Xue et al. Breast Cancer Research (2024) 26:128
https://doi.org/10.1186/s13058-024-01884-9

Breast Cancer Research

RESEARCH Open Access

Challenges and improvements in HER2 scoring and histologic evaluation: insights from a national proficiency testing scheme for breast cancer diagnosis in China

Xuemin Xue[1†], Lei Guo[1†], Changyuan Guo[1†], Liwei Xu[2†], Lin Li[1], Lin Yang[1], Xin Wang[1], Wei Rao[1], Pei Yuan[1], Jiali Mu[1], Jiangtao Li[1], Bingning Wang[1], Quan Zhou[1], Weicheng Xue[3], Fei Ma[4*], Wenjing Yang[5*] and Jianming Ying[1*]

Abstract

Background In 2022, our team launched the pioneering national proficiency testing (PT) scheme for the pathological diagnosis of breast cancer, rapidly establishing its credibility throughout China. Aiming to continuously monitor and improve the proficiency of Chinese pathologists in breast pathology, the second round of the PT scheme was initiated in 2023, which will expand the number of participating institutions, and will conduct a nationwide investigation into the interpretation of HER2 0, 1+, and 2+/FISH- categories in China.

Methods The methodology employed in the current round of PT scheme closely mirrors that of the preceding cycle in 2022, which is designed and implemented according to the "Conformity assessment—General requirements for proficiency testing"(GB/T27043—2012/ISO/IEC 17043:2010). More importantly, we utilized a statistics-based method to generate assigned values to enhance their robustness and credibility.

Results The final PT results, published on the website of the National Quality Control Center for Cancer (http://117.133.40.88:3927), showed that all participants passed the testing. However, a few institutions demonstrated systemic biases in scoring HER2 0, 1+, and 2+/FISH- with accuracy levels below 59%, considered unsatisfactory. Especially, the concordance rate for HER2 0 cases was only 78.1%, indicating challenges in distinguishing HER2 0 from low HER2 expression. Meanwhile, areas for histologic type and grade interpretation improvement were also noted.

Conclusions Our PT scheme demonstrated high proficiency in diagnosing breast cancer in China. But it also identified systemic biases in scoring HER2 0, 1+, and 2+/FISH- at some institutions. More importantly, our study

†Xuemin Xue, Lei Guo, Changyuan Guo and Liwei Xu contributed equally to this work.

*Correspondence:
Fei Ma
mcmf76@gmail.com
Wenjing Yang
wenjing86824@126.com
Jianming Ying
jmying@cicams.ac.cn

Full list of author information is available at the end of the article

highlighted challenges in the evaluation at the extreme lower end of the HER2 staining spectrum, a crucial area for further research. Meanwhile, it also revealed the need for improvements in interpreting histologic types and grades. These findings strengthened the importance of robust quality assurance mechanisms, like the nationwide PT scheme conducted in this study, to maintain high diagnostic standards and identify areas requiring further training and enhancement.

Keywords Breast pathology, Proficiency testing, China, HER2, Immunohistochemistry

Background

Breast cancer represents the most prevalent cancer among women worldwide and was the fifth leading cause of cancer-related deaths in 2020 [1]. In China, the incidence of breast cancer is rising, and survival outcomes remain inferior to those in the United States, underscoring a critical need for improved cancer care strategies [2, 3]. There are numerous factors significantly impact the management of cancer, among which the quality of pathology diagnosis is paramount. In order to conduct a baseline assessment of six key pathological items (histologic type, grade, ER, PR, HER2,and Ki67), as previously recommended [4–9], and to identify critical gaps in the application of standardized diagnosis guidelines/recommendations for breast cancer in China, the National Cancer Center/National Quality Control Center for Cancer (NCC/NQCCC) together with the China National Accreditation Service for Conformity Assessment (CNAS) launched the first national proficiency testing (PT) program for pathology diagnosis in 2022, which has rapidly established its credibility throughout the country [10].

Leveraging this foundational success, in 2023, the second round of the PT scheme was initiated to gain a more comprehensive understanding of pathologists' performance across China. This round of PT program planned to expand the number of participating institutions, and to include a nationwide investigation into the interpretation of HER2 0, 1+and 2+/FISH- categories in China. In addition, a statistics-based method was utilized to generate assigned values, aiming to enhance their robustness and credibility. Therefore, this round of PT scheme is essential for continuously monitoring and improving the proficiency of Chinese pathologists in breast pathology.

Methods

Program design and implement

This is the second round of nationwide PT program for pathological diagnosis in China, which was implemented by the NCC/NQCCC based on the "Conformity assessment—General requirements for proficiency testing" (GB/T27043—2012/ISO/IEC 17043:2010). The PT provider is the pathology department of the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, which has achieved accreditation (No. CNAS PT0128) by the CNAS. Additionally, this program included an expert panel consisting of three consultant pathologists specializing in breast pathology, named Q.Z., W.X., and Y.J., who are affiliated with various tertiary grade A hospitals.

The methodology employed in the current round of PT scheme closely mirrors that of the preceding cycle, including test samples selection, preparation, distribution, and report collection [10]. To ensure the uniformity of the pathology images, all digital copies were derived from identical digital sources, and the USB flash drives were procured from the same model and batch. Finally, a total of 173 USB flash drives were distributed simultaneously to 173 participating institutions via express delivery. Each drive contained the following files: (1) digital slides, (2) digital slide viewer software (KF-Viewer), (3) an operation guide, and (4) a confirmation form for receiving status. Upon confirmation of their receipt of intact USB drives containing the correct digital files, the participating institutions will report their evaluation results through an online reporting system specifically developed for our PT program.

As suggested by the expert panel, five cases of breast cancer (H&E, ER, PR, HER2 and Ki67) and fifteen cases of HER2 slides (H&E and HER2-IHC) were selected for testing. The histologic type of selected cases included (1) invasive breast carcinoma of no special type (IBC-NST), (2) invasive lobular carcinoma (ILC), (3) invasive micropapillary carcinoma (IMC), and (4) carcinoma with apocrine differentiation (C-AD). Antibodies against HER2 (clone 4B5), ER (clone SP1) and PR (clone 1E2) were purchased from Ventana/Roche; Antibodies against Ki67 (clone MIB-1) was purchased from Genetech (Shanghai, China). All immunohistochemistry (IHC) stainings were performed according to the manufacturer's instructions on an automated Ventana Benchmark ULTRA system. In addition, some essential immunohistochemical markers required for diagnosis, such as E-cadherin, were also provided to the participants.

It should be noted that the purpose of this national PT program is to assess the capacities of different institutions, rather than the individual pathologist skills within their respective institutions. However, the results submitted by each institution may represent a consensus among multiple pathologists within their own institutions.

