Experience with external quality control in spermatology

T.G.Cooper1, A.D. Atkinson2 and E.Nieschlag1,3

1Institute of Reproductive Medicine of the University, Domagkstraße 11, D-48129 Münster, Germany and 2UK NEQAS, Sub-fertility Laboratory, First Floor, Old Building, Saint Mary’s Hospital, Whitworth Park, Manchester M13 0JH, UK

3To whom correspondence should be addressed

Results are presented from participation in an external quality control (EQC) programme for semen analysis (UK NEQAS). Formalin-fixed semen samples and videotapes of motile spermatozoa were distributed four times a year over a 3–4 year period. Over the entire period there was close agreement for sperm concentration with, initially, the average of values from the other groups participating in the scheme, and later, values designated as reference values obtained from six laboratories of several chosen that consistently agreed with each other. The initial underestimation of the percentage of normal forms was abolished at the time of change in derivation of designated values and this largely eliminated the difference to establish closer agreement with the designated values. A consistent bias in the assessment of different categories of progressive sperm motility appeared to be resolved by a conscious decision to consider most spermatozoa as grade b and the exceptions as grade a, rather than the converse. Feedback of results to the technicians of the laboratory participating in an external quality control programme leads to reappraisal of subjective evaluation and to harmonization of results between laboratories.

Kew words: andrology/quality control/ring trials/semen analysis/spermatology

Introduction

Semen analysis comprises measurement of the concentration of spermatozoa and the percentages of cells with different morphologies and motilities. Although the former should be accurately assessed by volumetric techniques, the assessment of morphology and motility are more subjective. The appearance of a spermatozoon in semen smears in the light microscope is influenced by the fixation and staining techniques employed as well as by the quality of the microscope used; its lenses, the optics and the set-up of the instrument. Whether a spermatozoon is considered ‘normal’ by a technician is most subjective, although WHO guidelines (1992) are recommended worldwide as a source of information regarding the absolute and relative sizes of certain structures. Sperm motility is influenced by the temperature and the depth and nature of the chamber used to contain the spermatozoa, but again, the assessment is subjective. The concept of non-progressive motility is not uniformly agreed and whether a progressive spermatozoon is ‘fast’ or ‘slow’ can be subjectively influenced by the motion of spermatozoa in the surroundings. All subjective assessments can also be influenced by the mood of the observer.

Considering all the above, it should not be surprising to find differences between laboratories in their assessment of the same semen sample. Indeed, the limited experience with external quality control (EQC) for semen analysis has confirmed that large differences exist even between experienced laboratories in their concepts of normal forms and motility grades (Neuwinger et al., 1990; Matson, 1995; Jørgensen et al., 1997) and this awareness has provided impetus for improving the situation by calls for enrolling in ring trials or external quality control programmes (Cooper, 1996; Michelmann, 1997; Ochsendorf and Beschmann, 1998). Although the aim of this exercise must be the eventual agreement between participating centres, whether this is achieved in practice has not been demonstrated. At a users’ meeting in April 1997, it was clear that in many cases the results of EQC were often retained by the consultant in charge of the laboratory and not discussed with the technicians performing the tests. The results of such comparisons between laboratories indicate to the clinician where the results of his laboratory stand in relation to others, but where differences in assessment exist, changes in assessment of semen characteristics can only occur when technicians are aware of these comparisons.

Materials and methods

About the EQC scheme

The UK NEQAS (United Kingdom National External Quality Assessment Schemes) in Andrology (Scheme organiser Dr D.Critchlow, Scheme Manager Mrs A. Atkinson, address, UK NEQAS, Sub-fertility Laboratory, First Floor, Old Building, Saint Mary’s Hospital, Whitworth Park, Manchester M13 0JH, UK) were set up in 1994. This followed the British Andrology Society (BAS) pilot project (Matson, 1995) which clearly demonstrated the need for and the feasibility of such a scheme for sperm concentration, morphology and antibodies. The UK NEQAS Andrology scheme has been operated from Central Manchester Healthcare Trust laboratories since its inception. All UK NEQAS schemes are supported by a Steering Committee and the Andrology scheme also has a Specialist Advisor Group (SAG) which comprises experts in this field. This SAG meets periodically and operates to support and advise the scheme organiser and manager on such matters as the source and setting of target values, appropriateness of investigations and research and development. The SAG does not concern itself with the performance of individual participants. Information about participating laboratories is strictly
confidential and known only to the scheme staff and the laboratory concerned.

