Immunohistochemical HER2 Status Evaluation in Breast Cancer Pathology Samples: A Multicenter, Parallel-Design Concordance Study

Tülay Canda1, Ekrem Yavuz2, Necmettin Özdemir3, Sennur İlvan4, Serpil Sak Dizbay5, Merih Güray Durak1, Sıtkı Tuzlalı2, Osman Zekioğlu3, Atakan Demir4, Handan Onur5, Kasım Üstündağ6, Burçe Göktaş6

1Department of Pathology, Dokuz Eylül University School of Medicine, İzmir, Turkey
2Department of Pathology, İstanbul University, İstanbul School of Medicine, İstanbul, Turkey
3Department of Pathology, Ege University School of Medicine, İzmir, Turkey
4Department of Pathology, İstanbul University, Cerrahpaşa School of Medicine, İstanbul, Turkey
5Department of Pathology, Ankara University School of Medicine, Ankara, Turkey
6Clinical Research, Roche Preparations San. Inc., İstanbul, Turkey

ABSTRACT
Objective: As patients with increased human epidermal growth factor receptor (HER2) overexpression are more likely to benefit from trastuzumab treatment, the accuracy of HER2 receptor status in breast cancer patients is significant for appropriate disease management. However, this assessment is not harmonized and results may be highly variable between centers. The aim of this study was to investigate the degree of interlaboratory variability in the results of HER2 expression reported by 5 participating centers and to assess the concordance between these centers and a reference laboratory.

Materials and Methods: A total of 30 breast cancer samples were tested and scored for HER2 expression using immunohistochemical method in 5 centers from Turkey and in a reference laboratory from Netherlands (Academic Medical Center, Amsterdam). All the participating centers had an experience of more than 10 years regarding the HER2 testing. The results were compared both among the centers and with the reference laboratory.

Results: When the concordance of participating centers and the reference laboratory was evaluated regarding negative (0-1+), equivocal 2(+) and positive 3(+) classification of HER2 immunostaining, the highest concordance was found in Center-A, and the lowest in Center-C (Kendall’s tau-b concordance coefficient 0.911 and 0.724, respectively). The concordance of the centers with reference laboratory was 80.0% both in equivocal and positive samples, while it increased up to 91.8% in negative samples.

Conclusions: This study showed that in general there is sufficiently good agreement between the reference laboratory and the participating centers for immunohistochemical HER2 assessment.

Keywords: Immunohistochemistry, breast cancer, diagnosis, HER2

Introduction
Breast cancer, which constitutes about 25% of all cancers in women, is the most frequent cancer type in women worldwide, after skin cancer (1, 2). Epidermal growth factor receptor (ErbB) tyrosine kinase receptors (Type I tyrosine kinase receptors) comprise the most extensively studied growth factor receptor system, with the highest amount of information in breast cancer. The family of receptors in this group is made up of four homologous receptors: epidermal growth factor receptor (ErbB1/EGFR/HER1), HER2 (HER2/neu), ErbB3 (HER3) and ErbB4 (HER4) (3-5).

The studies have shown that the neu oncogene is an important mediator of cell proliferation and differentiation (6). This gene is localized on chromosome 17. HER2 positivity is mostly encountered in high grade breast cancers with high proliferation ratio, which demonstrate ER, PR negativity and lymph node positivity (6, 7). Amplification or overexpression of Her-2/neu in breast carcinoma is associated with poor prognosis, short disease-free interval, and short survival time in both node negative and -positive patients (8).

Moreover, patients with HER2 gene amplification or protein overexpression are more likely to benefit from single or combinational trastuzumab treatment (9). These developments show that, correct assessment of HER2 receptor status in cancer cells has a critical role in determining patients who are appropriate for trastuzumab treatment.

Both molecular and immunohistochemical methods are used to demonstrate HER2 status. In routine clinical practice, Her2/neu is evaluated by two methods. Protein receptors produced by oncogenes are assessed by immunohistochemistry (IHC) while gene amplification is evaluated by in situ hybridization (ISH) method.
Although IHC is a cheap and easy method to assess the HER2 status there may be discordance of HER2 scores between laboratories. This discordance may be related to various factors including choice of commercially available primary antibodies, duration of tissue fixation, level of experience about interpretation of the HER2 immunostaining etc.

The aim of this study is to assess concordance rates between 5 participating centers from Turkey and the reference laboratory about immunohistochemical scores of HER2 in 30 cases.

