Human Reproduction vol.14 no.4 pp.1005–1012, 1999

Establishment of predictive variables associated with testicular sperm retrieval in men with non-obstructive azoospermia

U.I.O.Ezeh[1], N.A.Taub[2], H.D.M.Moore[1,3,4] and I.D.Cooke[1]

[1]University Departments of Obstetrics and Gynaecology, Jessop Hospital for Women, Leavygreave Road, Sheffield S3 7RE, [2]Department of Epidemiology and Public Health, University of Leicester, Leicester LE1 6TP and [3]Department of Molecular Biology and Biotechnology, University of Sheffield S10 2TN, UK

[4]To whom correspondence should be addressed at: University Department of Obstetrics and Gynaecology, Jessop Hospital for Women, Leavygreave Road, Sheffield S3 7EN, UK

Although testicular biopsy for sperm extraction is a procedure with a potential for complications, sperm retrieval is successful in 30–70% of patients with non-obstructive azoospermia. In order to predict the probability of retrieving at least one testicular spermatozoon we conducted a prospective study of a set of variables in 40 patients with non-obstructive azoospermia. Using the receiver operating characteristic curves, we determined the probability estimates of testicular volume, plasma follicle stimulating hormone (FSH) concentration, Johnsen score and visualization of testicular spermatids in discriminating between patients with successful and failed testicular sperm extraction. Visualization of testicular spermatids provided the best estimate of success of testicular sperm extraction. Of the factors studied using logistic–regression analysis (age, maternal and paternal age at birth, body mass index, luteinizing hormone, testosterone, FSH, testicular volume, the presence of testicular spermatids and Johnsen score), only the presence of spermatids and Johnsen score were independent variables able to predict the success of testicular sperm extraction. The visualization of the presence of spermatids gave a correct prediction of 77% and Johnsen score of 71%. The diagnostic model derived from these independent predictors when validated in 40 patients using the Jackknife technique gave a correct overall prediction of 87%. The probability of successful testicular sperm extraction in patients with non-obstructive azoospermia could be objectively predicted on the basis of simple histopathological criteria represented by the visualization of testicular spermatids and Johnsen score.

Key words: azoospermia/diagnostic index/germ cells/prediction/sperm retrieval

Introduction

Testicular biopsy is a procedure with the potential for complications (Harrington *et al*., 1996; Schlegel and Su, 1997) while testicular sperm retrieval is successful in only 30–70% of azoospermic patients with defective spermatogenesis (Jow *et al*., 1993; Tournaye *et al*., 1997). Sperm extraction involves multiple testicular biopsies often scheduled with oocyte retrieval after ovarian stimulation and monitoring. A failed testicular sperm extraction therefore has significant emotional and financial implications for the couple (Tournaye *et al*., 1997). On the other hand, retrieval of a few living spermatozoa may suffice for successful intracytoplasmic sperm injection (ICSI) of oocytes. With the success of ICSI using testicular spermatozoa and the increasing use of testicular biopsy, accurate prediction of testicular sperm retrieval has become essential.

Whether or not to undertake a therapeutic sperm extraction procedure without a preliminary diagnostic testicular biopsy is debatable. Patients with non-obstructive azoospermia may undergo a prior diagnostic testicular biopsy (Silber *et al*., 1997; Tournaye *et al*., 1997) or may have multiple therapeutic testicular biopsies scheduled with oocyte retrieval to look for spermatozoa for ICSI with no preliminary biopsy (Devroey *et al*., 1995; Tournaye *et al*., 1996). In cases where no spermatozoa are retrieved the couples are prepared for donor insemination. Patients with non-obstructive azoospermia may have cryptozoospermia. Hence, a careful search of the ejaculate on the day of ovum retrieval may yield some spermatozoa without resorting to testicular extraction. Nagy *et al*. (1995) reported an ongoing pregnancy rate of 24.5% per cycle in a study of 57 cycles of the wives of men with cryptozoospermia, a value similar to that reported for ICSI with testicular and epididymal spermatozoa or ejaculated sperm from men with normozoospermia. Cryptozoospermia may occur in 10–35% (Lindsay *et al*., 1995; Ron-El *et al*., 1997) of patients with azoospermia, the exact proportion depending on the centrifugation and duration force used in the preparation of the semen sample. Moreover, cryptozoospermia may be intermittent, so that ejaculated spermatozoa may not be found on the day of oocyte retrieval. Conventional parameters associated with spermatogenesis, including plasma follicle stimulating hormone (FSH), testicular volume and testicular histology individually have failed to provide a good discrimination between patients who will have successful and failed testicular sperm extraction (Tournaye *et al*., 1996, 1997; Mulhall *et al*., 1997). It is possible that a combination of these variables may provide a better discrimination. Algorithms and predictive models have not been developed for the treatment with testicular biopsy and ICSI of men with azoospermia due to defective spermatogenesis, probably because this type of treatment is relatively new. Such models can be used to establish diagnosis, aid counselling of patients and reduce the money spent on unnecessary investigations.

