

ER/PgR guideline hones approach to ER-low positives

Anne Paxton

April 2020—The CAP and the American Society of Clinical Oncology released two years ago a focused update of their clinical practice guideline for HER2 testing for breast cancer, following an update in 2013. This year comes another update as the CAP and ASCO issue their latest guideline for estrogen receptor and progesterone receptor testing. The goal of the ER/PgR update: to continue improving the analytical performance, diagnostic accuracy, and clinical utility of hormone receptor testing.

“Things were changing in HER2 testing more rapidly than they were in ER testing,” says Kimberly Allison, MD, director of breast pathology at Stanford University Medical Center and co-chair of the multidisciplinary international expert panel in charge of the 2020 guideline update. “But when we were doing the 2018 HER2 update because there was new HER2 data emerging, we realized we needed to look at ER/PgR guideline updating as well.” In addition to a meta-analysis that had been published shortly after publication of the initial 2010 ER/PgR guideline, more data were emerging on the ER-low-positive group, she explains.

The expert panel Dr. Allison co-chaired, after several months of meetings, agreed to continue recommending ER testing of invasive breast cancers by validated immunohistochemistry as the standard for predicting which patients may benefit from endocrine therapy. But the new guideline contains three key changes: new reporting recommendations for low-positive ER results (one to 10 percent); recommendations that laboratories establish standard procedures to optimize performance, interpretation, and reporting of cases with low-positive or negative results; and a new testing recommendation for patients diagnosed with noninvasive ductal carcinoma in situ (Allison KH, et al. *Arch Pathol Lab Med*. Epub ahead of print Jan. 13, 2020. doi: 10.5858/arpa.2019-0904-SA).

“The good news is that a lot hasn’t changed,” Dr. Allison says of the latest recommendations. “The 2010 guideline set several standards that are the same. So we’re really fine-tuning the grey zone a little bit here, just as we did in the HER2 guideline update in 2018, to make sure the more unusual results are as reproducible and clinically useful as possible.”



Dr. Allison

While the CAP/ASCO guidelines are focused on setting a threshold for ER positivity based on its ability to predict potential benefit from endocrine therapy, she says, ER testing is used for other purposes—determining overall treatment pathways, eligibility for clinical trials, predicting molecular subtype—where the same threshold for positive might not be the best discriminator. Gene expression data and clinical neoadjuvant treatment have shown that a significant proportion of the ER-low-positive cancers tend to profile and behave more like typical triple-negative cancers, Dr. Allison says. “In addition, the data on the benefit of endocrine therapy in this group is more limited and challenging to extrapolate from original ligand binding assay data to IHC percents.” For these reasons, a specific “low positive” reporting category was created, along with a recommended comment explaining the more limited data on this group, for cases with low ER expression—one percent to 10 percent.

The new recommendations respond to the more limited data on endocrine responsiveness in this low-positive group and the overlapping features with ER-negative cancers, says expert panel member David L. Rimm, MD, PhD, director of Yale Pathology Tissue Services, Yale School of Medicine. “The previous guideline set a conservative cutpoint to maximally treat patients with endocrine therapy.” Now, he says, new management strategies for

cancer treatment have prompted a reexamination of what the cutpoint should be.

In 2010 there were fewer treatment options for patients with triple-negative results. “It was before the PARP inhibitors were available,” Dr. Rimm says, referring to the targeted breast cancer therapy poly (ADP-ribose) polymerase inhibitors. Thus, much of the focus of the guideline has changed, he says. “It’s become more important to not bias every diagnosis in favor of calling a result ER-positive.” If a patient is two percent positive and the tumor otherwise has features of a triple-negative cancer, “are we doing them an injustice by calling them an ER-positive? Because they aren’t going to get a PARP inhibitor from which they might benefit, we feel that if a patient is ER-negative, that needs to be recognized because there are now options for ER-negative patients.”

Many of the standard ideas about breast cancer treatment still rely, of necessity, on randomized double-blind endocrine therapy trials conducted in the 1970s, Dr. Allison says. A trial that assigned subjects to placebo rather than endocrine therapy would not be considered ethical now if the cancer has ER expression. “You’re not going to do a new trial where you say, ‘Let’s not give endocrine therapy to ER-positive cancers by IHC at different levels and see if there’s a better threshold.’ You’re kind of doing the best you can with the old data and translating it into today’s assay.”