Statistic-based assigned values generation

To determine the assigned values (gold standards) for testing, a two-stage process was implemented based on a full discussion and related guidelines of breast pathology [5–7, 11].

In the first stage, five breast cancer cases and fifteen HER2 IHC slides were evaluated by 22 attending surgical pathologists in our department. The assigned values for histologic type, histologic grade, and HER2-IHC score were generated based on the evaluation results that achieved a consensus of 70% or greater. If consensus did not exceed 70%, the two most frequent evaluation outcomes were both selected as assigned values. For example, the HER2 IHC score for case A1 resulted in 59% for HER2 2+and 41% for HER2 3+, thus, the assigned value was determined to be HER2 2+/3+. For ER, three distinct categories (<1%, 1 ~ 10%, and >10%) were delineated, and the assigned value range was established according to the category in which the median value of the results fell. For PR, two categories (<1% and ≥1%) were defined, and the similar approach was utilized to determine the assigned value range. For the Ki67 index, the assigned values were determined by employing the median value of the results, with the range of minimal and maximal values.

In the second stage, all pathological slides along with the statistically-derived assigned values mentioned above were submitted to the expert panel for confirmation and approval.

Data statistics

In our study, a Weibull distribution was employed to describe the probability of overall PT scores and HER2 IHC accuracy among participating institutions, as described by our previous study [10]. Herein, "fitdistr" analysis in "MASS" package to calculate the scale and shape values of Weibull distribution. "ggpubr" and "ggplot2" packages were used to plot data. Meanwhile, the final PT score of each participant was considered passed if it was between 70 and 80, good if it was between 80 and 90, and excellent if it was over 90.

All statistical analyses were performed in R (version 3.6.0, https://www.r-project.org/). The final PT results were published on the website of the NQCCC (http://117.133.40.88:3927) (Additional file S1).

Results

Significant expansion in participants from 2022 to 2023

Initially, 173 institutions signed up to participate in this national PT program. However, we finally collected valid results from 169 institutions across 30 provinces/municipalities/autonomous regions in China (Fig. 1a). It demonstrated an approximate 76% increase, from 96 participant institutions in 2022 to a striking 169 in 2023. Moreover, the geographic distribution of our PT scheme expanded from 26 to 30 provinces/municipalities/autonomous regions across China. Particularly, Guangdong province shows a dramatic increase from 4 to 19 participant institutions, matching Jiangsu province, that also increased to 19 participants in 2023 (Fig. 1b). Meanwhile, among the 169 institutions, 84% of them were general hospitals, and 15% were specialized hospitals, and the remaining one (1%) was an independent pathology center (Fig. 1c). Furthermore, 93% of the total participating institutions were tertiary grade A hospitals which represent the highest level of healthcare facilities in China, offering comprehensive and specialized medical services, advanced teaching, and research capabilities (Fig. 1d). This broader engagement affords a more comprehensive assessment of pathologists' diagnostic proficiencies in China.

Statistic-based assigned values generation

In this round of PT scheme, statistic-based assigned values were evaluated based on a two-stage process, as outlined in the methodology section. The detailed consensus of each testing item, assessed by 22 attending pathologists, is depicted in Fig. 2a and b. All five cases of histologic type and grade achieved a consensus of 70% or higher among the 22 attending pathologists. Regarding the HER2 IHC scoring, cases A1 and A3 did not reach the 70% agreement threshold, leading to the assignment of values of HER2 2+/3+, which were subsequently confirmed as HER2 FISH positive. Case A4 also exhibited less than 70% agreement, resulting in an assigned value of HER2 1+/2+. As to ER expression, case A2 exhibited an assigned range of <1%, whereas the remaining four cases displayed an assigned range of >10%. Similarly, for PR expression, case A2 showed an assigned range of <1%, while the other four cases exhibited an assigned range of ≥1% (Fig. 2a and b; Table 1). The representative images of A1 to A5 can be found in additional file (Additional file S2).

Among the fifteen cases of HER2 IHC slides evaluated, two cases (P1 and P9) displayed 100% consensus on a HER2 3+score. Five cases (P4 ~ P8) demonstrated a HER2 2+score with at least 70% agreement and were subsequently confirmed as HER2-FISH negative. Similarly, another five cases (P2, P3, P11, P13, and P15) achieved a HER2 1+score with at least 70% agreement. Two cases (P10, P12) were assigned a HER2 0 score, also achieving at least 70% agreement. However, the remaining one case (p14) did not reach the 70% agreement threshold, resulting in the assignment of an equivocal HER2 score of 0/1+ (Fig. 2c; Table 2). The representative HER2 images of P1 to P15 can be found in additional file (Additional file S3).

