

## New Semen Quality Scores Developed by Principal Component Analysis of Semen Characteristics

ASHOK AGARWAL,\* RAKESH K. SHARMA,\* AND DAVID R. NELSON†

From the \*Center for Advanced Research in Human Reproduction, Infertility, and Sexual Function, Urological Institute and Department of Obstetrics-Gynecology, and the †Department of Biostatistics and Epidemiology, The Cleveland Clinic Foundation, Cleveland, Ohio.

**ABSTRACT:** The purpose of this study was to determine whether semen characteristics can be reduced to 2 semen quality (SQ) scores and whether these new scores can help the clinician in assessing the reproductive outcome. A cross-sectional sample of 250 patients seeking infertility treatment were analyzed for semen characteristics. In addition, 177 male-factor patients (prostatitis with infection,  $n = 40$ ; varicocele,  $n = 77$ ; varicocele with infections,  $n = 11$ ; and vasectomy reversal,  $n = 43$ ) were also assessed. Sperm motion kinetics were measured by computer-assisted semen analysis (CASA) (concentration, percent motility, curvilinear velocity [VCL], straight-line velocity [VSL], average path velocity [VAP], linearity [LIN], and amplitude of lateral head displacement [ALH]). Sperm morphology was assessed by both World Health Organization (WHO) guidelines and Tygerberg strict criteria. The principal component analysis model was used to construct an SQ score and a relative semen quality (RQ) score. A separate set of 25 normal

donors was included as controls to determine normal ranges of the semen scores. Among the patient samples, SQ and RQ scores (median and 25% and 75% interquartile values) were 89.9, 25.1, and 130.4 and 106.1, 45.2, and 165.9, respectively. The SQ score for the varicocele and varicocele with infection groups was comparable ( $78.6 \pm 17.4$  and  $84.8 \pm 20.6$ ) but significantly different from the control ( $100 \pm 10$ ,  $P < .001$  and  $.03$ ). Vasectomy reversal patients had an SQ score of 78.2 plus or minus 16.8 that was significantly lower than controls ( $P < .001$ ). The correlation among semen characteristics allows for the efficient combining of semen measures. The composite scores can summarize overall SQ and quantity. Both SQ and RQ scores provide meaningful information on the quality of semen specimens for the clinician.

Key words: Computer-assisted semen analysis, infertility, spermatozoa, morphology.

J Androl 2003;24:343–352

Numerous studies have reported the predictive value of semen parameters such as concentration of motile spermatozoa (Fetterolf and Rogers, 1990; Liu et al, 1991), quantitative and computerized measurements of spermatozoa motility (Aitken et al, 1982; Irvine and Aitken, 1986; Jeulin et al, 1986; Barratt et al, 1992, 1993; Marshburn et al, 1992; Irvine et al, 1994; Krause, 1995; Macleod and Irvine, 1995), morphology (Kruger et al, 1988; Hinting et al, 1990; Menkveld et al, 1990; Enginsu et al, 1991; Check et al, 1992; Grow et al, 1994; Menkveld and Kruger, 1995; Eggert-Kruse et al, 1996), zona pellucida binding (Coddington et al, 1994), and occurrence of the acrosome reaction (Cummins et al, 1991; Aitken et al, 1994) to aid in the determination of in vitro fertilization and pregnancy outcome. After the introduction of computer-assisted semen analysis (CASA), the number of semen characteristics examined has increased to the extent

that each semen evaluation quantifies more than 10 semen characteristics. Although these characteristics are unique measures of semen quality (SQ), they are not independent of one another in the sense that patients with low motility tend to have low concentration and vice versa. Therefore, semen characteristics are positively correlated with each other. In biological systems, when there are several correlated variables, principal component analysis can be used to reduce these to 1 or 2 variables that are the linear functions of the original variables (Seber, 1984). This methodology has been widely used in disciplines ranging from gene expression data to scoring self-report surveys in studies of psychiatric patients, situations in which attempts are made to reduce the complexity of multiple correlated variables. This method also appears to be applicable to semen characteristics, in which several variables are recorded, with generally high correlation. Since many of these characteristics are interrelated, an overall semen score can be developed by an appropriate statistical model.

The present study was performed with the goal of providing a more efficient way of summarizing SQ and also predicting the outcome of natural conception and assisted conception techniques utilizing a 2-factor scoring system based on the results of a routine semen analysis. This was

This work was supported by funds from the Cleveland Clinic Foundation, Cleveland, Ohio.

Correspondence to: Dr Ashok Agarwal, Director, Center for Advanced Research in Human Reproduction, Infertility, and Sexual Function, Urological Institute, The Cleveland Clinic Foundation, 9500 Euclid Ave, Desk A19.1, Cleveland, OH 44195 (e-mail: agarwaa@ccf.org).

Received for publication September 13, 2002; accepted for publication November 12, 2002.

accomplished by applying principal component analysis to the original semen variables obtained by both CASA and conventional or manual semen analysis. The resultant scores may provide the clinicians with a more reliable and efficient method to predict SQ and the outcome of a specific assisted reproductive technology (ART) procedure.

## Materials and Methods

A cross-sectional sample of 250 consecutive patients seeking infertility treatment at the male infertility clinic of the Urological Institute provided semen samples. These samples were analyzed in the andrology laboratory to examine the variability and correlation of the patient semen characteristics. In addition, a separate set of 177 male-factor infertility patients with different clinical diagnoses was also assessed. These cases included prostatitis with infection ( $n = 46$ ); varicocele ( $n = 77$ ); varicocele with infection ( $n = 11$ ); and vasectomy reversal ( $n = 43$ ). A separate set of 25 donors was included in the study as controls to determine normal ranges of the semen scores. The Institutional Review Board approved this study.

