Interobserver Variation in the Assessment of Immunohistochemistry Expression Levels in HER2-Negative Breast Cancer

Baez-Navarro, Ximena; van Bockstal, Mieke R.; Nawawi, Diënna; Broeckx, Glenn; Colpaert, Cecile; Doebar, Shusma C.; Hogenes, Marieke C.H.; Koop, Esther; Lambein, Kathleen; Peeters, Dieter J.E.

Published in:
Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc

DOI:
10.1016/j.modpat.2022.100009

Mod Pathol 36 (2023) 100009

MODERN PATHOLOGY

USCAP UNITED STATES AND CANADIAN ACADEMY OF PATHOLOGY
Creating a Better Pathologist

Journal homepage: https://modernpathology.org/

Research Article

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

Ximena Baez-Navarro[a,*], Mieke R. van Bockstal[b], Diënna Nawawi[a], Glenn Broeckx[c], Cecile Colpaert[d,e], Shusma C. Doebar[f], Marieke C.H. Hogenes[g], Esther Koop[h], Kathleen Lambein[i,j], Dieter J.E. Peeters[k,l], Renata H.J.A. Sinke[m], Johannes Bastiaan van Brakel[n], José van der Starre-Gaal[o], Bert van der Vegt[p], Koen van de Vijver[i,q], Celien P.H. Vreuls[r], Willem Vreuls[s], Pieter J. Westenend[t], Carolien H.M. van Deurzen[a]

[a] *Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands;* [b] *Department of Pathology, Cliniques Universitaires Saint-Luc, Brussels, Belgium;* [c] *Department of Pathology, Antwerp University Hospital, Edegem, Belgium;* [d] *Department of Pathology, General Hospital Turnhout, Turnhout, Belgium;* [e] *Department of Pathology, University Hospital Leuven, Leuven, Belgium;* [f] *Department of Pathology, Spaarne Gasthuis, Haarlem Zuid, The Netherlands;* [g] *Department of Pathology, Laboratory Pathology East Netherlands, Hengelo, The Netherlands;* [h] *Department of Pathology, Gelre Hospital, Apeldoorn, The Netherlands;* [i] *Department of Surgical Oncology, Leuven University Hospitals, Leuven, Belgium;* [j] *Department of Pathology, Ghent University Hospital, Ghent, Belgium;* [k] *Department of Pathology, Sint-Maarten General Hospital, Mechelen, Belgium;* [l] *Department of Pathology, CellCarta NV, Antwerp, Belgium;* [m] *Department of Pathology, Pathan B.V., Rotterdam, The Netherlands;* [n] *Department of Pathology, Skåne University Hospital, Malmö, Sweden;* [o] *Department of Pathology, Isala Clinics, Zwolle, The Netherlands;* [p] *Department of Pathology & Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands;* [q] *Division of Diagnostic Sciences, Cancer Research Institute Ghent, Ghent University, Ghent, Belgium;* [r] *Department of Pathology, Utrecht University Medical Center, Utrecht, The Netherlands;* [s] *Department of Pathology, CWZ Hospital, Nijmegen, The Netherlands;* [t] *Department of Pathology, PAL Laboratory of Pathology, Dordrecht, The Netherlands*

ARTICLE INFO

Article history:
Received 29 June 2022
Revised 3 August 2022
Accepted 16 September 2022

Keywords:
breast cancer
interobserver agreement
HER2 low

ABSTRACT

The classification of human epidermal growth factor receptor 2 (*HER2*) expression is optimized to detect *HER2*-amplified breast cancer (BC). However, novel *HER2*-targeting agents are also effective for BCs with low levels of *HER2*. This raises the question whether the current guidelines for *HER2* testing are sufficiently reproducible to identify *HER2*-low BC. The aim of this multicenter international study was to assess the interobserver agreement of specific *HER2* immunohistochemistry scores in cases with negative *HER2* results (0, 1+, or 2+/in situ hybridization negative) according to the current American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines. Furthermore, we evaluated whether the agreement improved by redefining immunohistochemistry (IHC) scoring criteria or by adding fluorescent in situ hybridization (FISH).

We conducted a 2-round study of 105 nonamplified BCs. During the first assessment, 16 pathologists used the latest version of the ASCO/CAP guidelines. After a consensus meeting, the same pathologists scored the same digital slides using modified IHC scoring criteria based on the 2007 ASCO/CAP guidelines, and an extra "ultralow" category was added.

Overall, the interobserver agreement was limited (4.7% of cases with 100% agreement) in the first round, but this was improved by clustering IHC categories. In the second round, the highest reproducibility was observed when comparing IHC 0 with the ultralow/1+/2+ grouped cluster

* Corresponding author.
E-mail address: x.baeznavarro@erasmusmc.nl (X. Baez-Navarro).

https://doi.org/10.1016/j.modpat.2022.100009

(74.3% of cases with 100% agreement). The FISH results were not statistically different between *HER2*-0 and *HER2*-low cases, regardless of the IHC criteria used.

In conclusion, our study suggests that the modified 2007 ASCO/CAP criteria were more reproducible in distinguishing *HER2*-0 from *HER2*-low cases than the 2018 ASCO/CAP criteria. However, the reproducibility was still moderate, which was not improved by adding FISH. This could lead to a suboptimal selection of patients eligible for novel *HER2*-targeting agents. If the threshold between *HER2* IHC 0 and 1+ is to be clinically actionable, there is a need for clearer, more reproducible IHC definitions, training, and/or development of more accurate methods to detect this subtle difference in protein expression levels.

