ORIGINAL ARTICLE

Interobserver Variability in HER2 Breast Biomarker Reporting: Implications for Diagnostic Consistency and Treatment Precision

Hira Batool, Sameen Afzal, Fatima Khalid, Saira Javeed, Zonaira Rathore, Akhter Sohail Chughtai

ABSTRACT

Objective: To assess the interobserver variability in HER2 immunohistochemical stain interpretation by pathologists of different strata of experience.

Study Design: Cross-sectional observational study.

Place and Duration of Study: Chughtai Institute of Pathology at Lahore, during a three-month interval from 01/6/2024 to 31/8/2024.

Materials and Methods: Fifty cases (n=50) of invasive breast cancer were included by random probability sampling and blocks were retrieved through respective biopsy reports by the laboratory information system. All the cases of ductal carcinoma in situ (DCIS) and salivary gland-like tumors of the breast were excluded. HER2 antibody was applied to these cases and interpreted by four histopathologists with varying years of reporting experience. Interobserver variability was observed by using the Cohen's and Fleiss kappa tests. A *p-value* of ≤0.05 was considered statistically significant.

Results: All patients were female, with a mean age of 46.3 years ± 12.87. The highest concordance was observed at a score of 3, while the greatest discordance occurred at a score of 2, with kappa values of 0.81 and 0.35, respectively, and a p-value of <0.01.

Conclusion: The highest interobserver variability was observed in the assessment of HER2 score 2, highlighting the challenges in interpreting equivocal cases.

Key Words: *ASCO Guidelines, Equivocal Cases, Human Epidermal Growth Factor Receptor 2, Interobserver Variability, Invasive Breast Cancer.*

Introduction

Breast cancer is one of the most prevalent cancers in the world. Its incidence is 12% all over the world.[1] Various risk factors are associated with the occurrence of breast cancer, and they include age, with higher prevalence in women more than 50 years, family history, obesity, and excessive alcohol consumption along with hormone replacement therapies.[2-4] Multifaceted treatment options are available for breast cancer. One of the commonly employed treatment options is targeted therapy given against estrogen receptors (ER), progesterone receptors (PR), and human epidermal growth factor receptor 2 (HER2).[5,6] Various immunohistochemical tests are done on tissue biopsies of breast tumors to detect the expression of these biomarkers.[7]

Approximately 20-25% of breast cancers are positive for HER2 immunohistochemical stain.[8] HER2 is a transmembrane receptor with tyrosine kinase activity and plays a pivotal role in cell signaling, differentiation, and angiogenesis and therefore plays an important role in malignant transformation.[9, 10] Breast cancers can exhibit a 40 to 100-fold increase in HER2 protein.[11] According to 2018 the American Society of Clinical Oncology (ASCO) guidelines, HER2 expression has been categorized as a score of 0, 1, 2, or 3 based on the intensity of membranous expression. Recent 2023 ASCO Guidelines have suggested recategorization of HER2 scores to ultra-low expression at score 0, low HER2 expression at scores 1+ and 2+ In Situ Hybridization (ISH) negative, and high expression at score 3+ due to a recently conducted trial on the trastuzumab response in patients with HER2 Score 1 or 2. It was observed that those patients responded well to anti-HER2 therapy, and the overall survival rate was also increased.[12-13] However, the College of American Pathologists (CAP) and ASCO guidelines 2023 have not yet applied the

Department of Pathology
Chughtai Institute of Pathology, Lahore

Correspondence:
Dr. Hira Batool
Department of Pathology
Chughtai Institute of Pathology, Lahore
E-mail: hirabatoolgmc@gmail.com

