Immunohistochemical demonstration of oestrogen and progesterone receptors: correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries

A Rhodes, B Jasani, A J Balaton, K D Miller

Abstract

Aims—To investigate the sensitivity of immunohistochemical (IHC) assays for oestrogen receptors (ER) and progesterone receptors (PR) achieved by laboratories on breast tumours fixed and processed in their own department, and to compare this with the degree of sensitivity they achieve on tumours circulated as part of an external quality assessment (EQA) programme.

Methods—On 10 occasions between April 1994 and June 1998, histological sections from breast cancers showing various degrees of expression of ER and PR were circulated for IHC staining to laboratories participating in the UK national external quality assessment scheme for immunocytochemistry (UK NEQAS-ICC). The staining of these tumours, in addition to that of tumours fixed and processed in the participants own laboratories (in house tumours), was assessed by a panel of four assessors, using the established UK NEQAS-ICC scoring system. For a selected assessment run, the degree of expression of participants in house tumours was evaluated by means of the semiquantitative quick score method.

Results—Although the scores awarded for the staining of in house tumours were generally higher than those awarded for the staining of UK NEQAS tumours, there was also a significant positive correlation between the two sets of scores. Using the quick score method of evaluation for one of the assessment runs, 47% of in house tumours were classified as having a high degree of ER expression. Of the remaining cases, a significant proportion initially classified as having only low or medium expression of ER were found to have higher expression when stained by the optimally demonstrating laboratory. The UK NEQAS-ICC centre’s routine assay for hormonal receptors was found to be 90–100% efficient in achieving optimal demonstration of breast tumours from over 150 different laboratories.

Conclusions—The significant positive correlation between the results obtained on the UK NEQAS tumours and the in house tumours provides evidence for the view that results achieved on EQA material are accurate indicators of in house laboratory performance. Although most laboratories adequately detected tumours with high receptor expression, a large proportion of in house tumours classified initially by participants’ staining as being of low or medium ER expression had a higher degree of expression when stained by the UK NEQAS-ICC centre. The efficiency of the organising centre’s routine IHC method for ER and PR in optimally demonstrating participants in house breast tumours shows that variations in fixation and tissue preparation are not limiting factors preventing a different laboratory achieving optimal demonstration.

Keywords: immunohistochemistry; oestrogen receptors; progesterone receptors; external quality assessment

Recent leading articles have emphasised the importance of establishing the oestrogen receptor (ER) status of women with breast cancer. Other articles have reported on the degree of variability that exists between laboratories when demonstrating ER by immunohistochemistry (IHC) on the same cases. The largest of these studies looked at the results obtained by 200 different participants of an external quality assessment (EQA) programme on slides circulated to these laboratories by the EQA scheme and containing tumours with differing degrees of ER expression.

Although many view the results of EQA as a useful gauge of a laboratory’s ability to perform adequate staining for ER and progesterone receptors (PR) on paraffin wax embedded sections, it could be argued that the system suffers from a number of drawbacks. Distributed material is limited and consists of tissue that has been fixed, processed, and prepared under different conditions, no matter how slight, to those used by the participating laboratory. This is thought to be important because there is a view that the IHC assay optimised for use on in house material cannot be expected to produce results of the same quality on tissues fixed and processed in a different laboratory. Consequently, it is thought by some that the quality of IHC achieved on material distributed by an EQA scheme does not reflect
Materials and methods
TUMOURS CIRCULATED BY THE EQA SCHEME
Laboratories participating in the UK NEQAS-ICC programme for steroid hormonal receptors were sent, at each assessment, two unstained slides containing histological tissue sections of formalin fixed and paraffin wax processed breast tumours showing different degrees of hormonal expression. Each participant was asked to demonstrate ER and/or PR and to return the stained slide(s) to the UK NEQAS coordinating centre for assessment of staining quality. Table 1 shows details of the tumours circulated at each assessment from April 1994 to June 1998. Although tested in the organisers laboratory, most of these tumours were fixed and processed in the laboratories of participants, from where they were kindly donated. Whenever possible, tumours that also had their receptor status determined biochemically by the ligand binding assay (LBA) were used.

