Assessment of Interlaboratory Variation in the Immunohistochemical Determination of Estrogen Receptor Status Using a Breast Cancer Tissue Microarray

Robin L. Parker, MD,1 David G. Huntsman, MD,1,3 David W. Lesack, MD,4 James B. Cupples, MD,5 Dennis R. Grant, MD,6 Majid Akbari, MD,3 and C. Blake Gilks, MD1

Key Words: Breast cancer; Tissue microarray; Interlaboratory variation; Estrogen receptor; Immunohistochemistry

Abstract

The determination of tumor cell estrogen receptor (ER) expression status by immunohistochemical analysis has become standard practice, yet assay reproducibility has not been assessed adequately. By using a breast cancer tissue microarray, we examined interlaboratory variability in ER reporting. A 2-fold redundant tissue microarray block was created from 29 breast cancers. Unstained slides were distributed to 5 laboratories, and each laboratory immunostained and scored 1 slide for ER. Interlaboratory agreement ranged from moderate to high (overall kappa = 0.54 for 0-3+ grading; overall kappa = 0.84 for negative vs positive assessment of ER status). When 1 observer scored each of the 5 slides, interlaboratory agreement was slightly better (kappa = 0.63 for 0-3+ scoring; kappa = 0.96 for negative vs positive scoring). One laboratory, which had used an antibody and antigen retrieval method different from the others, demonstrated only fair concordance with the other 4 laboratories, but there was substantial intralaboratory interobserver agreement and excellent agreement with an outside observer reviewing the slide stained in that laboratory. The tissue microarray was an efficient and effective tool for identifying variability in ER reporting and should prove valuable in other external quality assurance programs.

The determination of estrogen receptor (ER) status has become a standard practice in the evaluation of patients with invasive breast cancer, having important prognostic and therapeutic implications.1 During the 1990s, immunohistochemical assays have replaced biochemical assays as the method of choice for the evaluation of ER levels, a change predicated on issues of decreasing tumor size at diagnosis because of earlier detection, cost, and convenience. Many studies have demonstrated good concordance between the 2 methods,2,3 and ER immunohistochemical analysis has been shown to be at least as good as biochemical assays at predicting response to adjuvant endocrine therapy.4

One aspect of the immunohistochemical assessment of ER that has not been assessed adequately is the extent to which the results are reproducible between different observers and different laboratories. As with any immunohistochemical assay, the results are subject to variability owing to differing antibodies, staining methods, and interpretation.5,6 Considering the importance of ER determination to treatment decisions for patients with breast cancer, external quality assessment of the immunohistochemical assay should be performed regularly to ensure that reporting is similar between laboratories.

Tissue microarrays are a recently developed technology that permits the assessment of hundreds of tissue cores on a single glass slide.7 With this technology, multiple tissue samples can be analyzed in parallel using immunohistochemical analysis, fluorescence in situ hybridization, or RNA in situ hybridization.8 The technology greatly increases the efficiency of tissue-based research. In addition to potential applications in the molecular profiling of tumors, tissue banking, and optimization of antibodies and probes, tissue arrays may be valuable in
external quality assessment programs for histology laboratories. Specifically, they could facilitate the standardization of immunohistochemical staining and interpretation.

The objectives of this study were to assess interobserver and interlaboratory variability in the immunohistochemical determination of ER using tissue microarray technology and to identify the major factor(s) contributing to variability in ER reporting.

Materials and Methods

Study Material

Study material was derived from 29 invasive breast cancers diagnosed at the University of British Columbia Hospital, Vancouver, between April and December 2000. The cases included 26 primary breast tumors, 2 axillary lymph node metastases, and 1 chest wall recurrence. We reviewed the ER reports and retrieved the H&E-stained slides and paraffin-embedded tissue blocks that corresponded to the whole sections on which the initial ER immunohistochemical assays were performed. ER staining was scored from 0 to 3+ based on the percentage of tumor nuclei staining and the staining intensity as follows: 0, 3% or fewer of tumor nuclei stained; 1+, more than 3% of nuclei stained weakly; 2+, more than 3% of tumor nuclei stained moderately intensely; and 3+, more than 75% of nuclei stained strongly. Of the 29 cases, 6 cases were interpreted originally as negative for ER (score = 0), and 23 were interpreted as positive (3 scored as 1+, 10 scored as 2+, and 10 scored as 3+).

Tissue Microarray

For each case, an area of invasive breast cancer was identified on the H&E-stained slide and circled. The marked slide was aligned with the surface of the corresponding donor block to guide sampling of the tumor. By using a tissue microarrayer (Beecher Instruments, Silver Spring, MD), the area of interest in the donor block was cored twice with a 0.6-mm diameter cylinder and transferred to a recipient paraffin block for construction of a 2-fold redundant tissue microarray block.