### Assessment of Semen Variables

Sperm samples were analyzed on a CASA instrument (IVOS, V10.7s, Hamilton Thorne Research, Beverly, Mass) as described in the European Society of Human Reproduction and Embryology (ESHRE) Andrology Special Interest Group guidelines (1998). For each measurement, a 5- $\mu$ L sample aliquot from either a control or infertile patient sample was loaded on a MicroCell slide (Conception Technologies, La Jolla, Calif). Sperm motion kinetics measured by CASA included the following:

- 1) Sperm concentration ( $\times 10^6/\text{mL}$ ).
- 2) Percent motility.
- 3) Curvilinear velocity (VCL; micrometers per second): the time-average velocity of a sperm head along its actual curvilinear path, as perceived in 2 dimensions in the microscope.
- 4) Straight-line velocity (VSL; micrometers per second): the time-average velocity of a sperm head along the straight line between its first detected position and its last.
- 5) Average path velocity (VAP; micrometers per second): the time-average velocity of a sperm of a sperm head along its average path. This path is computed by smoothing the actual path according to algorithms in the CASA instrument: these algorithms vary between instruments.
- 6) Linearity (LIN; percent): the LIN of a curvilinear path,  $\text{VSL}/\text{VCL}$ .
- 7) Amplitude of lateral head displacement (ALH; micrometers): the magnitude of lateral displacement of a sperm head about its average path. This can be expressed as a maximum or an average of such displacement.

In addition to the computerized results, manual results were also calculated for sperm concentration and motility.

Standardization, strict quality control, and quality assurance

are critical and are strictly enforced in our program. The calibration setup of the CASA was as follows: 2 wells, 20  $\mu\text{m}$ , duration of data capture (frames); 15 (whole semen); minimum motile speed (micrometers per second): 2 (raw); maximum burst speed (micrometers per second): 600 (raw); distance scale factor (micrometers per pixel): 0.9457; centroid cell size minimum (pixels): 2; centroid cell size maximum (pixels): 8; number of cells to find per well: 200; and minimum number of fields per sample: 3. The results of CASA were manually verified. If the difference between the CASA and manual readings was within 20%, the CASA results were deemed valid; however, the manual results were used if the difference was greater than 20%. A high degree of correlation was seen between CASA and manual sperm counts ( $r^2 = 1$ , slope = 1) and motility ( $r^2 = 0.97$ , slope = 0.97) ( $n = 111$ ) and established the accuracy of CASA measurements. The reproducibility of the semen analyzer was determined by a videotape recording for calibration. An intra-assay variation of less than 10% was seen in CASA sperm counts and motility. Rejection criteria were greater than 2 standard deviations. The sperm count was  $38.3\text{--}42.4 \times 10^6/\text{mL}$  (30 frames/s), and sperm motility was 60.6%–75.0%.

### Measurement of Sperm Morphology

For morphological evaluation, seminal smears were stained with Giemsa stain (Diff-Quik, Baxter Scientific Products, McGraw Park, Ill). Sperm morphology was assessed by World Health Organization (WHO) guidelines (1999) and Tygerberg strict criteria (Mortimer and Menkveld, 2001).

### Calculation of the Semen Scores and Statistical Analyses

A principal component analysis model can be employed to calculate overall semen scores that accounts for most of variability observed among the battery of interrelated semen variables. Based on the fact that semen variables are interrelated, the relationships among the semen parameters and the feasibility of composite scores that represent most of the overall variability of the semen parameters were tested. The semen scores were calculated as follows:

- 1) Nine semen parameters identified according to the WHO guidelines were included: concentration, motility, sperm morphology according to WHO guidelines and Tygerberg strict criteria, VCL, VSL, VAP, LIN, and ALH.
- 2) Rather than using the raw values, base 10 logarithms of all the variables after adding a value of 1 (eg,  $\log(\text{concentration} + 1)$ ) were used. The reasons for this conversion were to reduce the effect of high outliers, to make the variables more comparable in scale (puts all variable measures in millions and those recorded in percentages on a more level playing field), and to completely reduce the effect of recording variables as  $\times 10^6$  or  $\times 10^4$ .
- 3) Principal component analysis was applied to the covariance matrix of the 9 log-transformed semen parameters. This produced 9 new components, which were weighted sums of the original variables. In matrix algebra nomenclature, the weights are referred to as "eigenvectors." We were interested mainly in components that would account for at least 10%

of the overall variability of the 9 semen parameters. The use of a log transformation will result in variables with the highest coefficient of variation being assigned higher weights.

- 4) Since the correlations among the semen parameters were positive, it was anticipated that most of the variability would be explained by a weighted sum of the whole variables, which would be related to the overall SQ.
- 5) The first principal component accounted for 64.8%, or almost two thirds, of the overall variability observed among the 9 variables. Therefore, this first principal component summed up all the semen parameters with varying weights given to each component. This score was referred to as the “SQ score” to represent the overall SQ. The SQ score was calculated as follows:

SQ score

$$\begin{aligned}
 &= (\log \text{ conc.} \times 0.807647 + \log \text{ motility} \times 0.254114 \\
 &\quad + \log \text{ strict criteria morphology} \times 0.331037 \\
 &\quad + \log \text{ WHO morphology} \times 0.274769 \\
 &\quad + \log \text{ VCL} \times 0.146286 + \log \text{ VSL} \times 0.193481 \\
 &\quad + \log \text{ VAP} \times 0.175990 + \log \text{ LIN} \times 0.009109 \\
 &\quad + \log \text{ ALH} \times 0.090307) \times (10/0.2901900) \\
 &\quad - 20.1911369
 \end{aligned}$$

- 6) The second principal component accounted for 15.6% of the variability from the 35.2% not accounted for by the SQ score. This second variable was a weighted sum of 8 variables, which subtracts concentration multiplied by the weight. High scores on this scale are related to high motility, good morphology, or high scores on other parameters relative to their concentration. Therefore, this score was “relative SQ score” or RQ score. The calculation was as follows:

RQ score

$$\begin{aligned}
 &= (\log \text{ conc.} \times -0.567457 + \log \text{ motility} \times 0.147947 \\
 &\quad + \log \text{ strict criteria morphology} \times 0.605335 \\
 &\quad + \log \text{ WHO morphology} \times 0.418694 \\
 &\quad + \log \text{ VCL} \times 0.180399 + \log \text{ VSL} \times 0.202853 \\
 &\quad + \log \text{ VAP} \times 0.189159 + \log \text{ LIN} \times 0.009502 \\
 &\quad + \log \text{ ALH} \times 0.069321) \times (10/0.1813407) \\
 &\quad + 24.5183708.
 \end{aligned}$$

- 7) The 2 new semen scores, SQ and RQ, together account for more than 80% of the variability observed among the original 9 semen parameters. Because these scores were produced by principal component analysis, one feature that they have is that SQ and RQ scores are not correlated.

The newly constructed principal component scores were scaled on the basis of a sample of 25 normal, healthy donors to be converted to semen scores that would average 100 with a standard deviation of 10 among this donor group, such as that provided for intelligence quotient (IQ) scores.