Received: October 25, 2024 ; Revised: January 27, 2025
Accepted: February 03, 2025

https://doi.org/10.57234/jiimc.june25.2303

results of this study to the treatment regimen as this study excluded the treatment response of HER 2 score 0 as it is thought that in the future these tumors might show expression of HER2 by sensitive techniques. Due to the higher predictive power of HER2 immunostaining, the present practice of oncologists is to treat patients with trastuzumab who present with metastatic disease or early-stage breast cancer with HER2 score of 3 on immunohistochemistry and HER2 equivocal score, which is further amplified on fluorescence in situ hybridization (FISH) studies.[14,15] Scores 0 and 1 are contemplated as not eligible for treatment. However, these results are subject to variability owing to different aspects, such as preanalytical variables, the types of antibodies used, staining methods, and interpretation by different histopathologists. Due to the difference in ethnicity of our region, we expect different results and therefore, it is important to score this biomarker meticulously for optimal management.[16] It is evident that HER2 biomarker expression is of pivotal importance in the diagnosis and devising treatment regimen, however, the major caveat is the variation in the interrater variability in its scoring. In recent years, the issue has been studied well internationally, however, there is a dearth of research focused on the interrater variability in our region, where differences in laboratory practices, pathologist training, and, adherence to international guidelines can significantly influence diagnostic accuracy and consistency. Therefore, the study is novel and to the best of our knowledge, is the first of its kind in Pakistan. This study was designed to assess the inter-rater variation in the immunohistochemical staining interpretation of HER2 antibody among pathologists and to highlight the significance of accurately scoring this parameter according to internationally recommended guidelines to ensure optimal treatment of breast cancer patients.

Materials and Methods

This cross-sectional observational study was conducted at the Chughtai Institute of Pathology (CIP), Lahore lasting three months from 01/06/2024 to 31/8/2024 after getting IRB approval (IRB#1277/IRB/ CIP). Fifty diagnosed and reported cases of invasive breast cancer booked in the mentioned period were selected by random probability sampling techniques after retrieving respective laboratory reports by the laboratory information system. Cases of ductal carcinoma in situ (DCIS) and salivary gland-like tumors of the breast were excluded. HER2 immunohistochemical stains were employed on slides prepared from paraffin-embedded tumor blocks. The type of antibody applied was polyclonal rabbit anti-human c-erbB-2 oncoprotein using Dako Link 48 automatic stainer. External control was applied to each batch of cases for quality assurance. Four pathologists interpreted the results, of which three were general histopathologists and the fourth was breast subspecialist histopathologist. The three general histopathologists hold different years of reporting experience upto 5 years, are of Pakistani nationality, and have an age range from 35 to 40 years. These three general histopathologists independently scored all fifty slides of HER2 hormone receptor antibody according to 2018 ASCO guidelines which define a score 0 if there is no staining or less than 10% expression on tumor cell membranes, a score of 1 if the weak expression on membranes of more than 10% tumor cells, score 2 if there is moderate but patchy expression on membranes in more than 10% neoplastic cells, or absolute expression on membranes in more than 10% neoplastic cells, score 3 if there is peak, complete staining on membranes in more than 10% tumor cells and individually submitted their results.[9] All were blinded to each other's assessment and designated observers 1, 2, and 3. Then a consensus opinion on the HER2 score of each case was made by all three general histopathologists along with the most experienced fourth breast subspecialist histopathologist allotted as observer 4 having experience of up to 10 years in reporting breast hormone receptors, Pakistani nationality, and 45 years of age. This consensus score was considered a gold standard score, and results were submitted as reported by observer 4. The statistical analysis was done by entering the scores of all 50 cases submitted by all four observers into SPSS version 29. A *p-value* of ≤0.05 was considered statistically significant. The concordance between observers 1,2 and 3 scores was calculated with the consensus score conferred by observer 4 one by one by applying Cohen's kappa test. Further, interrater agreement among all the observers was measured by using the Fleiss multi-rater kappa test which is an

https://doi.org/10.57234/jiimc.june25.2303

extended form of Cohen's kappa. Kappa value was interpreted as no agreement at a score range of 0-0.20, minimal agreement with a score range of 0.21-0.39, weak agreement with a score range of 0.40-0.59, moderate agreement with a score range of 0.60-0.79, strong agreement with a score range of 0.80-0.90 and almost perfect agreement with a score above 0.90.

