Interobserver Concordance in Implementing the 2010 ASCO/CAP Recommendations for Reporting ER in Breast Carcinomas

A Demonstration of the Difficulties of Consistently Reporting Low Levels of ER Expression by Manual Quantification

Emily S. Reisenbichler, MD,[1] Susan C. Lester, MD, PhD,[1] Andrea L. Richardson, MD, PhD,[1] Deborah A. Dillon, MD,[1] Amy Ly, MD,[2] and Jane E. Brock, MBBS, PhD[1]

Department of Pathology, [1]Brigham and Women's Hospital, Boston, MA, and [2]Massachusetts General Hospital, Boston.

Key Words: Estrogen receptor; Immunohistochemistry; Concordance; Breast carcinoma

DOI: 10.1309/AJCP1RF9FUIZRDPI

ABSTRACT

Objectives: *Endocrine therapy reduces recurrence risk by 30% to 50% in estrogen receptor (ER)–positive breast cancer. The ER-positive threshold recommended by the American Society of Clinical Oncology/College of American Pathologists is 1% based on studies using the ER-6F11 antibody. ER-SP1 antibody has a higher sensitivity and is more widely used.*

Methods: *We report interobserver concordance manually measuring ER in 264 breast cancers using ER-SP1 and 1D5 and 2 scoring methods (H-score and Allred score).*

Results: *With both antibodies, 3% to 4% of cases have a low level of ER expression (1%-10%), more than previously reported (<1%). We find a high level of paired observer concordance with both antibodies and scoring methods (*κ *= 0.892-0.943) with no significant difference with method of scoring. Despite excellent concordance, positive/negative discordance was almost 5% among 3 observers using either antibody, an underappreciated clinically significant rate.*

Conclusions: *Discordance overwhelmingly reflected differing opinions recording the proportion of tumor cells positive with low levels of expression (<10% staining; 12/13 cases).*

Estrogen receptor (ER) and progesterone receptor (PR) expression levels in breast cancer guides hormone therapy treatment decisions. Endocrine therapy in ER-positive tumors reduces the overall recurrence risk by half in the first 4 years of therapy and then reduces the risk by one third 5 to 10 years after diagnosis. There is no similar benefit in ER-negative tumors. Early testing of ER used ligand-binding assays (LBAs) to quantify ER protein content in breast tumors, establishing the positive threshold in LBAs based on the odds of response to endocrine therapy in the metastatic setting using data collected before 1975.[1] Improved patient outcomes with hormone-targeted therapy using LBAs was observed to lie at the positive threshold for assay detection of 3 fmol/mg or more, although a more robust, statistically significant threshold and a common international laboratory standard for positive ER by LBA throughout the 1990s was 10 fmol/mg or more.[2] Some groups even used 20 fmol/mg or more as a positive cutoff.[2,3] To determine the corresponding immunohistochemical staining threshold for response to endocrine therapy with improved survival, both mouse monoclonal antibodies (mMabs) 1D5 and 6F11 were evaluated and scored using the Allred method, which combines the proportion of positive-staining tumor cells and the intensity of staining to give a score between 0 and 8.[2,4] Harvey et al[2] determined that the minimal Allred score of 3, seen in 6% of cases evaluated, was a highly significant cutoff point corresponding to improved disease-free survival and overall survival. Despite the data of Harvey et al, the ER-positive threshold with immunohistochemistry used in different clinical trials and in routine reporting varied among institutes both in the United States and around the world, ranging from any positive staining up to 10%.

In 2010, a consensus American Society of Clinical Oncology (ASCO)/College of American Pathologists (CAP) meeting of expert panelists, including Allred, aimed at standardizing procedures for testing and reporting prognostic markers in breast carcinomas. After that meeting, the consensus threshold for reporting ER as positive was set at 1% staining because an Allred score of 3 can be seen with as few as 1% to 10% weakly staining cells.[2,5] An Allred score of 3 can be achieved by either 1% to 10% of tumor cells showing weak staining or less than 1% of cells showing moderate to strong staining. However, in the Harvey et al[2] study of cases with Allred score 3 (n = 117, 6%), most showed 1% to 10% weak staining. The ASCO/CAP panel decided that a threshold of 1% would capture the very small number of tumors with low ER expression that are also endocrine sensitive.

A recently published study looked at survival in patients with tumors showing low levels of ER expression (between 0 and 10%).[6] This retrospective study reported overall survival and recurrence-free survival in a large number of low-level ER-expressing tumors, including a subset of patients who received endocrine therapy, albeit with limited follow-up. Raghav et al[6] found no significant recurrence-free survival advantage at 3 years in patients with tumors that stained less than 1% (n = 897) or between 1% and 5% (n = 241), while those with ER staining between 6% and 10% (n = 119) showed a recurrence-free survival advantage trend. The authors could not demonstrate that addition of endocrine therapy significantly improved overall survival in the ER-positive subgroup with 1% to 10% staining (n = 81) compared with those showing less than 1% staining (n = 37); however, a trend toward recurrence-free survival was noted with increasing ER expression. They concluded that there is no significant overall survival or recurrence-free survival difference between tumors with less than 1% staining and those with 1% to 5% staining regardless of whether the patient receives endocrine therapy.

