J Clin Pathol 1986;39:1347–1352

Morphometry of spermatozoa using semiautomatic image analysis

J R JAGOE,* N P WASHBROOK,† ELIZABETH A HUDSON†

*From the *Division of Medical Statistics and the †Department of Histopathology (Cytopathology), Northwick Park Hospital and Clinical Research Centre, Harrow, Middlesex*

SUMMARY Human sperm heads were detected and tracked using semiautomatic image analysis. Measurements of size and shape on two specimens from each of 26 men showed that the major component of variability both within and between subjects was the number of small elongated sperm heads. Variability of the computed features between subjects was greater than that between samples from the same subject.

The male component of infertile partnerships is often unexplained. After the history and physical examination semen is routinely examined for motility, density, and morphology. Modern techniques of time exposure photomicrography[1] and the sperm penetration assay[2] provide additional quantitative information of sperm function but these tests are not widely available at present. The value of sperm morphology has been acknowledged for more than 50 years[3-5] and was shown by Aitken[6] to be the most statistically predictive test of the routine semen analyses in male infertility. There are also data to suggest that change in sperm morphology may be an early sign of damage to testicular function by toxic environmental substances or drugs.[7 8]

From six to 60 morphological forms have been described,[3 4 9] but because of observer variation in the differential counts made with the light microscope[10 11] routine investigations are limited to the percentage of all abnormal forms. Abnormalities of the sperm head are the most common and of relatively greater prognostic importance than abnormalities of the body and tail.[3 9] The relative effectiveness of each head type cannot be measured objectively in terms of fertility, but computed image processing should provide a more consistent and objective analysis of sperm morphology than the subjective assessment in current use.

At present there are insufficient data on the distribution of shape and size of sperm heads in normal semen with which to compare infertile samples. We have used a semiautomatic method to detect and track sperm head boundaries to measure the features of head size and shape. The ranges and interrelations of these computed measurements were used to describe intrasubject and intersubject variations.

Accepted for publication 18 June 1986

Material and methods

SEMEN SPECIMENS

Slides prepared for morphological assessment of spermatozoa as part of routine semen analysis were taken from the laboratory files for this study. The samples had been obtained by masturbation after at least 48 hours' abstinence from intercourse. Motility and density estimations were made within two hours of ejaculation. The slide preparations were dried in air before fixation in methanol and staining by the Papanicolaou method. All the subjects had been investigated because they had infertile marriages.

Measurements were made on two specimens obtained on separate occasions from each of 26 subjects. Twenty subjects had at least two specimens with normal sperm count ($>40 \times 10^6/cm^3$) and morphology ($<30\%$ abnormal forms) and no record of abnormal semen indices. These results indicate potential fertility within the limitations of routine laboratory analysis. Two subjects had normal counts but abnormal morphology, two subjects had subnormal counts and abnormal morphology, and two subjects had subnormal counts and normal morphology for both specimens.

COMPUTER METHODS

The slides were placed under a Leitz Ortholux II microscope on which was mounted a standard 625 line monochrome TV camera connected to a Magiscan (Joyce-Loebl) image analysis system. The central area of the field of view was scanned and digitised to a 512×512 square matrix of six bit pixels (picture

1 Size = A max + A min

2 P2A = perimeter^2/4 × π × size

3 Min D

4 Max D

5 Max D: Min D

6 Askew = A max: A min

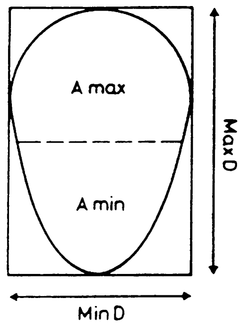

Fig 1 *Six features computed for each head.*

elements represented by discrete grey level values between 0 and 63 inclusive). Each sample was scanned using a ×100 oil immersion objective with a 1·5 magnifying stage, giving a resolution of 0·092 μm per pixel. A typical sperm head occupied an area of about 1000 pixels.

An area of the slide which showed well spread spermatozoa was selected for commencement of the measurements, and adjacent fields were scanned so that spermatozoa were not measured more than once. All heads, apart from those that overlapped were measured in each field. The sampling on each specimen ended when 400 heads had been located, with the exception of two low count specimens where only 200 heads were measured.