Materials and Methods

The study was planned as an epidemiologic, non-interventional study. Ethics committee approval was received for this study from the ethics committee of Dokuz Eylül University School of Medicine. Written informed consent was obtained from patients who participated in this study.

A total of 5 centers from Turkey have participated in the study. In these centers, 400-700 immunohistochemical breast cancer evaluations are performed annually. The experience of HER2 assessing pathologists involved in this study on breast cancer is more than 10 years. Each center prepared six tumor blocks (one tumor block per patient). Thus, 30 samples were evaluated. Additionally, a laboratory from Netherlands (Academic Medical Center, Amsterdam) has evaluated all the samples, as the reference center. Centers were labeled with letters A to E.

Six tumor blocks obtained from each center contained one “IHC 0”, two “IHC 1+”, two “IHC 2+”, and one “IHC 3+” stained samples.

The inclusion criteria of the study were: women of age ranging between 18 and 75 years; samples of primary lesions (except lymph nodes); samples fixed in 10% neutral-buffered formalin and embedded in paraffin.

Blocks were labeled with letters assigned to the centers (A-E), and were sent to the coordinator central laboratory (CCL). In CCL, 11 unstained sections were prepared from each block. IHC 3+ control sections (ISH confirmed) were placed on the slides as well. Along with the center letters, case numbers [1-6] were written on the slides (A-1, B-4, B-6 etc.).

CCL has sent 2 unstained sections of each block to the institutions (A-E) and 3 unstained sections to the reference laboratory for IHC testing. Thus, each center was able to apply the tests on the sections prepared from the same blocks and to report the results. Each institution has stained the sections with their routine technique, using the preferred antibody and kits, and have recorded the antibody and HER2 IHC kit used in the process. The immunohistochemical HER2 assessment was performed according to the ASCO/CAP guidelines (Table 1) (10).

All centers have sent data entry forms to the CCL (Figure 1).

Statistical analysis

All the samples that complied with the protocol have been included in the statistical analysis. The concordance of the centers was evaluated by calculating Kendall’s tau-b coefficient. Values of Tau-b have ranged from –1 (100% negative association) to +1 (100% positive association, or perfect agreement). A value of zero has indicated the absence of association. Statistical analyzes were performed using Statistical Package for the Social Sciences (V21.0) software.

Results

A total of 30 samples were included in the analysis.

Twenty-eight (93.3%) of the samples included in the study were invasive ductal carcinoma, two were mixed (invasive ductal+invasive lobular) carcinoma. According to the Modified Bloom-Richardson Grade system, 17 (56.7%) of the cases were evaluated as grade 2, 13 (43.3%) as grade 3. Axillary lymph node metastasis was found in 42.8% of the cases.

<table>
<thead>
<tr>
<th>Table 1. Immunohistochemical HER2 assessment (13)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Score 0 (Negative)</strong></td>
</tr>
<tr>
<td><strong>Score 1 (+) (Negative)</strong></td>
</tr>
<tr>
<td><strong>Score 2 (++) (Equivocal)</strong></td>
</tr>
<tr>
<td><strong>Score 3 (+++) (Positive)</strong></td>
</tr>
</tbody>
</table>

HER2: human epidermal growth factor receptor

Canda et al. HER2 Status Concordance Study

**Figure 1. Study design flow chart**
All centers used 10% buffered formalin for fixation and duration of fixation ranged from 24 to 72 hours. All the immunohistochemical staining process, including deparaffinization and antigen retrieval, was performed with a fully automatic immunohistochemical staining device (Ventana BenchMark XT, Ventana Medical Systems, Tucson, AZ). Regarding incubation period, average was 32 minutes. The primary antibody clone for HER2 was SP3 and Ultraview universal dab detection kit was used in all the participating centers.

Assessing the concordance of 5 centers with the reference center as 0, 1 positive, 2 positive and 3 positive according to the IHC results of 30 samples, the highest concordance was found in Center A, and the lowest in Center C (Kendall’s tau-b concordance coefficient 0.862 and 0.706, respectively) (Table 2).

The results were similar when concordance was assessed according to negative (0-1+), equivocal (2+), positive (3+) classification (Kendall’s tau-b concordance coefficient was 0.911 for Center A; 0.724 for Center C) (Table 3).

