

Interobserver variability in HER-2/neu reporting on immunohistochemistry in breast carcinomas

Nadira Mamoon,¹ Fareeha Naseer Syed,² Sajid Mushtaq,³ Humaira Nasir,⁴ Imran Nazir Ahmad⁵

Abstract

Objective: To evaluate human epidermal growth factor receptor 2 (HER2/neu) interobserver variability between specially trained and untrained general histopathologists.

Methods: The retrospective study cases of invasive breast carcinoma received at Shifa International Hospital, Islamabad, from June 2010 to December 2011, for assessment of HER2/neu status by immunohistochemistry which were retrieved from the files and, 30 consecutive cases each of score 0, 1+, 2+ and 3+ were selected for a total of 120 cases. Two groups of two histopathologists each examined the cases blindly. One group had attended a short course in Germany, while the other group comprised two qualified histopathologists who had not had any special training. Each histopathologist reported the cases independently according to standard guidelines. Kappa statistics were applied.

Results: The trained group of histopathologists showed agreement in 113 (94%) cases. Kappa value was calculated to be 0.96 which means 'perfect agreement'. In contrast the untrained group showed agreement in 83 (69%) cases with a kappa value of 0.59 which means 'moderate agreement'.

Conclusion: Interobserver variability in immunohistochemical scoring of HER-2/neu in breast carcinoma is high among untrained general histopathologists. This may adversely affect the selection of patients with cancers who could benefit from Herceptin therapy.

Keywords: Her2/neu, Interobserver variability, Breast carcinoma, Immunohistochemistry. (JPMA 64: 151; 2014)

Introduction

There are a number of methods available to assess human epidermal growth factor receptor 2 (HER2/neu) status. However, immunohistochemical analysis is the most popular and can be performed easily in every reasonably good pathology laboratory. Fluorescent in situ hybridisation (FISH) is more expensive and only recommended to be done in cases scored 2+.¹ HER2/neu interpretation by immunohistochemistry is difficult and, keeping in mind its role in treatment, it is of utmost importance that interobserver variation should be minimal.² The criteria for interpretation are laid down quite simply and are apparently easy to follow.³ However, some special short courses are available to train histopathologists in this aspect.

This study was undertaken to evaluate the HER2/neu interobserver variability between specially trained (those who had done a course) and untrained general histopathologists.

Materials and Methods

The correlational study based on non-probability

.....
^{1,2,4,5}Shifa International Hospital, Islamabad, ³Shaukat Khanum Memorial Cancer Hospital, Lahore, Pakistan.

Correspondence: Nadira Mamoon. Email: nmamoon@live.com

purposive sampling had the approval of the institutional ethics committee. Cases of invasive breast carcinoma received at the Shifa International Hospital, Islamabad, for the assessment of HER2/neu status by immunohistochemistry were retrieved from the files related to the preceding two years starting from June 2010 to December 2011. Out of those, 30 consecutive cases each of score 0, 1+, 2+ and 3+ were selected according to the inclusion criteria, making a total of 120 cases for the study. Slides of these cases were retrieved from the archives.

There were two groups of histopathologists who examined these 120 cases blindly without any accompanying details or knowledge of originally reported score. Group 1 comprised two histopathologists who had undergone a course of 4-day duration titled 'HER2/neu reporting by immunohistochemistry and FISH' at separate times arranged by Targos in Kassel, Germany. The other group (Group 2) comprised two general histopathologists who had not had any such training. Each histopathologist reported the cases independently according to the American Society of Clinical Oncology and College of American Pathologists (ASCO/CAP) recommended guidelines as follows:

Score 0 - No staining; Score 1+: Faint or barely perceptible

incomplete membrane staining in any percentage of tumour cells; Score 2+: Strong, complete, homogenous membrane staining in less than or equal to 30% of tumour cells OR weak to moderate, heterogenous complete membrane staining in at least 10% of tumour cells; Score 3+: Strong, homogenous, complete membrane staining in more than 30% of tumour cells.