Although, theoretically, anyone who is ER-positive is a candidate for endocrine treatment, Dr. Rimm says, the placebo studies showed that many patients didn’t need that treatment because even if they got the placebo they did well. “However, the patients who got the endocrine therapy did better than the placebo patients. So we had a reason to treat them in both the adjuvant setting and in the metastatic setting.”

In the course of his career, Dr. Rimm says, ER cutpoints have shifted substantially. “When I was a resident, the ER-positive cutpoint was an H-score of 75”—intensity times percentage of cells positive at intensity of 1+, 2+, or 3+. “So that would have meant more than 10 percent positive cells at 3+ intensity would be negative [H-score = 30]. Then they switched it down to 10 percent positive, and then down to one percent positive. So clearly we were always erring toward increasing the number of ER-positive patients because we had a therapy for those patients. But if they were ER-negative, until recently we didn’t have a therapy. We wanted to perhaps officially bias our cutpoint toward making sure that no patient missed the opportunity to benefit from therapy if they had even a whiff of ER presence.”

Now there are new therapies for triple-negative breast cancer and new trials underway that give more options to ER-negative patients than they used to have, Dr. Rimm says. “So we don’t want to be so quick to treat by ER-positive algorithms if the patient might actually be best treated as an ER-negative cancer. There is a grey zone near the threshold.” Among the expert panel members, “There was a lot of discussion about that,” he adds. “In the whole first recommendation and throughout the guideline document, there are nods to the fact that we need to be identifying patients who are ER-low but positive. Even though there are not a lot of patients in that category, clinicians should be aware of them.”

The guideline also recommends that the status of internal controls be reported for cases with zero percent to 10 percent staining, with a special comment included for those lacking internal controls.

Another concern the panel talked about was the possibility of assay drift, with increases in the numbers of ER-positive cancers over time, Dr. Rimm says. “In the 1980s, they used to do the dextran-coated charcoal assay, which was a biochemical assay that might have been pretty strict, and you needed a big chunk of tissue to do it. In the late 1990s, the standard was an ER greater than 10 percent or an H-score of about 75, so that was even more strict. We might have only had 70 percent or 75 percent of patients test ER-positive.”

“Then after we moved the cutpoint down to one percent,” he continues, “many labs switched antibodies. They went from the antibody clone 1D5 to the antibody clone SP1, which is more sensitive, so we then got even more patients who were in the ER-positive group because we kept changing the definition of what was ER-positive. We don’t really know how much assay drift may have occurred with these changes.” The recommendations therefore emphasize the importance of low-positive negative control tissue, such as tonsil, as well as good negative controls to ensure that assay drift is not occurring in the laboratory.

Fortunately, Dr. Allison says, it's rare to get an ER result in the one percent to 10 percent range. No more than two to four percent of cases tend to fall in that range. "So it's not as if suddenly a lot of patients will be having a challenging conversation with their oncologists on the pros and cons of endocrine therapy." The 2020 guideline update continues the 2010 guideline recommendation on such a discussion. "But we're bringing it more to the forefront as a bigger bullet point with a standardized comment and so on. Any time you have a really low result that is close to a threshold," she says, "more conversation is needed because the data is more limited."

Left unchanged in the guideline is the recommendation to continue using validated immunohistochemistry for primary screening for ER expression. "IHC is still the gold standard," Dr. Allison notes. "The big benefit of IHC is that you can visualize it. Because normal background breast expresses some hormone receptor, you can make sure you are looking at the invasive cancer and not hormone-expressing cells in the background. Ductal carcinoma in situ can also express hormone receptor and kind of mix in with your invasive cancer if you are grinding up tissue and running an assay. And that's how the old ligand-binding assay used to work, so you had those potential confounders. So IHC has an advantage in that you can look at it as an in situ assay, and visually know what you're scoring your assay for."

IHC does have drawbacks. "You have to have well-preserved tissue and consistency in staining, and your observers who are interpreting the assay have to be well trained and know how to score so that everybody scores the same," Dr. Allison notes. But while the expert panel considered other assay types, such as RNA-based tests like Oncotype DX and the Prosigna gene signature assay, and many RNA companies were listening carefully to see if RNA was going to be acceptable, Dr. Rimm says, "the committee didn't feel there was sufficient evidence to put one of those in a guideline."