Similarly, the same procedure was undertaken with the 4 non-

CASA characteristics: concentration, motility, and 2 morphology measures. Pearson product-moment correlations between the 9-characteristic and 4-characteristic semen scores were calculated. The broken-stick test was used to determine whether the principal component analysis provided variables that accounted for more overall variation than chance alone. Student’s *t* tests were used to compare semen scores of various diagnoses. Calculations were performed with SAS version 8.1 software (SAS Institute Inc, Cary, NC).

## Results

### *SQ and RQ as Measures of Semen Score*

The correlations among the semen samples were generally positive (Table 1). Among the kinetic variables, high correlation was seen between VAP and VSL ( $r = 0.98$ ); VAP and VCL ( $r = 0.96$ ); and VSL and VCL ( $r = 0.91$ ). The least nonsignificant positive correlation was observed between LIN and strict criteria morphology ( $r = 0.01$ ). Also, a weak nonsignificant negative correlation was observed between LIN and both VCL ( $r = -0.07$ ) and ALH ( $r = -0.03$ ). Given the overall positive correlation, it was anticipated that most of the variability would be explained by a weighted sum of the variables that would be related to the overall SQ.

Principal component analysis indicated that the 2 semen scores or components account for 80.3% of the variability observed among the 9 semen parameters, which was significantly better than expected by chance alone ( $P < .001$ ). The first principal component was termed the “SQ” score to represent “overall SQ.” This was a weighted sum of all semen characteristics and accounted for 64.8% or almost two thirds of the overall variability observed among the 9 variables. The SQ score weighs each semen parameter on the basis of the values listed in Table 2. For example, the variable with the greatest weight was sperm concentration (+0.81). All of the other scores also had positive weights. Therefore, the SQ score sums up all the semen characteristics with varying weights given to each score.

The second component was termed an “RQ” score since it represents “relative quality.” It assigns a negative weight for concentration and positive weights for all other variables. Therefore, an individual with relatively low concentrations but above-average values for other characteristics would have a high RQ score compared to what would be expected on the basis of other characteristics. In contrast, low scores on this scale will have poor motility, poor morphology, or low scores on other characteristics relative to their concentration. Therefore, the 2 new scores together accounted for more than 80% of the variability observed among the 9 semen characteristics. Because these 2 scores are produced by principal component analysis, SQ and RQ are not correlated.

Table 1. Summary statistics and correlation among 9 semen characteristics\*†

Variable	Original Variable (Mean ± SD)	Log Transformed (Mean ± SD)	Correlation Among Log-Transformed Variables (r)								
			Conc.	Motility (%)	VCL (μm/s)	VSL (μm/s)	VAP (μm/s)	Linearity (μm/s)	ALH (μm)	WHO	Strict Criteria
Conc. (×10 <sup>6</sup> /mL)	53.93 ± 57.08	1.48 ± 0.56	1.00	0.55	0.41	0.47	0.46	0.07	0.32	0.52	0.48
Motility (%)	58.01 ± 22.72	1.72 ± 0.23	...	1.00	0.52	0.63	0.59	0.35	0.26	0.43	0.45
VCL (μm/s)	41.01 ± 13.38	1.60 ± 0.15	...	...	1.00	...	...	...	0.32	...	...
VSL (μm/s)	22.98 ± 9.35	1.34 ± 0.18	...	...	0.91	1.00	...	...	0.32	...	...
VAP (μm/s)	27.74 ± 10.17	1.43 ± 0.17	...	...	0.96	0.98	1.00	...	0.32	...	...
Linearity (%)	52.70 ± 7.20	1.73 ± 0.06	...	...	-0.07	0.30	0.14	1.00	-0.03	0.03	0.01
ALH (μm)	4.10 ± 1.28	0.69 ± 0.16	...	...	0.45	0.46	0.46	...	1.00	1.00	...
WHO (%)	31.84 ± 13.74	1.47 ± 0.24	...	...	0.45	0.47	0.47	...	0.31	0.85	...
Strict criteria (%)	9.14 ± 5.12	0.93 ± 0.29	...	...	0.45	0.47	0.47	...	0.32	1.00	...

\* Log<sub>10</sub> transformation after adding "1" to the original value.

† Conc. indicates concentration (×10<sup>6</sup>/mL); VCL, curvilinear velocity; VSL, straight-line velocity; VAP, average path velocity; ALH, amplitude of lateral head displacement; and WHO, World Health Organization.

Subsequently, the SQ and RQ scores were applied to samples from normal, healthy donors such that these 2 scores could be scaled to a mean of 100 and a standard deviation of 10. This involved both addition and multiplication of the values by a number of constants to derive the standardized scale. A mean of 100 was chosen, simply because it provided a level of recognition greater than 100 and had a quality better than the average male, and vice versa for scores less than 100, similar to the IQ scores, which are also standardized to a mean of 100. The standard deviation of 10 was chosen because it provided an easily computable difference from 100 or any other value of interest. For example, a population of normal individuals can often fall into a range of +2 standard deviations from the mean; therefore, the semen score would generally be expected to fall within a range of 80–120 for a group of donors. The steps involved in the calculation of the scores are illustrated in Figure 1. Examples of the semen characteristics that produce the 5 highest and the 5 lowest SQ and RQ scores are shown in Table 3. The 5 lowest SQ scores represent poor SQ. The highest and lowest SQ scores are more obvious in interpretation. The sample with the highest SQ score had a concentration of 414 × 10<sup>6</sup>/mL, a motility of 95%, and a normal morphology (by WHO guidelines) of 40%, whereas the sample with the lowest SQ score had a concentration of 1.02 × 10<sup>6</sup>/mL, a motility of 6%, and a WHO morphology of 11%. In addition, the 5 highest values for RQ scores showed that the concentration was lower than expected on the basis of the other above-average semen parameters. The RQ score is not as obvious, because it is based on the relationship of 8 parameters compared to their concentration. For instance, the highest RQ score was a sample with a concentration of only 0.56 × 10<sup>6</sup>/mL but with a motility of 39% and a normal sperm morphology of 48%. In contrast, the individual with the lowest RQ score had a concentration of 24.3 × 10<sup>6</sup>/mL, with a motility of 36% and a normal sperm morphology of 1%. Among the entire group of 250 patient samples, the average SQ score (median and 25% and 75% interquartile values) was 89.9, 25.1, and 130.4, and the average RQ score was 106.1, 45.2, and 165.9, respectively. The distribution of the 250 patients and 25 donors for both SQ and RQ scores is illustrated in Figure 2.

