Small groups, big answers in HER2 testing

Karen Titus

July 2018—Breast cancer has a way of deflating conventional wisdom:

- Maybe it's not about the journey.
- Maybe more is better.
- Maybe communication is overrated.

Take the new ASCO/CAP guideline for HER2 testing (Wolff AC, et al. *Arch Pathol Lab Med*. Epub ahead of print May 30, 2018. doi:10.5858/arpa.2018-0902-SA). Since the first groundbreaking joint guideline appeared 11 years ago, the authors have made a habit of addressing cases that flummox pathologists, medical oncologists, and patients. Now, in 2018, they have clarified the diagnostic approach to in situ hybridization groups two, three, and four, rare cases that nonetheless cause an outsized share of headaches and worries. It also clarifies language from the 2013 guideline that had sent some labs astray, and it addresses the use of multiple alternative chromosome 17 probe assays.



As Dr. Kimberly Allison puts the new HER2 guideline into practice at Stanford, she's revising reporting templates and meeting with colleagues in the cytogenetics lab. "In our reporting," she says, "we want to reflect the additional workup and that some of these result categories are unusual." [Photo: Cindy Charles]

The previous guidelines turned out to be tough acts to follow—a bit like following Sean Connery in the role of James Bond—even as the new one benefits from new data. The first one plunged into the brave new world of companion diagnostics and marked the first time the CAP and ASCO joined forces on a guideline, recognizing the need for both specialties to partner on a systems level as well as in routine clinical practice, says lead author Antonio Wolff, MD, professor of oncology, Breast and Ovarian Cancer Program, Johns Hopkins Sidney Kimmel Comprehensive Cancer Center, and member, Miller Coulson Academy of Clinical

Excellence at Johns Hopkins.

As the guidelines evolved, from 2007 to 2013 and now most recently, "We have learned a lot more about how HER2 testing happens in daily practice," says Dr. Wolff, who, along with coauthor Elizabeth Hammond, MD (respectively, the lead oncologist and lead pathologist authors on all three guidelines), spearheaded the first document. That, in turn, enabled the current guideline's authors to refine their focus.

Did someone say Daniel Craig?

Because of the broad reach of earlier documents, the authors wanted to present not only updates—obviously part and parcel of any revised guideline—but also a clear overview of the topic. They did so in the summary tables and with comprehensive figures. "We didn't just want to say, 'This is what we changed from last time—read the old document again,'" says coauthor Kimberly Allison, MD, professor of pathology, director of pathology residency training, and director of breast pathology and breast pathology fellowship, Stanford University School of Medicine. "We're building on too many different things." Dr. Allison became involved at the tail end of the 2013 update, initially in the role of patient advocate (she notes she was treated for HER2-positive breast cancer in 2008); she's subsequently been involved because of her professional expertise.

In that sense, the guideline emphasizes what *hasn't* changed. "Table 1 is an important read," Dr. Allison says, "because it has all that laid out," showing the summary of all recommendations, both the original and focused updates. Moreover, a number of the figures sketch out diagnostic testing algorithms, "which have become somewhat complicated for certain groups of FISH testing."

The guideline could be summarized in the vein of the oft-recited political poem (albeit with a much happier result):

First they came for the false-positives, Then they came for the false-negatives, Now they come for the unusual results cases.

With the first guideline, says Dr. Wolff, "Our main concern was about an excessive number of false-positive test results," which had become apparent with the enrollment of patients in the first generation of adjuvant trials in HER2-positive disease. Rates ran between 25 and 30 percent. "We were in that moment concerned about the need to improve specificity."

Having succeeded in reducing the frequency of such results, "We then wanted to make sure we were not potentially missing false-negatives," he says. The swinging pendulum—between false-positives and false-negatives, specificity and sensitivity—explains the focus the authors took with the 2013 guideline in handling difficult cases. False-negative results were estimated by some to be as high as 10 percent, though the actual frequency was difficult to estimate because only the positive cases were being submitted for central test confirmation.

Laboratories have made significant strides, particularly with preanalytical (care of specimens) and analytical (assay standardization) issues. At this point, "We may have succeeded in improving the accuracy of HER2 testing to above 95 percent," Dr. Wolff says.