Before the publication of the 2010 guidelines, ER reporting methods varied from simple positive/negative reports to more complex evaluations including intensity and percentage of tumor cell staining. Based on several publications demonstrating that increasing expression of ER correlates with greater hormone therapy sensitivity, the 2010 ASCO/CAP guidelines recommend reporting percentage and average intensity of expression because this may provide valuable predictive and prognostic information to inform treatment strategies.[3,7-9] Percentages can be estimated or quantified either manually or by using image analysis, and intensity is reported as weak, moderate, or strong. ER expression, although considered a continuous variable, is bimodal in distribution, with Collins et al[10] reporting that more than 99% of tumors are either ER negative (Allred scores of 0 or 2) or strongly positive (Allred scores of 7 or 8) using mMabs. Numerous scoring methods have been reported. The most complex one, the H-score, attempts to capture the full range in percentage and intensity of staining seen in tumors rather than just the average intensity captured by the Allred score. The H-score ranges between 1 and 300, more heavily weighting stronger-intensity staining than weaker-intensity staining. A score of 1 or more corresponds to at least 1% weak expression and is considered positive. A recent study using the H-score determined that low levels of expression (defined as H-score ≤50) have lower overall and disease-free survival when treated with only endocrine therapy.[11]

In the last 3 years, the use of SP1 rabbit monoclonal antibody (rMab) has replaced mMabs in many laboratories across the United States. SP1 is superior to 1D5 in sensitivity and robustness, and a small but significant number of tumors (0.5%-8%) are classified as SP1-ER positive but 1D5 negative.[12,13] Such tumors appear to be hormone sensitive and cluster with outcome in the ER-positive group.

With the 2010 ASCO/CAP recommendations in mind and the publication by Raghav et al,[6] which highlights outcome in a large series of tumors with low ER expression, we aimed to review our concordance as a group of dedicated breast pathologists at reporting the intensity and percentage of ER expression using 2 different antibodies and 2 different scoring methods, the Allred and the more complex H-score methods. We wanted to determine the spread of ER expression with particular interest in the subset of tumors with low levels of ER expression and to establish our concordance at reporting ER in borderline threshold cases.

In our previous study reporting the comparison of ER-SP1 and ER-1D5, we did not collect either Allred/H-score data or interobserver variability.[13] In that study (with data collected before the release of the 2010 guidelines), tumors were routinely stained and reported simply as positive (>10%), positive-low (>1%-10%), or negative (<1%) by 1 of 4 breast pathologists. Only cases that showed a discrepancy between the 2 different antibodies, as noted by one observer, were subsequently reviewed by a second pathologist. Using these cases that were stained routinely for clinical use, we re-reviewed a subset of consecutive tumors from this data set, evaluating both Allred and H-score, to determine our interobserver variability in reporting ER across a full range of tumors, including those classified as ER negative and ER positive.

Materials and Methods

This study was approved by the Brigham and Women's Hospital institutional review board.

Cases of primary invasive breast carcinoma requiring immunohistochemical staining for routine clinical care were stained for ER using the primary antibodies ER-1D5 (mMab) and ER-SP1 (rMab). Four-micrometer-thick sections were

DOI: 10.1309/AJCP1RF9FUIZRDPI

immunostained according to the manufacturer's recommendations using the EnVision+ System-HRP (DAKO, Carpinteria, CA). In detail, slides were baked at 37°C overnight, then deparaffinized and rehydrated (100% xylene 4 times for 3 minutes each, 100% ethanol 4 times for 3 minutes each, and running water for 5 minutes). Endogenous peroxidase activity was blocked with 3% hydrogen peroxide in methanol for 10 minutes and washed under running water for 5 minutes. For heat-induced epitope retrieval, slides were placed in 10 mmol/L citrate buffer at a pH of 6.0 (Target Retrieval Solution, S1699, DAKO) and then pressure cooked (Biocare Medical, Concord, CA) at 122°C to between 14 and 17 psi. Each cycle lasted, on average, 45 minutes and had a cooldown period of approximately 20 minutes. Immunostains were performed on an automated instrument (DAKO Autostainer Plus, DAKO). A range of titers was tested for both antibodies, and titers were calibrated using internal positive control staining of normal breast epithelium. Primary antibodies ER-1D5 (dilution 1:100; DAKO), ER-SP1 (dilution 1:200; Lab Vision, Fremont, CA), and PR (PgR636, dilution 1:200; DAKO) were incubated for 40 minutes at room temperature. A DAKO polymer secondary antibody system was used (Envision Poly K4011 for the SP1 RabMab and Envision Mono K4007 for 1D5 and PR MMabs) and incubated for 30 minutes in a humid chamber at room temperature. Sections were developed using 3,3′-diaminobenzidine (Sigma Chemical, St Louis, MO) as a substrate and counterstained with Mayer hematoxylin. External positive controls were also run. The studies comparing ER antibodies were performed on the same day for a given case.

