

Low interobserver agreement among subspecialised breast pathologists in evaluating HER2-low breast cancer

Gulisa Turashvili ¹, Yuan Gao,¹ Di (Andy) Ai,¹ Abdulwahab M Ewaz,¹ Sandra Gjorgova Gjeorgjievski,¹ Qun Wang,¹ Thi T A Nguyen,¹ Chao Zhang,² Xiaoxian Li¹

¹Pathology and Laboratory Medicine, Emory University Hospital, Atlanta, Georgia, USA
²General Dynamics Information Technology Inc, Falls Church, Virginia, USA

Correspondence to
Xiaoxian Li; bill.li@emory.edu

Received 11 July 2023
Accepted 24 August 2023
Published Online First
15 September 2023

ABSTRACT

Aims Metastatic HER2-low breast cancer (HLBC) can be treated by trastuzumab deruxtecan. Assessment of low levels of HER2 protein expression suffers from poor interobserver reproducibility. The aim of the study was to evaluate the interobserver agreement among subspecialised breast pathologists and develop a practical algorithm for assessing HLBC.

Methods Six breast pathologists (4 juniors, 2 seniors) evaluated 106 HER2 immunostained slides with 0/1+ expression. Two rounds (R1, R2) of ring study were performed before and after training with a modified Ki-67 algorithm, and concordance was assessed.

Results Agreement with 5% increments increased from substantial to almost perfect (R1: 0.796, R2: 0.804), and remained substantial for three categories (<1% vs 1%–10% vs >10%) (R1: 0.768, R2: 0.764). Seniors and juniors had almost perfect agreement with 5% increments (R1: 0.859 and 0.821, R2: 0.872 and 0.813). For the three categories, agreement remained almost perfect among seniors (R1: 0.837, R2: 0.860) and substantial among juniors (R1: 0.792, R2: 0.768). Binary analysis showed suboptimal agreement, decreasing for both juniors and seniors from substantial (R1: 0.650 and 0.620) to moderate (R2: 0.560 and 0.554) using the 1% cut-off, and increasing from moderate to substantial (R1: 0.478, R2: 0.712) among seniors but remaining moderate (R1: 0.576, R2: 0.465) among juniors using the 10% cut-off. The average scoring time per case was higher (72 vs 92 s).

Conclusions Subspecialised breast pathologists have suboptimal agreement for immunohistochemical evaluation of HLBC using the modified Ki-67 methodology. An urgent need remains for a new assay/algorithm to reliably evaluate HLBC.

INTRODUCTION

Invasive breast carcinomas with amplification of human epidermal growth factor receptor 2 (*ERBB2*) gene or overexpression of the corresponding HER2 protein have distinct tumour biology, and HER2 status is a strong predictive biomarker for tumour response to HER2 targeted therapy.¹ HER2 status is usually determined by immunohistochemistry (IHC) and fluorescent in situ hybridisation (FISH).^{2,3} Based on the current American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) recommendations, HER2

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Targeted therapy has recently been shown to improve clinical outcomes of metastatic HER2-low breast cancer (HLBC). However, assessment of low levels of HER2 protein expression suffers from poor interobserver reproducibility.

WHAT THIS STUDY ADDS

⇒ Two rounds of ring study were performed before and after training with a practical algorithm developed for immunohistochemistry (IHC) evaluation of HLBC based on the modified Ki-67 methodology. Concordance among subspecialised breast pathologists did not significantly improve.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The proposed algorithm may help improve agreement among general anatomic pathologists. An urgent need remains for a new assay/algorithm to reliably evaluate HLBC.

IHC scores are classified as 0, 1+, 2+ and 3+ (table 1).^{2,4} Breast cancers with 0 or 1+ scores are currently defined as HER2-negative (HER2-) and do not require further evaluation by FISH. Breast cancers with 2+ (equivocal) scores should be subjected to reflex FISH analysis for determining gene amplification status. Breast cancers with either 3+ scores or *ERBB2* gene amplification by FISH are currently defined as HER2-positive (HER2+).

Traditionally, only patients with HER2+ breast cancers have been eligible for HER2 targeted therapy.^{5–7} Trastuzumab deruxtecan (T-DXd) is an antibody drug conjugate composed of trastuzumab, which recognises HER2 protein on tumour cell surface, and cytotoxic topoisomerase I inhibitor (payload) through an enzyme-cleavable linker.^{8,9} Following binding of T-DXd to tumour cell surface HER2, T-DXd undergoes internalisation and the linker is cleaved by lysosomal enzymes to release the topoisomerase I inhibitor payload component and inhibit tumour cell growth. The ratio of drug payload-to-antibody is approximately 8 and the release of the drug payload has cytotoxic effects on adjacent tumour cells regardless of their HER2 expression level, known as bystander effect.¹⁰ In April 2022, T-DXd was granted breakthrough



© Author(s) (or their employer(s)) 2024. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Turashvili G, Gao Y, Ai DA, et al. *J Clin Pathol* 2024;**77**:815–821.

