The Journal Of The International Federation Of Clinical Chemistry And Laboratory Medicine

eJIFCC

INTERPRETATION IN CLINICAL BIOCHEMISTRY: AN EXTERNAL QUALITY ASSURANCE SCHEME

Gordon Challand and Jacqui Osypiw
Consultant Clinical Biochemists
Department of Clinical Biochemistry,
Royal Berkshire Hospital,
Reading, Berkshire RG1 5AN, UK

The need for Departments of Clinical Biochemistry and of Laboratory Medicine to provide both pre- and post-analytical advice has never been greater. With increasing specialisation of Hospital Clinicians; with an increasing number of investigations being requested directly from Primary Care; and with an ever-increasing range of specialised tests being available, there are many possibilities of inappropriate investigation and of incorrect interpretation of unfamiliar tests by clinical staff. Obvious examples are a conclusion of digoxin toxicity from a sample taken too soon after the last dose of digoxin or of a low serum transferrin being interpreted as synonymous with iron deficiency. Our own Department receives some 2000 samples a day (over half of which come from Primary Care), from which more than 1000 reports contain significant abnormalities. It is totally impracticable for our Department to contact requesting Clinicians directly about every report that contains unexpected abnormalities or that may be liable to misinterpretation.

Today, most Departments of Clinical Biochemistry in the UK use a Duty Biochemist. Duty Biochemists (whether medically or scientifically qualified) are usually senior members of laboratory staff, working on a rota. Their function is to scan reports containing abnormal results (typically many hundreds each day), and identify reports containing unexpected abnormalities. 'Unexpected' can be in the context of previous results on the same patient, or abnormalities at variance with the clinical information given about the patient. Faced with such a report, the Duty Biochemist has several options: to let the report go without further action; to ask for analyses to be checked if there is doubt about quality control; to add further tests in the hope of eliciting further useful information; to telephone the responsible clinician; to visit the ward; or to add an interpretative comment to the report (1).

Thirty years ago, few Duty Biochemists added an interpretative comment to a report, but with increasing workload and increasing clinical specialisation, adding an interpretative comment to reports has become common. However few Duty Biochemists have received formal training in adding interpretative comments, they tend to work in isolation with little or no feedback from users, and often never become aware of clinical outcome.

Clinical Pathology Accreditation (UK) Ltd (the national accreditation organisation in the UK) produced, in 1992, standards required for Departments seeking accreditation: standard D4 stated *Interpretive reports are accurate, comprehensive and clinically relevant* and added *Reports should be subject to regular audit*. For Clinical Biochemistry, there was no external procedure available to check compliance with this standard.

The idea of circulating a set of results containing abnormalities and asking colleagues what interpretation they would add was conceived in 1997: the internet-based general discussion mailbase of the Association of Clinical Biochemists (acb-clin-chem-gen@jiscmail.ac.uk) was an ideal forum for this (2). The first 'Case for Comment' and the summarised responses is given in Table 1. We had assumed that a clear consensus would emerge on an appropriate response, but even on what was thought to be a straightforward case, there was a considerable disparity of response and opinion, and none of the comments matched the one produced by this Department*: 'Some proteinuria. Low volume of very concentrated urine: adequate water intake?'*

Table 1: the first Case for Comment

A female aged 21, an inpatient on a maternity ward.

24 hr urine volume:	709 ml
Urine creatinine:	14 100 µmol/L
Urine protein:	0.94 g/L
Serum creatinine:	52 µmol/L
Creatinine clearance:	136 ml/min

There were 24 replies

6 participants would make no comment, 2 would seek further clinical information before reporting.
3 would add further tests before reporting (2 urate, 1 urine glucose, 1 albumin excretion rate).
10 would suggest further tests (6 check blood pressure, 2 ask for a further 24 hr urine collection for protein, 1 check for diabetes).
11 comments mentioned proteinuria.
6 said 'normal clearance for pregnancy',
5 said 'low urine volume, complete collection?',
1 said 'clearance suggests borderline impairment'.

Cases drawn from the workload of this Department were then circulated weekly, and initial experience showed that there was seldom a consensus opinion. Even when there was a consensus, this often did not reflect the actual outcome of the case and there was a bewildering range of suggestions for follow-up. More seriously, on every case, some of the comments made were felt to be inappropriate or incorrect.