All tissue samples/blocks of invasive breast carcinoma received for assessment of HER2/neu status by immunohistochemistry were included in the study. Cell blocks made by fine needle aspiration (FNA) biopsies, poorly fixed specimens, tissue with scanty tumour cells and tissues showing autolysis or complete necrosis with no appreciable histological details were excluded from the study.

The cases were assigned numbers from 1 to 120 randomly. These cases were shown blindly, individually and sequentially to each of the four consultant histopathologists, belonging to either of the two groups, without any accompanying clinical details for independent HER2/neu scoring. Every histopathologist scored these cases blindly without knowing the previously reported score. The score awarded by each consultant was recorded on a proforma and analysed to determine interobserver variability of each group.

The data was entered in SPSS version 10.0. Frequency and percentage were calculated for qualitative variables like gender. Mean and standard deviation were calculated for quantitative variables like age and scoring done by each observer for all cases. The interobserver variability was evaluated by kappa statistics for both groups separately.

Cohen's kappa statistic measures the agreement between raters for classifying a number of data points into 2 or more categories. Kappa ranges can be used to establish agreement between tests, methods, or evaluators. Kappa values range from -1 to 1, with 1 representing perfect agreement between data sets.

Greater kappa values reflect stronger agreement between the raters.

A range from 0.81 to 1.00 implies almost perfect agreement; from 0.61 to 0.80 means substantial agreement; 0.41 to 0.60 implies moderate agreement; 0.21 to 0.40, means fair agreement; and, 0.00 to 0.20 means slight agreement.

Results

All the 120 patients were female (100%). The mean age of patients was 48.93 ± 13.02 years ranging from 26 to 92 years.

Table-1: Comparison of scoring by Group 1 histopathologists: Highlighted figures show cases with complete agreement.

	HER-2/neu score	Observer 2			Total	
		0	1+	2+		3+
Observer 1	0	54	2		56	
	1+		9		9	
	2+	1	1	26	3	31
	3+				24	24
	Total	55	12	26	27	113/120

Table-2: Comparison of scoring by Group 2 histopathologists. Highlighted figures represent cases with complete agreement.

	HER-2/neu score	Observer 4			Total	
		0	1+	2+		3+
Observer 3	0	19	14	4		37
	1+	6	17	7		30
	2+			20	3	23
	3+			3	27	30
	Total	25	31	34	30	83/120

Table-3: Paired kappa of four observers.

Level of agreement	Observer 1 vs Observer 3	Observer 1 vs Observer 4	Observer 2 vs Observer 3	Observer 2 vs Observer 4
Kappa	0.56	0.48	0.52	0.45

In group 1, 54 (45%) cases were scored by both as 0, 9 cases were scored 1+, 26 (21.6%) cases were scored 2+ and 24 (20%) cases were scored 3+. Complete agreement of the scoring results between the two pathologists was found in 113 (94.2%) of the 120 cases. Discrepancy was found in only 7 (5.85) cases, of which 3 cases were scored 2+ by observer 1 and 3+ by observer 2. Similarly 2 cases were scored 2+ by observer 1 and 0 and 1+ by observer 2. (Table-1).

In group 2, 19 (15.8%) cases were scored 0 by both observers, 17 (14.1%) cases were scored 1+, 20 (16.6%) cases were scored 2+ and 27 (22.5%) cases were scored 3+. Complete agreement was in 83 (69.2%) cases out of 120 while 37 (30.8%) showed discrepancy. There were 14 cases scored 0 by observer 3 and 1+ by observer 4, four cases were scored 0 by observer 3 and 2+ by observer 4, seven cases were scored 2+ by observer 4 and 1+ by observer 3, six cases were scored 0 by observer 3 and 1+ by observer 4, three cases were scored 3+ by observer 3 and 2+ by observer 4 and vice versa in three more cases (Tables-2).

Kappa value was 0.91 for group 1 which means "perfect agreement", whereas it was 0.59 for group 2 which means

"moderate agreement" (Table-3).