Table 1 Current ASCO/CAP recommendations for HER2 IHC evaluation

HER2 IHC score	ASCO/CAP recommendations
0	No staining or incomplete and faint membrane staining in $\leq 10\%$ of tumour cells
1+	Incomplete and faint membrane staining in $>10\%$ of tumour cells
2+	Weak to moderate complete membrane staining in $>10\%$ of tumour cells, or intense complete membrane staining in $\leq 10\%$ of tumour cells
3+	Intense complete membrane staining in $>10\%$ of tumour cells
ASCO/CAP, American Society of Clinical Oncology/College of American Pathologists; IHC, immunohistochemistry.	

therapy designation by the Food and Drug Administration (FDA) to treat patients with metastatic HER2-low breast cancer (HLBC)¹¹ based on the DESTINY-Breast04 trial which demonstrated improved clinical outcomes with addition of T-DXd to the therapies in patients with metastatic HLBC.^{12 13} In this trial, HLBC was defined as HER2 IHC scores of 1+ or 2+ and lacking gene amplification by FISH.

With this FDA approval, patients with HLBC can be treated with T-DXd as the targeted therapy. However, the difference between 0 and 1+ or 1+ and 2+ scores can be difficult to discern. There is no established protocol to guide evaluation of low HER2 levels, leading to high interobserver variability. A recent study showed poor interobserver agreement among 18 pathologists with only 26% concordance for 0 and 1+ scores.¹⁴ Furthermore, a multicentre international study demonstrated moderate agreement for distinguishing 0 and HER2-low scores, which did not improve by FISH.¹⁵ Inaccurate evaluation of HER2 IHC may lead to suboptimal treatment; for example, patients with metastatic HLBC would be ineligible for T-DXd if HER2 IHC were erroneously assessed. Therefore, there is an urgent unmet need to develop a standardised protocol for identifying HLBC.

The overarching goal of this study was to develop a practical algorithm for IHC evaluation of HLBC and examine its impact on interobserver agreement.

MATERIALS AND METHODS

Study cohort

A total of 106 HER2-negative breast carcinomas (67 biopsies and 39 excisions that met ASCO/CAP guidelines for cold ischaemia time and fixation time) were identified in the pathology archives in which HER2 status was assessed using the ASCO/CAP guidelines as part of routine clinical biomarker testing,¹⁶ including 60 cases reported as HER2-negative with IHC 1+ scores (weak incomplete staining in $>10\%$ of tumour cells) and 46 cases with reported as HER2-negative with IHC 0 scores, the latter group consisting of tumours without any staining and tumours with weak incomplete staining in $<10\%$ of tumour cells. HER2 IHC was performed using a laboratory developed

test.^{17–19} Briefly, slides were stained with the HercepTest HER2 antibody (Dako) using an automated Leica Bond III stainer (Leica Biosystems, Deer Park, Illinois, USA) following antigen retrieval in citrate buffer at pH 6.0 for 20 min. The slides were evaluated by six board certified pathologists with subspecialised practice in breast pathology. Of the six pathologists, two were seniors with ≥ 5 years of practice experience, and four were junior with < 5 years of experience.

Development of an algorithm for evaluating HER2 IHC scores

The proposed algorithm included: (1) evaluation of the whole slide at $\times 100$ magnification to identify staining heterogeneity; (2) in each area with similar staining percentage, evaluation of the % of incomplete membrane staining of any intensity at $\times 400$ magnification using the eyeballing method and (3) calculation of the global HER2 score. The algorithm is summarised in table 2. In tumours with homogeneous distribution of HER2 membrane staining, only two high power field estimations were performed, and the average was used to calculate a final HER2 IHC score. Figure 1 shows an example of evaluating HER2 IHC in cases with heterogeneous distribution of HER2 IHC staining.

Two rounds of HER2 IHC evaluation

In both rounds, HER2 IHC was evaluated as the percentage (%) of neoplastic cells with membrane staining in 5% increments: <1 , 1–5, 6–10, 11–15, 16–20, 21–25, 26–30, >30 (table 3). No additional specific instructions were given. A research assistant recorded the scores and the time spent per case.

Round 2 was performed using the same slides after 1 month of wash-out time. Training with the proposed protocol described above was provided to all pathologists before round 2. The training included a PowerPoint presentation and demonstration of HER2 IHC evaluation of 10 cases previously reported as HER2-negative (0/1+). These 10 training cases were not included in the study cohort. Within 1 week of the training session, the same six pathologists re-evaluated the same cases. The same research assistant recorded the scores and time spent per case.

Statistical analysis

We first evaluated the interobserver agreement in round 1 and 2 (before and after the training with the proposed algorithm). Three score systems were analysed: (A) incremental scores with 5% increase as $<1\%$, 1%–5%, 6%–10%, 11%–15%, 16%–20%, 21%–25%, 26%–30%, $>30\%$; (B) scores with three categories as $<1\%$ vs 1%–10% vs $>10\%$ and (C) scores with binary categories ($<1\%$ vs $\geq 1\%$ or 0%–10% vs $>10\%$) (figure 2).

Concordance was assessed using Kendall's coefficient (W) for 5% increments and three categories ($<1\%$ vs 1%–10% vs $>10\%$), and Fleiss kappa coefficient (K) for binary categories ($<1\%$ vs $\geq 1\%$ or 0%–10% vs $>10\%$). Kendall's coefficient of concordance is a nonparametric statistic related to Friedman's test and tests agreement of the raters' rankings of the items. It

Table 2 Stepwise approach for the proposed standardised HER2 IHC evaluation protocol

Step 1	Evaluate the whole slide at $\times 10$ to identify IHC staining heterogeneity
Step 2	Estimate the percentage of areas with different percentage of HER2 IHC staining (with $>10\%$ difference)
Step 3	Evaluate each area at high power field ($\times 40\times$) and estimate the % of incomplete and complete membranous staining using the eyeballing method
Step 4	Calculate the global HER2 complete and incomplete membranous staining: ((average % of positive cell in area 1 $\times$ % of area 1) + (average % of positive cells in area 2 $\times$ % of area 2) + (average % of positive cells in area 3 $\times$ % of area 3) + ... + (average % of positive cells in area N $\times$ % of area N)) $\times 100$ (%)
IHC, immunohistochemistry.	