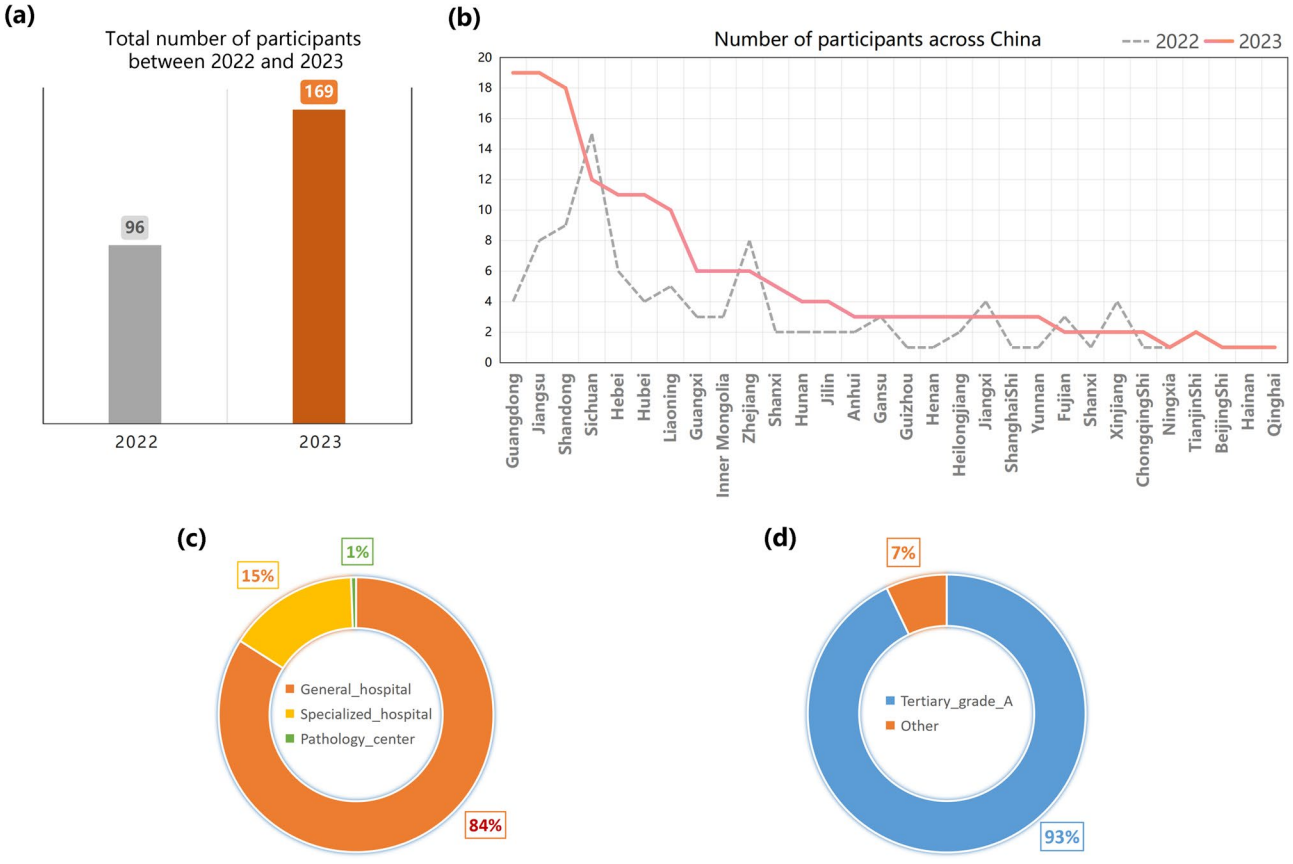


Fig. 1 Overall characteristics of participating institutions across China. **a** A noteworthy expansion in the number of participants is observed from 2022 to 2023. **b** The geographical distribution of the PT scheme has extended from 26 to 30 provinces/municipalities/autonomous regions across China. **c** The classification of institution types among the 169 participating entities is depicted in the donut charts. **d** The classification of institution levels among the 169 participants is presented in the donut charts

Concordant rates of six key items evaluated by participants and their overall performance scores

In the five cases (A1 ~ A5), the PT program demonstrated satisfactory overall concordant rates (OCR) for histologic type (84.9%), histologic grade (81.9%), ER (99.4%), PR (99.1%), HER2 (88.8%), and Ki67 (95.9%) (Fig. 3a). Regarding the individual cases, IMC (A1) and C-AD (A2) exhibited the lowest two OCRs for histologic type, achieving 63.3% and 72.2%, respectively. Similarly, both cases also exhibited the lowest two OCRs for histologic grade (71.0% and 74.6%, respectively). As to the HER2 IHC scoring, A5 demonstrated the least OCR, which is only 58.0% (Fig. 3a).

While, among the fifteen HER2 IHC slides, the OCR was 80.9% (Fig. 3b). Within the subset categorized as HER2 0, 1+and 2+/FISH-, there was a decrease in OCR, with a rate of 78.1%. As regarding to the individual HER2 slide, the lowest concordant rate was observed in cases P5 and P8 with HER2 2+score, showing 45.6% and 50.3%, respectively. Conversely, the remaining thirteen cases achieved a concordant rate exceeding 70% for their respective scores. As expected, the cases (P1 and P9) with HER2 3+score demonstrated excellent concordant rates, with rates exceeding 98% (Fig. 3b).

Among the 169 participants, the median overall performance scores were 90.0. The maximum and minimum scores were 99 and 72, respectively, displaying a left-skewed distribution that fit well with the Weibull distribution model, characterized by a shape parameter of 19.89 and a scale of 92.32 (Fig. 3c). Consequently, all participants successfully passed this round of the PT scheme. Notably, 57% of them showed excellent performance, and 38% showed good performance (Fig. 3d). However, as the Weibull distribution analysis, nine participants were identified with overall performance scores below 79.5, exhibiting statistical significance with $p<0.05$ (Fig. 3c). This finding highlighted the necessity for further identifying and addressing their specific areas of weakness to improve their diagnostic accuracy and competency.

Analysis of existing issues of HER2 0, 1+and 2+/FISH- from the PT results

In the evaluation of HER2 0, 1+and 2+/FISH- cases, some inaccuracies were observed among the 169

