IDENTIFICATION OF MALE FACTOR INFERTILITY USING A NOVEL SEMEN QUALITY SCORE AND REACTIVE OXYGEN SPECIES LEVELS

Kiran P. Nallella, Rakesh K. Sharma, Shyam S. R. Allamaneni, and Ashok Agarwal


PURPOSE: To determine whether patients with male factor infertility can be accurately identified by calculating a novel semen quality score and measuring levels of reactive oxygen species during routine infertility screening.

METHODS: Semen samples from 133 patients and 91 healthy donors were evaluated with manual and computer-assisted semen analysis. A principal component analysis model was employed to calculate a semen quality score. In brief, this score was calculated by base 10 logarithms multiplied by varying weights given to 9 sperm parameters. Reactive oxygen species levels were measured using chemiluminescence assay.

RESULTS: The semen quality score had a sensitivity of 80.45% and accuracy of 77% at a cutoff of 93.1 in identifying patients with male factor infertility. The area under the receiver operating characteristic curves for the semen quality score was 84.28% (95% CI: 65.22%-100%). Reactive oxygen species levels [log10 (reactive oxygen species +1)] were significantly higher in male factor infertility patients. Reactive oxygen species had a sensitivity of 83.47% and specificity of 60.52% with an accuracy of 75% at a cutoff of 1.25 in identifying male factor infertility patients. The area under the receiver operating characteristic curve for reactive oxygen species levels was 78.92% (95% CI: 72.60%-85.23%). semen quality scores were significantly and negatively correlated with reactive oxygen species levels in the donors and the male factor infertility patients.

CONCLUSIONS: The semen quality score and reactive oxygen species levels in semen samples appear to be strongly associated with male factor infertility. Because both of these parameters are more sensitive than individual sperm parameters in identifying male factor infertility, they should be included in routine infertility screening.

normal spermiograms in these cases is approximately 15%.

Identifying diagnostic measures for MFI that are easy to perform, relatively inexpensive, and able to provide an accurate diagnosis is necessary.

Because semen parameters are interrelated, they can be reduced to 2 semen scores termed the overall semen quality (SQ) and relative quality (RQ) scores. The SQ score was developed by principal component analysis of 9 individual sperm parameters, and has been reported as a highly reliable and efficient tool for clinicians who screen for and diagnose MFI.

Furthermore, studies have shown that 40% to 88% of nonselected infertile patients have high levels of seminal reactive oxygen species (ROS). Uncontrolled and excessive production of ROS may be one of the major factors leading to infertility. It appears, therefore, that the presence of oxidative stress in infertile normozoospermic men may help explain previously unexplained cases of infertility that were otherwise attributed to female factors.

The purpose of this study was to: 1) examine improved parameters in identification of MFI patients during infertility screening, 2) establish cutoff values for the SQ score and ROS levels that identify patients with MFI, and 3) determine the relationship between the SQ score and levels of ROS in patients with MFI.

METHODS

The Institutional Review Board of The Cleveland Clinic Foundation approved the study. Medical charts of the patients attending the infertility clinic for infertility evaluation were reviewed.

Study population

The patient population consisted of 133 MFI patients. All patients had a history of at least 1 year of primary or secondary infertility with their current partner and had completed a basic evaluation that included medical history, a physical examination, and at least 2 semen analyses. On occasion, the patient provided more than 1 semen sample. Semen samples (n = 264) were divided into 4 groups based on results from all semen analyses: oligozoospermic (n = 61), asthenozoospermic (n = 96), teratozoospermic (n = 69), and oligoasthenoteratozoospermic (OAT, n = 38). Subjects with semen samples containing >1 x 10^6 round cells/mL were excluded to avoid a potential source of ROS generation. All female partners had patent fallopian tubes and experienced regular ovulation. In addition, results of semen samples from 91 normal healthy volunteers (donors) were used as the control for this study.

Semen analysis

Semen was collected by masturbation after 2 to 3 days of sexual abstinence. After liquefaction, semen analysis was performed both manually and by computerized semen analysis (CASA) (IVOS, 10.7s, Hamilton Thorne Research, Beverly, MA). For each measurement, a 5 µL aliquot from either a control or infertile patient sample was loaded on a MicroCell slide (Conception Technologies, San Diego, CA). Sperm motion kinetics measured by CASA included: sperm concentration (10^6/mL), percent motility, curvilinear velocity (VCL; µm/sec), straight-line velocity (VSL; µm/sec), average path velocity (VAP; µm/sec), linearity (LIN; percent), and amplitude of lateral head displacement (ALH; µm). In addition to the computerized results, manual results were also calculated for sperm concentration and motility.