Introduction

For more than 2 decades, overexpression of the human epidermal growth factor receptor 2 (*HER2*) has been recognized as a negative prognostic biomarker and therapeutic target in invasive breast cancer (BC).[1] International guidelines were developed to standardize and optimize *HER2* testing because only those patients with *HER2* amplification were likely to respond to *HER2*-targeted treatment.[2,3]

During the updates of the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines, considerable changes were made in the immunohistochemistry (IHC) cut-off points.[4–6] According to the first version of the guideline, published in 2007, BC was categorized as IHC 0 (no staining), IHC 1+ (weak, incomplete membrane staining in any proportion of tumor cells or weak complete staining in <10% of cells), IHC 2+ (equivocal), or IHC 3+ (uniform intense membrane staining of the *HER2* protein in >30% of invasive tumor cells). For treatment considerations, *HER2* was defined as positive in the case of IHC 2+ with amplification after reflex testing or IHC 3+.[5] In the updated versions of the guidelines, published in 2013 and 2018, the definitions of the IHC scoring were modified. The cut-off point for IHC 3+ was changed to complete, intense staining of >10% of the tumor cells instead of 30%. The definition of IHC 0 was adapted to either no staining or incomplete membrane staining that is faint/barely perceptible and within <10% of the invasive tumor cells.[4,6] This change resulted in a substantial increase of IHC 0 cases according to the 2013 version compared with the 2007 guidelines.[7,8]

In recent years, the development of an emerging group of *HER2*-targeted drugs, the so-called antibody-drug conjugates, has led to a different view on this historical *HER2* classification system. In the ongoing clinical trials, these drugs have demonstrated efficacy and safety against *HER2*-positive metastatic BC.[9,10] Moreover, because of their favorable drug-to-antibody ratio and bystander killing effect, antibody-drug conjugates such as trastuzumab-deruxtecan (T-DXd) have also proved to have significant antitumor action in BCs with a low expression level of *HER2* (*HER2* low), comprising IHC 1+ and IHC 2+ cases without amplification.[9–15] The Destiny-Breast06 clinical trial is currently evaluating the effect of T-DXd in BC with even lower levels of *HER2* expression (IHC >0, <1+), the so-called *HER2* ultralow category. These *HER2* ultralow cases would have been classified as IHC 1+ according to the ASCO/CAP guidelines of 2007 but as IHC 0 according to the 2013 and 2018 editions. Obviously, these novel treatment options raise the question whether our current method of IHC testing, historically optimized to detect *HER2*-amplified BC, is sufficiently robust and reproducible to discern *HER2*-low BC as well.

It is worth mentioning that the terms *HER2* low and ultralow have been used so far, mainly in clinical trials.[11,15] Because the clinical treatment relevance of threshold is currently untested, a nonzero result is still necessary for treatment with T-DXd. Thus, until the actual clinical threshold is known, it will not be possible to draw a clear definition of these terms, thus making it more difficult to be recognized by any guideline.

Previous studies have evaluated the *HER2* IHC interobserver reproducibility using the 2013 and 2018 versions of the ASCO/CAP guidelines with inconsistent results.[16–19] In a recent study by Fernandez et al,[19] data were collected from around 1400 laboratories worldwide. The lowest IHC agreement was found between *HER2* 0 versus *HER2* 1+ (<70% agreement). An interobserver analysis of 92 cases graded as IHC 0 or IHC 1 resulted in a 90% agreement (17 of 18 pathologists) in only 24 of 92 (26%) cases.[19]

In a study by Schettini et al,[17] 5 specialized observers evaluated 100 BC cases using the 2018 guidelines. Overall, 35 of 100 cases were discordant, among which the highest disagreement was found between IHC 1+ vs IHC 0 (n = 15).[17] Interestingly, older studies that used the cut-off points of the 2007 version of the ASCO/CAP guidelines seem to perform better in differentiating IHC 1+ from 0.[20–22] Umemura et al[20] reported a good general agreement in 14 of 20 cases evaluated by 7 observers. In their study, IHC 2+ and 3+ were the cases with low concordance (55%-64%), whereas cases with *HER2* 0 and 1+ showed a high concordance (90%-100%).[20] Additionally, Thomson et al[21] assessed 127 cases scored by 3 observers and reported a high interobserver agreement (kappa = 77-95.6) for IHC 0 and 3+ cases, whereas it was generally poor (kappa = 32.8-59.1) for cases with 1+ and 2+ staining.[21]

Thus, we hypothesized that the IHC scoring criteria according to the 2007 version of the ASCO/CAP recommendations for *HER2* testing are likely to be more reproducible, in particular to distinguish between IHC 0 and 1+, compared with the 2013 and 2018 criteria. The primary objective of this multicenter international study was to quantify the interobserver agreement of *HER2* low scoring according to the current guidelines and to evaluate whether we could improve this agreement by redefining some of the current IHC scoring criteria or by adding in situ hybridization.