Slides were scored for ER positivity by 3 breast pathologists (E.S.R., A.L., and J.E.B.), and the combination of percentage of positive cells and intensity of expression was recorded to determine Allred and H-scores. Positive vs negative discordance cases among the original 3 observers were reviewed by 3 additional breast pathologists at our institute (S.C.L., D.A.D., and A.L.R.). An Allred score was calculated by scoring the proportion of positive cells on a scale of 0 to 5 (with 0, none; 1, <1/100; 2, 1/100 to 1/10; 3, >1/10 to 1/3; 4, >1/3 to 2/3; and 5, >2/3) combined with staining intensity scored on a scale of 0 to 3 (0, none; 1, weak; 2, intermediate; and 3, strong). The proportion and intensity were summed to produce total scores of 0 or 2 through 8. For H-score, evaluations were recorded as percentages of positively stained tumor cells in each of the 4 intensity categories denoted as zero (no staining), 1+ (weak but detectable), 2+ (moderately distinct), and 3+ (strong). For each tumor, a value was derived by summing the percentages of cells that stained at each intensity multiplied by the weighted intensity of staining: H-score = (0 × % at 0) + (1 × % at 1+) + (2 × % at 2+) + (3 × % at 3+). This score produces a continuous variable that ranges from 0 to 300.

Interobserver agreement was assessed using a nonweighted (for positive vs negative categories) and weighted (for multiple scoring categories) Fleiss κ statistic, performed using Stata 11.0 statistical software (StataCorp, College Station, TX). The student *t* test was used to compare continuous variables. κ agreement was grouped according to the following previously published categories[14]: 0.00 to 0.20, poor agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1.00, excellent agreement.

Results

Two-hundred sixty-four consecutive cases of invasive carcinoma were evaluated. Overall, 79% of cases were ER positive with ER-SP1. Allred scores ranged from 0 to 8 using both antibodies, with mean scores of 5.8, 5.9, and 6.0 for each respective observer with SP1 and 5.8, 6.0, and 5.9 for 1D5 **❚Figure 1A❚**. Mean H-scores were 172, 184, and 171 for SP1 and 172, 188, and 169 for 1D5, with scores ranging from 0 to 300 in both antibody groups **❚Figure 1B❚**. Cases scoring either Allred 0/2 (ER negative) or Allred 7/8 (ER strongly positive) ranged from 88% to 90% with SP1 and 90% to 92% with 1D5 **❚Figure 2❚**. The 2 antibodies resulted in a difference in Allred score of 1 point or less in 254 (96%), 250 (95%), and 252 (95%) cases for each respective observer and less than a 50-point intraobserver difference between H-scores in 232 (88%), 242 (92%), and 232 (88%) cases. The majority of cases (94% with SP1 and 93% with 1D5) showed no more than a 100-point H-score variability among all 3 observers. Only 1 point or less separated Allred scores by all observers in 92% and 91% of cases with SP1 and 1D5, respectively **❚Table 1❚**.

Pairwise κ agreement among observers ranged from 0.852 to 0.924 with SP1 and 0.824 to 0.932 with 1D5 when stratifying individual case scores into less than 1, 1 to 50, 51 to 100, 101 to 200, and 201 to 250 for H-score and 0 to 2, 3 to 4, 5 to 6, and 7 to 8 for Allred score **❚Table 2❚**. We found improved agreement, ranging from 0.863 to 0.924 with SP1 and 0.892 to 0.943 with 1D5 when dividing cases as either positive (Allred score >2; H-score ≥1) or negative (Allred score 0, 2; H-score <1). Overall agreement among the 3 observers for determining positive vs negative was similar regardless of antibody (Table 2).

All 6 observers found a higher rate of positive cases with SP1 than with 1D5. Thirteen cases (4.9%) showed discrepant positive/negative results among the original 3 observers with 1 or both antibodies. The staining of both antibodies in these discrepant cases was then evaluated by 3 additional observers **❚Table 3❚**. Of these 13 cases, 11 showed the interobserver H-scores to be discrepant by less than 10

