Automated Quantitative Analysis (AQUA) of In Situ Protein Expression, Antibody Concentration, and Prognosis

Anthony McCabe, Marisa Dolled-Filhart, Robert L. Camp, David L. Rimm

Background: Disparate results in the immunohistochemistry literature regarding the relationship between biomarker expression and patient outcome decrease the credibility of tissue biomarker studies. We investigated whether some of these disparities result from subjective optimization of antibody concentration. Methods: We used the automated quantitative analysis (AQUA) system and various concentrations of antibodies against HER2 (1:500 to 1:8000 dilutions), p53 (1:50 to 1:800 dilutions), and estrogen receptor (ER; 1:100 and 1:1000 dilutions) to assess expression of HER2 and p53 in a tissue microarray containing specimens from 250 breast cancer patients with long-term survival data available. HER2 expression in the tissue microarray was also assessed by conventional immunohistochemistry. Relative risk (RR) of disease-specific mortality was assessed for every cutpoint with the X-tile program. Cumulative disease-specific survival was assessed by the Kaplan-Meier method. All statistical tests were two-sided. Results: For HER2 and p53 and an optimal cutpoint, when a high antibody concentration (i.e., 1:500 dilution) was used with the AQUA system, low expression was associated with poorer survival than high expression; however, when a low antibody concentration (i.e., 1:8000 dilution) was used, high expression was associated with poorer survival. For example, for a 1:8000 dilution of HER2 antibody and high expression defined as the top 15% of HER2 expression, high HER2 expression was associated with increased disease-specific mortality (RR = 1.98, 95% confidence interval [CI] = 1.21 to 3.23; P = .007), compared with low expression. However, for a 1:500 dilution of HER2 antibody and high expression defined as the top 85% of HER2 expression, high HER2 expression was associated with decreased disease-specific mortality (RR = 0.47, 95% CI = 0.29 to 0.76; P = .002), compared with low HER2 expression. Conclusions: Biomarker antibody concentration appears to dramatically affect the apparent relationship between biomarker expression and outcome. [J Natl Cancer Inst 2005;97:1808–15]

Because immunohistochemistry provides information on the expression level and the localization of an antigen, it has become the standard in situ assay to assess protein expression. This method has been used in thousands of papers that show associations between protein expression and clinical outcome. However, immunohistochemistry is semiquantitative, subjective, and highly dependent on a range of poorly controlled variables, including antibody concentration. For example, various laboratories choose different antibody dilutions based on the subjective “by eye” optimization of the signal-to-noise ratio. This subjective process is the accepted standard method in the field, even though it lacks rigor and standardization. Thus, different laboratories optimize conditions by use of variable methods that generate results that can vary between studies (1,2). Immunohistochemistry analyses of p53 provide a striking example of this phenomenon; high-level expression may be associated with poor outcome in one study, whereas no association may be found in another study [for review, see Elledge and Allred (3)].

One method to reduce the number of variables in immunohistochemistry analysis is the use of tissue microarrays. Many laboratories have used this technology to study the expression of biomarkers in hundreds of tumor samples on the same slide (4,5). Comparisons between tissue microarrays containing breast cancer samples and corresponding whole-tissue sections have demonstrated that, for some tissues, only a single tissue microarray section (i.e., a histospot) is sufficient to identify antigen expression in the whole section when the heterogeneity is averaged across a whole population of tumor samples, with high concordance for common biomarkers and reproducible prognostic associations between staining levels and clinical outcomes (6,7). This approach eliminates differential antigen retrieval and staining conditions as possible variables. However, it does not eliminate the variability in scoring. The current standard methods for scoring protein expression by immunohistochemistry of traditional tissue sections or tissue microarrays is the 0–3 scale (in which 0 = no staining, 1 = weak staining, 2 = moderate staining, and 3 = strong staining) or the H score (a product of staining intensity and the percent of cells stained—for example, if 50% of the slides shows an intensity of more than 3, 20% show an intensity of more than 2, and 30% are negative, then the H score would be 150 + 40 = 190). These scoring methods are subjective and are subject to
human variability even within controlled settings, i.e., in which positive and negative controls are used (8).