Materials and Methods

Study Design and HER2 *Immunohistochemical Scoring*

We performed a multi-institutional study with 2 rounds of scoring, including 105 needle biopsies with invasive BC that were scored by 16 pathologists. These cases were a consecutive series of archived BC cases diagnosed in 2019 with a negative *HER2* status according to the original pathology report using the 2018 ASCO/CAP guidelines. Tissue sections were immunostained with the 4B5 HER2/neu antibody using an automatic immunostainer (Ventana BenchMark Ultra, Roche). All slides were scanned with the Nanozoomer 2.0-HT (Hamamatsu Photonics), which enabled

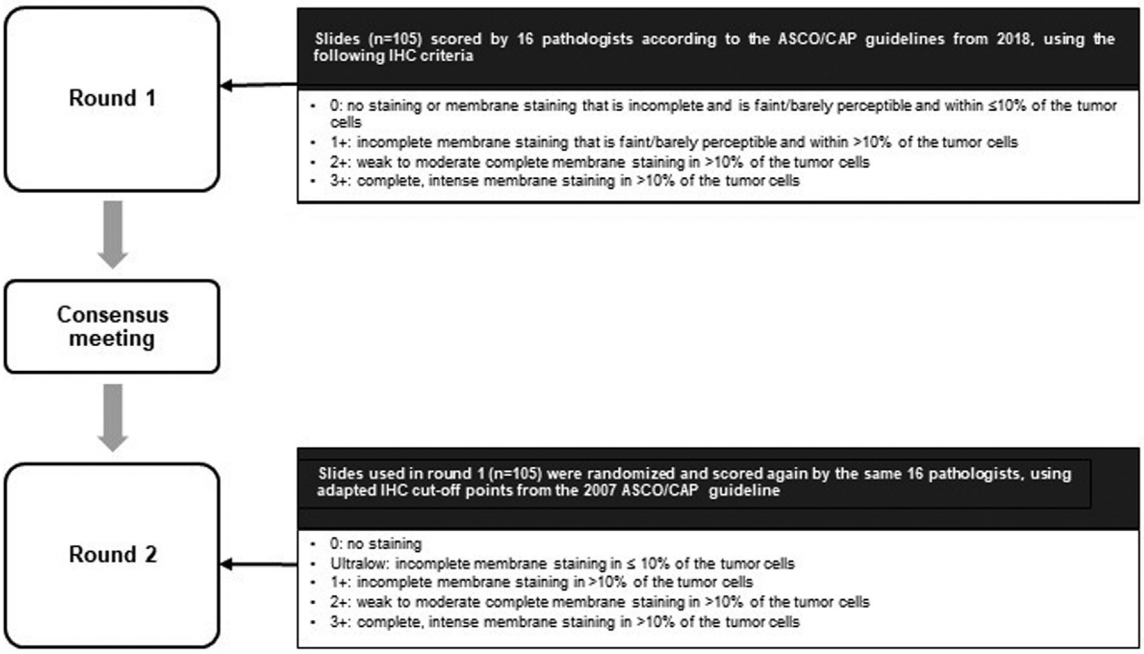

Figure 1.
Study design and immunohistochemistry (IHC) scoring criteria used for the first and the second scoring round. ASCO/CAP, American Society of Clinical Oncology/College of American Pathologists.

Z-stacking, and they were uploaded to Slide Score B.V. (version 1.2-2022-05-24T15:37:11) (Netherlands Cancer Institute), which allowed zooming to a high magnification (objective, ×40). This program blinded the case numbers and randomized the slides for the participants. The use of coded leftover patient material was in accordance with the code of conduct of the Federation of Medical Scientific Societies in The Netherlands.[23]

In the first round of scoring, 16 pathologists scored a total number of 105 slides according to the 2018 ASCO/CAP guidelines as IHC 0, 1+, 2+, 3+, or "too few tumor cells." Once the results from the first round were completed and analyzed, all pathologists participated in an online consensus meeting. This meeting included the following topics: presentation of the IHC criteria of the ASCO/CAP guidelines of 2007, 2013, and 2018; presentation and discussion of slides with good and poor agreement; and proposal of the criteria for use in the second round of scoring. Figure 1 provides an overview of the study design. For the second round of scoring, all pathologists scored the same slides again after randomization of the slide order in Slide Score. The IHC scoring criteria used in the second round of scoring were modified in accordance with the 2007 ASCO/CAP guidelines, supplemented with a separate category of *HER2* ultralow, as described in Figure 1. This resulted in the following categories: IHC 0, ultralow, 1+, 2+, 3+, or "too few tumor cells." For this second round, a document describing the new criteria and examples of each IHC category was sent to the pathologists.

HER2 *In Situ Hybridization*

To complement this study, fluorescence in situ hybridization (FISH) was performed on all 105 cases using the BenchMark Ultra (Roche). For the detection of *HER2*, the ZytoLight SPEC ERBB2/CEN 17 Dual Color Probe (Zytovision) was used according to the manufacturer's protocol. Signal numbers for the chromosomal region 17q12-q21.1 harboring the *HER2* gene (labeled with SPEC ERBB2, ZytoLight) and the alpha satellite centromeric region of chromosome 17 (*CEP17*; labeled with CEN 17, ZytoLight) were counted in at least 30 invasive tumor cells, and the ratio of *HER2* to *CEP17* signal numbers was calculated. The analysis of the FISH tests was performed by 1 observer. A second observer checked and approved the interpretation of the tests.

Statistical Analysis

The scoring option of "too few tumor cells" was considered missing data and excluded from the statistical analysis. Krippendorff alpha test was used to estimate the interobserver agreement in rounds 1 and 2.[24] The cut-off points of interobserver reliability for this test were alpha value of <0.67 (low), between 0.67 and 0.8 (moderate), and >0.8 (high).

To analyze the correlation between the IHC scores and FISH results (*HER2*/*CEP17* ratio and the average number of *HER2* copies/cell), we used the IHC score that was the most frequently scored by the pathologists in the first and second round. We calculated the mean, median, and range for continuous variables. The Shapiro-Wilk test was used to check the normal distribution. Kruskal-Wallis and Mann-Whitney *U* tests were used to study the association between the FISH results and IHC score, as these data were not normally distributed. A *P* value of <.05 was considered significant, except for the post hoc Mann-Whitney *U* tests following the Kruskal-Wallis tests, where a *P* value of <.016 was used (ie, Bonferroni correction for multiple testing).