Table 1 Details of tumours circulated by UK NEQAS-CC for assessments between April 1994 and June 1998

<table>
<thead>
<tr>
<th>Run no.</th>
<th>Date</th>
<th>Number of labs</th>
<th>Tumour type(s)</th>
<th>Receptor</th>
<th>Degree of expression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IHC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cytosol</td>
</tr>
<tr>
<td>26</td>
<td>4/94</td>
<td>37</td>
<td>IDC</td>
<td>ER</td>
<td>100%, +++</td>
</tr>
<tr>
<td>29</td>
<td>2/95</td>
<td>66</td>
<td>IDC (mets)</td>
<td>ER</td>
<td>100%, +++</td>
</tr>
<tr>
<td>32*</td>
<td>12/95</td>
<td>98</td>
<td>Inf./DCIS</td>
<td>ER</td>
<td>100%, ++</td>
</tr>
<tr>
<td>34*</td>
<td>8/96</td>
<td>105</td>
<td>IDC</td>
<td>ER</td>
<td>90-100%, +++</td>
</tr>
<tr>
<td>36</td>
<td>1/97</td>
<td>118</td>
<td>IDC</td>
<td>PR/ER†</td>
<td>90%, +++</td>
</tr>
<tr>
<td>38*</td>
<td>4/97</td>
<td>175</td>
<td>IDC</td>
<td>ER</td>
<td>90-100%, +++</td>
</tr>
<tr>
<td>39*</td>
<td>8/97</td>
<td>176</td>
<td>IDC</td>
<td>PR/ER†</td>
<td>90-100%, +++</td>
</tr>
<tr>
<td>40*</td>
<td>12/97</td>
<td>192</td>
<td>IDC</td>
<td>ER</td>
<td>100%, ++</td>
</tr>
<tr>
<td>41</td>
<td>4/98</td>
<td>178</td>
<td>IDC</td>
<td>ER and PR‡</td>
<td>90%, +</td>
</tr>
<tr>
<td>42*</td>
<td>6/98</td>
<td>205</td>
<td>IDC</td>
<td>ER</td>
<td>100%, ++</td>
</tr>
</tbody>
</table>

All tumours were fixed for 24 hours in 10% neutral buffered formalin. ER/PR expression of tumours by IHC is described in terms of the proportion of invasive nuclei staining (%) and the average staining intensity (+, ++, +++).

Materials and methods
TUMOURS CIRCULATED BY THE EQA SCHEME
Laboratories participating in the UK NEQAS-ICC programme for steroid hormonal receptors were sent, at each assessment, two unstained slides containing histological tissue sections of formalin fixed and paraffin wax processed breast tumours showing different degrees of hormonal expression. Each participant was asked to demonstrate ER and/or PR and to return the stained slide(s) to the UK NEQAS coordinating centre for assessment of staining quality. Table 1 shows details of the tumours circulated at each assessment from April 1994 to June 1998. Although tested in the organisers laboratory, most of these tumours were fixed and processed in the laboratories of participants, from where they were kindly donated. Whenever possible, tumours that also had their receptor status determined biochemically by the ligand binding assay (LBA) were used.

Table 2 Details, as given on the participants returning questionnaire, of the 152 in house breast carcinomas stained and submitted for assessment (run 41)

<table>
<thead>
<tr>
<th>Type of tumour</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma</td>
<td>5</td>
<td>3.3</td>
</tr>
<tr>
<td>Breast carcinoma (unspecified)</td>
<td>66</td>
<td>43.4</td>
</tr>
<tr>
<td>Ductal carcinoma (grade unspecified)</td>
<td>47</td>
<td>30.9</td>
</tr>
<tr>
<td>Infiltrating ductal carcinoma (grade I)</td>
<td>4</td>
<td>2.6</td>
</tr>
<tr>
<td>Infiltrating ductal carcinoma (grade II)</td>
<td>13</td>
<td>8.6</td>
</tr>
<tr>
<td>Infiltrating ductal carcinoma (grade III)</td>
<td>5</td>
<td>3.3</td>
</tr>
<tr>
<td>Colloid breast carcinoma</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Cribriform</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Infiltrating lobular carcinoma</td>
<td>4</td>
<td>2.6</td>
</tr>
<tr>
<td>Intracytic papillary</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Metastatic disease</td>
<td>4</td>
<td>2.6</td>
</tr>
<tr>
<td>Mucinous</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Tubular carcinoma</td>
<td>2</td>
<td>1.3</td>
</tr>
<tr>
<td>Total</td>
<td>152</td>
<td>100</td>
</tr>
</tbody>
</table>

IN HOUSE TUMOURS
In parallel with the immunostaining of slides circulated by the scheme, participants were asked to demonstrate the same receptor on their own in house tumour, and to submit this stained slide for assessment of staining quality, along with two additional unstained tissue sections from the same tissue block. Tables 2 and 3 give examples of the types of tumours submitted by participants and the various fixatives used. These unstained slides were then stained by the UK NEQAS organising laboratory along with its routine workload, utilising its routine methodology and reagents, for the demonstration of hormonal receptors (table 4). They were then coded and filed alongside the participants own immunostaining before routine assessment. The UK NEQAS-ICC organising centre also stained the participants’ in house tumours so that a comparison could be
made between this staining and that produced by the participant on the same case.