Patients are often classified as oligozoospermic, asthenozoospermic, or teratozoospermic on the basis of concentration, motility, and morphology or any of these combinations. To illustrate how these classifications are related to the semen scores, the 250 patients in our study were classified into 8 groups (Figure 3). Patients without any abnormal semen parameters had average SQ and RQ scores greater than 100. Patients that were either asthenozoospermic or teratozoospermic had an average SQ score of only 90. Since concentration is the parameter that is as-

Table 2. Weights to calculate semen quality and relative quality scores in 250 patients\*†

Variable	CASA-Derived Semen Score Weights		Non-CASA Semen Score Weights	
	SQ score	RQ score	SQ score	RQ score
Concentration (×10 <sup>6</sup> /mL)	0.81	-0.57	0.86	-0.47
Motility (%)	0.25	0.15	0.24	0.11
VCL (μm/s)	0.15	0.18	...	...
VSL (μm/s)	0.19	0.20	...	...
VAP (μm/s)	0.18	0.19	...	...
Linearity (%)	0.009	0.001	...	...
ALH (μm)	0.09	0.07	...	...
Sperm morphology				
WHO (%)	0.27	0.42	0.27	0.51
Strict criteria (%)	0.33	0.61	0.34	0.71
Percentage of total variability accounted for by each score				
	64.8	15.6	72.9	17.6

\* Scores computed by multiplying each log-transformed variable by the semen score weights.

† VCL indicates curvilinear velocity; VSL, straight-line velocity; VAP, average path velocity; ALH, amplitude of lateral head displacement; WHO, World Health Organization; CASA, computer-assisted semen analysis; SQ, overall semen quality score; and RQ, relative quality score.

signed the maximum weight in defining the SQ score, patients who were oligospermic had an average SQ score that was only slightly greater than 80, similar to the score in asthenoteratozoospermic patients. When patients with oligozoospermia were combined with those patients that were either asthenozoospermic or teratozoospermic, SQ values averaged between 70 and 80. Patients with all 3 combinations (oligoasthenoteratozoospermic) had an average SQ score of less than 60. Therefore, SQ scores decrease as the number of abnormal characteristics increases. This plot also helps in interpreting perhaps the less intuitive RQ score, which also is a measure of asthenoteratospermia, because it examines the quality of the non-

concentration characteristics. Patients identified as asthenoteratozoospermic had an average RQ score of less than 90 compared to oligozoospermic patients, who had an average RQ score of 130. The latter patients' RQ score is high, in the sense that their sperm concentration is poor, but other motility and morphology values are in the normal range. As in normal patients, the RQ score in oligoasthenoteratozoospermic patients was close to 100, because their motility and morphology were, as expected, based on their poor sperm concentration.

As part of the validation process, the 2 SQ scores were applied to patients with various clinical diagnoses and were also compared with the semen scores in donors.

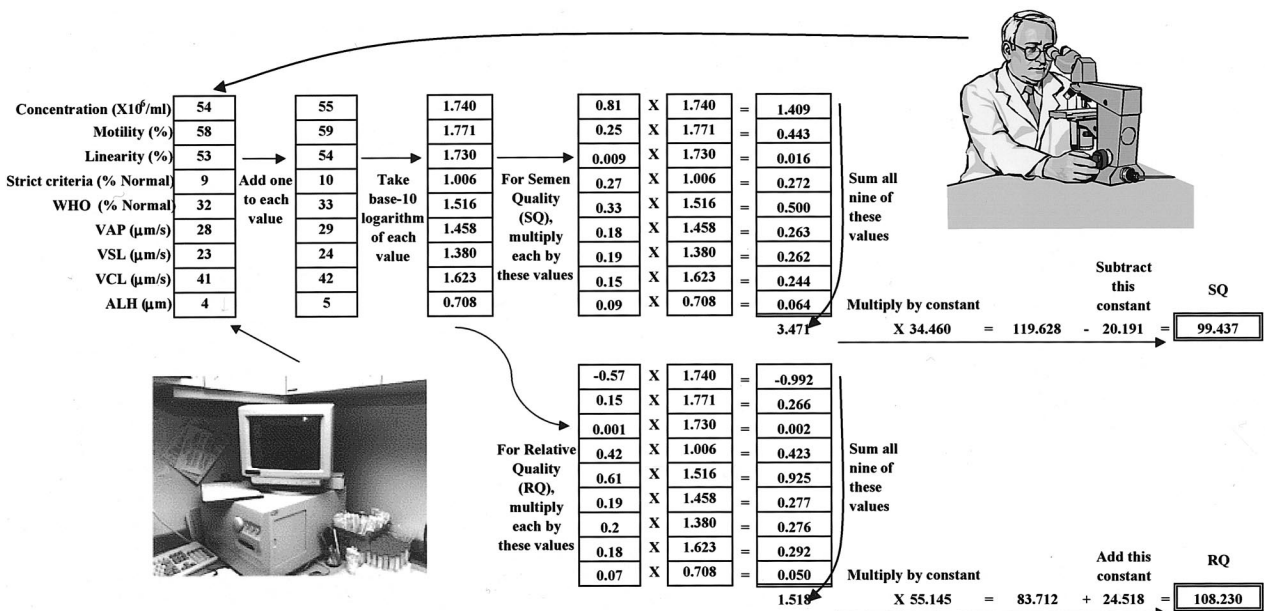


Figure 1. Steps involved in computing semen quality (SQ) and relative quality (RQ) scores.

