Semen parameters as predictors of in-vitro fertilization: the importance of strict criteria sperm morphology

Ahmed I.Vawda1,2, Joanne Gunby2 and Edward V.Younglai2

1Department of Zoology, University of Durban Westville, Private Bag X54001, Durban 4000, South Africa and 2Department of Obstetrics and Gynaecology, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

This study evaluated 120 couples undergoing in-vitro fertilization treatment to determine which semen parameter(s) predicted fertilization and whether there was any consistent relationship between strict criteria and standard assessment of sperm morphology. Strict criteria morphology was the only significant predictor of fertilization ($P = 0.0006$, $r^2 = 0.09$), with a sensitivity of 94% and a specificity of 40%. A 12% cut-off point presented a negative predictive value of 98% and a positive predictive value of 22%. The probability of satisfactory fertilization is 40% with morphology <4%, which increases to 97% with normal morphology ($\geq 12\%$). The receiver operating characteristic curve deviated significantly from the diagonal with a 76% area under the curve, making this a superior predictive test. This was augmented by likelihood ratios (LR) of 8.25 (LR+) for results with <4% normal morphology and 0.15 (LR−) for results with $\geq 12\%$ normal morphology by strict criteria. While there was some correlation between strict criteria and standard assessment of morphology ($r = 0.35$), the former explained only 12% ($r^2 = 0.12$) of the variability in the latter. This study concludes that strict criteria morphology predicts fertilization, while other semen parameters do not. A 12% cut-off point makes strict criteria morphology an excellent predictor of satisfactory fertilization, while a value <4% is a good predictor of poor fertilization.

Key words: IVF/receiver operating characteristic curve/semen parameters/spermatozoa/sperm morphology

Introduction

Routine semen analysis usually forms part of the initial investigation of male factor infertility. Although it is an important initial test and is widely used, it provides little information about fertility potential unless semen parameters are grossly abnormal (Dunphy et al., 1989; Aitken et al., 1991). Even in the latter case, the results of a semen analysis can sometimes be misleading because pregnancy can still occur (Silber, 1989; Seibel and Zilberstein, 1995). The advent of assisted reproductive technology and the lack of reliability of the semen analysis in providing prognostic information have shifted emphasis towards a more functional assessment of sperm morphology. Unfortunately, many of these sperm function tests are elaborate, expensive and require experienced technicians and/or specialized equipment, thereby placing them out of reach of the routine laboratory. However, recent evidence suggests that sperm morphology assessment according to strict criteria (Menkveld et al., 1990), which is relatively simple and inexpensive, provides prognostic information similar to that obtained from some of the more elaborate sperm function tests (Franken et al., 1990; Oehninger et al., 1992) and also predicts fertilization outcome in vitro and pregnancy rates in vivo (Van Zyl et al., 1990; Enginsu et al., 1991; Kobayashi et al., 1991; Grow et al., 1994). Hence, strict criteria sperm morphology assessment has the potential to replace some of the elaborate tests. In addition, an accurate prediction of in-vitro fertilization (IVF) outcome early in the infertile couple’s work-up is especially important in view of the expenditure incurred in the cycle before the semen sample is needed for the procedure itself.

The description of morphologically normal spermatozoa according to strict criteria was based initially on the examination of post-coital spermatozoa obtained from the internal os of the cervix. Most of the spermatozoa from such samples were similar and resembled those that bound the zona pellucida in the hemizona assay and in IVF. Therefore, these spermatozoa were presumed to be normal and were used to develop normal morphometric criteria (Menkveld et al., 1990). With these criteria, male factor patients could then be categorized into prognostic groups according to their sperm morphology and the rates of oocyte fertilization in their partners (Grow et al., 1994). In general, patients with a higher sperm morphology had better oocyte fertilization rates.

This study examined routine semen parameters, including strict criteria sperm morphology and combinations of parameters in IVF patients, to determine how they were related to oocyte fertilization rates and to select that parameter which, in our hands, best predicted fertilization outcome. Predictive values, optimum cut-off points, likelihood ratios (LR) and receiver operating characteristic (ROC) curves were computed and compared with other published values. An attempt was also made to determine whether there is a consistent proportional difference between strict criteria and standard assessment of morphology, owing to the controversy surrounding the use of strict criteria sperm morphology in predicting IVF outcome.