Now, in 2018, the focus has shifted once again, as more information has emerged about less common types of HER2 test results, as seen in dual-probe in situ hybridization (ISH) groups two, three, and four.

"We also began to realize that some labs were misinterpreting some of the recommendations from 2013," he says. The most obvious example involved cases of reference labs that performed only in situ hybridization. When some of those results were initially inconclusive, labs might re-test with up to five alternative chromosome 17 probes, leading to a significant concern about false-positives, says Dr. Wolff.

The most recent update ponders five clinical questions, three of which focus on so-called unusual results categories. These are the low-frequency, high-intensity cases—about five percent overall—that, like the piccolo in Beethoven's 9th, command attention despite appearing only briefly.

"We now have very specific guidance that takes into account primarily a joint review of immunohistochemistry and in situ hybridization," says Dr. Wolff, which "will help pathologists best resolve those difficult cases."

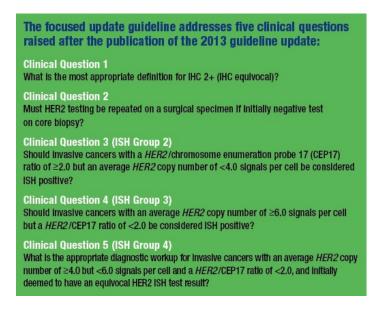
It comes at a compelling juncture. He and his clinical colleagues are starting to see that "in the treatment of HER2positive disease, outcomes are improving in a major way."

Dramatic evidence of this appears in the APT study (Tolaney SM, et al. *N Engl J Med.* 2015; 372[2]:134-141), which looked at treatment for patients with anatomic low-stage, HER2-positive disease, a group that had been ineligible for the pivotal trials of adjuvant trastuzumab. Patients received weekly paclitaxel and trastuzumab treatments for 12 weeks, followed by nine months of trastuzumab monotherapy. (Endocrine therapy was added in ER-positive cases.) The strategy of deescalating chemotherapy plus trastuzumab from two to three cytotoxic drugs to just one worked, Dr. Wolff says. The study has since been updated—seven-year data is available but not yet published, he says. "The survival in that initial paper [median follow-up: four years] was excellent, and we are now able to offer the potential of excellent survival outcome to patients treated with HER2-targeted therapies while avoiding excessive chemotherapy toxicity," says Dr. Wolff, a coauthor of the study.

But for deescalation to work, "we need to be cautious," he says. "We need to have as much certainty about the accuracy of the HER2 testing as possible." If a candidate patient actually has triple-negative disease, rather than HER2-positive/estrogen-receptor-negative disease, chemotherapy regimens with two or three cytotoxic drugs would be the recommended adjuvant regimen, without any anti-HER2 targeting drugs like trastuzumab.

It's possible, says Dr. Wolff, looking even further ahead, that deescalation might benefit patients with higher anatomic stage cancer. Patients who have meaningful response to perioperative therapy, who at the time of surgery have evidence of a pathologic complete response, might benefit from an approach similar to that used in the APT study. "In the next generation of HER2-target trials, we plan to integrate pathologic response to neoadjuvant therapy as a functional prognostic biomarker of outcome along with HER2 testing as a predictive biomarker of response to targeted therapy, and design studies that better enrich patients for therapy deescalation or escalation strategies, and ultimately be able to offer just the right treatment for the right patient."

"So we really need to make the diagnosis as accurate as possible," Dr. Wolff reiterates. He sends an approbative look toward the laboratory. "I think we are there."



Overall, IHC maintains its pole position as an excellent test for initial screens, assuming the test is well validated, says Dr. Allison.

That's not an assumption the guidelines have ever taken lightly, the authors say. And as Dr. Allison notes, "All our same rules apply for test validation and proficiency testing, including fixation and preanalytical variable controls. Those haven't changed. All those recommendations from the 2013 update still stand."

With the new guideline, "We actually bring IHC testing into more prominence now that the quality control has gone up. Protein expression is a very important aspect of testing," she says. "It's really the phenotype you're treating,

which has a very tight association with HER2 gene amplification."

While in situ hybridization testing is used by some laboratories as an initial test, most use it as a second-line test when IHC results are equivocal, says Dr. Allison. "And more and more labs are also doing dual testing—IHC and FISH testing on all cases.