Results

The mean age of patients in 50 selected cases was 46.3 years, with a standard deviation of 12.87. All patients belonged to the female gender (100%). The most frequent breast cancer in this study was invasive breast carcinoma (IBC) of No Special Type (NST) 47 (94%), and the commonest grade was II 29 (58%) (Table I). According to the categorical scale mentioned in materials and methods, the interpretation of Cohen's kappa value following analyzing interobserver concordances of HER 2 scores (0-3) among observers 1, 2, and 3 was compared to the interpretation of results by observer 4. The results showed weak agreement between observers 1 and 3, with observer 4 having Cohen's kappa values of 0.54 and 0.57, respectively, and moderate agreement between observers 2 and 4 with a Cohen's kappa value of 0.71 (Table II). The Kappa value for the multi-rater Fleiss Kappa analysis was 0.65, which indicated moderate agreement among all observers (Table III). Interrater agreement on HER2 scores 0,1,2 and 3 was also acquired. The results illustrated nearly perfect agreement on score 3, moderate agreement on scores 0 and 1, and weak agreement on an equivocal score of 2 (p value<0.01). (Table IV).

Table I: Histopathological Features of Breast Cancer Cases Included in This Study

Attributes		N (%)
Histologic type	IBC NST	47 (94%)
	Mucinous carcinoma	1 (2%)
	Pleomorphic lobular carcinoma	1 (2%)
	Invasive lobular carcinoma	1 (2%)
Histologic grade	Grade I	2 (4%)
	Grade II	29 (58%)
	Grade III	19 (38%)
Estrogen receptor (ER)	Positive	31 (62%)
	Negative	18 (36%)
	Not applied	1 (2%)
Progesterone receptor (PR)	Positive	30 (60%)
	Negative	19 (38%)
	Not applied	1 (2%)

IBC NST: Invasive Breast Carcinoma, No Special Type

Table II: Interobserver/Interrater Cohen's Kappa Value for The Degree of Concordance Between Observers 1-3 With Observer 4

Inter-observer comparison	Cohen's Kappa value	p-value*
Observer 1 and 4 concordances	0.54	<0.01
Observer 2 and 4 concordances	0.71	<0.01
Observer 3 and 4 concordances	0.57	<0.01

***Statistically significant p-value <0.05**

Table III: Multi-Rater Fleis Kappa Statistics Value for The Overall Degree Of Concordance

Overall comparison	Fleiss Kappa value	p-value*
Overall concordance	0.65	<0.01

***Statistically significant p-value <0.05**

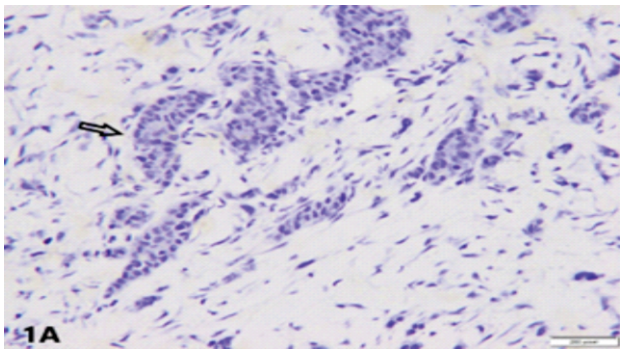
Table IV. Individual HER2 Score Agreement Status by Fleiss Kappa Statistics Value

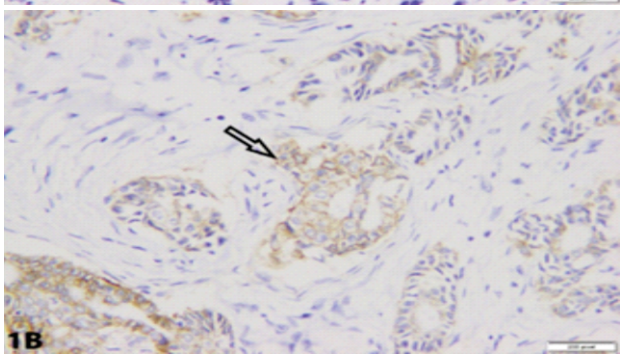
HER2 Score	Fleiss Kappa value	p-value*
0	0.73	<0.01
1	0.65	<0.01
2	0.35	<0.01
3	0.81	<0.01