DOI: 10.1309/AJCP1RF9FUIZRDPI

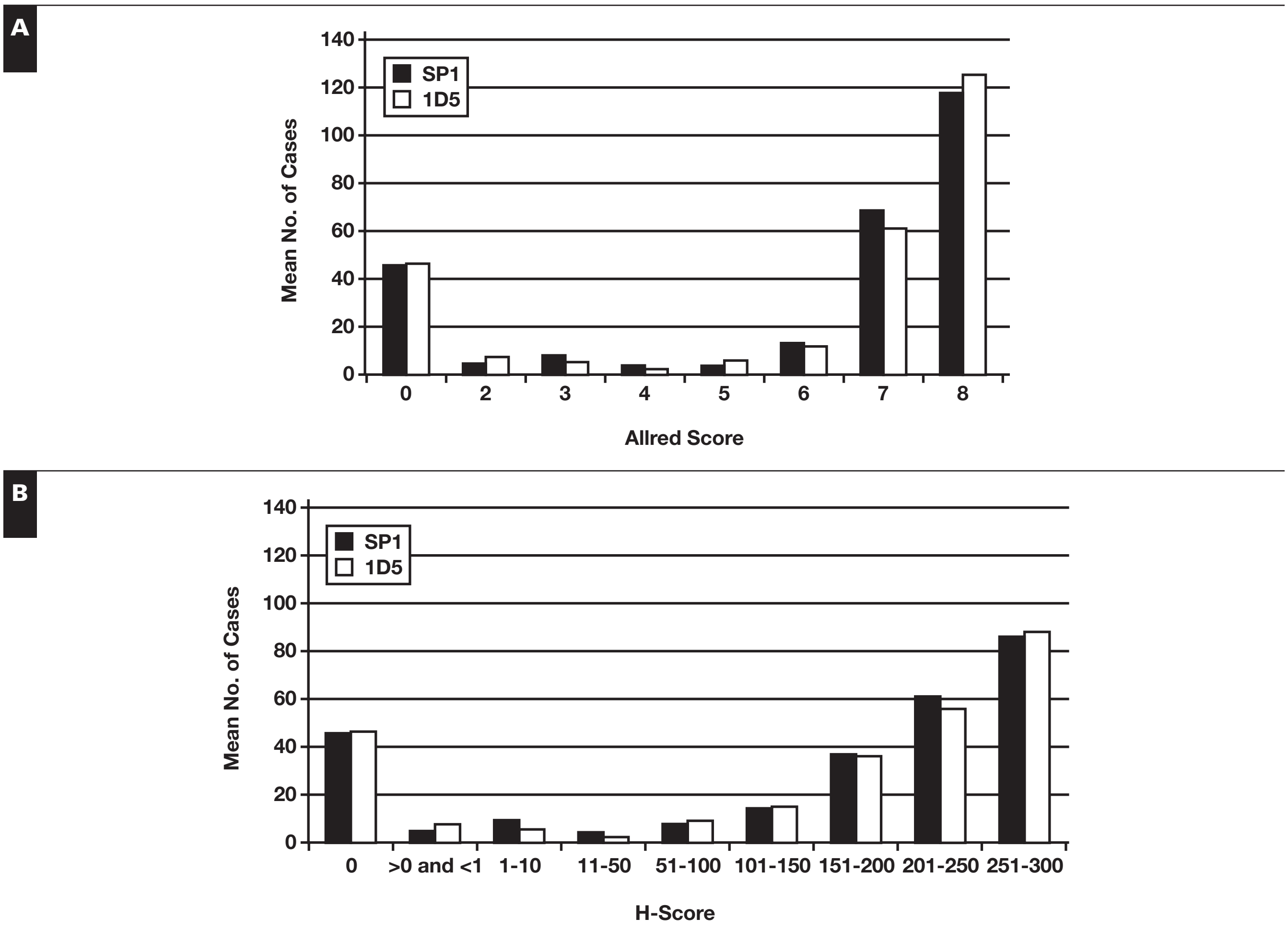

Figure 1 Frequency distribution of the mean Allred scores (**A**) and H-scores (**B**) for 3 observers.

points with both SP1 and 1D5. For 2 cases, the interobserver H-scores were discrepant by up to 49 points with SP1 and 50.5 points with 1D5. Allred scores were discrepant by 1 to 2 points in 11 of 13 cases and in 2 cases by up to 4 and 5 points with SP1 and 1D5, respectively (data not shown). All of these cases were also HER2/neu negative (data not shown), and 10 of 13 had been classified as ER negative when originally reviewed by a single pathologist for routine clinical care. Four were reported as showing less than 1% positive staining, and 3 were classified as positive-low (1%-10% staining). All tumors were poorly differentiated ductal carcinomas, and patients received chemotherapy.

On average, 3% (range, 2.6%-3.8%) of tumors showed ER expression in the 1% to 10% range, with more than 75% of these low expression cases falling in the 1% to 5% range for all observers with both antibodies **Table 4**. Using an ER expression cutoff of 6% or more resulted in fewer cases with a positive/negative discrepancy (7 cases; 2.7%) among all 3 observers **Table 5**. Paired observer concordance is higher using a cutoff of 6% or more rather than a cutoff of 1% or more for any pair **Table 6**.

Discussion

Despite excellent paired observer concordance at reporting ER regardless of antibody type and scoring method, we found that almost 5% of cases (13/264) evaluated by experienced and dedicated breast pathologists had a low level of ER expression, which proved difficult to classify as ER positive or negative using manual quantification of percentage and intensity of staining. Our series of routinely stained cases appears representative of the typical distribution of invasive breast carcinomas, with 79% of tumors being ER positive and 14% (37/264) classified as triple negative (ER negative, PR negative, and HER2/neu negative; PR and HER2/neu data not shown).