In May 1994 the first semen samples, fixed in formalin, were sent out. On this occasion four laboratories participated. In July 1998 there were 148 participants. The target value used for the first five distributions was an all laboratory trimmed mean (ALTM) but this was deemed inappropriate especially for morphology where the spread of results returned was very wide. Sixteen reference laboratories were chosen on the strength of their consistently good performance over the previous distribution. On each occasion six of these laboratories are randomly chosen and the mean of these results is given as the target value. A noticeable improvement in the results returned has been apparent since this change was incorporated. The method of setting target values will be kept under review and a return to consensus may occur in future if the SAG believes this to be appropriate. Results are returned to participants using histograms and graphs depicting cumulative scores of mean running variance index (MRVIS) and mean running bias index (MRBIS). These graphs are shaded and participants should aim to remain consistently within the unshaded area.

An assessment of sperm motility is an essential part of any semen analysis and the need for a method to assess the quality of results was needed. Given the large numbers involved and the distance from Manchester of many, it would obviously be impossible to send viable spermatozoa to all participating units. The motility and indeed viability of samples on arrival would be inconsistent and would therefore invalidate the scheme. Videotaping (VHS) of samples under phase contrast microscopy was considered to be the most appropriate because some units might not have access to more sophisticated information technology. Although looking at spermatozoa on a television screen is not the same as down a microscope the scheme is enjoying similar success to the concentration and morphology scheme and is attracting new participants. Currently there are 98 units enrolled. Consensus values are used as target values for each of the four categories of motility. Results are returned using histograms, MRVIS and MRBIS graphs as in the concentration and morphology scheme.

Methodology

The Institute of Reproductive Medicine has been enrolled in the UK NEQAS Andrology Quality Control Scheme for sperm concentrations and morphology since December 1994 and in their scheme for sperm motility since December 1995. For sperm concentration and morphology four aliquots of semen containing 100 µl formalin (10%, v/v) per ml were provided four times a year so that a total of up to 63 samples is analysed in this paper. They were handled in the same way as all the other specimens received in the clinic by one technician (but not the same one each time) and assessed as described in the WHO Manual (1992). They were known to be quality control samples since they were provided in different (1.5 ml) tubes than regular samples and were of lower volume. For sperm motility, videotapes of phase contrast images of spermatozoa were provided four times a year (a total of 48 samples) and assessed visually in the same way as microscopic images of normal semen preparations.

Feedback from the UK NEQAS Centre was regularly provided in the form of graphical data and summaries. The designated values of sperm concentration, normal morphology and motility were tabulated and correlated with the values provided by the technicians. Histograms depicting the number of laboratories obtaining certain sperm concentrations, percentage normal forms and percentage of the WHO (1992) motility grades a (‘excellent’), b (‘sluggish’), c (‘non progressive’) and d (‘immotile’) were provided with an arrow indicating in which bin the value from our laboratory was found. The results from the laboratory, together with the designated value, were given as well as bias (difference from designated values), percentage bias, variance index score (indicating the total error/difference from target value calculated from the bias index score), mean running variance index score and mean running bias index score.

The results from each distribution were discussed with the technicians; possible reasons for agreement and disparities were considered and a change in discrimination between categories of normal forms and motility was implemented. Results were also presented to the staff of the Institute at regular meetings.

Results

Sperm concentration

There was excellent agreement between the sperm concentrations arrived at by technicians and the designated values provided by NEQAS (Figure 1). Overall the linear correlation was \( y = 0.986x + 2.42, R^2 = 0.952, n = 62 \).

Sperm morphology

The first three distributions of semen analysed (distributions 3–5, sample numbers 9–20) were assessed by our technicians to contain far lower percentages of spermatozoa with normal morphology than those indicated by the designated values (Figure 2). Values were clustered below the line of identity and the slope of the regression line was almost parallel to the abscissa. In the next three distributions (sample numbers 21–32) values were closer to the designated values with regression lines close to and approaching parallelism with the line of identity. This improvement was maintained in the subsequent six sample distributions (9 to 14, sample numbers 33–56) with results falling both sides of the identity line, although large scatter was evident for some distributions (Figure 2). In the two most recent distributions a good correlation and regression line closer to the line of identity was helped by the large range of morphological forms (Table 1).

Another way of plotting these results is provided in Figure 3, where the absolute differences between the technicians’ assessments and the designated values are given for all samples.
assessed. The general tendency to underestimate morphologically normal forms as we entered the trial is evident, as is the long time it took for values to approach those designated. The marked improvement that occurred with sample 21 reflects the change to a designated value determined from consensus and not from trimmed means. From then the running average of the differences hovered around the designated values in the most recent samples.

**Sperm motility**

The largest discrepancy of the three sperm parameters was found for the rating of sperm motility. Throughout the observation period there was quite close agreement between estimates of progressive cells (grades a + b), those considered motile but non-progressive (grade c) and immotile cells (grade d) and for these regression lines were close to the line of identity, although all showed a tendency to fall below that line at higher motility ratings (Table II). However, there was a large difference in the assessment of spermatozoa graded as a or b. Grade a was grossly underestimated and grade b reciprocally overestimated (Figure 4).