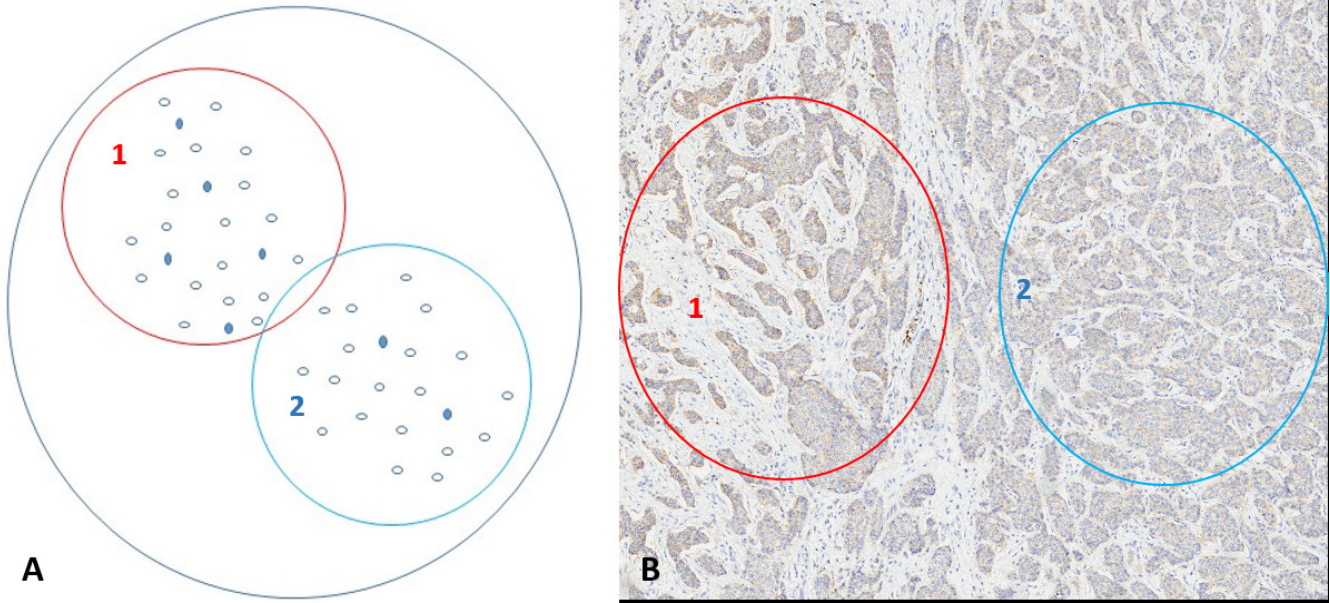


Figure 1 Schematic (A) and microscopic (B) examples of evaluating HER2 immunohistochemical (IHC) staining in tumours with heterogeneous HER2 protein expression using the eyeballing method. Each blue dot represents a tumour cell with incomplete IHC membrane staining of any intensity and each empty dot represents a negative tumour cell. In this example, circle 1 has 20% positive cells and comprises 50% of the entire tumour area; circle 2 has 10% positive cells and comprises 50% of the entire tumour area. The global % of incomplete HER2 membrane staining is $[(50\% \times 20\%) + (50\% \times 10\%)] \times 100 (\%) = 5\%$. IHC, original magnification $\times 100$ (B).

is available when the response is ordinal and numerically coded. P values of <0.05 were considered significant. When using the Kendall's coefficient analysis, agreement is assessed as fair (0.20 to <0.40), moderate (0.40 to <0.60), substantial (0.60 to <0.80) and perfect (≥ 0.80). In Kappa coefficient analysis, agreement is fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) or perfect (0.81–1.00). Statistical analysis was conducted by using SAS V.9.4.

RESULTS

Interobserver variability without training with the proposed algorithm

In round 1, the six pathologists assessed HER2 expression in 5% increments. Average scoring time per case was 72 s (range 34–122). Interobserver agreement was substantial for ordinal scale ratings analysis with 5% increments and three categories ($<1\%$, $1\%–10\%$ and $>10\%$) ($W=0.796$ and $W=0.768$, respectively). Dichotomous scale analysis also demonstrated substantial agreement for the 1% cut-off (0% vs $\geq 1\%$) ($K=0.650$) and moderate for the 10% cut-off ($0\%–10\%$ vs $>10\%$) ($K=0.569$) (table 4).

