UK NEQAS
Immunocytochemistry
& In-Situ Hybridisation

Proficiency testing for HER2-low expression in breast cancer: results of the UK NEQAS ICC & ISH programme examining technical and interpretive accuracy

Suzanne Parry[1]*, Abeer Shaaban[2], Bharat Jasani[3], Andrew Dodson[1]

1). UK National External Quality Assessment Scheme for Immunocytochemistry & In-Situ Hybridisation (UK NEQAS ICC & ISH), 5 Coldbath Square, London EC1R 5 HL, UK. 2). University Hospitals Birmingham NHS Foundation Trust, Birmingham B15 2GW, UK. 3). University of Cardiff, Cardiff, CF10 3AT Wales
* The lead author declares that they have no conflicts of interest to declare. Email address: sparry@ukneqasiccish.org

2023 ESMO BREAST CANCER Annual Congress
BERLIN GERMANY
11-13 MAY 2023
Poster: 54P (ID274)

BACKGROUND

Trastuzumab deruxtecan (Enhertu®) has been shown to have significant anti-tumour efficacy against HER2-low-expressing metastatic breast cancer,[1] and has recently received approval for use in the USA and in Europe in this clinical setting.

Identification of patients with the potential to respond to the therapy relies on accurate assessment of low levels of HER2 protein expression (HER2 0, 1+ and 2+). Previously, the HER2 1+ category has not been considered clinically significant and evidence on the ability of laboratories to identify it reproducibly is lacking.

We report the first results of an external quality assessment (EQA) programme specifically designed to examine testing proficiency in this area.

METHODS AND MATERIAL

Study cohort

The UK National External Quality Assessment Scheme for Immunocytochemistry & In-Situ Hybridisation (**UK NEQAS ICC & ISH**) recruited laboratories from amongst its participants who were enrolled in the Scheme's EQA programme for HER2 protein expression assessment. All of whom regularly carry-out HER2 predictive testing in breast cancer (BC) in the clinical setting.

Breast cancer samples and test tissue microarray construction

Formalin-fixed paraffin-embedded (FFPE) BC resection samples were independently tested in two different laboratories, both of which used an FDA approved immunohistochemical (IHC) assay (PATHWAY anti-HER-2/neu (4B5) Rabbit Monoclonal Primary Antibody, Roche Diagnostics) according to the manufacturer's instructions.

Three different BC samples were identified in which HER2 protein expression levels were reproducibly shown to be: HER2 1+ in tumour A, HER2 0 in tumour C and HER2 2+ in tumour D. In these three samples, expression was homogeneous throughout the tumours. In a fourth BC sample (B) expression was initially shown to be HER2 1+. But following its incorporation into the test tissue microarray (TMA) its expression at deeper levels changed to HER2 0. This was identified by the Scheme's standard quality control (QC) procedures. The test TMA was constructed using 1.2mm cores taken from the four different BC samples.

Participant staining

Participating laboratories were provided with 3µm sections cut from the TMA and asked to demonstrate HER2 protein using their standard IHC methodology. Stained slides were then submitted for central assessment. Laboratories were also required to submit details of their staining methodology and optionally, their own interpretation of each samples staining.

Assessment procedure

A panel of four assessors comprising two expert pathologists (AS, BJ) and two biomedical scientists with extensive experience in assessing HER2 staining in the EQA setting (SP, AD) concurrently but independently assigned a HER2 category to the returned slides, variance was resolved by discussion. Assessment of reference laboratory-stained slides cut at the closest matched section level gave the expected category for comparison.

RESULTS

Participation

Forty-five laboratories submitted stained slides for assessment, 36 (80%) of whom supplied their interpretation of HER2 expression in each of the supplied BC samples.

IHC staining methodologies

Table 1 gives details of the methods used. These comprised three companion diagnostic assays (CE-IVD or FDA approved) and three laboratory developed tests. Where the assay was not performed according to the manufacturer's recommended instructions, we classified that method as an LDT. This was applicable to the 4B5 (Roche Diagnostics) product only and it was due to the use of a non-assay version of the 4B5 antibody. Methods submitted by the users of the Hercep Test (Agilent Dako) and Oracle (Leica Biosystems) methods did not indicate any deviations from the respective recommended protocols.

