

Standardization of Estrogen Receptor Measurement in Breast Cancer Suggests False-Negative Results Are a Function of Threshold Intensity Rather Than Percentage of Positive Cells

Allison W. Welsh, Christopher B. Moeder, Sudha Kumar, Peter Gershkovich, Elaine T. Alarid, Malini Harigopal, Bruce G. Haffty, and David L. Rimm

See accompanying editorial on page 2955

From Yale University School of Medicine, New Haven, CT; McArdle Laboratory for Cancer Research, University of Wisconsin–Madison, Madison, WI; and The Cancer Institute of New Jersey, New Brunswick, NJ.

Submitted September 28, 2010; accepted March 7, 2011; published online ahead of print at www.jco.org on June 27, 2011.

Supported by Grant No. NIH R33 CA 106709 from the National Institutes of Health (D.L.R.) and a US Army predoctoral fellowship (A.W.W.).

Presented in part at the 32nd Annual San Antonio Breast Cancer Symposium, December 9-13, 2009, San Antonio, TX.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: David L. Rimm, MD, PhD, Department of Pathology, BML 116, Yale University School of Medicine, 310 Cedar St, PO Box 208023, New Haven, CT 06520-8023; e-mail: david.rimm@yale.edu.

© 2011 by American Society of Clinical Oncology

0732-183X/11/2922-2978/\$20.00

DOI: 10.1200/JCO.2010.32.9706

A B S T R A C T

Purpose

Recent misclassification (false negative) incidents have raised awareness concerning limitations of immunohistochemistry (IHC) in assessment of estrogen receptor (ER) in breast cancer. Here we define a new method for standardization of ER measurement and then examine both change in percentage and threshold of intensity (immunoreactivity) to assess sources for test discordance.

Methods

An assay was developed to quantify ER by using a control tissue microarray (TMA) and a series of cell lines in which ER immunoreactivity was analyzed by quantitative immunoblotting in parallel with the automated quantitative analysis (AQUA) method of quantitative immunofluorescence (QIF). The assay was used to assess the ER protein expression threshold in two independent retrospective cohorts from Yale and was compared with traditional methods.

Results

Two methods of analysis showed that change in percentage of positive cells from 10% to 1% did not significantly affect the overall number of ER-positive patients. The standardized assay for ER on two Yale TMA cohorts showed that 67.9% and 82.5% of the patients were above the 2-pg/ μ g immunoreactivity threshold. We found 9.1% and 19.7% of the patients to be QIF-positive/IHC-negative, and 4.0% and 0.4% to be QIF-negative/IHC-positive for a total of 13.1% and 20.1% discrepant cases when compared with pathologists' judgment of threshold. Assessment of survival for both cohorts showed that patients who were QIF-positive/pathologist-negative had outcomes similar to those of patients who had positive results for both assays.

Conclusion

Assessment of intensity threshold by using a quantitative, standardized assay on two independent cohorts suggests discordance in the 10% to 20% range with current IHC methods, in which patients with discrepant results have prognostic outcomes similar to ER-positive patients with concordant results.

J Clin Oncol 29:2978-2984. © 2011 by American Society of Clinical Oncology

INTRODUCTION

It is widely recognized that the immunohistochemistry (IHC) test has significant limitations in accuracy because of a wide range of variables.¹ These issues were highlighted by a recent incident in Canada that revealed a 40% misclassification rate between local and central laboratories² and raised urgent awareness of the limitations of estrogen receptor (ER) measurement.³⁻⁶ To address this issue, the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) convened

an expert panel that ultimately issued a series of guidelines.⁷ Most significantly, the guidelines lowered the standard for ER positivity from 10% positive nuclei to 1% positive nuclei, but they did not address the issue of intensity or threshold (that is, what actually constitutes a "positive" nucleus). They define positivity as "immunoreactivity... in the presence of expected reactivity of internal (normal epithelial elements) and external controls."

Although these guidelines may represent the state of the art for assessment of immunoreactivity, they lack a mechanism for universal standardization.

Since the amount of ER is scored qualitatively by eye, there is variability and lack of reproducibility between pathologists. Different laboratories use different antibodies, reagents, and protocols to prepare ER slides for interpretation. To compound the problem, there has been a broad shift to core biopsy over the last few years, so specimens are commonly too small to have normal epithelial elements on the same slide. Here we describe a potential method for standardization of ER measurement on a slide. We use quantitative immunofluorescence (QIF), now commercialized as automated quantitative analysis (AQUA) technology (HistoRx, New Haven, CT). This method calculates marker expression on a continuous scale by using pixel intensity and is shown to be widely applicable for biomarker analysis.⁸⁻¹⁴ Previous measurements of ER by AQUA have correlated well with IHC analysis on tissue from two large clinical trials and have also predicted response to tamoxifen.^{15,16}

In an attempt to both quantify and standardize the measurement of ER in patient tissue, we first sought to define an ER cut point with biologic and clinical relevance. This was done by using a control (index) tissue microarray (TMA) containing 40 patient controls alongside a panel of cell lines (prepared as tissue and built onto the TMA). This index array was used as a standard and was stained alongside every cohort that was assessed for ER to allow reproducible selection of the threshold for positivity. Finally, we used this standardized assay on two independent archival Yale cohorts to estimate the level of discordance as a function of intensity threshold (rather than percent positive) in sample populations.