Most of the heads were detected automatically. Unsuccessful detections were traced manually by the operator using a light pen system. Partially successful detections were patched using an edit facility. The boundary information on each head was stored on disc file for subsequent size and shape analysis.

Six size and shape dependent features were computed from these boundary coordinates (fig 1). They were: Size (the area contained by the boundary); P2A (the perimeter squared divided by 4 × π × size); MinD (the minimum head diameter); MaxD (the maximum head diameter); MaxD:MinD (the ratio of the two diameters); Askew (the ratio of the larger to the smaller head areas obtained by bisecting the head at right angles to the maximum diameter).

STATISTICAL METHODS

The values of the computed size and shape features for each head were transmitted via a data link to a DEC-20 computer for analysis using the BMDP[12] statistical package.

The distributions of the measured features within any sample were generally non-normal in a statistical sense. Techniques exist to transform skewed data to normality, but often a transform that normalises a distribution from one sample will not do so for another—that is, no single transform is likely to produce normality for the values of a given feature for all samples. Thus the information cannot be completely summarised by the two measures of mean and standard deviation. As an alternative the information was summarised approximately using the nine decile cut point values (the values of the 10th to 90th centiles).

Results

SEMIAUTOMATIC DETECTION AND TRACKING

The program automatically located and tracked 60·4% of all 20 400 sperm heads. The success rate was variable between samples. Thirty per cent of heads were partially tracked and manually edited, while only 9·2% needed to be traced entirely manually. In 46 of the 52 samples the mean Size of the manually traced heads was larger than in those automatically tracked. The average value of this difference was 7·5%. The difference between the number of manually traced heads below and above the median size of all heads, however, was very small (955 and 925, respectively), indicating that the size differences within samples reflected operator characteristics, rather than the failure of the autotracking procedure more often on larger heads. The Wilcoxon two sampled signed ranks test showed no significant difference between manual and autotracked head shape as measured by MaxD:MinD.

RANGES OF THE MEASURED FEATURES

Fig 2 shows the variability of the sample distributions. This shows the 52 samples in order of median values for all six features. The upper and lower indicators on each vertical line represent the 10th and 90th centile values in each sample. The largest median sperm head size was 9·78 μm, which was 73% larger than the smallest median head size (5·65 μm). The largest median diameter ratio was 2·2, which was 50% larger than the smallest median ratio (1·48). Specimens from the six subjects identified as having abnormal indices were scattered throughout the ranges.

RELATIONS BETWEEN THE COMPUTED FEATURES

Within all samples

For each sample the 15 interfeature rank correlation coefficients (Kendall's τ) were computed. Table 1 shows the mean value of these coefficients and their standard deviations using all 52 samples. The correlation between Size and shape, as measured by MaxD:MinD, was not significant, but this only implies lack of a monotonic relation between Size and shape. A table of the number of heads in each of the Size tenths against each of the MaxD:MinD tenths (using cut points specific to each sample) should have roughly equal counts (204) in each category if there is