***Statistically significant p-value <0.05**

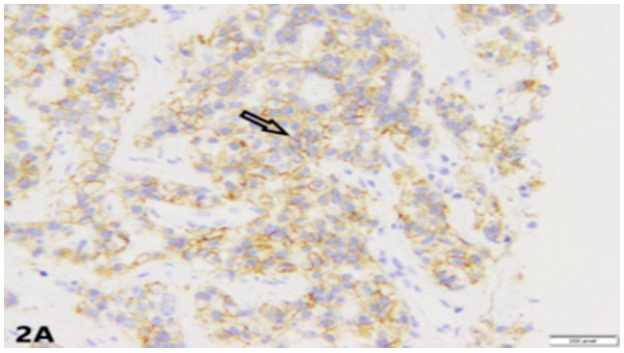
Discussion

The scoring of HER2 was done following the categorization by ASCO guidelines 2018 as negative at scores 0 (Figure 1A) and 1 (Figure 1B), equivocal at score 2 (Figure 2A), and positive at score 3 (Figure 2B).[17,18] Overall agreement by the Fleiss kappa test revealed moderate agreement among all the observers with a kappa value of 0.65 (p value < 0.01). The results showed the greatest conflict at HER2 score 2 (kappa value of 0.35; p value < 0.01) and the highest agreement at HER2 score 3 (kappa value of 0.81; p value < 0.01). Therefore, reporting score 2 cases can be the most challenging and this area needs to be improved as if scored erroneously this could affect the individual management of breast cancer patients.





1B. (400X magnification) manifests weak incomplete membranous staining of HER2 antibody in >10% neoplastic cells, scored 1



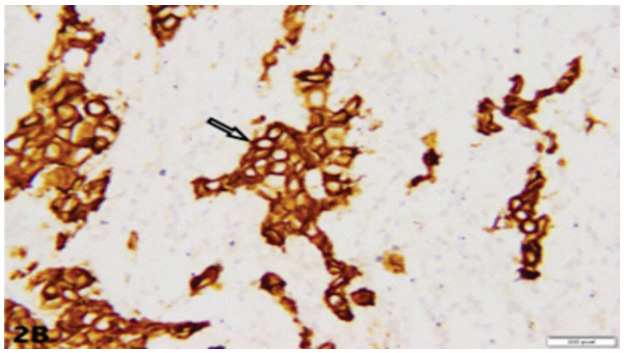


Figure 2A. (400X magnification) reveal moderate incomplete membranous staining in >10% neoplastic cells for HER2 antibody, scored 2 and Figure 2B. (400X magnification) depicts strong membranous immunostaining in >10% neoplastic cells for HER2 antibody, scored 3

Casterá and Bernet [19] found low to moderate interobserver concordance, with a kappa value of 0.2 to 0.4 among five observers, and also compared HER2 immunostain expression with ISH results. Concordance of HER2 immunostain was higher in ISH-positive cases as opposed to ISH-negative cases. The probable reason for the difference in opinion was the personalized interpretation of this biomarker. The analysis done by Sun *et al.,*[20] showed excellent agreement among observers and with digital image analysis as all the pathologists had expertise in breast cancer reporting in contrast to our study showing moderate concordance among general histopathologists as well as breast expert histopathologist with an overall kappa value of 0.65.

A study by Baez-Navarro *et al.,*[21] revealed low agreement at scores of 1 due to difficulty in the distinction between score 0 and 1 nonspecific staining and evaluating staining in 10% of tumor cells. This was in contrast to our study which had a moderate level of agreement at a score of 1. Umemura *et al.,*[22] conducted a research study and established a good level of agreement in immunostain interpretation among 10 laboratories and disagreement in results was due to different staining procedures in institutions. In contrast, our research study calculated interobserver concordance in one institution. Cross-institutional assessment by Robbins *et al.*[23] found substantial discordance at scores 1 and 2 using the 4-category scoring system (0,1,2, and 3) and higher agreement using the 3-category scoring system (0, low, and 3). The higher concordance in the latter 3-tier scoring system was due to the assembling of scores 1 and 2 as low scores, which made the interpretation untroubled.