METHODS

Details regarding all methods are provided in the Appendix (online only).

Cell Line Panel and Culture

A panel of American Type Culture Collection (ATCC) breast cancer cell lines was chosen to span a range of ER expression. We also included Puro9 cells (MCF-7 cells with tetracycline-inducible ER- α overexpression),¹⁷ maintained as six separate cultures (treated with 0, 0.01, 0.1, 0.5, 1, and 5 mg/mL doxycycline).

Quantitative Immunoblotting

The amount of ER was quantified as a concentration (picograms of ER per microgram of total protein) for each cell line, by using 1D5 antibody (DAKO, Copenhagen, Denmark).

Immunofluorescent Staining

TMA were stained for 4,6-diamidino-2-phenylindole (DAPI), cytokeratin, and ER (1D5 antibody) by using a standard protocol developed in our laboratory. IHC assessment of ER was performed by two board-certified pathologists at Yale (M.H. and D.L.R.) or at The Cancer Institute of New Jersey who used the 1D5 antibody and standard IHC methods (new 1% cutoff guidelines for YTMA 49 and 10% cutoff for YTMA 130). These IHC assessments were performed on the same TMAs used for analysis by the AQUA assay and used the same core from each patient.

AQUA Analysis

ER immunofluorescence (IF) was quantified in tumor nuclei by using AQUA technology, which was previously developed in our laboratory.

Patient Cohorts

Two large cohorts of archival breast cancer samples from Yale were used: YTMA 49 (patients diagnosed from 1962 to 1982; $n = 619$) and YTMA 130 (patients diagnosed from 1976 to 2005; $n = 390$). Tissues were collected in accordance with consent guidelines in protocol 8219 issued to D.L. Rimm from the Yale Human Investigation Committee (institutional review board).

Clinicopathologic characteristics of both cohorts are found in Appendix Table A1 (online only).

Statistical Analysis

All analyses were performed by using the StatView software platform (SAS Institute, Cary, NC). Box plots, analysis of variance (ANOVA) tests, and Kaplan-Meier survival analyses were performed on each cohort (disease-free survival [DFS] or recurrence-free survival [RFS]), and statistical significance was assessed by using the log-rank test.

RESULTS

Assessment of Discordance As a Function of the Change From 10% to 1% Immunoreactive Cells

Although it has been a relatively short time since our institution has adopted the new ASCO/CAP guidelines for percent positivity, we have a sufficient volume of patient data to address the effect on ER-positive classification. By using a custom-designed retrospective search of the Yale Copath database, we determined the percentage of total patients called ER-positive by the 10% standard for each year since 2000. We then compared this number to the percentage of patients called positive since April 2010 (when the 1% standard came into effect). By using χ^2 analysis, we determined that there is not a significant difference in the percentage of patients who are called positive when using the adopted 1% standard compared with using the 10% standard when performing pairwise comparison of patient samples read in 2010 according to the new standard with those of any previous year (Table 1).

To test this difference in an experimental setting, three observers (two pathologists and one student) scored the conventionally stained TMA according to the new ASCO/CAP guidelines, including both an intensity score and a percentage score. There is almost no difference (approximately 1% of cases) in the percentage of cases called ER positive using the 10% or 1% cutoff (Table 2).

Table 1. Number of Patients With Invasive Breast Carcinoma Diagnosed As ER Positive at Yale-New Haven Hospital From 2000 to 2010

Year	Patients With Invasive Carcinoma		χ^2 P for Pairwise Comparison With 2010 Data
	No. With ER Results	ER-Positive Results No. %	
2000	246	189 76.83	.29
2001	268	212 79.10	.60
2002	264	196 74.24	.09
2003	298	226 75.84	.18
2004	332	266 80.12	.79
2005	455	342 75.16	.11
2006	491	406 82.69	.64
2007	497	395 79.48	.64
2008	502	411 81.87	.82
2009	550	450 81.82	.83
From April 2010	180	146 81.11	—

NOTE. Data from 2010 includes only April 1 through August 31. Note that over the last 10 years there has been a statistically significant trend toward increase in ER in the population seen at Yale (Mantel-Haenszel χ^2 P = .0036). Abbreviation: ER, estrogen receptor.

Table 2. Number of Patients With Invasive Breast Carcinoma Scored As ER Positive on YTMA 49

Scorer	Total No. of Patients With Invasive Carcinoma With ER Results	Patients Scored As ER Positive			
		Using 10% Cutoff		Using 1% Cutoff	
		No.	%	No.	%
D.L.R.	526	312	59.31	318	60.46
M.H.	462	293	63.42	293	63.42
A.W.W.	502	335	66.73	340	67.73

NOTE. Excluded patients could not be scored because of insufficient tumor, infiltration, or out-of-focus tissue. D.L.R. and M.H. are board-certified pathologists; A.W.W. is a graduate student in pathology.
Abbreviations: ER, estrogen receptor; YTMA 49, Yale tissue microarray [cohort].

Development of an Immunoblot-Standardized Method for Quantification of ER

To allow reproducible and quantitative selection of an ER cut point, we sought to create a control array (which we call the index TMA) that would serve as a standard curve for ER expression and include both a panel of cell lines (prepared as patient tissue) and 40 patient controls. The goal of using a panel of cell lines was to perform quantitative western blotting (provides ER measurement as a concentration) in parallel with quantitative IF (provides ER measurement as an AQUA score) to create a conversion from AQUA scores to concentrations that could be applied to the 40 patient controls.

