Immunohistochemical Staining of Estrogen and Progesterone Receptors: Aspects for Evaluating Positivity and Defining the Cutpoints

PÄIVI JALAVA1,2, TEIJO KUOPIO3, RIIKKA HUOVINEN4, JUHANA LAINE1 and YRJÖ COLLAN1

1Department of Pathology, University of Turku, Turku; 2Department of Internal Medicine and 3Department of Pathology, Jyväskylä Central Hospital, Jyväskylä; 4Department of Oncology, Turku University Central Hospital, Turku, Finland

Abstract. Background: Estrogen (ER) and progesterone (PgR) receptors are predictive and prognostic factors in breast cancer. The most suitable immunohistochemical cutpoints for dividing the tumors in hormone receptor-negatives and -positives may, however, need more consideration. We examined the association between breast cancer survival and cutpoints assessed by four different models. We looked for evidence for which patient subgroups could be handled best through applying different cutpoints. Materials and Methods: Three hundred and twenty-four samples of invasive breast cancer were immunohistochemically stained for ER and PgR, bcl-2 and erbB2. Fractions of ER- and PgR-positive cells and also ER and PgR staining scores were assessed. The fractions of stained cells and staining scores, respectively, were determined on the whole section area, and the area of most intense staining. Candidate cutpoints, dividing the patients into good and poor prognosis groups, were tested among all patients group, N+ and N– groups, premenopausal and postmenopausal patient groups. The correlation between immunohistochemistry results of ER, PgR, bcl-2 and erbB2 as well as SMI (standardized mitotic index), patient age, tumor size and axillary lymph node status were tested. Results: The ER score was correlated with age, SMI and bcl-2 positivity. The PgR score was correlated with erbB2 and bcl-2. Lobular carcinomas had higher staining scores of ER and PgR than ductal carcinomas. Conclusion: In this material, ER was correlated with factors reflecting the differentiation of the tumor. On the basis of the ER and PgR immunohistochemistry cutpoint analysis, we found that the optimal cutpoints in different patient groups may not be the same.

Correspondence to: Päivi Jalava, Department of Pathology, University of Turku, Finland. e-mail: paivi.jalava@utu.fi

Key Words: Immunohistochemistry, breast cancer, estrogen and progesterone receptors, cutpoint analysis, prognosis.

Adjuvant systemic chemotherapy and endocrine therapy improve survival of breast cancer. Adjuvant treatment options are selected according to the age of the patients and according to the prognostic factors, of which estrogen (ER) and progesterone (PgR) receptors are most important (1). Assessment of ER and PgR receptors predict response to adjuvant therapy. Many studies have shown that the receptor status may not be an effective prognosticator of relapse and breast cancer death.

Before the time of immunohistochemical (IHC) assays, ER and PgR used to be measured by the dextran-coated charcoal method. The cutoff for positivity was considered 10 fmol or more ER per 1 mg cytosol (1). Nowadays, ER and PgR are assessed on formalin-fixed, paraffin-embedded tissue sections, and tumors are classed as either positive or negative according to the immunohistochemical staining characteristics of tumor cell nuclei (staining intensity, fraction of stained nuclei). Because there are no unambiguous cutpoints available for hormone receptor positivity based on IHC, we decided to study how the association between hormone receptor status and survival of breast cancer depended on the model of determining the positivity of tumors. The selection of cutpoints may have an important role in defining the importance of a staining model in prognostication (2). It is possible that the optimal cutpoint might not be the same in all patient subcategories.