We have recently developed an automated scoring system for assessing biomarker expression in tissue sections called the automated quantitative analysis (AQUA) system (9). The AQUA system is linked to a fluorescent microscope system that detects the expression of biomarker proteins by measuring the intensity of antibody-conjugated fluorophores within a specified subcellular compartment (typically including the nucleus, cytoplasm, and plasma membrane) within the tumor region of each tissue microarray spot. The result is a quantitative score of immunofluorescence intensity for the tumor. An AQUA analysis removes the subjectivity of the traditional scoring system and provides more continuous and reproducible scoring of protein expression scoring in tissue samples (9).

In this study, we use the AQUA system to measure the expression of HER2, estrogen receptor (ER), and p53 proteins in breast cancer tumors and to measure the expression of HER2 in a series of control cell lines to determine its absolute expression in each cell line. Our goal was to use the AQUA system to investigate how the antibody concentration used in immunohistochemical staining affects the association between apparent protein expression in the tumor and patient outcome. We also used a range of HER2 antibody concentrations on a series of cell lines that express HER2 to assess the dynamic range of HER2 protein expression (i.e., the range of HER2 concentrations in cells from those with low HER2 expression to those with high HER2 [i.e., HER2 amplified] expression) in a population of tumors.

**Patients and Methods**

**Tissue Microarray Construction**

Tissue microarrays were constructed from tumor tissue core samples. Core samples were obtained from representative regions of sections from each tumor that had been selected by use of the corresponding full sections stained with hematoxylin and eosin. The tissue microarray was constructed with single core samples (0.6-mm diameter) from each tumor that were spaced 0.8 mm apart in a grid format by using a tissue microarrayer (Beecher Instruments, Silver Spring, MD), as previously described (6,7). The tissue microarray was cut into 5-μm sections with a microtome, and the sections were adhered to the slide by means of an adhesive tape-transfer method, as described by the manufacturer (Instrumedics, Inc., Hackensack, NJ), and cross-linked to the slide with UV irradiation according to manufacturer’s instructions.

**Patient Cohort Characteristics**

The patient cohort consisted of a total of 250 patients with invasive breast carcinoma (125 with lymph node–negative disease and 125 with lymph node–positive disease) with tumor tissue available in formalin-fixed paraffin-embedded tissue blocks. These 250 patient samples were selected at random and on the basis of adequate tumor availability from a collection of more than 700 patient samples obtained from archives at the Yale University Department of Pathology; the original tumors had been resected between January 1, 1962, and December 31, 1977, and attempts were made to collect every tumor sample during that period. However, because of missing or exhausted tissue blocks, some samples could not be retrieved and included in this study. The follow-up time ranged from 2.4 months to 41.5 years (median = 8.3 years), and age at diagnosis ranged from 24 years to 86 years (median = 59 years). More detailed information about this cohort is published elsewhere (10,11). All personal health information was collected under the approval of Yale University Human Investigation Protocol 8219, which approved the informed consent signed at the time of surgery.