Immunohistochemical Analysis

We cut 2-μm-thick sections from the tissue microarray block and transferred them to slides using a water-bath technique. One H&E-stained section and 3 unstained sections on 3-amino propyltriethoxy-silane–coated slides were distributed to each of 5 hospital laboratories in British Columbia. The laboratories included 1 university hospital and 4 community hospitals.

Each laboratory was asked to immunostain 1 of the unstained slides for ER using the antibody and staining method normally used in that laboratory (the additional unstained sections were provided for negative controls and backup if technical problems occurred). At least 1 pathologist in each laboratory interpreted the staining, applying the reporting system normally used by the laboratory. The ER-stained slides were returned to the university hospital, where 1 pathologist, blinded to the results of the other observers, also scored the staining for each slide.

Results

Of the 58 cores of invasive breast cancer represented on the tissue microarray, 5 were reported as inadequate for evaluation by at least 3 of the 5 laboratories and excluded from the analyses of interobserver and interlaboratory agreement. The reasons given for inadequacy included insufficient numbers or absence of tumor cells in the core section for evaluation (n = 2) and absence of the core section on the immunostained slide (n = 3). Image 1 shows H&E- and ER-stained core sections from the tissue microarray.

Pathologists from 4 of the laboratories applied a similar 4-point scoring system to that used in the evaluation of the original ER-stained whole sections (see the “Materials and Methods” section) in which the proportion and intensity of positively stained nuclei are assessed. The cutoff point for positivity ranged from 3% to 10% of tumor nuclei staining. Table 1 shows the distribution of ER scores reported by the 4 laboratories. While the distributions of scores were similar for 3 of the laboratories, one laboratory (laboratory 4) reported fewer cores as 3+ and more as 0, 1+, and 2+ (ie, generally weaker staining of the cores).

When a 4-point scoring system for ER staining is used in British Columbia, cases scored as 0 are interpreted as negative for ER and those scored as 1+, 2+, or 3+ as positive. The fifth laboratory scored cases only as negative (3%) or fewer of tumor nuclei stained) or positive (>3% of tumor...
nuclei stained). Table 2 shows the number of cores reported as positive or negative by each of the 5 laboratories. More cases were reported as negative by 1 laboratory (laboratory 4) compared with the others.

Assessment of interlaboratory agreement for the 4 laboratories whose pathologists applied the 0 to 3+ scoring system revealed an overall kappa of 0.54. Pairwise kappa analyses showed good concordance (kappa = 0.63-0.69) between 3 of the laboratories (1, 2, and 3), but only fair concordance between these laboratories and laboratory 4 (kappa = 0.41-0.46). As demonstrated in Figure 1, laboratory 4 reported lower intensity ER staining for several of the tumor cores compared with laboratory 1, resulting in a lower kappa value. This also was observed when laboratories 2 and 3 were compared with laboratory 4. Translation of the scores to negative or positive ER determinations and inclusion of the fifth laboratory resulted in excellent agreement (overall kappa = 0.84), although laboratory 4 again demonstrated less agreement with the other laboratories (kappa = 0.64-0.77). As shown in Figure 2, laboratory 4 reported several tumor cores as negative for ER that laboratory 1 reported as positive, resulting in a lower kappa value. The same was observed when laboratory 4 was compared with laboratories 2, 3, and 5.

To further assess interlaboratory differences in ER staining without the variable of different observers, 1 observer scored the ER-stained slides returned by the 5 laboratories. The interlaboratory agreement was somewhat better, with an overall kappa of 0.63 when staining was scored 0 to 3+ and an overall kappa of 0.93 when staining was interpreted as negative or positive. Again, there was poor concordance between the scores for the microarray section stained by laboratory 4 compared with those stained by the other laboratories, with kappa values ranging from 0.40 to 0.48.

To assess whether the variability in ER determination by laboratory 4 was due to technical factors or interpretation, we assessed interobserver agreement. The slide stained for ER by laboratory 4 was reviewed independently by 4 other pathologists in that laboratory, who scored the staining as 0 to 3+. The agreement among the 5 pathologists was high (overall kappa = 0.76). In addition, there was good agreement (kappa = 0.85) between the original observer in laboratory 4 and an outside observer. Thus, interobserver variability in the interpretation of ER reporting was low, suggesting that methodologic differences rather than variation in interpretation were the main contributors to the observed interlaboratory variability in ER reporting.

To determine what technical factors may have contributed to variability in ER reporting, we examined the staining method used by each of the 5 laboratories and found substantial differences between the method used by laboratory 4 and the methods used by the other 4 laboratories.
Whereas laboratory 4 used antibody ID5 (DAKO, Glostrup, Denmark), the other laboratories used antibody 6F11 (Novocastra, Newcastle upon Tyne, England). For antigen retrieval, laboratory 4 preheated a glycerin solution to 120°C in a microwave oven, then placed the slide in the heated solution on a hotplate for 7 minutes. The other laboratories used citrate buffer and pressure cooker techniques (1 laboratory used a bench top autoclave, and 3 used plastic pressure cookers in microwave ovens), with times ranging from 7 to 20 minutes in the pressure cooker. Other differences identified were that laboratories 1 through 4 used automated staining and detection systems according to the manufacturer’s recommendations, whereas laboratory 5 stained manually.