Since our technicians were considering the majority of spermatozoa to be grade b and only exceptionally fast spermatozoa classified as grade a our physicians were aware that a man...
presenting a high percentage of grade a, spermatozoa was out of the ordinary. We have attempted to reverse our concept and consider that most spermatozoa are ‘rapid, progressive’ and that the slow ones are to be graded b. Preliminary observations after this changes have demonstrated a marked improvement in the grading of spermatozoa as a or b (Figure 5).

Discussion

The results demonstrate that good agreement between some laboratories can exist for certain parameters and also that where values are not in agreement initially, a change in assessment can occur when the results of the trial are discussed with the technicians responsible for analysing the semen. There was good agreement between estimates of sperm concentration made by the participating laboratory and the target values given by NEQAS. Since sperm concentrations assessed by volumetric measurement are in good agreement with values determined from flow cytometry (Cooper et al., 1992), the designated values provided by the EQC centre are accurate by implication. Thus the proper use of haemocytometer chamber and pipettes can be used to determine sperm concentration both accurately and reproducibly.

Agreement on the results of assessment of sperm morphology was more of a problem with severe underestimation of normal forms. Nevertheless, with time, and after some delay and after a period of overestimation, a move towards closer agreement with the target values was observed. However, as the major improvement occurred at the same time as the designated values were provided by consensus, rather than from trimmed means of values provided by all participants in the scheme, it is difficult to be certain whether feedback of results to the technicians helped alter the subjective evaluation of sperm morphology. It also emphasizes the responsibility of the EQC centre in maintaining consistent designated values.

The poor agreement on certain aspects of sperm motility has several causes. Assessment of videotapes is clearly a different experience than looking through a microscope eyepiece; nevertheless, for less subjective categories [grade d (immotile spermatozoa), grade c (flagellating but non-progressive spermatozoa) and grades a + b (progressive spermatozoa)] agreement was reasonable over the period of study. The major discrepancy between our technicians’ results and the designated values was in categorizing grades a and b. The reason for this discrepancy is known to be the high threshold of sperm velocities (~70 µm/s) above which our technicians grade a spermatozoon as being grade a (Yeung et al., 1997).

Until recently no guidelines were recommended for the velocities of ‘fast progressive’ (grade a) spermatozoa, although the forthcoming 4th edition of the WHO Manual will suggest the value of 25 µm/s or more recommended by Mortimer (1994) and NEQAS and utilized by MacLeod et al. (1994) and others. Sperm velocities below this value are associated with reduced fertility resulting from insemination of frozen semen (Holt et al., 1989) and fertilization in vitro (Holt et al., 1985). With a view to adopting this recommendation, and to conform with the other laboratories participating in the NEQAS scheme, the concept of making a grade a spermatozoon an exceptionally fast one to considering a grade b spermatozoon an exceptionally slow one helped. Here, the drastic improvement in agreement of our laboratory with the designated values after the change in assessment was not due to changes in the method of generating the designated value.

This preliminary exercise in evaluating the results of participation in an external quality control programme has indicated its usefulness in reducing differences between assessments made by different laboratories. All laboratories currently taking part in such ring trials are encouraged to evaluate their data in a similar way to see if improvements are a consistent feature of such schemes. With such a demonstration of moves to agreement the standing of semen analysis should be raised from that of a much neglected test (Chong et al., 1983) to being on a par with hormone assays. Clinical andrology is still necessary, despite the widespread introduction of ICSI in male infertility treatment (van der Ven and Haidl, 1997). Currently this discipline is being reassessed and new training procedures need to be introduced (Jequier and Cummins, 1997; Patrizio and Kopf, 1997; Tournaye, 1997) which should include basic performance of routine semen work-up and the concepts of internal and external quality control. Together with guidelines for the application of computer-assisted semen analysis to analysis of spermatozoa (ESHRE, 1998), this convergence of views should lead to a uniformity of semen assessment that are necessary for multicentre studies and provide useful information on the predictive values of these semen parameters for fertilization in vitro or in vivo or meaningful prospective studies on secular trends in sperm counts (Nieschlag and Lerchl, 1966).

Acknowledgements

We thank Barbara Hellenkemper, Heidi Beering, Elke Börger, Raphaele Kürten, Sabine Rehr and Katrin Wardecki for the semen analysis. This work was supported by the German Federal Health Ministry (Bonn) and Deutsche Forschungsgemeinschaft Confocal.
Research Grant Ni 130/-15 ‘The Male Gamete: Production, Maturation, Function’.

References


Received on September 3, 1998; accepted on December 4, 1998