In 1998, participation in these Cases was accepted for Continuing Professional Development in the United Kingdom by the Royal College of Pathologists, but it was felt that specific guidance should be given to participants on which comments were most or least appropriate. There appeared to be five main methods of doing this:

a. **The outcome of the Case.** At the time a Duty Biochemist makes a comment, the outcome is not known, and it seemed unfair to judge a comment upon retrospective information particularly when the outcome is a rarity rather than a common condition which might present biochemically in a similar way. In addition, the outcome of many Cases is never known, particularly with patients being investigated in the Primary Care sector.

b. **The consensus comment.** As stated above, there was seldom a consensus comment.

c. **The opinion of the Organiser.** This would be invidious, and no one is an expert on all areas of Clinical Biochemistry.

d. **The pooled opinion of 'experts'.** Selection of appropriate experts would in itself be difficult; and it was felt that introducing 'expert opinion' would have introduced an unacceptably long turn-round time for each Case. Subsequent experience from a similar Australian scheme has shown that even experts can differ in opinion (S Vasikaran, personal communication).

e. **Anonymous peer review of each comment by a panel of assessors.** This seemed the most appropriate way forward. Assessors (each holding Membership of the Royal College of Pathologists or equivalent qualification, equally split between medical and scientific backgrounds and between Teaching and District General Hospitals) were asked to score components of comments on a numerical scale between -2 (highly inappropriate) to +2 (highly appropriate), their mean score giving an overall assessment of value (3). Components were scored to simplify the process because although every comment was different, many contained common components. Table 2 shows an example of scored assessments of components. With these, a short summary was distributed for each Case.

Table 2. Case 21 for Comment

A 25 year old female seeing her GP. Clinical information on the request form is 'dentist says calcium deficient'.

Serum results are

Normal U & E, LFTs.

Total calcium: (2.10 - 2.55)	2.36 mmol/L
Phosphate: (0.81 - 1.45)	0.61 mmol/L
Alkaline phosphatase:	75 IU/L (<126)

Scored components of comments

Recheck calcium and phosphate on a fresh fasting sample (+1.2)

Post-prandial sample? (+0.7)

Low phosphate, significance uncertain (+0.7)

Pregnant? (-0.3)

Measure PTH (-0.8)

Early hyperparathyroidism? (-1.0)

Osteomalacia or osteoporosis? (-1.2)

By 2001, 100 Cases had been distributed through the Internet. Most of these dealt with analytical interpretation, but pre- and post-analytical issues were also covered. There had been more than 400 different individual participants from 29 countries, and 'Cases for Comment' has been translated into Italian, French and Mandarin. The Cases were widely acclaimed for their educational value, and we are aware of them being used for teaching in sites ranging from the South Island of New Zealand through Chengdu in China to Prince Edward Island in the Gulf of St Lawrence, Canada. Despite all this, there was little evidence of improvement in commenting practice, and the number of participants on each Case began to the scheme unmanageable. The decision was therefore made to move to a formal EQAS run through a pre-programmed web page; initial funding was granted by Clinical Pathology Accreditation (UK) Ltd. Anonymised peer review was continued, but using assessment of whole comments rather than comment components, to eliminate the subjective element inherent in breaking down comments into individual components. Assessors were asked to score each comment on the basis of its appropriateness, taking into account the results, the clinical information given, and the intended recipient of the report, so that effectiveness of communication was included in the assessment. A positive score loosely equates to the comment 'adding value' to the report. The scheme was given Pilot UK NEQAS status, and the first distribution using the new format was made in July 2001 (4). A similar scheme has since been established in Australia (5).

Cases are made available fortnightly through the home page of UK NEQAS (www.ukneqas.org.uk). Each participant (protected by individual password) logs on to the web page, sees the Case, and has 2 weeks to make a succinct comment. Assessors then have 1 week to score each comment on a scale from -1 (inappropriate) to +3 (highly appropriate). The mean score given by the assessors to each comment enables ranking of all comments. The organisers then make a summary of the Case available to participants through the web page: an example of a participant's summary is given in Figure 1. This includes the Case, the comment made by the participant, the mean score given to this, the distribution of scores given to all participants, the participant's average score over the previous 6 months, and an outline of the Case which includes examples of low-, median-, and high-scoring comments. Users of our service have agreed with the utility of this marking.

In the first three years, some 70 Cases have been distributed. There are currently around 300 individual and group participants. There have been more than 40 000 visits to the web site. It is widely used as an educational resource (6), and in questionnaires, the scheme has been awarded an 80% rating in terms of its educational value. Some individual participants have been monitored (with their permission), and their scores for each Case have improved over time. The proportion of participants receiving zero or negative scores for each Case has markedly reduced. Nonetheless, doubts have been expressed about the validity of the peer review process used to award marks to each comment.

