Using a Higher Cutoff for the Percentage of HER2+ Cells Decreases Interobserver Variability in the Interpretation of HER2 Immunohistochemical Analysis

Omar Hameed, MBChB, Amy L. Adams, MD, Allyson C. Baker, MD, Nicole E. Balmer, MD, Walter C. Bell, MD, Holly N. Burford, MD, David C. Chhieng, MD, MBA, MSHI, Nirag C. Jhala, MD, Michael J. Klein, MD, and Thomas Winokur, MD

Key Words: Breast carcinoma; HER2; Immunohistochemistry; Interobserver variability; Cutoff value

DOI: 10.1309/KETM3HANLUQHW7YV

Abstract

The effect of using a 30% cutoff for the proportion of HER2+ cells on the interobserver variability in the interpretation of HER2 immunohistochemical results was evaluated. Immunostained sections from 96 cases of breast carcinoma were reviewed by 10 pathologists and scored as positive (3+) when uniform strong membranous staining was identified in at least 10% of tumor cells; the actual percentage of cells with such staining was also estimated. The agreement rates and the \( \kappa \) values using a 30% cutoff were compared with those using a 10% cutoff. These proved to be higher in 62% and 66% of measurements, respectively, with average interobserver rates and \( \kappa \) values of 72% and 0.54 using the 30% cutoff and 70% and 0.49 using the 10% cutoff (\( P = .001 \) for all comparisons). Using a 30% cutoff for the percentage of HER2+ cells by immunohistochemical analysis modestly decreased interobserver variability in the interpretation of HER2 immunohistochemical results.

Analysis of HER2 status in breast carcinoma specimens has become routine practice in surgical pathology laboratories throughout the world mainly because of its value as a predictive marker for the response of breast cancer to adjuvant chemotherapeutic regimens, including to trastuzumab.\(^1\)\(^2\) Although there are several ways to evaluate HER2 status, assessment of protein overexpression by immunohistochemical analysis is the most widely used method.\(^5\)

In an effort to improve the accuracy and consistency of HER2 testing, an expert panel from the College of American Pathologists (CAP) and the American Society of Clinical Oncology (ASCO) recently recommended that at least 30% (vs 10%) of tumor cells need to show uniform intense membranous staining for an HER2 immunohistochemical result to be considered positive (3+).\(^5\) We have previously shown that using this higher cutoff increases the specificity of HER2 immunohistochemical analysis and its concordance with results of HER2 amplification by fluorescent in situ hybridization (FISH).\(^6\) This study was performed to examine the effect of using this 30% cutoff on interobserver variability in the interpretation of HER2 immunohistochemical results.

Materials and Methods

A series of 96 retrospectively identified and previously studied\(^6\) cases of breast carcinoma formed the basis of this study. The H&E-stained slides and the HER2 immunostained sections (clone CB11, Ventana Medical Systems, Tucson, AZ, prediluted; or polyclonal A0485, dilution 1/400 dilution; DAKO, Carpinteria, CA) of these cases were circulated among the 10 study coauthors who have variable experience
in breast pathology. The participants were asked to interpret the HER2 immunostains as positive (3+) for overexpression after reviewing a PowerPoint presentation (Microsoft, Redmond, WA) prepared by one of us (O.H.) that included images of representative cases with different levels of expression and the qualitative criteria for a positive (3+) result based on the immunohistochemical interpretation criteria published by the CAP/ASCO panel.\(^2\) Participants were instructed to score cases as positive when sufficiently strong membranous staining was identified in at least 10% of tumor cells and also to estimate the actual percentage of cells with such staining.

The latter data were used by one of us (O.H.) to derive a second set of results whereby only cases with strong membranous staining in more than 30% of tumor cells were classified as positive. To determine consistency among pairs of observers, each participant’s interpretation of the HER2 immunostains was compared with that of the other participants in a pairwise manner to generate 2 sets of interobserver agreement rates and \(\kappa\) values—a set using the 10% cutoff, and another using the 30% cutoff. The paired sample \(t\) test was then used to compare the pairwise agreement rates and the pairwise \(\kappa\) values between the 2 sets of data. A \(P\) value of .05 was considered statistically significant. Calculation of the \(\kappa\) values and the statistical analysis was performed by using SPSS for Windows statistical program (SPSS, Chicago, IL).

**Results**

Comparison of the HER2 interpretation results among the different observers in a pairwise manner generated 45 interobserver agreement rates using each cutoff. These rates ranged from 47% to 82% when the 10% cutoff was used (mean, 70%) and from 54% to 84% when the 30% cutoff was used (mean, 72%). The agreement rates obtained using the 30% cutoff were higher than, lower than, or similar to those obtained using the 10% cutoff in 28 (62%), 10 (22%), and 7 (16%) paired observations, respectively Table 1. The difference in the agreement rates between paired observations using the 2 cutoffs was statistically significant (\(P = .001\)).