Materials and Methods

The data consists of 324 female primary breast cancer patients, all treated at Turku University Central Hospital, Finland, from 1988-91. All patients had undergone surgical operation. Most of the patients (275) were treated by breast ablation and some (46) by breast conserving technique. Adjuvant chemotherapy was given to 24 (7.4%) out of 323 patients (one patient data missing), whereas adjuvant endocrine treatment was given to 56 (17.3%) patients. Two patients underwent surgical removal of the ovaries.
Table I. Basic clinical and histological information on the studied breast cancers (n=324).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean (SD)</th>
<th>Median</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor size (cm)</td>
<td>2.35 (1.5)</td>
<td>2.00</td>
<td>0.17-12</td>
</tr>
<tr>
<td>Median age</td>
<td>56.4</td>
<td>56.3</td>
<td>31-87</td>
</tr>
<tr>
<td>Range (years)</td>
<td>12.4</td>
<td>12.4</td>
<td>0.17-12</td>
</tr>
<tr>
<td>Axillary lymph node status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive, n (%)</td>
<td>106 (32.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative, n (%)</td>
<td>218 (67.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histological type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ductal</td>
<td>272 (84.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lobular</td>
<td>28 (8.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>24 (7.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histological grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I, n (%)</td>
<td>68 (21.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II, n (%)</td>
<td>177 (56.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III, n (%)</td>
<td>71 (22.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not known</td>
<td>8 (2.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The maximum diameter of the tumor was measured by the operating surgeon or, in many cases of non-palpable lesions, by the pathologist. Histological type and grade were evaluated according to the WHO classification (3). Standardized mitotic index (SMI; M/Vindex) was evaluated from the most cellular areas at the periphery of the tumor (4). The mean of SMI (mitoses/mm²) was 12.8, SD 16.7, range 0-117, and median 6.63. The follow-up data and the causes of death were based on patient files, autopsy reports and death certificates. The mean follow-up time was 5.8 years. The basic data about the tumors is shown in Table I.

The IHC was done according to Taylor et al. (5). Bcl-2, erbB2 and ER were stained manually, as described by Jalava et al. (6, 7). The commercial mouse monoclonal antibodies were used for ER (ER, DAKO M7047, dilution 1:500) and for PgR (PgR, Novacstra, NCL-PgR, and dilution 1:20). The PgR staining was done by using the automated TechMate™500 Plus (DAKO, BioTek, Solutions Inc., model TM500-220, TMS20185, USA) stainer in routine use at the Department of Pathology, Turku University Central Hospital. The principles of the latter were similar to those used in the manual staining procedure. The tissue sections were first deparaffinized with xylene and then washed in descending series of ethanol and TBS-buffer (0.05 M TRIS-buffered physiological saline). Then, the sections were subjected to microwave oven treatment for the retrieval of antigen epitopes (two times for 5 minutes) in 10 mM aqueous sodium citrate. After that, the antibody was applied.

Immunohistochemically-stained ER and PgR were analyzed, respectively, using similar principles. First, the most intensively stained area was chosen using 10x objective magnification. Four staining categories were used: 0=negative staining, only blue background color could be seen on nuclei; 1=positive brown staining could be seen with difficulty and background color could be seen easily; 2=moderate positivity, brown color could be seen better than the blue background color; 3=strong positivity: background color could not be seen. Then, using 40x objective magnification, the fractions of tumor cells representing each staining category were evaluated. The formula for staining score (S) was used: S=p1x1+p2x2+p3x3 in which p1, p2 and p3 represented fractions of cancer cells with respective staining categories 1-3 (8). We registered both the fractions of stained cells and the score values. After that, the whole section area was evaluated and the fraction of stained cells (the sum of the fractions of different staining categories), as well as the score based on the whole section, was calculated. To summarize: we had four different measures for both hormone receptors: fractions and score analyzed from the most intensely stained area (AA=area analysis) and fractions and score analyzed from the whole section area (WSA=whole section area analysis). The score values were used in all ER and PgR analyses, and the fractions of stained cells only in the ER and PgR cutpoint analyses.

Bcl-2 positivity was analyzed in microscopy in four categories, and the cutpoint between second and third categories divided the material into two categories, “positive” and “negative” tumors (6). The erbB2 immunostaining index was calculated according to the formula by Lippomen et al. described above (7, 8).

Statistical analyses. Three hundred and twenty-four tumor samples were available for ER analysis and 264 out of them for PgR analysis. All statistical investigations were performed using the SAS program package for Windows release 6.12 (SAS Institute, Inc., Cary, NC, USA). Axillary lymph node-positive (N+) and -negative (N–) groups were compared in respect to ER and PgR scores. The correlation of ER and PgR scores with SMI, tumor size, age and erbB2 score (7), were analyzed and Pearson’s correlation coefficients determined. The ER and PgR scores were then studied in respect to the most common histological breast cancer types (ductal infiltrating and lobular infiltrating carcinomas) and histological grade. ER and PgR scores (WSA) and fractions of stained cancer cells (WSA) were compared with bcl-2 staining using Mann-Whitney U-test. The distribution of hormone receptor staining scores were also analyzed, respectively (Figure 1).