**ASSESSMENT OF SLIDES**

An expert panel of four, comprising consultant pathologists and biomedical and clinical scientists, assessed the quality of the IHC independently on a single blind basis, with each assessor awarding marks out of 5 for each of the coded slides. The four individual marks were then added together to give a total mark out of 20. Marks were awarded by comparing the proportion and intensity of tumour nuclei staining in the participant’s slide, to that achieved on duplicate sections of the same cases by the UK NEQAS-ICC organising centre. A total mark > 12 out of 20 indicates acceptable immunostaining and a pass at assessment, a mark of 10–12 out of 20 is considered to be suboptimal and borderline, whereas a total mark < 10 out of 20 is given for staining that is of unacceptable quality and represents a failure at assessment. One of the main criteria by which staining is deemed unacceptable is when < 10% of receptor positive tumour nuclei are clearly demonstrated in a tumour that has been shown by the UK NEQAS-ICC organising centre to express > 10% ER or PR positive nuclei. To ensure assessor concordance, within ± 1 mark, the slides were marked in batches of 20. On the marking of the 20th slide, all scores were read out, when there was a difference of greater than 1 mark between any of the assessors' individual marks, the respective slide was reviewed until a consensus was reached—that is, all the assessors gave the same mark, within ± 1 mark.

For a selected assessment (run 41), the participants in house staining and that achieved by the organising centre on sections from the same tumours was assessed using the semi-quantitative quick score method of evaluation. With this method, the intensity of the immunohistochemical reaction as viewed under the light microscope was recorded as either: 0, negative (no staining of any nuclei even at high magnification); 1, weak (only visible at high magnification); 2, moderate (readily visible at low magnification); or 3, strong (strikingly positive even at low power magnification). The proportion of tumour nuclei showing positive staining was also recorded as either: zero (0), approximately 1–25% (1), 26–50% (2), 51–75% (3), or 76–100% (4). The score for intensity was then added to the score for proportion, giving the quick score with a range of 0–7. In the case of composite blocks (n = 17), the tumour showing the lowest amount of expression was assessed using this method, and where participants had submitted normal breast tissue, these were not submitted for quick score evaluation. Before the assessors’ (AR, BJ) evaluation of these slides, all were randomised using numbers generated by a Microsoft Excel program. Determination of the degree of concordance between the quick scores of the participants' immunostaining of their own in house tumour and the quick scores of the UK NEQAS laboratory’s immunostaining of the same tumour was established using Cohen’s κ coefficient. The proportion of cases that showed the same or a higher degree of expression when stained by the organising laboratory was analysed by means of the χ² test, as was the proportion of cases that showed the same or a higher degree of expression when stained by the participant's laboratory. The degree of assessor concordance when using the quick score method to evaluate stained slides was measured using Goodman and Kruskal’s γ statistic.

**VALIDATION OF ASSESSMENT THRESHOLDS AND OPTIMAL SENSITIVITY**

Of the participating laboratories, six were identified as having published studies clinically
Immunohistochemistry for oestrogen and progesterone receptors

Passed rate refers to the proportion of participants achieving a score > 12 out of 20 (acceptable staining).
NEQ (E) refers to UK NEQAS slides circulated to each participating laboratory.
In house (F) refers to the participants' own in house tumours.

Results

Figure 1  Box plot to show the relation between the UK NEQAS scores achieved by participants on the slides circulated by UK NEQAS (unshaded boxes labelled with run number and the letter “E”) and the scores achieved by the same participants on their own in house control slides (shaded boxes labelled with the run number and the letter “F”). The maximum score attainable was 20 and the minimum score attainable was 4. The bold line across each box indicates the median score. N, number of laboratories participating in each assessment run.