The objective of this pilot study was to explore the diagnostic value of current parameters used in the management of azoospermia, including testicular volume, plasma FSH concentration, Johnsen score and the visualization of testicular spermatids. Logistic regression analysis was used to develop a diagnostic model, based on a set of variables associated with spermatogenesis, that could be applied to individual patients, thereby limiting testicular biopsy to those patients with the best chance of yielding testicular spermatozoa.

Materials and methods

Study population

The population comprised 40 consecutive patients with non-obstructive azoospermia who underwent a diagnostic testicular biopsy for genetic studies and evaluation of spermatogenesis at the Jessop Hospital for Women in Sheffield, UK. Testicular biopsies and sperm extraction were not synchronized with oocyte retrieval and ICSI cycles which were performed 3–6 months later. Spermatozoa obtained were cryopreserved. The study was approved by the South Sheffield Hospitals Ethics Committee, UK. The developmental, social, family, medical and reproductive histories as well as a history of urological operations and exposure to gonadotoxins were documented in every patient. Each patient underwent a physical examination which included an evaluation of secondary sexual characteristics, examination of the penis, vasa deferentia, epididymides and a rectal examination to exclude prostatic pathology. The testicular volume was estimated with a Prader orchidometer. Data on height and weight were used to calculate the body mass index (BMI; kg/m^2). Blood samples were obtained for the determination of their karyotype using standard techniques (Verma *et al*., 1995), plasma concentrations of FSH and luteinizing hormone (LH; normal value: 1–12 IU/l) were measured by immunoassay and plasma testosterone (normal value: 9.4–37.0 nmol/l) by radioimmunoassay, according to the manufacturer's instructions (Diagnostic Products Ltd, Wales, UK). The intra-assay and inter-assay coefficients of variation did not exceed 6.5%. In cases where where congenital cystic fibrosis was suspected because of low sperm volume, acidic pH or absence of vasa deferentia, a cascade screening of common CF gene mutations was undertaken using standard techniques (Chillion *et al*., 1994) The objective of this was to exclude patients with obstructive azoospermia from the test group. Semen samples were produced by masturbation after 3–5 days of sexual abstinence, collected into sterile containers, allowed to liquefy at room temperature for 30 min and analysed according to World Health Organization criteria (WHO, 1992). The diagnosis of azoospermia was made from at least two semen analyses when no spermatozoa were found in the pellet obtained following semen centrifugation at 1500 *g* for 10 min. Patients for the study were selected according to eligibility criteria shown in Table I.

Table I. Eligibility of patients

Not eligible
Obstructive azoospermia
vasectomy
vasectomy reversal
sexually transmitted disease
swollen testis
distended epididymis
absence of vasa deferentia/epididymides
semen pH <7
presence of sperm agglutinins
Retrograde ejaculation
low semen volume (<1 ml)
diabetes mellitus
retroperitoneal lymphadenectomy
bladder neck surgery
spinal injury
Hypogonadotrophism
Kallmann's syndrome
endocrine disorder
low FSH/LH
Eligible
Testicular failure
idiopathic
chemotherapy
radiotherapy
malignant disease
cryptorchidism
orchidopexy
torsion of the testis
mumps orchitis
abnormal karyotype
testicular atrophy (⩽12 ml)
high FSH/LH

FSH = follicle stimulating hormone; LH = luteinizing hormone.

Testicular biopsy

Bilateral outpatient testicular biopsy was performed under general anaesthesia or with intravenous midazolam, spermatic cord block with 0.5% bupivacaine solution and local infiltration of the scrotum with 2% lignocaine in a 1 in 200 000 solution of adrenaline. To avoid multiple biopsies and potential testis damage (Schegel and Su, 1997) and to ensure that sufficient testicular tissue was obtained, a single large biopsy (1.5–2 cm) was taken from each testis. The testicular spermatozoa obtained were cryopreserved as back-up for a subsequent testicular extraction procedure 3–6 months performed later which was synchronized with ICSI–in-vitro fertilization (IVF). Spermatid extraction was not performed since this is not permitted in the UK. Where adequate viable thawed sperm were obtained at the time of the planned ICSI–IVF procedure, the second testicular extraction was not performed. Those patients in whom no spermatozoa were found during the initial exploratory extraction did not undergo a second procedure but were counselled for donor insemination. This arrangement maximized the patient's chance of using his own spermatozoa for ICSI and ensured that patients were better prepared for donor insemination treatment in case of unsuccessful sperm extraction. The initial diagnostic biopsy was also used to exclude carcinoma *in situ* of the testis and for evaluation of the status of spermatogenesis. A tissue wedge (~25 mg) from each testis was placed in phosphate-buffered saline (PBS) and immediately transported to the laboratory. For histopathological examination a biopsy was fixed in Bouin's solution. This study reports only the data obtained from the initial exploratory diagnostic testicular biopsy and sperm extraction procedures.