**Tissue Microarray Immunohistochemical Staining**

The tissue microarray slides were deparaffinized by two xylene rinses followed by two rinses with 100% ethanol. Antigen retrieval was performed by boiling the slides in a pressure cooker filled with 7.5 mM sodium citrate (pH 6.0). After rinsing briefly in 1× Tris-buffered saline (TBS) at pH 8, slides were incubated for 30 minutes in 2.5% hydrogen peroxide in methanol to block endogenous peroxidase activity. Slides were then incubated with 0.3% bovine serum albumin in 1× TBS for 1 hour at room temperature to reduce nonspecific background staining and then subjected to washes in 1× TBS, in 1× TBS containing 0.01% Triton, and then in 1× TBS, each 2 minutes long (hereafter referred to as TBS rinses). Slides were incubated first with a mouse anti-cytokeratin monoclonal antibody (clone AE1/AE3, DAKO, Carpinteria, CA; diluted 1:200) for the HER2 slides or with a rabbit anti-cytokeratin polyclonal antibody (Zymed, South San Francisco, CA; diluted 1:50) for the ER and p53 slides overnight at 4°C to define the epithelial mask. Slides were rinsed in 1× TBS and then incubated with HER2 antibody (c-erbB-2 oncoprotein, DAKO; diluted 1:500 through 1:800), ER antibody (ER-α mouse anti-human clone ID5, DAKO; diluted 1:100 or 1:1000), or p53 antibody (mouse anti-human clone DO-7, DAKO; diluted 1:50 through 1:800) for 1 hour at room temperature. Slides were rinsed in TBS as described above and incubated with secondary antibodies for 1 hour at room temperature: biotin anti-mouse or biotin goat anti-rabbit secondary antibodies (Vector Laboratories, Burlingame, CA; diluted 1:200) or mouse or rabbit secondary antibodies attached to a dextran-polimer backbone that was decorated with more than 100 molecules of covalently attached horseradish peroxidase (called “Envision,” DAKO) for HER2, ER, and p53 studies. The slides were washed with the TBS rinses described above and incubated for 30 minutes at room temperature with Alexa 546-streptavidin (Molecular Probes, Eugene, OR; diluted 1:200) to label the cytokeratin and then for 10 minutes with Cy-5 tyramide (NEN Life Science Products, Boston, MA) to allow coupling of Cy-5 dyes adjacent to the horseradish peroxidase–conjugated secondary antibody (tyramide is activated by horseradish peroxidase, and the activated form interacts covalently with adjacent protein molecules) for HER2, ER, and p53 studies. The emission peak of Cy-5 falls outside the tissue autofluorescence spectrum, thus minimizing background fluorescence for more accurate quantification of signal. The slides were then stained with the DNA-staining dye 4′,6-diamidino-2-phenylindole (DAPI) for 10 minutes, mounted with an antifade medium containing 0.6% n-propyl gallate in glycerol, and covered with a cover slip.

For the diaminobenzidine-based brown staining for HER2 expression, slides were prepared as described for the primary antibody incubations and washes, and then slides were incubated with Envision (DAKO) for 1 hour at room temperature. This incubation was followed by TBS rinses, visualization with
diaminobenzidine (diaminobenzidine chromogen, DAKO), and then counterstaining with ammonium hydroxide–acified hematoxylin. The slides were mounted with ImmunoMount (Shandon, Pittsburgh, PA) and then analyzed by use of a conventional four-point scoring system for HER2 membrane expression (0 = no staining, 1 = weak staining, 2 = moderate staining, and 3 = strong staining). Slides were read by two observers who were blinded to the outcome data for each slide. A high concordance (>90% exact agreement) was found between their scoring.

Cell Lines

The cell lines SK-BR-3, BT-474, MDA-MB-453, MDA-MB-435S, BT-549, T-47D, SW-480, MCF-7 MDA-MB-231, and MDA-MB-361 were purchased from the American Type Culture Collection (Manassas, VA), and all are breast cancer lines except for SW-480 cells, which are a colon cancer line. BAF3 cells, an interleukin 3–dependent cell line, were obtained from a laboratory in the Department of Genetics at Yale University. SK-BR-3, BT-474, SW-480, MCF-7, MDA-MB-453, and MDA-MB-435S cells were routinely cultured in Dulbecco’s modified Eagle medium containing 10% fetal bovine serum and 1% penicillin–streptomycin (Life Technologies, Inc., Grand Island, NY). T-47D and MDA-MB-231 cells were cultured in RPMI 1640 medium containing 10% fetal bovine serum and 1% penicillin–streptomycin (Life Technologies). BAF3 cells were cultured in RPMI 1640 medium containing 10% fetal bovine serum, 1% penicillin–streptomycin, and 10% WEHI-cell conditioned medium (12), as a source of interleukin 3. All cell lines were maintained in a 37°C incubator with 5% CO₂–95% air (Life Technologies).