Concordance was further analysed based on the pathologists' seniority. For senior pathologists, interobserver agreement was almost perfect for ordinal analysis including 5% increments and 3 categories of $<1\%$, $1\%–10\%$ and $>10\%$ ($W=0.859$ and $W=0.837$, respectively). Binary analysis also

demonstrated substantial agreement for the 1% cut-off (0% vs $\geq 1\%$) ($K=0.620$), and moderate for the 10% cut-off ($0\%–10\%$ vs $>10\%$) ($K=0.478$). For junior pathologists, interobserver agreement was almost perfect for ordinal analysis including 5% increments ($W=0.821$) and substantial for 3 categories of $<1\%$, $1\%–10\%$ and $>10\%$ ($W=0.792$). Binary analysis demonstrated substantial agreement with the 1% (0% vs $\geq 1\%$) cut-off ($K=0.650$) and moderate agreement with the 10% ($0\%–10\%$ vs $>10\%$) cut-off ($K=0.576$) (table 5).

Interobserver variability after training with the proposed algorithm

In round 2, HER2 IHC was assessed using the proposed protocol. Average scoring time per case was 92 s (range 33–209) compared with 72 s in round 1. Interobserver agreement was almost perfect for ordinal analysis including 5% increments ($W=0.804$) and substantial for 3 categories of $<1\%$, $1\%–10\%$ and $>10\%$ ($W=0.764$). Binary analysis demonstrated moderate agreement when using the 1% (0% vs $\geq 1\%$) ($K=0.590$) or 10% cut-offs ($0\%–10\%$ vs $>10\%$) ($K=0.549$) (table 4).

Concordance was further analysed based on the pathologists' seniority. For senior pathologists, interobserver agreement was almost perfect for ordinal scale ratings analysis including 5% increments and 3 categories of $<1\%$, $1\%–10\%$ and $>10\%$ ($W=0.872$ and $W=0.860$, respectively). Dichotomous scale analysis demonstrated moderate agreement for the 1% cut-off (0% vs $\geq 1\%$) ($K=0.554$), and substantial for the 10% cut-off ($0\%–10\%$ vs $>10\%$) ($K=0.712$). For junior pathologists, interobserver agreement was perfect for ordinal analysis with 5% increments ($W=0.813$) and substantial for 3 categories of $<1\%$, $1\%–10\%$ and $>10\%$ ($W=0.768$). Dichotomous scale analysis demonstrated moderate agreement with both the 1% (0% vs $\geq 1\%$) and 10% ($0\%–10\%$ vs $>10\%$) cut-offs ($K=0.560$ and $K=0.465$, respectively) (table 2).

Table 3 Example HER2 IHC scoring sheet	
	%
Case no	<1 1–5 6–10 11–15 16–20 21–25 26–30 >30
1	
2	
IHC, immunohistochemistry.	

