Lack of compliance by UK andrology laboratories with World Health Organization recommendations for sperm morphology assessment

Denise Riddell¹, Allan Pacey² and Kate Whittington¹,3

¹University of Bristol, Clinical Science at South Bristol (Obstetrics & Gynaecology), St Michael’s Hospital, Southwell Street, Bristol and ²Academic Unit of Reproductive and Developmental Medicine, Level 4, The Jessop Wing, Royal Hallamshire Hospital, Sheffield, S10 2SF, UK
³To whom correspondence should be addressed. E-mail: K.Whittington@bris.ac.uk

BACKGROUND: Sperm morphology is known to correlate with the probability of conception both in vitro and in vivo, but the assessment of sperm morphology in the laboratory remains problematic. The 4th edition (1999) of the World Health Organization (WHO) Laboratory Manual has attempted to improve matters by giving rigorous recommendations regarding sperm morphology assessment. However, it is unknown how well these recommendations have been implemented in practice. METHODS: A survey of the methods used to undertake the assessment of sperm morphology during semen analysis was undertaken in 37 laboratories in the UK. RESULTS: In total, only two laboratories (5%) were compliant with all current WHO guidelines regarding morphology assessment, including methods of staining and observation, classifying and sampling methods, and the participation in internal and external quality control programmes. CONCLUSION: These results illustrate an urgent need for education and training initiatives to encourage laboratories to become compliant with current WHO guidelines for sperm morphology assessment.

Key words: compliance/morphology/semen analysis/sperm/World Health Organization

Introduction

The analysis of ejaculated semen is the most important means of identifying male infertility [World Health Organization (WHO), 1993]. However, although semen analysis can identify gross sterilizing abnormalities such as azoospermia or globozoospermia, it remains controversial as to how predictive of a man’s fertility the results of a semen analysis actually are when moderate numbers of motile sperm are present.

Although a number of previously conducted studies have shown that the variables of sperm concentration and motility are correlated with the probability of conception following IVF (Kruger et al., 1986; Grow et al., 1994) as well as in unassisted conception in cohorts of pregnancy planners (Bonde et al., 1998; Guzick et al., 2001), it is still controversial as to whether there exists an obvious threshold between fertile and infertile men in terms of their semen quality. This is disappointing because the information would be of immense value to clinicians when advising patients as to the probability of achieving a pregnancy, by whatever mode of conception was considered appropriate.

Interestingly, in their study of 696 fertile and 765 infertile men Guzick et al. (2001) found that although there was significant overlap between the semen variables of the two groups, sperm morphology was the greatest discriminator between them. This is perhaps of no surprise since it has been known for some time that only morphologically normal sperm can pass through mid-cycle cervical mucus (Katz et al., 1990) and the human zona pellucida selectively binds sperm with normal morphology (Liu and Baker, 1992) such that men with severe teratozoospermia frequently have defective sperm–zona pellucida interaction (Liu and Baker, 2003). However, since the pioneering work of MacLeod (1956), it is the techniques of assessment of sperm morphology that have undergone the most change with successive versions of the WHO Laboratory Manual (WHO, 1980, 1987, 1992, 1999) providing increasingly stringent methodological approaches for the identification and classification of sperm with normal morphology. Therefore, with this knowledge it might be assumed that the techniques for sperm morphology assessment were becoming easier to implement and as such providing more robust data for the development of thresholds for use in the clinic.

Interestingly, however, a global survey performed by Ombelet et al. (1997) illustrated the relatively poor compliance of laboratories in performing sperm morphology according to the then relevant WHO (1992) guidelines for sperm morphology. More recently Keel et al. (2002) illustrated the general lack of standardization in laboratories across the USA, with only 85% of laboratories performing sperm morphology assessment as part of semen analysis and of these only 23% using the WHO (1999) criteria for identifying normal sperm. Clearly, therefore, there is
some way to go before the techniques of sperm morphology assessment are implemented consistently around the world.