Table 5  Assessment runs April 1994 to June 1998; distributions of differences between the same in house tumours

<table>
<thead>
<tr>
<th>Run no.</th>
<th>Wilcoxon (Z)</th>
<th>p Value</th>
<th>Spearman’s correlation (r)</th>
<th>SE</th>
<th>p Value</th>
<th>Pass rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>−0.189</td>
<td>0.850</td>
<td>0.436</td>
<td>0.152</td>
<td>0.097</td>
<td>73%</td>
</tr>
<tr>
<td>29</td>
<td>−3.612</td>
<td>&lt; 0.0001</td>
<td>0.314</td>
<td>0.126</td>
<td>0.018</td>
<td>73%</td>
</tr>
<tr>
<td>32</td>
<td>−5.705</td>
<td>&lt; 0.0001</td>
<td>0.362</td>
<td>0.149</td>
<td>0.007</td>
<td>60%</td>
</tr>
<tr>
<td>34</td>
<td>−4.962</td>
<td>&lt; 0.0001</td>
<td>0.348</td>
<td>0.156</td>
<td>0.007</td>
<td>60%</td>
</tr>
<tr>
<td>36</td>
<td>−7.333</td>
<td>&lt; 0.0001</td>
<td>0.336</td>
<td>0.160</td>
<td>0.007</td>
<td>70%</td>
</tr>
<tr>
<td>38</td>
<td>−7.990</td>
<td>&lt; 0.0001</td>
<td>0.307</td>
<td>0.099</td>
<td>0.001</td>
<td>68%</td>
</tr>
<tr>
<td>39</td>
<td>−6.464</td>
<td>&lt; 0.0001</td>
<td>0.356</td>
<td>0.069</td>
<td>0.001</td>
<td>73%</td>
</tr>
<tr>
<td>40</td>
<td>−10.705</td>
<td>&lt; 0.0001</td>
<td>0.376</td>
<td>0.068</td>
<td>0.001</td>
<td>64%</td>
</tr>
<tr>
<td>41</td>
<td>−10.345</td>
<td>&lt; 0.0001</td>
<td>0.304</td>
<td>0.068</td>
<td>0.001</td>
<td>56%</td>
</tr>
<tr>
<td>42</td>
<td>−10.769</td>
<td>&lt; 0.0001</td>
<td>0.327</td>
<td>0.064</td>
<td>0.001</td>
<td>43%</td>
</tr>
</tbody>
</table>

Pass rate refers to the proportion of participants achieving a score > 12 out of 20 (acceptable staining).
NEQ (E) refers to UK NEQAS slides circulated to each participating laboratory.
In house (F) refers to the participants’ own in house tumours.
COMPARISON OF THE UK NEQAS SCORES AWARDED AT ASSESSMENT FOR THE QUALITY OF IMMUNOSTAINING ON THE SLIDES CIRCULATED AND THOSE ACHIEVED BY THE SAME PARTICIPANTS ON IN HOUSE SLIDES FROM APRIL 1994 TO JUNE 1998

For nine out of the 10 assessments analysed, the median for the scores that participants achieved on in house tumours was higher than the median for the scores achieved on UK NEQAS tumours (fig 1). The Wilcoxon signed ranks test showed a highly significant difference in the distribution of marks for these nine runs (p < 0.0001; two tailed; table 5). The interquartile range was also frequently smaller for the scores achieved on in house sections, indicating less spread in the results. The proportion of participants achieving acceptable staining was always greater on the in house tissues than on the tumours circulated at assessment. These differences ranged from just 8% for run 26 to 44% for run 42 (table 5). However, Spearman’s test showed a significant positive correlation between the routine scores awarded for the staining of UK NEQAS tumours and the staining of in house tumours. This relation is seen for all the assessments for ER/PR conducted between April 1994 and June 1998 (table 5; figs 2–9).

Figure 2 Results of immunohistochemistry for oestrogen receptors (ER) performed by the UK NEQAS organising laboratory on the low expressing (cytosol assay ER, 10 fmol/mg protein), ER positive, infiltrating ductal carcinoma circulated by UK NEQAS-ICC for assessment run 41.

Figure 3 High power detail of the same section shown in fig 2.

Figure 4 Results of immunohistochemistry for oestrogen receptors (ER) performed by laboratory “X” on the low expressing infiltrating ductal carcinoma shown in figs 2 and 3. The UK NEQAS score awarded to laboratory “X” for this staining was 8 out of 20. Laboratory “X” considered this tumour to be ER negative.

Figure 5 High power detail of the same section shown in fig 4.

Figure 6 Results of immunohistochemistry for oestrogen receptors (ER) performed by laboratory “X” on the high expressing in house tumour submitted by laboratory “X” for run 41 (UK NEQAS score, 16 out of 20, quick score, 4).