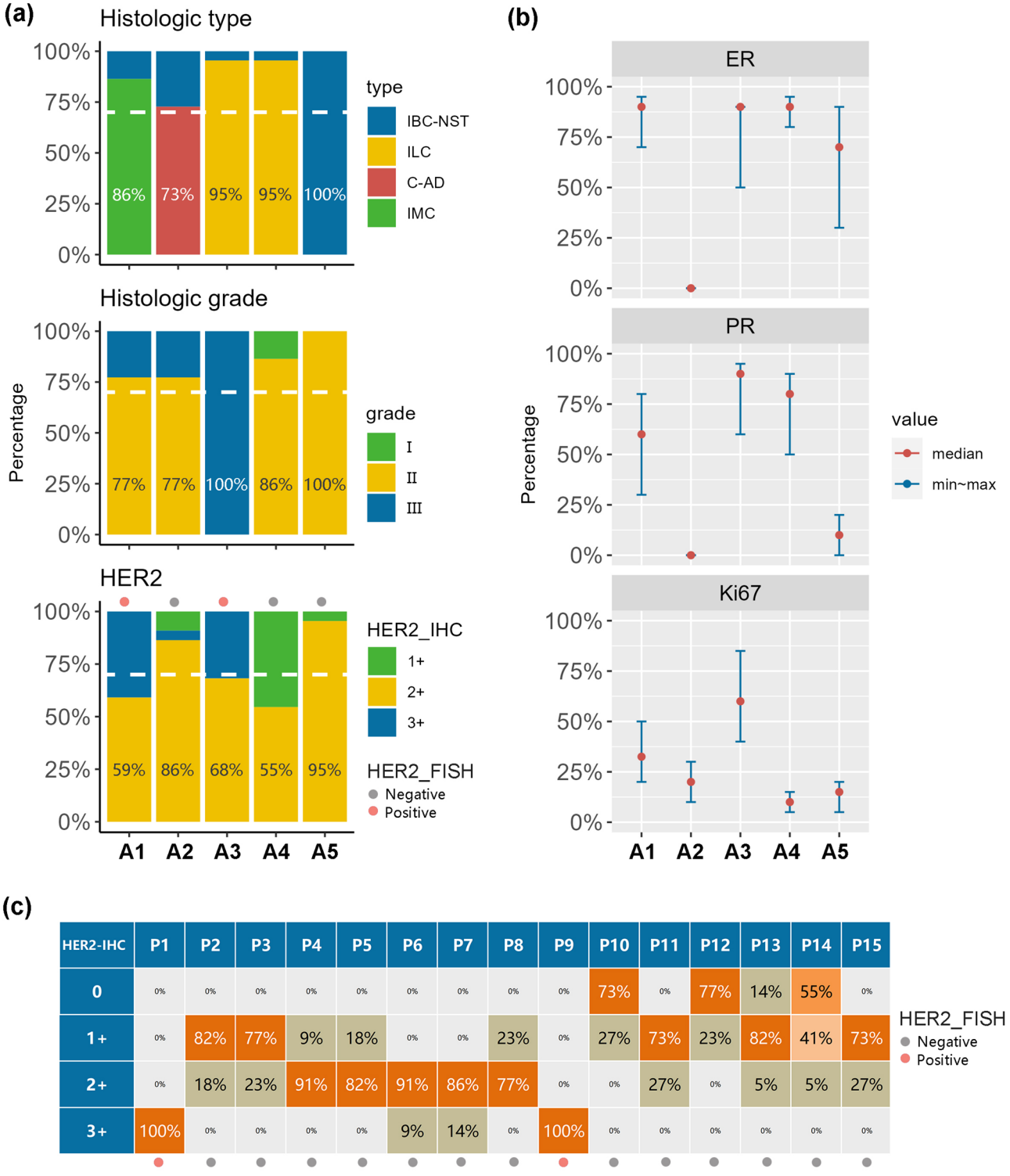


Fig. 2 The detailed consensus of each testing item assessed by 22 attending pathologists. **a** The detailed consensus of histologic type, grade and HER2 scoring across five cases (A1 ~ A5). **b** The median values along with the minimum and maximum values of ER, PR, and Ki67 for the same five cases (A1-A5). **c** The detailed consensus of HER2-IHC scoring among the fifteen cases of HER2 IHC slides

Table 1 Assigned values of six key pathological items across five cases

Case	Histologic type	Histologic grade	ER*	PR*	Ki67#	HER-2
A1	IMC	II	90%	60%	32.5% [20-50%]	2+/3+
A2	C-AD	II	0%	0%	20% [10-30%]	2+
A3	ILC	III	90%	90%	60% [40-85%]	2+/3+
A4	ILC	II	90%	80%	10% [5-15%]	1+/2+
A5	IBC-NST	II	70%	10%	15% [5-20%]	2+

*, median value; #, median value with range of min and max

participants. For the HER2 0 cases, an average of 21.6% of participants erroneously classified those as HER2 1+. In the HER2 1+category, an average of 3.8% of participants misclassified these as HER2 0, and 16.2% misclassified them as HER2 2+. Furthermore, for the HER2 2+category, an average of 24.9% of participants incorrectly classified these as HER2 1+, and 2.5% misclassified them as HER2 3+(Fig. 4a). The detail results of each category can be found in Fig. 4b.

Regarding the accuracy of sixteen cases of HER2 0, 1+and 2+/FISH-, the median accuracy was 81.2% (13/16) (Fig. 4c). When it come to the individual participant, two institutions provided correct answers only for 8 out of 16 cases, achieving the lowest accuracy rate of 50% (Fig. 4c). Conversely, the other four institutions correctly answered all 16 cases, reaching the highest accuracy rate of 100% (16/16) (Fig. 4c). Based on the Weibull distribution, characterized by a shape parameter of 8.58 and a scale of 83.21, accuracy levels below 59% were considered unsatisfactory since they were statistically significant ($p<0.05$) (Fig. 4c). Accordingly, ten participants were identified.

Interestingly, upon detailed analysis of these ten participants, systemic biases in the evaluation of HER2 0, 1+and 2+/FISH- cases were demonstrated (Fig. 4d). Among them, four participants (U8, U9, U10, and U11) exhibited a tendency to overestimate the HER2 score, frequently classifying HER2 1+as HER2 2+, and HER 0 as HER2 1+, as depicted in Fig. 4d. Conversely, two participants (U7 and U12) tended to classify HER2 2+as HER2 1+ (Fig. 4d). Therefore, these findings indicated the presence of systemic bias in the evaluation of HER2 0, 1+and 2+/FISH- in a few institutions.

Meanwhile, our PT scheme also revealed that combining HER2 1+and 2+/FISH- categories together could result in a higher concordance rate. For the combined group (HER2 1+ & 2+/FISH-), an average of 96.8% of participants demonstrated correct classification. While, for the HER2 0 cases, an average of 78.1% of participants demonstrated accurate classification (Fig. 5a). The detail results can be found in Fig. 5b.

Analysis of existing issues of histologic type and grade from the PT results

There were several issues identified in the evaluation of histologic type and grade. Regarding histologic type, 31.6% of participants incorrectly diagnosed IMC as IBC-NST, and 24.0% misclassified C-AD as IBC-NST (Fig. 6a). Notably, one participant (U1) exhibited a low concordant rate of 20% for histologic type evaluation, diagnosing all cases as IBC-NST (Fig. 6b).

For histologic grading, an average of 7.1% of participants inaccurately classified grade 2 tumors as grade 1, while an average of 13.1% misclassified them as grade 3. The detailed misclassification percentages for the five cases were depicted in Fig. 6c. As to the individual institution, participant U5 did not grade ILC and assigned a grade 3 to all the other three cases (Fig. 6d). Another participant (U6) consistently graded all five cases as grade 3 (Fig. 6d). Conversely, participant U7 tended to assign grade 1 to all five cases, representing the opposite extreme in grading tendencies (Fig. 6d). Therefore, these findings indicated the presence of systemic bias in the diagnosis of histologic type and grade in a few institutions.

Discussion

In 2022, our team launched the pioneering national PT scheme for the pathological diagnosis of breast cancer. Since its inception, the credibility of this scheme has been widely recognized by pathology institutions across China. The geographic distribution expanded from 26 to 30 provinces/municipalities/autonomous regions, encompassing 169 participating institutions in 2023, which contributed to offering a more comprehensive view of the diagnostic capabilities of pathologists in China.

Currently, low HER2 expression plays an increasingly essential role in guiding breast cancer treatment, particularly with the emergence of HER2 antibody drug conjugate (ADC) highlighted by the encouraging results of Trastuzumab deruxtecan (T-DXd) in the DestinyBreast04 study [12]. However, this subgroup is not yet distinctly separated from HER2 0 counterpart in routine clinical practice in China. As a result, data on the proficiency of HER2 0, 1+and 2+/FISH- scoring among Chinese pathologists remains limited. To address this gap,

Table 2 Assigned values of HER2-IHC scoring among fifteen cases of HER2 IHC slides

Case	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
HER2	3+	1+	1+	2+	2+	2+	2+	2+	3+	0	1+	0	1+	0/1+	1+

our PT scheme conduct a nationwide baseline assessment of HER2-IHC classification proficiency in China. As expected, the IHC evaluation of HER2 3+cases showed excellent OCR, achieving 99.4% (Fig. 3b). This high concordance reflects the well-established criteria for HER2 3+categorization, which are consistently implemented by Chinese pathologists.