"But that's not a requirement in any way," she continues. "Most labs can still follow the algorithm of IHC first." If the result is equivocal (IHC 2+), subsequent ISH testing should follow. Results will unfold in several directions.

One possibility: getting clearly amplified cases (group-one results) that have elevated *HER2* signals per cell (\geq 4.0), and a *HER2*/chromosome enumeration probe 17 ratio greater than or equal to 2.0. These are obvious ISH positives that correlate well with higher levels of protein expression by IHC (2–3+). In addition, cases with <4.0 *HER2* signals per cell and ratios <2.0 are also obvious negatives (group-five results), and these correlate well with the absence of protein overexpression.

And then there are three groups that make up the gray zone. The 2013 guideline acknowledged their existence and proposed result categories for them, but also acknowledged the very limited to nonexistent evidence of their frequencies and clinical-pathologic features and behavior and leaving group-four results in the "no man's land," Dr. Allison says, of an equivocal or unresolved result.

Dual-probe ISH groups two, three, and four are uncommon (less than five percent of all cases) but require precision, like *passé simple* French verb conjugations. They are addressed in clinical questions three, four, and five in the guideline.

Cases in group two (the Sheldons of the world might be irritated that these are addressed in clinical question No. 3) have low *HER2* copy number (<4.0 signals per cell) but a *HER2*/CEP17 ratio \geq 2.0 These cases tend to have loss of the centromeric control signal. They also would have been considered eligible for the original clinical trials for HER2-targeted therapy, given their ratio-positive status, so the 2013 guideline considered them positive as well. "We initially didn't want to exclude them from the therapy option," Dr. Allison says.

The authors of the new guideline looked at recent data on the rates of concordance between IHC positivity versus negativity. "The findings were that these were largely IHC-negative cases," Dr. Allison says. They're frequently ER positive, and they typically don't have biologic features that would suggest they behave more aggressively than an ER-positive, HER2-negative cancer.

"You would expect a HER2-positive truly amplified, truly protein-overexpressing cancer to behave in a more aggressive way than a HER2-negative cancer that's ER positive," she says. But the (admittedly limited) data available from the first generation of adjuvant trastuzumab trials suggest patients in group two didn't receive much benefit from HER2-targeted therapy. Hence the new guideline's recommendation: For group-two ISH cases, do a concurrent IHC test. "Some labs might just currently do FISH without looking at protein expression in cases like this," Dr. Allison explains. IHC is a useful guide for scoring ISH/FISH, she says, particularly because it helps recognize if there is regional heterogeneity, but it also can flag discordant results with unusual ISH result categories.

If the IHC result is 3+ positive, "you can go ahead and result the ISH test as HER2 positive," Dr. Allison says. If it's IHC negative (zero or 1+), then it should be called HER2 negative. The guideline provides a suggested comment that can be modified and included in the report to note the aforementioned limitations of the early trials, including the enrollment of only small subsets of group-two cases.

And if the blurriness continues, in the form of a 2+ result? "It's recommended to double-check the initial in situ hybridization result by having a second observer count at least 20 cells that are in that area of 2+ IHC staining," she says. The second reviewer should be blinded to the first reviewer's counts. If the result remains in the group-two category, with a 2+ IHC result, "it's going to be negative overall result for group-two cases," she says, and should be noted as such with comment. If the new count moves the result into a different ISH category, the final

category should be resolved by internal procedures, the guideline notes.

Dr. Allison characterizes these steps as "a little additional workup." Labs can choose to adopt the additional ISH counting workup for all group-two cases, if that's easier for the testing parameters. But at minimum, group-two cases that are IHC 2+ should get a second count. Essentially, these cases—again, ratio-positive, signals-per-cell-negative—are considered positive only if additional IHC testing is 3+ positive.

Clinical question No. 4 addresses cases in group three (cue coughing noises from the overfastidious). These are the opposite of group-two cases, in that they are signals-per-call high (\geq 6.0) but with a *HER2*/CEP17 ratio <2.0. Should these be considered ISH positive?

