Results of a questionnaire on sperm morphology assessment

W.Ombelet, H.Pollet, E.Bosmans and A.Vereecken

1Genk Institute for Fertility Technology, ZOL Campus St Jan, Schiepse Bos 6, B-3600 Genk, 2Eurogenetics, Tessenderlo and 3Leuven Institute for Fertility and Embryology, Tiensevest 168, B-3000 Leuven, Belgium

This survey describes the results of a questionnaire on the methodology of sperm morphology assessment. A questionnaire form was sent to 410 fertility centres. A total of 170 answer forms (41.5%) from 40 different countries was evaluated. Most responding centres (147 or 86.5%) treat more than 200 new couples per year. According to our results, a wide and complex variation in different methods of sperm preparation, staining procedures and classification systems is observed world wide. WHO recommendations for sperm preparation seem to be poorly followed. Only 86 centres (50.6%) reported the use of a single approach to both semen preparation and sperm morphology evaluation. Our results indicate an urgent need for standardization and consensus on sperm morphology methodology to regain the power of this important sperm parameter.

Key words: human spermatozoa/laboratory standardization/questionnaire/sperm morphology

Introduction

Although the management of male subfertility has changed dramatically following the introduction of assisted fertilization (i.e. ICSI), routine semen examination still remains a very important tool not only for the diagnosis, but also for the choice of treatment in human subfertility cases.

The general poor level of understanding of the aetiological backgrounds of male infertility and the diagnostic limitations, particularly with respect to the reliance of semen analysis, creates an atmosphere of unease and inadequacy in many andrology laboratories.

The World Health Organization (WHO, 1982, 1987, 1992) tried to standardize the performances of semen analysis and related procedures in order to reduce variation in the results obtained. There is general agreement that standardization of semen analysis is useful and essential, but this is where the consensus ends. According to a previously reported questionnaire (Helmerhorst et al., 1995), there is a wide divergence of opinion as to what can be considered as normal semen. Sperm analysis results show a wide range of values for any given sample, most probably not only related to different methodolo-

Materials and methods

A questionnaire (Figure 1) was sent to 410 ‘fertility centres’ all over the world. The 410 addresses included participants of the first and/or second edition of the International Symposium ‘Andrology in the Nineties’ (April 1993, October 1995, Genk, Belgium) and a selected number of well-known in-vitro fertilization (IVF) centres attending the ESHRE workshops on ICSI (Free University of Brussels, 1994, 1995). The questionnaire was designed to obtain information on the size of the centre (questions 1 and 2), different aspects of the methodology of sperm morphology assessments (questions 3–10) and on various aspects of intra-laboratory quality control (questions 11–13).

A dendrogram (cluster analysis) was made by summarizing the various methods of sperm morphology evaluation, including different methods of preparation, staining and classification.

The questionnaires were distributed between December 1995 and April 1996. Data were stored in Excel spreadsheets (Microsoft Excel Version 5.0a) and analysed using Microsoft Access (Version 2.0).

A total of 170 answer forms (41.5%) from 40 different countries was received. A further 57 centres (13.9%) responded that they did not perform assisted reproduction techniques and 18 questionnaires (4.4%) were sent back too late (after evaluation of the results).
1. How many new couples (all fertility investigation and treatment) do you see per year in your centre?
   - O less than 100
   - O 100-200
   - O 200-500
   - O more than 500

2. In 1994: How many cycles were performed with
   - O IUI
   - O GIFT
   - O IVF
   - O ICSI

3. Do you perform sperm morphology assessment on most semen samples?
   - O No
   - O Yes
   - O on washed semen
   - O on unwashed semen
   - O O wet preparations
   - O O air dried slides
   - O O fixed preparations

4. Which staining-method do you use for sperm morphology assessment?
   - O Papanicolaou
   - O Diff-Quik
   - O Giemsa method
   - O SpermMac
   - O Other: __________________________