Further research would be needed to justify use of tests about which there is provocative data in the field but not enough data to support a change in the guideline. "One of them would be the RNA tests," he suggests. "I wouldn't be surprised in the next set of guidelines if it's okay to use RNA, not just IHC, because RNA is more reproducible and less subjective. As we make precision medicine more and more precise, we need to squeeze out inter-operator variability or subjectivity. It's still pretty prominent in pathology, but we try to remove it where we can. And RNA tests, especially some of the closed-system RNA tests, have small inter-operator variability of zero or .001. So I wouldn't be surprised to see more research in that space."

However, disadvantages also exist in that these assays need fairly pure cancer samples to avoid contamination with other tissues that could influence results. "In fact, current data suggests that IHC low-positive cases are frequently negative by mRNA assay," Dr. Allison says. "So they are not currently recommended as 'tiebreakers' in IHC results close to the threshold. It is harder to know exactly what an mRNA assay is including in its results. Things like normal tissues and DCIS can influence things."

Standardization of evaluation of intensity is another area where further research will be important. "A nucleus that is low-intensity positive in one lab might be negative in another lab," Dr. Rimm says, "because there is still variation among autostainers. And research has shown there is some degree of variability when you quantitate autostainers from week to week and day to day. So, in the future—and the [CAP] IHC Committee is working on this—there will be ways of standardizing every test, as we do in laboratory medicine." Rather than the binary resulting standards now used in IHC, "I think we might see continuous standards introduced to increase the accuracy of ER and PgR, and probably HER2 as well as any IHC test that needs to be quantitative."

A second major change in the guideline update is the recommendation that laboratories adopt specific standard operating procedures to ensure validity of low ER-positive or ER-negative interpretations and results. "We want to ensure that a lab result that is a negative or a low positive is a valid and reproducible result, not a false-negative or falsely low," Dr. Allison says. "While many labs already have their own process to check for the quality of a result, such as checking internal and external controls, correlating with the histology, or getting a second read from another pathologist for results close to a threshold, we are recommending formalizing and standardizing these." An example standard operating procedure is provided in the guideline supplement.

Of concern to the expert panel were variations in results due to performance problems, especially close in the low positive or negative results, Dr. Rimm says. He coauthored a study, published Feb. 5, suggesting that some samples, when they are retested, turn out to be higher in ER expression (Caruana D, et al. *npj Breast Cancer*. 2020;6:5. doi.org/10.1038/s41523-020-0146-2).



Dr. Rimm

"In our studies, when you adjust for the intensity in normal breast cancer cells, some of these that were called one percent look like they might have been 10 percent or 20 percent. But it's just that the stainer was having a bad day and the stain was weak that day. That's something laboratories always have to be on guard for," Dr. Rimm says. "Pathologists are always told to look at a normal duct in the same specimen to make sure the normal ducts are roughly the same intensity. If they see normal ducts that are negative in the specimen, that's a big red flag that something is wrong. But if the stain intensity is low, this is a problem that labs are not yet equipped to deal with." That may change in the future, he says, when the guideline is updated again.

There is no laboratory-required quality control for the intensity of the stain, and that is something that is not addressed in the guideline update, Dr. Rimm says, although it is a discussion he and other members of the CAP Immunohistochemistry Committee are having, with the possibility of standardization being introduced in the next six to eight years. "I expect linear standardization as opposed to binary standardization to become a very important parameter for labs, and once those standardization rules are put into place, they will become part of the validation technology that labs will have to learn how to do and prepare for." For now, using good low positive and negative external on-slide controls is the recommended practice.

Stain intensity has generated many questions from commenters, Dr. Allison says. "The threshold for positive is set around percent of cells staining, but labs report both percent and intensity. And we still recommend that. The intensity of the stain, it was felt, would tell you a little more about how well your assay worked. If it was weaker than expected, maybe your stain didn't work well. If it's nice and strong, you're pretty confident in your results, as long as it's not looking like it's staining things you didn't expect it to. So it's really a quality control measure."