Materials and methods

Patients

A total of 120 consecutive couples who enrolled for IVF therapy in a tertiary infertility unit were included in a prospective cohort study.
Semen analysis

Semen was collected by masturbation after 3–5 days of abstinence. The sample was allowed to liquefy at 37°C before a routine semen analysis was performed according to standard World Health Organization (WHO, 1992) criteria. The following parameters were determined by standard assessment: sperm concentration, motility, viability, morphology, total number of motile spermatozoa, semen pH and viscosity. In addition, a semen smear was prepared for a strict criteria sperm morphology evaluation (Menkveld et al., 1990). Briefly, 5 μl semen were thinly smeared on a glass microscope slide and allowed to air dry. The smears were stained according to the Papanicolaou (WHO, 1992), Spermac (Fertility Technologies Inc., San Diego, CA, USA) or Diff-Quik (Baxter Health Care Corporation, Miami, FL, USA) method, depending on when the couple underwent IVF during the 9 month period. Unpublished data from our laboratory indicated that morphology scores were comparable from any one of the three methods. The major limitation of the Papanicolaou method was the long processing time (60 min), but this was offset by excellent quality slides. Spermac required ~10 min for processing a morphology slide and Diff-Quik 5 min. In all, 100 spermatozoa were examined by an experienced assessor under oil immersion on a light microscope and classified into one of the following eight categories: normal, amorphous, tapering head, large head, small head, double head, tail or midpiece defect. Throughout this manuscript, sperm morphology is expressed as the percentage of morphologically normal spermatozoa.

Table I. Semen parameters and fertilization rates in vitro (n = 120)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean ± SEM</th>
<th>Median</th>
<th>25–75%*</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration/mL (×10⁶)</td>
<td>42.7 ± 1.4</td>
<td>42</td>
<td>31–55</td>
<td>10</td>
<td>77</td>
</tr>
<tr>
<td>Progressive motility (%)</td>
<td>57.1 ± 1.7</td>
<td>59</td>
<td>45–71</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>Viability (%)</td>
<td>76.1 ± 1.0</td>
<td>77</td>
<td>70–85</td>
<td>31</td>
<td>96</td>
</tr>
<tr>
<td>Volume (ml)</td>
<td>3.5 ± 0.2</td>
<td>3.2</td>
<td>2–4.5</td>
<td>0.5</td>
<td>9</td>
</tr>
<tr>
<td>Fertilization rate per oocyte (%)</td>
<td>65.4 ± 2.8</td>
<td>75</td>
<td>44–92</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

*Interquartile ranges.

The remainder of the semen sample was processed for IVF by either routine swim-up or centrifugation through a discontinuous Percoll density gradient.

Ovulation induction

Female partners underwent ovarian stimulation with luprolide acetate (Lupron; Abbott Laboratories, London, Canada) and high purity follicle stimulating hormone (Serono, Norwell, USA). Follicular development was monitored by serum oestradiol concentrations and ultrasonography. Human chorionic gonadotrophin (HCG) was administered when more than two follicles achieved a diameter of 18 mm and serum oestradiol concentration was >5000 pmol/L. Follicles were aspirated under transvaginal ultrasound guidance 34–36 h after HCG administration. Oocytes were inseminated with 50 000 spermatozoa per oocyte in all but 11 cases where either patients had previously failed fertilization, or a combination of at least two of the following conditions existed: oligozoospermia, asthenozoospermia and teratozoospermia. In the latter cases, 500 000 spermatozoa were used to inseminate each oocyte. Fertilization was defined as the presence of two pronuclei in the oocyte 18 or 24 h after insemination and/or signs of cleavage 48 h after insemination.

Statistical analysis

Data were analysed using standard methods for univariate analyses, and with the use of stepwise linear regression, stepwise logistic regression and correlation coefficients. The Kappa test was used to evaluate agreement. Initially all measured semen parameters and some combinations of parameters (i.e. the total number of motile spermatozoa and the product of the total number of motile spermatozoa and standard criteria sperm morphology) were tested; only those showing acceptable correlations were computed further. The possible influence of the overall results by the inclusion of the study of 11 cases with a higher insemination concentration was assessed by the χ² test. Data were also analysed to confirm that the fertilization rate was independent of female age and diagnostic classification.