Discussion

The HER2/neu gene has proved to be a significant prognostic and predictive biological marker in breast cancer. HER-2/neu overexpression in breast carcinoma is associated with a poor prognosis in terms of lack of response to conventional chemotherapy. Therefore HER-2/neu testing is of paramount importance as treatment decisions are based on it and it is the standard of care.

The most popular, inexpensive, easily available and usually the first method of choice to detect HER-2/neu overexpression is immunohistochemistry (IHC). A lot of interobserver variability has been observed in IHC assessment of HER-2/neu, especially in assigning a 2+ score.⁴

There are a number of reasons for this interobserver variation. These include technical reasons as well as observer-dependent issues. The technical reasons include problems with interpretation of tissue taken from the edge of the lesion, tissue with retraction and crushing artefacts and improper staining technique. Overstaining at the edge of the tissue section can lead to overinterpretation. Loss of tissue antigenicity due to improper fixation can lead to uneven staining and observer variation in interpreting the results. Interobserver variability also reflects histopathologists' experience, including their special interests and how long and where they have practised.

The only factor that was looked into in this study was the efficacy of Continuing Medical Education (CME) as the educational course attended by two of the consultants was the major difference between the two groups. Working hours, working schedules, volume of work per person were almost similar for all the consultants. As the cases were seen by the consultants at their own convenience and there was no time limit, fatigue was not a relevant factor.

Initially the cutoff limit for assigning a positive HER2/neu score was 10%.^{5,6} In 2007 HER2/neu testing algorithm integrating both the IHC and the FISH algorithms were been updated by ASCO/CAP. Accordingly, the cutoff for positivity for IHC result is now defined as >30% strong complete membrane staining.⁷ This has led to improvement in the assessment of HER2/neu scoring on IHC.

There are certain guidelines for interpreting the HER2/neu IHC results⁸ such as scoring the percentage and intensity of only the cells showing complete membrane staining.

Cytoplasmic staining should not be included when

interpreting results. Staining should be assessed in the invasive component, but not in the in situ component. Normal epithelial cells should not stain. If staining is noted, the test should be rejected. Retraction artifacts may be falsely interpreted as positive.

Several studies have been done on the subject of standardisation of IHC protocols, but training of pathologists remains the main issue to be addressed. A study clearly demonstrated interobserver variability in IHC assessment of HER2/neu in breast cancers.⁹ It studied 24 cases of invasive breast carcinoma where HER2/neu was scored by 5 pathologists. Kappa values were high for negative and positive reporting (without a numerical score) i.e; ranging from 0.74 - 0.87. However kappa values for numerical scores (0, 1+, 2+, 3+) were low (0.31, 0.17, 0.43, 0.40, 0.25), indicating poor agreement regarding staining.⁹

The current study is the first to address interobserver variability in HER2/neu scoring between specially-trained and untrained groups of histopathologists. Results of the study showed a perfect level of agreement between the trained histopathologists (kappa; 0.96). However, a great deal of variation existed between untrained histopathologists (kappa: 0.59). Score 0 and 1+ are considered negative for HER2/neu. Therefore, even if variability exists between 0 and 1+ scores, it is clinically insignificant. Most important is the discrepancy which was seen in assigning a 2+ score. There was an overlap in assigning 1+/2+ scores and in 2+/3+ scores. This is the critical step because 2+ (equivocal) cases require confirmation by FISH, whereas, 3+ cases are supposedly definitely positive for HER2/neu and eligible for Herceptin therapy. Nowadays many clinicians also request confirmation of 3+ cases by FISH before embarking on Herceptin therapy for these very reasons.

It is possible that the difference in opinion is due to the difference in interpretation of the intensity of membrane staining. The agreement among pathologists concerning strong intensity was not perfect. The paired kappa ranged from 0.45 to 0.56 among the pathologists as a whole.

