ORIGINAL ARTICLE

The NHS breast screening programme (pathology) EQA: experience in recent years relating to issues involved in individual performance appraisal

D M Parham, D Coleman, S Kodikara, S Moss, I O Ellis, S Al-sam, N Anderson, L Bobrow, I Buley, C E Connolly, N S Dallimore, S Hales, A Hanby, S Humphreys, F Knox, J Lowe, J Macartney, R Nash, J Patnick, S E Pinder, C M Quinn, A J Robertson, J Shrimankar, R A Walker, C Wells, R Winder, N Patel

J Clin Pathol 2006;59:130–137. doi: 10.1136/jcp.2004.025619

See end of article for authors' affiliations

Correspondence to:
Dr D M Parham,
Department of Pathology,
Royal Bournemouth
Hospital, Castle Lane East,
Bournemouth, Dorset BH7
7DW, UK; consultant.
histopathologist@rbch.nhs.uk

Accepted for publication
27 May 2005

Background: The original role of the National Health Service breast screening programme (pathology) external quality assessment (EQA) scheme was educational; it aimed to raise standards, reinforce use of common terminology, and assess the consistency of pathology reporting of breast disease in the UK.

Aims/Methods: To examine the performance (scores) of pathologists participating in the scheme in recent years. The scheme has evolved to help identify poor performers, reliant upon setting an acceptable cutpoint. Therefore, the effects of different cutpoint strategies were evaluated and implications discussed.

Results/Conclusions: Pathologists who joined the scheme improved over time, particularly those who did less well initially. There was no obvious association between performance and the number of breast cancer cases reported each year. This is not unexpected because the EQA does not measure expertise, but was established to demonstrate a common level of performance (conformity to consensus) for routine cases, rather than the ability to diagnose unusual/difficult cases. A new method of establishing cutpoints using interquartile ranges is proposed. The findings also suggest that EQA can alter a pathologist's practice: those who leave the scheme (for whatever reason) have, on average, marginally lower scores. Consequently, with the cutpoint methodology currently used (which is common to several EQA schemes) there is the potential for the cutpoint to drift upwards. In future, individuals previously deemed competent could subsequently be erroneously labelled as poor performers. Due consideration should be given to this issue with future development of schemes.

The success of the UK National Health Service breast cancer screening programme (NHSBSP) is dependent on a high quality service. This programme required from the outset a commitment from the NHS to a systematic programme of quality assurance involving all stages of the screening process. For pathologists in particular, this required for the first time mandatory participation in an external quality assessment (EQA) scheme. The scheme was introduced nationally in 1990, with the objective of improving the consistency of pathologists reporting breast disease.

'The Royal College of Pathologists' recommendations are not prescriptive with regard to the type of scoring system to be used'

In the late 1990s the Department of Health, in conjunction the Royal College of Pathologists (RCPATH) and the Association of Clinical Pathologists, established a working group to make recommendations for the future of EQA (in histopathology and cytopathology). Interestingly, the RCPATH recommendations¹ are not prescriptive with regard to the type of scoring system to be used. Indeed, because of the variation in the case mix and case difficulty, different methods are in operation in different schemes.² A key problem with EQA in histopathology is defining the correct diagnosis. In most schemes, the consensus answer is regarded as the correct diagnosis. Furthermore, for most schemes only cases that achieve 80% or more agreement among participants with regard to the diagnosis are used to determine the performance scores of candidates. Cases with

less than 80% agreement are rejected for the purposes of performance assessment.

The RCPATH report does not recommend a single minimum score. However, the document did recommend that the best approach was to compare an individual's score with those of their peers. There remains controversy over what can be deemed a poor level of performance in EQA.^{3–4}

Apart from the NHSBSP EQA, which is compulsory for pathologists who participate in breast screening, other histopathology EQAs have been voluntary. This has started to change, and in future participation in EQAs will be required by the General Medical Council as part of a consultant's portfolio to demonstrate revalidation.⁵ Furthermore, because the NHSBSP EQA is about to obtain CPA (Clinical Pathology Accreditation) accreditation (in 2005), the scheme will be able to implement the action points for identified poor performance. This involves a graded incremental response that includes advice, education, and if necessary investigation and intervention.

Here, we document our experience with the scheme in recent years and discuss some of the implications for the future.

MATERIALS AND METHODS

Data for our study were obtained from the NHSBSP EQA. A general description and details of standard operating procedures have been published recently.⁶ In brief, over 50

Abbreviations: EQA, external quality assessment; NHSBSP, National Health Service breast screening programme; RCPATH, Royal College of Pathologists

Table 1 Scoring matrix

Diagnostic category			
Benign	Atypical	In situ	Invasive
3	2	1	0
2	3	2	1
1	2	3	2
0	1	2	3

sets of 12 cases are circulated to pathologists in the UK who report breast pathology over a three month period. Each case comprises one haematoxylin and eosin stained slide representative of the lesion removed from a patient. The cases are submitted by participants for use in the scheme. The participating pathologist independently examines the slides and for each case completes a tick box proforma, which includes their opinion on the diagnosis. Where appropriate (for example, for carcinomas), additional information regarding tumour size and grade can be recorded. Each pathologist has a unique identification code number, which they record on each submitted proforma. These completed proformas are returned to the cancer screening evaluation unit (Institute of Cancer Research, Sutton, UK), where the participants' opinions for each case are collated. This procedure is repeated twice a year.