Statistical analyses were performed in SPSS (IBM Corp; version 28.0, released 2021.). Additionally, a macro was downloaded from http://afhayes.com/spss-sas-and-r-macros-and-code.html to perform the Krippendorff alpha test in SPSS.

IHC combinations					K-alpha	100% (16/16 pathologists) agreement	87.5% (14/16 pathologists) agreement
Round 1							
0 vs 1+ vs 2+ vs 3+	0	1+	2+	3+	α= 0.63	4.7% (5 of 105)	30.4% (32 of 105)
0 vs 1+ and 2+	0	1+ and 2+		3+	α= 0.56	33.3% (35 of 105)	76.2% (80 of 105)
Round 2							
0 vs ultralow vs 1+ vs 2+	0	ultra low	1+	2+	α=0.32	9.5% (11 of 105)	33.9% (39 of 105)
0 vs ultralow and 1+ vs 2+	0	ultralow and 1		2+	α= 0.71	26% (30 of 105)	53.9% (62 of 105)
0 vs ultralow, 1+, and 2+	0	ultralow, 1+ and 2+			α= 0.68	74.3% (78 of 105)	80% (84 of 105)
0 and ultralow vs 1+ vs 2+	0 and ultralow		1+	2+	α= 0.73	20.9% (22 of 105)	50.4% (53 of 105)
0 and ultralow vs 1+ and 2+	0 and ultralow		1+ and 2+		α= 0.65	52.4% (55 of 105)	74.3% (78 of 105)

Figure 2.
Krippendorff alpha values and percentage of agreement per IHC combination in the first and second round. IHC, immunohistochemistry.

Results

Interobserver Agreement First Round

In both scoring rounds, each of the 16 pathologists evaluated 105 BC cases. The results of the first round are presented in Supplementary Table S1. A total number of 6 cases were scored as "too few tumor cells" and were considered missing data, resulting in 1674 scores. Of these, 383 (22.8%) were scored as IHC 0, 749 (44.6%) were scored as IHC 1+, 526 (31.3%) were scored as IHC 2+, and 16 (1%) were scored as IHC 3+. The Krippendorff alpha for estimating the interobserver agreement from the first scoring round showed low agreement for the categories IHC 0, 1+, 2+, and 3+ ($\alpha = 0.63$). A consistently low agreement ($\alpha = 0.56$) was found when we grouped the categories 1+ and 2+ together.

Furthermore, the percentage of complete agreement, where all 16 pathologists grouped the tumors into the same IHC category (0, 1+, 2+, and 3+), was achieved in only 5 of 105 (4.7%) cases. If we considered an agreement of 87.5% as an acceptable consensus (14 of 16 pathologists), 30.4% (32 of 105) of the cases were grouped into the same category. Figure 2 presents the level of agreement using different combinations of IHC clusters. After clustering IHC 1+ and IHC 2+ together, the percentage of agreement increased to 33.3% (35 of 105) for a complete agreement (all pathologists) and 76.2% (80 of 105) for an agreement among 14 of 16 pathologists, which highlights that the distinction between 1+ and 2+ cases is problematic.

Consensus Meeting

During the online consensus meeting after the first round of scoring, a representative subset of cases with good and poor interobserver agreement was discussed. Figure 3A, B represents 2 examples of cases with low agreement. Several issues caused doubt when distinguishing between IHC 0 and 1+, including difficulty in discriminating nonspecific staining from membranous staining. In addition, the application of the term "barely perceptible" was considered highly subjective by most participants. Another difficulty involved the correct estimation of 10% of tumor cells. The use and resolution of digital slides were also considered an obstacle by some pathologists. Moreover, several pathologists were not used to evaluate the 4B5 *HER2* antibody because they used another antibody in their laboratory. Finally, a small subset of cases was scored as 3+ in the first round because some pathologists did not realize that only *HER2*-negative cases were included. Considering all the aforementioned difficulties, we performed a second round using modified criteria based on the 2007 ASCO/CAP guidelines (Fig. 1), as described in the Materials and Methods section.

Interobserver Variation Second Round

The results of the second round are presented in Supplementary Table S2. Three cases were scored as "too few tumor cells," resulting in 1677 scores. Of these, 275 (16.4%) were

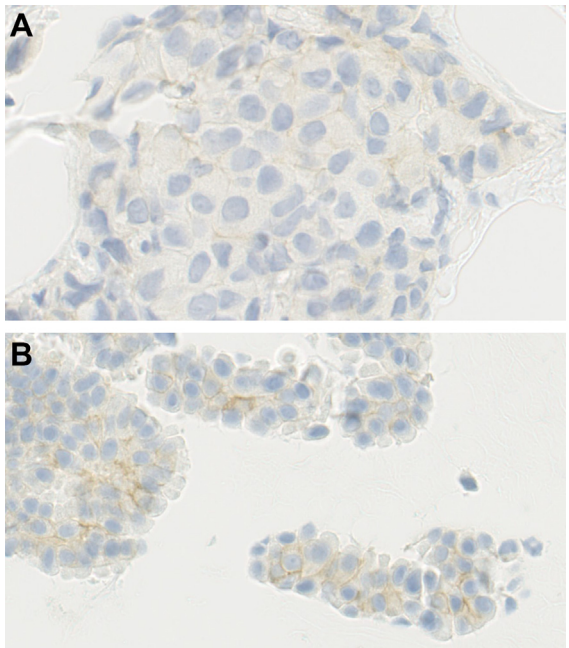

Figure 3.
Examples of cases with a low agreement in the first round (amplification, ×80). (A) Seven pathologists scored immunohistochemistry (IHC) 0 and 9 scored IHC 1+. (B) Nine pathologists scored IHC 0 and 7 scored IHC 1+.