Antibody Name	Antibody Type	Supplier	Assay Name	Method Type	Count (%)
Hercep Test	Rb poly	Agilent Dako	Hercep Test	Assay	1 (2.2%)
CB11	Ms mono	Leica Biosystems	Oracle	Assay	2 (4.4%)
4B5	Rb mono	Roche Diagnostics	PATHWAY/Confirm	Assay	39 (86.7%)
4B5	Rb mono	Roche Diagnostics	NA	LDT*	1 (2.2%)
SP3	Rb mono	Cell Marque	NA	LDT	1 (2.2%)
c-erbB-2	Rb poly	Agilent Dako	NA	LDT	1 (2.2%)

Table 1. Primary antibody and IHC method type with data on usage in the study.
Rb = rabbit; Ms = mouse; poly = polyclonal; mono = monoclonal; LDT = laboratory developed test; LDT* = non-assay version of 4B5 antibody used; NA = not applicable.

HER2 category agreement in locally versus centrally stained core samples

The expert panel's assessment of HER2 category in each of the four core samples was compared with the expected category obtained by the same panel's assessment of centrally stained references (in duplicate). The results are shown in Table 2 below.

	Assay (4B5)		Assay (CB11)		Assay (Hercep)		LDT		Totals	
	N	%	N	%	N	%	N	%	N	%
Sample A										
Agree	29	74.4	0	0.0	0	0.0	3	100.0	32	71.1
Disagree	10	25.6	2	100.0	1	100.0	0	0.0	13	28.9
Totals	39	100.0	2	100.0	1	100.0	3	100.0	45	100.0
Sample B										
Agree	23	59.0	0	0.0	0	0.0	0	0.0	23	51.1
Disagree	16	41.0	2	100.0	1	100.0	3	100.0	22	48.9
Totals	39	100.0	2	100.0	1	100.0	3	100.0	45	100.0
Sample C										
Agree	39	100.0	2	100.0	1	100.0	3	100.0	45	100.0
Disagree	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Totals	39	100.0	2	100.0	1	100.0	3	100.0	45	100.0
Sample D										
Agree	33	84.6	0	0.0	1	100.0	2	66.7	36	80.0
Disagree	6	15.4	2	100.0	0	0.0	1	33.3	9	20.0
Totals	39	100.0	2	100.0	1	100.0	3	100.0	45	100.0

Table 2. Levels of agreement for local versus central staining.
LDT = laboratory developed test.

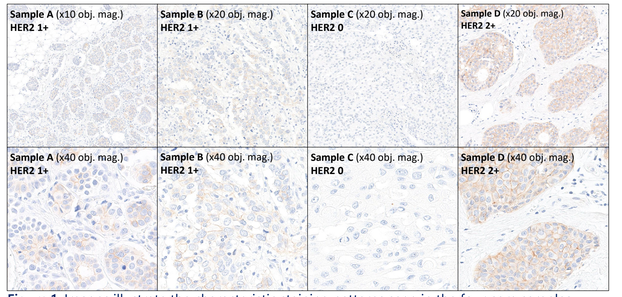

Figure 1. Images illustrate the characteristic staining patterns seen in the four core samples.
Top row = low power photomicrographs (taken using either x10 or x20 objective lens), bottom row are high power views selected from the area of tumour illustrated in the matching low power image (all taken using x40). Obj. mag. = Objective magnification.

HER2 category agreement in locally versus centrally assessed core samples

The levels of interpretive agreement were examined in each of the four core samples independently. This was done by comparing the expert panel's HER2 category assignment with that of the local observer examining the same core (locally stained). The results are shown in Table 3 below.