**Cell Line Immunofluorescent Staining for HER2**

SK-BR-3, BT-474, MDA-MB-361, and MDA-MB-453 cells were trypsinized and plated at intermediate cell density into six-well plates (Life Technologies), grown overnight, and then fixed with neutral-buffered 10% formalin. BT-549, T-47D, SW-480, MCF-7, MDA-MB-231, BAF3, and MDA-MB-435S cells were trypsinized and then fixed in Shandon cytospin collection fluid (Thermo Electron, Pittsburgh, PA). Cells were stained as described above for the tissue microarray sections for HER2.

**HER2 Detection by Enzyme-Linked Immunosorbent Assay**

To measure the concentration of HER2 protein in cell lines, HER2 protein was detected with the DuoSet IC Human Total ErbB2 enzyme-linked immunosorbent assay (ELISA) (R&D Systems, Minneapolis, MN), in which lysate production and the assay were as described by the manufacturer’s protocol. HER2 concentrations were calculated from the resultant standard curve and are expressed as picograms per microgram of total cell lysate. All standards and samples were measured in duplicate.

**Evaluation of Staining by the AQUA System**

The AQUA system was used for automated image acquisition and analysis as previously described (9). In brief, images of the tissue microarray core sections (histospots) and cell lines were captured with an Olympus BX51 microscope and analyzed with the AQUA software. For each histospot, areas of tumor are distinguished from stromal elements by creating an epithelial tumor mask from the antikeratin protein signal, which was visualized via the Alexa 546 fluorophore. The tumor mask was determined by gating the pixels in this image, in which an intensity threshold was set by visual inspection of histospots, and each pixel was recorded as “on” (tumor) or “off” (nontumor) by the software on the basis of the threshold. The DAPI image, which was used to identify the nuclei, was subjected to a rapid exponential subtraction algorithm that improves signal-to-noise ratio by subtracting the out-of-focus image from the in-focus image. After application of the rapid exponential subtraction algorithm, the signal intensity of the target antigen (e.g., HER2, p53, or ER), which was acquired under the Cy5 signal, was scored on a scale of 0–255. The AQUA score within the subcellular compartments (i.e., nucleus and membrane) was calculated by dividing the signal intensity by the area of the specified compartment. The AQUA score for the cell lines was determined by dividing the signal intensity by the total area under the tumor mask.

**Statistical Analysis**

Statistical analyses for HER2 and p53 protein expression in the tumor microarrays were completed with X-tile (13) and StatView version 5.0.1 (SAS Institute, Inc., Cary, NC). Survival was calculated by the Kaplan–Meier method, and statistical significance of the association between biomarker expression and survival was determined by the Mantel–Cox log-rank test. Plots of univariate relative risks (RRs) of mortality as a function of cutoffpoint of biomarker expression and antibody concentration were constructed by X-tile. All statistical tests were two-sided.