On the first draft of the laboratory procedure recommendation in the guideline update, a long standard operating procedure for cases with less than 10 percent or weak staining was proposed, Dr. Allison says. "When we put that out for public comment, we got a lot of feedback about how specific we were, and so we decided it would be better for labs to come up with their own standard operating procedures for how to handle the validity of results when they're negative or low positive."

As suggested guidance, the supplement to the update includes a sample SOP with a number of steps. "It calls for either a review by a second pathologist or, if you're using a well-validated digital image analysis platform, you can use that as a second review, plus a double-check on controls, making sure the assay works, considering a repeat of the test if there are no controls, and so on. We didn't want to get too prescriptive because there are so many different pathways and scenarios you could go down," Dr. Allison says. At Stanford, a reproducibility study found that the high positives and negatives (the zeros) were highly reproducible in their group of breast pathologists interpreting the stains, but the one percent to 10 percent group had much more variability. "And so now we have an internal policy that you have to do a second review by another pathologist when you get a one percent to 10 percent result, in addition to correlating with histology, checking internal and external controls, and correlating with any prior results on the patient."

Many labs probably already have some form of standard operating procedure, Dr. Allison says. "I think our making

a bigger statement about it is going to help ensure that labs are doing it and that they're thinking carefully about these results to make sure there's not a false-negative that's missed and that the low positives are at least a reproducible interpretive result between pathologists in your practice."

The guideline update recommends that patients with newly diagnosed ductal carcinoma in situ have ER testing, but it considers PgR testing optional. (In the 2010 guideline, there was limited guidance on testing DCIS.) PgR testing for invasive cancers was a somewhat controversial topic on the expert panel, but there was agreement that it adds prognostic value to the panel of tests performed in invasive breast cancers. PgR is not considered so much for therapy as for prognosis, Dr. Rimm says. "There was discussion of its being time to get rid of PgR, that it's not necessary, but others felt it's useful because of its prognostic value, so we kept it in place. I'm not sure how long it will survive, but it survived this round."

Another reason for its survival: "If you do have a PR-positive result in the setting of a concurrent ER-negative result," Dr. Allison says, "some folks felt that then PR might serve as a reasonable quality control measure to consider that maybe your ER stain didn't work. Basically, using it as a second test that would help avoid false-negative ERs in some instances." But PgR is not the primary screen for whether the patient will benefit from endocrine therapy, she says. "ER is. We stopped short of recommending the same standard operating procedures and reporting recommendations for PgR for that reason, although it is good practice to include quality control measures for every IHC stain. Labs have the option to use the same low-positive reporting category for PgR results in the one to 10 percent range, but they should not include the recommended comment that applies only to ER results in invasive cancers. But we were more focused on ensuring the validity of ER as a predictive marker, so that is where we focused."

On the whole, the quality of hormone receptor testing is very good and has improved, Dr. Allison says. "You can look at clinical trial data such as local versus central laboratory review where they have repeated the ER testing, and while the triple-negative trials were the ones that had the most discordance, probably because there were more low positives in this group, there isn't a concerning level of discordance between local and central labs as much as in the past." Proficiency testing data also indicate an improvement. "The CAP has seen a huge increase in the number of labs enrolled in CAP PT programs for hormone receptor testing, and the number of labs that pass has increased not just for CAP but internationally for programs like NordiQC and UK NEQAS."

How the ASCO/CAP guideline is applied will be up to laboratories and clinicians, Dr. Rimm says, although he believes both large and small labs will be interested in following the updated recommendations. "We know from the CAP Surveys that about 1,400 labs around the world do ER testing and maybe 800 or 1,000 are in the U.S. And there aren't 1,000 big labs in the U.S., so we know it's not just big labs that are doing it."

"But a lot of the rest of the world also looks to the U.S. to see what guidelines we're using and then they adopt them." The first ER/PgR guideline published in 2010 "dramatically improved the quality of our testing for patients," Dr. Rimm says, making it less subjective and more accurate. "So even though we're not finished yet, the guideline has made a big difference, and we believe we need to continue that. The CAP will revisit and update the guideline as new technologies come out and as things get better, because the College realizes how important guidelines are in improving patient care."□

Anne Paxton is a writer and attorney in Seattle.