5. How many (mean number) spermatozoa are evaluated per slide?
   - O 100
   - O 200
   - O 300
   - O 400
   - O more

6. What criteria do you use for sperm morphology assessment?
   - O WHO 1980
   - O WHO 1987
   - O WHO 1992
   - O David
   - O Strict Tygerberg criteria
   - O Düsseldorf criteria
   - O modified Williams system
   - O Other: __________________________

7. Do you use a CASA system for sperm morphology？
   - O No
   - O Yes

8. How many years of experience do they have？(mean value)... years

9. Did your technicians get experience through:
   - O literature
   - O training-course

10. What is your cut-off for normality？
    - O 50%
    - O 30%
    - O 14%
    - O 10%
    - O Other: __________________________

11. Is this cut-off value based on:
    - O the literature
    - O own study on a reference population

12. Do you presently use an "internal quality control system"？
    - O Yes
    - O No

13. Are you interested in an "external quality control programme"？
    - O Yes
    - O No

---

**Results**

**Geographical distribution and response rate of participating centres**

The response rate varied substantially between different areas and countries although the difference in response rate was minimal between the continents (Table I). The best recovery rate of data was obtained from Belgium (78.9%), the Eastern part of Europe (75%), the Nordic countries (64.7%) and the Netherlands (64.3%). Mediterranean countries scored low, with a response rate of only 19.6%.

**Size of the participating centres**

In 147 centres (86.5%), >200 new infertility couples were treated per year. Of these 147 centres, 80 treated >500 new couples a year (80/170 or 47.1%).

Only four participating centres treated <100 couples.

With regard to the level of therapeutic activities in the participating departments, three centres performed only intrauterine insemination (IUI), while in the other centres, in addition to IUI, gamete intra-Fallopian transfer (GIFT), IVF, and/or intracytoplasmic sperm injection (ICSI) procedures were also offered to the patients.

In 1994, IUI was performed in 158 centres (mean 292 cycles, range 2–1850) while GIFT was performed in only 60 centres (mean 57 cycles, range 1–700). A mean of 356 IVF cycles (range 10–1560) were performed in each of 165 centres. Of these 165 centres, 119 (72.1%) performed ICSI (mean 150 cycles, range 4–1578).

In three centres, sperm morphology was never assessed in the routine exploration of the subfertile couple. Overall, 149 centres (87.6%) answered that they examined sperm morphology in most semen samples.

**Methodology of sperm morphology assessments**

Table II shows the different methodology used in the participating centres.

Only 17 centres (10%) always washed the semen before preparation of the smears. In 36 laboratories (21.2%) wet
Sperm morphology assessment

Table I. Questionnaires sent to 410 centres: geographical distribution and rate of response

<table>
<thead>
<tr>
<th>Continent</th>
<th>Country(ies)</th>
<th>Questionnaires (number)</th>
<th>Answers (number)</th>
<th>Answers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>Austria–Switzerland</td>
<td>18</td>
<td>5</td>
<td>27.8</td>
</tr>
<tr>
<td></td>
<td>Belgium</td>
<td>19</td>
<td>15</td>
<td>78.9</td>
</tr>
<tr>
<td></td>
<td>Eastern Europe countries</td>
<td>4</td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>37</td>
<td>12</td>
<td>32.4</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>59</td>
<td>16</td>
<td>27.1</td>
</tr>
<tr>
<td></td>
<td>Mediterranean countries</td>
<td>46</td>
<td>9</td>
<td>19.6</td>
</tr>
<tr>
<td></td>
<td>Nordic countries</td>
<td>34</td>
<td>22</td>
<td>64.7</td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
<td>14</td>
<td>9</td>
<td>64.3</td>
</tr>
<tr>
<td></td>
<td>United Kingdom</td>
<td>48</td>
<td>17</td>
<td>35.4</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>279</td>
<td>108</td>
<td>38.7</td>
</tr>
<tr>
<td>Africa</td>
<td>South Africa</td>
<td>8</td>
<td>3</td>
<td>37.5</td>
</tr>
<tr>
<td>Oceania</td>
<td>Australia, New Zealand</td>
<td>17</td>
<td>7</td>
<td>41.2</td>
</tr>
<tr>
<td>Asia</td>
<td></td>
<td>35</td>
<td>20</td>
<td>57.1</td>
</tr>
<tr>
<td>South America</td>
<td></td>
<td>9</td>
<td>5</td>
<td>55.6</td>
</tr>
<tr>
<td>North America</td>
<td></td>
<td>62</td>
<td>27</td>
<td>43.5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>410</td>
<td>170</td>
<td>41.5</td>
</tr>
</tbody>
</table>