Peer review was introduced to guide the Organisers on which comments were more or less appropriate. In this process, valid

differences in opinion can occur: these can include differences in interpretation (particularly when there is little or no evidence base); in professional or ethical issues; and in the comparative weight given to 'good' and 'bad' components included within a comment. An otherwise good comment can be ruined by an inappropriate suggestion for follow-up. In addition, often quite minor differences in phraseology can significantly affect the score given to a comment: dogmatic statements of a diagnosis tend to score worse than suggestions of possible diagnoses. This reflects the weight given by assessors to communication as well as interpretational skill. However, clinicians shown the summaries have totally agreed with the ranking given to individual comments quoted in the Case summary.

There is no gold standard regarding the appropriateness of a comment on a Clinical Biochemistry report, nor to what extent the marks given to a series of comments by a participant might be regarded as 'poor performance'. Because of this, poor performance in the scheme is solely defined as active participation (i.e. submitting a comment which then receives a Continuing Professional Development credit) in less than 50% of the distributed Cases. However, even passive participation (looking at the Case and its summary) is of educational value. It is debatable to what extent the scheme might be used to identify poorly performing participants on the basis of the numerical scores allocated to their comments, particularly as there is no way of ensuring that the comments made to an EQAS Case reflect the comments which a participant makes in real life (however, exactly the same criticism can be levelled at conventional analytical EQA). Concerns have been expressed about the scheme's potential uses in a revalidation process and in identifying poorly performing participants. However, there would have to be considerable discussion and widespread agreement with professional and regulatory bodies before any such use could be put into effect. On the positive side, participation in the EQAS is concrete evidence that an individual is submitting himself to an audit process required both for laboratory accreditation and for personal appraisal: as such, it can only be of benefit to the individual and to the community.

Early in 2004, doubts were expressed, through the general discussion mailbase of the Association of the Clinical Biochemists, about the entire utility of a Duty Biochemist scanning reports containing abnormalities. The 'antagonists' felt that the possibility of a Duty Biochemist making a mistake through insufficient clinical information or knowledge of the patient was high; and that scanning each day many hundreds of reports containing abnormalities was a questionable use of the time of highly skilled laboratory personnel, which would be better spent in direct contact with Clinicians. In addition, there is no evidence that a Duty Biochemist improves patient care. The 'protagonists' felt that to maximise most good to most patients, a proactive Duty Biochemist service for both the pre- and post-analytical phases was essential; and in addition suggested that a formal study to establish benefit to patients would be both unethical and impracticable. There is only a little evidence that interpretative comments change clinical practice for the better (e.g. 7) but there is considerable indirect evidence that they are appreciated, for example views expressed through user questionnaires. Neither side of this discussion has questioned the undeniable educational value of such an EQAS. The topic is obviously of major importance to the future role of professional staff working in Clinical Biochemistry and in Laboratory Medicine, and it is possible that a formal debate on the topic may be held at the next meeting of the International Federation of Clinical Biochemistry (Glasgow 2005).

Acknowledgements

Parts of this article have previously been published in the UK in the Association of Clinical Pathologists' News: they are here reprinted by kind permission of its Editor, Emyr Wyn Benbow. Thanks are also due to Jane French and Finlay MacKenzie for managing all operational aspects of the UK NEQAS for Interpretative Comments in Clinical Chemistry, and to UK NEQAS for permission to reproduce the copyright Case Summary included as Figure 1.

References

1: The provision of interpretative comments on biochemical report forms: a personal view. WJ Marshall and GS Challand; Ann Clin Biochem 37: 758 - 763, 2000.

2: Cases for comment, education and audit. GS Challand; JIFCC 10: 53 - 55; 1998

3: Experience with assessing the quality of comments on Clinical Biochemistry reports. Ping Li and GS Challand; Ann Clin Biochem 36: 759 - 765, 1999

