

No ifs, ands, or buts on IHC assay validation

Karen Titus

March 2014—Like Gypsy Rose Lee, tests and their true nature reveal themselves bit by bit. For immunohistochemistry, this unhurried disclosure has meant evolving ideas of whether these tests must indeed be validated and, if so, then how, exactly. The discussion recently culminated in a new CAP guideline for laboratories.

“Principles of Analytic Validation of Immunohistochemical Assays” was scheduled to be published March 19 online ahead of print in *Archives of Pathology & Laboratory Medicine* (<http://tinyurl.com/ihcguideline>). It’s a pioneering effort to address an area overlooked in anatomic pathology.



Dr. Patrick Fitzgibbons, chair of the group that wrote the new guideline, says the feedback helped the group reconsider the discretion lab directors have. “We’re basically stressing, more than we had initially, that the lab director has to be responsible for making some of the decisions,” he says.

While laboratories have known for years that assays need to be validated before being put into clinical service—it’s part of CLIA, after all—not everyone has appreciated that tests that essentially resemble special stains need to be scrutinized, too.

“Pathologists have learned that validation of immunohistochemical assays is a little bit more important than they might have thought five or six years ago,” says Paul E. Swanson, MD, a member of the workgroup that produced the guideline and professor of pathology, University of Washington School of Medicine, Seattle. Some laboratories may have figured it out on their own by being attentive to the model proposed in the HER2 guidelines and following through with ER and PR testing, says Dr. Swanson, who was formerly the director of anatomic pathology

at UW. “But they weren’t entirely sure whether it applied, for example, to a structural protein that defined a pattern of differentiation rather than a possible target for therapy.”

Patti Loykasek, HTL(ASCP), QIHC(ASCP), another member of the workgroup, says that she does at least one CAP inspection a year and that she too sees a gap. “I think labs have gotten a *little* better about knowing they need to validate tests, but I think it’s done a little haphazardly. The results aren’t always well-documented, and final data collation and sign-off by the medical director are often missing,” says Loykasek, test development technologist at RML (Regional Medical Laboratory), Tulsa, Okla.



Loykasek

While the CAP checklists ask if antibodies are validated, says Loykasek, they give no specific parameters for how to validate, leaving much open to interpretation. “Most people are going to do the least amount of work possible, because they’re busy,” she says. “We’re always asking them to do more work with fewer people.”

“There was definitely a need for a set of guidelines,” she adds.

Three IHC tests have already run the validation gauntlet and are the subject of their own guidelines: HER2, ER, and PgR. (These three markers are thus not covered in this most recent document.) But some pathologists had long suspected that apart from this trio, IHC validation was a hazy concept for many labs.

Hunches gave way to proof with a recent study, says Patrick Fitzgibbons, MD, who chaired the workgroup. It’s the fourth reference in the guideline (Hardy LB, et al. *Arch Pathol Lab Med*. 2013; 137[1]:19–25), which detailed a CAP survey looking at IHC validation procedures and practices in 727 laboratories. (Dr. Fitzgibbons and another workgroup member, Jeffrey D. Goldsmith, MD, were coauthors.)

“What we learned,” says Dr. Fitzgibbons, who also chairs the CAP Cancer Biomarker Reporting Committee, “is that there really is not a consistent mechanism for validating immunohistochemistry assays.”

As the workgroup searched the literature, further inconsistencies became apparent. Some papers recommended validation sets of 20 positive cases and 20 negative cases. Others suggested more cases, and still others, fewer.

Beyond that basic—*how many?*—lay others. Should all assays be validated the same way? Or were there differences?

“Let’s say you use a different fixative, or let’s say you decalcify a specimen, because it’s bone tissue,” Dr. Fitzgibbons says. “Does that affect the validation?”

What about antigens that are extremely difficult to find, so-called rare antigens? If a validation set requires 40 cases, “There may not be a lab in the country that can get 40 of these, if they’re that rare. What do you tell labs in that setting? How do you validate assays for rare infectious organisms?” asks Dr. Fitzgibbons, a pathologist at St. Jude Medical Center, Fullerton, Calif.

