent that merits

From the *Department of Pathology, Beth Israel Deaconess Medical Center, Children's Hospital Boston, and Harvard Medical School, Boston, MA; †Department of Pathology, St. Jude Medical Center, Fullerton, CA; and ‡Department of Pathology and Laboratory Medicine, University of Calgary and Foothills Medical Centre, Calgary, AB, Canada.

In various combinations, authors have received honoraria and travel expense reimbursement for speaking at the 2011 (PS), 2013 (PF, JG), 2014 (PF, JG), and 2015 (PF, JG) College of American Pathologists Annual Meetings regarding The Guidelines cited in this article. J.D.G. and P.E.S. spoke at the 2014 annual meeting of the American Society of Clinical Pathology regarding these guidelines. J.D.G. also spoke at the 2014 annual meetings of the American Society of Cytopathology and the United States and Canadian Academy of Pathology. In addition, all authors were on the committee that produced the aforementioned guidelines and received reimbursement from the College of American Pathologists for expenses incurred as a result of committee participation.

Reprints: Jeffrey D. Goldsmith, MD, Department of Pathology, Beth Israel Deaconess Medical Center, 330 Brookline Ave., Boston, MA 02215 (e-mail: jgoldsmi@bidmc.harvard.edu).

some sort of revalidation (see below). For example, if a manufacturer discontinues a primary antibody and it is replaced with a different primary antibody clone, this change is considered a fundamental modification to the assay that requires complete revalidation. In this circumstance, the use of results obtained from previously validated assays from the same laboratory as a comparator would be a reasonable approach.

(3) Another laboratory's results from the same validation set using a previously validated assay: This method is particularly useful for assays that are difficult to validate. In this situation, interlaboratory comparison allows the laboratory to directly compare results from a previously validated assay on the same tissues.

(4) Previously validated results from a sufficiently validated nonimmunohistochemical assay: As noted above, this comparator applies to very few assays, but is often the most robust validation method. Examples include chromogenic or fluorescent in situ hybridization (CISH/FISH) for Her2-neu as applied to Her2 IHC, flow cytometric analyses for markers such as CD3, CD20, and other common hematopoietic analytes, and mutation testing for the *BRAF* V600E mutation as compared with mutation specific b-raf IHC.

CONCORDANCE AND SIZE OF VALIDATION SET

The desired level of concordance between the new assay and the comparator is tightly related to the size of the validation set. This is due to the fact that both of these parameters have a hand in determining the confidence interval for a particular level of concordance. The confidence interval, generally set at 95%, is the statistical value that determines the level of confidence that the observed concordance level reflects the true performance of the test. Thus, as the size of the validation set increases, the level of confidence that the observed concordance is the true value increases. For an example, see Table 1; this table shows that the 95% confidence intervals are smaller and overall confidence levels higher with a validation set composed of 40 cases compared with 20 case validation set. Thus, as a general rule, a larger validation set is desirable, whenever possible. Of course, larger validation sets can be difficult to obtain, especially in smaller laboratories.

The size of the validation set should also be dictated by the intended clinical use of the assay. The clinical use of IHC assays fall into 2 general groups. The first are markers that are interpreted in the context of the morphologic findings and are typically used as ancillary stains for diagnosis (eg, cytokeratin 7, cytokeratin 20, TTF-1, GATA-3, etc.). The second group of stains includes those that are interpreted without regard to the histologic context; many of these markers give predictive information about the sensitivity of a tumor to various treatments (eg, Her2 IHC on breast carcinoma, b-raf mutation-specific IHC on melanoma). Markers that are used for histologic diagnosis and are interpreted in a histologic context have less direct clinical impact than predictive markers that result in an actionable result that is independent of the morphologic context. Thus, the size of the validation set for a predictive marker should be larger than that prepared for a diagnostic marker. The expansion of the size of the validation set for predictive markers increases the confidence that the observed concordance level truly reflects the desired level of concordance. As such, the CAP Guideline mandates that the size of the validation set for predictive markers should contain at least 40 challenges, whereas nonpredictive/diagnostic markers should have at least 20 challenges. Depending on the resources available, expansion of the size beyond the prescribed amount of the validation set is optimal and would add additional assurance that the assay will behave as expected.