	Assay (4B5)		Assay (CB11)		Assay (Hercep)		LDT		Totals	
	N	%	N	%	N	%	N	%	N	%
Sample A										
Agree	12	40.0	1	50.0	0	0.0	0	0.0	13	36.1
Disagree	18	60.0	1	50.0	1	100.0	3	100.0	23	63.9
Totals	30	100.0	2	100.0	1	100.0	3	100.0	36	100.0
Sample B										
Agree	14	46.7	0	0.0	1	100.0	2	66.7	17	47.2
Disagree	16	53.3	2	100.0	0	0.0	1	33.3	19	52.8
Totals	30	100.0	2	100.0	1	100.0	3	100.0	36	100.0
Sample C										
Agree	25	83.3	1	50.0	1	100.0	3	100.0	30	83.3
Disagree	5	16.7	1	50.0	0	0.0	0	0.0	6	16.7
Totals	30	100.0	2	100.0	1	100.0	3	100.0	36	100.0
Sample D										
Agree	20	66.7	0	0.0	1	100.0	2	66.7	23	63.9
Disagree	10	33.3	2	100.0	0	0.0	1	33.3	13	36.1
Totals	30	100.0	2	100.0	1	100.0	3	100.0	36	100.0

Table 3. Levels of agreement for local versus central assessment.
LDT = laboratory developed test.

HER2 category agreement for local versus central staining and assessment

This analysis looked at the combined effects of local staining and assessment on the level of interpretive agreement. It was examined in each of the four core samples independently. This was done by comparing the local HER2 category assignment with that of the expert panels assessment of a closely matched, centrally stained comparator section. The results are shown in Table 4 at the top of the next column.

	Assay (4B5)		Assay (CB11)		Assay (Hercep)		LDT		Totals	
	N	%	N	%	N	%	N	%	N	%
Sample A										
Agree	7	23.3	0	0.0	1	100.0	1	33.3	9	25.0
Disagree	23	76.7	2	100.0	0	0.0	2	66.7	27	75.0
Totals	30	100.0	2	100.0	1	100.0	3	100.0	36	100.0
Sample B										
Agree	14	46.7	2	100.0	0	0.0	1	33.3	17	47.2
Disagree	16	53.3	0	0.0	1	100.0	2	66.7	19	52.8
Totals	30	100.0	2	100.0	1	100.0	3	100.0	36	100.0
Sample C										
Agree	25	83.3	1	50.0	1	100.0	3	100.0	30	83.3
Disagree	5	16.7	1	50.0	0	0.0	0	0.0	6	16.7
Totals	30	100.0	2	100.0	1	100.0	3	100.0	36	100.0
Sample D										
Agree	20	66.7	2	100.0	1	100.0	3	100.0	26	72.2
Disagree	10	33.3	0	0.0	0	0.0	0	0.0	10	27.8
Totals	30	100.0	2	100.0	1	100.0	3	100.0	36	100.0

Table 4. Levels of agreement for locally versus centrally stained and assessed samples.
LDT = laboratory developed test.

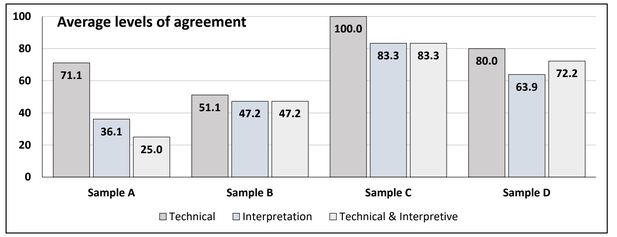

Figure 2. Average levels of agreement across all staining method types for each core sample.

CONCLUSIONS

Analysis of this study's data indicates that in the HER2-low expression range:

- breast cancer samples, when tested for HER2 expression in clinical laboratories well-practised in HER2 testing, produce category scores that show poor agreement levels (here between 25.0 and 47.2%) when compared with those obtained using a well-validated assay interpreted by experts.
- variance in the technical procedures used in the sample staining and inconsistences in interpretive assessment contributed in more or less equal measure to the variability that underlies the poor agreement levels seen.

References: 1. Modi S, et al. N Engl J Med. 2022; 387(1):9-20. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer.

The authors wish to thank all participating laboratories and to acknowledge the invaluable contribution they have made to the study.