The survey also showed that many laboratories were unaware when assay revalidation is needed, says Dr. Fitzgibbons. What requires full revalidation (equivalent to initial assay validation) and what requires only confirmation that the assay is working as intended?

These were among the issues facing the workgroup as they put together the guideline.

The guideline's 14 recommendations should give laboratories a solid push out of the starting blocks.

The first recommendation sets matters straight: Laboratories must validate all immunohistochemical tests before placing them into clinical service. Per the guideline, means include (but aren't limited to):

1. Correlating the new test's results with the morphology and expected results;
2. Comparing the new test's results with the results of prior testing of the same tissues with a validated assay in the same laboratory;
3. Comparing the new test's results with the results of testing the same tissue validation set in another laboratory using a validated assay;
4. Comparing the new test's results with previously validated nonimmunohistochemical tests; or
5. Testing previously graded tissue challenges from a formal proficiency testing program (if available) and comparing the results with the graded response.

Beyond that declaration, the guideline's authors highlight some other critical areas:

- For initial validation of assays used clinically (apart from HER2, ER, and PgR), labs should achieve at least 90 percent overall concordance between the new test and the comparator test or expected results. "It could be another IHC test done at a different laboratory, or another marker or another methodology, like in situ hybridization," says Dr. Fitzgibbons. The most common scenario would be a lab using a new antibody for a marker it has offered in the past, he says. "Because antibodies change all the time. If you have a completely new antibody clone, you should revalidate it."
"We also allow labs to use just expected results," he continues. "Because sometimes you don't have another test, but from the literature you know what the results ought to be."
- For predictive marker assays (again, with the exception of HER2, ER, and PgR), labs should test a minimum of 20 positive cases and 20 negative cases. If the lab's medical director decides that a validation set of fewer than 40 cases is sufficient, he or she will need to document the rationale.
- For nonpredictive factor assays, the guideline recommends a smaller validation set: a minimum of 10 positives and 10 negatives. Again, lab directors who decide that a smaller validation set is appropriate need to document their reasons.

In essence, there are two levels of validation. Why would one test require less stringent validation than another?

Dr. Swanson traces the answer back to the early practice of immunohistochemistry. Before the advent of predictive and prognostic markers, IHC focused on giving information that helped to resolve a reasoned, histologic diagnosis, a practice that remains largely true today. When tests are an ancillary element of analysis, he says, “they are, I think, quite reasonably seen as less risk to a patient.” He points to a similar line of reasoning at the FDA, which considers risk to patients when determining approvals and clearances of IHC reagents and other medical devices. “With that difference in mind, we felt a less stringent approach to a diagnostic validation was appropriate,” he says.

While the workgroup was willing to recommend a smaller validation set for nonpredictive markers, Dr. Swanson makes clear they nonetheless still require a higher standard than labs might have previously thought. “You might say, ‘Well, let’s do three cases, because I know from my experience these cases should be positive, and maybe a couple negatives—and everything will be fine.’ But that’s not true,” Dr. Swanson says. “Anybody who does laboratory medicine knows that you can’t establish a reference range or an expected outcome for a given test unless you’ve looked at enough samples to achieve a credible level of reproducibility.”

The committee thus wanted to provide a guideline that had, as Dr. Swanson puts it, “statistical meat to it” but could still be attained by the typical laboratory.

Where did those numbers come from? “Sometimes people think these numbers are pulled out of the air. I know I did when I read previous guidelines,” says Dr. Goldsmith, director of the surgical pathology laboratory at Beth Israel Deaconess Medical Center, Boston, and assistant professor of pathology, Harvard Medical School. “So it’s worth mentioning that we deliberated for a long time, walking the fine line between doing the right thing and not making it overly onerous on the labs. At the end of the day, the number that we came up with for a typical validation set was supported by statistics,” which are provided in the guideline’s supplemental material. “It would be better to have 50 cases,” he continues, “but everyone knows if we had 50 cases in the validation set, no one would ever do it.”