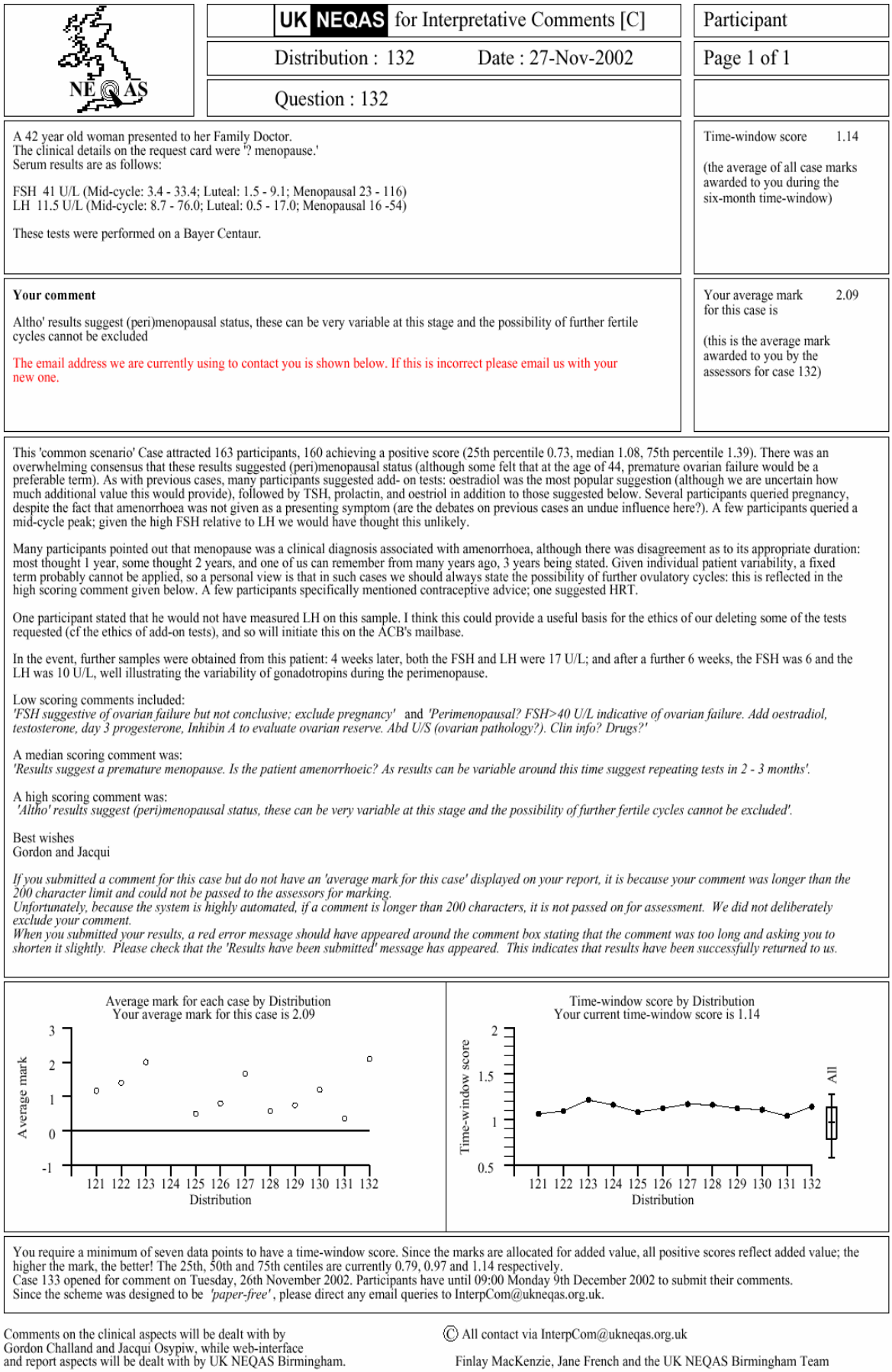
4: Web based external quality assessment of individual skills - the UK NEQAS for interpretative comments in Clinical Chemistry. GS Challand, JC Osypiw, F MacKenzie, JG Middle et al; Abstract S259; Euromedlab, Prague, 2001

5: Quality assessment of interpretative commenting in Clinical Chemistry. EM Lim, KA Sikaris, J Gill, J Calleja et al; Clin Chem 50: 632-637, 2004

6: Cases for Comment on the ACB mailbase: were they educational? JC Osypiw, GS Challand; Proceedings of the ACB National Meeting, Glasgow 2002, A106, p86

7: Do interpretative comments work? MA Giles and GE Curtis. Proceedings of the ACB National Meeting, Birmingham 2004, A93, p116.