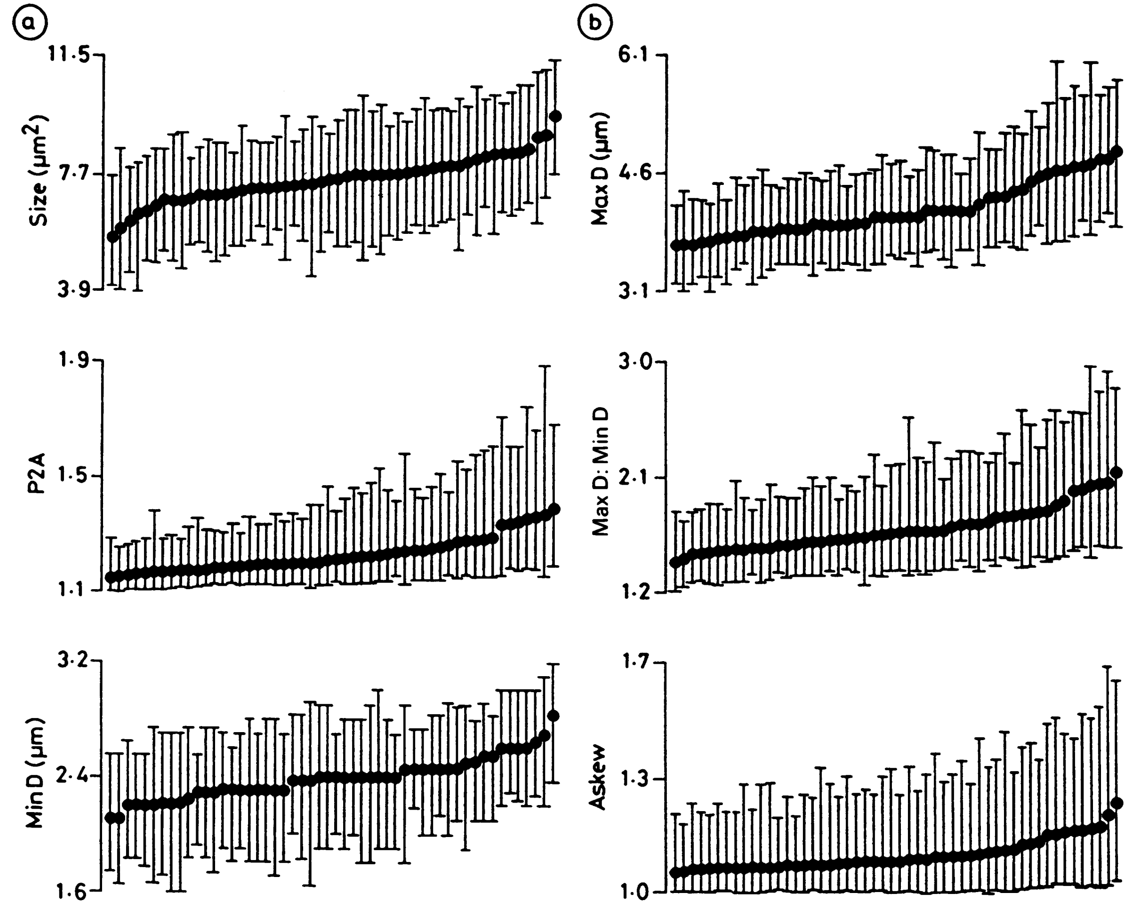

Fig 2 *Tenth, 50th, and 90th centile values for 52 samples arranged in ascending order of median values.*

no relation of any kind. Table 2 shows the difference between observed category sizes and their expected sizes. Grouped in this way there is a negative rank correlation of −0·36 between size and shape. The most noticeable feature is the large excess of small elongated heads (top right hand corner). There was also an excess of small round heads in the lowest 10th intervals of Size and MaxD:MinD. The larger heads tended to be rounder. In this context the descriptions small, large, round and elongated are applied relative to a sample. Intermediate head sizes are most likely to be oval. This is regarded as the normal shape in subjective assessment.

Between samples from different subjects

Using a single sample from each subject the median values of each feature in a sample were used to compute the 15 rank correlation coefficients between the

Table 1 *Mean (SD) values of between feature rank correlation coefficients within samples*

	Size	*P2A*	*MinD*	*MaxD*	*MaxD:MinD*
P2A	−0·10 (0·14)				
MinD	0·63 (0·07)	0·44 (0·08)			
MaxD	0·50 (0·06)	0·41 (0·12)	0·14 (0·08)		
MaxD:MinD	−0·12 (0·12)	0·79 (0·03)	−0·51 (0·07)	0·39 (0·10)	
Askew	0·02 (0·15)	0·26 (0·07)	−0·04 (0·11)	0·16 (0·15)	0·017 (0·06)

Table 2 *Observed minus expected numbers of heads in each of Size tenth and MaxD:MinD tenth intervals*