However, the interpretation accuracy for HER2 0, 1+and 2+/FISH- cases was markedly lower, with an OCR of only 78.1%. More concerning, a few participants in the PT scheme showed systemic bias in the evaluation of sixteen cases with HER2 0, 1+and 2+/FISH-. They exhibited accuracy levels below 59% with $p<0.05$ that were considered unsatisfactory (Fig. 4c). Among them, four participants (U8, U9, U10, and U11) demonstrated a tendency to overestimate HER2 score, frequently classifying HER2 1+as HER2 2+, and HER 0 as HER2 1+. One possible reason for this systemic bias is that the evaluation of low HER2 expression is a new clinical concern, which is not widely accepted by all institutions in China. Consequently, a few institutions may primarily focus on the assessment of HER2 2+and 3+cases, which have been emphasized for their clinical significance for many years in China, while paying less attention to accurate evaluation of low HER2 expression. As a result, they may not strictly adhere to the criteria for evaluating HER2 0, 1+, and 2+/FISH- as recommended by currently-used 2018 ASCO/CAP guidelines [11]. To address these systemic bias issues, implementing continuous quality assurance mechanisms within pathology institutions in China is essential. Regular proficiency testing, like the scheme conducted in this study, is critical for upholding high standards of diagnostic accuracy. By identifying systemic biases and individual discrepancies in HER2 scoring, our PT schemes help to pinpoint specific areas where further training are needed.

Additionally, our PT scheme also revealed that combining HER2 1+and 2+/FISH- into a single group could enhance the diagnostic concordance, showing an impressive average concordance rate of 96.8% among participants (Fig. 5a and b). These findings were aligned with the Zaakouk M, et al., study [13]. However, it should be noted that the direct use of "combined group" or similar nomenclature that integrates HER2 2+/FISH- and 1+into a single category is not recommended in the currently routine practice. The HER2 2+category remains crucial for determining further FISH testing for HER2 amplification currently and should, therefore, be interpreted separately. Simply combining HER2 2+/FISH- and 1+into a single category may lead to potential misunderstanding or misguidance regarding the necessity of HER2 2+scoring independently.

In fact, for the emerging clinical treatment strategies involving HER2-ADC [12, 14], the primary challenge

Fig. 3 Concordant rates of six key items evaluated by participants and their overall performance scores. **a** Overall concordant rates (OCR) for histologic type, histologic type, ER, PR, HER2, and Ki67 across five cases (A1 ~ A5). **b** OCR for HER2 IHC scoring among the fifteen HER2 IHC slides. **c** The Weibull distribution of overall performance scores across 169 participants. **d** The overall performance scores are divided into three categories, delineating pass, good, and excellent levels

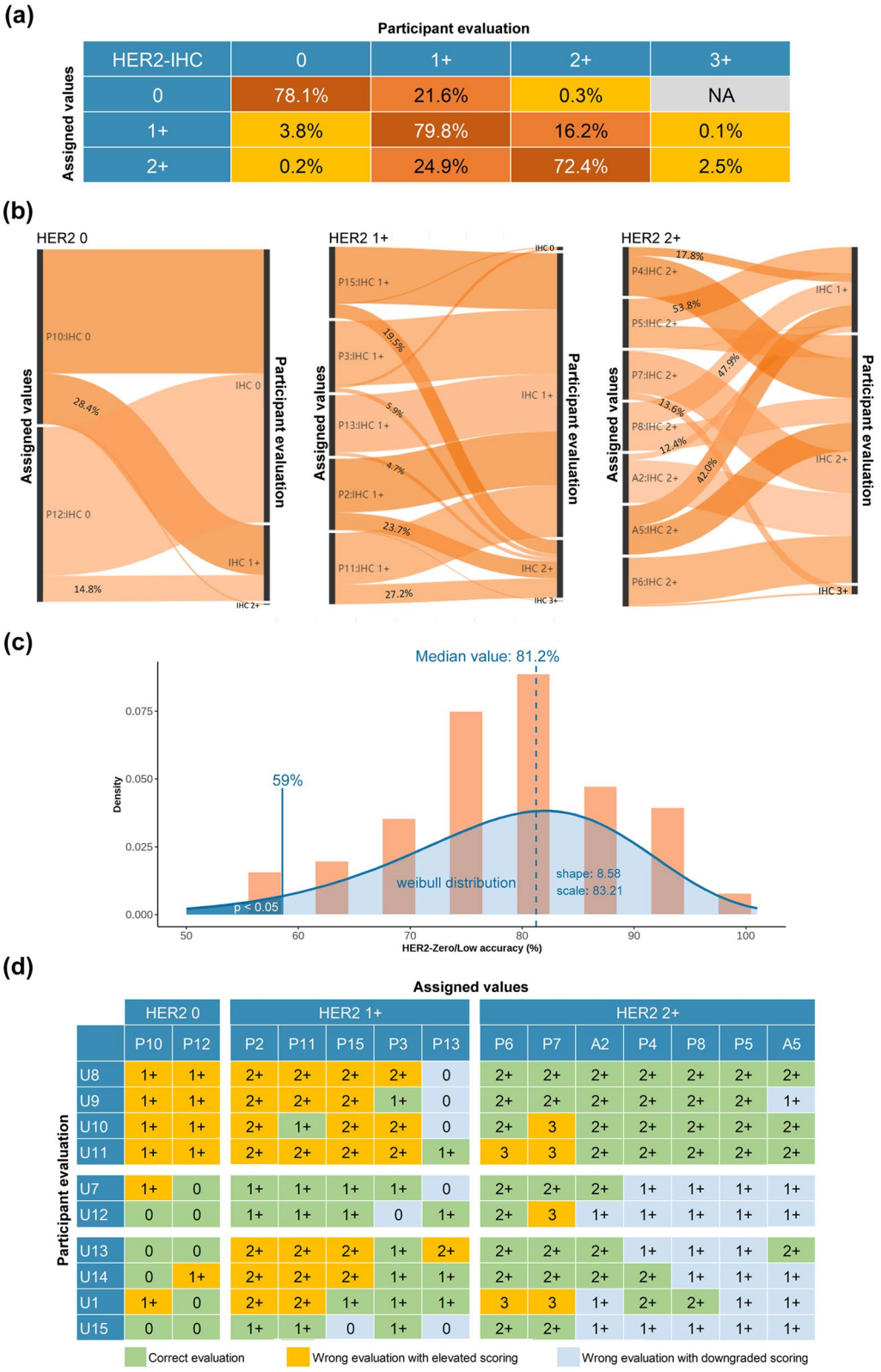


Fig. 4 Analysis of existing issues of HER2 0, 1 + and 2+. **a** Average percentages of 169 participants evaluation in individual categories of HER2 0, 1 + and 2+. **b** The detailed misclassification percentages of HER2 0, 1 + and 2 + evaluation among the 169 participants. **c** The Weibull distribution of accuracy of sixteen cases with HER2 0, 1 + and 2+. **d** Detailed scoring results of sixteen cases evaluated by ten participants who showed accuracy rates below 59%