Table 3. Semen parameters of the 5 highest and 5 lowest scores observed for SQ and RQ among the samples of 250 patients\*

Concentration ( $\times 10^6/\text{mL}$ )	Motility (%)	VCL ( $\mu\text{m}/\text{s}$ )	VSL ( $\mu\text{m}/\text{s}$ )	VAP ( $\mu\text{m}/\text{s}$ )	Linearity (%)	ALH ( $\mu\text{m}$ )	WHO (%)	Strict Criteria (%)	SQ Score	RQ Score
Five lowest values of SQ score (overall semen quality)										
1.02	6	9.1	18	7.2	45	0	11	0	25.08	81.07
0.4	32	7.1	11.7	6	71	0	5	2	27.23	97.11
0.36	22	7.9	14.4	4.9	46	0	15	2	29.70	106.39
0.48	42	18.3	43.5	12.8	50	0	6	0	31.10	95.48
0.96	29	9.7	15.8	8.2	54	2.2	10	1	35.14	98.07
Five highest values of SQ score (overall semen quality)										
198	91	45.8	54	42.3	69	2.9	54	17	124.12	106.96
230.8	89	37.3	52	31.7	53	4.5	50	16	124.19	101.30
389.4	97	34.8	52.6	28.7	53	4.3	34	7	125.05	79.03
193.7	80	62.9	88.2	54.9	57	5.8	62	19	127.69	115.31
414	95	44	59.7	37.6	56	4.2	40	12	130.38	89.58
Five lowest values of RQ score (relative semen quality)										
24.3	36	18.4	27	15.6	55	3.3	1	0	61.20	45.17
9.6	27	13.5	21.3	10.8	55	5.3	3	0	50.74	59.63
21.3	49	16.7	24	12.4	53	3.5	8	0	65.95	61.16
7	36	17.4	27.8	13.7	51	4.9	2	0	48.94	64.68
77.1	87	35	49.5	28.3	54	3.9	10	2	95.31	74.27
Five highest values of RQ score (relative semen quality)										
3	39	22.1	32.5	18	55	3.7	39	13	65.96	141.18
0.2	50	28.3	47.3	14.2	36	2.3	26	7	48.19	147.24
2.04	65	38.2	60.3	33.4	55	5.6	32	10	67.46	149.72
5.8	56	44.1	61.1	40.7	56	5.1	65	14	81.82	151.12
0.56	39	26.3	45.4	19.5	37	7.2	48	21	59.73	165.94

\* VCL indicates curvilinear velocity; VSL, straight-line velocity; VAP, average path velocity; ALH, amplitude of lateral head displacement; WHO, World Health Organization; SQ, overall semen quality score; and RQ, relative quality score.

Nonsignificant differences in SQ would indicate a lack of validity of the scores, whereas significant differences would provide initial evidence of their applicability. All 4 clinical diagnoses had significantly lower SQ scores

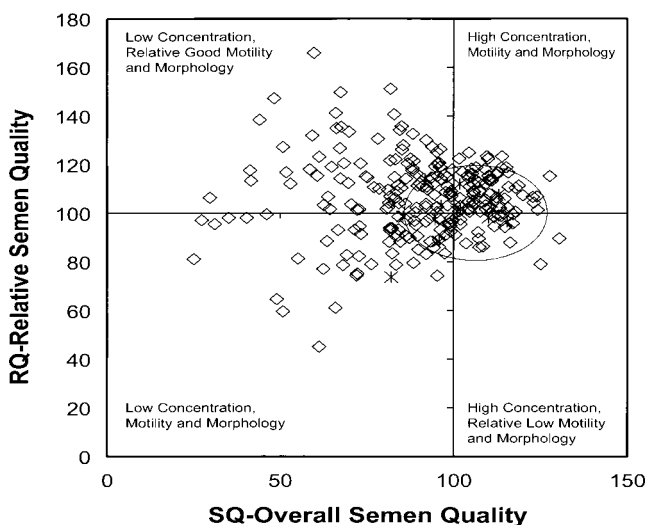


Figure 2. Distribution of semen scores of 250 infertile patients assessed for male infertility ( $\diamond$ ) and 25 donors (\*). The circle encompasses 2 standard deviations from 100, where most donors' scores should fall.

than the donor control group ( $P < .03$ ) (Table 4). These groups averaged SQ values 15–22 points below donors. Using multivariate logistic regression analysis to determine the ability of the semen scores to discriminate between donors and patients, the SQ score had an odds ratio of 2.53 (95% CI, 1.36–4.72;  $P < .004$ ); the RQ score had an odds ratio of 0.86 (95% CI, 0.51–1.46;  $P < .58$ ).

We next calculated the conventional semen score. In this score only (conventional or manual), semen characteristics (ie, sperm concentration, percent motility, and percent normal forms according to both WHO morphology and Tygerberg strict criteria) were used. The principal component results for these 4 variables are shown in Table 3. A highly significant positive correlation was seen between the CASA SQ score and the conventional SQ score ( $r = 0.99$ ). Similarly, a very strong correlation was seen for the RQ score between the CASA RQ score and the conventional RQ score (4 variables) ( $r = 0.96$ ).

## Discussion

Manual semen analysis using light microscopy has been the standard method for semen analysis in most labora-

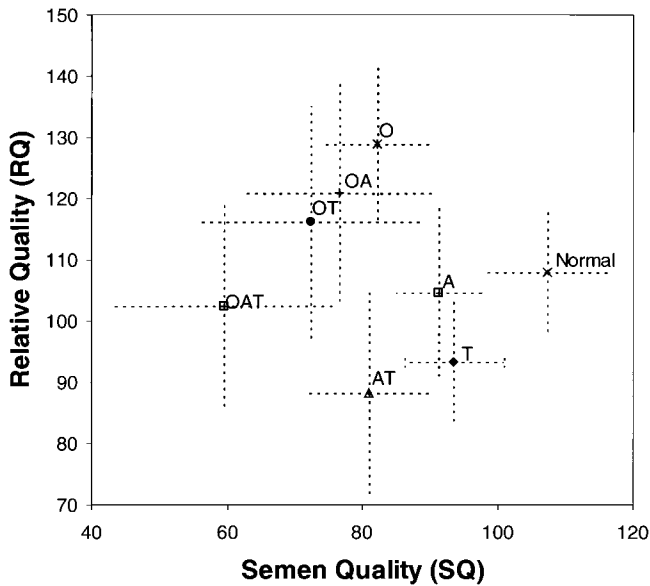


Figure 3. Distribution of mean and standard deviations of the 20 men classified as normal, asthenozoospermic (A), teratozoospermic (T), oligozoospermic (O), oligoasthenozoospermic (OA), oligoteratozoospermic (OT), asthenozoospermic (AT), or oligoasthenoteratozoospermic (OAT).