Only a few rare cases like this were included in one of the original trials, and only if very high *HER2* copy number, Dr. Allison notes, since they were ratio negative. But in the 2013 update these cases were considered positive, based on high copy number alone; the ratio, she says, was originally designed to control for polysomy. That places the picture slightly askew. "So instead of true gene amplification causing protein overexpression, is it a cancer that just has multiple copies of chromosome 17? That might not be a true HER2-positive case was the original thinking during the initial trials." But more recent data, including studies that looked at scrutinizing other areas of the chromosome, showed that the majority of these cases have co-amplification of large regions of chromosome 17, rather than true polysomy, Dr. Allison says.

"This group was a little more controversial than group two," she says, given that this is likely a more heterogeneous result category. Some data groups indicated the majority of such cases are IHC negative, with a handful of IHC positives that appeared to have very high-level *HER2* amplification. Other groups showed a large number of IHC 3+ cases and high levels of amplification.

"Again, it's a rare group," says Dr. Allison, which means these cases will be very sample dependent. "It seems like it's a mixed group. And that includes some more aggressive features and more frequent ER-negativity. Some of these cases have really high-level *HER2* amplification, even though they're also ratio negative, and they have the positive protein overexpression by IHC to support true HER2-positive phenotype." Again, the recommendation is for concurrent IHC in the workup of an initial group-three ISH result. A negative IHC means pathologists can call the case negative, although, Dr. Allison adds, "I would also double-check a negative result," to ensure there were no problems with fixation, decalcification, and so on.

Positive IHC results mean the case can be resulted as positive. "That's a nice concordant result," she says.

And for those troublesome 2+ IHC cases? "Again, additional counting by a second observer in the areas that are 2+ staining," Dr. Allison says. In contrast to group two, because of the evidence supporting that many of these cases have co-amplification of the HER2 and centromeric control signals, the guideline's authors wanted to give the benefit of the doubt to patients who have 2+ IHC results and high levels of *HER2* copy number, but are ratio negative. Cases that are 2+ and 3+ are both considered positive. As with group two, the guideline provides a comment for inclusion in pathologists' reports. Dr. Allison says she starts her comments by noting the results are unusual/infrequent, and that additional testing was done per the new guideline, to explain what's been done and how that might differ from the way similar cases were previously handled.

On to group four (clinical question No. 5). Can you bear to hear the word "equivocal" one more time?

In the 2013 guideline, these cases were breast cancer's interminable trip on a Great Plains interstate, with their often equivocal IHC results, followed by an equivocal FISH, IHC, or both. Do we retest? Do we do alternative probes? Do we do IHC? *Do we take another sample to get us out of the loop of continuous equivocal results?* "It seemed endless," Dr. Allison says.

The majority of these cases (*HER2*/CEP17 ratio <2.0; average *HER2* copy number \geq 4.0 and <6.0 signals per cell) are zero to 2+ by IHC, while 3+ cases are rare. And, as with group three, their ratio negative status meant they

were not included in the original trials. Based on data from other trials, however, it doesn't appear that this group does any worse than other HER2-negative cases that are ER positive (as most of these cases are).

Retesting these cases can feel like flipping a coin multiple times, says Dr. Allison. "And if you're close to a threshold for a result, it's sort of chance whether you end up one side or the other." Yet again, the recommendation is to use IHC to help decide the final ISH result. The algorithm is similar to that used for group two. IHC 3+ is called positive, though these cases "are extraordinarily rare," says Dr. Allison. IHC negatives are called negative, with a recommended comment about the uncertainty of benefit and chance involved in retesting results. If the 2+ result persists on recount, call it negative, with an explanatory comment. "The majority of these cases will now be considered negative," she says.

Are you now ready to pop Champagne corks? With this now rigorously defined group, "There will be no equivocal category anymore for ISH testing," says Dr. Allison. It's as if the Académie Française had banned the use of a loanword.

Is this as exciting as it sounds? "It is!" she says. "I think that is one of the biggest accomplishments" of the new guideline.

"'Equivocal' was a troublesome word," she says. "It implied more needed to be done. And that wasn't always the case." The original intent wasn't to doggedly pursue an elusive, perhaps nonexistent truth. "It's a gray zone, and sometimes they exist, and then difficult clinical decisions need to be made within that context. So we're trying to disperse the gray zone a bit."