For the cell line panel, we chose ATCC breast cancer cell lines representing the range of ER levels. To expand the ER dynamic range so it more closely mirrored that seen in patients, we used MCF-7 cells stably transfected with a tetracycline-inducible ER overexpression system (cultured at 0, 0.01, 0.1, 0.5, 1, and 5 mg/mL doxycycline) as previously described.¹⁷ ER was measured in this panel of cell lines by quantitative western blot (Fig 1A) alongside a standard curve of recombinant ER to determine absolute concentration of ER in picograms per microgram of total protein. Cell lines were also prepared as tissue (pelleted, formalin-fixed, paraffin-embedded, and cored) and placed on the index TMA alongside 40 patient controls for quantitative IF analysis by AQUA (scores shown in Fig 1B). The same ER antibody (1D5) was used for both western blot and IF analysis. Combining the AQUA and quantitative ER determination from select cell lines, absolute concentrations of ER (in picograms per microgram) were correlated with ER AQUA scores, and the regression (Fig 1C) was used to determine concentrations of ER (picograms per microgram) from AQUA scores in the cell line panel. Known ER expression in these cell lines allowed us to determine the cut point between the highest ER-negative cell line and the lowest ER-positive cell line to be 2 pg/ μ g.

This cut point was applied to the panel of 40 patient controls on the index TMA, whose ER concentrations (picograms per microgram) were calculated from their AQUA scores using the same regression (Fig 1C). There was one patient who did not have sufficient tissue for AQUA analysis, and thus the final panel consisted of 39 patient controls (Fig 1D). We further validated this threshold of 2 pg/ μ g by eye, contracting the dynamic range of the grayscale image (adjusted maximum red-green-blue [RGB] input level from 255 to 16 by using Adobe Photoshop) to visualize low levels of specific nuclear staining as

well as nonspecific background. Corresponding images for the highest negative control case (blue arrow in Fig 1D) and the lowest positive control case (gold arrow in Fig 1D) are shown in Figure 1E.

This index TMA is incorporated as a key component of the ER AQUA assay and is stained as a control in every experiment to determine a cut point and to standardize scores between users, machines, and sites. It is assessed for reproducibility with each staining run and, over the course of eight individual runs, has displayed an average coefficient of determination (r^2) of 0.902 ($r = 0.950$).

Comparison of ER Quantification by QIF Versus Pathologist Review

To determine the effects of a standardized threshold compared with current standard methods, we used our assay to measure ER on two independent retrospective breast cancer cohorts from Yale. For each cohort, ER status was determined using the standardized assay described in Figure 1 (using the index TMA) and compared with ER status as determined by IHC review (read by two independent pathologists; 0 is negative and 1-3 is positive). The first cohort (YTMA 49) is a retrospective collection from Yale consisting of 619 patients, with median follow-up time of 104.1 months (clinicopathologic characteristics are provided in Appendix Table A1). Because of TMA exhaustion, valid data for ER expression at two-fold redundancy were obtained on 280 patients. We saw a high overall concordance between the QIF assay and IHC review (Appendix Fig A1A, online only). Of a total of 252 patients, 33 (13.4%) had discordant results and 23 (9.1%) were ER positive by QIF analysis and ER negative by IHC review (QIF-positive/IHC-negative; Table 3).

Quantification of ER revealed a unimodal distribution with 67.9% of cases above the 2-pg/ μ g threshold which were thus defined as positive (Fig 2A). The distribution of discordant cases showed that many of them fell around the 2-pg/ μ g threshold (Fig 2A), as expected. To examine the significance of this discordance with respect to patient prognosis, we performed Kaplan-Meier survival analysis by using DFS as an end point. Stratifying patients by using both methods of ER analysis (Fig 2B), we found that the patients with discrepant ER status (ER-positive by QIF; ER-negative by IHC) displayed survival behavior that aligned with that of patients who were ER positive by both assays (QIF-positive/IHC-positive). To further validate the 2-pg/ μ g threshold on this cohort, we visually examined images of ER QIF staining for patient samples that fell on either side of the cut point (Fig 2C). We confirmed specific nuclear staining seen above the threshold at 4.5 pg/ μ g, in contrast to low levels of nonspecific background seen below the threshold at 0 pg/ μ g.

The second cohort (YTMA 130) is a newer retrospective collection from Yale consisting of 390 patients, 49% of whom had received tamoxifen, with a median follow-up time of 80 months (clinicopathologic characteristics are described in Appendix Table A1). Of these, 234 patients had valid data on ER status by the QIF assay. Again we saw a strong correlation between IHC review and QIF analysis (Appendix Fig A1B), but a total of 47 patients (20.1%) still had discordant results, with 98% (46 of 47) of patients being QIF-positive/IHC-negative (Table 3). Representative AQUA/IF images of ER staining for each of these classifications are shown in Appendix Figure A1C, confirming specific nuclear staining in patients considered positive by QIF analysis but negative by IHC review. Similarly, we saw nonspecific background staining in patients who were classified as QIF-negative/IHC-positive.