It was in the knowledge of these observations, in addition to the fact that, anecdotally, laboratory personnel often report the assessment of sperm morphology as one of the hardest parts of semen analysis to perform, that a survey of current laboratory practice within the UK was conducted. The focus of this survey was to determine if the implementation of the methods described in the 4th edition of the WHO Laboratory Manual (WHO, 1999) has indeed reduced the high degree of variability in the techniques used to assess sperm morphology and thereby potentially increased the clinical usefulness of morphology as a predictive indicator of male fertility.

Materials and methods

Personal contact was made by telephone to 53 laboratories requesting their agreement to complete a questionnaire regarding the methods they used to assess sperm morphology. Questionnaires were subsequently sent to these laboratories by e-mail or fax. Of the 53 laboratories, 24 were embryology and/or andrology laboratories either performing semen analysis prior to assisted conception treatments or recognized as specialized laboratories performing semen analysis for secondary as well as primary referrals for infertility assessments and/or treatment purposes. These were termed ‘specialist laboratories’ (SL group). A further 29 laboratories were located in district general hospitals and were performing semen analysis solely for primary referrals. These were termed ‘DGH laboratories’ (DGH group). The questionnaire was designed to obtain information on the staff complement, the number of semen analyses performed, the methodology used for morphology staining and assessment, and the implementation of quality control procedures.

Results

Of the 53 laboratories initially contacted, 37 (70%) returned their completed questionnaires. This comprised 19 ‘specialist laboratories’ and 18 ‘DGH laboratories’. In both groups a total of 124 members of staff was involved in undertaking sperm morphology assessment and their qualifications are shown in Table I. Of the 71 staff performing semen analysis in ‘specialist laboratories’, nine (13%) were scientists with doctorates (PhD), 38 (53%) were qualified embryologists [Association of Clinical Embryologists (ACE) certificate/diploma or equivalent and state registration], 17 (24%) were trainee embryologists, and seven (10%) were biomedical scientists (BMS: degree in biomedical sciences or equivalent and state registration). In comparison, of the 53 staff that performed semen analysis in ‘DGH laboratories’, the majority, 45 (85%), were qualified BMS, four (7.5%) were BMS trainees and four (7.5%) were cytoscreeners (certificate in cytology screening). The higher number of staff in ‘specialist laboratories’ reflects the higher number of semen analyses performed by these laboratories per week [median 30 (range 5–100)] as opposed to a median number of 10 (range 3–40) semen analyses performed per week by the ‘DGH laboratories’.

Staining methods

When asked whether samples were stained prior to morphology assessment, 14 (74%) of the ‘specialist laboratories’ and two (11%) of the ‘DGH laboratories’ reported that they observed unstained preparations. However, two of the laboratories in each group reported that they observed both unstained and stained preparations. Therefore, 25 out of 37 of all laboratories (68%) assessed sperm morphology on stained preparations: seven (37%) of the ‘specialist laboratories’ and 18 (100%) of the ‘DGH laboratories’ (Table II).

When these 25 laboratories were asked about the staining methods they used (Table II), only five ‘specialist laboratories’ and 12 ‘DGH laboratories’ used either Papanicolaou or Diff Quick as recommended by WHO (1999). Therefore 65% of all laboratories, and 22% of laboratories that did stain samples prior to analysis, were following procedures that were not compliant with current WHO guidelines for the staining of semen smears.

Classification of sperm morphology

The WHO (1999) criteria for sperm classification (size parameters) were the most popular, being used by 32 of the 37 laboratories (86%) (i.e. 16 laboratories from each group). The now outdated WHO (1992) criteria were being used by two laboratories: one from each group. Two ‘specialist laboratories’ used the Kruger criteria and one ‘DGH laboratory’ used the Mortimer (1994) classification.