The proforma returns are analysed after the closure date for slide circulation.

Cases are scored as follows.

Each diagnosis is categorised into one of four possibilities. A maximum of three points is awarded when the participant agrees with the consensus answer. A score of two points is allocated if the participant deviates by one diagnostic group and one point if the participant deviates by two groups (table 1).

Thus, for an invasive carcinoma (consensus diagnosis), a participant would score three points if this was the offered diagnosis or two points if in situ carcinoma was offered. One point would be awarded for atypical hyperplasia and none if a benign diagnosis was tendered.

For each participant, the points for each case are totalled and the results (final score) expressed as a percentage as follows:

$$\text{score (\%)} = \frac{\text{total points}}{(\text{number of cases} \times 3)} \times 100.$$

Scores are rounded to the nearest whole number. For the 12 cases, possible scores are 100 (36/36), 97 (35/36), 94 (34/36), 92 (32/36), etc. The fact that there are no scores of 95, 96, 98, or 99 etc. is an artefact of the scoring system. Very occasionally, a few pathologists do not examine all 12 cases and this can lead to other scores, because the divisor will be less than 36.

Participants receive feedback in the form of a document after each EQA circulation. This document details individual scores in the form of a list of participant code numbers and score obtained, together with consensus diagnosis and specialist opinion of the cases. There is also a regional feedback session at which pathologists meet and are able to review cases retrospectively, together with the consensus diagnosis and specialist opinions.

Cutpoints

These can be established using rank order and percentile methodologies. Both these methods are in use in histopathology EQAs in the UK. Since 2003, the fifth percentile method has been used by the NHSBSP EQA.

In the rank order method, scores are placed in ascending order and the cutpoint established by the number of

participants in the bottom X% (for example, 2.5%). In contrast, in the (per)centile method, the cutpoint is defined as the score below (but not at) a designated value. Centiles are obtained by calculating:

$q = k(n + 1)/100$, where k is the centile and q is the ranked observation. For example, the fifth centile of 149 observations is $q = 5 \times 150/100 = 7.5$. Therefore, the fifth centile is half way between the seventh and eighth observations. If the seventh observation is 97 and the eighth is 107 the fifth centile is 102. Both methods suffer from the problem that the cutpoint is often in the middle of a group, so the cutpoint is taken below that group.

We present the data for the percentile method. The rank order and percentile methods usually give similar results, although very occasionally small differences can occur. When this is experienced this is highlighted in the results tables.

A new method for establishing cutpoints has recently been proposed using outlier statistics (DM Parham. Performance assessment of histopathologists. Dissertation, University of Warwick, 2003),³ and this approach is also shown here. This method was originally derived from the statistical graphing method known as the box plot. In brief, the modified box plot rule uses the interquartile range ($Q3-Q1$) as the basis for setting the upper and lower limits (fences) used to detect outliers. For performance in EQA the lower fence is used to establish the cutpoint and this is defined as $Q1-k(Q3-Q1)$, where k is the outlier coefficient. Typical values of k are either 1.5 or 3 although other values could be used. When $k = 3$ the modified box plot rule is quite conservative at identifying outliers. Indeed, it can be calculated that for scores that follow a normal distribution, if there are 100 participants the probability of erroneously being labelled as an outlier is 0.3% (when k is smaller this percentage increases).

Additional data analysis was undertaken using Windows based software: Microsoft Excel 2000 (Microsoft Corporation) and software accompanying the book *Confidence interval analysis*.⁷

RESULTS

Performance profiles of pathologists in the NHSBSP EQA

Ten circulations of 12 slides were examined between the years 1998 and 2002. Each set was circulated to between 407 and 485 pathologists at six month intervals. Only cases that obtained a minimum of 80% agreement among coordinators with regard to diagnosis are used in the NHSBSP EQA assessment of performance (cases not achieving 80% consensus on the diagnosis are excluded from the measurement of performance). As shown in table 2, in some circulations this meant that up to three cases were excluded.

Table 2 Number of participants and cases in each circulation, with number of cases achieving an 80% consensus

Circulation number	Number of participants	Number of cases (slides)	Number of cases achieving 80% consensus
2002, no. 2	478	12	11
2002, no. 1	485	12	10
2001, no. 2	467	12	12
2001, no. 1	476	12	12
2000, no. 2	466	12	10
2000, no. 1	431	12	12
1999, no. 2	470	12	9
1999, no. 1	472	12	12
1998, no. 2	464	12	10
1998, no. 1	407	12	9

Figure 1 Profile of scores for all participants and all cases in circulation 2002 no. 2. The x axis class boundaries are ≤ 80 , 81–82, 83–84, 85–86, etc.