Diagnostic test properties, including sensitivity [likelihood of an abnormal (positive) test result amongst patients with unsatisfactory fertilization], specificity [likelihood of a normal (negative) test result amongst patients with satisfactory fertilization], positive predictive value [likelihood of unsatisfactory fertilization amongst patients with a positive (abnormal) test result] and negative predictive value [likelihood of satisfactory fertilization amongst patients with a negative (normal) test result], were obtained from 2×2 contingency tables of sperm morphology against fertilization in vitro. Because previous unpublished data (n = 576) from our unit had indicated a low pregnancy rate (per embryo transfer) of 6% for fertilization rates <30% and a higher rate of 18% for fertilization rates ≥30%, we used this 30% cut-off value to indicate satisfactory fertilization and tested the predictive value of sperm morphology results against this
outcome. The diagnostic value of sperm morphology assessment in predicting fertilization outcome was also ascertained with the aid of ROC curves (Peng et al., 1987). A number of sperm morphology cut-off points were selected to draw ROC curves and from this the optimum cut-off point was determined. LR were also used to determine test quality, whereby the ratios denoted the likelihood of a normal (LR−) or abnormal (LR+) sperm morphology result in patients with poor fertilization over those with good fertilization. LR+ is generally ≥1, and a ratio >5 was taken as indicating a good test. LR− is usually ≤1, and a good test was accepted as a ratio <0.2 (Collins, 1989).

**Results**

The mean fertilization rate per oocyte was 65.4 ± 2.8%; 8% of patients had no fertilization, and 85% had ≥30% of their oocytes fertilized. The mean proportion of spermatozoa considered to be normal by strict criteria was 10.0 ± 0.5% (range 1–29), while by standard assessment the mean proportion of spermatozoa with normal morphology was 42.7 ± 1.4% (range 10–77). Additional semen parameters are listed in Table I. The overall pregnancy rate for couples in this study was 9.2% per cycle.

Correlations of semen parameters with fertilization rate and with each other are shown in Table II. Strict criteria sperm morphology was more strongly correlated with fertilization rate than was morphology assessed by standard methods (r = 0.31, P = 0.0006 versus r = 0.20, P = 0.03). Of the various semen parameters measured, the only significant predictor of fertilization rate was strict criteria sperm morphology (stepwise linear regression, P = 0.0006, β coefficient = 1.78, adjusted r² = 0.09). In a stepwise logistic regression analysis, using fertilization ≥ or <30% as the dependent variable, strict criteria sperm morphology was again the only significant predictor of fertilization. Other variables included in the analysis were standard assessment morphology, progressive motility, total number of motile spermatozoa, the product of total number of motile spermatozoa and strict criteria morphology, diagnostic category of female infertility, female age, cycle number and the concentration of spermatozoa used for insemination. None of the female factors recorded for our patients significantly influenced the prediction of fertilization rate (r = 0.52). Similarly, female age and cycle number did not significantly affect fertilization (χ², P = 0.54 and χ², P = 0.11 respectively). Figure 1 shows the probabilities of satisfactory fertilization predicted from the logistic regression model.

In 11 cases with anticipated low fertilization because of a previously failed IVF attempt or combined male factor conditions, an insemination concentration of 500 000 spermatozoa per oocyte was used. Only two of these cases also had strict criteria morphology <4%. However, this compensation was inadequate because only 55% of patients in this group had satisfactory fertilization compared with 88% for patients receiving 50 000 spermatozoa per oocyte (χ², P = 0.003). Thus the number of spermatozoa inseminated is not regarded as a confounding variable.

The ROC curve for strict criteria sperm morphology and satisfactory fertilization depicted an 8% cut-off value as optimal in maximizing sensitivity and specificity. The curve’s deviation from the diagonal indicated that, overall, strict criteria sperm morphology assessment was a good predictor of IVF outcome (Figure 2). In addition, the area under the curve was calculated as 76%, making this value the proportion of successful predictions of fertilization. From a clinical standpoint, however, double cut-off values of 12 and 4% appeared
fertilization. A discriminator diagnostic test. Neither a normal nor an abnormal test value significantly changed the probability of good discrimination. The calculated LR make standard sperm morphology assessment a non-predictive value of a normal test 86%, predictive value of an abnormal test 20%, LR+ 1.4 and LR- 0.9. The calculated diagnostic test characteristics for standard assessment sperm morphology were: sensitivity 28%, specificity 80%, predictive value of a normal test 86%, predictive value of an abnormal test 20%, LR+ 1.4 and LR- 0.9. The calculated LR make standard sperm morphology assessment a non-discriminatory diagnostic test. Neither a normal nor an abnormal test value significantly changed the probability of good fertilization.