**Results**

To determine whether the antibody concentration used in assays to assess tissue biomarker status affects the association between tumor biomarker status and patient outcome (i.e., disease-specific survival), we used a tissue microarray cohort of 250 breast cancer samples from breast cancer patients with long-term follow-up and three biomarker antigens—HER2, ER, and p53—that had commercially available, well-characterized, and widely used antibodies, one a polyclonal antibody (against HER2) and two monoclonal antibodies (against ER and p53). In these experiments, we used various concentrations of each antibody on different tissue microarray sections and then scored each slide by use of the AQUA system, as described previously (9). We then constructed histograms showing levels of protein expression (data not shown) and used the X-tile program (13) to select the optimal single cutoffpoint for each antigen to separate tumor tissue into a group with high expression and a group with low expression. This software allows optimal parsing of continuous data and then uses the resulting cutpoints in a Kaplan–Meier analysis to examine the association between survival of the corresponding patients and HER2 expression in tumor tissue. When we used a high HER2 antibody concentration (i.e., a dilution of 1:500), we found that low HER2 expression was associated with decreased survival (Fig. 1, A). In contrast, when we used a low HER2 antibody concentration (i.e., a dilution of 1:8000), we found that high HER2 expression was associated with decreased survival (Fig. 1, B). We (14) and others (15) have previously described this effect. Results for p53 were similar to those for HER2; however, results for ER were different. ER expression...
To determine whether conventional immunohistochemistry could detect this effect, we assessed HER2 expression on the same tissue microarray sections, but we used dianaminobenzidine-based brown staining as in conventional assays with antibody dilutions of 1:8000 (which is the current clinical standard concentration for this antibody) and 1:500 (which is a very high concentration for this antibody) and then conducted a Kaplan–Meier analysis (Fig. 2). Although results with neither antibody concentration indicated that the expression of HER2 was statistically significantly associated with breast cancer survival in our cohort, at the 1:8000 antibody dilution, high HER2 expression (i.e., grouping scores of more than 2 and more than 3 together to mimic current clinical practice) was associated with decreased survival (Fig. 2, B). In contrast, at the 1:500 antibody dilution, no association between HER2 expression and outcome was found (Fig. 2, A). Nearly complete overlap of all curves was observed, suggesting that conventional immunohistochemistry did not detect the relationship between antibody dilution and outcome detected with the AQUA-based assay.

Although the Kaplan–Meier analysis of data from the AQUA-based assay showed the effect of antibody concentration on outcome, this illustration was not statistically rigorous because the