scored as IHC 0, 259 (15.4%) were scored as ultralow, 715 (42.6%) were scored as IHC 1+, and 428 (25.5%) were scored as IHC 2+. No cases were scored as IHC 3+. By analyzing all IHC categories separately, the Krippendorff alpha showed poor agreement ($\alpha = 0.32$). The agreement was moderate after clustering other IHC groups together ($\alpha = 0.65$-0.73), and the percentage of agreement remained poor to moderate (20.9%-80%; Fig. 2). Noticeably, the highest percentage of agreement was achieved when we clustered ultralow, IHC 1+, and IHC 2+ together and compared it with IHC 0; the complete agreement (all pathologists) was 74.3% (78/105), and an agreement among 14 of 16 pathologists was 80% (84/105). Figure 4 presents 2 examples of cases with low agreement in the second round, likely because of difficulty distinguishing between nonspecific staining versus true membrane staining (Fig. 4A) and/or the use of the 10% cut-off point (Fig. 4B).

Fluorescence In Situ Hybridization Analysis and Differences Between Immunohistochemistry Categories

The FISH assay was performed for all 105 BC cases. The *HER2*/*CEP17* ratios and average number of *HER2* copies/cell were non-normally distributed. Overall, the median *HER2*/*CEP17* ratio and mean *HER2* copy number were 1.15 (range, 0.43-4.19) and 1.79 (range, 1.07-4.23), respectively.

The FISH results, according to the IHC consensus of round 2, are presented in Figure 5. For the analysis of the mean *HER2* copy number according to the IHC scores, an overall statistically significant difference was found ($P < .001$) when *HER2* 0 and ultralow scores were grouped together versus *HER2* 1+ and 2+ scores (Fig. 5A). In the post hoc analyses, no significant difference was found in the *HER2* copy number between *HER2* 0 and ultralow and 1+ ($P = .141$), whereas significant differences were found between *HER2* 0 and ultralow and *HER2* 2+ ($P < .001$) and between *HER2* 1+ and 2+ ($P < .001$).

In the second analysis, we grouped *HER2* ultralow and 1+ together (Fig. 5B). We also found a significant difference between the ranks of *HER2* copy number ($P < .001$) when comparing ultralow and *HER2* 1+ together vs *HER2* 0 and 2+ scores. In the post hoc analyses, no significant difference in the mean *HER2* copy number was found between *HER2* 0 and ultralow and 1+ ($P = .149$). Significant differences in the mean *HER2* copy number were observed between *HER2* ultralow and 1+ and 2+ ($P < .001$), and between *HER2* 0 and 2+ ($P < .001$). Regarding the *HER2*/*CEP17* ratios, no significant differences were found between the different IHC categories when clustering 0 and ultralow together ($P = .315$) or grouping ultralow and 1+ together ($P = .7$). These results were consistent when performing the same analyses using the results of the consensus IHC scores of the first scoring round (data not shown). Therefore, the FISH results seem to have limited additional value to differentiate between the *HER2* 0 and *HER2*-(ultra)low cases.

Discussion

The precision of *HER2* IHC scoring is essential in selecting patients with BC for existing and novel *HER2*-targeting agents. Because T-DXd showed effectiveness in patients with *HER2*-low tumors (IHC 1+ or 2+/in situ hybridization negative), the necessity to distinguish between 0 and *HER2* 1+ expression has become clinically relevant, although the clinical validity of IHC 0 vs 1+ threshold and treatment response remains currently untested.[11,15] In this study, there was a decrease in the number of cases scored as *HER2* 0 from the first (22.8%) to the second (16.3%) round, which is in line with the findings

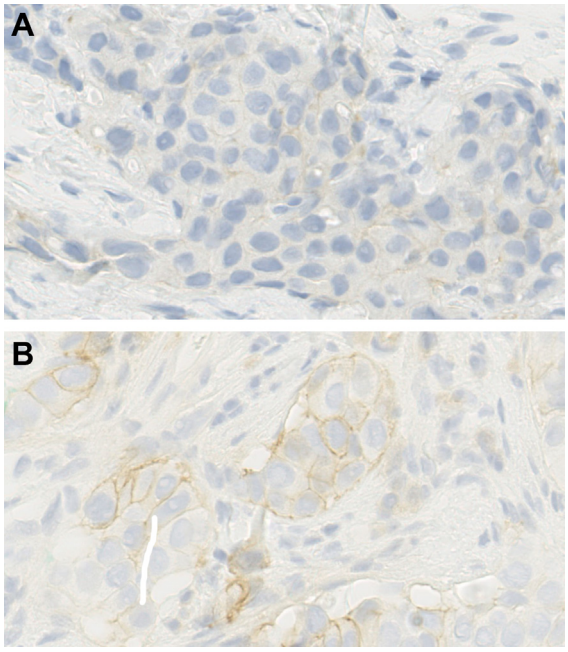

Figure 4.
Examples of cases with a low agreement in the second round (amplification, ×80). (A) Seven pathologists scored immunohistochemistry (IHC) 0 and 9 scored IHC ultralow. (B) Six pathologists scored IHC ultralow and 10 scored IHC 1+.

reported in the literature.[7,22,25] Noticeably, the proportion of IHC 2+ cases was also higher (31.3%) with the current criteria compared with the 2007 criteria (25.5%), as described previously.[7]