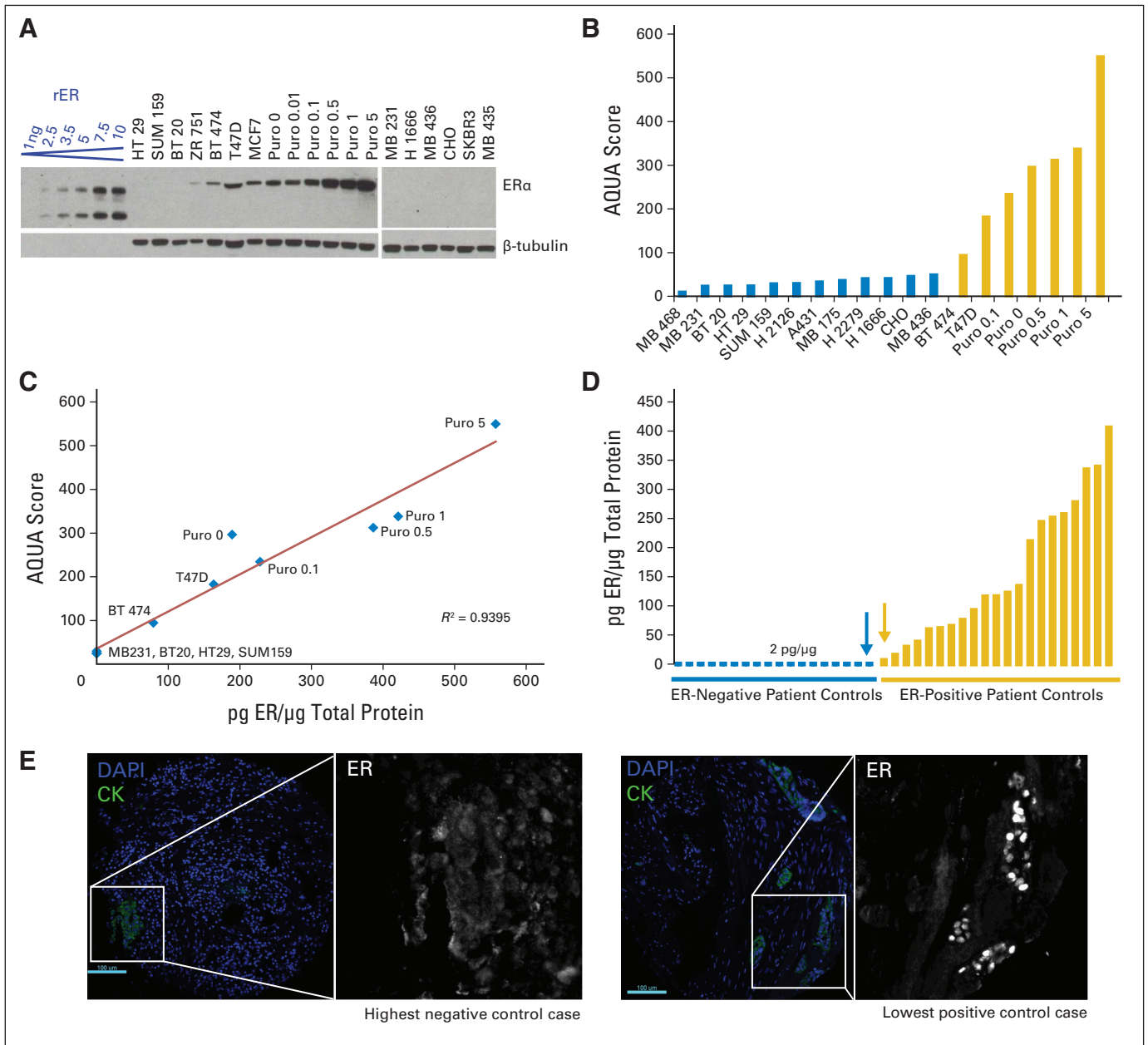


Fig 1. Method for quantification of estrogen receptor (ER) using an immunoblot-standardized automated quantitative analysis (AQUA) assay. (A) ER was measured in a panel of cell line controls by western blot (1D5 antibody; DAKO, Copenhagen, Denmark) alongside a standard curve of recombinant ER (rER) to determine absolute concentration in picograms per microgram of total protein. Cell lines included Puro9 cells, which are MCF-7 cells with doxycycline-induced overexpression of ER (0, 0.01, 0.1, 0.5, 1, and 5 mg/mL doxycycline). Cell lines were also pelleted, cored, and placed on the index tissue microarray for immunofluorescence and AQUA analysis. (B) Absolute concentrations of ER (picograms per microgram) were correlated to ER expression by immunofluorescence (AQUA score using 1D5), and the regression (C) was used to convert AQUA scores to concentrations of ER (picograms per microgram) in the set of patient controls present on the same index tissue microarray; (D) shows picograms per microgram distribution. Immunofluorescent AQUA images of ER in the highest negative control case (blue arrow) and the lowest positive control case (gold arrow) are also shown in (E) to validate the cut point. Cytokeratin (CK) was used as a mask to define regions of tumor. For ER, we contracted the dynamic range of the grayscale image (adjusted maximum red-green-blue input level from 255 to 16 by using Adobe Photoshop) to visualize low levels of specific nuclear staining as well as nonspecific background. DAPI, 4,6-diamidino-2-phenylindole.