In our study, the performance scores of pathologists was examined using this measure (retained cases). In addition, performance scores using all cases circulated were calculated.

The most recent circulation examined (2002, no. 2) is typical of the score profile for participants. Figure 1 shows the score distribution (for all cases), and demonstrates a range of scores, with one possible outlier scoring 86. However, there is a larger group of “borderline scores”; this group could be part of a “normal distribution” of scores and the lack of participants with a score of 95 or 96 could result from an “artefact” in the scoring scheme. The interquartile range methodology with both a stringent and non-stringent k value locates one participant below the cutpoint. The rank and percentile methodologies group 10 participants below the cutpoint (table 3).

When the data are examined using the retained cases only, the distribution of scores is tighter (fig 2). Both the rank order and percentile methodologies place a group of 10 participants below the cutpoint. This group is distinct from most of the other results. Because of the tightness of the distribution, it is not possible to obtain an interquartile range.

Figure 2 Profile of scores for all participants and retained cases in circulation 2002 no. 2. The x axis class boundaries are ≤ 80 , 81–82, 83–84, 85–86, etc.

However, using a 10–90th range with an appropriately adjusted (DM Parham. Performance assessment of histopathologists. Dissertation, University of Warwick, 2003)³ stringent k value, this method also identifies a similar group of participants below the cutpoint (table 4).

Tables 5 and 6 summarise the number of participants failing to make the cutpoints by the different methodologies for all cases and retained cases, respectively. Overall, there is no consistent difference between the numbers failing to make the cutpoints using the different methods and comparable thresholds.

Table 7 shows the average score for coordinators and non-coordinators for all cases; the difference in score between the two groups is an average of one percentage point.

Educational value of EQA

Between EQA circulations 1998 no. 1 and 2002 no. 1, inclusive, 199 new participants joined the EQA scheme. Data were obtained over subsequent EQAs for each participant. To preserve anonymity, their first result was zeroed and subsequent results were expressed either as an increase or

Table 3 Effects of different cutpoint methods on circulation 2002 no. 2 (all cases)

Method	Calculation	Cutpoint	No. below cutpoint	Comment
Percentile	2.5th percentile	Score 97	10	10 scored 94 or less, next 52 scored 97 As above
Percentile	5th percentile	Score 97	10	
IQR (non-stringent k)	$Q1 - (3 \times IQR)$	Score 88	1	
IQR (stringent k)	$Q1 - (1.5 \times IQR)$	Score 92	1	

IQR, interquartile range.

Table 4 Effects of different cutpoint methods on circulation 2002 no. 2 (retained cases)

Method	Calculation	Cutpoint	No below cutpoint	Comment
Percentile	2.5th percentile	Score 97	10	10 scored 94 or less, next 52 scored 97 As above
Percentile	5th percentile	Score 97	10	
IQR (non-stringent k)	$Q1 - (3 \times IQR)$	–	–	
IQR (stringent k)	$Q1 - (1.5 \times IQR)$	–	–	
10–90th range (non-stringent k)	10th centile – $(1.34 \times 10\text{–}90\text{th range})$	Score 93	1	
10–90th range (stringent k)	10th centile – $(0.55 \times 10\text{–}90\text{th range})$	Score 95	10	

IQR, interquartile range.

Table 5 Number failing to make cutpoint in each EQA circulation by the different methods (all cases)

	2.5th percentile	IQR (k = 3)	5th percentile	IQR (k = 1.5)
Circulation 2002, no. 2	10	1	10	1
Circulation 2002, no. 1	12	0	16	2
Circulation 2001, no. 2	11	11	24	24
Circulation 2001, no. 1	7	24	24	44
Circulation 2000, no. 2*	1	0	3	0
Circulation 2000, no. 1	10	4	10	10
Circulation 1999, no. 2	12	0	12	12
Circulation 1999, no. 1	11	11	11	26
Circulation 1998, no. 2	5	1	14	2
Circulation 1998, no. 1	8	6	14	37
Total	87	58	138	158

*Equivalent result for corresponding rank order method (2000, no. 2)

EQA, external quality assessment; IQR, interquartile range.

Table 6 Number failing to make the cutpoint in each EQA circulation by the different methods (retained cases)

	2.5th percentile	IQR (k = 3)	5th percentile	IQR (k = 1.5)
Circulation 2002, no. 2	10	1	10	10
Circulation 2002, no. 1	5	2	5	2
Circulation 2001, no. 2; no cases excluded	11	11	24	24
Circulation 2001, no. 1; no cases excluded	7	24	24	44
Circulation 2000, no. 2	6	6	15	42
Circulation 2000, no. 1; no cases excluded	10	4	10	10
Circulation 1999, no. 2	7	7	7	7
Circulation 1999, no. 1; no cases excluded	8	6	14	37
Circulation 1998, no. 2	6	5	14	14
Circulation 1998, no. 1	9	0	11	9
Total	79	66	129	199

EQA, external quality assessment; IQR, interquartile range.