In theory, the level of aggregated positive and negative concordance between the new test and the comparator should be 100%. However, this is not practically obtainable due to a number of factors including, but not limited to, preanalytic factors, intratumoral heterogeneity of analyte expression and the quality of the originally validated method or comparator set. For these reasons, the Guideline set the desired level of concordance at 90%; this was chiefly based on evidence from concordance data between Her2 IHC and Her2-neu FISH, in which concordance levels higher than 90% were not feasible for a majority of laboratories.[5–7]

COMPOSITION OF THE VALIDATION SET

As a general rule, the composition of the validation set should reflect the intended clinical use of the assay. Not only should relevant positive cases be included, but also judicious inclusion of cases that show lack of expression of the analyte of interest should be part of the validation set. For example, TTF-1 is a transcription factor that is often used as an ancillary test in the workup of metastatic carcinoma of unknown origin. It is expressed in a majority of small cell carcinomas of the lung, most primary pulmonary adenocarcinomas, and many types of primary epithelial tumors of the thyroid gland. Inclusion of tumor types that are known to be positive for TTF-1 should be part of the validation set. In addition, tumor types that are known to be TTF-1 negative and are in the histologic differential diagnosis of either metastatic pulmonary adenocarcinoma and metastatic thyroid carcinoma should be included. Such examples of clinically relevant TTF-1 negative carcinomas might include ductal carcinoma of the breast, colorectal carcinoma, and pancreatic ductal adenocarcinoma. Inclusion of clinically relevant cases in the validation set adds additional assurance that the validation accurately reflects the performance of the assay when performed on patient samples.

Occasionally, assays are used in more than 1 clinical context. In this circumstance, it would be wise to tailor the validation set to reflect all potential clinical uses. For example, CD30 is a marker that is often used to diagnose

TABLE 1. Comparison of Concordance Rates and 95% Confidence Intervals for Validation Sets Composed of 20 and 40 Tissues

No. Validation Tissues	Concordance for 0 Discordant Cases	Concordance for 1 Discordant Case	Concordance for 2 Discordant Cases
20	100% (81%-100%)	95% (75%-100%)	90% (69%-98%)
40	100% (90%-100%)	97.5% (86%-100%)	95% (83%-99%)

Hodgkin lymphoma and various germ cell tumors, such as embryonal carcinoma. In this circumstance, the validation set should include cases of Hodgkin lymphoma, in which Reed-Sternberg cells are the expected positive cells, and cases of embryonal carcinoma. Relevant negative cases include nodular lymphocyte-predominant Hodgkin lymphoma and CD30-negative primary mediastinal diffuse large B-cell lymphoma, which are known mimics of Hodgkin lymphoma and are CD30 negative. In addition, expected CD30-negative cases in the differential diagnosis of embryonal carcinoma should be included in the validation set that might include seminoma and yolk sac tumor.

FORMAT OF VALIDATION CHALLENGES

Classically, validation is achieved by applying single tissue sections on slides analogous to the practice on patient samples. More recently, tissue microarrays have been used as a more efficient and cost-effective method of displaying multiple challenges on a single microscopic slide.[8–12] Tissue microarrays are usually an acceptable method of validation. However, caution should be exercised with assays that are known to show significant heterogeneity of staining. Examples of this include bcl-6 staining in normal tonsillar tissue. Bcl-6 expression is limited to germinal center cells; as such, a tissue microarray would not be an effective method. Similarly, CD15 and CD30 validation using classic Hodgkin lymphomas should not be performed using tissue microarrays, as the CD15-positive and CD30-positive Reed-Sternberg cells are very often heterogenously distributed within lesional tissue in this tumor type.

PREANALYTIC CONSIDERATIONS

Once tissues become devitalized at the time of biopsy or resection, they are fixed, processed, and prepared for microscopic diagnosis. This process can differ between laboratories and, in fact, may vary within a particular laboratory depending on the specimen type. These variations in tissue processing and handling may have dramatic effects on IHC results. For example, for some antibodies, acidic decalcification solutions can change the avidity of the primary antibody for its epitope(s).[13] Although it is impossible to control for all possible preanalytic factors during validation, attention to major causes of preanalytic variation should be taken into account. Some of the major preanalytic factors that may impact results include fixative type and preparation method (ie, cytologic preparations vs. formalin-fixed, paraffin-embedded tissues).[14,15]

If tissues fixed in alternative fixatives or tissues exposed to decalcifying solutions are to be used for IHC, efforts should be made to ensure that the results are clinically valid. A reasonable approach would be to validate a subset of assays that are often run on decalcified samples. Examples of such assays might include cytokeratins, CD45, S-100, and estrogen receptor.

Similarly, if IHC is run on cytologic preparations, including smears, cytospins, cell blocks, and ThinPrep preparations (or core samples submitted with aspirate fluid or other preparation to the cytology laboratory in CytoLyt or other nonformalin solutions), reasonable efforts should be made to assure that these assays perform adequately before they are used on patient samples. The selection of markers tested and number of cases included in these separate validation studies must be determined by the laboratory medical director.

REVALIDATION AFTER CHANGES TO ASSAY CONDITIONS

Once initial assay validation is successfully completed and a test is placed in clinical service, it is common for assay conditions to change. When that occurs, some sort of revalidation is needed to assure that the assays perform as expected. In general, changes to assay conditions fall into 3 categories. The first, and perhaps most straightforward, is a change to the antibody clone. As different antibody clones target different epitope(s), changes in antibody clone are considered a fundamental change to the assay. In this circumstance, full analytic revalidation is required.