Figure 2. Distribution of the numbers of observers per centre.

Figure 3. Distribution of the years of experience per observer.

preparations were used, while air-dried slides were used for staining in 30 centres (17.6%). Most responders (52.9%) made use of fixed preparations, as recommended in the WHO manuals (WHO, 1980, 1987, 1992).

Staining of the smears was never performed in 13 centres (7.6%) and in the remaining 157 clinics 11 different staining methods were used. Papanicolaou staining was most popular (33.5%), followed by Diff-Quik (22.9%) and SpermMac (7.6%).

We found no evidence in geographical differences in the use of staining methods although the Shorr staining was popular only in France (55.5%), and a surprisingly high number (8/17 or 47.0%) of centres in the United Kingdom used no staining method.

Regarding the criteria used for evaluation, strict Tygerberg and WHO 92 criteria were used as the only method in respectively 40.1 and 36.4% of centres (Table II). Less popular were the WHO 1987 (7.4%), the David (5.6%) and the Düsseldorf criteria (1.2%). Some centres (9%) claimed to use two different criteria (in most cases strict criteria and WHO 1992).

The strict criteria were most popular in South Africa (100%), Belgium (60%) and Asia (55%). WHO criteria were most

Table II. Description of the different methods used for preparation and evaluation of semen samples in all responding centres. Figures in parentheses are percentages

<table>
<thead>
<tr>
<th>Semen preparations (170)*</th>
<th>Staining methods (170)*</th>
<th>Number of spermatozoa counted (164)*</th>
<th>Criteria (162)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washing</td>
<td>Fixation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>None</td>
<td>10 (5.9)</td>
<td>None (3.6)</td>
</tr>
<tr>
<td>Washed</td>
<td>Wet</td>
<td>17 (10.0)</td>
<td>Papanicolaou (33.5)</td>
</tr>
<tr>
<td>Unwashed</td>
<td>Air dried</td>
<td>99 (58.2)</td>
<td>Diff-Quik (22.9)</td>
</tr>
<tr>
<td>Both</td>
<td>Fixed</td>
<td>44 (25.9)</td>
<td>SpermMac (7.6)</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>13 (7.6)</td>
<td>Shorr or more (1.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Test Simples (4.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other methods (29)</td>
</tr>
</tbody>
</table>

*Number of responding centres.
often followed in the United Kingdom (WHO, 1987: 15.8% and WHO, 1992: 57.8%). The David criteria were used in eight out of 12 centres in France.

We could find no correlation between morphological criteria and staining methods in different centres. The same applies for the number of spermatozoa counted with the exception of the Düsseldorf criteria. Following these criteria, >300 sperm cells are always evaluated. For the other criteria, 100 or 200 cells are examined per slide in more than 90% of centres.

CASA systems were available in 10 centres (5.9%) and in all cases the system was not routinely used, but was only available for research purposes. The distribution of the number of observers per centre (mean: 3.0, range: 1–15) and the average years of experience per observer (mean: 5.9 years, range: <1–28 years) are shown in Figures 2 and 3.