Figure 7 Results of immunohistochemistry for oestrogen receptors (ER) performed by the UK NEQAS organising laboratory on the tumour of laboratory “X” shown in fig 6 (UK NEQAS score, 20 out of 20, quick score, 7).
COMPARISON OF THE QUICK SCORES (RUN 41) ON PARTICIPANTS’ STAINING OF THEIR OWN IN HOUSE TUMOURS AND THE QUICK SCORES OF THE SAME TUMOURS WHEN STAINED BY THE UK NEQAS-ICC ORGANISING CENTRE

The number of participants who submitted two unstained slides containing sections of breast tumour, along with their own laboratories’ immunostaining of that tumour, for run 41 was 152 (85% of the total returns). The remaining 26 participants (15%) did not provide unstained slides, or only ones of normal breast tissue. Table 2 details the types of tumours submitted, as described in the returned questionnaires. The initial Wilcoxon test indicated a highly significant difference between the quick scores for the in house tumours when stained by the participant and when stained by the organisation laboratory (Z = −6.814; p < 0.0001; two tailed). Table 6 shows the degree of expression of the 152 tumours as evaluated using the quick score method on both the slides stained by the participants and duplicate slides stained by the organising laboratory. The proportion of cases designated as high expressers was 51.3% (n = 78) by the participants’ staining and 80.9% (n = 123) by the UK NEQAS organising laboratory, with concordance on 72 cases (55.8%; κ coefficient, −0.091; p = 0.043). The proportion of cases designated as medium expressers was 27.6% (n = 42) by the participants’ staining and 12.5% (n = 19) by the organising laboratory, with concordance on seven cases (13.0%; κ coefficient, −0.495; p < 0.0001). The proportion of cases designated as low expressers was 15.8% (n = 24) by the participants’ staining and 4.6% (n = 7) by the organising laboratory, with concordance on just three cases (10.7%; κ coefficient, −0.316; p < 0.0001). Lastly, the proportion of cases designated as negative was 5.3% (n = 8) by the participants’ staining and 2.0% (n = 3) by the organising laboratory, with concordance on three cases (37.5%; κ coefficient, not applicable). Overall, there was agreement on the degree of expression, as defined by the participants’ staining and that of the organising laboratory, in 96 of the 152 cases (63.2%; κ coefficient, −0.026; p = 0.291).

Table 7 details the analysis of the tumours showing less than high ER expression by the participants’ IHC. Of the 42 cases classified as medium ER expressers and having quick scores of 4 and 5, 69% (p = 0.014) were shown to have higher expression when stained by the UK NEQAS organising laboratory and achieved quick scores that were higher by 2 marks or more. Of the 24 cases initially classified as low expressers and having quick scores of 3 or 2, 83% (p = 0.001) were shown to have higher expression when stained by the UK NEQAS organising laboratory, with quick scores that were higher by 2 marks or more (table 7). Lastly, five of the eight cases classified by the participants’ IHC as being ER negative and having quick scores of zero were shown to be ER positive when stained by the UK NEQAS reference laboratory, with one having a quick score of 2, two quick scores of 3, and two quick scores of 6.

EVALUATION OF THE EFFICIENCY OF THE UK NEQAS ORGANISING LABORATORY’S ROUTINE METHOD IN STAINING TUMOURS SUBMITTED FROM PARTICIPATING LABORATORIES

Using the standard UK NEQAS scoring system, for six of the seven assessment runs at

Table 6  The degree of oestrogen receptor (ER) expression of 152 tumours from 152 laboratories participating in assessment run 41, as defined by the participants’ IHC assays and the UK NEQAS organising laboratory’s IHC assay

<table>
<thead>
<tr>
<th>Degree of ER expression</th>
<th>Participant laboratory</th>
<th>Organising laboratory</th>
<th>Level of concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n %</td>
<td>n %</td>
<td>n %</td>
</tr>
<tr>
<td>High expressing tumours (quick scores of 7 and 6)</td>
<td>78 51.3</td>
<td>123 80.9</td>
<td>72 55.8</td>
</tr>
<tr>
<td>Medium expressing tumours (quick scores of 5 and 4)</td>
<td>42 27.6</td>
<td>19 12.5</td>
<td>7 13.0</td>
</tr>
<tr>
<td>Low expressing tumours (quick scores of 3 and 2)</td>
<td>24 15.8</td>
<td>7 4.6</td>
<td>3 10.7</td>
</tr>
<tr>
<td>Negative tumours (quick scores of 0)</td>
<td>8 5.3</td>
<td>3 2.0</td>
<td>3 37.5</td>
</tr>
<tr>
<td>All tumours (quick scores the same, within ±1 mark)</td>
<td>96 63.2</td>
<td>63.2</td>
<td>96 63.2</td>
</tr>
</tbody>
</table>