For morphological evaluation, seminal smears were stained with Giemsa stain (Diff-Quik, Baxter Healthcare Corporation, McGraw Park, IL), and the percent sperm with normal morphology was assessed by WHO guidelines and Kruger’s strict criteria.

Measurement of Reactive Oxygen Species

Aliquots of liquefied semen were centrifuged at 300x g for 7 minutes. The sperm pellet was washed twice with phosphate buffered saline (PBS), pH 7.4, and resuspended in the same medium at a concentration of 20 x 10^6 sperm/mL. ROS production was measured by the chemiluminescence assay method using luminol (5-amino-2, 3-dihydro 1, 4-phthalazinedione; Sigma Chemical Co., St. Louis, MO) as the probe. Ten mL of 5 mM luminol prepared in dimethyl sulfoxide (DMSO; Sigma Chemical Co.) was added to 400 mL of the washed sperm suspension. The ROS levels were determined by measuring chemiluminescence with a luminometer (Autolumat LB 953, Berthold technologies, Bad-Wildbad, Germany) in the integrated mode for 15 minutes. Results were expressed as 10^4 counted photons per minute (cpm)/20 x 10^6 sperm. These were log transformed [log (ROS + 1)], hereafter referred as ROS for simplicity, and were used in statistical analysis.

Statistical Analysis

A principal component analysis model was employed to calculate an overall SQ score that accounts for most of variability observed among the battery of interrelated semen variables. Details of the SQ score calculation are described in our previous study. In brief, this score was calculated by base 10 logarithms multiplied by varying
weights given to the 9 sperm parameters: concentration, motility, sperm morphology according to WHO guidelines, and Tygerberg strict criteria, VCL, VSL, VAP, LIN, and ALH.

The SQ score and ROS level comparisons between groups were made using unpaired t tests, while correlations between variables were assessed using Pearson’s correlation coefficient. In addition, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were also calculated. Receiver operating characteristic (ROC) curves, such as the area under the curve (AUC), were calculated to summarize the inherent capacity of the sperm quality variables to discriminate patients with MFI from the control donors. The SQ score along with ROS and sperm parameters were compared using De Long’s nonparametric comparisons. Calculations were performed with GraphPad InStat version 3.00 statistical software (GraphPad Software, Inc., San Diego, CA) and StatsDirect (StatsDirect Ltd., Gresham Way, UK). A P value < .05 was considered statistically significant.

RESULTS

Identification of MFI patients with SQ score

The SQ scores (mean ± standard deviation) for the donors and patients are shown in Table 1. The SQ scores were significantly higher in the donors than in the MFI patients (P < .001). Significantly lower SQ scores were observed in all groups of MFI patients compared with donors. The lowest SQ score was seen in the patients with OAT.

In order to determine whether the SQ score could differentiate MFI patients from control donors, we examined various cutoff values to determine the SQ score that would have the highest sensitivity. Table 2 displays various predictors of semen quality in 91 donors and 133 MFI patients using different cutoff values of the SQ score. Using a cutoff of 100 amongst the patient population, 93.23% (124 of 133) of men had a SQ score < 100, and only 7% (9 of 133) of the patients had a SQ score > 100. At this cutoff, the sensitivity was high (93.23%), and the PPV (or the probability that a person having the disease given a positive test) was 70%, with an accuracy (defined as the positively diagnosed MFI patients and correctly excluded donors) of 72%. The specificity was, however, very low (40.65%) at this cutoff. Lower cutoff values resulted in an increase in specificity with a corresponding decrease in sensitivity.

A cutoff of ≤ 93.1 provided the optimum sensitivity of 80.45% and specificity of 70.32%. When a cutoff of ≤ 93.1 was used, the SQ score was able to correctly identify 80.45% of the patients as being MFI patients. Using this cutoff, the overall accuracy in differentiating the donors from the patients was 77% i.e. 150 of the 194 individuals