To find the optimal cutpoints [the cutpoint with the highest statistical significance (lowest p-value) in respect to survival] for both hormone receptor scores (ER, PgR), all quartiles and 1%, 5%, 10%, 90% and 99% score values were tested one by one in univariate analysis and, in this way, candidate cutpoints found. The candidate cutpoints for fractions of stained cancer cells were also tested (1%, 5%, 10%, 15%, ..., 85%, 90% and 95%) in a similar way. All of these analyses considered both WSA and AA, and all of them were studied among all patients group, N+ and N–patients, as well as among patients ≤52 (premenopausal) and >52 (postmenopausal) years of age. The cutpoints with lowest p-values were considered most optimal for dividing patients into two groups differing in survival. After defining the optimal cutpoints, the survival analyses were performed with Cox proportional hazard
model, and $p$-values, hazard ratios and confidence intervals were estimated for each tested cutpoint.

**Results**

The age of the patients and the ER score, based on evaluation of both WSA and AA, were positively correlated, but Pearson’s correlation coefficients were low ($r=0.25$, $r=0.29$, respectively, $p<0.0001$). The PgR score did not show any kind of correlation with the age of the patient.

The SMI and ER scores were inversely correlated ($r=-0.16$, $p=0.0015$ and $r=-0.18$, $p=0.004$, respectively), irrespective of the method of evaluation. The PgR scores did not show any correlation with this proliferation parameter. Tumor size had no correlation with hormone receptor scores.

The medians of both hormone receptor scores were higher among tumors with low histological grade. ErbB2 and PgR were inversely correlated but ER and erbB2 had no significant correlation ($p=0.5122$ in WSA). There was no correlation between axillary lymph node status and hormone receptor statuses. Lobular carcinomas had higher ER and PgR scores than ductal carcinomas on average. The Mann-Whitney $U$-test for bcl-2 and hormone receptor status gave a statistically significant difference in both cases. ($p<0.001$, respectively).

In the cutpoint analysis for the fraction of ER-stained cells, two minimum $p$-values were found among all patients. In the WSA, the ER fractions of 15% and 30% gave the lowest $p$-values ($p=0.037$ and $p=0.0682$, respectively) which, however, were not dramatically different from $p$-values at adjacent tried cutpoints. The similar analysis from the AA also gave the $p$-value minima at 15% and 30% which, however, were not statistically significant. Among N- patients, no real candidate cutpoints were found at all with either method of analysis, whereas among N+ patients, 30-35% gave the lowest $p$-values ($p=0.0487$ for both) in WSA. The AA showed the same two minima which, however, were not statistically significant. Among postmenopausal patients, there were no significant candidate cutpoints, but among premenopausal patients, significant cutpoints were found at 45% and 65% ($p$-values 0.0364 and 0.0439, respectively) in WSA, and at 50% ($p=0.0364$) in AA.
The ER score of the WSA had a p-value minimum at 0.75 (p=0.0402) among all patients. Basically, this score corresponds to cases showing the 15-30% fraction cutpoint. The analysis based on AA did not improve significance. Among N– patients, there were no significant cutpoints (as there were not in the analysis of the fraction of stained cells). N+ patients, on the other hand, had a significant minimum at 0.75 (p=0.0186; WSA), see Figure 2. Postmenopausal patients did not show statistically significant minima. In the analysis of the ER score, patients under 52 years of age had almost significant p-values in the WSA at 0.75 (p=0.0773), and in AA at 0.83 (p=0.0783). Table II shows ER scores and respective p-values as an example in table format.

Table II. ER scores and respective survival-associated p-values among different patient subgroups assessed by the whole section analysis (WSA). The most significant p-values are shown in bold.