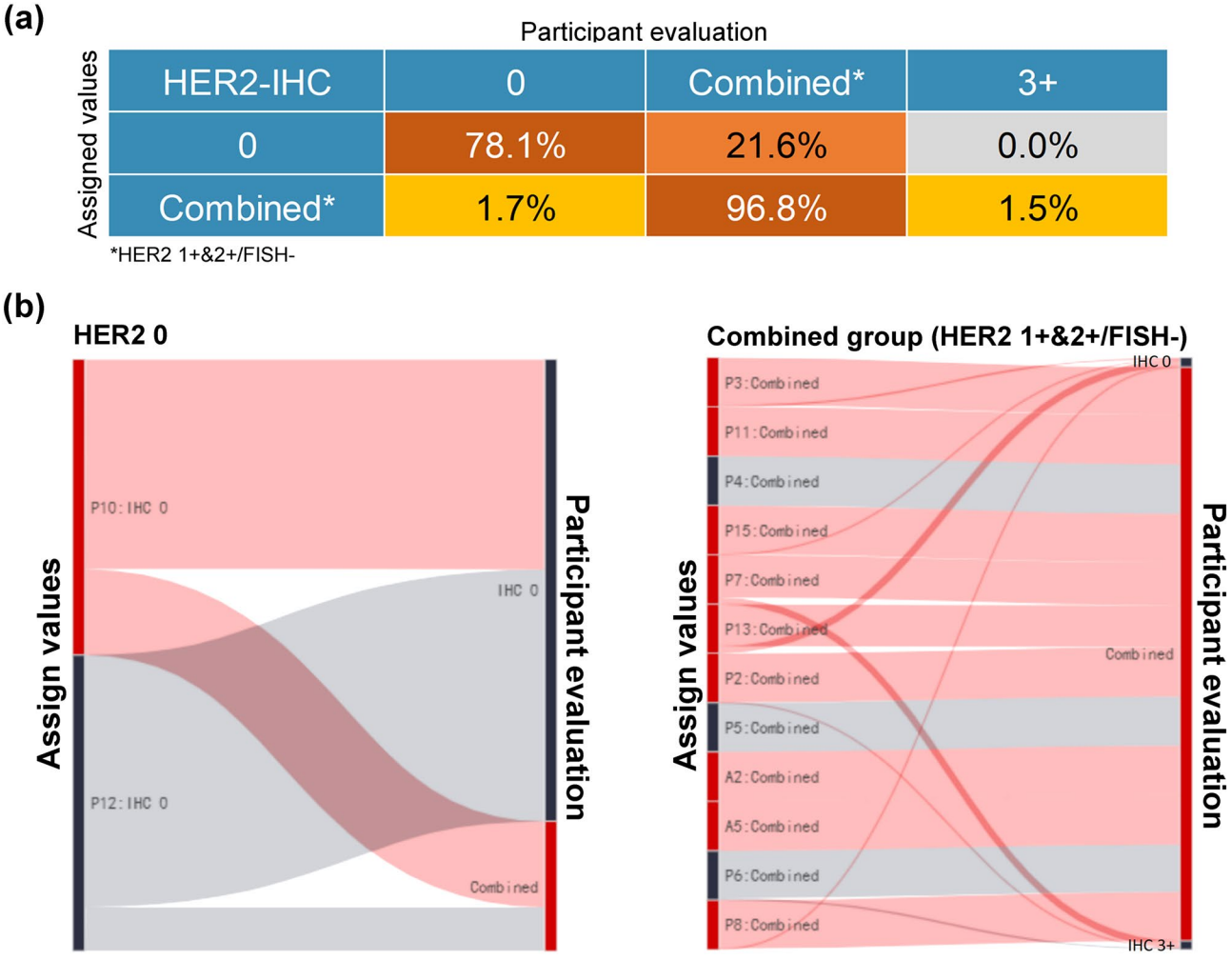
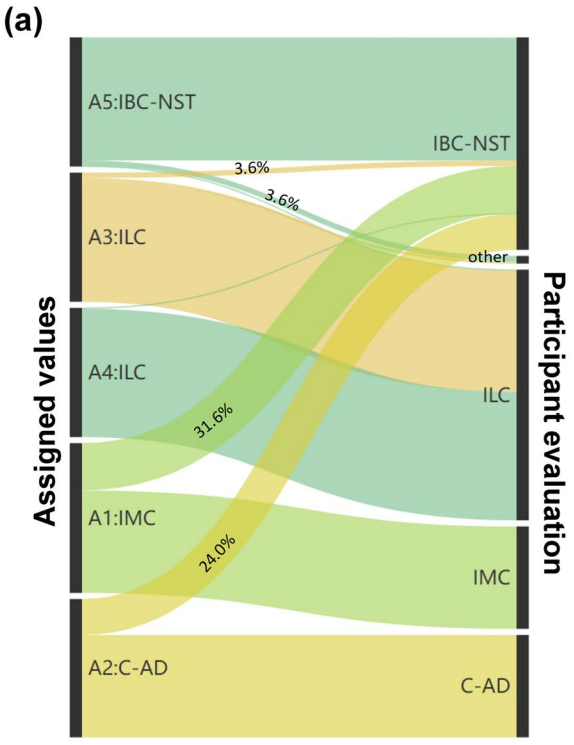
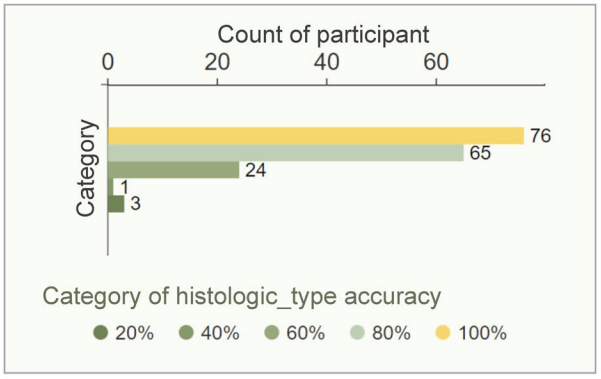


Fig. 5 Analysis of the concordant rate of cases in HER2 0 and combined (HER2 1+&2+/FISH-) groups. **a** Average percentages of evaluation by 169 participants in individual categories of HER2 0 and combined groups. **b** Detailed misclassification percentages of HER2 0 and combined groups among the 169 participants