	Size tenths									
MaxD:MinD tenths	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1	61	−79	−96	−85	−77	−65	−36	3	78	309
2	−14	−46	−63	−50	−19	3	14	12	52	119
3	−35	−34	−23	2	−11	−10	37	34	38	7
4	−41	−16	21	−12	− 2	−11	3	39	27	− 5
5	−17	− 8	2	− 2	5	− 8	20	23	12	−37
6	−17	32	10	42	9	10	10	−14	6	−77
7	− 8	24	36	17	33	39	− 4	−21	−35	−75
8	−10	21	33	24	21	31	9	−11	−62	−71
9	23	50	40	59	21	4	−16	−38	−68	−86
10	53	73	25	15	12	5	−21	−36	−51	−85

Table 3 *Between sample rank correlation coefficients of measured features*

	Size	*P2A*	*MinD*	*MaxD*	*MaxD:MinD*
P2A	0·13				
MinD	0·54	−0·36			
MaxD	0·56	0·59	0·07		
MaxD:MinD	0·10	0·92	−0·38	0·56	
Askew	0·25	0·74	−0·20	0·60	0·71

features (table 3). There was no significant correlation between median Size and median MaxD:MinD (n = 26).

VARIATION WITHIN AND BETWEEN SUBJECTS

Within subjects

The nine decile cut points for the features Size and MaxD:MinD were estimated for each subject by averaging the two sets of values for each sample. Ten × 10 tables of the number of heads in each of the Size and MaxD:MinD 10ths were computed for each sample using the estimated cut points for the subject. Subtracting the two tables for each subject gave a difference table. The absolute values of the differences were used and the variance of each entry in the 26 difference tables was computed. For a reference comparison the same process was applied five times to pseudo samples generated by a random selection of two equal sized groups from the combined samples of each subject. The mean values of the variances of the difference tables was computed. The values of the variances of the difference table entries derived from between subject samples was divided by their corresponding entries in the variance table derived from the pseudo samples. The largest variance ratio was 12·0 (DF = 25·25; $p < 0·001$), which corresponded to values derived from heads in the smallest 10% Size group and the largest 90% MaxD:MinD group. The next largest variance ratio was 8·3 ($p < 0·001$) for heads in the smallest 10% Size group and the 80–90% MaxD:MinD group. Thus the major component of between samples variation within subjects was found in the numbers of small elongated sperm heads.

BETWEEN SUBJECTS

An analysis of variance on the median head Size values showed that the mean square variation between subjects was four times greater than that within sub-

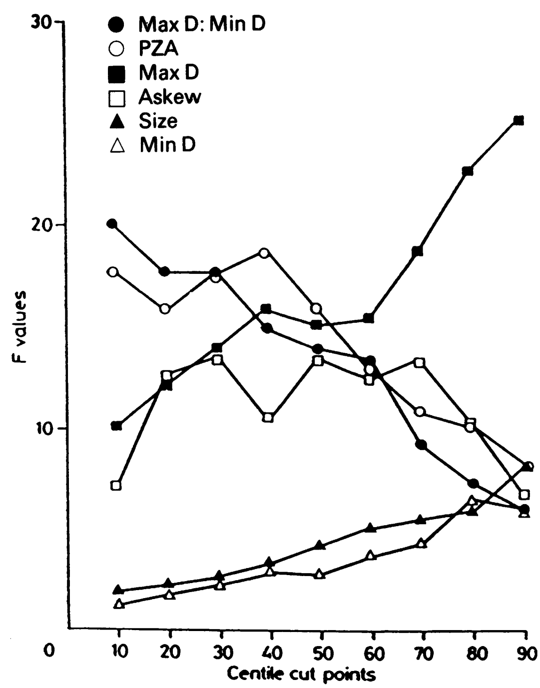

Fig 3 *Between subject discriminating power of each of measured features at each decile cut point.*

jects. A similar analysis for the shape feature MaxD:MinD showed that the mean square variation between subjects was 14 times greater than within subjects—that is, head shape was more characteristic of a subject than size.