Table 7 Difference between mean score (all cases) of coordinators and non-coordinators

Circulation	Coordinators	Non co-ordinators	Difference between means (95% CI)
1998, no. 1	95.9 (SD = 2.7, 23 participants)	94.9 (SD = 3.6, 384 participants)	1.0 (−0.47 to 2.57)
1998, no. 2	97.3 (SD = 2.8, 25 participants)	96.3 (SD = 3.2, 464 participants)	1.0 (−0.27 to 2.27)
1999, no. 1	98.6 (SD = 2.9, 24 participants)	97.7 (SD = 3.8, 448 participants)	0.9 (−0.61 to 2.49)
1999, no. 2	95.4 (SD = 3.4, 21 participants)	93.4 (SD = 3.8, 449 participants)	2.0 (0.34 to 3.70)
2000, no. 1	99.6 (SD = 1.0, 22 participants)	98.5 (SD = 2.9, 409 participants)	1.1 (−0.08 to 2.34)
2000, no. 2	92.9 (SD = 5.7, 22 participants)	93.4 (SD = 5.6, 444 participants)	−0.5 (−2.9 to 1.90)
2001, no. 1	99.6 (SD = 1.7, 23 participants)	97.1 (SD = 4.5, 453 participants)	2.5 (0.69 to 4.41)
2001, no. 2	100 (SD = 0.0, 23 participants)	99.4 (SD = 1.7, 464 participants)	0.6 (0.05 to 1.05)
2002, no. 1	96.2 (SD = 4.4, 21 participants)	94.9 (SD = 4.0, 464 participants)	1.4 (−0.28 to 3.11)
2002, no. 2	98.3 (SD = 2.2, 23 participants)	98.6 (SD = 2.2, 455 participants)	−0.3 (−1.20 to 0.66)
Overall mean score	97.4 (SD = 3.69)	96.4 (SD = 4.24)	1.0 (0.48 to 1.60)

*CI, confidence interval.

decrease in score. For participants who joined at EQA circulation 2002 no. 1, there would be only one subsequent EQA score to follow.

Table 8 Average change in score in subsequent EQAs compared with the first one (includes difficult cases)

Circulation	Number of participants	Mean improvement in score (% points)	95% CI for mean
2nd	199	0.16	−0.8 to 1.11
3rd	152	0.26	−0.67 to 1.19
4th	124	1.02	0.83 to 3.0
5th	99	1.81	0.69 to 2.91
6th	76	1.09	−0.18 to 2.36
7th	52	0.92	0.66 to 2.42
8th	33	1.0	−0.92 to 2.92

CI, confidence interval; EQA, external quality assessment.

Table 8 shows the improvement in scores for all new participants (based on all cases). Similar results (not shown) were obtained when the analysis was confined to the retained cases only (that is, cases obtaining greater than 80% agreement). In both cases, there was a marginal improvement in performance scores between the first and subsequent EQAs. In most subsequent EQAs the confidence interval for the mean improvement bridges zero.

It was possible to obtain separate data for the new participants with the lowest scores (defined subjectively from examination of the distribution of scores obtained from new participants). For the performance scores based on all cases, over the period examined, there was a group of 11 low scorers. The scores for these individuals were zeroed and their improvement (or deterioration) tracked over subsequent EQAs. Table 9 shows a consistent improvement in the second and subsequent EQA scores compared with their first scores. The results were reanalysed using only the test cases that

Table 9 Improvement in scores for lowest performers* (includes difficult cases)

Circulation	Number of participants	Mean improvement in score (% points)	95% CI for mean
2nd	11	12.27	9.6 to 14.94
3rd	8	10.0	5.79 to 14.21
4th	6	9.5	6.29 to 12.71
5th	5	12	4.4 to 19.6
6th	3	10.67	7.99 to 29.33

*11 participants with the lowest scores of all new participants between 1998 no. 1 and 2002 no. 1.
CI, confidence interval.

Table 10 Improvement in scores for lowest performers* (excluding difficult cases)

Circulation	Number of participants	Mean improvement in score (% points)	95% CI for mean
2nd	12	11.25	9.22 to 13.28
3rd	10	9.1	4.05 to 14.15
4th	8	10.25	7.19 to 13.31
5th	6	12.33	10.62 to 14.04
6th	4	8.5	1.56 to 14.44

12 participants with the lowest scores of all new participants between 1998 no. 1 and 2002 no. 1.
CI, confidence interval.

achieved an 80% agreement among coordinators with regard to diagnosis. Review of performance scores showed a group of 12 low scorers. Table 10 shows the results in subsequent EQAs when these individuals were followed.