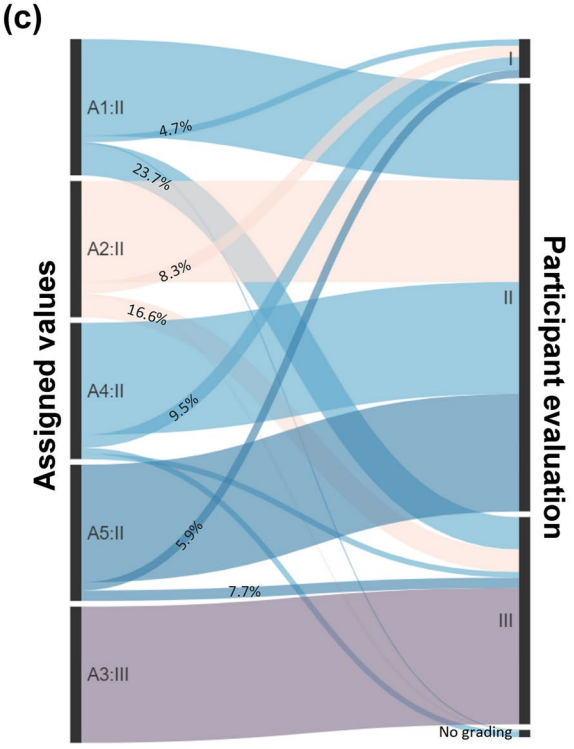
lies in the evaluation at the extreme lower end of the HER2 staining spectrum. Our results demonstrated a relatively low average concordance rate of 78.1% among 169 participants in accurately classifying HER2 0 cases, highlighting the increasing pressure on pathologists. According to currently-used 2018 ASCO/CAP guidelines [11], a defined threshold of 10% of invasive cancer cells with faint or barely detectable incomplete HER2 membrane staining produces a great difficulty for pathologists to clearly and reliably distinguish HER2 1+from HER2 0 [15]. Furthermore, the phase 2 DAISY trial conducted by Mosele F et al. demonstrated promising modest anti-tumor activity in patients categorized as HER2 0 [16]. Consequently, defining the complete HER2 0 category emerges as a particularly interesting research topic, pending outcomes from the DESTINY-Breast06 trial (NCT04622319), which aims to explore the efficacy of T-DXd in breast cancer with HER2 IHC>0<1+expression.

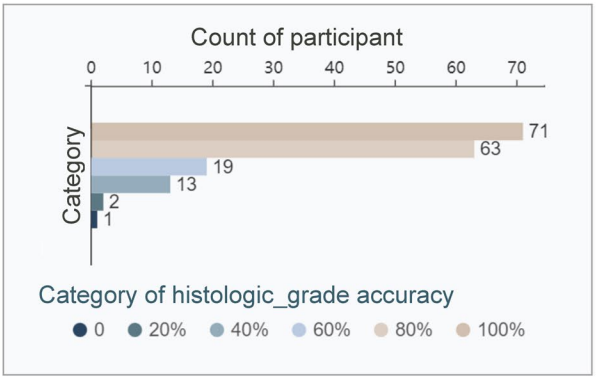
Meanwhile, areas for improvement in the interpretation of histologic type and grade were also identified. For histologic type, systemic biases were particularly notable in certain cases, such as IMC and C-AD that were incorrectly diagnosed as IBC-NST by 31.6% and 24.0% of participants, respectively (Fig. 6a). This kind of incorrect interpretations was especially prominent in some institutions, where one participant (U1) demonstrated a notably low accuracy rate of 20%, diagnosing all cases as IBC-NST (Fig. 6b). In terms of histologic grading, a few participants displayed obvious biases. Participant U6 consistently assigned a grade 3 to all cases, while participant U7 frequently assigned a grade 1, representing the opposite extreme in grading practices (Fig. 6c and d). These findings indicated that accurate diagnosis of histologic type and grade for breast cancer was less prioritized





Case	Evaluation results of four participants			
	U1	U2	U3	U4
A1:IMC	IBC-NST	IBC-NST	IBC-NST	IBC-NST
A2:C-AD	IBC-NST	IBC-NST	IBC-NST	IBC-NST
A3:ILC	IBC-NST	Carcinoma	Carcinoma	ILC
A4:ILC	IBC-NST	ILC	ILC	ILC
A5:IBC-NST	IBC-NST	Carcinoma	Carcinoma	Carcinoma
Categroy	20%	20%	20%	40%





Case	Evaluation results of three participants		
	U5	U6	U7
A1:II	III	III	I
A2:II	III	III	I
A3:III	No_grading	III	III
A4:II	No_grading	III	I
A5:II	III	III	I
Categroy	0%	20%	20%

Fig. 6 Analysis of existing issues of histologic type and grade. **a** The detailed misclassification percentages of histologic type among the 169 participants for each case (A1 ~ A5). **b** The categories of histologic type among 169 participants, and their detailed evaluation results of those participants categorized in the least two accuracy groups. **c** The detailed misclassification percentages of histologic grade among the 169 participants for each case (A1 ~ A5). **d** The categories of histologic grade among 169 participants, and their detailed evaluation results of those participants categorized in the least two accuracy groups

by some participants. Therefore, enhanced education on relevant guidelines and recommendations is crucial to address these discrepancies and refine diagnostic practices.

Despite the issues we discuss above, the general results of this round PT scheme highlighted the commendable overall competency and proficiency of the participants in the diagnosis of breast cancer in China. All participants achieved overall performance scores surpassing 70, successfully passing the testing, with 57% demonstrating excellent performance and 38% showing good performance (Fig. 3c and d). The IHC evaluations for ER, PR and Ki67, which were underscored by the NCCN guidelines [4] and reinforced by previous well-established Chinese External Quality Assessment programs [17], showed outstanding results in this round of PT (Fig. 3a and b).

There are several limitations in our study. Firstly, although we conducted training on interpreting pathological changes on digital slides after the initial round of PT scheme, evaluating such changes remains challenging for many pathologists currently. However, given the increasing incorporation of digital slides in routine clinical practice, our nationwide PT scheme holds potential for promoting widespread utilization of digital slides, and facilitating the transition from traditional to digital pathology in the future. Secondly, although the current PT scheme demonstrated a relatively satisfactory interpretation of the most common histologic subtypes of breast cancer, this focus may not sufficiently represent the diverse scenarios encountered in clinical practice. Therefore, in the future round of PT scheme, we will expand diversity of cases within the testing protocol to achieve a more comprehensive view of diagnostic capabilities among pathologists in China.