Collins et al[10] previously reported that ER staining demonstrates a bimodal distribution, with 99% of cases being negative (Allred 0 or 2) or positive (Allred 7 or 8). They reported tumors in the 1% to 10% category as very rare, with the remaining 1% of cases falling in the 20% to 60% category. By contrast, we find 94% of cases either negative or strongly positive, and for each of our observers, approximately 6% of cases fall outside the Allred 0 to 2 or 7 to 8 bimodal categories

DOI: 10.1309/AJCP1RF9FUIZRDPI

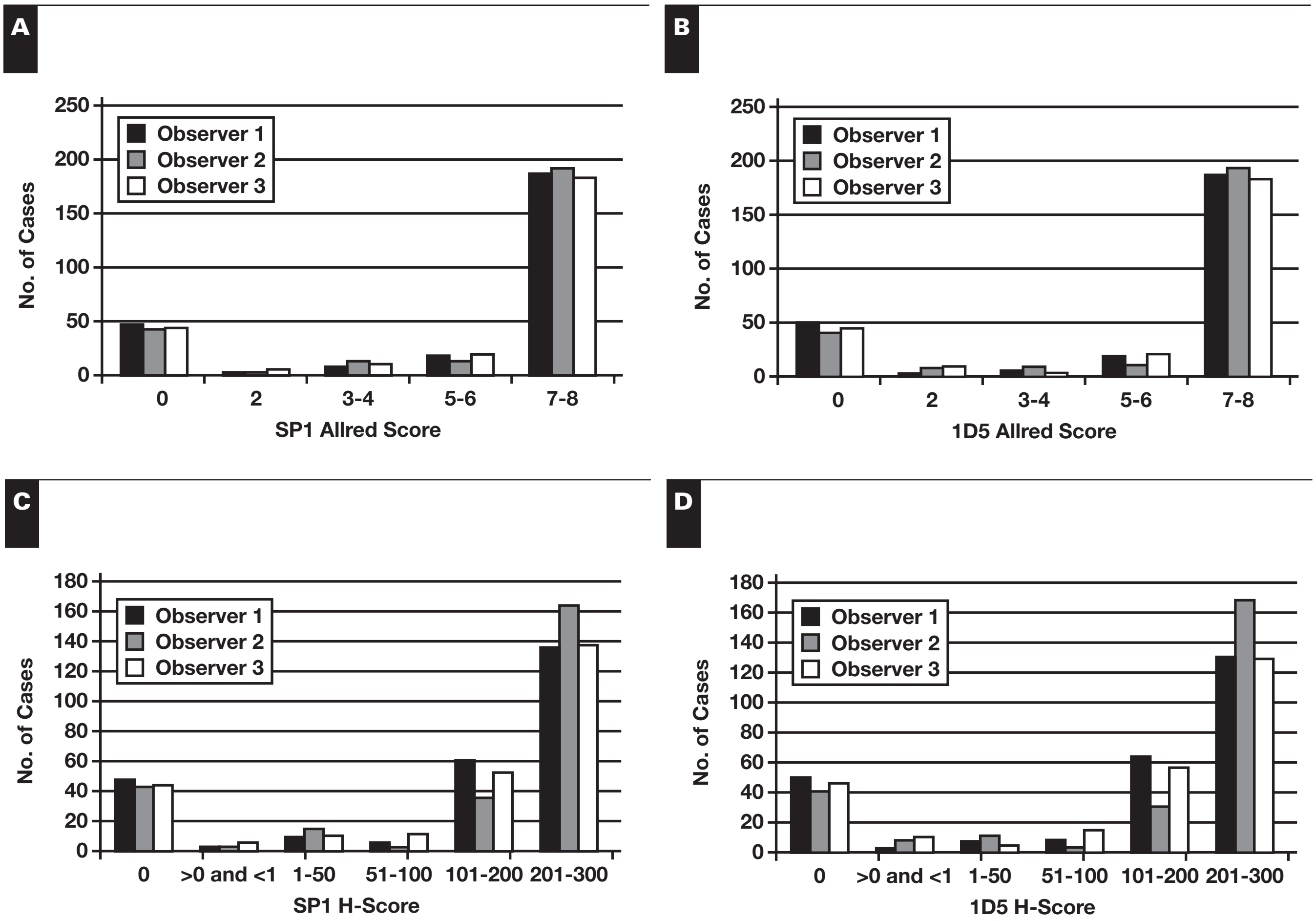

Figure 2 Frequency distribution of the Allred scores (**A**, **B**) and H-scores (**C**, **D**) with the SP1 (**A**, **C**) and 1D5 (**B**, **D**) antibodies among 3 observers.