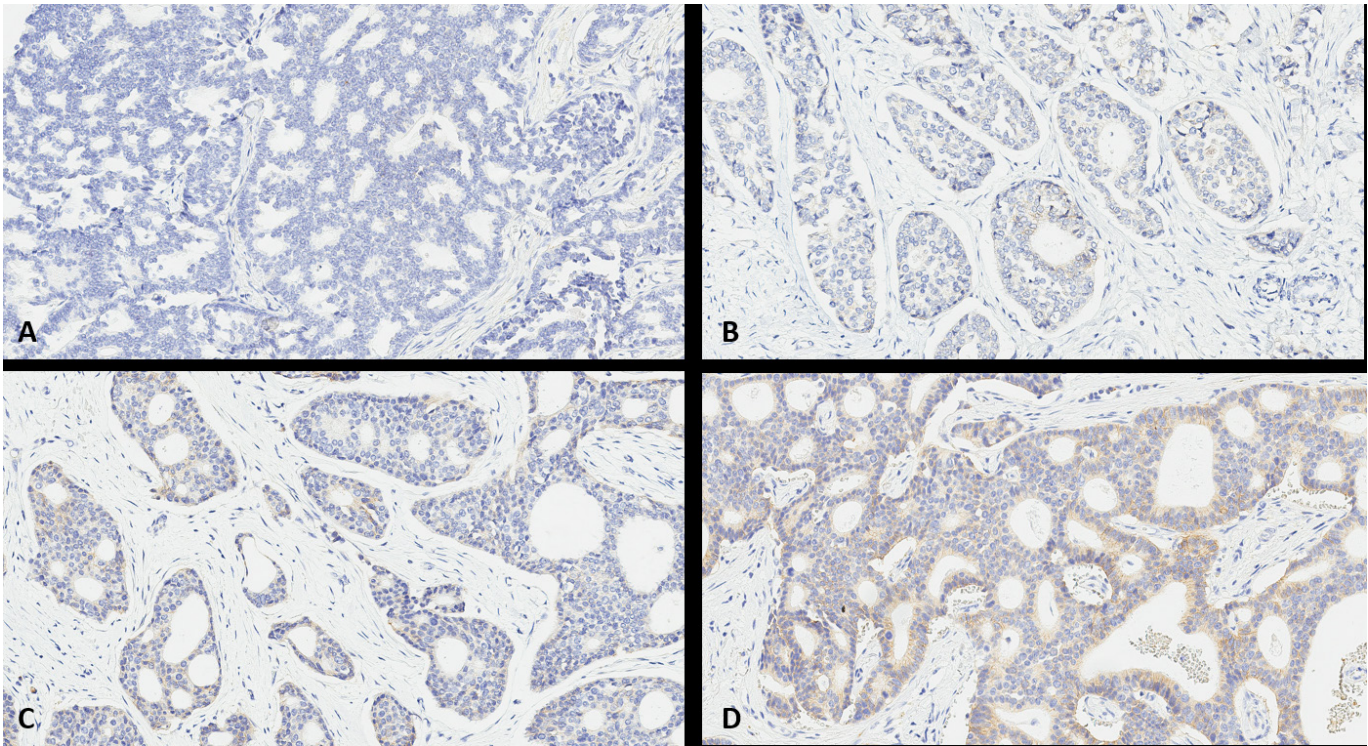


Figure 2 Examples of HER2 immunohistochemical (IHC) staining with no expression (A), <5% of tumour cells with weak incomplete membranous staining (B), 15% of tumour cells with weak incomplete membranous staining (C), and >30% of tumour cells with weak to moderate incomplete membranous staining (D). IHC, original magnification $\times 400$ (A–D). IHC, immunohistochemistry.

Comparison of interobserver variability before and after training with the proposed algorithm

Average scoring time per case was slightly higher in round 2 compared with round 1 (92 s vs 72 s). Interobserver agreement for 5% increments was substantial in round 1 ($W=0.796$) and increased to almost perfect ($W=0.804$) in round 2. The agreement slightly increased but remained substantial for three categories (<1% vs 1%–10% vs >10%) in round 2 ($W=0.764$) from round 1 ($W=0.768$). Junior and senior pathologists had almost perfect agreement for 5% increments in round 1 ($W=0.821$ and $W=0.859$) and round 2 ($W=0.813$ and $W=0.872$). For the three categories, agreement in round 2 slightly increased but remained almost perfect among seniors ($W=0.837$ vs $W=0.860$) and slightly decreased but remained substantial among juniors ($W=0.792$ vs $W=0.768$). For binary analysis using the 1% cut-off, agreement between juniors and seniors was substantial in round 1 ($K=0.650$ vs $K=0.620$) but decreased to moderate in round 2 ($K=0.554$ vs $K=0.560$). When using the 10% cut-off, the agreement of seniors increased from moderate ($K=0.492$) in round 1 to substantial ($K=0.712$) in round 2 and remained moderate among juniors ($K=0.576$ and $K=0.465$).

DISCUSSION

Although conventionally only patients with HER2+ breast cancer have been eligible for HER2 targeted therapy,^{20 21} antibody drug conjugates such as T-DXd provide opportunities for patients with HLBC.^{22–24} At least 50% of breast cancers can be categorised as HER2-low, that is, showing IHC 1+ or 2+ scores with negative FISH.^{25–27} T-DXd had clinical benefits even in 30% of patients with HER2 IHC 0 scores in the DAISY trial, which is almost identical to that of patients with HER2 IHC 1+ scores²⁸; this may be due to inclusion of HER2 ultra-low tumours (defined as incomplete and weak membranous staining in $\leq 10\%$ of tumour cells) in the IHC 0 group, pathologists’ poor agreement of HER2 IHC evaluation or similar HER2 levels in HER2 ultra-low and HER2 1+ breast cancers.^{25 26 29 30} Accurate assessment of HER2 expression is therefore crucial for optimal clinical management.

Previous studies have demonstrated low reproducibility of HLBC assessment. The CAP surveys over 2 years from 1391 to 1452 laboratories of 40 cases from each laboratory (20 cases biannually for a total of 80) demonstrated $\leq 70\%$ concordance for 0 vs 1+ scores in 19% of cases.¹⁴ In a cohort of 170 scanned breast biopsies, concordance among 18 pathologists was only

Table 4 Interobserver agreement in round 1 and round 2			
Question	Statistical parameter	Round 1	Round 2
Ordinal analysis including all categories	Kendall’s W	0.796	0.804
Ordinal analysis of 3 categories <1% vs 1%–10% vs >10%	Kendall’s W	0.768	0.764
Binary analysis using one as cut-off (0 vs all others ($\geq 1\%$))	Kappa	0.650 (0.559, 0.740)	0.590 (0.499, 0.681)
Binary analysis using 10% as cut-off (0%–10% vs >10%)	Kappa	0.569 (0.468, 0.670)	0.549 (0.451, 0.647)

Table 5 Interobserver agreement in round 1 and round 2 by seniority

Question	Statistical parameter	Round 1		Round 2	
		Senior	Junior	Senior	Junior
Ordinal analysis including all categories	Kendall's Coefficient	0.859	0.821	0.872	0.813
Ordinal analysis of 3 categories <1% vs 1%–10% vs >10%	Kendall's Coefficient	0.837	0.792	0.860	0.768
Binary analysis using one as cut-off (0% vs all others (≥1%))	Kappa coefficient	0.620 (0.467, 0.783)	0.650 (0.547, 0.753)	0.554 (0.389, 0.718)	0.560 (0.451, 0.669)
Binary analysis using 10 as cut-off (0%–10% vs >10%)	Kappa coefficient	0.478 (0.302, 0.654)	0.576 (0.459, 0.693)	0.712 (0.566, 0.857)	0.465 (0.350, 0.580)