ICH, immunohistochemistry; NA, not applicable.
which the UK NEQAS organising laboratory stained participants’ in house slides, the median of the scores awarded to the organising laboratory was either equal or greater to the median of the scores awarded to the participants on these same in house slides (fig 10). The interquartile range for the scores achieved by the organising laboratory was smaller than the interquartile range of the participants scores on all seven occasions, indicating less spread in the results. The Wilcoxon signed rank test showed that the distribution of scores awarded to the UK NEQAS organising laboratory was significantly higher overall \( (Z = -6.190; p < 0.0001; \text{two tailed}) \), and individually in four of the seven assessments. For the remaining three runs, there was no significant difference between the two sets of scores (table 8).

Using the quick score method of evaluation, the technique used by the UK NEQAS organising laboratory was 99% efficient \( (p < 0.0001) \) in demonstrating the 152 different tumours submitted by participants for run 41, at either the first or second attempt. The overall efficiency achieved by participants using various different methods was 65% \( (p < 0.0001; \text{table 9}) \).

Discussion

For the accurate assessment of the results achieved by different laboratories participating in EQA it is essential to validate the standards against which optimal sensitivity is defined. In our study, we have sought to validate these standards in various ways. Comparison of the results deemed to be optimal by the organising centre with those achieved by participants of the scheme who are known to have clinically validated their results has revealed many similarities, both in the proportion of the UK NEQAS tumours confirmed to be receptor positive and in the quick scores generated on the in house tumours for run 41. Of the tumours used for assessment by the scheme between April 1994 and June 1998, 81% had been initially tested using both the LBA and IHC. Of these, all were similarly receptor positive or negative with either assay, using a threshold value of 10% or greater of invasive tumour nuclei stained by IHC and 10 fmol/mg protein or greater with the LBA, as designating receptor positive status. Although the use of any threshold value is arbitrary, we have used this cut off point because of its use in several studies that correlate IHC receptor assay results with clinical and biochemical values.\(^{16-20}\)

We have also shown previously that this is the threshold most commonly used by the laboratories participating in UK NEQAS-ICC.\(^1\) Also imperative to our study is the reproducibility of the methods of evaluation used to assess the quality of IHC. The reproducibility of the routine UK NEQAS scoring system was ensured at assessment by the checking of assessor concordance after every 20 slides. A highly significant degree of concordance with the quick score evaluations was confirmed by Goodman and Kruskal’s \( \gamma \) statistic.

Between April 1994 and June 1998, UK NEQAS-ICC conducted 10 assessment runs for ER or PR. During this period, the pass rate on in house tumours remained high (81–97%), whereas that on the distributed UK NEQAS

\[\text{Table 7} \quad \text{The proportion of the 74 participating laboratories in house tumours submitted for run 41 that showed a higher degree of oestrogen receptor (ER) expression* when tested by the UK NEQAS organising laboratory} \]

<table>
<thead>
<tr>
<th>ER expression (participant)</th>
<th>n</th>
<th>ER expression (organising lab)</th>
<th>n</th>
<th>Quick score (participant)</th>
<th>Quick score (organising lab)</th>
<th>%</th>
<th>( \gamma ) Value</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium</td>
<td>42</td>
<td>High</td>
<td>29</td>
<td>5</td>
<td>7</td>
<td>69%</td>
<td>6.095</td>
<td>0.014</td>
</tr>
<tr>
<td>Low</td>
<td>24</td>
<td>High/Medium</td>
<td>20</td>
<td>3</td>
<td>&gt; 7</td>
<td>83%</td>
<td>10.667</td>
<td>0.001</td>
</tr>
<tr>
<td>Negative</td>
<td>8</td>
<td>High</td>
<td>5</td>
<td>0</td>
<td>&gt; 6</td>
<td>63%</td>
<td>0.500</td>
<td>0.480</td>
</tr>
</tbody>
</table>

*Achieved a quick score that was greater by 2 marks or more.
TABLE 8 The differences in distributions of the routine marks awarded to participants for the staining of their own in house tumours and those awarded to the UK NEQAS-ICC organising laboratory for staining of the same tumour (February 1995 to June 1998)