Figure 1. A participant's summary sheet from the EQAS



UK NEQAS for Interpretative Comments [C]

Participant

Distribution : 132 Date : 27-Nov-2002

Page 1 of 1

Question : 132

A 42 year old woman presented to her Family Doctor.
The clinical details on the request card were '? menopause.'
Serum results are as follows:

FSH 41 U/L (Mid-cycle: 3.4 - 33.4; Luteal: 1.5 - 9.1; Menopausal 23 - 116)
LH 11.5 U/L (Mid-cycle: 8.7 - 76.0; Luteal: 0.5 - 17.0; Menopausal 16 -54)

These tests were performed on a Bayer Centaur.

Time-window score 1.14

(the average of all case marks awarded to you during the six-month time-window)

Your comment

Altho' results suggest (peri)menopausal status, these can be very variable at this stage and the possibility of further fertile cycles cannot be excluded

The email address we are currently using to contact you is shown below. If this is incorrect please email us with your new one.

Your average mark for this case is 2.09

(this is the average mark awarded to you by the assessors for case 132)

This 'common scenario' Case attracted 163 participants, 160 achieving a positive score (25th percentile 0.73, median 1.08, 75th percentile 1.39). There was an overwhelming consensus that these results suggested (peri)menopausal status (although some felt that at the age of 44, premature ovarian failure would be a preferable term). As with previous cases, many participants suggested add- on tests: oestradiol was the most popular suggestion (although we are uncertain how much additional value this would provide), followed by TSH, prolactin, and oestriol in addition to those suggested below. Several participants queried pregnancy, despite the fact that amenorrhoea was not given as a presenting symptom (are the debates on previous cases an undue influence here?). A few participants queried a mid-cycle peak; given the high FSH relative to LH we would have thought this unlikely.

Many participants pointed out that menopause was a clinical diagnosis associated with amenorrhoea, although there was disagreement as to its appropriate duration: most thought 1 year, some thought 2 years, and one of us can remember from many years ago, 3 years being stated. Given individual patient variability, a fixed term probably cannot be applied, so a personal view is that in such cases we should always state the possibility of further ovulatory cycles: this is reflected in the high scoring comment given below. A few participants specifically mentioned contraceptive advice; one suggested HRT.

One participant stated that he would not have measured LH on this sample. I think this could provide a useful basis for the ethics of our deleting some of the tests requested (cf the ethics of add-on tests), and so will initiate this on the ACB's mailbase.

In the event, further samples were obtained from this patient: 4 weeks later, both the FSH and LH were 17 U/L; and after a further 6 weeks, the FSH was 6 and the LH was 10 U/L, well illustrating the variability of gonadotropins during the perimenopause.

Low scoring comments included:
'FSH suggestive of ovarian failure but not conclusive; exclude pregnancy' and *'Perimenopausal? FSH>40 U/L indicative of ovarian failure. Add oestradiol, testosterone, day 3 progesterone, Inhibin A to evaluate ovarian reserve. Abd U/S (ovarian pathology?). Clin info? Drugs?'*

A median scoring comment was:
'Results suggest a premature menopause. Is the patient amenorrhoeic? As results can be variable around this time suggest repeating tests in 2 - 3 months'.

A high scoring comment was:
'Altho' results suggest (peri)menopausal status, these can be very variable at this stage and the possibility of further fertile cycles cannot be excluded'.

Best wishes
Gordon and Jacqui

If you submitted a comment for this case but do not have an 'average mark for this case' displayed on your report, it is because your comment was longer than the 200 character limit and could not be passed to the assessors for marking.
Unfortunately, because the system is highly automated, if a comment is longer than 200 characters, it is not passed on for assessment. We did not deliberately exclude your comment.
When you submitted your results, a red error message should have appeared around the comment box stating that the comment was too long and asking you to shorten it slightly. Please check that the 'Results have been submitted' message has appeared. This indicates that results have been successfully returned to us.

You require a minimum of seven data points to have a time-window score. Since the marks are allocated for added value, all positive scores reflect added value; the higher the mark, the better! The 25th, 50th and 75th centiles are currently 0.79, 0.97 and 1.14 respectively.
Case 133 opened for comment on Tuesday, 26th November 2002. Participants have until 09:00 Monday 9th December 2002 to submit their comments.
Since the scheme was designed to be *'paper-free'*, please direct any email queries to InterpCom@ukneqas.org.uk.

Comments on the clinical aspects will be dealt with by
Gordon Challand and Jacqui Osypiw, while web-interface
and report aspects will be dealt with by UK NEQAS Birmingham.

© All contact via InterpCom@ukneqas.org.uk

Finlay MacKenzie, Jane French and the UK NEQAS Birmingham Team

Published at 09:45 on Friday 7 May 2004