Cluster analyses were performed, summarizing the different methods of sperm morphology assessment (Table III). We found a surprisingly high number of different techniques and interpretations worldwide. Only 86 out of the 170 centres (50.6%) used one approach to sperm morphology preparation and evaluation. Only these 86 centres were taken into consideration for cluster analysis. The wide variation in methodology was most striking, with the exception of 16 centres using the same method of slide preparation and sperm morphology assessment, namely fixed, unwashed smears, Papanicolaou stained and evaluated according to strict Tygerberg criteria.

### Table III. Cluster analysis of sperm morphology methodology in 86 centres responding to use of only one method of slide preparation and sperm morphology evaluation

<table>
<thead>
<tr>
<th>Staining method</th>
<th>Criterion</th>
<th>Air dried washed</th>
<th>unwashed</th>
<th>Fixed washed</th>
<th>unwashed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papanicolaou</td>
<td>WHO 1987</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>WHO 1992</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Strict</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>David</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Düsseldorf</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Diff Quik</td>
<td>WHO 1992</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Strict</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>SpermMac</td>
<td>WHO 1992</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Strict</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Test Simples</td>
<td>WHO 1987</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>WHO 1992</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Strict</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Shorr</td>
<td>WHO 1992</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Strict</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>David</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Total (n = 86)</td>
<td></td>
<td>18</td>
<td>8</td>
<td>51</td>
<td>9</td>
</tr>
</tbody>
</table>

### Discussion

The interpretation of semen analysis results remains, even in the era of ICSI, the most important laboratory practice undertaken in the investigation of the male partner in a subfertile couple. According to the literature, the importance of sperm morphology as a single and independent predictor of in-vivo and in-vitro fertilization seems to be proven (Fraser and DasGupta, 1993; Ombelet et al., 1994, 1995, 1997; Kruger et al., 1996). However, the usefulness of sperm morphology assessments as a predictor of a man’s fertilizing potential has often been challenged due to different classification systems, various slide preparation techniques and inconsistency of analyses within and between laboratories (Ombelet et al., 1995). Differences in methodology of sperm morphology assessments between laboratories are well known, but until now we had no clear idea of the real situation worldwide.

Although our questionnaire was sent to or answered only by centres performing assisted reproductive technologies, we believe that our results do indeed reflect the present laboratory practice concerning sperm morphology assessments, since decision-making in infertility practice is mostly done in the so-called ‘fertility centres’ and most responding centres in this study are considered to be leading centres in their country.

The value of washing the smears before preparation is not cut-off value for normality, 103 responders (64%) followed the literature without performing a study on own-reference population. Fifty-eight centres (36%) answered that they had performed a study on a reference population to evaluate the real cut-off value for normality in their population. For those centres using the strict criteria, a reference population was studied in 41.5 compared to 25.9% of centres using the WHO 1992 criteria. In none of the laboratories applying the WHO 1987 criteria was such a population study performed.

Most centres (82.4%) were candidates for potential participation in an external quality control system.

### Intra-laboratory quality control

A total of 94 centres (56.3%) answered positively to the use of an internal quality control system. In the evaluation of the cut-off value for normality, 103 responders (64%) followed the literature without performing a study on an own-reference population. Fifty-eight centres (36%) answered that they had performed a study on a reference population to evaluate the real cut-off value for normality in their population. For those centres using the strict criteria, a reference population was studied in 41.5 compared to 25.9% of centres using the WHO 1992 criteria.
and quality control purposes and (iii) the necessity of high quality phase contrast optics.

On the other hand, a surprisingly high number of centres did not follow the methods of slide staining as proposed by the World Health Organization (WHO, 1980, 1987, 1992). The Papanicolaou staining was used only in 33.5% of centres and the most alarming finding was that 11 different staining methods were reported, not to mention the fact that a substantial number of centres (16) reported the use of a modified version of one of the more common staining methods. Unfortunately, we still face a lack of studies investigating the influence of different staining techniques, if the classification system remains the same. Nevertheless, we fully agree with Barratt (1995) that it is very important to consider that different preparation methods and staining procedures will alter cell size and therefore influence semen analysis results. Because cell size limits are extremely important in most classification systems, one can easily understand the immense problem of standardization.