26% for 0/1+ scores compared with 58% for 2+/3+ scores.¹⁴ Of note, the study pathologists were unaware of the importance of identifying HLBC. Blinded analysis of 200 scanned HER2 IHC stained slides from 100 independent cases including all 4 HER2 categories (0, 1+, 2+, 3+) by 5 breast pathologists revealed substantial agreement (K=0.79) with a 35% overall discordance rate. The discordant cases consisted of 15 1+ vs 0 scores, 12 1+ vs 2+ scores, 1 2+ vs 0 score, 1 3+ vs 1+ score and 6 3+ vs 2+ scores. The agreement was almost perfect for the 0 and 3+ scores (K=0.82 and K=0.92, respectively), but only substantial for the 1+ and 2+ scores (K=0.67 and K=0.74, respectively).²⁶

There are no guidelines or recommendations on evaluating HLBC. The International Ki67 in Breast Cancer Working Group has developed a Ki67 evaluation protocol with high interobserver agreement which requires evaluation of the percentage (%) of areas with low, medium and high Ki67 index and then counting 100 cells in each area to calculate the final global Ki67 score.³¹ Despite significant improvement of interobserver agreement, each case takes a median of 9 min.³¹ No such protocol exists for HER2 IHC. Our simplified version of the Ki67 protocol increased the average scoring time per case from 72 to 92 s. The interobserver agreement for 5% increments increased from substantial to almost perfect (W=0.796 vs W=0.804), and remained substantial for 3 categories of <1% vs 1%–10% vs >10% (W=0.768 vs W=0.764). When analysing the data based on the pathologists' seniority, the agreement remained almost perfect for both juniors (W=0.821 vs W=0.813) and seniors (W=0.859 vs W=0.872) for 5% increments. It also remained almost perfect among seniors (W=0.837 vs 0.860) and substantial among juniors (W=0.792 vs W=0.768) for the three categories. For binarised scores based on the 1% cut-off, agreement decreased from substantial to moderate among both juniors (K=0.650 vs K=0.560) and seniors (K=0.620 vs K=0.554). The agreement remained moderate among juniors (K=0.576 vs K=0.465) but increased from moderate to substantial (K=0.492 vs K=0.712) among seniors for binarised scores using the 10% cut-off. Of note, all study pathologists are subspecialised and were aware of the importance of separating 1+ vs 0 scores. Although the proposed protocol may be useful for uniform HER2 IHC assessment, the interobserver agreement in evaluating HLBC is suboptimal among subspecialised breast pathologists even after training with the protocol. The interobserver variability may be higher among general anatomic pathologists.

Several studies investigated the role of artificial Intelligence (AI) in evaluating HLBC. The performance of the AI digital image analysis (DIA)-assisted workflow was recently assessed in a cohort of 67 primary breast carcinomas and 30 metastases in which 3 breast pathologists independently assessed HER2 expression first visually (ground truth) and after being provided the results of the DIA. There was moderate agreement (K=0.59) between the ground truth and AI, with most discrepancies

occurring between 0 and 1+ scores.³² Another study included 363 cases with HER2 IHC scores 0, 1+ and 2+ (without HER2 gene amplification) and available HER2 mRNA level. Artificial neural network analysis was then used to distinguish 0 vs 1+ scores. Score 1+ was refined as either faint staining in ≥20% of cells irrespective of the circumferential completeness, weak complete staining in ≤10% of tumour cells, or weak incomplete staining in >10% and moderate incomplete staining in ≤10% of tumour cells. Based on the refined criteria, 63% of cases were reclassified as HER2-low, and the refined scores showed perfect agreement with the original clinical scores.³³ However, AI-based analysis may require substantial infrastructure investment and expertise. Evaluation under light microscope by pathologists is still the standard practice and a practical HER2 IHC evaluation algorithm will be of great help.

The interpretation of HLBC has been addressed by the recently published updated ASCO/CAP guidelines, which recommend including the following comment in biomarker reports: 'patients with breast cancers that are HER2 IHC 3+ or IHC 2+/ISH amplified may be eligible for several therapies that disrupt HER2 signalling pathways. Invasive breast cancers that test 'HER2-negative' (IHC 0, 1+ or 2+/ISH not-amplified) are more specifically considered 'HER2-negative for protein overexpression/gene amplification' since non-overexpressed levels of the HER2 protein may be present in these cases. Patients with breast cancers that are HER2 IHC 1+ or IHC 2+/ISH not amplified may be eligible for a treatment that targets non-amplified/non-overexpressed levels of HER2 expression for cytotoxic drug delivery (IHC 0 results do not result in eligibility currently)'.⁴ However, the most recent CAP breast biomarker protocol from March 2023 includes the following comment: 'Breast cancers with HER2 IHC score 1+ or HER2 IHC score 2+ and a negative ISH result are eligible for clinically appropriate HER2-targeted therapy and may be reported as 'HER2 Low'.³⁴ ASCO/CAP 2023 guidelines and CAP biomarker reporting guidelines will yet have to be fully aligned.