<table>
<thead>
<tr>
<th>ER score</th>
<th>All patients</th>
<th>N−</th>
<th>N+</th>
<th>≤52 years</th>
<th>&gt;52 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=324</td>
<td>n=218</td>
<td>n=106</td>
<td>n=109</td>
<td>n=215</td>
</tr>
<tr>
<td>0.00</td>
<td>0.0577</td>
<td>0.0547</td>
<td>0.4826</td>
<td>0.0793</td>
<td>0.4588</td>
</tr>
<tr>
<td>0.75</td>
<td>0.0402</td>
<td>0.4560</td>
<td>0.0186</td>
<td>0.0773</td>
<td>0.3738</td>
</tr>
<tr>
<td>1.55</td>
<td>0.5928</td>
<td>0.4619</td>
<td>0.3840</td>
<td>0.9906</td>
<td>0.2187</td>
</tr>
<tr>
<td>2.05</td>
<td>0.5818</td>
<td>0.9907</td>
<td>0.9521</td>
<td>0.9906</td>
<td>0.9407</td>
</tr>
<tr>
<td>2.45</td>
<td>0.2745</td>
<td>0.9927</td>
<td>0.2656</td>
<td>-</td>
<td>0.7916</td>
</tr>
</tbody>
</table>

In the analysis of PgR positivity, no statistically significant cutpoints were found among all patient groups. Among N– patients, there were no potential cutpoints in either WSA or AA. PgR positivity of N+ patients, on the other hand, showed a significant minimum at 5% in the WSA (p=0.0425) and at 55% when the evaluation was based on the AA (p=0.0301). In the latter analysis, 5% was almost significant (p=0.0504). Among premenopausal patients, no significant minima were found. The postmenopausal patient...
group had a significant minimum p-value \((p=0.0327)\) at 50\% (WSA), Figure 3.

The PgR score in the WSA showed the best cutpoint at 0.04 among all patients \((p=0.0377)\), Figure 4, and among N+ patients \((p=0.0220)\), Figure 5. Other significant cutpoints were not found.

Discussion

In this study, there was a slight positive correlation between the age of the patients and ER status (the younger the patient, the lower the ER content). Postmenopausal patients often have immunohistochemically ER-positive tumors (9-11), which may reflect the different biological background of these tumors compared to the tumors of premenopausal patients. Generally, the oldest patient group (>70-75 years), as well as the youngest one (<35 years), tend to have the poorest prognosis (12-16). In our study, regardless of the method of analyzing the section (WSA vs. AA), ER positivity seemed to be correlated with the older age of the patients. PgR did not show correlation with age. This is one of the reasons why the biological significance of ER and PgR determinations cannot necessarily be considered equal. Also, others have reported similar results (10, 17).

SMI is a very powerful prognostic factor (18). In this study, SMI and ER were inversely correlated, which is understandable because ER-positive patients tend to have better prognosis than ER-negative ones (19, 20). Tumors having high proliferating activity (e.g. high SMI) are less differentiated and have poorer prognosis. There was no statistically significant correlation between PgR and SMI in our material, but the trend seemed to be inverse. Perhaps in larger material, the result might have reached statistical significance. Also in previous studies, hormone receptors have reflected well the differentiation of tumors which is related to low proliferation activity (9, 21-23).

Well-differentiated (grade 1) breast cancers are often bcl-2-positive, which is associated with favorable outcome, even though bcl-2 is not considered an independent prognostic factor (24-26). In this study, bcl-2 positivity showed strong association with both ER and PgR, which is in line with previous findings (6, 27). It is, however, hormone receptor status rather than bcl-2 status which is believed to influence response to endocrine treatment (28-30).

It is interesting that we found no correlation between ER and erbB2. Others have reported negative correlation between these prognostic markers (31-35) and even supposed that estrogens could in some way specifically suppress erbB2 expression (36, 37). These two factors seem to be independent in their effect on neoplastic proliferation. ErbB2 and PgR were inversely correlated in our material, which requires explanation. Perhaps, in some way, PgR positivity protects the tumor from erbB2 mutation. A negative correlation between PgR and erbB2 has also been reported by others (33, 35). On the other hand, PgR activity may reflect differentiation more than proliferation. In this sense, PgR-positive tumors could slow down the induction of proliferation by erbB2.