<table>
<thead>
<tr>
<th>Study population</th>
<th>SQ score</th>
<th>P value</th>
<th>ROS levels</th>
<th>Correlation coefficient</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donors</td>
<td>97.07 ± 10.76 (n = 91)</td>
<td>-</td>
<td>1.20 ± 0.80* (n = 76)</td>
<td>-0.45</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>MFI patients</td>
<td>75.56 ± 18.55 (n = 133)</td>
<td>&lt; .001</td>
<td>2.29 ± 1.05 (n = 121)</td>
<td>-0.36</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Oligozoospermic</td>
<td>64.70 ± 14.93 (n = 61)</td>
<td>&lt; .001</td>
<td>2.70 ± 1.18 (n = 57)</td>
<td>-0.17</td>
<td>.20</td>
</tr>
<tr>
<td>Asthenozoospermic</td>
<td>70.55 ± 18.00 (n = 96)</td>
<td>&lt; .001</td>
<td>2.30 ± 1.10 (n = 90)</td>
<td>-0.39</td>
<td>.0002</td>
</tr>
<tr>
<td>Teratozoospermic</td>
<td>66.82 ± 16.50 (n = 69)</td>
<td>&lt; .001</td>
<td>2.40 ± 1.19 (n = 66)</td>
<td>-0.39</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>OAT</td>
<td>56.01 ± 12.69 (n = 38)</td>
<td>&lt; .001</td>
<td>2.82 ± 1.21 (n = 35)</td>
<td>-0.16</td>
<td>0.32</td>
</tr>
</tbody>
</table>

*Values are mean ± SD; OAT = oligoasthenoteratozoospermic; SQ = semen quality; 1P < .05 was considered significant comparing SQ score between donors and different groups of infertile patients; 2P < .05 was considered significant comparing ROS levels between donors and different groups of infertile population; 3P < .05 was considered significant using Pearson correlation coefficient between SQ score and ROS levels; 4Log (ROS + 1) were used

Table 2 - Identification of male factor infertility patients using prediction parameters with different cutoff values

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cutoff value</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQ score cutoff</td>
<td>100</td>
<td>93.23</td>
<td>40.65</td>
<td>69.66</td>
<td>80.43</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>93.1</td>
<td>80.45</td>
<td>70.32</td>
<td>79.85</td>
<td>71.11</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>75.19</td>
<td>76.92</td>
<td>82.64</td>
<td>67.96</td>
<td>76</td>
</tr>
<tr>
<td>ROS levels</td>
<td>1</td>
<td>89.26</td>
<td>42.10</td>
<td>71.05</td>
<td>71.11</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>1.25</td>
<td>83.47</td>
<td>60.52</td>
<td>77.09</td>
<td>69.70</td>
<td>74.61</td>
</tr>
</tbody>
</table>

SQ = semen quality; ROS = reactive oxygen species; ROS values are log (ROS + 1); PPV = Positive predictive value; NPV = Negative predictive value
in our study population could be correctly categorized with this test (true positive and true negative).

Using an SQ score cutoff of $\leq 93.1$, we compared SQ score in this study population with normal sperm parameters with cutoff values established by WHO guidelines (2) i.e. sperm concentration ($\geq 20 \times 10^6$/mL), motility ($\geq 50\%$) and WHO morphology ($\geq 30\%$ normal forms) (Table 3). Motility showed a sensitivity of 72.18% and specificity of 83.51%. Using the WHO classification for sperm morphology, the sensitivity was 51.87% and the specificity was 87.91%. On the other hand, Kruger’s morphology had sensitivity of 82.70% but a very poor specificity (26.37%).

**Identification of MFI patients using ROS levels**

Significantly higher levels of ROS [log (ROS + 1)] were seen within MFI patients, as well as in all the 4 subgroups, compared to donors ($P < .001$) (Table 1). The highest levels of ROS were seen in oligoasthenoteratozoospermic (OAT) patients. A strong negative correlation was seen between the SQ score and levels of ROS for donors ($r = -0.45$, $P < .001$) and MFI patients ($r = -0.36$, $P < .001$). A negative correlation was also seen in the asthenozooospermic and teratozoospermic patients (Table 1). Using an ROS cutoff of 1, the sensitivity was 89.26%, but the specificity in correctly identifying the infertile patients was poor (42.10%). When the ROS cutoff was increased to 1.25, the sensitivity decreased to 83.47%, but the specificity increased along with accuracy (Table 1).

**Receiver operating characteristic (ROC) curves**

The effectiveness of the SQ score in differentiating the MFI patients from the normal healthy donors was studied by generating ROC curves (Table 4). Using a SQ score cutoff of $\geq 93.1$, the AUC was 84.28% with a 95% confidence interval (CI) of 65.22% to 100%. The AUC using different sperm parameters and ROS cutoff of 1.25 is shown in Table 4. Sperm concentration and percent motility had a similar AUC. Both the AUC and 95% CI were much lower for sperm morphology both by WHO criteria and Kruger’s strict criteria (Fig. 1). The AUC for SQ score was higher (84.28%) compared with that for ROS (78.92%) (Fig. 2).