In the meantime, she adds, "We'll continue to collect the data, and if these results need to be fine-tuned even more, we can do that." But for now, the rallying cry is *More:* more IHC, more observers, more comments.

The further good news is that these are rare groups. "Infrequently frequent," Dr. Allison calls them. "But there should be confidence that the majority of HER2 tests are clear-cut results [groups one or five]. We're not debating those at all."

Other areas of the guideline didn't require a tune-up. The authors wondered whether to re-address the matter of HER2 heterogeneity—also unusual—but decided the earlier guidance, from 2013, was still sufficient: The area of interest remains clustered areas of overexpression or gene amplification. If that's seen, the area should be counted separately by ISH and scored separately. Scattered, intermixed heterogeneity is likely not clinically relevant and doesn't need to be reported.

Heterogeneous cases should still be reported as amplified if they have greater than 10 percent clustered heterogeneity. "But you should include in your report the percent that's amplified overall," Dr. Allison says. IHC is a useful technique for discovering heterogeneity (though it's not a requirement) because it enables pathologists to more easily see it at a lower power. "And then you'd want to score a FISH area separately in that setting." The other two clinical questions in the new guideline could be considered semantic as well as clinical in nature.

The first was a revision of what the 2+ by IHC category meant. Dr. Allison says the 2013 guideline suffered from a somewhat confusing definition, which was addressed in an earlier correspondence (Wolff AC, et al: Reply to E.A. Rakha, et al. *J Clin Oncol.* 2015;33[11]:1302–1304). In 2013, says Dr. Allison, "We didn't mean to change the definition of 2+ from the FDA definition." The newest guideline makes clear the same descriptors should be used. "This is not controversial—it's just fixing the wording."

Use of the word "must" also created confusion in the 2013 guideline, evidence that three little words—"must," "should," and "may"—can change the course of testing, if not romance. The context here: calling for repeat HER2 testing on surgical specimens that are initially negative on core biopsy. The 2018 guideline has been revised to say repeat testing "may" be ordered.

The 2013 guideline discussed this issue in the accompanying data supplement. "The problem with the data supplement is it's going to be the rare aficionado who's going to read it," says Dr. Wolff. The discussion has now

moved to the body of the document, which should draw a larger readership.

Earlier concerns about false-negatives led the authors to take a conservative approach and use the word "must." But, says Dr. Wolff, experience shows that a negative core biopsy, if done appropriately, is likely to be negative in the surgical specimen as well.

Both these changes "were easy fixes," says Dr. Allison. "They didn't require a lot of discussion and had already been addressed" in the earlier correspondence.

As she puts the guideline into clinical practice at Stanford, Dr. Allison says she's revising reporting templates and meeting with colleagues in the cytogenetics lab. "And then it's a little more complicated reporting for these rare cases, because you're folding in another test to your final result."

On a practical level, such testing might already be happening, in a hidden-part-of-the-iceberg sort of way. "I think a lot of ISH labs already did this in the background, without reporting it," she suggests. She plans to include the IHC findings for cases in the unusual results categories, although the guideline does not make this a requirement. By showing how she arrived at the final result, "Everybody understands what your workup is," she says.

"In our reporting we want to reflect the additional workup and that some of these result categories are unusual," Dr. Allison says. That makes for a more complex report, with an additional result in some cases, and more comments.

Based on the group's discussions, she's confident that clinicians will find this useful. "They didn't like treating the group-two cases that had ratio positive but really low signals per cell, usually IHC negative," she says. These were more often low-grade cancers, more often in older patients. "You don't want to treat a low-grade biology with chemotherapy plus Herceptin if it's unclear there's going to be benefit. Potentially you're harming them." On the other hand, they like the idea of giving the benefit of the doubt to ISH group three, IHC 2+ and 3+ cases, she says, since these appear to be more aggressive cancers.

And for patients, it should help alleviate the angst that often accompanies an equivocal result. While it will mean fewer patients will be considered HER2 positive, it also means nonaggressive cases won't be overtreated. And more aggressive cases will be treated accordingly.

"Everybody should be happy there's no equivocal anymore," she says.



Dr. Wolff

Dr. Wolff agrees the added information in the reports will likely be helpful, especially for medical oncologists who aren't breast specialists. Faced with unusual results groups, "they need even more guidance from the pathologist about whether the patient is likely to benefit from HER2-targeted therapy."