Similarly, comparison of the HER2 interpretation results among the different observers also generated 45 \(\kappa\) values using each cutoff. These values ranged from 0.18 to 0.70 when the 10% cutoff was used (mean, 0.49) and from 0.24 to 0.77 when the 30% cutoff was used (mean = 0.54). The \(\kappa\) values obtained using the 30% cutoff were higher than, lower than, or similar to those obtained using the 10% cutoff in 30 (67%), 8 (18%), and 7 (16%) paired observations, respectively Table 2. The difference in the \(\kappa\) values between paired observations using the 2 cutoffs was statistically significant (\(P = .001\)).

**Discussion**

Original US Food and Drug Administration– and manufacturer-proposed criteria for an HER2 immunohistochemical test to be considered positive required the presence of moderate (2+) or strong (3+) membranous staining in 10% of tumor cells. The weaker response to trastuzumab (or lack thereof) in cases scored as 2+ compared with 3+ cases,\(^3\) as well as low rates of concordance with gene amplification, has led to reclassification of 2+ cases as equivocal for HER2 overexpression with only 3+ cases considered positive.\(^7,8\) More recently, the ASCO/CAP expert panel stated that this scoring system “provide[s] insufficient specificity”\(^9\) and has recommended that at least 30% (vs 10%) of tumor cells need to show uniform intense membranous staining for an HER2 immunohistochemical result to be considered positive.\(^1\) After finding that using this higher cutoff increases the specificity of HER2 immunohistochemical analysis and its concordance with results of HER2 amplification by FISH,\(^6\) we wanted to evaluate its effect, if any, on the interobserver

**Table 1**

Comparison of Pairwise Agreement Rates Between Pathologists (A-J) in the Evaluation of HER2 Immunohistochemical Results Using 10%* and 30%† Cutoffs for the Proportion of Cells With Uniform Strong Membranous Staining\(^2\)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>82</td>
<td>69</td>
<td>79</td>
<td>69</td>
<td>75</td>
<td>80</td>
<td>78</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>79</td>
<td>64</td>
<td>47</td>
<td>73</td>
<td>72</td>
<td>81</td>
<td>84</td>
<td>79</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>68</td>
<td>62</td>
<td>73</td>
<td>73</td>
<td>54</td>
<td>62</td>
<td>59</td>
<td>62</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>79</td>
<td>75</td>
<td>66</td>
<td>73</td>
<td>65</td>
<td>70</td>
<td>68</td>
<td>72</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>65</td>
<td>61</td>
<td>78</td>
<td>69</td>
<td>60</td>
<td>62</td>
<td>66</td>
<td>66</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>69</td>
<td>73</td>
<td>47</td>
<td>65</td>
<td>48</td>
<td>72</td>
<td>82</td>
<td>78</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>71</td>
<td>73</td>
<td>66</td>
<td>69</td>
<td>69</td>
<td>69</td>
<td>81</td>
<td>83</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>72</td>
<td>80</td>
<td>52</td>
<td>68</td>
<td>68</td>
<td>82</td>
<td>72</td>
<td>83</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>71</td>
<td>77</td>
<td>57</td>
<td>71</td>
<td>71</td>
<td>77</td>
<td>73</td>
<td>82</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>78</td>
<td>76</td>
<td>77</td>
<td>72</td>
<td>72</td>
<td>60</td>
<td>70</td>
<td>63</td>
<td>68</td>
<td></td>
</tr>
</tbody>
</table>

* Lower left portion of table.
† Upper right portion of table.
\(^3\) Agreement rates are given as percentages. Bold type indicates an increased agreement rate for 30% vs 10% cutoffs.
variability in the evaluation of HER2 immunohistochemical results, especially because, to our knowledge, this has not been previously investigated.

We have indeed shown that using a 30% instead of a 10% cutoff for the proportion of cells with strong, uniform membranous staining for a result to be considered positive resulted in a modest but statistically significant increase in the agreement rates and κ values between pathologists with variable experience in breast pathology. These findings are consistent with those of Hsu et al, who also found that using a higher cutoff (>50% of cells in their study) for “strongly positive” cases led to better interobserver agreement in distinguishing between 2+ and 3+ cases and increased the generalized κ value from 0.38 to 0.78.

One issue this study did not address, which to the best of our knowledge has not been well addressed in the literature, is the possible effect of pathologist experience on the interobserver variability in the interpretation of HER2 immunohistochemical results. Believing that this is a separate (and probably more significant) question deserving of special consideration, we are currently analyzing it in a separate study along with the effect of pathologist experience on the interobserver variability in histologic grading of breast carcinoma.

Based on the findings in this study, one can conclude that in addition to improving concordance with FISH, using the newly proposed 30% cutoff for the proportion of cells with strong, uniform membranous staining, as suggested by the ASCO/CAP guideline recommendations, leads to a modest decrease in the interobserver variability in interpretation of HER2 immunohistochemical results. Nevertheless, continuous testing and additional validation of this, and other ASCO/CAP recommendations, is needed to further improve the accuracy and consistency of HER2 testing in breast carcinoma.

Presented in part at the 97th Annual Meeting of the United States and Canadian Academy of Pathology; March 2008; Denver, CO.

Address reprint requests to Dr Hameed: Division of Anatomic Pathology, University of Alabama at Birmingham, 619 19th St South, NP 3550, Birmingham, AL 35294-6823.

References