tories. Although an easy test to perform, meticulous attention to details and techniques is essential in order to obtain an accurate and reproducible analysis (Keel and Webster, 1990; Mortimer, 1990). However, manual analyses can be very subjective and prone to within- and between-observer technical error (Keel and Webster, 1990). Any laboratory analysis is subject to 2 potential sources of error: random (sampling error) and systematic (observer bias). At present, methods used for human semen evaluation vary substantially. These methods range from those recommended by the WHO (1999) to detailed CASA of sperm motion characteristics, morphometric evaluation of sperm shape, and various physical and biochemical analyses (Boyle et al, 1992; Barratt et al, 1993; Macleod and Irvine, 1995). These measures lack sufficient corresponding information for accurate characterization of their relationships with fertility. According to WHO guidelines (1999), a sperm concentration below  $20 \times 10^6/\text{mL}$  is abnormal. However, these guidelines are not

based on studies of fertility but are arbitrarily defined by a committee of international experts. The so-called normal values provided by the WHO manuals for the basic semen parameters (ie, volume and qualitative and quantitative motility and morphology) were obtained mostly through studies of the so-called fertile populations (WHO, 1987, 1992). This may be a reason why the clinical value of traditional semen characteristics in the assessment of male infertility is a subject of considerable debate (Tomlinson et al, 1999). Previous reference ranges for normal men were considerably stricter and delineated the lower range for sperm concentration at  $60 \times 10^6/\text{mL}$ . Multiple studies have demonstrated the degree to which sperm characteristics may be of predictive value (ESHRE, 2000).

Sperm motility characteristics are important semen characteristics, which, before the 1980s, could only be evaluated subjectively (Aitken et al, 1982; Mortimer et al, 1986; Wichmann et al, 1994). The introduction of CASA made it possible to measure motility characteristics of individual spermatozoa. CASA is widely accepted as providing a rapid and objective measurement of individual "classical" sperm characteristics such as sperm count and sperm movement. CASA also increases the accuracy and reproducibility of sperm count and motility (Keel and Webster, 1990) and allows the determination of sperm motion characteristics, termed "kinematics," which cannot be obtained by microscopic observation, and may be important in determining the fertility potential of sperm (Seber, 1984). A comparison of the measurements obtained by CASA with those obtained by conventional means has shown the potential for discrepancy due to the methodology used (ESHRE, 2000). These sperm motion characteristics particularly have been correlated with the sperm penetration of cervical mucus, the sperm penetration of human oocytes, and the results of in vitro fertilization (Aitken et al, 1982). It is often useful to make comparisons of semen parameters prior to and after treatments such as varicocele or medical/hormonal treatment of the infertile male. It may also be useful to compare sperm characteristics and response to

Table 4. Semen scores in 177 men with clinical diagnoses and 19 healthy donors (controls)\*

Diagnosis	SQ (Mean ± SD)	P†	RQ (Mean ± SD)	P‡
Control (n = 19)	100.0 ± 10.0	...	100.0 ± 10.0	...
Prostatitis with infection (n = 46)	83.3 ± 18.0	.001	102.4 ± 19.9	.63
Varicocele (n = 77)	78.6 ± 17.7	<.0001	104.0 ± 15.7	.41
Varicocele with infection (n = 11)	84.8 ± 20.6	.03	109.4 ± 15.4	.17
Vasectomy reversal (n = 43)	78.2 ± 16.8	<.0001	98.7 ± 18.0	.80

\* SQ indicates overall semen quality score; RQ, relative quality score.

† Pairwise comparisons for SQ between control and other groups.

‡ Pairwise comparisons for RQ between control and other groups.

sperm treatments (such as sperm washing methodology) over lengthy time periods of inseminations or treatments.

Perhaps the most widely utilized semen characteristic is sperm count. Men with less than  $20 \times 10^6$  spermatozoa/mL are typically deemed subfertile, and men with counts less than  $5 \times 10^6$  spermatozoa/mL are often considered infertile. Semen samples containing less than 14% normal forms by Tygerberg strict criteria are reported as subfertile, and those containing less than 5% normal forms are considered severely impaired, causing some centers to recommend couples discontinue treatment and move on to donor insemination. However, like its predecessors, strict sperm morphology is not absolutely accurate in predicting fertility. In recent years, numerous authors have realized the deficiencies in using these measures of semen parameters (Huszar et al, 1988a,b; Aitken et al, 1992; Davis and Katz, 1993; Mortimer, 1994; Clements et al, 1995). The reported predictive value of spermatozoa concentration, for example, has been criticized because of the natural day-to-day variations that occur in spermatozoa concentration (Huszar et al, 1988a,b). The use of CASA technology has been criticized because of difficulties in achieving optimum set-up procedures (Davis and Katz, 1993; Mortimer, 1994; Clements et al, 1995) and because of the operational difficulties of using routine CASA technology (Boone et al, 2000; Carrell, 2000; Oehninger et al, 2000). Many spermatozoa function tests today are included as part of a conventional semen analysis designed to detect deficiencies of the fertilizing potential of mature spermatozoa that otherwise may show optimal motility, morphology, and concentration values (Cross et al, 1986; Barratt et al, 1992, 1993; Marshburn et al, 1992; Benoff et al, 1993; Coddington et al, 1994; Irvine et al, 1994; Krause, 1995; Macleod and Irvine, 1995; Moutaffian and Parinaud, 1995). It is a basic principle of statistical modeling that a set of predictive characteristics must be tested on another set of data to determine if the analysis is valid and hence useful.

In this study, we examined a statistical model utilizing the principal components to narrow down the elaborate measures of semen analysis to 2 main components or scores. The results demonstrate that SQ and RQ scores were able to account for the majority of the variability of 9 individual semen characteristics. A majority of the variables demonstrated positive correlation with these scores. Among the patients, the average SQ score (median and 25% and 75% interquartile value) was 89.9, 25.1, and 130.4, respectively. In addition, using different subsets of infertile patients with various clinical diagnoses, we were able to measure the SQ and RQ scores. Therefore, SQ scores decrease as the number of abnormal parameters increases. A low SQ score reflects low concentration, motility, and morphology, whereas a high SQ score repre-

sents high concentration and relatively good motility and morphology.

Significantly lower SQ scores were seen in men with varicocele and in patients following vasectomy reversal than in controls. Therefore, SQ scores were significantly poor in all clinical diagnoses, but they did not differ among themselves. No significant differences were observed between donors and the clinical diagnosis groups for the RQ score, indicating that their concentration levels were not significantly different from that expected on the basis of their other characteristics. From our results, we can conclude that the SQ score can significantly discriminate between the normal control donors and patients, whereas RQ was not a good discriminator of SQ in these patient groups; this also confirms our univariate findings.