<table>
<thead>
<tr>
<th>Run no.</th>
<th>Wilcoxon (Z)</th>
<th>p Value</th>
<th>Participant</th>
<th>NEQAS organising lab*</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>-2.079</td>
<td>0.038</td>
<td>93%</td>
<td>100%</td>
</tr>
<tr>
<td>36</td>
<td>-2.752</td>
<td>0.006</td>
<td>97%</td>
<td>98%</td>
</tr>
<tr>
<td>38</td>
<td>-0.105</td>
<td>0.916</td>
<td>93%</td>
<td>98%</td>
</tr>
<tr>
<td>39</td>
<td>-4.416</td>
<td>&lt; 0.0001</td>
<td>94%</td>
<td>97%</td>
</tr>
<tr>
<td>40</td>
<td>-0.606</td>
<td>0.544</td>
<td>99%</td>
<td>97%</td>
</tr>
<tr>
<td>41</td>
<td>-5.982</td>
<td>&lt; 0.0001</td>
<td>97%</td>
<td>99%</td>
</tr>
<tr>
<td>42</td>
<td>-0.026</td>
<td>0.979</td>
<td>87%</td>
<td>90%</td>
</tr>
</tbody>
</table>

*The scores awarded to the UK NEQAS-ICC organising laboratory for immunohistochemical staining of participants in house tumours was performed for research purposes only.

slides, over the range of antigen expression included in the tumours examined.

Although the results achieved by participants on the UK NEQAS tumours were significantly different to those obtained on their own in house tumours, Spearman's coefficient revealed that there was a significant positive correlation between the two sets of scores (table 5). To understand why there is a significant difference, and yet still a significant correlation, it is necessary to consider the relation between IHC assay sensitivity and the degree of receptor expression by the tumours under investigation. Suboptimal IHC assay sensitivity when used to stain ER or PR in a low expressing receptor positive tumour (for example, a UK NEQAS tumour) usually results in <10% of invasive nuclei being stained and a failure at assessment. When applied to a high expressing in house tumour, this same degree of IHC sensitivity is also suboptimal because some invasive receptor positive nuclei that should be demonstrated are not. However, this is unlikely to result in <10% of the tumour nuclei being stained, purely on the basis of the large number of receptor epitopes available. Consequently, participants who fail on the low expressing receptor positive UK NEQAS tumour tend to achieve lower scores than they should on their high expressing in house tumour. However, they do not usually fail (a score <10 out of 20) if the proportion of nuclei demonstrated is equal to, or greater than, the designated 10% threshold. Conversely participants with high IHC assay sensitivity, who achieve a relatively high score on the low expressing receptor positive UK NEQAS tumour, tend to score very high marks on their in house tumour. This relation between assay sensitivity and the proportion of nuclei demonstrated in low expressing ER positive tumours and high expressing ER positive tumours is illustrated in figs 2–9.

The main implication of this correlation is that the IHC sensitivity achieved by laboratories on tumours circulated by UK NEQAS-ICC at assessment is a reflection of the sensitivity that the same laboratories achieve on tumours fixed and processed in their own laboratory (in house tumours). This is the first time evidence has been obtained in support of the view that the IHC results achieved on EQA material are accurate indicators of in house laboratory performance.

It has been shown previously that there is a significant positive correlation between the sensitivity achieved by the same laboratories on tumours of differing expression when these tumours are stained as a composite block. In
our study, we show that there is a similar correlation between suboptimal demonstration of in
house tumours and suboptimal demonstration (< 10% nuclear staining) of relatively low
expressing ER positive tumours circulated by an EQA scheme.

At present, there is a tendency to overlook
the implications of suboptimal staining of
tumours with relatively high amounts of ER or
PR expression. Reiner and colleagues' and
MacGrogan and colleagues' showed that
patients whose carcinomas contained high
numbers of hormone receptor positive cells
(> 30%, > 50%, > 70%) had a better overall
survival than those patients whose tumours
had fewer receptor expressing cells. This was in
general agreement with the results of Barnes
and colleagues and Walker, who showed a high rate of recurrence occurring in
patients whose tumours contained high propor-
tions of ER negative cells. Hawkins has
suggested that it is possible for ER IHC results
to be divided into a minimum of four
categories (negative, low, medium, and high)
and still provide prognostic/predictive infor-
mation similar to that provided, as a con-
tinuum, by a sensitive and quantitative bio-
chemical assay. Our present study found that
69% (p = 0.014) of tumours initially classified
as medium expressers by participants' staining
were subsequently shown to be high expressers
when stained by the organising centre, and that
83% (p = 0.001) of those classified as low
expressers were shown to be medium or high
expressers (table 7). In addition, five of eight
completely ER negative tumours were found to be
ER positive. Two of these tumours were
subsequently classified as high expressers, with
quick scores of 6, and three low expressers,
with quick scores of 3 or 2, although all with 10%
or greater of the tumour nuclei staining.
Obviously, the clinical importance of produc-
ing false negative staining is greater than that of
suboptimally staining relatively high expressing
ER positive tumours. However, all these
tumours were from patients who, according to
the criteria of Reiner and colleagues, Barnes and
Walker, and Hawkins, would have a
better overall survival than that predicted by
the initial in house IHC assays.