Table 1
Interobserver Agreement

	SP1, No. (%) of Cases			1D5, No. (%) of Cases		
Observers	**Full**	**± 1**	**± 2**	**Full**	**± 1**	**± 2**
Allred score						
1 and 2	197 (75)	252 (95)	258 (98)	191 (72)	245 (93)	256 (97)
2 and 3	177 (67)	254 (96)	262 (99)	176 (67)	247 (94)	258 (98)
1 and 3	175 (66)	251 (95)	260 (98)	168 (64)	252 (95)	262 (99)
All 3	146 (55)	244 (92)	257 (97)	141 (53)	239 (91)	255 (97)
H-score						
1 and 2	59 (22)	234 (89)	261 (99)	57 (22)	230 (87)	263 (99)
2 and 3	53 (20)	205 (78)	252 (95)	51 (19)	207 (78)	249 (94)
1 and 3	56 (21)	220 (83)	259 (98)	58 (22)	221 (84)	258 (98)
All 3	42 (16)	176 (67)	248 (94)	39 (15)	179 (68)	246 (93)

(Figure 2). Our study found, on average, that 3% of cases had low levels of ER expression (1%-10%). Published studies report variable percentages of tumors showing low levels of ER expression. Harvey et al[2] reported 6% of cases with Allred scores of 3 using ER-6F11 (1%-10% weak staining). Diaz et al[15] reported a series of 70 invasive carcinomas evaluated with ER-6F11 and found 3% of cases with low levels of ER expression (1%-9%). Using ER-1D5, Nadji et al[16] reported unquantified "focal" staining in 8% of tumors evaluated, but they attributed the lack of diffuse staining to poor tissue fixation or necrosis rather than a true low level of ER expression. Raghav et al[6] set the ER-negative threshold in their study

DOI: 10.1309/AJCP1RF9FUIZRDPI

Table 2
κ Agreement Among Observers

	H-Score (5 categories)		Allred Score (4 categories)		Positive vs Negative[a]	
	Obs 2	Obs 3	Obs 2	Obs 3	Obs 2	Obs 3
SP1 antibody						
Obs 1	0.861	0.852	0.904	0.924	0.863	0.916
Obs 2		0.854		0.918		0.924
1D5 antibody						
Obs 1	0.853	0.840	0.919	0.932	0.892	0.943
Obs 2		0.824		0.906		0.906

Obs, observer.
[a] Negative, <1% staining, any intensity; positive, ≥1% staining, any intensity.

at 10% or less and had 40% (360/897) of cases with well-documented low levels of ER expression between 1% and 10%; however, we are not aware of the proportion of total ER-positive tumors this might represent.

Our study showed no overall skew toward higher scores with SP1 across the range of ER expression with multiple observers, although as previously reported, SP1 had a higher sensitivity with more cases staining positive (79%) than with 1D5 (78%). No statistically significant interobserver difference was seen between the 2 antibodies using either the Allred score or H-score across all tumors reviewed. A higher correlation was seen with the Allred scoring method than with H-score when categorized into 4 or 5 scoring subgroups, respectively. Although the H-score takes into account the variation in ER staining intensity commonly seen in tumors, weighting stronger staining more highly, it is more complicated and time consuming. Its greater complexity is reflected in a lower pair correlation than the Allred score.

Few previous studies have specifically addressed concordance at low levels of ER expression using a manual count. In the studies by Harvey et al[2] and Allred et al,[4] which used 6F11 mouse monoclonal antibody, 11% of cases were initially independently evaluated by 2 observers and had a reported concordance rate of 0.87. They reported a 1% positive/negative discrepancy (2 of 220 cases). Because of this high concordance rate, the remaining cases in the study were evaluated by only 1 observer. Using the same Allred scoring method and different antibodies (ER-1D5 and ER-SP1), we have similarly high concordance rates for the designation of ER positive vs negative, with κ scores ranging from 0.892 to 0.943 for ER-1D5 and 0.863 to 0.924 for ER-SP1 between any paired

Table 3
Discrepant Positive vs Negative Cases Using 1% Cutoff[a]

	SP1 Antibody						1D5 Antibody					
Case No.	Obs 1	Obs 2	Obs 3	Obs 4	Obs 5	Obs 6	Obs 1	Obs 2	Obs 3	Obs 4	Obs 5	Obs 6
1	+	+	–	+	+	–	+	–	–	–	–	–
2	+	–	–	–	–	–	+	–	–	–	–	–
3	–	+.	+	–	–	–	–	+	–	–	–	–
4	–	+	+	+	+	–	–	+	–	–	–	–
5	+	+	+	+	+	+	–	+	+	+	+	–
6	–	+	–	–	+	–	–	+	–	–	+	–
7	–	+	–	+	–	–	–	+	–	+	+	–
8	–	+	–	–	+	–	–	+	–	–	–	–
9	–	+	+	–	+	–	–	–	–	–	+	–
10	+	–	+	–	+	–	–	–	–	+	+	–
11	–	+	–	–	–	–	–	–	–	–	–	–
12	–	–	–	–	–	–	–	+	–	–	–	–
13	+	+	+	+	+	+	+	+	–	–	+	+
No. (%) of positive cases	5 (38)	10 (77)	6 (46)	5 (38)	8 (62)	2 (15)	3 (23)	8 (62)	1 (8)	3 (23)	6 (46)	1 (8)

Obs, observer.
[a] –, Negative, <1% staining, any intensity; +, positive, ≥1% staining, any intensity.