Fig 3 shows how well each of the measured features at each decile cut point performs as a discriminator between subjects. The vertical axis (F value) indicates the between to within sample variance ratio. These results fit well with the finding that the greatest source of variation between samples from the same subject is the number of small elongated heads. Increasing cut points of MaxD:MinD are more likely to be influenced by this variation and therefore less likely to be good discriminators between subjects. P2A, which is strongly related to MaxD:MinD, also follows this trend. Increasing cut points for Size, MinD, and MaxD are less influenced by this variation and therefore likely to be increasingly better at between subject discrimination. The feature Askew does not show any consistent trend except that the lower and upper cut points are less useful between subject separators than the intermediate cut points.

A more complete analysis using a stepwise linear discriminant program (BMDP7M[12]) showed that a function based on the two 90% cut points of MaxD and MinD ($2 \times MaxD_{90} - 3 \times MinD_{90}$) gave the best between subject separation. Using this function the mean square variation between subjects was 67 times the within subject variation.

Discussion

TECHNIQUES

The limitations of the traditional method of assessing sperm morphology have been shown by Jequier and Ukombe[11] who asked 26 professional observers to assess the morphology of one semen specimen. They rated the abnormal forms between 5% and 85%, and three technicians who were highly experienced in this field made assessments between 25% and 40%.

Measurement has been applied to the morphology of spermatozoa by several techniques, but so far these have not shown sufficient speed and accuracy for wide application. Sperm size and density measurements have been made rapidly using a Coulter counter but the accuracy is poor when sperm counts are low,[13] and these may be the samples which require most exact evaluation. Measurements of sperm have also been made using a digitising tablet[14] and from scanning electron microscope photographs.[15-17]

Our semiautomatic method is relatively rapid and objective. The performance can be improved by alternative methods of preparation and staining. Subsequent experience using gallocyanin,[18] which gives better contrast between the sperm heads and the background, increased the success rate of the automatic method. A technique[19] of making smears, which avoids clumping and overlapping of spermatozoa, should also improve the detection rate and reduce manual editing with the light pen. The larger average size noted for the manually traced heads corresponded to a radial difference of about 1·5 pixels. This can probably be accounted for by the tendency of an operator to trace around the outside of an edge whereas the automatic tracking program detects and follows points of maximum gradient that generally lie inside the outer edge. The difference, however, is small compared with the differences both within and between samples. It is worth noting that the features which turned out to be most informative could be estimated by manually digitising the four points delineating the maximum and minimum diameters.

RESULTS

It would be expected for objects like sperm heads that the three shape measurements P2A, MaxD:MinD, and Askew might be highly related. In the case of P2A and MaxD:MinD this is true, but within samples Askew is comparativly unrelated to either. Askew does not seem to be an important measure in its ability to characterise subjects.

The most notable finding is that on average more than twice as many of the heads in the smallest tenth category of a sample are likely to fall into the largest MaxD:MinD tenth of that sample than if they were uniformly distributed within the shape range. The variation between samples from the same subject is also mainly due to this small elongated type of head. The difference between the numbers of this type of cell in samples from the same subject over an interval of time could possibly be related to the interval since last emission; this was not known in this study. Freund,[20] however, using subjective assessment, found that frequency of emission had no effect on morphology, although numbers of sperm decreased. If morphological abnormality is related to the number of these forms it could mean that male fertility is not constant over time.

The sperm with the small elongated head is described as the commonest abnormal morphological form where subjective differential counts have been done on semen from subfertile males.[3 4] It is possible that this form, which we have identified in this study by size and the MaxD:MinD ratio, is mainly responsible for the relation between subjective assessment of abnormal morphology and fertility.

Differences within and between subjects are most pronounced at the extreme ends of the size and shape ranges. If these features are also the most sensitive to adverse effects due—for example, to treatment with drugs[21]—the implication is that a large number of heads must be measured to obtain satisfactory estimates of these values.

We have shown that computed image processing can be applied usefully to measurements of sperm morphology using slide preparations and light microscopy. It has a potential for both study of male infertility and for routine laboratory semen analysis.

We thank Mr D Altman and Dr ID Hill for their advice on the statistical analysis, Mr L Pratsis for technical assistance and Mr AM Fisher for permission to study material from his patients. NPW was financially supported by the North West Thames Regional Health Authority.

References

1 Overstreet JW, Katz DF, Hanson FW, Fonseca JR. A simple inexpensive method for objective assessment of human sperm movement characteristics. *Fertil Steril* 1979;**31**:162–72.