### Table IV. In-vitro fertilization and pregnancy results according to strict criteria sperm morphology groupings

<table>
<thead>
<tr>
<th>Strict criteria sperm morphology (%)</th>
<th>No. of patients</th>
<th>Mean fertilization rate (%)</th>
<th>Patients with satisfactory fertilization (%)</th>
<th>Pregnancy rate per cycle (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>≥30</td>
<td>0</td>
</tr>
<tr>
<td>1-3</td>
<td>10</td>
<td>34</td>
<td>40</td>
<td>60*</td>
</tr>
<tr>
<td>4-11</td>
<td>68</td>
<td>64</td>
<td>84</td>
<td>91*</td>
</tr>
<tr>
<td>12-29</td>
<td>42</td>
<td>75</td>
<td>98</td>
<td>100*</td>
</tr>
<tr>
<td>Overall data</td>
<td>120</td>
<td>65</td>
<td>85</td>
<td>92*</td>
</tr>
</tbody>
</table>

*Also denotes the proportion of patients who received an embryo transfer.

### Table V. 2×2 Contingency tables for fertilization rates and strict criteria sperm morphology

<table>
<thead>
<tr>
<th>Test result</th>
<th>12% cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unsatisfactory fertilization</td>
</tr>
<tr>
<td>Fertilization rate ≥ 30%</td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td>17</td>
</tr>
<tr>
<td>Normal</td>
<td>1</td>
</tr>
<tr>
<td>Fertilization rate &gt; 0%</td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td>10</td>
</tr>
<tr>
<td>Normal</td>
<td>0</td>
</tr>
</tbody>
</table>

Discussion

This study investigated a variety of semen parameters and combinations of parameters to determine which was/were best able to predict fertilization rates in vitro. Of all the semen parameters measured and calculated, strict criteria sperm morphology was the only significant predictor of fertilization outcome. The clinical significance of sperm morphology in infertility diagnosis and treatment has been discussed recently by several authors (Aitken et al., 1995; Barratt, 1995; Oehninger and Kruger, 1995; Seibel and Zilberstein, 1995; Seracchioli et al., 1995). While it is clearly evident that because of the complexity of events leading up to and in accomplishing fertilization, no single sperm parameter can be used conclusively to predict IVF outcome, several studies have shown that strict criteria sperm morphology is a powerful predictor of fertility potential. On the other hand, several studies dispute the importance of sperm morphology in identifying male factor infertility and its use in predicting IVF and pregnancy (Seracchioli et al., 1995). This disparity is most probably a result of inter-laboratory variations in morphology assessment methods and study profiles. The importance of objectivity and minimal inter- and intra-assessor variability in scoring strict morphology cannot be over-emphasized (Barratt, 1995). Furthermore, evidence that strict criteria sperm morphology does not affect the outcome of intracytoplasmic sperm injection but does affect that of IVF (Oehninger and Kruger, 1995) emphasizes the importance of morphologically normal spermatozoa in overcoming barriers leading to fertilization when spermatozoa are not injected into the egg.

We were able to group our patients distinctly into three categories, in much the same way as the pioneers who described the relationship between strict criteria sperm morphology and fertilization (Kruger et al., 1988; Grow et al., 1994). However,
they selected a cut-off point >14% for normal sperm morphology, with corresponding fertilization rates of 64 (Kruger et al., 1988) and >85% (Grow et al., 1994), while our study selected a value of ≥12%, which led to a mean fertilization rate of 75% in this group. Patients with morphology values <4% generally had poor fertilization rates and no pregnancies, although there were instances of satisfactory fertilization in this group. The LR+ of 8.25 for an abnormal result (LR+) and a positive predictive value of 60% make this cut-off point a good predictor of unsatisfactory fertilization. Conversely, patients with morphology values ≥12% had a good prognosis for successful fertilization, as indicated by a LR- of 0.15 and a negative predictive value of 98%. A third category of patient with sperm morphology values between 4 and 11% demonstrated fertilization rates (64%) similar to the overall rate of 65% in this study. Therefore, these results present an ideal characteristic for a good diagnostic test, i.e. adjusting the predicted fertilization rate towards 100% with increasing sperm morphology values and adjusting it towards 0% with decreasing values. Furthermore, the discriminatory power of a diagnostic test may be judged by the mean value of the sensitivity and specificity, with 50% indicating a non-discriminatory test and 100% an ideal test (Peng et al., 1987). Our data produced values of 65 and 67% for the 4 and 12% cut-off points respectively, again depicting the usefulness of strict criteria sperm morphology assessment in predicting the outcome of fertilization in vitro.