![Fig. 1. HER2, p53, and estrogen receptor (ER) antibody concentrations and survival. A tissue microarray containing samples from 250 breast cancers was stained with antibodies against HER2, p53, or ER, at the antibody dilutions indicated, was analyzed with the automated quantitative analysis (AQUA) system. The number of tumors in each group is shown. Solid squares = censored times in the high-level expresser group; solid circles = censored times in the low-level expresser group. Time is shown in months. A) Low (1:500) dilution of HER2 antibody or high concentration. Median cumulative disease-specific survival of patients with tumors with high-level HER2 expression at 5 years was 74.3% (95% confidence interval [CI] = 67.8% to 80.7%; 127 patients at risk), at 10 years was 59.8% (95% CI = 52.4% to 67.2%; 88 patients at risk), and at 20 years was 52.2% (95% CI = 44.4% to 60.1%; 55 patients at risk). Median survival of patients with tumors with low-level HER2 expression at 5 years was 54.3% (95% CI = 37.8% to 70.8%; 19 patients at risk), at 10 years was 24.9% (95% CI = 9.6% to 40.2%; seven patients at risk), and at 20 years was 14.9% (95% CI = 4.6% to 40.2%; four patients at risk). B) High (1:8000) dilution of HER2 antibody. Median survival of patients with tumors with high-level HER2 expression at 5 years was 50.2% (95% CI = 31.0% to 69.5%; 13 patients at risk), at 10 years was 28.5% (95% CI = 10.3% to 46.7%; six patients at risk), and at 20 years was 28.5% (95% CI = 12.0% to 45.7%; three patients at risk). Median survival of patients with tumors with low-level HER2 expression at 5 years was 73.9% (95% CI = 67.6% to 80.2%; 19 patients at risk), at 10 years was 60.1% (95% CI = 52.9% to 67.3%; 92 patients at risk), and at 20 years was 52.8% (95% CI = 45.2% to 60.5%; 57 patients at risk). C) Low (1:500) dilution of p53 antibody. Median survival of patients with tumors with low-level p53 expression at 5 years was 59.5% (95% CI = 46.1% to 72.9%; 30 patients at risk), at 10 years was 38.7% (95% CI = 25.0% to 52.3%; 17 patients at risk), and at 20 years was 33.5% (95% CI = 19.9% to 47.1%; 11 patients at risk). Median survival of patients with tumors with high-level p53 expression at 5 years was 74.6% (95% CI = 68.0% to 81.1%; 122 patients at risk), at 10 years was 60.4% (95% CI = 52.8% to 68.0%; 83 patients at risk), and at 20 years was 54.0% (95% CI = 46.0% to 62.0%; 51 patients at risk). D) High (1:800) dilution of p53 antibody. Median survival of patients with tumors with high-level p53 expression at 5 years was 54.1% (95% CI = 35.7% to 72.5%; 15 patients at risk), at 10 years was 28.8% (95% CI = 12.0% to 45.7%; eight patients at risk), and at 20 years was 28.8% (95% CI = 12.0% to 45.7%; four patients at risk). Median survival of patients with tumors with low-level expression at 5 years was 73.8% (95% CI = 67.7% to 79.9%; 142 patients at risk), at 10 years was 58.9% (95% CI = 51.8% to 66.0%, 94 patients at risk), and at 20 years was 51.3% (95% CI = 43.8% to 58.8%; 59 patients at risk). E) Low (1:100) dilution of ER antibody. Median survival of patients with tumors with high-level ER expression at 5 years was 88.7% (95% CI = 78.3% to 99.2%; 29 patients at risk), at 10 years was 74.4% (95% CI = 61.8% to 87.0%; 32 patients at risk), and at 20 years was 63.1% (95% CI = 46.3% to 79.9%; 12 patients at risk). Median survival of patients with tumors with low-level ER expression at 5 years was 68.5% (95% CI = 61.8% to 75.2%), at 10 years was 52.3% (95% CI = 44.9% to 59.7%; 79 patients at risk), and at 20 years was 46.3% (95% CI = 38.7% to 53.9%; 50 patients at risk). F) Low (1:100) dilution of ER antibody. Median survival of patients with tumors with high-level ER expression at 5 years was 87.8% (95% CI = 78.6% to 97.0%; 42 patients at risk), at 10 years was 74.4% (95% CI = 61.8% to 87.0%; 32 patients at risk), and at 20 years was 63.6% (95% CI = 49.0% to 78.2%; 16 patients at risk). Median survival of patients with tumors with low-level ER expression at 5 years was 66.6% (95% CI = 59.5% to 73.7%; 11 patients at risk), at 10 years was 50.1% (95% CI = 42.3% to 57.8%; 68 patients at risk), and at 20 years was 44.4% (95% CI = 36.5% to 52.4%; 44 patients at risk). All statistical tests were two-sided.
To more rigorously assess the effect of antibody concentration on disease-specific mortality to include all possible cutpoints, we conducted a univariate analysis by creating two populations for each biomarker antigen (one with high expression and one with low expression), and then we plotted the univariate relative risk as a function of each binary cutpoint (Fig. 3).

For example, with a cutpoint that defines the top 15% of HER2 expression as high expression and the other 85% of HER2 expression as low expression and with a low HER2 antibody concentration (i.e., high antibody dilution of 1:8000), high HER2 expression was strongly associated with increased disease-specific mortality (i.e., maximal RR = 1.98, 95% confidence interval [CI] = 1.21 to 3.23), compared with low HER2 expression, as expected. When we used this cutpoint and progressively higher HER2 antibody concentrations (i.e., a 1:2000 dilution and a 1:500 dilution), high HER2 expression was progressively less strongly associated with increased disease-specific mortality, compared with low HER2 expression (for example, for a 1:500 dilution, RR = 0.47, 95% CI = 0.29 to 0.76; P = .002). However, when we used the bottom 15% of HER2 expression as the cutpoint and a high HER2 antibody concentration, the relationship between HER2 expression and disease-specific mortality was reversed, with high HER2 expression being associated with decreased disease-specific mortality, compared with low HER2 expression. Similar relationships were found for p53 expression (Fig. 3, B) but not for ER expression (Fig. 3, C).