Analyzing the concordance rate according to the immunohistochemical HER2 positivity rate, the average concordance of the centers with the reference center was found to be higher in 2+ and 3+ results, both 80.0%, and lower in 0 and 1+ results (61.5% and 60.0%, respectively). On the other hand, when assessed according to the negative (0-1+), equivocal (2+) and positive (3+) classification, average concordance rates were found to be 80.0% in equivocal samples, 91.8% in negative samples, and 80.0% in positive samples.

Analyzing the concordance between 5 institutions, the highest concordance was found between centers A and B (Kendall’s tau-b coefficient 0.764), and the lowest concordance was found between centers B and D (Kendall’s tau-b coefficient 0.567).
Analyzing the concordance between the reference center and the study centers based on specimens with the negative (0-1+), equivocal (2+), positive (3+) distribution, for 18 specimens all centers showed concordance, for 8 specimens 4 centers, for 3 specimens 3 centers, and for one specimen none of the centers’ results showed concordance with the reference center.

Discussion and Conclusion

In this study which the same pathology slides were simultaneously assessed in different centers, an average of 69.3% concordance rate was found between study centers and the reference center in determining immunohistochemical staining of HER2. This rate was found to be 60.0% for 1+ samples, and 81.0% for 3+ samples. When analyzed according to negative (0-1+), equivocal (2+), positive (3+) classification, the average concordance rates naturally increased up to 89.6%.

As the average concordance of the study centers were found to be higher for 2+ and 3+ results, and lower for 0 and 1+ results, suggests that making the right decision gets easier as the protein overexpression increases. This result may also be related to tendency of interpreting pathologists to focus on clinically important scores since the differentiation of scores 0 and 1 from each other has no importance. Yet when average concordance rates are evaluated according to negative (0-1+), equivocal (2+), positive (3+) classification, protein overexpression being none or little (concordance 91.8%) facilitates making decisions, whereas for equivocal (2+) and positive (3+) cases (both concordance 80.0%) the decision is harder and therefore the concordance decreases.

Many previous studies dealt with concordance of HER2 analysis with immunohistochemistry [12-15]. In the GEFPICS study, which was a multicenter French study conducted in 2006, the authors reported poor agreement in the score 2+ group (kappa=0.38) and excellent agreement for the 0/1+ (kappa=0.85) and 3+ (kappa=0.82) groups (11). On the other hand, Thomson’s study showed that the interobserver agreement for staining intensity for each antibody was good for 0+ and 3+ groups but poor for 1+ and 2+ groups (12). In accordance with GEFPICS study (12), two studies analyzing the concordance between different centers have shown that samples showing 100% concordance are positive or negative samples, and that equivocal (2+) samples were not fully concordant (13, 14). In an inter-laboratory concordance study conducted in 2007, concordance between the laboratories was identified in terms of immunohistochemical scoring as a result of the assessment by IHC of 20 samples in five centers. Of the 20 specimens, four were scored negative (0/1+) and five positive (3+) in all centers while eight specimens were found negative or questionable (2+) and three were found positive or questionable (13). In an inter-laboratory concordance study conducted in 2007, concordance between the laboratories was identified in terms of immunohistochemical scoring as a result of the assessment by IHC of 20 samples in five centers. Of the 20 specimens, four were scored negative (0/1+) and five positive (3+) in all centers while eight specimens were found negative or questionable (2+) and three were found positive or questionable (13). Additionally, in the NSABP B–31 trial, Paik and colleagues found aHER-2 discordance rate of 18% between local small-volume laboratories and a reference laboratory, proposed the idea of a central testing facility for HER-2 (15). Another study which investigated the role of digital

