Evidence of significant differences in the quality of PD-L1 immunohistochemical testing of triple negative breast cancer: results from external quality assessment

Suzanne Parry[1] and Andrew Dodson[1]. 1. UK NEQAS ICC & ISH, 5 Coldbath Square, London EC1R 5HL UK.

Corresponding author: sparry@ukneqasiccish.org.

UK NEQAS
International Quality Expertise

Layman's Summary

Triple negative breast cancer (TNBC) is an aggressive type of breast cancer. It is composed of cells that do not express oestrogen receptor, progesterone receptor or human epidermal growth factor-2 (HER2) – hence 'triple-negative'. The lack of these targets means the cancer is unlikely to respond to two highly effective breast cancer treatments - endocrine therapy and HER2-targeted therapy,. Until recently, this left chemotherapy as the only remaining treatment option. Unfortunately, it is known that chemotherapy has limited success in treating TNBC.

Over the last few years, a revolutionary new type of cancer treatment called immunotherapy has started to be used in a wide range of hard-to-treat cancers, often with great success. Immunotherapy uses the bodies own defense system to target and destroy the cancer cells. For it to work effectively a protein called PD-L1 must be present on the tumour cells, or the tumour associated immune cells.

A laboratory-based test called immunohistochemistry (IHC) is used to measure the amount of the PD-L1 protein present on a sample of the patient's tumour. This tells the clinician if immunotherapy is likely to work for that patient. It is therefore very important that the test produces accurate results. We have identified some features of the IHC method that need to be carefully controlled if the test is to work well. These are:

- the identity of the primary antibody used to detect the PD-L1 protein
- a key step in the IHC method - antigen retrieval, which must be done in exactly the right way
- use of an additional step in the method which amplifies the signal and makes it easier to see

We present evidence gathered as a part of our routine quality testing of large numbers of hospital laboratories performing the test which proves these are significantly important.

This information will help laboratories to optimize those important features of the test and so improve the accuracy of their results.

Introduction

PD-L1 inhibitors are used as first line treatment options for patients with advanced disease in Triple Negative Breast Cancer (TNBC). Companion and complementary diagnostic PD-L1 immunohistochemistry (IHC) assays are available for use as predictive diagnostic tests to aid in treatment decisions when selecting patients who may benefit from treatment. Laboratories can also develop and validate their own IHC method (laboratory developed test, LDT).
The UK National External Quality Assessment Scheme for Immunocytochemistry and In-situ Hybridisation (UK NEQAS ICC & ISH) regularly carries out external quality assessment (EQA) of laboratories that carry-out IHC testing for PD-L1 in TNBC and has done so at quarterly intervals since 2021.
We examined the data that we have collected from all the assessment runs conducted to date, looking for trends over time and associations between assessment score achieved and methodological parameters. Here we report the results of that examination.

Materials and Methods

At each assessment survey, the UK NEQAS ICC & ISH PD-L1 in TNBC EQA programme prepared and distributed unstained composite slides consisting of formalin-fixed-paraffin-embedded (FFPE) tonsil, TNBC tissues, and cell lines of known PD-L1 expression. Participants were asked to stain the slides using their routine clinical PD-L1 IHC method and return the slides for a central assessment by an expert panel of four assessors working independently. Participating laboratory's slides were qualitatively assessed for technical quality, by comparing to the staining achieved in reference slides prepared both in the Scheme's own testing facility and externally in partner reference laboratories.

Data was collected on antibodies and methods used and other methodological parameters.

Data from each of the PD-L1 TNBC survey sets were collated and retrospectively analysed looking for significant associations between methodological parameters and test accuracy.

Results

Between 2020 and 2023, 11 assessment runs were conducted at 3-monthly intervals, with the total number of submissions being 352. The mean number of laboratories subscribed per run was 32 (range: 16 - 40). Participating laboratories were located in 24 different countries, with UK-based laboratories contributing 141 (41%) of submissions overall, this being the largest contribution from any one country. The average quality score for all submissions was 12.8 (scored range: 4 to 20, with four being unacceptable and 20 excellent), a score of 12.8 is in the low acceptable range.