Recently conducted studies by Baez Navarro et al.,[21] and Robbins et al.,[23] exhibited low concordance at scores 1 and 2 similar to our study which manifests discordance at scores 1 and 2, maximum at the latter. Among the 50 cases, 22 (44%) had discrepancies for multiple reasons encountered by the pathologists including faint membranous staining, conglomeration of the tumor, missing maximum interest area of HER2 antibody expression, and lack of experience. Among all the cases with ambiguity in our study, equivocal score 2 was the most controversial, because score 2 lies in the borderline

https://doi.org/10.57234/jiimc.june25.2303

category, neither positive nor negative making it a difficult zone for overall general pathologists, suggesting that such cases should be reported after receiving a second opinion from another pathologist having expertise in reporting breast pathology to counteract incorrect scoring practices and acquiring efficient training for HER2 reporting skills.

Recommendations and Limitations:

We did not correlate HER2 equivocal/score 2 with the fluorescence in situ hybridization (FISH) technique. So, we recommend that future studies should be conducted with a correlation of score 2 with the FISH studies.

Conclusion:

Based on our study objective, we found moderate interobserver concordance among pathologists with varying years of reporting experience in interpreting HER2 immunostain from scores 0 to 3. HER2 equivocal score 2 results, however, indicated significant interobserver variability. Thus, in order to minimize diagnostic discrepancies and to ensure optimal diagnosis, pathologists should be properly trained and consult interdepartmentally before interpreting HER2 immunostains.

Conflict of interest: Authors declared no conflict of interest.

Grant support and financial disclosure: Authors have declared no specific grant for this research from any funding agency in the public, commercial, or nonprofit sector.

Data sharing statement: The data that support the findings of this study are available from the corresponding author on the request.

REFERENCES

1. Arzanova E, Mayrovitz HN. The Epidemiology of Breast Cancer. In book:*Breast Cancer*; Mayrovitz, HN., Ed.; Exon Publications: Brisbane, Australia, 2022; Chapter 1: 1-20. doi:10.36255/exon-publications-breast-cancer-epidemiology.
2. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast cancer-epidemiology, risk factors, classification, prognostic markers, and current treatment strategies-an updated review. Cancers. 2021 Aug; 13(17):4287. doi:10.3390/cancers13174287.
3. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al. Breast Cancer Statistics, 2022. CA Cancer J Clin. 2022 Nov; 72(6): 524-41. doi:10.3322/caac.21754.
4. Collins A, Politopoulos I. The genetics of breast cancer: risk factors for disease. Appl Clin Genet. 2011 Jan; 4:11-9. doi:10.2147/TACG.S13139.
5. Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, et al. Risk Factors and Preventions of Breast Cancer. Int J Biol Sci. 2017 Nov; 13(11): 1387-97.doi:2017;13(11):1387.
6. Vinogradova Y, Coupland C, Hippisley-Cox J. Use of hormone replacement therapy and risk of breast cancer: nested case-control studies using the QResearch and CPRD databases. BMJ. 2020 Oct; 371: m3873. doi:10.1136/bmj.m3873.
7. Zaha DC. Significance of immunohistochemistry in breast cancer. World J Clin Oncol. 2014 Aug; 5(3):382-92. doi: 10.5306/wjco.v5.i3.382.
8. Li SG, Li L. Targeted therapy in HER2-positive breast cancer. Biomed Rep. 2013 Jul; 1 (4): 499–505. doi: 10.3892/br.2013.95.
9. Ahn S, Woo JW, Lee K, Park SY. HER2 status in breast cancer: changes in guidelines and complicating factors for interpretation. J Pathol Transl Med. 2020 Jan; 54(1):34-44. doi: 10.4132/jptm.2019.11.03.
10. Iqbal N, Iqbal N. Human epidermal growth factor receptor 2 (HER2) in cancers: overexpression and therapeutic implications. Molecular biology international. 2014; 1:852748. doi:10.1155/2014/852748.
11. Gutierrez C, Schiff R. HER2: biology, detection, and clinical implications. Arch Pathol Lab Med. 2011 Jan; 135(1):55-62. doi:10.5858/2010-0454-RAR.1.
12. Ivanova M, Porta FM, D'Ercole M, Pescia C, Sajjadi E, Cursano G, et al. Standardized pathology report for HER2 testing in compliance with 2023 ASCO/CAP updates and 2023 ESMO consensus statements on HER2-low breast cancer. Virchows Arch. 2024 Jan; 484(1):3-14. doi:10.1007/s00428-023-03656-w.
13. Zaakouk M, Quinn C, Provenzano E, Boyd C, Callagy G, Elsheikh S, et al. Concordance of HER2-low scoring in breast carcinoma among expert pathologists in the United Kingdom and the republic of Ireland -on behalf of the UK National Coordinating Committee for Breast pathology. The Breast. 2023 Aug; 70: 82-91. doi:10.1016/j.breast.2023.06.005.
14. Wynn CS, Tang SC. Anti-HER2 therapy in metastatic breast cancer: many choices and future directions. Cancer Metastasis Rev. 2022 Mar; 41(1):193-209. doi: 10.1007/s10555-022-10021-x.
15. Bradley R, Braybrooke J, Gray R, Hills R, Liu Z, Peto R, et al. Early Breast Cancer Trialists' Collaborative Group (EBCTCG: Trastuzumab for early-stage, HER2-positive breast cancer: a meta-analysis of 13 864 women in seven randomized trials. Lancet Oncol. 2021 Aug; 22(8):1139-1150. doi:10.1016/S1470-2045(21)00288-6.
16. Acs B, Fredriksson I, Rönnlund C, Hagerling C, Ehinger A, Kovács A, et al. Variability in Breast Cancer biomarker assessment and the effect on oncological treatment decisions: A nationwide 5-year population-based study. Cancers. 2021 Mar; 13(5): 1166. doi:10.3390/cancers 13051166.
17. Gordian-Arroyo AM, Zynger DL, Tozbikian GH. Impact of the 2018 ASCO/CAP HER2 Guideline Focused Update. Impact of the 2018 ASCO/CAP HER2 Guideline Focused Update. Am J Clin Pathol. 2019 Jun; 152(1):17-26. doi:10.1093/ajcp/aqz012.
18. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et