Conclusions

In summary, our PT scheme has quickly established its credibility across China. The results of the current PT program highlighted the commendable overall competency and proficiency of participants in diagnosing breast cancer. Nonetheless, the findings also identified specific areas requiring improvement, particularly systemic biases in scoring for HER2 0, 1+, and 2+/FISH- noted in a few institutions. Additionally, the relatively low concordant rate of 78.1% in classifying HER2 0 category underscores the challenges in evaluating the extreme lower end of the HER2 staining spectrum, suggesting a critical focus for further research. Meanwhile, areas for improvement in the interpretation of histologic type and grade were also identified. All in all, our findings suggested the need for implementing robust quality assurance mechanisms, like the nationwide PT scheme conducted in this study, to uphold high diagnostic standards, and to pinpoint specific aspects for further training and improvement.

Abbreviations

PT	proficiency testing
NCC	National Cancer Center
NQCCC	National Quality Control Center for Cancer
CNAS	China National Accreditation Service for Conformity Assessment
IBC-NST	Invasive breast carcinoma of no special type
ILC	Invasive lobular carcinoma
IMC	Invasive micropapillary carcinoma
C-AD	Carcinoma with apocrine differentiation
IHC	Immunohistochemistry
OCR	Overall concordant rates
ADC	Antibody drug conjugate
T-DXd	Trastuzumab deruxtecan

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13058-024-01884-9.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Author contributions

X.X. was involved in the study design, data collection, and statistical analysis and drafted the manuscript. L.G., C.G. and L.X. designed the study, prepared the samples, collected the data, and critically reviewed the manuscript. L.L. and L.Y. contributed to the sample preparation and online report system development. X.W., W.R., and P.Y. contributed to data collection. J.M., J.L., and B.W. contributed to the sample preparation, release, and preservation. Q.Z. and W.X. contributed to the determination of the assigned values and uncertainty. F.M, W.Y. and J.Y. contributed to study design, coordination, discussion, and critical manuscript editing. All authors approved the content of this manuscript.

Funding

Dr. C.G. has received research funding from the National Anti-Tumor Drug Surveillance System of National Cancer Center (Grant No.DSS-YSF2022-08), which supported this study.

Data availability

The data of this study are available from the corresponding author J.Y. email: jmying@cicams.ac.cn, upon request.

Declarations

Ethics approval and consent to participate

This is a retrospective study which did not affect the patient's therapy, and no patient privacy information is involved. Therefore, this study was approved by the ethical approval from the Independent Ethics Committee of Cancer Hospital, Chinese Academy of Medical Sciences, National GCP Center for Anticancer Drugs (22/499–3701), with waiver of informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China
[2]Department of Medical Affairs, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

[3]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Pathology, Peking University Cancer Hospital & Institute, Beijing 100142, China
[4]Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China
[5]Office for Cancer Diagnosis and Treatment Quality Control, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China

Received: 17 May 2024 / Accepted: 20 August 2024
Published online: 03 September 2024

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. Cancer J Clin. 2021;71(3):209–49.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. Cancer J Clin. 2016;66(2):115–32.
3. Xia C, Dong X, Li H, Cao M, Sun D, He S, Yang F, Yan X, Zhang S, Li N, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. Chin Med J (Engl). 2022;135(5):584–90.
4. National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology: Breast Cancer.Version 4.2022; 2022.
5. WHO Classification of Tumours Editorial Board. Breast tumours. Lyon (France): International Agency for Research on Cancer. (WHO classification of tumours series, 5th ed.; vol.2) https://publications.iarc.fr/581; 2019.
6. Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, Hayes DF, Lakhani SR, Chavez-MacGregor M, Perlmutter J, et al. Estrogen and progesterone receptor testing in breast Cancer: ASCO/CAP Guideline Update. J Clin Oncology: Official J Am Soc Clin Oncol. 2020;38(12):1346–66.
7. Recommended by Breast Cancer Expert Panel. [Guideline for HER2 detection in breast cancer, the 2019 version]. Zhonghua Bing Li Xue Za Zhi. 2019;48(3):169–75.
8. Biganzoli L, Marotti L, Hart CD, Cataliotti L, Cutuli B, Kuhn T, Mansel RE, Ponti A, Poortmans P, Regitnig P, et al. Quality indicators in breast cancer care: an update from the EUSOMA working group. Eur J Cancer. 2017;86:59–81.
9. Del Turco MR, Ponti A, Bick U, Biganzoli L, Cserni G, Cutuli B, Decker T, Dietel M, Gentilini O, Kuehn T, et al. Quality indicators in breast cancer care. Eur J Cancer. 2010;46(13):2344–56.
10. Xue X, Guo L, Guo C, Li L, Yang L, Wang X, Rao W, Yuan P, Mu J, Li J et al: Proficiency testing of diagnosis in histopathology and immunohistochemistry of breast pathology in China: results from a pilot work of National Single Disease Quality Control Program for breast cancer. BMC Cancer 2024, 24(1):23.
11. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, Bilous M, Ellis IO, Fitzgibbons P, Hanna W, 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. J Clin Oncology: Official J Am Soc Clin Oncol. 2018;36(20):2105–22.
12. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, Tsurutani J, Ueno NT, Prat A, Chae YS, et al. Trastuzumab Deruxtecan in previously treated HER2-Low advanced breast Cancer. N Engl J Med. 2022;387(1):9–20.
13. Zaakouk M, Quinn C, Provenzano E, Boyd C, Callagy G, Elsheikh S, Flint J, Millican-Slater R, Gunavardhan A, Mir YJTB. 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. 2023, 70:82–91.
14. Mosele F, Deluche E, Lusque A, Le Bescond L, Filleron T, Pradat Y, Ducoulombier A, Pistilli B, Bachelot T, Viret F, et al. Trastuzumab Deruxtecan in metastatic breast cancer with variable HER2 expression: the phase 2 DAISY trial. Nat Med. 2023;29(8):2110–20.
15. Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, Marra A, Viale G, Trapani D, Cardoso F, et al. HER2-Low breast Cancer: pathological and clinical Landscape. J Clin Oncology: Official J Am Soc Clin Oncol. 2020;38(17):1951–62.
16. Mosele F, Deluche E, Lusque A, Le Bescond L, Filleron T, Pradat Y, Ducoulombier A, Pistilli B, Bachelot T, Viret F et al. Trastuzumab Deruxtecan in metastatic breast cancer with variable HER2 expression: the phase 2 DAISY trial. Nat Med 2023.
17. Pu T, Shui R, Shi J, Liang Z, Yang W, Bu H, Li Q, Zhang Z. External quality assessment (EQA) program for the immunohistochemical detection of ER, PR and Ki-67 in breast cancer: results of an interlaboratory reproducibility ring study in China. BMC Cancer. 2019;19(1):978.

Publisher's note