We were also interested to see whether we could arrive at the best cutoff values for sperm parameters in identifying MFI patients compared to the well-established WHO values for normal sperm parameters. By giving equal weight to sensitivity and specificity, the best cutoff values were provided by the statistical program. Using this method, a significantly different cutoff value was obtained for sperm concentration compared to the WHO cutoff value (Table 3). Using a cutoff value of $< 49.80 \times 10^6$/mL for sperm concentration, the best sensitivity (79.69%) was seen compared to 45.86% at the WHO defined cutoff of $< 20 \times 10^6$/mL.

**Table 3 - Characteristics in correctly identifying the male factor infertility patients using calculated and World Health Organization (WHO) established cutoff for various sperm parameters**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cutoff value</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sperm count (X10^6/mL)</td>
<td>$&lt; 20^a$</td>
<td>45.86</td>
<td>94.50</td>
<td>92.42</td>
<td>54.43</td>
</tr>
<tr>
<td></td>
<td>$\leq 49.80^b$</td>
<td>79.69</td>
<td>71.42</td>
<td>80.30</td>
<td>70.65</td>
</tr>
<tr>
<td>Motility (%)</td>
<td>$&lt; 50^a$</td>
<td>72.18</td>
<td>83.51</td>
<td>86.48</td>
<td>67.25</td>
</tr>
<tr>
<td></td>
<td>$\leq 50^b$</td>
<td>74.43</td>
<td>81.31</td>
<td>85.34</td>
<td>68.51</td>
</tr>
<tr>
<td>Morphology (% normal forms)</td>
<td>WHO morphology (%)</td>
<td>$&lt; 30^a$</td>
<td>51.87</td>
<td>87.91</td>
<td>86.25</td>
</tr>
<tr>
<td></td>
<td>$\leq 29^b$</td>
<td>51.87</td>
<td>87.91</td>
<td>86.25</td>
<td>55.55</td>
</tr>
<tr>
<td></td>
<td>Kruger’s morphology (%)</td>
<td>$&lt; 14^a$</td>
<td>82.70</td>
<td>26.37</td>
<td>62.14</td>
</tr>
<tr>
<td></td>
<td>$\leq 7^b$</td>
<td>54.13</td>
<td>89.01</td>
<td>87.80</td>
<td>57.04</td>
</tr>
</tbody>
</table>

$^a$WHO cutoff values; $^b$Cutoff values given by the statistical program; PPV = positive predictive value; NPV = negative predictive value

**Table 4 - Areas under the curve (AUC) for semen quality (SQ) score and various sperm parameters**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cutoff value</th>
<th>AUC (%)</th>
<th>95% CI for AUC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQ cutoff score</td>
<td>93.1</td>
<td>84.28</td>
<td>65.22–100</td>
</tr>
<tr>
<td>Log (ROS+1)</td>
<td>1.25</td>
<td>78.92</td>
<td>72.60–85.23</td>
</tr>
<tr>
<td>Sperm count (x10^6/mL) $^a$</td>
<td>20</td>
<td>81.31</td>
<td>63.62–99.99</td>
</tr>
<tr>
<td>Motility (%) $^a$</td>
<td>50</td>
<td>82.29</td>
<td>64.13–100</td>
</tr>
<tr>
<td>Morphology (% normal forms)</td>
<td>WHO morphology (%)</td>
<td>30</td>
<td>68.16</td>
</tr>
<tr>
<td></td>
<td>Kruger’s morphology (%)</td>
<td>14</td>
<td>70.39</td>
</tr>
</tbody>
</table>

WHO = World Health Organization; $^a$Cutoff values established by the WHO guidelines; ROS = reactive oxygen species; CI = confidence interval
The cutoff values obtained by the statistical program for both motility and sperm morphology, however, were similar (49.9% and 29%) to the WHO defined values of 50% and 30%. These values were also comparable for sensitivity, specificity, positive predictive value, and negative predictive value. For sperm morphology by Kruger’s strict criteria, a cutoff value of 7% had greater sensitivity and specificity than a cutoff of 14% in correctly identifying the MFI patients.