al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. N Engl J Med. 2022 Jul; 387(1):9-20. doi: 10.1056/NEJMoa2203690.
19. Casterá C, Bernet L. HER2 immunohistochemistry inter-observer reproducibility in 205 cases of invasive breast carcinoma additionally tested by ISH. Ann Diagn Pathol. 2020 Apr; 45: 151451. doi:10.1016/j.anndiagpath.2019.151451.
20. Sun H, Kang EY, Chen H, Sweeney KJ, Suchko M, Wu Y, et al. Immunohistochemical assessment of HER2 low breast cancer: interobserver reproducibility and correlation with digital image analysis. Breast Cancer Res Treat . 2024 Jun; 205(2):403-11. doi:10.1007/s10549-024-07256-3.
21. Baez-Navarro X, van Bockstal MR, Nawawi D, Broeckx G, Colpaert C, Doebar SC, et al. Interobserver variation in the assessment of immunohistochemistry expression levels in HER2-negative breast cancer: Can we improve the identification of low levels of HER2 expression by adjusting the criteria? An international interobserver study. Mod Pathol. 2023 Jan; 36(1):100009. doi:10.1016/j.modpat.2022.100009.
22. Umemura S, Osamura RY, Akiyama F, Honma K, Kurosumi M, Sasano H, et al. What causes discrepancies in HER2 testing for breast cancer? A Japanese ring study in conjunction with the global standard. Am J Clin Pathol. 2008 Dec; 130(6):883-91. doi:10.1309/AJCP5UUMFMA5ZKII.
23. Robbins CJ, Fernandez AI, Han G, Wong S, Harigopal M, Podoll M, et al. Multi-institutional assessment of pathologist scoring HER2 immunohistochemistry. Mod Pathol. 2023 Jan; 36(1):100032. doi:10.1016/j.modpat.2022.100032.

CONFLICT OF INTEREST

Authors declared no conflicts of Interest.

GRANT SUPPORT AND FINANCIAL DISCLOSURE

Authors have declared no specific grant for this research from any funding agency in public, commercial or nonprofit sector.

DATA SHARING STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.