Dr. Swanson

For those who find it difficult to obtain the required number of cases for validation sets, it’s possible, says Dr. Swanson, that three or four smaller labs could join efforts, sharing information and material, for, say, rare antigen. “It’s a little extra work; in my mind, it’s not a lot of extra work,” he says. And while the committee members discussed the importance of having validation tissues handled, processed, fixed, and stained in the same way clinical materials are, “We also know that that’s not always practical even for large reference labs, because they are often working with materials that were not processed in their lab.”

The guideline also makes clear, says Dr. Swanson, that labs will sometimes use smaller validation sets. “It’s not that labs can throw the recommendations out the window or neglect them,” says Dr. Swanson, “but using the 20-case validation set could be altered to fit certain clinical circumstances at the laboratory director’s discretion.”

With this approach, however, 10 different medical directors might approach the problem of rare antigen, for example, in 10 different ways. “Can you be sure that the quality of that stain in those 10 laboratories is comparable?” Dr. Swanson asks. “The answer is no.” That’s why the guideline requires directors to document their alternative validation and demonstrate objectively its validity. While not stated in the guideline, the implication is clear, says Dr. Swanson: “If you can’t establish validity of a test, you shouldn’t do the test.”

- The recommendations on revalidation address three possible changes:
 1. When a new lot of antibody is opened, “We don’t believe you need to completely revalidate the assay,” says Dr. Fitzgibbons. “We just think you need to confirm that the new antibody is working as expected.” One known positive and one known negative case should suffice.
 2. If there are minor changes to the assay itself—antibody dilution, using the same antibody clone but purchasing it from a different company, or making changes in incubation times—labs should run two known positives and two known negatives. “Slightly more stringent,” as Dr. Fitzgibbons puts it. “But still a fairly easy confirmation that the assay performs as expected.”
 3. When a lab uses an entirely different clone or antibody, the assay needs to be completely revalidated, as if it were a brand-new assay.
- The committee spent a fair amount of time discussing the best approach for specimens other than routinely formalin-fixed paraffin-embedded tissues, including cytology specimens and decalcified specimens. They eventually decided not to specify the number of cases needed for validation sets. Given the wide variety of cytology specimens, for instance, it was too hard to come up with a number that worked for all situations. Labs do need to take steps to prove that the assays work on the alternative specimen types, however.
- Tissue microarrays, also known as multitissue blocks, presented another mind-bender. Labs are increasingly using tissue microarrays as an efficient means of validating assays, Dr. Fitzgibbons notes. But does the literature support that? “Our conclusion was these are acceptable specimens for validation purposes for the majority of cases,” he says, “though there are limitations to their use.”
- A final highlight, says Dr. Fitzgibbons, is also the most obvious. Recommendation No. 14 reminds labs that they need to document all validations and verifications in compliance with regulatory and accreditation requirements.

“It’s a no-brainer,” he says. But it’s worth noting because it was the least controversial item when the guideline was put out for public comment.

Not every recommendation met with such genial response, which made the workgroup sit up and take notice.

The guideline's first incarnation had 18 recommendations. The group winnowed it down after the public comment period, which garnered some 1,000 comments from more than 200 individuals, Dr. Fitzgibbons reports. "We deleted some; we consolidated some others. And we really rewrote quite a few of them."

"The guideline was made better by the comment period," says Dr. Goldsmith, noting that in many cases the feedback focused on practical concerns.

Loykasek was the only laboratory technologist in the guideline group. As such, she also brought pragmatic views to the discussions. "Sometimes things on paper sound very doable, but in practicality, in the lab, it's almost impossible," says Loykasek, who previously was involved in validating new IHC assays at PhenoPath Laboratories, Seattle.