DISCUSSION

The role of traditional semen analysis and individual sperm parameters in identifying fertile and infertile men is a matter of ongoing debate. Several studies have reported the predictive values of individual sperm parameters such as concentration of motile spermatozoa20,21 and computerized measurements of different patterns of spermatozoa motility22-24 and morphology25-28 to aid in the determination of male fertility and infertility. There is no consensus within the literature on the cutoff values of any individual parameter in defining patients with MFI from fertile males.29 It appears that there is a need to establish a combined value for all these sperm parameters (concentration, motility, and morphology) into a single score that can explain semen analysis results effectively and help establish the status of the individual who comes to the infertility screening. This approach would be helpful both to the clinician and the patient.

Because a screening test is used to identify a maximum number of patients during routine evaluation, the test must be sensitive so that it can identify all true positives (number of patients identified as being patients) at a given cutoff point. Our results show that the SQ score was able to predict MFI patients with the best sensitivity (80.45%) along with ROS (83.47%) at given cutoff points of 93.1 and 1.25, respectively. All the sperm parameters had lower sensitivities in identifying the MFI patients when WHO-defined cutoff points were used versus the SQ score and ROS.

Using the 10th percentile for the donors, we found that the lower limit of normality for the SQ score was 84.38 while the upper limit (90th percentile) in MFI patients was 96.79. This indicates that only 10% (9 of 91) of the donors have a score < 84.38 and that 11% of MFI patients (14 of 133) had an SQ score > 96.79. Overlapping of normal and abnormal SQ scores between 2 groups is unavoidable using sperm parameters and SQ scores. Using a SQ cutoff score of 93.1, only 12% (29/224) of our study population had false positive results; that is, subjects who tested positive but were actually negative. These individuals will, however, be considered MFI patients. Similarly, by using this cutoff, only 12% (26 of 224) of the patient population fell in the category of normal healthy donors. Therefore, an SQ score with a cutoff of 93.1 can better discriminate MFI patients from normal, healthy donors. As a result, this cutoff can serve as an effective screening tool during routine infertility evaluation. Furthermore, using the AUC for the SQ score indicated the probability of correctly identifying MFI patients.

RESUMO


OBJETIVO: Determinar se pacientes portadores do fator de infertilidade masculina podem ser precisamente identificados através do cálculo de um novo escore de qualidade de sêmen e pela medida de espécies reativas de oxigênio durante uma avaliação rotineira de infertilidade.

MÉTODOS: Amostras de sêmen de 133 pacientes e de 91 doadores saudáveis foram avaliadas através de análise manual e computadorizada de sêmen. Um modelo de análise...
do componente principal foi empregado para calcular o escore de qualidade de sêmen, utilizando logaritmos base 10, multiplicados por ponderações variáveis de 9 parâmetros espermáticos. Os níveis de espécies reativas de oxigênio foram medidos através de testes de quimiluminescência.

RESULTADOS: O escore de qualidade de sêmen apresentou sensibilidade de 80.45% e precisão de 77% para um “cutoff” de 93.1 na identificação do fator de infertilidade masculina. A área sob a curva “receiver operating characteristic” para o escore de qualidade de sêmen foi de 84.28% (95% intervalo de confiança: 65.22%-100%). Os níveis de espécies reativas de oxigênio [log10 (espécies reativas de oxigênio +1)] foram significativamente mais elevados nos pacientes portadores de fator de infertilidade masculina. A medica de espécies reativas de oxigênio apresentou sensibilidade de 83.47% e especificidade de 60.52% com uma precisão (definida como pacientes portadores do fator de infertilidade masculina com diagnóstico positivo e doadores corretamente excluídos) de 75% para um “cutoff” de 1.25 na identificação de pacientes portadores do fator de infertilidade masculina. A área sob a curva “receiver operating characteristic” para níveis de espécies reativas de oxigênio foi de 78.92% (95% intervalo de confiança: 72.60%-85.23%). Os escores de qualidade de sêmen correlacionaram negativamente com os níveis de espécies reativas de oxigênio tanto nos doadores e nos pacientes portadores do fator de infertilidade masculina.

CONCLUSÕES: O escore de qualidade de sêmen e os níveis espécies reativas de oxigênio nas amostras de sêmen parecem associar-se fortemente com o fator de infertilidade masculina. Na medida em que os dois parâmetros mostraram-se mais sensíveis que parâmetros espermáticos individuais na identificação do fator de infertilidade masculina, deveriam ser incluídos na avaliação rotineira de infertilidade.


REFERENCES