Testicular sperm extraction

A piece of biopsy tissue (~5 mg) was minced into fine pieces (<1 mm^3) using a pair of sterile glass microscope slides for ~25 min. The resultant suspension was transferred in PBS to a conical centrifuge tube (Falcon type, 100 mm; Fahrenheight Laboratory Supplies, Rotherham, UK), vortexed for 5 min and centrifuged for 5 min at 500 *g*. The pellet was resuspended in 2 ml of erythrocyte-lysing buffer (155 mM NH_4Cl, 10 mM $KHCO_3$, 2 mM EDTA, pH 7.2) and allowed to stand for 5 min. After centrifugation at 500 *g* for 5 min, the pellet was resuspended in Earle's balanced salt solution (EBSS) supplemented with 2.26% human serum albumin (Gibco BRL, Life Technologies, Paisley, UK) and transferred to a Petri dish containing 3 ml of EBSS. The medium was overlayered with 1 ml paraffin

oil. The preparation was immediately examined under an inverted microscope equipped with Hoffman modulation contrast photomicroscopy at magnifications ×200 and ×400 for the presence of testicular sperm and then incubated at 37°C in an atmosphere of 5% CO_2 in air for up to 72 h. This testicular culture was examined daily for the presence of testicular spermatozoa.

Histopathology

Semi-thin paraffin wax testicular tissue sections (4 μm thick) were stained with haematoxylin and eosin and then examined under a light microscope at ×100 and ×400 magnification using standard techniques. Testicular histology was classified into hypospermatogenesis (reduction in the number of normal spermatogenetic cells), maturation arrest (an absence of the later stages of spermatogenesis), Sertoli cell-only (the absence of germ cells in the seminiferous tubules), and tubular sclerosis (no germ cells or Sertoli cells present in the tubules). The presence of an occasional seminiferous tubule with a few spermatids in a field of seminiferous tubules that otherwise exhibited maturation arrest, Sertoli cell-only pattern or tubular sclerosis was classified as focal spermatogenesis. One hundred tubules were examined per slide and each slide was scored using the Johnsen score (Johnsen, 1970), whereby seminiferous tubules are assessed on a scale of 1–10, with tubules having a complete lack of cells scored as 1 and those with a maximum sperm presence (at least five or more spermatozoa in the lumen) scored as 10. In order to obtain a mean score for each patient, the number of tubules recorded at each Johnsen score point was multiplied by the corresponding Johnsen score, and the sum of all multiplications was divided by the total number of tubules recorded. The average of the scores for both testes was obtained to get the overall score for each patient. The slides were examined by three different assessors who were unaware of the results of the testicular sperm retrieval. Discussion among the assessors resolved any discrepancies and enabled a final diagnosis to be reached, so that the coefficient of variation was not calculated. Quantitative analysis of spermatid count was not performed (Silber *et al*., 1997). Instead the finding of at least one round, elongating or elongated spermatid on histological sections was defined as 'positive spermatid on histology'.

Design

The endocrine profile (plasma FSH, LH and testosterone concentrations) and biophysical profile (patients' age, paternal and maternal ages at birth, occupation, risk factors for azoospermia, height, weight, BMI, and the sum of volume of both testes), whether or not testicular spermatids were visualized, and Johnsen score were used as predictors of the likelihood of retrieval of at least one testicular spermatozoon. Consecutive patients were studied.

Blinding of variables and end-points

The persons performing the semen analysis, testicular sperm extraction and clinical evaluation were blinded from each other. Each of these procedures was performed by the same person on each occasion.