Leavers

We analysed the scores of participants leaving the EQA scheme. Information regarding the reason for leaving (for example, retirement, change of employment, change of job description) is not collected by the data centre. Most individuals cease to participate without informing the data collection centre of their intention not to continue. For our study, individuals were deemed to have left the EQA scheme if they had not participated in three consecutive EQA circulations up to circulation 2003 no. 1. The last score for individuals meeting these criteria was obtained. For each EQA circulation the average score for leavers was calculated and compared with non-leavers. As shown in table 11, the average score of leavers tended to be slightly lower than that for non-leavers. However, table 12 shows that there has been no corresponding upward drift in the cutpoint over the circulations under consideration.

Table 12 The cutpoint value for each circulation using the conventional methodology

Circulation	Cutpoint
1998, no. 1	92
1998, no. 2	90
1999, no. 1	89
1999, no. 2	93
2000, no. 1	89
2000, no. 2	90
2001, no. 1	86
2001, no. 2	94
2002, no. 1	92
2002, no. 2	94

The cutpoint in the table is calculated from the restricted set, which would be used for performance monitoring.

Workload data

For EQA circulation 2002 no. 2, information regarding the number of breast cancers that the individual participants report annually was requested. This information was returned by 305 participants. Some pathologists in the scheme report only breast biopsies and resected breast specimens, whereas others are generalists reporting specimens from a wide variety of body sites including breast. This difference in work is reflected in the variation in the number of breast cancers reported (range, 45–900/year). No data were collected on the total workload of individuals. From fig 3 it is apparent that there is no correlation between the number of

Mean no of cancers reported annually	254
SD	145
Minimum	45
Maximum	900
Number of pathologists returning workload data	305

Figure 3 Relation between workload and score.**Table 11** Comparison of scores of leavers and non-leavers in different EQA circulations

EQA circulation	Mean score leavers	Mean score non-leavers	Difference	95% CI for the difference between means
1998, no. 1	95.75 (8)	97.18 (399)	-1.43	-3.43 to 0.58
1998, no. 2	99.18 (11)	98.71 (453)	0.47	-1.26 to 2.20
1999, no. 1	94.23 (17)	97.86 (455)	-3.61	-5.41 to -1.80
1999, no. 2	99.81 (21)	99.65 (449)	0.16	-0.54 to 0.86
2000, no. 1	97.18 (11)	98.56 (420)	-1.38	-3.07 to 0.31
2000, no. 2	99.3 (20)	98.76 (446)	0.54	-0.68 to 1.76
2001, no. 1	96.00 (24)	97.29 (452)	-1.29	-3.12 to 0.55
2001, no. 2	98.89 (27)	99.51 (460)	-0.63	-1.28 to 0.03
Mean score	97.73	98.46	-0.73	-0.74 to -0.72

Numbers in parenthesis refer to the number of participants.
CI, confidence interval; EQA, external quality assessment.

breast cancers reported and an individual's performance score.

DISCUSSION

The results obtained in the EQA do not necessarily reflect the routine performance of pathologists in the breast screening programme or an individual breast screening unit. EQA cases may be assessed by some pathologists in a more diligent fashion than routine cases. It is also unusual for a pathologist to make a diagnosis solely on the basis of a single haematoxylin and eosin section, or not to have access to clinical information or the opportunity to take additional sections or to perform special stains that could aid in diagnosis. In addition, as part of the breast screening programme, cases are reviewed at multidisciplinary meetings, and the final classification takes account of this interaction and availability of additional information. Despite this false nature of the EQA environment, the methods used by the UK NHSBSH EQA are typical of many other histopathology schemes, and benefit from a simple standard methodology, which appears to be acceptable to participants.

During the five year period of study, 11 of the 120 cases used in the EQA circulations did not achieve an 80% consensus among more than 400 participants. This is not unexpected. In an Italian study using 81 breast cases, only 27% achieved a 100% consensus as to the diagnosis (using a similar four group classification) among 16 histopathologists.⁸ However, the remit of our study was not to analyse the error rate in histopathology.

"A fundamental issue with these tests is setting the standard between acceptable and unacceptable levels of performance, and this should be taken in the context of what is achievable and practicable, not setting a cutpoint that is too high or too low, but realistic"

The EQA reveals the consistency of reporting among the participants and gives a range of scores for pathologists reflecting their closeness to the peer group. However, it does not measure expertise: although there is a tail of scores representing those who perform less well (with regard to consensus agreement), there is little separation or discrimination between most pathologists. Furthermore, the difference (1%) in the scores obtained between the coordinators (who have a special interest in breast pathology) and other participants was small (although this suggests there may be room for improvement). However, this relative lack of difference could result from the fact that the EQA policy is that the consensus answer should be deemed the correct answer; this may not necessarily be the same as an "expert answer". A negative aspect of this policy is that although interobserver consistency may be reassuring there is a risk of groupthinking⁹—the participants may all agree but be wrong (equally an expert opinion is not necessarily always correct!). This indicates that the EQA is not a measure of expertise but a measure of a common level of attainment (conformity to consensus). To balance this risk before formal publication of the results each case is reviewed by the coordinating group midway through a round, when approximately 50% of participants' scores are available for evaluation. This review serves to give feedback to participants on the issues deemed relevant for each case. In addition, in the very rare circumstances where a case is believed to have been classified incorrectly by the majority, the case can be withdrawn from individual performance appraisal.