In view of some existing controversy regarding the usefulness of strict criteria versus standard assessment of sperm morphology, we also investigated the relationship between the two to establish whether there was a consistent proportional difference in scores between the two methods, and which of the methods predicted fertilization outcome. Although there was a fairly good correlation between standard and strict criteria morphology, we ascribed this to the fact that gross abnormalities, e.g. large, small and double heads, are identified by both methods. Despite this similarity in the scoring procedures, <12% of the variability was explained, and there was very little agreement beyond chance in identifying samples as normal or abnormal. The reason for this is that, in our opinion, strict criteria morphology assessment is objective while standard sperm morphology assessment is generally subjective, which would lead to substantial observer variability. Standard assessment sperm morphology is not predictive of fertilization outcome. Both LR+ and LR- classified standard morphology assessment as a non-discriminatory test in predicting the fertilization rate.

Despite it being biologically accurate to accept any rate >0% as denoting the occurrence of fertilization, we opted to use a cut-off value of 30% to indicate satisfactory fertilization. The main reason for this was that our unit has recorded previously a low pregnancy rate of only 6% for fertilization rates <30% and a higher rate of 18% for fertilization rates >30%. In our opinion, using the 30% value for our computations was more useful clinically in predicting the outcome of IVF from strict criteria sperm morphology scores.

Our selection of 12% normal sperm morphology as a cut-off point distinguishing normal from abnormal morphology scores is based on an assessment of satisfactory fertilization rates above and below various cut-off points. With this cut-off value only one couple out of 42 with a sperm morphology >12% presented with unsatisfactory fertilization (<30% fertilization rate). Had we selected 8% as the cut-off value, as promulgated by the ROC curve (Figure 2), a total of six couples out of 79 with normal morphology scores would have presented with unsatisfactory fertilization rates.

In conclusion, data in the literature suggest that strict criteria sperm morphology scores are more predictive of fertilization in vivo and in vitro than morphology scores determined by other methods (Van Zyl et al., 1990; Enginsu et al., 1991, Kobayashi et al., 1991, Grow et al., 1994). Furthermore, a positive correlation between strict criteria sperm morphology and some sperm function tests, e.g. the hypo-osmotic swelling test (Zaneveld et al., 1990), the hemizona assay (Franken et al., 1990), zona pellicula binding (Oehninger et al., 1992) and the chromatin condensation test (Claassens et al., 1992), enhances the legitimacy of the morphology test. However, it is important to accept that strict criteria sperm morphology, like any other currently used diagnostic test, does not evaluate all of the many prerequisites for fertilization, and therefore cannot conclusively predict whether fertilization will occur (Seibel and Zilberstein, 1995). Nevertheless, it provides a very sensitive measure of the fertilizing potential. In our study, it predicted both poor fertilization rates in patients with <4% normal forms and high fertilization rates in patients with a sperm morphology ≥12%. The high predictive value of strict criteria sperm morphology in IVF outcome and its correlation with other elaborate tests, coupled with the fact that the method is simple and inexpensive, make it a potential substitute for these tests.

In conclusion, our results suggest that of all semen parameters measured during routine semen analysis, only sperm morphology evaluation according to strict criteria predicts IVF outcome in a manner that is clinically meaningful. Our study suggests a cut-off value of 12% normal forms to differentiate between normal and abnormal sperm morphology. Patients with values ≥12% have an excellent prognosis, those with values of 4–11% have a lower but good chance of success, while those with values <4% have a poor prognosis. There is little agreement between standard and strict criteria assessment of sperm morphology in identifying abnormal samples, and standard morphology assessment does not predict IVF outcome. Strict criteria sperm morphology evaluation represents a simple, inexpensive test with high prognostic value comparable with some of the more elaborate sperm function tests. It should be included in any diagnostic work-up of an infertile couple.

Acknowledgements
The authors thank the staff of the Gamete Biology Laboratory at Chedoke–McMaster Hospitals for their assistance. This study was supported in part by a grant from the Foundation for Research Development, South Africa to A.I.V.

References
A.I. Vawda, J. Gunby and E.V. Younghal


Received on February 2, 1996, accepted on May 13, 1996