For both HER2 and p53, we obtained U-shaped relationships between expression levels and survival. U-shaped relationships occur when, for example, both extremely high and extremely low biomarker expression levels were associated with decreased survival, compared with that of middle-level biomarker expression. These U-shaped relationships have been observed for HER2 and are illustrated here for both HER2 and p53 expression, but not for ER expression, which appeared to be linear with respect to outcome (the more protein expressed, the better the outcome). However, as illustrated in Fig. 3, C for ER expression, the relative risk still varied as a function of cutpoint, and thus it was possible to select cutpoints that would diminish the strength of the association between ER expression and disease-specific mortality; i.e., ER expression is linearly associated with improved survival.

For both HER2 and p53, we used AQUA score, biomarker protein concentration, and antibody concentration in cell lines, we used HER2 because its expression has been measured in many cell lines (16, 17). Published data (16, 17) indicate that an effective assay for HER2 protein must be able to detect from less than 10^3 HER2 molecules per cell to greater than 10^6 HER2 molecules per cell. However, because levels of HER2 expression vary among cell lines, we selected a series of cell lines (Table 1), measured
HER2 protein levels in these cell lines by ELISA, and compared our results with previously published gene amplification data (18–21). One group of cell lines had high HER2 protein expression and multiple copies of the HER2 gene (i.e., B-T-474, SK-BR-3, MDA-MB-361, and MDA-MB-453), and the other group of cell lines lacked or had low HER2 protein expression and no detectable HER2 gene amplification (i.e., T-47D, MDA-MB-435S, BT-549, SW-480, MCF-7, MDA-MB-231, and BAF3). The results of our ELISAs for the construction of the standard curves were more or less consistent with published values (16, 17).

We then used AQUA to analyze these standardized cell lines at two antibody dilutions. At a high concentration of antibody (i.e., a dilution of 1:250), the cell lines segregated into two groups—one with low HER2 expression and one with high HER2 expression. The level of HER2 protein among cells in group with high HER2 expression appears to have saturated the assay. When HER2 concentration in lysates was plotted against AQUA score, the AQUA score of cells with low HER2 expression increased linearly with HER2 concentration, but those of cells with low HER2 expression were indistinguishable from background, probably because of the insensitivity of the assay to detect low concentrations of HER2 protein at the high antibody dilution.

Thus, antigens with a broad range of expression in tumors, such as HER2, may require the use of more than one antibody concentration to accurately assess antigen expression levels in cells. This observation suggests that tumors with very low levels of HER2 expression would not be accurately evaluated at the high antibody dilutions typically used in the clinical setting.

**DISCUSSION**

By careful quantification of the exact levels of expression of a series of tissue biomarkers, we have illustrated key weaknesses in immunohistochemistry approaches, the current standard for evaluation of protein expression in situ. Specifically, we found that the antibody concentration (the dilution), which is typically defined at the relative risk of disease-specific mortality of 1.0. The value of the y axis is the relative risk for disease-specific mortality for the high group (the patients above the defined cutpoint) compared with the reference group (below the defined cutpoint). A) Solid arrow = cutpoint at the top 15% of HER2 expression; open arrow = cutpoint at the bottom 15% of HER2 expression.
empirically determined, can dramatically affect, and even reverse, apparent relationships between biomarker and outcome. The reason for this observation is that the concentration range of the antibody-based immunohistochemistry assay is insufficient to span the range of expression of some biomarker proteins in various cells and tissues. When a part of the range of expression of a biomarker is obscured by the insensitivity or saturation of an assay, some specimens are essentially excluded from analysis, so that the entire population cannot be evaluated. When a U-shaped relationship exists between biomarker expression and outcome (when outcome is plotted as a function of protein expression level), evaluation of only the lower expression range of the population may result in associations between biomarker expression and survival that are opposite from those obtained with only the higher expression range.