Our next objective was to examine how this score could be utilized effectively in those programs for which CASA is not available and, secondly, how much information is really provided by the sperm kinematics associated with CASA. As seen from our findings, very little information will be lost if the CASA characteristics are excluded and only conventional or manual scores are used. Therefore, programs that are not equipped with more sophisticated computer analysis systems for semen analysis can still have an effective means of interpreting the results of semen analysis. However, it is imperative that these programs have high standards of quality control and efficient and well-trained technical staff to minimize within-laboratory variations in semen analysis results (Clements et al, 1995; Cooper et al, 1999; Auger et al, 2000; Keel et al, 2000).

The semen score can provide important information on the SQ and the likelihood of establishing a pregnancy. Also, semen scores provide more meaningful information than the individual semen characteristics. However, it is important to understand the underlying cell biology of the spermatozoon. The ejaculate is composed of subpopulations of spermatozoa that differ in maturity, physiological status with respect to the ability to undergo capacitation and acrosome reaction, and fertilizing ability (Huszar et al, 1988a; Aitken et al, 1992).

What are the clinical applications of our findings? The utility of SQ and RQ scores for clinical practice may differ depending on the clinical situation. Because all semen characteristics were included in the derivation of the SQ score, this score may have greater utility in assessing the male fertility status for natural conception. On the other hand, the RQ score is derived from the measurement of morphology, motility, and motion parameters after adjusting for concentration. Therefore, the RQ score may be more helpful in artificial reproduction methods, where concentration may not be as essential as the other parameters.

In conclusion, we have demonstrated that semen char-

acteristics can be reduced to 2 SQ scores, which account for more than 80% of the variability expressed by all of the semen characteristics individually. We believe that reducing the 9 semen characteristics to 2 scores will be more efficient by allowing quick comparisons of SQ. In addition, the semen scores may provide improved assessment of male fertility. Patient scores less than 80 are below the expected normal range of donors. Similar information is obtained using either CASA or conventional or manual semen analysis variables. Given the outcome of our studies, SQ scores may provide a more reliable alternative to the prediction of ART outcome, particularly in couples with male-factor or idiopathic infertility. This can be accomplished by examining a database that includes 2 sets of patients undergoing IUI, one that resulted in successful pregnancies and the second that failed to establish pregnancies. In the accompanying article, we are extending our semen score model in establishing the clinical usefulness of these new measures of SQ compared to simple routine variables of semen analysis. The clinical value of this analysis, such as optimum criteria that can maximize the clinical outcome of predicting in vivo and in vitro conception, is discussed.

## Acknowledgments

The authors wish to thank the staff of the Clinical Andrology Laboratory for their technical assistance and Robin Verdi for her secretarial assistance.

## References

- Aitken RJ, Best FSM, Richardson DW, Djahanbakhch O, Mortimer D, Templeton AA, Lee MM. An analysis of sperm function in cases of unexplained infertility: conventional criteria, movement characteristics, and fertilizing capacity. *Fertil Steril.* 1982;38:212–221.
- Aitken RJ, Buckingham D, West K, Wu FC, Zikopoulos K, Richardson DW. Differential contribution of leukocytes and spermatozoa to the generation of reactive oxygen species in the ejaculates of oligospermic patients and fertile donors. *J Reprod Fertil.* 1992;94:451–462.
- Aitken RJ, West K, Buckingham D. Leukocytic infiltration into the human ejaculate and its association with semen quality, oxidative stress, and sperm function. *J Androl.* 1994;15:343–352.
- Auger J, Eustache F, Ducot B, et al. Intra- and inter-individual variability in human sperm concentration, motility and vitality assessment during a workshop involving ten laboratories. *Hum Reprod.* 2000;15:2360–2368.
- Barratt CL, McLeod ID, Dunphy BC, Cooke ID. Prognostic value of two putative sperm function tests: hypo-osmotic swelling and bovine sperm mucus penetration test (Penetrak). *Hum Reprod.* 1992;7:1240–1244.
- Barratt CLR, Tomlinson MJ, Cooke ID. Prognostic significance of computerized motility analysis for in vivo fertility. *Fertil Steril.* 1993;60:520–525.
- Benoff S, Cooper GW, Hurley I, Napolitano B, Rosenfeld DL, Scholl GM, Hershlag A. Human sperm fertilizing potential in vitro is correlated with differential expression of a head-specific mannose-ligand receptor. *Fertil Steril.* 1993;59:854–862.
- Boone WR, Jones JM, Shapiro SS. Using videotaped specimens to test quality control in a computer-assisted semen analysis system. *Fertil Steril.* 2000;73:636–640.
- Boyle CA, Khoury MJ, Katz DF, Annett JL, Kresnow M-J, DeStefano F, Schrader S. The relation of computer-based measures of sperm morphology and motility to male infertility. *Epidemiology.* 1992;3:239–246.
- Carrell DT. Semen analysis at the turn of the century: an evaluation of potential uses of new sperm function assays [review]. *Arch Androl.* 2000;44:65–75.
- Check JH, Bollendorf A, Press M, Blue T. Standard sperm morphology as a predictor of male fertility potential. *Arch Androl.* 1992;28:39–41.
- Clements S, Cooke ID, Barratt CL. Implementing comprehensive quality controls in the andrology laboratory. *Hum Reprod.* 1995;10:2096–2106.
- Coddington CC, Oehninger SC, Olive DL, et al. Hemizona index (HZI) demonstrates excellent predictability when evaluating sperm fertilizing capacity in in vitro fertilization patients. *J Androl.* 1994;15:250–254.
- Cooper TG, Atkinson AD, Nieschlag E. Experience with external quality control in spermatology. *Hum Reprod.* 1999;14:765–769.
- Cross NL, Morales P, Overstreet JW. Two simple methods for detecting acrosome-reacted human sperm. *Gamete Res.* 1986;15:213–226.
- Cummins JM, Pember SM, Jequier AM, Yovich JL, Hartmann PE. A test of the human sperm acrosome reaction following ionophore challenge. Relationship to fertility and other seminal parameters. *J Androl.* 1991;12:98–103.
- Davis RO, Katz DF. Operational standards for CASA instruments. *J Androl.* 1993;14:385–394.
- Eggert-Kruse W, Schwarz H, Rohr G, Demirakca T, Tilgen W, Runnebaum B. Sperm morphology assessment using strict criteria and male fertility under in-vivo conditions of conception. *Hum Reprod.* 1996;11:139–146.
- Enginsu ME, Dumoulin JC, Pieters MH, Bras M, Evers JL, Geraedts JP. Evaluation of human sperm morphology using strict criteria after Diff-Quik staining correlation of morphology with fertilization in vitro. *Hum Reprod.* 1991;6:854–858.
- ESHRE Andrology Special Interest Group. Guidelines on the application of CASA technology in the analysis of spermatozoa. *Hum Reprod.* 1998;13:142–145.
- ESHRE Capri Workshop Group. Optimal use of infertility diagnostic tests and treatments. *Hum Reprod.* 2000;15:723–732.
- Fetterolf PM, Rogers BJ. Prediction of human sperm penetrating ability using computerized motion parameters. *Mol Reprod Dev.* 1990;27:326–331.
- Grow DR, Oehninger S, Seltman HJ, Toner JP, Swanson RJ, Kruger TF, Muasher SJ. Sperm morphology as diagnosed by strict criteria: probing the impact of teratozoospermia on fertilization rate and pregnancy outcome in a large in vitro fertilization population. *Fertil Steril.* 1994;62:559–567.
- Hinting A, Comhaire F, Vermeulen L, Dhont M, Vermeulen A, Vandekerckhove D. Value of sperm characteristics and the result of in-vitro fertilization for predicting the outcome of assisted reproduction. *Int J Androl.* 1990;13:59–66.
- Huszar G, Corrales M, Vigue L. Correlation between sperm creatine phosphokinase activity and sperm concentrations in normospermic and oligospermic men. *Gamete Res.* 1988a;19:67–75.
- Huszar G, Vigue L, Corrales M. Sperm creatine phosphokinase activity as a measure of sperm quality in normospermic, variable spermic, and oligospermic men. *Biol Reprod.* 1988b;38:1061–1066.