One lesson from PhenoPath, she says, was that validation requires labs to think about specificity. It's easy to fall into the trap of looking only for an antibody to stain a specific cell type. "You need to look beyond that—what should this antibody be negative on, and can we prove that it's indeed negative?" Labs also need to look for cross-reactivity, Loykasek says, given that an antibody will oftentimes stain more than one thing. "See how your tissues are fixed and processed, and what kinds of cross-reactivities you'll have. And document those."

She says that whenever a new antibody came onboard at PhenoPath, it was always assigned to one technologist and one pathologist who would do the workup together. Likewise, she says, "Technologists can play a huge role" in helping labs follow the new guideline. At Pheno-Path, she says, validation was most successful when technologists were involved and when the process was well organized. "Before they started, they knew how they were going to capture and track their data and had the forms ready."

Dr. Swanson urges medical directors to involve all laboratory personnel in the design of validation protocols. "They're invested in the quality of the lab, and they want to understand why we're making changes." That's another responsibility for the lab director, in fact—making the argument clearly to others in the lab and being receptive to feedback and suggestions. For example, he says, laboratorians might more readily recognize that a 10-positive and 10-negative validation set doesn't accurately represent the expected stain distribution of a given marker in the clinical population tested in their lab. "Maybe you want to do 12 positives and eight negatives to better reflect that distribution," Dr. Swanson says. "This is the sort of conversation we've had in our laboratory. It gives the laboratory director greater insight into the nuances of the testing environment, and provides a bigger role in the validation process to those who actually run the tests."

The feedback during the comment period helped the group reconsider the discretion laboratory directors have in ensuring validation. "We're basically stressing, more than we had initially, that the lab director has to be responsible for making some of the decisions," Dr. Fitzgibbons says.

Commenters also took issue with the numbers used for the validation sets. "People didn't like having a minimum," says Dr. Fitzgibbons. Some wanted no number given at all; others said one or two cases should suffice. "We had some individuals comment that as long as you're doing positive and negative controls, you don't need to validate your assay, which we of course disagreed with," says Dr. Fitzgibbons.

There were also some comments that fell along the "state's rights" axis. "A lot of the negative comments focused on not having an organization like the CAP tell a lab how to do its business," says Dr. Fitzgibbons, who adds, "We anticipated that there would be people who don't like guidelines at all. But there were quite a few comments to that effect."

Dr. Swanson offers advice to those naysayers, which, in blunt terms, is: Get used to it. With more interdisciplinary guidelines likely to appear—the ASCO/CAP collaborations on HER2, ER, and PgR testing are prime examples—there will be added pressure on lab directors to more objectively define how they determine the quality of their IHC assays, he says.

Then there were the comments that revealed a lack of understanding about basic validation tenets. "Some

people didn't recognize it's a CLIA requirement—they thought it was more a discretionary thing," says Dr. Fitzgibbons.

Is it surprising that some labs view validation as optional? "I don't really know the answer to that, because we were surprised, too," says Dr. Fitzgibbons. The guideline became more than an attempt to bring order out of chaos. It's also an effort to build something from nothing. "Some labs weren't validating their assays at all," says Dr. Fitzgibbons.

He and others in the workgroup turn to history for answers—specifically, the history of special stains. Not everyone views, say, a keratin stain as a laboratory test, but rather as a special stain. "With these, we're usually referring to histochemical stains like trichrome and PAS, stains that have been around for a hundred years," says Dr. Fitzgibbons. Some pathologists may not view them as tests because they're stains that permit better assessment of tissue but don't provide stand-alone results. Some may then reason that validation isn't needed. But, says Dr. Fitzgibbons, "There are good reasons why that's not true."

Twenty-five years ago, at the dawning of the IHC era, pathologists—who already had plenty of experience doing special stains—didn't consider the new assays to be all that different. IHC was seen more as a special stain than a quantitative analyte such as serum glucose.