The literature regarding the setting of cutpoints for satisfactory performance is well established and has been reviewed by Cizeck.¹⁰ There is no gold standard method for

Take home messages

- Pathologists who joined the National Health Service breast screening programme external quality assessment (EQA) scheme showed improvement over time, particularly those who did less well initially
- There was no obvious association between the number of breast cancer cases that participants report each year and performance, probably because the EQA does not measure expertise, but was established to demonstrate a common level of performance (conformity to consensus) for routine cases, rather than the ability to diagnose unusual/difficult cases
- We propose a new method of establishing cutpoints using interquartile ranges, which can help identify poor performers
- Those who leave the scheme (for whatever reason) have marginally lower scores

establishing cutpoints. Jaeger compared 32 different methods and found little agreement regarding the level of the cutpoint established.¹¹ This study also found differences in the numbers failing to reach the cutpoints by the different methodologies examined (the variation between different circulations could result from differences in the degree of difficulty of the slides used in each circulation).

A fundamental issue with these tests is setting the standard between acceptable and unacceptable levels of performance, and this should be taken in the context of what is achievable and practicable, not setting a cutpoint that is too high or too low, but realistic. As stated by Kane *et al*, the standard should be high enough to provide protection for the public while not unnecessarily restricting the supply of qualified practitioners or excluding competent workers.¹² The precision and objectivity of the cutpoint is crucial. In the UK, the present method of defining poor performance (and hence cutpoints) in histopathology EQA has received much criticism in the professional press.^{3, 13} Failing the bottom 2.5% or 5% in each EQA is arbitrary and may not be appropriate, because it does not take into account the distribution of scores and what is achievable or practicable. If one assumes that most participants within the EQA are competent, and that the cutpoint should be based on the performance profile of the majority, the use of the interquartile method based on the scores obtained from the middle majority has a reassuring logic and rationality. This method would seem to solve many of the problems associated with the current methodology, as suggested elsewhere.³

The principal function of EQA when first introduced was educational, derived through the content and personal feedback, the aim of the EQA scheme being to raise standards. The potential educational value of EQA schemes was recognised by the Department of Health and the RCPATH.¹⁴ Evidence to support this assertion to date is limited. Data from our study have shown that the educational value for most participants is marginal when considering the overall diagnosis (for example, benign versus malignant), and most noticeable after completing three circulations. Nevertheless, it has been shown that in specific areas (such as tumour grading) improved performance can be demonstrated by the EQA.¹⁵ However, an important feature demonstrated by our study is that individuals who perform less well in their first EQA improve in subsequent EQAs. This supports the key aim of EQA of raising standards.

This improvement could be educational, through test content itself or subsequent self education. There are other possibilities for the improved scores and these include better familiarisation with the EQA scheme, the marking procedures, or the statistical phenomenon known as "regression to the mean".

A potential unintended consequence of an EQA scheme is that, rather than striving to improve their performance, individuals with low scores would withdraw from the scheme. This possibility has not been examined in detail in the past. In our study, the scores of those leaving the scheme were, on average, marginally lower than those who remained in the scheme. Although there is the potential, with the cutpoint methodologies presently used, for an upward drift in the cutpoint, we have not observed such a drift in the scheme so far, possibly because of the tendency of new participants to have lower scores. If there were to be an upward drift then participants who originally performed adequately would subsequently risk being caught. Although this may not be a bad scenario—"inadvertently" raising minimum standards—it could be viewed as an inappropriate means of obtaining this gain. It should be noted that these data cover all leavers.

Unfortunately, no information was available as to the reason that these individuals left. Some would leave because of retirement or job changes, but a few could have left because of their low score in the scheme. This issue is more likely to become more problematical as the action points for poor performance start to become implemented from 2005. Because the interquartile methodology³ uses cutpoints based on the performance of the middle majority, an upward drift in the cutpoint is less likely to occur if those with low scores withdraw from the scheme.

It is widely believed that workload (both excessive and insufficient) in histopathology may affect performance.¹⁶ In our study, there was no relation between workload in breast pathology and performance in the EQA. However, no data were available on the total workload undertaken by individual pathologists. It is therefore not possible to comment on the relation of specialised workload (or the effect of total workload) on performance in the EQA. Nevertheless, as noted earlier, the EQA is not a measure of expertise; it is designed to assess competence in routine pathology, not rare unusual or difficult cases. Therefore, there may be no relation between the number of breast specimens a pathologist sees annually and their performance in the EQA. However, this needs further investigation.