Magnification and sample size

When asked about the microscope magnification used for carrying out a morphology assessment and whether a graticule

| Table I. The grade of staff undertaking sperm morphology assessment in 19 ‘specialist laboratories’ (SL group) and laboratories in 18 district general hospitals (DGH group) across the UK |
|---|---|---|
| Staff grade | SL group | DGH group |
| PhD | 9 (13) | 0 (0) |
| Embryologist | 38 (53) | 0 (0) |
| Trainee embryologist | 17 (24) | 0 (0) |
| Biomedical scientist | 7 (10) | 45 (85) |
| Trainee biomedical scientist | 0 (0) | 4 (7.5) |
| Cytoscreener | 0 (0) | 4 (7.5) |
| Total | 71 | 53 |

Values shown are numbers of staff in each grade with the values in parentheses representing the percentage of the total staff.

| Table II. The proportion of laboratories observing stained and unstained sperm morphology preparations in 19 ‘specialist laboratories’ (SL group) and 18 ‘district general hospital laboratories’ (DGH group) across the UK |
|---|---|---|
| | SL group | DGH group |
| No staining methods | 14 | 2 |
| Papanicolaou or Diff Quick | 5 | 12 |
| Other histological stains | 2 | 6 |
| Total no. of laboratories | 19* | 18* |

Values shown are the number of laboratories in each group reporting that they observed unstained and/or stained morphology preparations.

*Two of the laboratories in each group reported observing both stained and unstained preparations.
(or equivalent) was used to determine sperm dimensions, 32 out of 37 (86%) of the participants completed this section of the questionnaire. Of the 17 ‘specialist laboratories’ that replied, only three (18%) were using WHO (1999)-recommended optics (×1000 magnification with oil immersion and graticule) compared to two (13%) of the 15 ‘DGH laboratories’ that responded to the question. Furthermore, only 30% of laboratories (11 out of 37) reported that they classify ≥200 sperm during morphology assessment with the majority (69%) choosing to classify only 100 (or fewer) sperm.

**Reporting morphology**

Unfortunately, the WHO (1999) Laboratory Manual fails to provide a ‘reference value’ for sperm morphology assessment to assist clinicians in discriminating between fertile and subfertile men. Therefore, laboratories performing semen analysis must provide their own. A total of 32 laboratories (19 ‘specialist laboratories’ and 13 ‘DGH laboratories’) responded to this section of the questionnaire (Figure 1) and reported what reference value they provided to clinicians on their semen analysis report form. In brief, a total of 19 (51%) laboratories (10 ‘specialist laboratories’ and nine ‘DGH laboratories’) use a reference value of 15% normal forms. This is in line with the footnote to Table 1b of the WHO (1999) Laboratory Manual, which suggests that data from assisted reproductive technology programmes indicates that fertilization rates in vitro will decline when sperm morphology falls to <15% normal forms. By comparison, four (11%) laboratories have taken a stricter view with a reference value of <14% or <10% with a further eight (22%) laboratories taking a more lenient view with threshold values of between 20 and 30% normal forms. Finally, one of the ‘specialist laboratories’ included in this survey had identified their own reference range based on their local fertile population. This group reports ‘% oval forms’ and had previously found that ‘<5% oval forms is consistent with fertility impairment as determined from dimensional parameters recorded via scanning electron micrographs (SEM) on donor sperm known to have produced pregnancies within six treatment cycles’.

**Quality control**

A greater proportion of ‘DGH laboratories’ (16 out of 18) participated in an external quality control scheme in comparison to the ‘specialist laboratories’ (13 out of 19). Participation in the scheme organized by UK NEQAS (St Mary’s Hospital, Manchester, UK) was the option all laboratories subscribed to, with one specialist laboratory also participating in the scheme organized by ESHRE. Surprisingly, the levels of compliance with internal quality control were notably lower with under half of all laboratories reporting that they were undertaking internal quality control: eight out of 18 (44%) of ‘DGH laboratories’ and eight out of 19 (42%) ‘specialist laboratories’.