We now know, of course, that they're identifying specific analytes and even quantifying those analytes," says Dr. Fitzgibbons. IHC tests are different from special stains, in other words, especially with predictive markers, where a single test result can drive therapy. And validation is critical.

The 2007 HER2 guideline toppled the first domino, asking labs to validate IHC tests like they would any other clinical lab test. "In other words, doing everything the right way," says Dr. Fitzgibbons.

Initially, predictive markers were thought to be more important from a validation standpoint, which is partly borne out by the aforementioned CAP survey. Validation of nonpredictive markers was much less consistent, he says. "It's not like the predictive markers were perfect," he says. "But clearly we were further along in that category."

At the same time, the boundary between predictive and nonpredictive markers is a fluid one, much like it can be hard to define what, exactly, is a molecular test. Some nonpredictive markers are used individually, "and they may make a huge difference," says Dr. Fitzgibbons. A keratin stain alone might be used on an undifferentiated malignant tumor to identify a poorly differentiated carcinoma; the patient would then be treated for carcinoma, not lymphoma. "It's not a simple adjunct," Dr. Fitzgibbons argues. "It's completely changed how you interpreted the case."

As the relationship between the diagnosis and targeted therapy becomes more precise, the traditionally nonpredictive lineage selective markers effectively become predictive in certain clinical settings. So it's reasonable, Dr. Swanson says, to keep open the discussion about whether a "diagnostic" test is less risk to a patient than a predictive one. "You can still make that argument in most settings, but it is becoming increasingly difficult as the lines between predictive and nonpredictive markers are blurred," he says.

CD117 (c-kit) offers one well-characterized example, says Dr. Swanson, noting the marker is considered both diagnostic of gastrointestinal stromal tumor and predictive, generally, of response to anti-c-kit (Gleevec) therapy.

The lines could very easily blur even more with the rise in targeted therapies, based on molecular and morphologic analysis. The guideline says that for a marker with predictive and nonpredictive applications, labs should validate it as a predictive marker when used as such.

The guideline doesn't purport to have all the answers, and, being a guideline, it is by definition something that will be revised. Dr. Swanson is fine with that. "Basically, we want to get our foot in the door and remind laboratories of their responsibility to the patient in providing an assay with reproducibility and high predictive value." And, he adds, "This was designed, in part, to make it as palatable as possible to a laboratory, allowing it to comply with what we regard as reasonable expectations for developing clinically precise and confident assays."

The guideline will, the group hopes, stimulate more research. The need is there, says Dr. Fitzgibbons, noting, “We didn’t have the strength of evidence for most of these [recommendations] that we hoped to find. There isn’t a lot of level one evidence for IHC.”

Adds Loykasek: “When papers are published on new antibodies, they tend to gloss over how they were validated.”

The HER2 guideline, once again, could be a good model to follow. “When that was published, there was a bit of an uproar,” Dr. Goldsmith recalls. “And as a result, people started publishing research that addressed the various points of contention, and the guideline changed.” In its first incarnation, for example, the guideline called for a fixation interval of between six to 48 hours. “Almost everyone in pathology thought that was too strict, that there were no downsides to testing specimens fixed for longer than 48 hours,” Dr. Fitzgibbons recalls. “But we could not prove it.” Since then plenty of published evidence has made the case for a longer interval, and the updated guideline recommends a six- to 72-hour interval.

The IHC guideline isn’t meant to usher in a Day of Wrath for labs. Dr. Swanson notes that when the group began its work, the goals were to re-emphasize the notion that all tests have to be validated and to provide basic guidance for the general immunohistochemistry laboratory.

“I would hope,” says Dr. Swanson, “that people would look at this as something that will help them do their job well.”

[hr]

Karen Titus is CAP TODAY contributing editor and co-managing editor. Jeffrey Goldsmith, MD, will present a CAP webinar on the guideline for analytic validation, to take place April 1 from 11 am to noon CDT. Register at <https://www1.gotomeeting.com/register/801536592>.