Quantification of ER on this cohort revealed a unimodal distribution with 82.5% of cases above the 2-pg/μg threshold (Fig 3A). Examining the distribution of discordant cases again showed that many were around the threshold, but some were also at the high range of expression. Kaplan-Meier analysis was performed by using RFS instead of DFS because data on patient recurrence was available on this cohort and also because tamoxifen treatment reduced the overall

number of deaths. Stratification of patients by using both methods of ER analysis (Fig 3B) showed that the patients with discordant ER status (QIF-positive/IHC-negative) displayed survival behavior that was similar to that of the double ER-positive population. As we did previously, we visually validated the 2-pg/μg AQUA threshold on patients on either side of the cut point (Fig 3C), confirming specific nuclear staining at 3.8 pg/μg but nothing specific detectable at 0.4 pg/μg.

Table 3. Comparison of ER Status by IHC Review Versus AQUA Assay for YTMA 49 and YTMA 130

ER Status		YTMA 49 (1962-1982)		YTMA 130 (1976-2005)	
AQUA (positive > 2 pg/μg; negative < 2 pg/μg)	IHC Review (positive = 1-3, negative = 0)	No.	%	No.	%
Positive	Positive	148	58.7	147	62.8
Positive	Negative	23	9.1	46	19.7
Negative	Positive	10	4.0	1	0.4
Negative	Negative	71	28.2	40	17.1
Total		252		234	

Abbreviations: ER, estrogen receptor; IHC, immunohistochemistry; AQUA, automated quantitative analysis; YTMA, Yale tissue microarray [cohort].

DISCUSSION

The two key findings of this study are (1) that the threshold of immunoreactivity appears to be more important than the percentage positive in generation of discordant or false-negative assays and (2) that the standardization method by using the QIF assay appears to be more

sensitive than the traditional IHC assay, even though the same antibody is used for detection of ER (1D5). In support of the first finding, although some pathologists report calling more cases positive as a result of the change in the guidelines, the two data collections examined in this study suggest that false negatives, like those reported in the Canadian incident,² are unlikely to be due to percentage-positive issues.

False-negative cases may be a significant problem at other sites around the world as well. Recently presented data on the ER false-negative rate in the Breast International Group 1-98 (BIG 1-98) population and Adjuvant Lapatinib and/or Trastuzumab Treatment Optimisation (ALTTO) trial also suggested that between 15% and 20% of case analyses performed in local laboratories may be falsely assigned a negative score. Other studies in the United States have much more modest disagreement between centralized versus local laboratories,^{18,19} but essentially no laboratories in the United States or elsewhere use a standard curve to assess the ER detection threshold. The current standard in most laboratories is to use a single strongly positive example case as a control for stainer runs. Other laboratories rely on intrinsic controls provided by adjacent normal ducts. Neither of these methods specifically assesses the threshold of positivity.