Run Number	Count of Laboratories (N)	Average of Quality Score
132	16	10.6
133	29	12.8
134	32	13.4
135	33	12.5
136	32	12.4
137	35	11.2
138	40	12.7
139	36	13.8
140	37	13.2
141	32	14.6
142	30	12.6
Overall	**352**	**12.8**

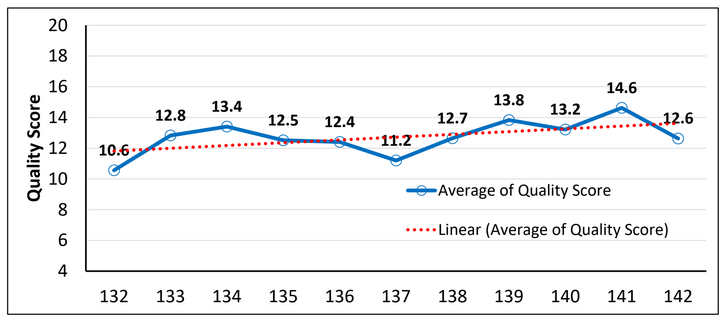


Table 1 and Chart 1. Descriptive statistics for each Assessment Run.
Count of laboratories submitting and the average quality score at each Run are shown. The trendline indicates an improvement in the quality of submissions over time.

A companion diagnostic assay based on the rabbit monoclonal SP142 antibody (Ventana) was the test employed by most laboratories, N = 292 (82.9%). This test showed a mean quality score of 13.1 (range 10.9 – 15.4). The next most commonly used test was the 22C3 antibody (Dako) used as a companion diagnostic (CDx). In this format it was used by 24 of laboratories (6.8%). The mean quality score was 12.3 (range = 8.0 – 16.0). Table 2 gives a full breakdown of antibody usage and Table 3 shows quality

Primary Antibody	Method Type	Count of Submissions (N)	Proportion of Submissions (%)
22C3 (Concentrate, Dako)	LDT	7	2.0
22C3 (GE006 PharmDx, Dako)	CDx	1	0.3
22C3 (SK006 PharmDx, Dako)	CDx	23	6.5
73-10 (PA0832, Leica)	LDT	1	0.3
E1L3N (13684, Cell Signaling)	LDT	1	0.3
Not stated		23	6.5
QR1 (Quartett)	LDT	1	0.3
SP142 (740-4859, Ventana)	CDx	37	10.5
SP142 (741-4860, Ventana)	CDx	255	72.4
SP263 (790-4905, Ventana)	LDT	3	0.9
Grand Total		**352**	**100.0**

Table 2. Primary Antibody Usage.
Six different antibody clones were used by participants. Their regulatory status is indicated in the second column (CDx = companion diagnostic assay; LDT = laboratory developed test.

Results (continued)

Method	Average of Quality Score	Count (n)
CDx		
22C3 (GE006 PharmDx, Dako)	19.0	1
22C3 (SK006 PharmDx, Dako)	12.6	15
SP142 (740-4859/741-4860, Ventana)	14.5	53
Whole group	**14.2**	**69**
CDx used off-label		
SP142 (740-4859/741-4860, Ventana) (no amp)	13.4	43
SP142 (740-4859/741-4860, Ventana) (incorrect AR & no amp)	13.3	40
SP142 (740-4859/741-4860, Ventana) (incorrect AR)	13.0	107
Whole group	**13.2**	**190**
LDT		
22C3 (Concentrate, Dako)	10.7	7
22C3 (SK006 PharmDx, Dako)	12.3	7
73-10 (PA0832, Leica)	12.0	1
E1L3N (13684, Cell Signaling)	8.0	1
QR1 (Quartett)	8.0	1
SP142 (740-4859/741-4860, Ventana)	9.3	3
SP263 (790-4905, Ventana)	8.0	3
Whole group	**10.5**	**23**
Grand Total	**13.2**	**282**

Table 3. Primary Antibody Usage combined with Details of Method.

Methods have been separated by type (CDx/LDT). Assays have been further divided according to their methodological details into those used in-line with manufacture's recommendations (CDx) and those in which the method had been altered (CDx used off-label).

Submissions from 70 participants had to be excluded due to supply of incomplete methodological details.