Which classification system to use is another difficult issue. When the main goal is the achievement of a very easy and simple test to predict the fertilizing potential of a man’s semen, the more strict criteria (Tygerberg strict criteria; WHO, 1992) are probably the best choice. For the strict Tygerberg criteria (Kruger et al., 1986; Menkveld et al., 1990), a few studies indicate a good predictive value in vivo (Toner et al., 1994; Ombelet et al., 1997). Most studies also confirm the initial reports of Kruger et al. (1986, 1988) showing a good predictive potential of sperm morphology in vitro (Enginsu et al., 1992, 1993; Grow et al., 1994; Ombelet et al., 1994; Al-Hasani et al., 1996), although Robinson et al. (1994) reported normal fertilization rates in teratozoospermic cases. For the WHO 1992 criteria (WHO, 1992), we are still waiting for reference population studies, but we believe that both classification systems are very similar and probably comparable regarding predictive value for in-vivo and in-vitro fertilization. On the other hand, Davis and Gravance (1994) reported that, although the WHO 1992 and Tygerberg strict criteria are internally consistent, a non-constant variance was found when >3700 digitized sperm heads were studied (allometric modelling).

In our survey, 84.4% of centres used the so-called strict criteria (strict Tygerberg and/or WHO 1992 criteria). Other interesting classification systems are the David and Düsseldorf criteria (David et al., 1975; Hofmann and Haider, 1985; Hofmann et al., 1985), but their use is limited to France and Germany respectively. The reason why these criteria are not commonly used is probably their complexity and time-consuming aspect, although the association of certain sperm abnormalities with clinical disturbances is highlighted using these classification systems.

Therefore, the benefit of a more pronounced pathophysiologic approach to sperm morphology assessment has to be weighed against the much easier strict criteria.

Concerning the number of spermatozoa counted per slide, a wide variation exists, even when the same criteria are applied. The only exceptions are those centres using the Düsseldorf classification system (always >300). Although there are no studies available estimating the importance of the number of spermatozoa counted, we can expect that intra- and inter-laboratory assessment variance will reduce with increasing numbers, especially when the cut-off values for normality are low. In other words, for a lower mean percentage of normal forms, a higher number of spermatozoa has to be evaluated to obtain the same coefficient of variation (CV) after repeated examinations. Since most centres nowadays use classification systems with low cut-off values for normality (WHO, 1992: 30%, strict Tygerberg criteria: 14%, Düsseldorf criteria: 30%) it may be logical that more spermatozoa need to be counted to overcome this problem of reproducibility, as suggested by Davis and Gravance (1993) using computerized systems.

Our results also indicate that confidence in automated image processing methods (CASA systems) is still limited. In only 10 out of 170 centres (5.9%) was a CASA system available, and it was only used for research purposes. This also means that in spite of hopeful reports (Davis et al., 1992, 1995; Kruger, 1995; Hofmann et al., 1996; Kruger et al., 1996), the validity and usefulness of CASA systems remain to be proven in clinical studies with special attention to availability, cost-benefit and reliability advantages over visual methods.

To conclude, our results highlight the importance of standardization of laboratory methods, i.e. slide preparation and classification systems. A uniform approach to sperm morphology methodology should be adopted. Assessment variance can only be reduced both within and between laboratories if the methodology is standardized and therefore regular expert meetings are mandatory. If a consensus is possible, regular training courses, conducted by different andrology and fertility societies, can be organized all over the world. Only if experts succeed in finding the gold standard between different approaches, based on well-controlled multicentre quality studies, will there be a happy ending to the debate on sperm morphology confusion.

Acknowledgements
The authors would like to express their appreciation to Annemie Cox, Petra Jacobs, Mia Janssen, Sien Ombelet and Hilde Vanderwegen for technical support in preparing this study. We also express our gratitude to the many participants in the study.

References
1020


Received on November 14, 1996; accepted on February 7, 1997

1020