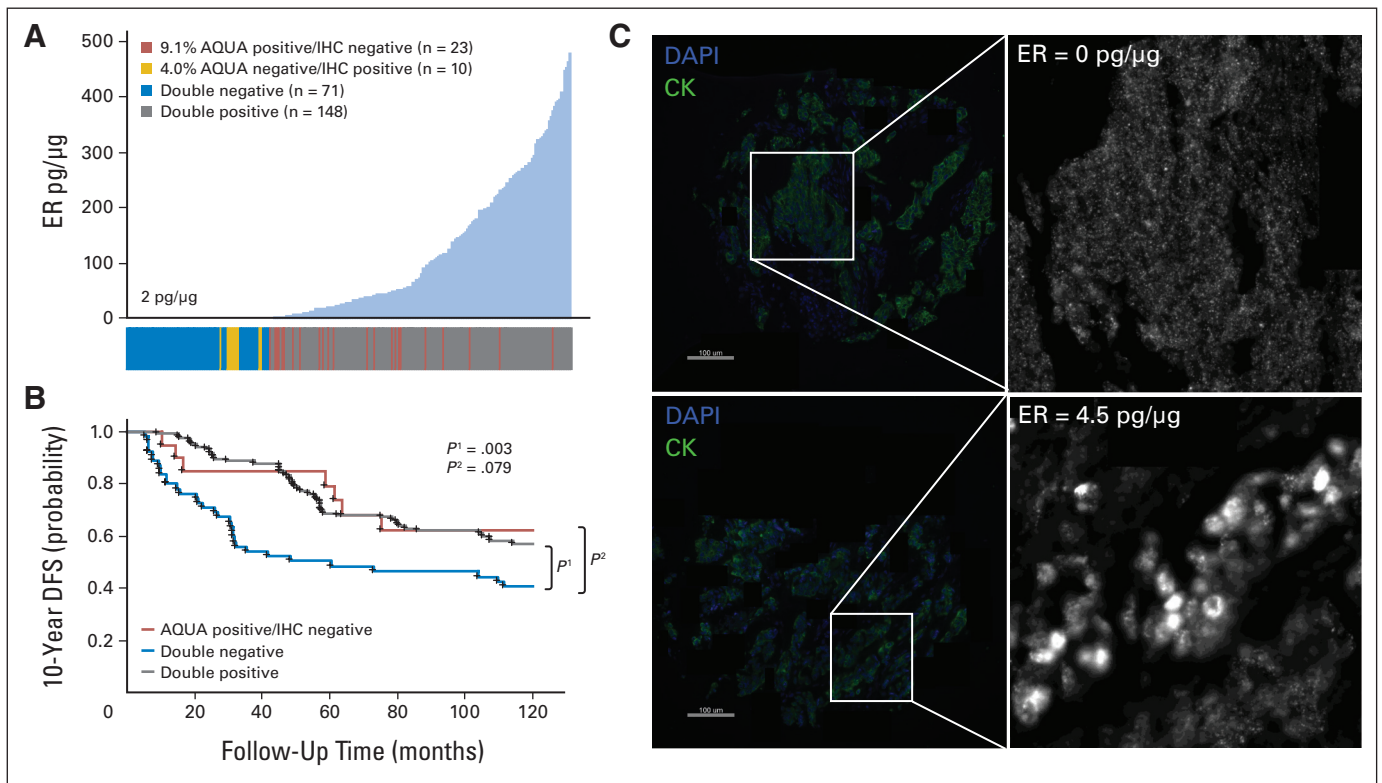


Fig 2. Discordant classification of estrogen receptor (ER) status in YTMA 49 cohort. (A) ER status was determined by immunofluorescence and automated quantitative analysis (AQUA) in Yale retrospective breast cancer cohort YTMA 49 (diagnosed between 1962 and 1982; clinicopathologic characteristics provided in Appendix Table A1) and compared with ER status as determined by immunohistochemistry (IHC), read by two certified pathologists (M.H. and D.L.R.) using the current 1% positive nuclei cutoff guidelines. Distribution of ER by AQUA (picograms per microgram standardized as illustrated in Fig 1) is shown in which each case is color coded in the bar below, according to its ER status by both AQUA and IHC. (B) Kaplan-Meier curves show 10-year disease-free survival (DFS), in which patients are grouped according to the classifications shown in (A). The AQUA-negative/IHC-positive group (n = 10) was excluded from survival analysis on account of small size and insufficient power. (C) To confirm and further validate the AQUA cut point of 2 pg/μg in this cohort, representative immunofluorescence images of ER staining for patients on either side of the cut point are shown (right panels). Cytokeratin (CK) was used as a mask to define regions of tumor (green, left panels). For ER, we contracted the dynamic range of the grayscale image (adjusted maximum red-green-blue input level from 255 to 16 by using Adobe Photoshop) to visualize low levels of specific nuclear staining as well as nonspecific background. DAPI, 4,6-diamidino-2-phenylindole.