It has also been postulated that such interobserver variability may result from the difference in individual perception of light and colour as each human being's eye has a slightly different range of perception of light of different wavelengths and colours. To obviate the human factor, computer-assisted and image analysis systems have been devised^{10,11} with improved accuracy in scoring and better results, but these systems are not readily available everywhere, specially in a developing country like ours.

IHC is popular and easily available. However, variation in staining methods, antibodies used and subjective scoring systems make FISH a more reliable method.^{9,12} Unfortunately, FISH is a more expensive, delicate, sophisticated and technically demanding technique compared to IHC. According to international protocols, FISH is mandatory only in 2+ reports and not in 3+ or 0/1+ (negative) reports. Asking for this test remains the prerogative of the oncologist and also depends on the affordability factor.

Conclusion

Interobserver variability was significant in pathologists without specific training in HER-2/neu scoring even though the recommended criteria for scoring seem to be quite straightforward. Specialised training had a significant effect on minimising interobserver variability. It is important to assess this important tumour marker with the help of a trained histopathologist so as to improve the diagnostic accuracy.

References

1. Nisa A, Bhurgri Y, Raza F, Kayani N. Comparison of ER, PR and HER-2/neu (C-erb B 2) reactivity pattern with histologic grade, tumor size and lymph node status in breast cancer. *Asian Pac J Cancer Prev* 2008; 9: 553-6.
2. Cserni G, Kálmán E, Kulka J, Orosz Z, Udvarhelyi N, Krenács T. Quality control of HER2 immunohistochemistry--results from a Hungarian study. *Magy Onkol* 2007; 51: 23-9.
3. Hanna W, O'Malley F, Barnes P, Berendt R, Gaboury L, Magliocco A et al. Updated recommendations from the Canadian National Consensus Meeting on HER2/neu testing in breast cancer. *Current Oncol* 2007; 14: 149-153.
4. Hsu CY, Ho DM, Yang CF, Lai CR, Yu IT, Chiang H. Interobserver Reproducibility of Her2/neu Protein Overexpression in Invasive Breast Carcinoma Using the DAKO Hercep Test. *Am J Clin Pathol* 2002; 118: 693-8.
5. Fitzgibbons PL, Page DL, Weaver D, Thor AD, Allred DC, Clark GM, et al. Prognostic factors in breast cancer. College of American pathologists consensus statement 1999. *Arch Pathol Lab Med* 2000; 124: 966-78.
6. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/ College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 2007; 25: 1-28.
7. Atkinson R, Mollerup J, Laenkholt A, Verardo M, Hawes D, Commins D, et al. Effects of change in cut off values for Human Epidermal Growth factor Receptor 2 status by Immunohistochemistry and Fluorescent in situ hybridization. *Arch pathol Lab Med* 2011; 135: 1010-16.
8. Bilous M, Dowsett M, Hanna W, Isola J, Lebeau A, Moreno A, et al. Current perspectives on HER2 Testing: A Review of National Testing Guidelines. *Mod Pathol* 2003; 16: 173-82.
9. Nichols DW, Self S, Metcalf JS, Jacobs DD, Hall RK, Cate JC. A Testing Algorithm for Determination of HER2 Status in Patients with Breast Cancer. *Ann Clin Lab Sci* 2002; 32: 3-11.
10. Odze RD, Goldblum J, Noffsinger A, Alsaigh N, Rybicki L, Fogt F. Interobserver variability in the diagnosis of ulcerative colitis associated dysplasia by Telepathology. *Mod Pathol* 2002; 15: 379-86.
11. Thomson T, Hayes M, Spinelli J, Hilland E, Sawrenko C, Phillips D, et al. HER-2/neu in Breast cancer: Interobserver variability and performance of Immunohistochemistry with 4 antibodies compared with Fluorescent in situ hybridization. *Mod Pathol* 2001; 14: 1079-86.
12. Isola J, Tanner M, Forsyth A, Cooke TG, Watters AD, Barlett JMS. Interlaboratory comparison of HER2/neu oncogene amplification as detected by chromogenic and fluorescent in situ hybridization. *Clin Cancer Res* 2004; 10: 4793-8.