A bit facetiously, Dr. Wolff says that as the guideline is refined, "The best situation is you don't need to talk to anybody, right?" He laughs. "You can simply get on with the job you need to do, comfortable with the information you have and that everything is working well." Indeed, if the test is done optimally each and every time, "then to a degree the system is working"—the goal of that first joint guideline. "We would expect actually a lesser need for oncologists to pick up the phone and call the pathologist. The number of times there's a difficult case, that emails will be flying back and forth between oncologists and pathologists, or among oncologists at different institutions, asking for help on how to deal with these more complex cases—I think we're going to make things a lot easier for everybody."

The equivocal category, he says, was in the beginning viewed as a tool to increase communication, to alert pathologists to the need for additional testing. "But what we realized after the fact is we were being left with an occasional case that the pathologist wasn't able to resolve."

Searching for answers, pathologists would then perform additional tests, hoping to get an answer. The test of choice: alternative probes. "When you do that, especially when you are close to the line of positive/negative, it doesn't take much for a single test by chance alone to give you an answer that is different from the one before." And by default, that testing population was enriched for more complex cases. Using an aphorism he favors, he says, "Multiple equivocals do not add up to a positive." It's simply a positive by chance, which meant uncertain oncologists ran the risk of treating even triple-negative patients with HER2-targeted therapy.

There was debate about how much to focus on single-probe versus dual-probe assays. Since the guideline doesn't recommend the use of single-probe assays, the authors wanted to minimize discussion. One-probe assays appear to be heading the way of the checkbook; dual-probe assays might be useful for labs that have limited resources.

The new guideline, says Dr. Wolff, strongly discourages the routine use of alternative probes. Some expert labs will continue to use them on very difficult cases, he acknowledges. "But in that case these labs are likely to be the exception, not the norm, and they're going to report more carefully, which is going to hopefully curtail the epidemic of alternative probe testing that occurred after 2013, which was really not our intention—was truly kind of an unintended consequence of how some labs interpreted what to do."

Not to be overlooked is the reminder that regardless of the HER2 testing result, anatomic pathology still rides shotgun. The authors emphasized this point in 2013, and Dr. Allison says it remains important. No one should look at a case strictly through the HER2 peephole. "If you're resulting a grade-one cancer as HER2 positive, you need to stop and rework your case," is how she puts it. "A pure mucinous carcinoma shouldn't be HER2 positive.

"You should still be a responsible pathologist," she says.

In the meantime, as new HER2-targeted therapies are developed, HER2 testing will continue to be the standard to qualify patients for these drugs, says Dr. Allison. "You want the HER2 testing to be able to apply to all HER2-targeted therapies. I don't think we're going to be revising the way we test based on just more drugs on the market."

There may be a role for add-on tests based on certain combinations of therapies. And HER2 testing will remain crucial as medical oncologists explore deescalation and other approaches. "We now can begin to explore additional biomarkers that can help identify what we believe is a group of patients who appear to have disease that is exquisitely sensitive to HER2-targeting manipulations," she says. "At the end of the day, having very tight assays for HER2 assessment allows us to test all of these strategies going forward. It's a big deal."

But for now, HER2 serum testing and gene expression arrays—to name just two possibilities—are more promise than product.

This is not a disappointment. Dr. Allison calls it "comforting to know we've come such a long way and standardized this HER2 testing really well. We don't want to throw that by the wayside.

"It's an amazing story, actually," she continues. It's good, she says, to be reminded of that first burst of glory—the light hasn't dimmed. She still recalls the excitement of having a test that qualified patients for a targeted therapy, and the many steps along the way to make the process even better, including multiple guidelines. "It's nice to see we're really at the point of fine-tuning rather than reworking the whole system over and over."

Karen Titus is CAP TODAY contributing editor and co-managing editor.

Resources are online to help laboratories implement the recommendations (<u>http://bit.ly/cap-asco-her2</u>). In addition, the 2018 Laboratory Accreditation Program checklists (anatomic pathology, cytogenetics, molecular pathology), to be released in August, will contain the guideline updates, as will the breast biomarker reporting template.