The traditional “brown stain” immunohistochemistry analysis, although robust and well accepted, is designed to detect context (i.e., tissue regions that are stained are compared with those regions that are not stained) rather than to linearly measure intensity across a broad dynamic range. The use of the AQUA assay expands the range of immunoassays and allows better quantification. However, conventional brown stain–based and immunofluorescence-based assays have a limited concentration range because of the enzyme amplification used and other factors. We estimate the functional range for conventional brown stain–based assays is between 1 and 1.5 orders of magnitude (data not shown). The functional range for fluorescence-based assays is between 1.5 and 2.5 orders of magnitude (see Fig. 4). This limitation may decrease the accuracy of the analysis and also obscure associations with expression because of the saturation or insufficient sensitivity of the assay. This problem may explain some of the variability in the p53 and HER2 literature and may also explain the U-shaped relationship that has been previously observed between HER2 expression and outcome (13, 14). However, this study has limitations. Specifically, the number of patients in each study set is relatively small, and thus the results should be validated in both larger cohorts and with other antibody–antigen pairs.

In summary, this study highlights the problems of subjective analysis of immunohistochemistry studies. When a high concentration of antibody was used, subtle differences in the level of HER2 expression in cells with low-level HER2 expression could be discerned, but because of assay saturation, cells with high-level HER2 expression could not be measured. Conversely, when a low concentration of antibody was used, HER2 levels in cells with a high-level HER2 expression could be distinguished from those in cells with very high-level HER2 expression; however, this concentration of antibody could not distinguish HER2 levels in cells with low-level HER2 expression from background. A similar relationship was observed for p53 and may also be apply to other biomarkers, although not to ER. The nonquantitative nature of conventional immunohistochemistry studies and the lack of validated control cell lines that can be used to make standard curves for biomarker expression levels can result in incomplete or even conflicting conclusions.

The implications of our results for current practice are unclear. This study illustrates that a greater awareness of an antigen’s dynamic concentration range is needed to rigorously assess tissue biomarker associations with outcomes. However, it is not clear that conventional brown stain immunohistochemistry can be used for this task. As the requirement for increased accuracy for measurement of in situ protein concentration increases, new methods, such as the AQUA system, may be used.
more often. We believe it is likely that exact amounts of multiple proteins in specific subcellular compartments will be measured to assist in the biological characterization of a tumor, such as identifying pathways that are activated or inactivated in cancer, to aid in the selection of targeted therapies. We will probably see the increased use of validated control cell lines to ensure that the appropriate concentration of antibody can be used to detect the antigen of interest in the linear range of the assay. This level of accuracy, quality control, and quality assurance has been established in many areas of evidence-based medicine, so that we believe that it should be only a matter of time before anatomic pathology establishes the same standards for its assays.

REFERENCES


NOTES

Editor’s note: Drs. Camp and Rimm are founders, stockholders, and consultants to HistoRx, a private company to which Yale University has given exclusive rights to produce and distribute the software and technologies embedded in AQUA; Yale University retains patent rights for the AQUA technology. A. McCabe and M. Dolled-Filhart contributed equally to this work.

Maria Dolled-Filhart is supported by the United States Army Breast Cancer Research Grant DAMD17-03-1-0349. Dr. Camp is supported by NIH Grant K0-8 ES11571, and both Drs. Rimm and Camp are supported by the Breast Cancer Alliance of Greenwich, CT. Dr. Rimm is supported by a grant from the Patrick and Catherined Weldon Donaghue Foundation for Medical Research, NIH grant NCI R21 CA100825, and United States Army Grant DAMD-17-02-0463. Funding to pay the Open Access publication charges for this article was provided by the Department of Pathology, Yale University School of Medicine. Manuscript received March 24, 2005; revised October 7, 2005; accepted October 28, 2005.