The present study suggests that the interobserver agreement for IHC interpretation could be improved by adapting the criteria for immunohistochemical assessment. The best (or, rather, least poor) reproducibility was seen in the second scoring round when comparing IHC 0 with the cluster of ultralow/1+ and 2+. This supports the hypothesis that identifying *HER2* 0 cases according to the 2007 guidelines (complete lack of *HER2* expression or the "all or nothing" principle) is easier than using a 10% cut-off point according to the 2018 guidelines. In addition, the use of "barely perceptible," as described in the 2018 guidelines, is perceived as subjective by most pathologists participating in this study. Difficulties in distinguishing the nonspecific background or cytoplasmic staining from true membrane staining could explain the discrepancies between IHC 0 and ultralow categories. These results are in line with the findings of the previous studies evaluating the interobserver agreement for each of these ASCO/CAP guidelines.[16,20,21] Our results support the idea that the current ASCO/CAP IHC criteria, which are optimized to select *HER2* amplified cases, are not very robust in distinguishing cases without from those with low levels of *HER2* expression.

In this study, we demonstrated that the reproducibility between *HER2* ultralow and 1+ is moderate, which also illustrates the difficulty of using the 10% cut-off point. The clinical relevance of *HER2* ultralow tumors is uncertain. A recent study by Diéras et al[26] demonstrated that T-DXd was effective even in *HER2* 0 tumors, but it is unclear whether these cases were completely negative or had protein expression levels that were less than 1+ or "ultralow." Future research should focus on whether the *HER2* IHC 0 group contains patients who can benefit from T-DXd and if novel methods to distinguish the subtle differences between zero and very low levels of protein expression can reproducibly and accurately identify these potential treatment responders. Some studies

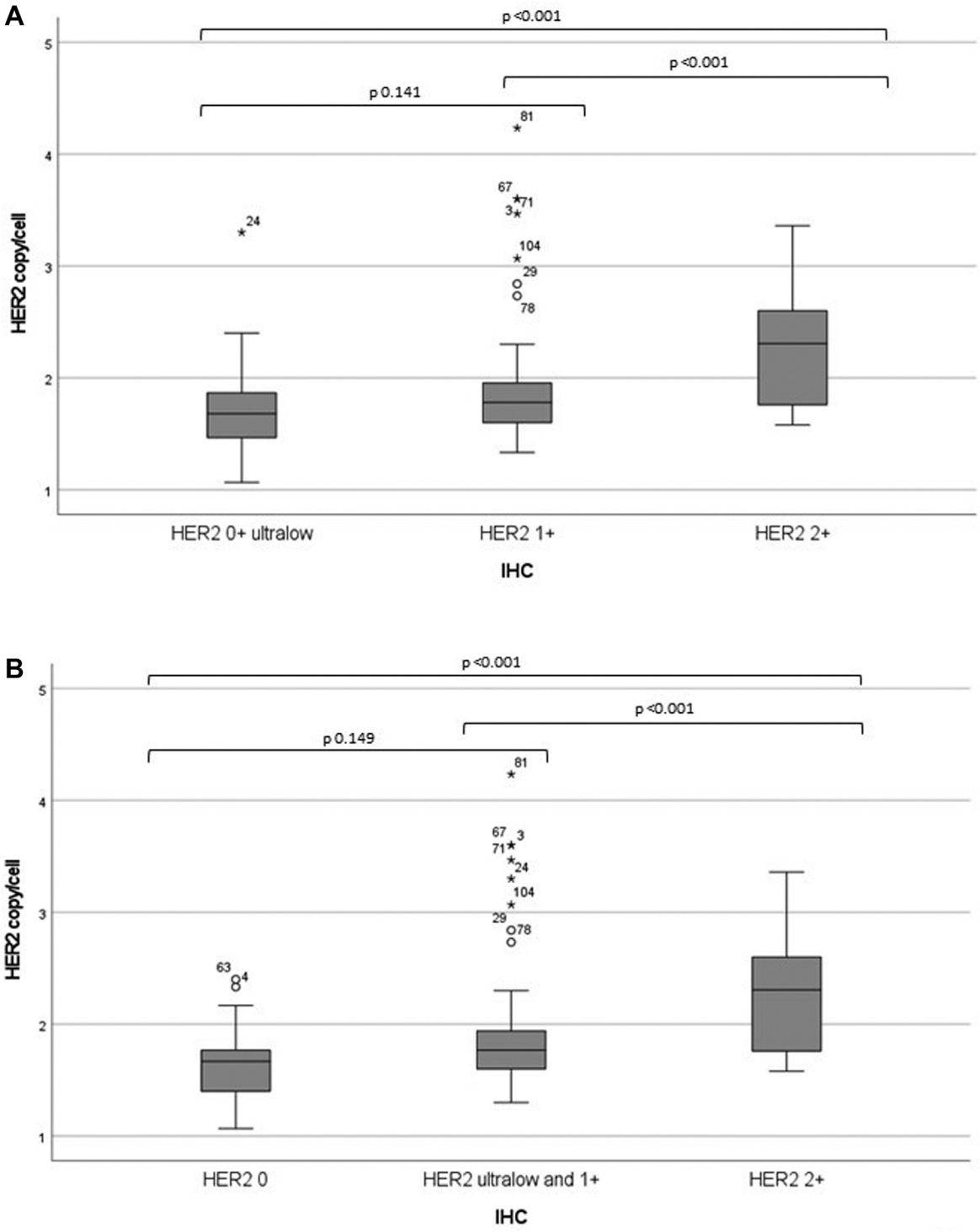

Figure 5.
Boxplot of fluorescent in situ hybridization results according to the consensus immunohistochemistry (IHC) categories of round 2 by clustering human epidermal growth factor receptor 2 (*HER2)* ultralow with IHC 0 (A) or by clustering *HER2* ultralow with IHC 1+ (B).

have already developed artificial intelligence algorithms and advanced techniques of targeted mass spectrometry for this purpose; however, the studies in this area are still limited and require validation in larger, independent cohorts.[27–30] Additionally, continuous training of practicing pathologists seems important, as highlighted by the high degree of variability in our study.