Of note, the currently used HER2 IHC assays were designed to identify HER2+ tumours. The concept of HLBC is evolving, and whether these assays are ideal for identifying HLBC remains controversial. However, as long as IHC is used to evaluate HER2 expression, a standardised protocol is needed to improve interobserver variation. Furthermore, updated guidelines for IHC interpretation (analytical phase) as well as accurate and reproducible testing methodologies and strategies (preanalytical phase) are necessary to improve diagnostic sensitivity and avoid under-reporting or over-reporting of HLBC.^{35–36} Preanalytical factors are particularly complex and include length of fixation, antigen retrieval, antibody clones (eg, 4B5, CB11, HercepTest) and dilution, incubation time, temperature.^{37–38} These factors may have a tremendous impact on the accuracy and reproducibility of HER2 IHC results and identification of HLBC.^{37–39} Therefore, each laboratory should have well thought out and rigorous quality control procedures in place for the preanalytical

phase combined with well-defined guidelines for HER2 IHC assessment for HLBC.^{35 40}

Our study has limitations. Although IHC has been the standard method of evaluating HER2 expression in invasive breast cancer for years, it may not be the most optimal technique for semiquantifying low levels of HER2 protein expression. A broad range of HER2 mRNA expression has been identified in tumours without detectable HER2 protein by IHC.⁴¹ Distinguishing between 0 and 1+ scores appeared challenging with the proposed algorithm, and the clear cut-off point to determine the eligibility for T-DXd remains to be determined. Additionally, the Ventana HER2 4B5 antibody clone has been shown to identify a higher proportion of HER2-low tumours compared with HercepTest (27.4% vs 9.2%).⁴² It is possible that the use of the HercepTest antibody may have affected the incidence of HLBC; however, the main aim of the study was to assess interobserver variability which is unlikely to be dependent on the antibody clone.

In conclusion, this analytical validation study indicates that subspecialised breast pathologists have suboptimal agreement in evaluating HLBC. Although our proposed algorithm using the modified Ki-67 assessment methodology did not significantly improve interobserver variability among breast pathologists for IHC evaluation of HLBC, it may help improve interobserver agreement among general anatomic pathologists. There is an urgent need to develop a new assay or algorithm to reliably evaluate HLBC.

Handling editor L C Collins.

Contributors Conception or design of the work: XL. Assessment of immunohistochemical staining: GT, DAA, AME, SGG, QW and XL. Data collection: YG and TTAN. Statistical analysis: CZ. Data interpretation: GT and XL. Drafting the article: GT and XL. Critical revision of the article: all authors. Final approval of the version to be published: all authors. Author responsible for the overall content as the guarantor: XL.

Funding This study was funded by the College of American Pathologists Foundation 2022 Young Investigator HER2 Research Award, sponsored by AstraZeneca and Daiichi Sankyo (XL).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the institutional ethics review board. (protocol #RB00077806)