Table 4
Interpretation of Estrogen Receptor Expression at Low Levels by 3 Observers

	SP1 Antibody, No. (%)				1D5 Antibody, No. (%)			
Observer	<1%	1%-5%	6%-10%	>10%	<1%	1%-5%	6%-10%	>10%
1	51 (19)	6 (2.3)	1 (0.4)	206 (78)	54 (20)	5 (1.9)	2 (0.8)	203 (77)
2	48 (18)	9 (3.4)	1 (0.4)	206 (78)	50 (19)	8 (3)	1 (0.4)	205 (78)
3	51 (19)	7 (2.7)	1 (0.4)	205 (78)	57 (22)	1 (0.4)	1 (0.4)	205 (78)

DOI: 10.1309/AJCP1RF9FUIZRDPI

Table 5
Discrepant Positive vs Negative Cases Using 6% Cutoff[a]

	SP1 Antibody			1D5 Antibody		
Case No.	**Obs 1**	**Obs 2**	**Obs 3**	**Obs 1**	**Obs 2**	**Obs 3**
1	+	+	+	−	+	+
2	−	+	−	−	−	−
3	+	+	−	−	−	−
4	−	+	−	−	−	−
5	−	−	−	−	+	−
6	+	−	+	+	+	+
7	+	−	+	+	−	+
No. (%) of positive cases	4 (57)	4 (57)	3 (43)	2 (29)	3 (43)	3 (43)

Obs, observer.
[a] –, Negative, <6% positive staining, any intensity; +, positive, ≥6% positive staining, any intensity.

observers, but a higher positive/negative discrepancy rate among multiple observers for both antibodies. A possible similarly high discrepancy rate was reported by Barnes et al[3] using ER-1D5 and comparing immunohistochemical scoring methods with LBA data. They reported discordance at less than 5% between 2 observers, but the precise number is not clarified further or expanded upon in their publication. Diaz et al[15] compared manual reading by 3 observers using image analysis with ER-6F11 in 70 cases of invasive carcinoma, but they combined the mean percentage staining obtained by the 3 observers to compare findings with image analysis rather than evaluating interobserver variability. Another study also commented on the presence of discordance in reading ER staining obtained by 2 observers but did not specify the discordance rate; in that study, observers resolved discordances by reviewing cases together.[17]

Our discrepant cases were typically a difference of opinion among all observers when calculating overall proportion of tumor cells in the 0 to 5% range (12/13 cases). In the remaining case, staining with 1D5 produced a disagreement evenly split among the 6 observers as to whether a faint nuclear blush seen in a larger number of tumor cells (30%-50%) was "real" weak staining and of clinical significance (the corresponding SP1 stain was considered positive by all observers). The ASCO/CAP panel recommends that all tumor-containing areas of the tissue be evaluated and the percentage arrived at by estimation or quantification either manually by counting cells or by image analysis.[5] We do not use image analysis in our ER evaluation. The ASCO/CAP panel cited controversy about how and whether image analysis should be implemented. Of note, studies comparing ER and Ki-67 immunohistochemical scores in breast carcinomas have documented that image analysis typically scores tumor cell percentage markedly lower than a manual count, and with Ki-67 this is calculated to be by a factor of 2.5.[15,18-20] Image-guided ER analysis is unlikely to be the current solution to this difficulty in classifying low-level ER expression.

The appropriate positive threshold for ER has been debated many times, not least by the ASCO/CAP panel. Most recently, Raghav et al[6] could not demonstrate any 3-year survival advantage in a large subset of tumors with low ER expression (1%-5%; n = 241) compared with tumors with less than 1% expression (n = 897) either with or without endocrine therapy. They did report a trend for a recurrence-free survival advantage with tumors in the 6% to 10% range of staining (n = 119). In the study by Raghav et al,[6] ER results were based on central review, but the authors do not comment on whether their results arise from a single observer reviewing all cases or if any interobserver discrepancy existed in the reporting of results in the case of multiple observers. Although some might advocate raising the threshold to 6% or higher based on the data of Raghav et al,[6] doing so will not completely remove the discrepancy problems highlighted in our study. When we use 6% or higher as a threshold, we almost halve (to 2.7%), but do not eliminate, discordance rates in our series. Similar to discrepancies seen around the 1% threshold, discrepancies at the 6% threshold typically reflect differences in opinion regarding the proportion of tumor cells positive with low levels of expression.

This study serves to highlight the hitherto underappreciated prevalence of tumors with a low level of ER expression that can be difficult to classify. This study should heighten the awareness of pathologists and oncologists of the potential for observer discordance in reporting low levels of expression. All can agree that the positive cutoff should include all women who have the chance of responding to endocrine therapy even

Table 6
κ Agreement Among Observers at Different Positive vs Negative Cutoffs

	SP1 Antibody				1D5 Antibody			
	1% (pos/neg)		6% (pos/neg)		1% (pos/neg)		6% (pos/neg)	
	Obs 2	**Obs 3**	**Obs 2**	**Obs 3**	**Obs 2**	**Obs 3**	**Obs 2**	**Obs 3**
Obs 1	0.863	0.916	0.955	0.989	0.892	0.943	0.967	0.989
Obs 2		0.924		0.944		0.906		0.978

neg, negative; Obs, observer; pos, positive.