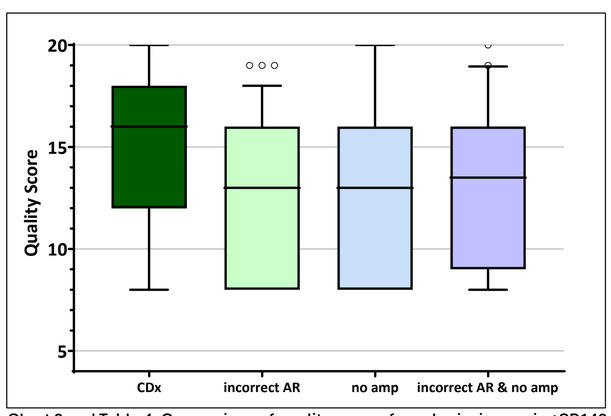
With the exception of SP142 most individual methods were represented by small numbers of submissions and results for these must be treated with caution.

Statistical analysis was performed on combined groups and showed:

The distributions of the CDx group (mean 14.2, 95%CIs: 13.2 – 15.2) was not significantly different to that of the CDx (off label) group (mean 13.2, 95%CIs: 12.6 – 13.7)

However, both the CDx group and the CDx (off label) group were associated with significantly higher quality scores when compared to the distribution for submissions stained using LDT methods (mean 10.5, 95%CIs: 9.1 – 11.9) at P = 0.0002 and 0.0013 respectively.

Table 3 shows a detailed breakdown of the methodologies employed. For SP142 in particular this indicates a substantial number of laboratories were not using the CDx assays according manufacturer's recommendations. This is true both for the antigen retrieval and for the use of an amplification reagent which is mandated in the recommended method. Comparisons showed a significant association of those laboratories using the assay as indicated with higher quality scores.



	CDx	incorrect AR	no amp	incorrect AR & no amp
Count (N)	53	107	43	40
Minimum	8.0	8.0	8.0	8.0
25% Percentile	12.0	8.0	8.0	9.0
Median	16.0	13.0	13.0	13.5
75% Percentile	18.0	16.0	16.0	16.0
Maximum	20.0	19.0	20.0	20.0
Range	12.0	11.0	12.0	12.0
Mean	14.5	13.0	13.4	13.3
Lower 95% CI	13.4	12.3	12.2	12.1
Upper 95% CI	15.7	13.7	14.7	14.5

Chart 2 and Table 4. Comparison of quality scores for submissions using SP142 companion diagnostic assay with and without method alteration.
Use of the method according to manufacturer's recommendations (CDx) is clearly associated with higher quality scores (dark green box in Chart) compared to submission in which the method has been altered. This difference was statistically significant for CDx compared to the group using incorrect antigen retrieval (AR), P = 0.014, but did not achieve significance in comparisons with the no amplification group (P = 0.173), or those using incorrect AR and omitting amplification (P = 0.117). In the chart the whiskers represent 5 and 95% distribution limits.