The second category includes modifications to assay conditions that are common to all assays in the laboratory. Examples include changes to detection chemistry, water supply, antigen retrieval solution(s), and tissue processing equipment. When such changes occur, it is not necessary to fully revalidate all assays affected by the change. Instead, it is reasonable to choose a representative sample of assays run in the laboratory and compare cases prepared with the modified assay conditions with examples representative of original conditions. If the subset of modified assays performs as expected, it would be reasonable to assume that the remaining assays will perform adequately. If, however, significant changes to the assay conditions are necessary to achieve expected results, more extensive revalidation may be necessary.

The final set of condition changes that merit revalidation are changes that apply to single assays. Examples of this might include changes to antibody lot, primary antibody dilution, primary antibody incubation time, and change of primary antibody vendor using the same clone. Of these changes, a new antibody lot (same clone) often results in minimal perturbation of the assay. As such, verification of continued expected assay results is achieved by running 1 known positive and 1 known negative case. It may be judicious to include a third case that shows a low-positive reaction as an additional indication of appropriate assay performance. Changes to primary antibody dilution, incubation time, and vendor are more substantive changes to the assay. In these circumstances, it is reasonable to run 2 known positive and 2 known negative cases to assure continued assay performance; again, it may be wise to run a fifth, low positive, case to assure appropriate assay sensitivity.

CONCLUSIONS

IHC is a critical ancillary test in the modern anatomic pathology laboratory that often has significant impact on patient care. To be assured of accurate results, robust analytic validation must be performed on all assays before their use on clinical samples. This review summarizes best practices for analytic validation for IHC assays and outlines an approach for revalidation necessitated by changes to assay conditions after successful completion of initial validation procedures.

REFERENCES

1. US Department of Health and Human Services. Clinical laboratory improvement amendments of 1988: Final Rule. *Fed Regist*. 1992;57:7001–7186.

2. Hardy LB, Fitzgibbons PL, Goldsmith JD, et al. Immunohistochemistry validation procedures and practices: a College of American Pathologists survey of 727 laboratories. *Arch Pathol Lab Med*. 2013;137:19–25.
3. 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.
4. Institute CLS. *Quality Assurance for Design Control and Implementation of Immunohistochemistry Assays: Approved Guideline (CLSI Document I/LA28-A2)*, 2nd ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2011.
5. Mayr D, Heim S, Werhan C, et al. Comprehensive immunohistochemical analysis of Her-2/neu oncoprotein overexpression in breast cancer: HercepTest (Dako) for manual testing and Her-2/neuTest 4B5 (Ventana) for Ventana BenchMark automatic staining system with correlation to results of fluorescence in situ hybridization (FISH). *Virchows Arch*. 2009; 454:241–248.
6. Rhodes A, Jasani B, Anderson E, et al. Evaluation of HER-2/neu immunohistochemical assay sensitivity and scoring on formalin-fixed and paraffin-processed cell lines and breast tumors: a comparative study involving results from laboratories in 21 countries. *Am J Clin Pathol*. 2002;118:408–417.
7. van der Vegt B, de Bock GH, Bart J, et al. Validation of the 4B5 rabbit monoclonal antibody in determining Her2/neu status in breast cancer. *Mod Pathol*. 2009;22:879–886.
8. Fons G, Hasibuan SM, van der Velden J, et al. Validation of tissue microarray technology in endometrioid cancer of the endometrium. *J Clin Pathol*. 2007;60:500–503.
9. Gulbahce HE, Gamez R, Dvorak L, et al. Concordance between tissue microarray and whole-section estrogen receptor expression and intratumoral heterogeneity. *Appl Immunohistochem Mol Morphol*. 2012;20:340–343.
10. Kwon MJ, Nam ES, Cho SJ, et al. Comparison of tissue microarray and full section in immunohistochemistry of gastrointestinal stromal tumors. *Pathol Int*. 2009;59:851–856.
11. Thomson TA, Zhou C, Chu C, et al. Tissue microarray for routine analysis of breast biomarkers in the clinical laboratory. *Am J Clin Pathol*. 2009;132:899–905.
12. Dessauvagie BF, Thomas C, Robinson C, et al. Validation of mitosis counting by automated phosphohistone H3 (PHH3) digital image analysis in a breast carcinoma tissue microarray. *Pathology*. 2015;47:329–334.
13. Athanasou NA, Quinn J, Heryet A, et al. Effect of decalcification agents on immunoreactivity of cellular antigens. *J Clin Pathol*. 1987;40:874–878.
14. Hanley KZ, Birdsong GG, Cohen C, et al. Immunohistochemical detection of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 expression in breast carcinomas: comparison on cell block, needle-core, and tissue block preparations. *Cancer*. 2009;117:279–288.
15. Nishimura R, Aogi K, Yamamoto T, et al. Usefulness of liquid-based cytology in hormone receptor analysis of breast cancer specimens. *Virchows Arch*. 2011;458:153–158.