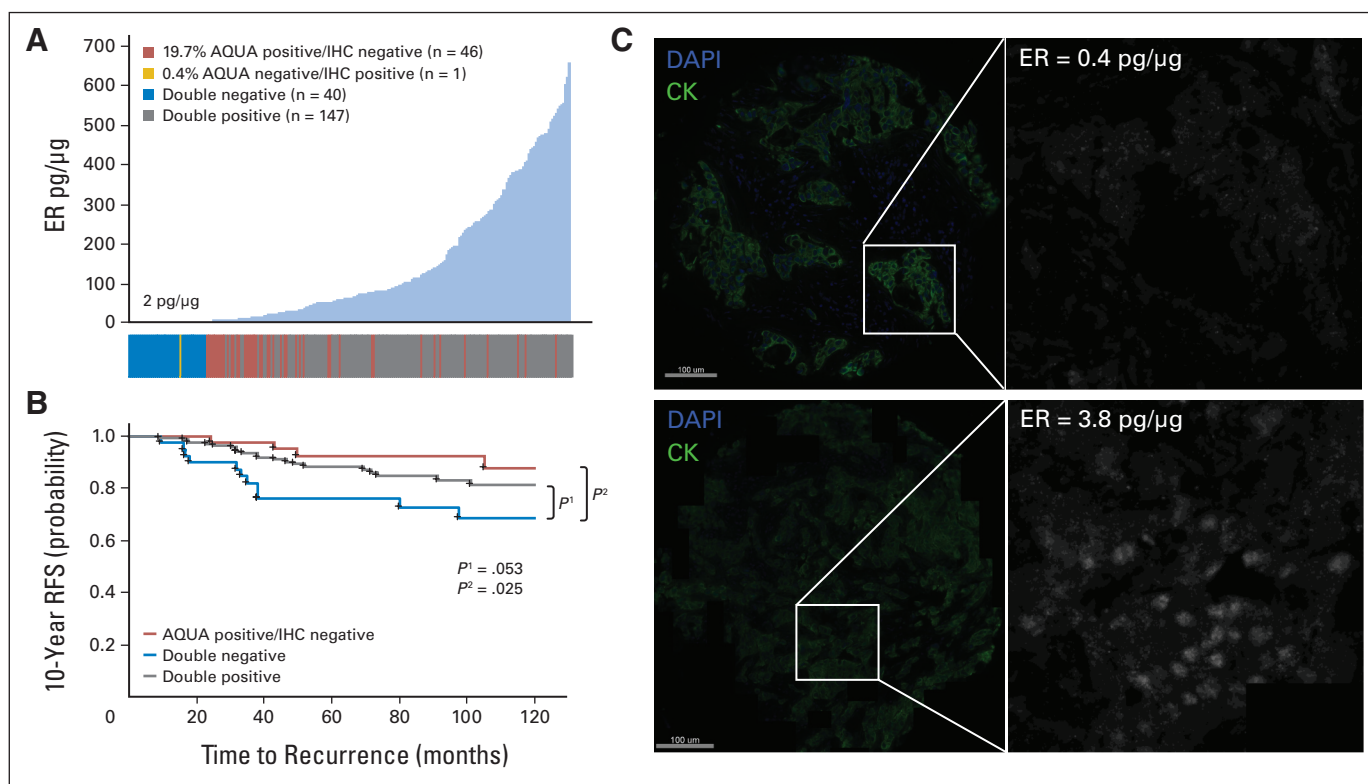


Fig 3. Discordant classification of estrogen receptor (ER) status in the YTMA 130 cohort. (A) ER status was determined by immunofluorescence and automated quantitative analysis (AQUA) in Yale retrospective breast cancer cohort YTMA 130 (diagnosed between 1976 and 2005; clinicopathologic characteristics in Appendix Table A1) and compared with ER status as determined by immunohistochemistry (IHC) using the 10% positive nuclei cutoff guidelines. A distribution of ER by AQUA (picograms per microgram standardized as illustrated in Fig 1) is shown, in which each case is color coded according to its ER status by both AQUA and IHC. (B) Kaplan-Meier curves show 10-year recurrence-free survival (RFS), in which patients are grouped according to the classifications shown in (A). The AQUA-negative/IHC-positive group (n = 1) was excluded from survival analysis on account of its size and insufficient power. (C) The AQUA cut point of 2 pg/μg was further validated in this cohort by examining representative immunofluorescent images of ER staining for patients on either side of the cut point (right panels). Cytokeratin (CK) was used as a mask to define regions of tumor (green, left panels). For ER, we contracted the dynamic range of the grayscale image (adjusted maximum red-green-blue input level from 255 to 16 by using Adobe Photoshop) to visualize low levels of specific nuclear staining as well as nonspecific background. DAPI, 4,6-diamidino-2-phenylindole.

The second key finding of this study is that the use of a standardized method results in a reproducible system for assessment of that threshold. Furthermore, it reveals a threshold that by QIF appears to be more sensitive than by traditional IHC. This may be due to the use of the hematoxylin counterstain that, when applied too heavily, can obscure faint staining, as has been previously described for other tumors.²⁰ Examples of two discordant cases that were QIF-positive/IHC-negative are shown in Appendix Figure A2 (online only). Some automated technologies claim to be able to “unmix” the colors, and they may have similar capacity and sensitivity. However, to the best of our knowledge, a head-to-head comparison has not yet been done.

There are several limitations in the conclusions that can be drawn from this study. Perhaps the most important is that we are unable to determine ground truth for ER status. Although we can assess test discordance and compare discordant cases to concordant cases with respect to survival, we have no absolute way of determining the true ER expression status of each patient. The best method to adjudicate this would be response to endocrine therapy. That information is not available for this study, although studies are planned to test this assay in clinical trial specimens in which that information is available.

The assay we developed represents our best attempt to accurately measure ER protein in tissue, but any assay can only measure protein that is present on the slide. Preanalytic factors, most significantly cold

ischemic time, can decrease the amount of ER epitope present on the slide and may account for some level of misclassification in the clinical setting.¹ However, in this study, both assays were performed on the same tissue specimens; thus, preanalytic variation is unlikely to contribute to the observed discordance. Another limitation of this study is that the cohort analyses were done on TMAs rather than on whole sections as used in the clinical setting. Although TMAs have been shown to be representative, they may have a limitation with respect to assessment of sufficient area. TMAs may also have a limitation in that the heterogeneity seen in a tissue section is unlikely to be completely represented in a TMA. In cases of discordance distant from the threshold, the cause could be tumor heterogeneity.

In this study, our goal was to derive a biologically relevant cut point and a method of standardization that could be used in clinical laboratories. Using cell lines allowed us to convert patient ER expression to an absolute concentration within a field of view. An absolute concentration, along with a confidence interval for measurement, is a standard readout for many laboratory tests based on fluid specimens, and thus it is a reasonable goal for ER. The use of cell lines may be a good future universal standard. However, we have found that, even if authenticated, cell line expression can vary as a function of confluence, passage number, and other variables that are yet to be determined. Studies are underway in the laboratory to develop alternative universal

standards. We believe the best current standard, even though it is not perfect, can be derived from a set of control cases in conjunction with a standardized set of cell lines. The index TMA in this article included samples from 39 patients with a range of ER expression, represented cases around the threshold, and showed strong run-to-run reproducibility ($r > 0.9$). It is a good example of a standard array that could be processed with each stainer run to ensure reproducibility around the ER threshold.