**Cumulative compliance with WHO 1999 Guidelines for Morphology**

Table III illustrates the number of laboratories that are compliant with each of the components of the guidelines for sperm morphology assessment as published by WHO (1999). In summary, when all aspects of staining methods, classification criteria, magnification and objectives and quality control are taken into account, only two of the 37 laboratories surveyed (5%) were compliant with all aspects of the current WHO (1999) guidelines for sperm morphology assessment.

**Discussion**

Although many studies have demonstrated the predictive power of sperm morphology assessment for both the outcome of IVF (Kruger et al., 1986; Enginsu et al., 1991; Grow et al., 1994) and the probability of unassisted conception in vivo (Bonde et al., 1998; Guzick et al., 2001), studies of laboratory practice have shown that sperm morphology assessment is one of the most difficult areas of semen analysis to perform (Matson, 1995; Ombet et al., 1995). As such, the 4th edition of the WHO Laboratory Manual (WHO, 1999) introduced a more rigorous framework for morphology assessment in the

| Criteria                      | SL group (n = 19) | DGH group (n = 18) | Total (n = 37) | Cumulative compliance
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Staining methods</td>
<td>7</td>
<td>18</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Papanicolaou or Diff Quick</td>
<td>5</td>
<td>12</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Sperm dimensions</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td>14</td>
</tr>
<tr>
<td>≥100 oil immersion</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Count ≥200 sperm</td>
<td>6</td>
<td>5</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>External quality control</td>
<td>12</td>
<td>16</td>
<td>28</td>
<td>3</td>
</tr>
<tr>
<td>Internal quality control</td>
<td>8</td>
<td>8</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>No. of compliant laboratories</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

*Running total of how many laboratories are compliant with WHO (1999) guidelines as each criterion is applied. The provision of a reference range on the semen analysis report form has been omitted from this analysis due to the lack of clear criteria within the current WHO (1999) Manual.*
hope that this would ultimately lead to the standardization of laboratory techniques around the world. It is therefore very disappointing, and somewhat surprising, that more than 5 years after the publication of the Manual the survey described here should find that only 5% of the UK laboratories that responded to the questionnaire were compliant with all of the WHO recommendations for sperm morphology assessment. However, this result is similar to that found in a worldwide survey (Ombelet et al., 1997) on compliance with the WHO (1992) criteria and more recently to the situation found in the USA where 15% of laboratories do not even report morphology as part of semen analysis (Keel et al., 2002). The lack of compliance with WHO (1999) methodology described in this report is more than just of academic interest. Each area examined by this survey, if badly implemented in the laboratory, has the potential to undermine the accuracy of sperm morphology assessment and hence result in incorrect diagnostic information being generated for the patient. For example, 43% of the laboratories surveyed in this report indicated that they estimated sperm morphology on the basis of observing unstained preparations. This is in spite of the fact that the pitfalls of doing so had been described previously (Mortimer, 1994) and the fact that the published length and width measurements for human sperm heads are based on the measurement of sperm stained with Papanicolaou (Katz et al., 1986) which are appreciably smaller than live (unstained) sperm. As such, for laboratories to apply the WHO (1999) size–shape definitions to unstained preparations uncritically would undoubtedly lead to some error in their measurements. This criticism can also be levelled at the 22% of laboratories that were observing smears that had been stained using a method that was not recommended by WHO (1999), as the size–shape characteristics of sperm stained by other methods have not been appropriately described and validated.