- Irvine DS, Aitken RJ. Predictive value of in-vitro sperm function tests in the context of an AID service. *Hum Reprod.* 1986;1:539–545.
- Irvine DS, Macleod IC, Templeton AA, Masterton A, Taylor A. A prospective clinical study of the relationship between the computer-assisted assessment of human semen quality and the achievement of pregnancy in vivo. *Hum Reprod.* 1994;9:2324–2334.
- Jeulin C, Feneux D, Serres C, Jouannet P, Guillet-Rosso F, Belaisch-Allart J, Frydman R, Testart J. Sperm factors related to failure of human in-vitro fertilization. *J Reprod Fertil.* 1986;76:735–744.
- Keel BA, Quinn P, Schmidt CF, Serafy NT Jr, Serafy NT Sr, Schalue TK. Results of the American Association of Bioanalysts national proficiency testing programme in andrology. *Hum Reprod.* 2000;15:680–686.
- Keel BA, Webster BW. The semen analysis. In: Keel BA, Webster BW, eds. *CRC Handbook of the Laboratory Diagnosis and Treatment of Infertility.* Boston, Mass: CRC Press; 1990:27–69.
- Krause W. Computer-assisted semen analysis systems: comparison with routine evaluation and prognostic value in male fertility and assisted reproduction. *Hum Reprod.* 1995;10:60–66.
- Kruger TF, Acosta AA, Simmons KF, Swanson RJ, Matta JF, Oehninger S. Predictive value of abnormal sperm morphology in in vitro fertilization. *Fertil Steril.* 1988;49:112–117.
- Liu DY, Clarke GN, Baker HW. Relationship between sperm motility assessed with the Hamilton-Thorne motility analyzer and fertilization rates in vitro. *J Androl.* 1991;12:231–239.
- Macleod IC, Irvine DS. The predictive value of computer-assisted semen analysis in the context of donor insemination programme. *Hum Reprod.* 1995;10:580–586.
- Marshburn PB, McIntire D, Carr BR, Byrd W. Spermatozoal characteristics from fresh and frozen donor semen and their correlation with fertility outcome after intrauterine insemination. *Fertil Steril.* 1992;58:179–186.
- Menkveld R, Kruger TF. Advantages of strict (Tygerberg) criteria for evaluation of sperm morphology [review]. *Int J Androl.* 1995;18:36–42.
- Menkveld R, Stander FS, Kotze TJ, Kruger TF, van Zyl JA. The evaluation of morphological characteristics of human spermatozoa according to stricter criteria. *Hum Reprod.* 1990;5:586–592.
- Mortimer D. Objective analysis of sperm motility and kinematics. In: Keel BA, Webster BW, eds. *CRC Handbook of the Laboratory Diagnosis and Treatment of Infertility.* Boston, Mass: CRC Press; 1990:97–133.
- Mortimer D, ed. *Practical Laboratory Andrology.* New York, NY: Oxford University Press; 1994.
- Mortimer D, Menkveld R. Sperm morphology assessment—historical perspectives and current opinions. *J Androl.* 2001;22:192–205.
- Mortimer D, Shu MA, Tan R. Standardization and quality control of sperm concentration and sperm motility counts in semen analysis. *Hum Reprod.* 1986;1:299–303.
- Moutaffian H, Parinaud J. Selection and characterization of human acrosome reacted spermatozoa. *Hum Reprod.* 1995;10:2948–2951.
- Oehninger S, Franken DR, Sayed E, Barroso G, Kolm P. Sperm function assays and their predictive value for fertilization outcome in IVF therapy: a meta-analysis. *Hum Reprod Update.* 2000;6:160–168.
- Seber GAF. Multivariate observations. In: Seber GAF, ed. *Dimension Reduction and Ordination.* New York, NY: Wiley; 1984:175–278.
- Tomlinson MJ, Kessopoulou E, Barratt CLR. The value of diagnostic and prognostic value of traditional semen parameters. *J Androl.* 1999;20:588–593.
- Wichmann L, Isola J, Tuohimaa P. Prognostic variables in predicting pregnancy. A prospective follow-up study of 907 couples with an infertility problem. *Hum Reprod.* 1994;9:1102–1108.
- World Health Organization. *WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interactions.* 2nd ed. New York, NY: Cambridge University Press; 1987.
- World Health Organization. *WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interactions.* 3rd ed. New York, NY: Cambridge University Press; 1992.
- World Health Organization. *WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interactions.* 4th ed. New York, NY: Cambridge University Press; 1999.