The axillary lymph node status showed no correlation with hormone receptor statuses, which is in line with previous findings (9, 21, 38, 39). This is relevant because the ability of the breast cancer tumors to metastasize depends on features other than hormone receptor status only. Lobular carcinomas are well-differentiated tumors and tend to be ER-positive (9, 40), although not all investigators have obtained this result (41).

Among all patients, the cutpoints of 15\% and 30\% of ER-positive cancer cells were the best to predict survival. In this light, the widely used cut-off value of 20\% may be a reasonable choice for ER analyses among all patients. The best ER score cutpoint of 0.75 reached statistical significance, possibly reflecting the same 30\% cutpoints as found in the percentage analysis (20\% and 30\% ER-positive cancer cells representing staining intensity 3, would give scores 0.6 and 0.9, respectively).

Among N+ patients in WSA, 30\% seemed to be a good cutpoint for dividing the patients into two groups which differed in survival. The ER score 0.75, among N+ patients in WSA, also reached statistical significance, reflecting the relevance between the fraction analysis and scoring system. We could find no good cutpoints for ER among the postmenopausal patient group, either using the proportion of positively-stained cancer cells or the scoring system. Among premenopausal patients, 45\% and 65\% were
statistically significant in the WSA and 50% when the AA was used. However, calculating the ER scores in this group of patients did not confirm this, but gave nearly significant cutpoints at scores 0.75 and 0.83.

The fractions of PgR-stained cells and the cutpoint analysis judged on the basis of p-values did not seem to be reasonable in either the all patients group, or N- patients or premenopausal patients: there were no good cutpoints. Among N+ patients, however, PgR positivity of 5% in the WSA and 55% in the AA gave the best cutpoints. Might this mean that even a small amount of strongly PgR-positive cancer cells in these N+ breast cancers could be a favorable feature and this could be seen as a lot of lower percentage when the whole sample area was considered?

ER among premenopausal patients gave the best cutpoints at quite high fractions (45% and 65%), and PgR among postmenopausal patients had the statistically significant minimum at as high as 50% (WSA). In the PgR score analysis in WSA, N+ patients showed significant differences in survival at score 0.04. This is reflected in all patients group in the PgR score analysis. All the PgR results, both scores and fraction of stained cancer cells, seemed to be clearly lower than the ER values, which may partly depend on the immunohistochemical antibodies used and partly on the true biological difference in the expressions of these receptor proteins. The cutpoint analysis did not seem reasonable among N- patients, but among N+ patients, on the other hand, it seemed promising. ER analysis seemed to be useful for premenopausal patients and PgR analysis for postmenopausal patients.

The score and fraction analyses gave quite similar results, which is to be expected because the same biological variable (ER or PgR staining) was measured and the score was calculated on the basis of the fraction. The score takes the staining intensity into consideration and in this way, and by multiplying by 2 and 3, gives more weight to higher staining intensities. As a whole, the authors' impression was that the multiplying by 2 and 3, gives more weight to higher staining intensity into consideration and in this way, and by calculating the ER scores in this group of patients did not confirm this, but gave nearly significant cutpoints at scores 0.75 and 0.83.

The purpose of our study was to investigate steroid hormone receptors in respect to other prognostic factors and also to investigate potential cutpoints for ER and PgR immunohistochemistry using different methods of analysis. We believe that we have shown, in our very theoretic assessment, that the currently prevalent cut-off value of 20% seems to be quite reasonable for dividing patients into ER-positive and -negative groups. However, there may be patient subgroups among whom a better cut-off would be other than 20%. The concern of the clinician is that no patient, who would benefit from endocrine treatment, should be left without it. In this report, it does not seem relevant to consider higher cutpoints than these presently used. The limitation of our study is that some patients were actually given tamoxifen. On the other hand, the best prognostic cutpoint may not be the best predictive cutpoint. Our study demonstrates the importance of studying the question of optimal cutpoint for steroid hormone receptors separately in respect to disease-free survival (predictive cutpoint) and survival (prognostic cutpoint).

References


Received September 21, 2004
Accepted January 11, 2005