In addition to selecting the correct staining methods, WHO (1999) also recommends that ×100 oil immersion brightfield objective (with at least a ×10 ocular) should be used to observe the stained smear. Yet only 14% of all laboratories in this survey reported using such optics. Similarly, 69% of laboratories reported that they were observing ≤100 sperm in order to derive a value for sperm morphology. The need for compliance with WHO (1999) recommendations for these parameters is more than a matter of simply ‘following the rules’ as both have the potential to introduce significant sampling error into sperm morphology assessments, although the impact of sample size is arguably more profound. The influence of sample size on the estimation of sperm morphology has recently been the subject of a more detailed analysis by Kuster et al. (2004) and although this paper is primarily aimed at the veterinary field where animal species have higher percentages of morphologically normal sperm in their ejaculates, the theory is equally applicable to human samples as is illustrated in the chapter on quality control (QC) in the WHO (1999) manual. Briefly, the fewer sperm that are counted during a morphology assessment, the wider the corresponding 95% confidence interval of the result obtained. Ironically, more sperm need to be evaluated in order to obtain the same confidence with regard to sperm morphology in a specimen with a low percentage of normal forms, such as exists in human males in comparison to many animal species. WHO (1999) have determined that the counting of 200 sperm is the absolute minimum requirement on which to base a sperm morphology assessment. If an assessment is made on fewer than this number, it is entirely plausible that teratozoospermia warranting the use of ICSI may be diagnosed, whereas in reality the sample is quite normal. How well laboratory scientists understand the effect on the reported result of classifying too few sperm is unclear, but needs to be addressed in training programmes and in future editions of the WHO Laboratory Manual.

A new topic introduced into the WHO (1999) Laboratory Manual was the importance of QC within the diagnostic laboratory. Although both internal and external quality control are now integral to laboratory accreditation (Burnett, 1996), it was surprising to find that although 78% of laboratories were taking part in a external quality assurance scheme, only 43% undertook internal quality control. Although internal quality control in Andrology can be considered cumbersome and difficult to implement (Clements et al., 1995) its benefits are well known and WHO (1999) recommends that between 1 and 5% of samples should be used for internal QC. The success of external quality control schemes is well recognized (Cooper, 1996; Jorgensen et al., 1997; Auger et al., 2000).

The omission of a normal reference range for morphology in Table 1a of WHO (1999) has certainly added to the ambiguity surrounding morphology assessment and further increased confusion as to its clinical relevance. Although the Manual did include a footnote to indicate that ‘as sperm morphology falls to <15% normal forms the fertilization rate decreases’, presumably based on the work of Kruger et al. (1986), the results of this survey (Figure 1) suggest that many UK laboratories have taken a pragmatic decision to cite this value as the reference range for fertile men. Although this decision has now been given some validity following the study of pregnancy planners by Guzick et al. (2001), the usefulness of this decision in the many laboratories that are poorly implementing sperm morphology assessment should be questioned.

Finally, it is interesting to consider whether the differences in the approach of sperm morphology assessment, and therefore compliance with WHO (1999) guidelines, differs between laboratories where andrology is carried out in a mixed discipline environment ‘DGH laboratories’ as opposed to that performed in ‘specialist laboratories’ usually associated with assisted conception units. The possibility for different approaches may come from the fact that the staff mix in these two types of laboratories is quite different (Table I) with typically more embryologists and PhD scientists working in ‘specialist laboratories’ as opposed to staff in ‘DGH laboratories’ being almost entirely biomedical scientists, with quite different training routes. Whilst it could be argued that ‘DGH laboratories’ are arguably closer to compliance with WHO (1999) recommendations than ‘specialist laboratories’ linked to assisted conception units (by virtue of more ‘DGH laboratories’ observing stained smears rather than unstained wet preparations), in reality only one laboratory from each group was fully compliant with all of the WHO (1999) recommendations (Table III). Although it is perhaps understandable that andrology laboratories linked to embryology laboratories (specialist labo-
ratories) would want to avoid the use of unnecessary volatile compounds, such as solvents, contaminating the atmosphere in which embryos are cultured (Elder and Dale, 2000) it is of concern that this may mean that diagnostic semen analysis is being performed inappropriately or attempts are being made to correlate sperm morphology results with treatment outcomes such as fertilization rate or pregnancy.

In conclusion, these results illustrate that there remains a great need for improvements in the standardization of methodology in the UK with regard to sperm morphology assessment. Improvements in andrology training and education are urgently needed to address these issues.

References


Submitted on May 26, 2005; resubmitted on June 15, 2005; accepted on June 24, 2005