To assess intralaboratory and intraobserver variation, one tissue microarray section was stained and interpreted for ER by the same pathologist and laboratory that assessed the original tumor sections for ER. Six cores were inadequate for assessment on this section. When scored 0 to 3+, the core results were the same as the original reported results in 35 (67%) of 52 cases. When assessed as positive or negative, the core results agreed with the original results in 50 (96%) of 52 cases. There was good concordance between the 2 cores from each tumor: 84% when scored 0 to 3+ and 99% when assessed as negative or positive.

**Discussion**

During the 1990s, immunohistochemical assays largely replaced biochemical methods, such as the dextran-coated charcoal assay, for determining the ER status of patients with breast cancer. A number of studies have demonstrated good agreement between the 2 methods, yet few studies have examined the reproducibility of the immunohistochemical assays. Treatment of breast cancer in the province of British Columbia is uniform and protocol driven; therefore, it is important to ensure that the reporting of ER status is similar between different observers and different laboratories.

The use of a breast cancer tissue microarray facilitated this assessment. With 58 cores on a single glass slide, interpretation of variable degrees of ER staining, as performed by 5 laboratories, was readily assessed. We showed significant variance by 1 laboratory, which reported weaker staining for several of the cores. Compared with using whole tissue sections for quality assessment, this method was more efficient and less costly, requiring less time on the part of the technical staff and pathologists and less antibody and reagent. The core ER results (ie, positive vs negative) were the same as the original reported results, based on whole sections of the tumors, in 96% of cases (50/52). A recent study that examined validation of tissue microarray technology for immunohistochemical assays, including ER, also found that the analysis of 2 core sections from 1 case was comparable to the analysis of whole tissue sections in more than 95% of cases. In another study of more than 2,000 bladder cancers assessed for histologic grade and proliferative index, information obtained from four 0.6-mm cores per case was highly concordant with that obtained from the whole sections. These findings demonstrate that intratumoral heterogeneity should not be a major impediment to use of arrays in quality assurance studies.
 Ideally, the assessment of ER staining and interpretation would be standardized, so that results are reproducible between different observers and different laboratories. In addition to using the same antibody and staining method, a single reproducible scoring system would be applied. At this time, no universally accepted immunohistochemical method for ER determination exists. The present study demonstrated that laboratories that use the same antibody and similar staining methods have similar results. The laboratory that used a different antibody and antigen retrieval technique had weaker staining and fewer positive cases compared with the other laboratories. One or both of these variables were certainly the major contributing factors to the observed variability in ER reporting. In another recent study, antigen retrieval was identified as the single most important factor contributing to the overall reliability of ER determination by immunohistochemical analysis.6 To ensure that ER reporting is similar among laboratories in British Columbia and that patient management is consistent, the laboratory that was using a different antibody and antigen retrieval technique has since opted to use the same antibody and antigen retrieval method as that used by the other laboratories in this study.

While interpretation of staining also could contribute to variability in ER reporting, we found that intralaboratory interobserver and interlaboratory interobserver agreement were generally high. These results not only confirm that variation in staining methods was the main contributor to variability in ER reporting but also indicate that both the 2-point (ie, positive vs negative) and the 4-point (ie, 0-3+) scoring systems are quite reproducible. This is despite the fact that cutoff points for positivity varied between 3% and 10% of tumor nuclei staining. Just as there is a lack of standardization of staining methods, there is a lack of standardization of stain interpretation. Selected cutoff points for ER positivity have been variable among different studies.1,3,5,14 As emphasized by others,1,3 a defined cutoff point is difficult to establish, and the relevant question of what levels of ER staining are associated with better prognosis and are predictive of response to endocrine therapy has not been answered definitively. In a study involving more than 1,000 patients with breast cancer and the 6F11 antibody, one group found that tumors with as few as 1% positive tumor cells were associated with improved clinical outcome.15

The tissue microarray proved to be an effective and efficient tool for assessing interlaboratory variation in ER staining and interpretation. To our knowledge, this is the first report of a tissue microarray being used for this purpose. From our results, we would advocate the use of tissue microarrays in external quality assessment programs such as this one. This could facilitate the standardization of immunohistochemical, fluorescence in situ hybridization, and other molecular assays, so that results are reproducible between laboratories. The tissue microarray also could be used in internal quality control (eg, quarterly or monthly) to ensure that there is no drift of staining results or reporting within a laboratory. We believe that the tissue microarray would be an improvement over the current practice of using a single strong positive control for quality assessment.

From the 1Department of Pathology and Genetic Pathology Evaluation Centre, Vancouver Hospital and University of British Columbia, Vancouver; 2Department of Pathology, British Columbia Hospital

© American Society for Clinical Pathology
References