An additional aim of our study was to assess whether the FISH data and mean *HER2* copy numbers, in particular, could improve the discrimination between *HER2* 0 and *HER2*-(ultra)low tumors. However, we did not observe any statistical differences in the mean *HER2* copy number between IHC 0 and 1+, regardless of the inclusion of the ultralow group. Therefore, using FISH as a

companion test for IHC does not seem to have additional value to separate *HER2* 0 from *HER2*-low tumors. These data are in contrast with those reported in the previous studies,[22,31] which found that tumors with an IHC score of 0 had a lower *HER2* copy number than those with an IHC score of 1+. In our study, we did not observe this effect. Although the *P* value was insignificant, the difference in the mean rank (Md) was higher in the ultralow and 1+ group (Md = 41.45, $P = .149$), compared with the ultralow with 0 (Md = 34.31, $P = .141$). The lack of statistical significance between the IHC 0 and 1+ groups in this study could be due to the smaller sample size compared to other studies or due to tumor heterogeneity.[22,31,32]

To our knowledge, this is the first *HER2* interobserver study focusing on the novel subgroup of *HER2*-low BC that estimated interobserver agreement before and after a consensus meeting and adjusting IHC criteria. In addition, we also assessed the association between IHC and FISH data. In this study, we used the 4B5 HER2/neu clone on the Ventana platform (Roche), which is also used in the DESTINY-Breast trials. Another strength of this study is the large number of specialized pathologists participating in both rounds. Furthermore, by including only *HER2*-negative cases, we prevented a false improvement in the agreement, because the previous reports already demonstrated that the assessment of IHC 3+ cases shows a better concordance. Our study also has some limitations. First, not all pathologists were used to score digital slides in daily practice, which could have affected our results. Second, we used needle biopsies for the primary tumor, while in the clinical setting, *HER2* status of the metastases is also relevant. However, because the scoring criteria of metastases do not differ from primary tumors, this is less likely to have affected our results.

The present study suggests that the 2007 ASCO/CAP criteria were more reproducible in distinguishing *HER2* 0 cases than the 2018 ASCO/CAP criteria. However, the reproducibility is still only moderate, and performing FISH does not provide additional support to discriminate between very low levels of *HER2* expression. This could lead to the suboptimal selection of patients who could benefit from novel treatment options like T-DXd. Our results reinforce the need to develop clearer, more reproducible definitions for IHC scoring and training of pathologists to diagnose *HER2*-low BC, adapted to the results of ongoing clinical trials. Future research in this field should also focus on the development of novel and more accurate methods to quantify the level of *HER2* expression so that these methods can be tested for their potential clinical relevance in selecting patients for *HER2*-targeted drug delivery.

Acknowledgments

The authors thank Hein Sleddens and Ludo Uytdewilligen for performing and scoring the FISH assays and J. Hudeck for his help with using Slide Score.

Author Contributions

C.H.M.v.D performed the study concept and design. X.B.-N. drafted the manuscript. C.H.M.v.D X.B.-N., M.R.v.B., and D.N. developed the methodology and reviewed and revised the paper. G.B., C.C., S.C.D., M.C.H.H., E.K., K.L., D.J.E.P., R.H.J.A.S., M.R.v.B., J.B.v.B., J.v.d.S.-G., B.v.d.V., K.v.d.V., C.P.H.V., W.V., P.J.W. performed the slides score and participated in the consensus meeting. X.B.-N. performed the data interpretation and statistical analysis. C.H.M.v.D provided the tissue material. All authors read and approved the final paper.

Data Availability Statement

The data sets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

This work was funded by AstraZeneca. The funder had no role in the design of the study, the interpretation of the results, or the writing of the article.

Declaration of Competing Interest

C.H.M.v.D. was on the advisory board of AstraZeneca/Daiichi Sankyo and received research funding from Roche and AstraZeneca. B.v.d.V. reported honoraria received by University Medical Center Groningen for expertise or scientific advisory board/consultancy (on request) from Visiopharm, Philips, MSD/Merck, and Daiichi Sankyo/AstraZeneca and speaker's fee from Visiopharm, Diaceutics, and MSD/Merck, all unrelated to the current work.

Ethics Approval and Consent to Participate

According to the Code of Conduct of the Federation of Medical Sciences in the Netherlands, no informed consent or ethical approval was needed for this study. This study was performed in accordance with the Declaration of Helsinki.

Supplementary Material

The online version contains supplementary material available at https://doi.org/10.1016/j.modpat.2022.100009