### Table 3. HER2 positive assessment results of the reference center and the other 5 centers

<table>
<thead>
<tr>
<th></th>
<th>Negative</th>
<th>Equivocal</th>
<th>Positive</th>
<th>Total</th>
<th>Kendall’s tau-b*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Center A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>16</td>
<td>94.1%</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
</tr>
<tr>
<td>Equivocal</td>
<td>1</td>
<td>5.9%</td>
<td>5</td>
<td>83.3%</td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>16.7%</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>100.0%</td>
<td>6</td>
<td>100.0%</td>
<td>7</td>
</tr>
<tr>
<td>Center B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>17</td>
<td>100.0%</td>
<td>2</td>
<td>33.3%</td>
<td>1</td>
</tr>
<tr>
<td>Equivocal</td>
<td>0</td>
<td>0.0%</td>
<td>4</td>
<td>66.7%</td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>100.0%</td>
<td>6</td>
<td>100.0%</td>
<td>7</td>
</tr>
<tr>
<td>Center C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>14</td>
<td>82.4%</td>
<td>1</td>
<td>16.7%</td>
<td>1</td>
</tr>
<tr>
<td>Equivocal</td>
<td>3</td>
<td>17.6%</td>
<td>4</td>
<td>66.7%</td>
<td>0</td>
</tr>
<tr>
<td>Positive</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>16.7%</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>100.0%</td>
<td>6</td>
<td>100.0%</td>
<td>7</td>
</tr>
<tr>
<td>Center D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>15</td>
<td>88.2%</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
</tr>
<tr>
<td>Equivocal</td>
<td>2</td>
<td>11.8%</td>
<td>6</td>
<td>100.0%</td>
<td>2</td>
</tr>
<tr>
<td>Positive</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>100.0%</td>
<td>6</td>
<td>100.0%</td>
<td>7</td>
</tr>
<tr>
<td>Center E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>16</td>
<td>94.1%</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
</tr>
<tr>
<td>Equivocal</td>
<td>1</td>
<td>5.9%</td>
<td>5</td>
<td>83.3%</td>
<td>0</td>
</tr>
<tr>
<td>Positive</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>16.7%</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>100.0%</td>
<td>6</td>
<td>100.0%</td>
<td>7</td>
</tr>
</tbody>
</table>

*For all p=0.000; HER2: human epidermal growth factor receptor
microscopy and computer-aided reading to diminish the interobserver difference in immunohistochemical HER-2 analysis, showed that the use of computer aided reading mode significantly improved the interobserver and intra-observer agreement in evaluation of preselected image fields (16).

Use of routine ISH method, the gold standard for HER2 testing, to assess HER2 may be suggested. However, ISH method is expensive and not widely available. For the immunohistochemistry is a cheap and widely performed method, it has to be refined for HER2 testing in breast cancer cases. GEPFICS study indicated that interobserver reproducibility can greatly be improved with adherence to national guidelines and incorporating a quality assurance process (11).

In the current study, the concordance rates regarding HER2 scoring using immunohistochemistry between the participating centers and the reference center are better than above-mentioned studies. This may be related to well-experience level of the pathologists in the study. The absence of ISH method as control of immunohistochemistry in this study may be regarded as a weakness. However, the participating centers had previously performed another study comparing ISH results on their cases which were negative by immunohistochemistry (either score 0 or 1) and it is another reason of getting experience in refinement of immunohistochemical testing of HER2 (17).

One of the reasons for positive samples to be low in concordance in our study is that, specimen A5 assessed as positive in the reference center, was not found to be positive in any of the study centers. Thus, the average positive concordance rate decreased. This specimen, that showed full discordance, was reported to have a fixation artefact by 3 study centers (Center B over fixation; Center C scattering; Center D tissue folding). The result of the reference center was positive for this specimen, whereas two of the study centers reported equivocal, and three centers negative results.

In our study, out of 30 specimens 12 were assessed as negative in all centers, 4 as positive, 2 as equivocal, 8 as equivocal or negative, and 4 as equivocal or positive.

In conclusion, our study has shown that there is good/remarkably good agreement (Kendall's tau-b coefficient 0.724 - 0.911) between the reference center and the study centers for immunohistochemical HER2 assessment.

For an efficient diagnostic evaluation, two factors appear to be of critical importance: 1) that all pathologists working in clinical pathology laboratories have a regular continuous professional education, and 2) all laboratories have a quality control program (either by participating in national or international programs or by defining and adhering to their own quality standards). Developing quality indicators for all steps in the testing process, and to establish related quality specifications, may enable clinical laboratories to compare, monitor and improve their performances in the every-day practice.

**Ethics Committee Approval:** Ethics committee approval was received for this study from the ethics committee of Dokuz Eylul University School of Medicine.

**Informed Consent:** Written informed consent was obtained from patients who participated in this study.

**Peer-review:** Externally peer-reviewed.

**Conflict of Interest:** The authors have no conflicts of interest to declare.

**Financial Disclosure:** The author declared that this study has received no financial support.

**References**

compared with fluorescent in situ hybridization. Mod Pathol 2001; 14: 1079-1086. (PMID: 11706067) [CrossRef]