The effect of specialist workload on performance would be an important area to investigate because there is debate as to whether pathologists should remain specialists or generalists.^{17–19} Knowledge of such information would have important implications for the shaping of service provision in the future. At present in the UK most pathologists are generalists. Most hospitals have a few general pathologists covering all areas of the work. If it were shown that specialist reporting influenced EQA scores positively, this would be a driver for more specialist (and fewer general) pathologists. This in turn would require an "exponential" increase in the number of pathologists with the possible concentration of specialists to fewer sites.

In conclusion, this study has highlighted some interesting phenomena occurring among participants in the EQA scheme. As yet, there has been no formal audit of participant's experience within the NHSBSP EQA. Further data need to be collected regarding the reason for participant withdrawal from the scheme, and detailed examination of the effect of total workload and specialist workload on performance needs to be undertaken. These findings could have implications for shaping the provision of pathology services in the UK.

Authors' affiliations

D M Parham, Department of Pathology, Royal Bournemouth Hospital, Bournemouth, Dorset BH7 7DW, UK
D Coleman, **S Kodikara**, **S Moss**, Institute of Cancer Research, Cancer Screening Evaluation Unit, Sutton, Surrey SM2 5NG, UK
I O Ellis, Department of Histopathology, Nottingham City Hospital, Nottingham NG5 1PB, UK
S Al-sam, Department Histopathology, Broomfield Hospital, Chelmsford, Essex CM1 7ET, UK
N Anderson, Department of Pathology, Royal Group of Hospitals Trust, Belfast BT12 6BA, Northern Ireland, UK
L Bobrow, **S E Pinder**, Department of Histopathology, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK
I Buley, Department of Histopathology, The John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK
C E Connolly, Pathology Department, University College Hospital, Galway, Ireland
N S Dallimore, Department of Histopathology, Llandough Hospital, Penarth, S Glamorgan CF64 2XX, Wales, UK
S Hales, Department of Pathology, Countess of Chester Hospital, Chester, Cheshire CH2 1UL, UK
A Hanby, Department of Histopathology, St James' University Hospital, Leeds LS9 7TF, UK
S Humphreys, Medical Solutions, Harley Street, London W1G 9QP, UK
F Knox, Department of Histopathology, Wythenshaw Hospital, Wythenshaw, Manchester M23 9LT, UK
J Lowe, **J Shrimankar**, Department of Histopathology, University Hospital of North Tees, Hardwick, Stockton on Tees TS19 8PE, UK
J Macartney, Department of Histopathology, University Hospitals Coventry and Warwickshire, Coventry, West Midlands CV1 4FH, UK
R Nash, Department of Cellular Pathology, St Georges Hospital, Tooting, London SW17 0QT, UK
J Patnick, NHS Cancer Screening Programmes, Sheffield ST10 3TH, UK
C M Quinn, Department of Histopathology, St Vincent's University Hospital, Dublin 4, Ireland
A J Robertson, Department of Pathology, University of Dundee, Dundee, Tayside DD1 9SY, UK
R A Walker, Department of Cancer Studies and Molecular Medicine, University of Leicester, Leicester Royal Infirmary, Leicester LE1 5WW, UK
C Wells, Department of Histopathology, St Bartholomew's Medical School, London EC1A 9DS, UK
R Winder, NHS Breast and Cervical Screening Programmes, Sheffield ST10 3TH, UK
N Patel, Department of Histopathology, Royal Sussex County Hospital, Brighton, East Sussex BN2 5BE, UK

The authors of this paper are members of the external quality assessment (EQA) scheme management group of the UK national coordinating committee for breast pathology, which is responsible for pathology quality assurance in the UK National Health Service breast screening programme and preparation of minimum dataset standards in breast cancer pathology for the Royal College of Pathologists. The committee acts as the steering committee for the UK national breast screening histopathology EQA scheme.

Membership of the UK national coordinating committee for breast pathology: Dr Al-Sam, Dr N Anderson, Dr L Bobrow, Dr I Buley, Mr D Coleman, Professor C E Connolly, Dr N S Dallimore, Professor I O Ellis, Dr S Hales, Professor A Hanby, Dr S Humphreys, Dr F Knox, Mrs S Kodikara, Professor S Lakhani, Dr J Lowe, Dr J Macartney, Dr S Moss, Dr R Nash, Dr D Parham, Mrs J Patnick, Dr S E Pinder, Dr C M Quinn, Dr A J Robertson, Dr J Shrimankar, Professor R A Walker, Dr C A Wells, and Mr R Winder.