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iD

Gulisa Turashvili <http://orcid.org/0000-0001-6125-5865>

REFERENCES

- Gampenrieder SP, Rinnerthaler G, Tinchon C, et al. Landscape of HER2-low metastatic breast cancer (MBC): results from the Austrian AGMT_MBC-Registry. *Breast Cancer Res* 2021;23:112.
- Wolff AC, Hammond MEH, Allison KH, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American Pathologists clinical practice guideline focused update. *Arch Pathol Lab Med* 2018;142:1364–82.
- Wolff AC, Hammond MEH, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American Pathologists clinical practice guideline update. *Arch Pathol Lab Med* 2014;138:241–56.
- Wolff AC, Somerfield MR, Dowsett M, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology-college of American Pathologists guideline update. *JCO* 2023;41:3867–72.
- Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344:783–92.
- von Minckwitz G, Procter M, de Azavedo E, et al. Adjuvant pertuzumab and trastuzumab in early Her2-positive breast cancer. *N Engl J Med* 2017;377:122–31.
- Martínez-Sáez O, Prat A. Current and future management of HER2-positive metastatic breast cancer. *JCO Oncol Pract* 2021;17:594–604.
- Doi T, Shitara K, Naito Y, et al. Safety, pharmacokinetics, and antitumour activity of trastuzumab deruxatecan (DS-8201), a HER2-targeting antibody-drug conjugate, in patients with advanced breast and gastric or gastro-oesophageal tumours: a phase 1 dose-escalation study. *Lancet Oncol* 2017;18:1512–22.
- Keam SJ. Trastuzumab deruxatecan: first approval. *Drugs* 2020;80:501–8.
- Chen Y-F, Xu Y-Y, Shao Z-M, et al. Resistance to antibody-drug conjugates in breast cancer: mechanisms and solutions. *Cancer Commun (Lond)* 2023;43:297–337.
- Enhertu granted breakthrough therapy designation in the US for patients with HER2-low metastatic breast cancer. 2022. Available: <https://www.astrazeneca.com/media-centre/press-releases/2022/enhertu-granted-btd-her2-low-breast-cancer.html>
- Modi S, Saura C, Yamashita T, et al. Trastuzumab deruxatecan in previously treated HER2-positive breast cancer. *N Engl J Med* 2020;382:610–21.
- Cortés J, Kim S-B, Chung W-P, et al. Trastuzumab deruxatecan versus trastuzumab emtansine for breast cancer. *N Engl J Med* 2022;386:1143–54.
- Fernandez AI, Liu M, Bellizzi A, et al. Examination of low ERBB2 protein expression in breast cancer tissue. *JAMA Oncol* 2022;8:1–4.
- Baez-Navarro X, van Bockstal MR, Nawawi D, 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;36:100009.
- Wolff AC, Hammond MEH, Allison KH, et al. Her2 testing in breast cancer: American society of clinical oncology/college of American pathologists clinical practice guideline focused update summary. *J Oncol Pract* 2018;14:437–41.
- Zhao J, Krishnamurti U, Zhang C, et al. HER2 immunohistochemistry staining positivity is strongly predictive of tumor response to Neoadjuvant chemotherapy in HER2 positive breast cancer. *Pathol Res Pract* 2020;216:153155.
- Meisel JL, Zhao J, Suo A, et al. Clinicopathologic factors associated with response to neoadjuvant anti-HER2-directed chemotherapy in HER2-positive breast cancer. *Clin Breast Cancer* 2020;20:19–24.
- Schneider F, Jin Y, Van Smaalen K, et al. The FDA-approved breast cancer HER2 evaluation kit (Herceptest; Dako) may miss some HER2-positive breast cancers. *Am J Clin Pathol* 2019;151:504–10.
- Swain SM, Shastry M, Hamilton E. Targeting HER2-positive breast cancer: advances and future directions. *Nat Rev Drug Discov* 2023;22:101–26.
- Tarantino P, Curigliano G, Parsons HA, et al. Aiming at a tailored cure for ERBB2-positive metastatic breast cancer: a review. *JAMA Oncol* 2022;8:629–35.
- Venetis K, Crimini E, Sajjadi E, et al. HER2 low, ultra-low, and novel complementary biomarkers: expanding the spectrum of HER2 positivity in breast cancer. *Front Mol Biosci* 2022;9:834651.
- Indini A, Rijavec E, Grossi F. Trastuzumab deruxatecan: changing the destiny of HER2 expressing solid tumors. *Int J Mol Sci* 2021;22:4774.
- Modi S, Park H, Murthy RK, et al. Antitumor activity and safety of trastuzumab deruxatecan in patients with HER2-low-expressing advanced breast cancer: results from a phase IB study. *J Clin Oncol* 2020;38:1887–96.
- Banerji U, van Herpen CML, Saura C, et al. Trastuzumab duocarmazine in locally advanced and metastatic solid tumours and HER2-expressing breast cancer: a phase 1 dose-escalation and dose-expansion study. *Lancet Oncol* 2019;20:1124–35.
- Schettini F, Chic N, Brasó-Maristany F, et al. Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. *NPJ Breast Cancer* 2021;7:1.
- Tarantino P, Hamilton E, Tolane SM, et al. HER2-low breast cancer: pathological and clinical landscape. *J Clin Oncol* 2020;38:1951–62.
- Diéras V, Deluche E, Lusque A, et al. Trastuzumab deruxatecan (T-Dxd) for advanced breast cancer patients (ABC), regardless HER2 status: a phase II study with biomarkers analysis (DAISY). *Cancer Res* 2022;82(4_Supplement):D8–O2.
- Tarantino P, Jin Q, Tayob N, et al. Prognostic and biologic significance of ERBB2-low expression in early-stage breast cancer. *JAMA Oncol* 2022;8:1177–83.
- Denkert C, Seither F, Schneeweiss A, et al. Clinical and molecular characteristics of HER2-low-positive breast cancer: pooled analysis of individual patient data from prospective, neoadjuvant clinical trials. *Lancet Oncol* 2021;22:1151–61.
- Nielsen TO, Leung SCY, Rimm DL, et al. Assessment of Ki67 in breast cancer: updated recommendations from the international Ki67 in breast cancer working group. *J Natl Cancer Inst* 2021;113:808–19.
- Palm C, Connolly CE, Masser R, et al. Determining HER2 status by artificial intelligence: an investigation of primary, metastatic, and HER2 low breast tumors. *Diagnostics (Basel)* 2023;13:168.
- Atallah NM, Toss MS, Green AR, et al. Refining the definition of HER2-low class in invasive breast cancer. *Histopathology* 2022;81:770–85.
- College of American Pathologists. Breast biomarker reporting V1.5.0.1. 2023. Available: <https://www.cap.org/protocols-and-guidelines/cancer-reporting-tools/cancer-protocol-templates>
- Sajjadi E, Venetis K, Ivanova M, et al. Improving HER2 testing reproducibility in HER2-low breast cancer. *Cancer Drug Resist* 2022;5:882–8.
- Zhang H, Moisini I, Ajabnoor RM, et al. Applying the new guidelines of HER2 testing in breast cancer. *Curr Oncol Rep* 2020;22:51.

- 37 Zhang H, Katerji H, Turner BM, *et al.* HER2-low breast cancers. *Am J Clin Pathol* 2022;157:328–36.
- 38 Sapino A, Goia M, Recupero D, *et al.* Current challenges for HER2 testing in diagnostic pathology: state of the art and controversial issues. *Front Oncol* 2013;3:129.
- 39 Schrohl A-S, Pedersen HC, Jensen SS, *et al.* Human epidermal growth factor receptor 2 (HER2) immunoreactivity: specificity of three pharmacodiagnostic antibodies. *Histopathology* 2011;59:975–83.
- 40 Angerilli V, Galuppini F, Pagni F, *et al.* The role of the pathologist in the next-generation era of tumor molecular characterization. *Diagnostics (Basel)* 2021;11:339.
- 41 Xu K, Bayani J, Mallon E, *et al.* Discordance between immunohistochemistry and ERB-B2 receptor tyrosine kinase 2 mRNA to determine human epidermal growth factor receptor 2 low status for breast cancer. *J Mol Diagn* 2022;24:775–83.
- 42 Scott M, Vandenberghe ME, Scorer P, *et al.* Prevalence of HER2 low in breast cancer subtypes using the VENTANA anti-HER2/neu (4B5) assay. *JCO* 2021;39.(15_suppl)