DOI: 10.1309/AJCP1RF9FUIZRDPI

though we know the response rate is unlikely to be high at low levels of expression. A threshold of 1% or higher achieves this, but so might a threshold of more than 5%. Pathologists face diagnostic dilemmas daily in the reporting of borderline cases and this study illustrates that experienced observers can disagree and may have biases.

Address reprint requests to Dr Reisenbichler: Dept of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, MCN C-3321, 1161 21st Ave S, Nashville, TN 37232;Emily.s.reisenbichler@vanderbilt.edu.

References

1. McGuire W, Carbone PP, Vollmer EP, eds. *Estrogen Receptors in Human Breast Cancer*. New York, NY: Raven Press; 1975.
2. Harvey JM, Clark GM, Osborne CK, et al. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol*. 1999;17:1474-1481.
3. Barnes DM, Harris WH, Smith P, et al. Immunohistochemical determination of oestrogen receptor: comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. *Br J Cancer*. 1996;74:1445-1451.
4. Allred DC, Harvey JM, Berardo M, et al. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol*. 1998;11:155-168.
5. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch Pathol Lab Med*. 2010;134:e48-e72.
6. Raghav KP, Hernandez-Aya LF, Lei X, et al. Impact of low estrogen/progesterone receptor expression on survival outcomes in breast cancers previously classified as triple negative breast cancers. *Cancer*. 2012;118:1498-1506.
7. Elledge RM, Green S, Pugh R, et al. Estrogen receptor (ER) and progesterone receptor (PgR), by ligand-binding assay compared with ER, PgR and pS2, by immuno-histochemistry in predicting response to tamoxifen in metastatic breast cancer: a Southwest Oncology Group Study. *Int J Cancer*. 2000;89:111-117.
8. Dowsett M, Allred C, Knox J, et al. Relationship between quantitative estrogen and progesterone receptor expression and human epidermal growth factor receptor 2 (HER-2) status with recurrence in the Arimidex, Tamoxifen Alone or in Combination trial. *J Clin Oncol*. 2008;26:1059-1065.
9. Dowsett M, Salter J, Zabaglo L, et al. Predictive algorithms for adjuvant therapy: TransATAC. *Steroids*. 2011;76:777-780.
10. Collins LC, Botero ML, Schnitt SJ. Bimodal frequency distribution of estrogen receptor immunohistochemical staining results in breast cancer: an analysis of 825 cases. *Am J Clin Pathol*. 2005;123:16-20.
11. Morgan DA, Refalo NA, Cheung KL. Strength of ER-positivity in relation to survival in ER-positive breast cancer treated by adjuvant tamoxifen as sole systemic therapy. *Breast*. 2011;20:215-219.
12. Cheang MC, Treaba DO, Speers CH, et al. Immunohistochemical detection using the new rabbit monoclonal antibody SP1 of estrogen receptor in breast cancer is superior to mouse monoclonal antibody 1D5 in predicting survival. *J Clin Oncol*. 2006;24:5637-5644.
13. Brock JE, Hornick JL, Richardson AL, et al. A comparison of estrogen receptor SP1 and 1D5 monoclonal antibodies in routine clinical use reveals similar staining results. *Am J Clin Pathol*. 2009;132:396-401.
14. Collins LC, Connolly JL, Page DL, et al. Diagnostic agreement in the evaluation of image-guided breast core needle biopsies: results from a randomized clinical trial. *Am J Surg Pathol*. 2004;28:126-131.
15. Diaz LK, Sahin A, Sneige N. Interobserver agreement for estrogen receptor immunohistochemical analysis in breast cancer: a comparison of manual and computer-assisted scoring methods. *Ann Diagn Pathol*. 2004;8:23-27.
16. Nadji M, Gomez-Fernandez C, Ganjei-Azar P, et al. Immunohistochemistry of estrogen and progesterone receptors reconsidered: experience with 5,993 breast cancers. *Am J Clin Pathol*. 2005;123:21-27.
17. Yamashita H, Yando Y, Nishio M, et al. Immunohistochemical evaluation of hormone receptor status for predicting response to endocrine therapy in metastatic breast cancer. *Breast Cancer*. 2006;13:74-83.
18. Barton S, Zabaglo L, A'Hern R, et al. Assessment of the contribution of the IHC4+C score to decision making in clinical practice in early breast cancer. *Br J Cancer*. 2012;106:1760-1765.
19. Cuzick J, Dowsett M, Pineda S, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol*. 2011;29:4273-4278.
20. Gokhale S, Rosen D, Sneige N, et al. Assessment of two automated imaging systems in evaluating estrogen receptor status in breast carcinoma. *Appl Immunohistochem Mol Morphol*. 2007;15:451-455.

DOI: 10.1309/AJCP1RF9FUIZRDPI