REFERENCES

- 1 **Royal College of Pathologists**. Recommendations for the development of histopathology/cytopathology external quality assurance systems, 1998. London: Royal College of Pathologists.
- 2 **Coghill S**. EQA in histopathology and cytopathology [letter]. *The Bulletin of the Royal College of Pathologists* 1998;**104**:39–40.
- 3 **Parham DM**. Defining PPP in EQA. *Pathology EQA schemes revisited. The Bulletin of the Royal College of Pathologists* 2003;**124**:7–9.
- 4 **Macartney JC**. An update on EQA in histopathology and cytology. *The Bulletin of the Royal College of Pathologists* 2003;**124**:9–12.
- 5 **General Medical Council**. *Good medical practice*, 2001. London: GMC.
- 6 **NHS BSP Publication No 57**. External quality assessment scheme for breast screening histopathology. General description and standard operating procedures. NHS Cancer Screening Programmes 2003 (<http://www.cancerscreening.nhs.uk>).

- 7 **Altman D**, Machin D, Bryant T, *et al*. *Statistics with confidence*, 2nd ed. London: BMJ Publishing Group, 2000.
- 8 **Palli D**, Galli M, Bianchi S, *et al*. Reproducibility of histological diagnosis of breast lesions: results of a panel in Italy. *Eur J Cancer* 1996;**32A**:603-7.
- 9 **Janus IL**. Groupthink. In: Leavitt HJ, Pondy LR, Boje DM, eds. *Readings in managerial psychology*, 4th ed. Chicago: University of Chicago Press, 1989:439-50.
- 10 **Cizek GJ**. *Concepts, methods and perspectives*. London: Lawrence Erlbaum Associates, 2001.
- 11 **Jaeger RM**. Certification of student competence. In: Linn RL, ed. *Education measurement*, 3rd ed. New York: MacMillan, American Council on Education, 1989:485-514.
- 12 **Kane MT**, Crooks TJ, Cohen AS. Designing and evaluating standard-setting procedures for licensure and certification tests. *Advances in Health Sciences Education* 1999;**4**:195-207.
- 13 **Geddy PM**. Poor performance in histology or cytology EQA schemes [letter]. *The Bulletin of the Royal College of Pathologists* 1999;**105**:56.
- 14 **Royal College of Pathologists**. Guidance on substandard performance and participation in diagnostic histopathology and cytopathology external quality assessment (EQA) schemes, April 2002 (<http://www.rcpath.org>).
- 15 **Ellis I**, Coleman D, Wells C, *et al*. Impact of a national external quality assessment scheme for breast pathology in the UK. *J Clin Pathol* 2006;**59**:138-45.
- 16 **Royal College of Pathologists**. Professional standards unit and professional performance panel. Substandard professional performance: guidance for trusts and pathologists, 2002 (<http://www.rcpath.org>).
- 17 **Black-Schaffer WS**, Young RH, Harris NL. Subspecialization of surgical pathology at the Massachusetts General Hospital. *Am J Clin Pathol* 1996;**106**(suppl 1):S33-43.
- 18 **Prat J**. Subspecialization in pathology. *Am J Clin Pathol* 1996;**106**(suppl 1):S43-4.
- 19 **Watts JC**. Is there still a need for the general surgical pathologist? *Am J Clin Pathol* 1996;**106**(suppl 1):S74-6.

BMJ Publishing Group

BMJ

Publishing Group

The BMJ is an equal opportunities employer committed to provide equality of opportunity in all of our functions

London

Closing date
28 Feb 2006

Associate editors: Journal of Clinical Pathology

The *Journal of Clinical Pathology* wishes to appoint two Associate Editors to oversee the peer review and development of its Chemical Pathology and Molecular Pathology sections.

The successful applicant will be driven, dynamic, enthusiastic, willing to work in a team, share ideas, prepared to actively solicit papers/reviews, promote the journal worldwide and invite opinion shapers in their area to contribute to the journal. He/she will have a recognised track (publishing) record in their field and regular contact with colleagues around the world. He/she will be expected to increase the number of submissions to the journal in their field and to commission and accept at least 4 reviews on hot topics in their area each year. Capacity to devote 1-2 hours/day to the journal, willingness to attend policy meetings in London at least once a year and experience of online submission and peer review systems are essential. Full training on the journal's online peer review system will be provided. Previous editor/editorial board experience is desirable but not essential. Tenure of the associate editor position is usually for 1 year and may be extended to a further 2 years by mutual agreement. We offer a small honorarium, based on the number of manuscripts handled, which is paid quarterly.

To apply, please email your CV, a short treatise on why you think you would make a good associate editor, how you would accomplish the aims of the journal in general and the requirements of the Associate Editor position in particular, and provide suggestions on how you could help the *Journal of Clinical Pathology* to become the pre-eminent pathology journal, to the Editor: Professor Runjan Chetty, Director of Surgical Pathology, University of Toronto (Runjan.Chetty@uhn.on.ca; tel: 001-416-946-2876).

Closing date for applications: 28 February 2006.