This raises the important question as to
whether anything short of optimal immuno-
staining is acceptable for ER testing, the results
of which are likely to influence overall clinical
management. A number of other immunocyto-
chemical markers also appear to fall into this
category, c-erb-2 being one of the best
examples.

Evaluation of the efficiency of the UK
NEQAS reference laboratory's routine tech-
nique in analysing the in house tumours
submitted between February 1995 and June
1998 using the standard UK NEQAS scoring
system gives an efficiency ranging from 90% to
100% (table 8). The in depth analysis of run
41, using the quick score method to evaluate
staining of 152 in house tumours, shows the
organising laboratory's routine method to be
99% efficient in achieving an equivalent or
greater sensitivity to that of the participating
laboratory from where the tumours were
submitted (table 9). This degree of sensitivity
was achieved on the first, or second, attempt.

These results clearly indicate that the
variations in fixation and paraffin wax process-
ing that have been used by participating
laboratories on the tumours submitted for
assessment to date are not limiting factors pre-
venting a different laboratory achieving a simi-
lar or greater degree of sensitivity for hormonal
receptors. This is supported by a detailed
analysis conducted by Williams et al., which
showed that variations in immunostaining as a
result of variations in fixation and processing
regimens could be overcome by heat mediated
antigen retrieval.

The organising laboratory’s standard tech-
nique uses routine commercial antibodies and
reagents—that is, the same antibodies or
reagents used by numerous participants at
assessment and by some of the laboratories that
have clinically validated their results (table 4).

Although some participants use different clones
to the ones used by the UK NEQAS organising
laboratory, the technical comparisons per-
formed after recent assessments do not show
one clone to ER or PR to be significantly supe-
rior to another. The same applies for the
differing secondary detection systems. This
leaves the efficiency of the heat mediated
antigen retrieval step as the most likely factor
preventing some participants from achieving
optimal demonstration of hormonal receptors.
Because all the clones to ER currently used at
assessment, and most of those used to PR,
necessitate heat mediated antigen retrieval for
use on routinely processed tissues, the
degree of sensitivity ultimately achieved with
these clones is directly dependant on how well
the heat mediated antigen retrieval step has
been performed. A multicentre study, involving
15 French laboratories, found the duration of
the antigen retrieval step to be the crucial factor
preventing some of the participating laborato-
ries producing adequate results for ER on
tissues fixed in a different laboratory. Subse-
quently, extension of the heat mediated antigen
retrieval time allowed these laboratories to
achieve optimal results. This supports the find-
ings of our study, which suggest inefficiencies
in the heat mediated antigen retrieval step might
be the most important factor responsible for
poor IHC demonstration of hormonal recep-
tors. It is beyond the remit of our present inves-
tigation to provide an in depth technical analy-
sis of the different variables that affect the
efficiency of heat mediated antigen retrieval. To
date, the few publications on this subject have
mainly restricted their investigations to the effi-
ciencies of the buffers used. Other papers
have compared the efficiency of the various
heating methods available—for example, micro-
wave ovens versus pressure cookers. How-
ever, a comprehensive study is clearly required
to compare the relative merits of all the different
systems, particularly with respect to their
efficiencies in the demonstration of low ER/PR
positive breast tumours that have been fixed
and processed under differing conditions. This
would provide valuable information in the
formulation of recommended technical guidelines for optimal IHC demonstration of ER and PR. In turn, this should help in the standardisation of the technique, increase the sensitivity of detection for some laboratories, and ultimately help ensure that hormonal receptor positive cases are not erroneously reported as hormonal receptor negative.

We thank E Anderson, D Barnes, R Baumann, L Bobrow, V LeDoussal, and R Golouh for providing us with invaluable assistance, and all the participants of UK NEQAS-IHC, without whom this study would not have been possible.

2 Ellis RM, Osborne CK. Oestrogen receptors and breast cancer: it is time for individualised treatment based on oestrogen receptor status. BMJ 1997;314:1843-44.