Overall, our results suggest that use of a standardized, quantitative, IF-based assay could significantly improve the way ER status is evaluated, overcoming the limitations of IHC by providing a method for reproducible assessment of the threshold. Furthermore, they suggest potential biologic relevance for low levels of ER expression and reinforce our need to adopt a standardized assay that can discern this subtle, but potentially important phenomenon. The AQUA method for analysis of patients' specimens has now be implemented by a Clinical Laboratory Improvement Amendments (CLIA) laboratory in an effort to offer a more accurate and reproducible test for ER, progesterone receptor, and human epidermal growth receptor factor 2 (HER2). Studies are now needed to confirm that using this test in routine practice will result in improved patient outcome.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject

matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment or Leadership Position: None **Consultant or Advisory Role:** David L. Rimm, HistoRx (C), Genentech (C), DAKO (C) **Stock Ownership:** David L. Rimm, HistoRx **Honoraria:** David L. Rimm, Genoptix **Research Funding:** None **Expert Testimony:** None **Other Remuneration:** None

AUTHOR CONTRIBUTIONS

Conception and design: Allison W. Welsh, Elaine T. Alarid, Bruce G. Haffty, David L. Rimm

Financial support: David L. Rimm

Provision of study materials or patients: Elaine T. Alarid, Bruce G. Haffty, David L. Rimm

Collection and assembly of data: Allison W. Welsh, Christopher B. Moeder, Sudha Kumar, Peter Gershkovich, Bruce G. Haffty, David L. Rimm

Data analysis and interpretation: Allison W. Welsh, Christopher B. Moeder, Sudha Kumar, Peter Gershkovich, Elaine T. Alarid, Malini Harigopal, Bruce G. Haffty, David L. Rimm

Manuscript writing: All authors

Final approval of manuscript: All authors

REFERENCES

- Gown AM: Current issues in ER and HER2 testing by IHC in breast cancer. *Mod Pathol* 21:S8-S15, 2008 (suppl 2)
- Hede K: Breast cancer testing scandal shines spotlight on black box of clinical laboratory testing. *J Natl Cancer Inst* 100:836-837, 2008
- Allred DC: Commentary: Hormone receptor testing in breast cancer—A distress signal from Canada. *Oncologist* 13:1134-1136, 2008
- Matthews A: Bad cancer tests draw scrutiny. *Wall Street Journal*, January 4, 2008:B1
- Allison KH: Estrogen receptor expression in breast cancer: We cannot ignore the shades of gray. *Am J Clin Pathol* 130:853-854, 2008
- Harris L, Fritsche H, Mennel R, et al: American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 25:5287-5312, 2007
- 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. *J Clin Oncol* 28:2784-2795, 2010
- McCabe A, Dolled-Filhart M, Camp RL, et al: Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis. *J Natl Cancer Inst* 97:1808-1815, 2005
- Berger AJ, Davis DW, Tellez C, et al: Automated quantitative analysis of activator protein-2alpha subcellular expression in melanoma tissue microarrays correlates with survival prediction. *Cancer Res* 65:11185-11192, 2005
- Camp RL, Dolled-Filhart M, King BL, et al: Quantitative analysis of breast cancer tissue microarrays shows that both high and normal levels of HER2 expression are associated with poor outcome. *Cancer Res* 63:1445-1448, 2003
- Dolled-Filhart M, McCabe A, Giltane J, et al: Quantitative in situ analysis of beta-catenin expression in breast cancer shows decreased expression is associated with poor outcome. *Cancer Res* 66:5487-5494, 2006
- Giltane JM, Rydén L, Cregger M, et al: Quantitative measurement of epidermal growth factor receptor is a negative predictive factor for tamoxifen response in hormone receptor positive premenopausal breast cancer. *J Clin Oncol* 25:3007-3014, 2007
- Gustavson MD, Bourke-Martin B, Reilly D, et al: Standardization of HER2 immunohistochemistry in breast cancer by automated quantitative analysis. *Arch Pathol Lab Med* 133:1413-1419, 2009
- Moeder CB, Giltane JM, Harigopal M, et al: Quantitative justification of the change from 10% to 30% for human epidermal growth factor receptor 2 scoring in the American Society of Clinical Oncology/College of American Pathologists guidelines: Tumor heterogeneity in breast cancer and its implications for tissue microarray based assessment of outcome. *J Clin Oncol* 25:5418-5425, 2007
- Bartlett JM, Brookes C, Robson T, et al: The TEAM trial pathology study identifies potential prognostic and predictive biomarker models for postmenopausal patients treated with endocrine therapy. 32nd Annual San Antonio Breast Cancer Symposium, San Antonio, TX, December 9-13, 2009 (abstr 75)
- Harigopal M, Barlow WE, Tedeschi G, et al: Multiplexed assessment of the Southwest Oncology Group-directed Intergroup Breast Cancer Trial S9313 by AQUA shows that both high and low levels of HER2 are associated with poor outcome. *Am J Pathol* 176:1639-1647, 2010
- Fowler AM, Solodin N, Preisler-Mashek MT, et al: Increases in estrogen receptor-alpha concentration in breast cancer cells promote serine 118/104/106-independent AF-1 transactivation and growth in the absence of estrogen. *Faseb J* 18:81-93, 2004
- Ma H, Wang Y, Sullivan-Halley J, et al: Breast cancer receptor status: Do results from a centralized pathology laboratory agree with SEER registry reports? *Cancer Epidemiol Biomarkers Prev* 18:2214-2220, 2009
- Collins LC, Marotti JD, Baer HJ, et al: Comparison of estrogen receptor results from pathology reports with results from central laboratory testing. *J Natl Cancer Inst* 100:218-221, 2008
- Rothberg BE, Moeder CB, Kluger H, et al: Nuclear to non-nuclear Pmel17/gp100 expression (HMB45 staining) as a discriminator between benign and malignant melanocytic lesions. *Mod Pathol* 21:1121-1129, 2008