Optimal and Sub-optimal PD-L1 Staining Results

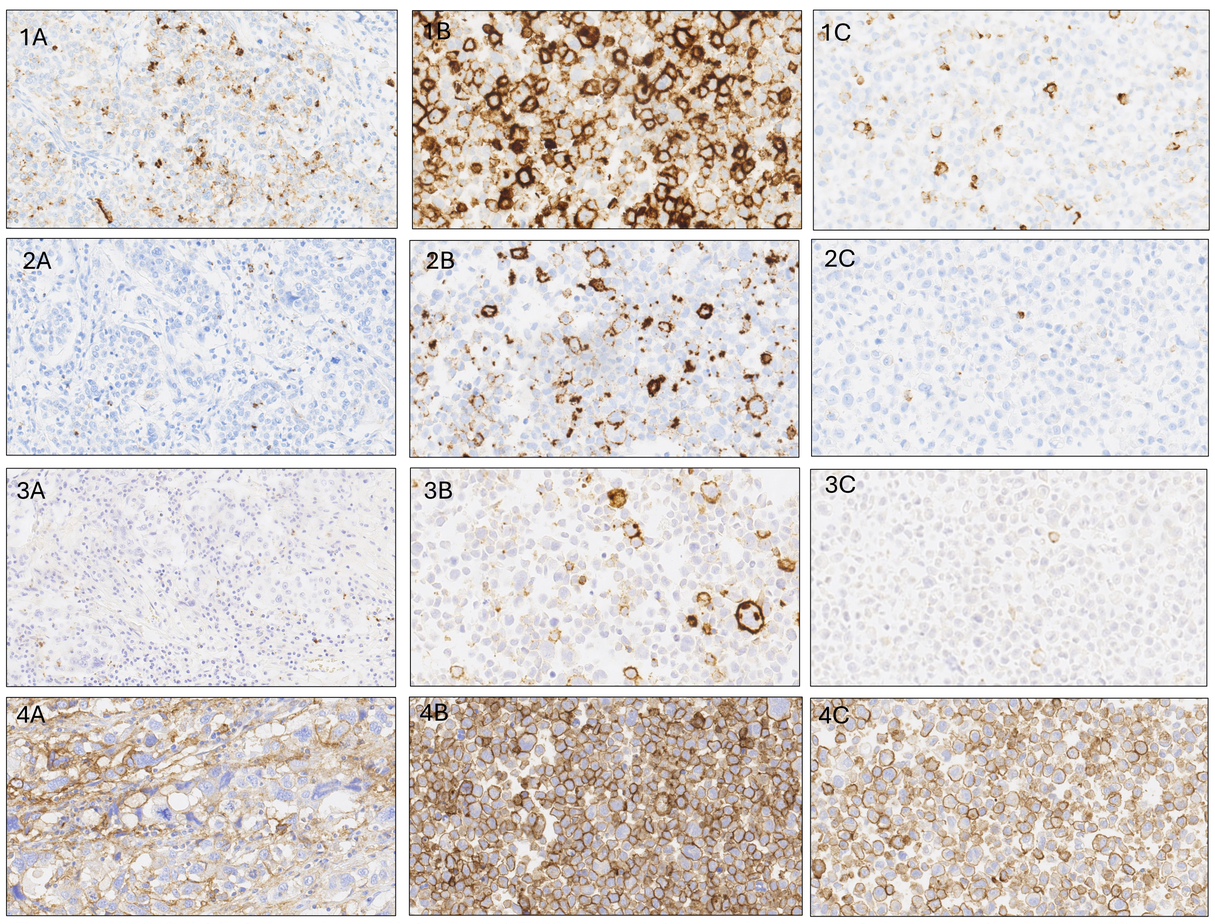


Figure 1: UK NEQAS distributed samples used in the Run 142 TNBC assessment survey, showing expected and sub-optimal levels of PD-L1:
1(A)-1(C) Expected level of PD-L1 staining in the Positive TNBC tissue (A), Positive Cell Line with high expression (B), Positive Cell Line with low expression (C). Stained with the Roche SP142 CDx assay correct protocol.
2(A)-2(C): Sub-optimal staining of PD-L1 in the TNBC positive tissue (A), Positive Cell Line with high expression (B), Positive Cell Line with low expression (C), showing lower levels of staining than expected. The Roche SP142 antibody, used as an LDT with an incorrect antigen retrieval protocol and no amplification.
3(A)-3(C): Sub-optimal staining of PD-L1 in the TNBC positive tissue (A), Positive Cell Line with high expression (B), Positive Cell Line with low expression (C), showing lower levels of staining than expected. The Roche SP142 antibody, used as an LDT with no amplification.
4(A)-4(C):) Expected level of PD-L1 staining in the Positive TNBC tissue (A), Positive Cell Line with high expression (B), 2nd Positive Cell Line with high expression (C). Stained with the Dako 22C3 CDx assay correct protocol.

Conclusions

Analysis of this large dataset has revealed statistically significant differences in the ability of the commonly used primary antibody clones to produce high quality staining.

SP142 when used according to the manufacturer's recommendations consistently out-performed all other clones and methods.

When using SP142, alteration to the method had a detrimental effect on quality scores:

- changes to the antigen retrieval method and/or the omission of signal-amplification produced results that were associated with inferior quality, and in particular with weak staining leading to the potential for false-negative results.

UK Interdisciplinary Breast Cancer Symposium

Poster number: 63. Birmingham, 2024