References

1. Tsang JYS, Tse GM. Molecular classification of breast cancer. *Adv Anat Pathol*. 2020;27(1):27–35.
2. Early Breast Cancer Trialists' Collaborative group (EBCTCG). Trastuzumab for early-stage, HER2-positive breast cancer: a meta-analysis of 13 864 women in seven randomised trials. *Lancet Oncol*. 2021;22(8):1139–1150.
3. Slamon D, Eiermann W, Robert N, et al. Adjuvant trastuzumab in HER2-positive breast cancer. *N Engl J Med*. 2011;365(14):1273–1283.
4. 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(11):1364–1382.
5. Wolff AC, Hammond ME, Schwartz JN, 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):118–145.
6. Wolff AC, Hammond ME, 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. *J Clin Oncol*. 2013;31(31):3997–4013.
7. Lambein K, Van Bockstal M, Denys H, Libbrecht L. 2013 update of the American Society of Clinical Oncology/College of American Pathologists guideline for human epidermal growth factor receptor 2 testing: impact on immunohistochemistry-negative breast cancers. *J Clin Oncol*. 2014;32(17):1856–1857.
8. Hanna WM, Slodkowska E, Lu FI, Nafisi H, Nofech-Mozes S. Comparative analysis of human epidermal growth factor receptor 2 testing in breast cancer according to 2007 and 2013 American Society of Clinical Oncology/College of American Pathologists guideline recommendations. *J Clin Oncol*. 2017;35(26):3039–3045.
9. Modi S, Saura C, Yamashita T, et al. Trastuzumab deruxtecan in previously treated HER2-positive breast cancer. *N Engl J Med*. 2020;382(7):610–621.
10. Tamura K, Tsurutani J, Takahashi S, et al. Trastuzumab deruxtecan (DS-8201a) in patients with advanced HER2-positive breast cancer previously treated with trastuzumab emtansine: a dose-expansion, phase 1 study. *Lancet Oncol*. 2019;20(6):816–826.

11. Modi S, Park H, Murthy RK, et al. Antitumor activity and safety of trastuzumab deruxtecan in patients with HER2-low-expressing advanced breast cancer: results from a phase Ib Study. *J Clin Oncol*. 2020;38(17):1887–1896.
12. AztraZeneca. Enhertu significantly improved both progression-free and overall survival in DESTINY-Breast04 trial in patients with HER2-low metastatic breast cancer. *AztraZeneca*. February 21, 2022. Accessed April 5, 2022. https://www.astrazeneca.com/media-centre/press-releases/2022/enhertu-improves-pfs-and-os-in-her2-low-bc.html#:~:text=Positive%20high%2Dlevel%20results%20from,low%20unresectable%20and%2For%20metastatic
13. Eiger D, Agostinetto E, Saúde-Conde R, de Azambuja E. The exciting new field of HER2-low breast cancer treatment. *Cancers (Basel)*. 2021;13(5):1015.
14. 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(8):1124–1135.
15. Modi S, Jacot W, Yamashita T, et al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. *N Engl J Med*. 2022;387(1):9–20.
16. Casterá C, Bernet L. HER2 immunohistochemistry inter-observer reproducibility in 205 cases of invasive breast carcinoma additionally tested by ISH. *Ann Diagn Pathol*. 2020;45:151451.
17. 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):1.
18. Md Pauzi SH, Masir N, Yahaya A, et al. HER2 testing by immunohistochemistry in breast cancer: a multicenter proficiency ring study. *Indian J Pathol Microbiol*. 2021;64(4):677–682.
19. Fernandez AI, Liu M, Bellizzi A, et al. Examination of low ERBB2 protein expression in breast cancer tissue. *JAMA Oncol*. 2022;8(4):1–4.
20. Umemura S, Osamura RY, Akiyama F, 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;130(6):883–891.
21. Thomson TA, Hayes MM, Spinelli JJ, 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(11):1079–1086.
22. Lambein K, Van Bockstal M, Vandemaele L, et al. Distinguishing score 0 from score 1+ in HER2 immunohistochemistry-negative breast cancer: clinical and pathobiological relevance. *Am J Clin Pathol*. 2013;140(4):561–566.
23. Human tissue and medical research: code of conduct for responsible use. Federa FMWV. Accessed April 15, 2022. https://www.bbmri.nl/sites/bbmri/files/styles/Federa_code_of_conduct_english.pdf
24. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Meas*. 2007;1(1):77–89.
25. D'Alfonso T, Liu YF, Monni S, Rosen PP, Shin SJ. Accurately assessing her-2/neu status in needle core biopsies of breast cancer patients in the era of neoadjuvant therapy: emerging questions and considerations addressed. *Am J Surg Pathol*. 2010;34(4):575–581.
26. Diéras V, Deluche E, Lusque A, et al. Trastuzumab deruxtecan (T-DXd) for advanced breast cancer patients (ABC), regardless HER2 status: a phase II study with biomarkers analysis (DAISY). Paper presented at: San Antonio Breast Cancer Symposium, 7-10, December, 2021; 2022; San Antonio, TX.
27. Kennedy JJ, Whiteaker JR, Kennedy LC, et al. Quantification of human epidermal growth factor receptor 2 by immunopeptide enrichment and targeted mass spectrometry in formalin-fixed paraffin-embedded and frozen breast cancer tissues. *Clin Chem*. 2021;67(7):1008–1018.
28. Yue M, Zhang J, Wang X, et al. Can AI-assisted microscope facilitate breast HER2 interpretation? A multi-institutional ring study. *Virchows Arch*. 2021;479(3):443–449.
29. Qaiser T, Mukherjee A, Reddy Pb C, et al. HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology*. 2018;72(2):227–238.
30. Tewary S, Mukhopadhyay S. HER2 molecular marker scoring using transfer learning and decision level fusion. *J Digit Imaging*. 2021;34(3):667–677.
31. Lambein K, Praet M, Forsyth R, et al. Relationship between pathological features, HER2 protein expression and HER2 and CEP17 copy number in breast cancer: biological and methodological considerations. *J Clin Pathol*. 2011;64(3):200–207.
32. Marchiò C, Annaratone L, Marques A, Casorzo L, Berrino E, Sapino A. Evolving concepts in HER2 evaluation in breast cancer: heterogeneity, HER2-low carcinomas and beyond. *Semin Cancer Biol*. 2021;72:123–135.