2 Yanagimachi R, Yanagimachi H, Rogers BJ. The use of zona-free animal ova as a test system for the assessment of fertilizing capacity of human spermatozoa. *Biol Reprod* 1976;**15**:471–6.

3 Moench GL, Holt H. Sperm morphology in relation to fertility. *Am J Obstet Gynecol* 1931;**22**:199–210.

4 Macleod J. The semen examination. *Clin Obstet Gynaecol* 1965;**8**:115–27.

5 Sherins RJ, Brightwell D, Sternthal PM. Longitudinal analysis of semen of fertile and infertile men. In: Troen P, Nankin HR, eds. *The testis in normal and infertile men.* New York: Raven Press, 1977:473–87.

6 Aitken RJ, Best FSM, Richardson DW, *et al.* An analysis of sperm function in cases of unexplained infertility: conventional criteria, movement characteristics, and fertilising capacity. *Fertil Steril* 1982;**38**:212–21.

7 Young IT, Gledhill BL, Lake S, Wyrobek AJ. Quantitative analysis of radiation-induced changes in sperm morphology. *Anal Quant Cytol* 1978;**4**:207–16.

8 Wyrobek AJ, Gordon LA, Burkhart JG, *et al.* An evaluation of human sperm as indicators of chemically induced alterations of spermatogenic function. *Mutat Res* 1983;**115**:73–148.

9 Belsey MA, Eliasson R, Gallegos AJ, Moghissi KS, Paulsen CA, Prasad MRN. *Laboratory manual for the examination of human semen and semen-cervical mucus interaction.* Singapore: Press Concern, 1980.

10 Freund M. Standards for the rating of human sperm morphology. *Int J Fertil* 1966;**11**:97–180.

11 Jequier AM, Ukombe EB. Errors inherent in the performance of a routine semen analysis. *Br J Urol* 1983;**55**:434–6.

12 Brown MB, Engelman L, Frane JW, Hill MA, Jennrich RI, Toporek JD. In: Dixon WJ, ed. *BMDP Statistical software 1981.* Berkeley: University of California Press, 1981.

13 Brotherton J, Barnard G. Estimation of number, mean size and size distributions of human spermatozoa in oligospermia using a Coulter counter. *J Reprod Fertil* 1974;**40**:341–57.

14 Schmassmann A, Mikuz G, Bartsch G, Rohr H. Spermiometrics: objective and reproducible methods for evaluating sperm morphology. *Eur Urol* 1982;**8**:274–9.

15 Dadoune JP, Fain-Maurel MA, Guillaumin M, Guillaumin D. Scanning electron microscopic morphometry of a discriminated population of elongated human spermatozoa. *Int J Fertil* 1980;**25**:18–27.

16 Liakatas J, Williams AE, Hargreave TB. Scoring sperm morphology using the scanning electron microscope. *Fertil Steril* 1982;**38**:227–32.

17 Hoidas S, Williams AE, Tocher JL, Hargreave TB. Scoring sperm morphology from fertile and infertile cigarette smokers using the scanning electron microscope and image analysis. *Fertil Steril* 1985;**43**:595–8.

18 Husain OAN, Watts KC. Rapid demonstration of nucleic acids using "oxidised" gallocyanin and chromic potassium sulphate: methods and applications. *J Clin Pathol* 1984;**37**:99–101.

19 Watts KC, Husain OAN, Tucker JH, *et al.* The use of cationic polyelectrolytes in the preparation of cell monolayers for automated cell scanning and diagnostic cytopathology. *Anal Quant Cytol* 1984;**6**:272–8.

20 Freund M. Interrelationships among characteristics of human semen and factors affecting semen specimen quality. *J Reprod Fertil* 1962;**4**:143–59.

21 Hudson E, Dore C, Sowter C, Toovey S, Levi AJ. Sperm size in patients with inflammatory bowel disease on sulfasalazine therapy. *Fertil Steril* 1982;**38**:77–84.

Requests for reprints to: Dr EA Hudson, Department of Histopathology, Northwick Park Hospital and Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, England.