Statistical analysis

The end-point was retrieval of at least one testicular spermatozoon. The sensitivity (the proportion of patients with successful sperm extraction identified by the predicting test) and specificity (the proportion of patients without testicular spermatozoa correctly identified by the predicting test) of testicular volume, plasma FSH concentration, Johnsen score and presence of testicular spermatids in predicting the probability of testicular sperm extraction were calculated. The accuracy of prediction of the likelihood of successful testicular sperm extraction and the best cut-off level for the variables in predicting the likelihood of testicular sperm retrieval were assessed by the area under the receiver operating characteristic (ROC) plot (Zweig and Campbell, 1993), a plot of true positive (sensitivity) against false positive (1 – specificity) rates for each possible cut-off point. Because the sensitivities and specificities were calculated separately, but using test results from two different subgroups (i.e. the sensitivity was calculated from the subgroup with testicular sperm present and specificity from the subgroup without testicular sperm), the ROC plot was independent of the prevalence of spermatozoa in the testis. The area under the ROC plot was determined statistically using the Wilcoxon and Mann–Whitney *U*-statistics (Hanley and McNeil, 1982) (Appendix 1). The best cut-off value was considered to be that which maximized the sum of sensitivity and specificity, roughly indicated on the graph by the shortest distance between the ROC curve and the top left-hand corner (where sensitivity = 1 and specificity = 1).

Stepwise logistic regression analysis was performed with forward selection to identify which combination of the potentially predictive variables was most predictive of testicular sperm retrieval. This technique was used to develop a mathematical equation to calculate the probability of sperm extraction in any patient on the basis of his own results.

The performance of the diagnostic model was then tested in a study sample using the Jackknife technique (Efron and Gong, 1983), in order that the prediction rule used for a given patient was derived independently of that patient's own data. Jackknife technique is an estimation of the error rate of a prediction model which gives a result similar to that obtained by testing the model in a different independent population. The method consists of (i) first deleting the patients x_1 from the data sample, (ii) using logistic regression analysis to recalculate the prediction rule on the basis of the remaining $n - x_1$ patients, (iii) using the new prediction rule to predict the outcome of x_1 and (iv) returning the patient x_1 to the group, and repeating the same procedure separately for patients x_2 to x_n (x_{40}). Analysis was performed using the 50% cut-off point to ensure that the likelihood of predicting successful sperm extraction was similar to the observed success. The misclassification rate, sensitivity, specificity as well as the positive and negative predictive values of the diagnostic index were calculated.

Results

Each patient's age, the ages of his parents at the time of his birth, his BMI, plasma LH, FSH and testosterone concentrations, testicular volume, the outcome of the assessment of testicular spermatids, Johnsen score and the results of testicular sperm extraction were available for all patients in the study. Patients with gonadal failure were selected on clinical criteria (Table I). All the patients showed histological evidence of testicular failure. Hypospermatogenesis was found in 10 (25%) patients, maturation arrest in seven (18%), Sertoli cell-only pattern in 12 (30%), and focal spermatogenesis occurred in 11 (28%) patients. A block in spermatogenesis at the level of spermiogenesis was not encountered in any of the patients. In those with hypospermatogenesis, round spermatids were always present in association with the long spermatids. The median values for age, combined testicular volume and plasma FSH concentration of the study population were 34 years, 32 ml, and 18 IU/l, respectively. Approximately 79% of the patients identified themselves as Caucasian, 8% as Asians, 8% as Arabs, and 5% as Afro-Caribbean. The mean weight (± SD) of tissue taken for histopathology was 585 (± 20) mg and for

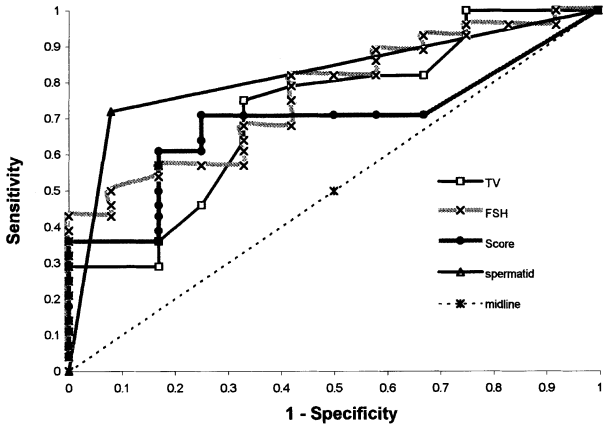

Figure 1. Predictive value of testicular volume (TV), plasma follicle stimulating hormone (FSH) concentration, Johnsen score and testicular spermatids using the receiver operating characteristic (ROC) curves. The best discriminating value is 1 and the worst discriminating value is 0.5. A test with perfect discrimination has an ROC plot that passes through the upper left corner, where the sensitivity (true positive) is 1 (perfect sensitivity) and the 1 – specificity (false positive) is 0 (perfect specificity). The theoretical plot for a test with no discrimination is a diagonal line from the lower left corner to the upper right corner. The more the ROC curve moves towards the top and left corner of the boundaries of the ROC graphs, the more accurate the individual curve. Qualitatively, the closer the plot is to the upper left corner the higher the overall accuracy of the test variable. The best cut-off value is roughly indicated on the graph by the shortest distance between the ROC curve and the top left-hand corner (where sensitivity = 1 and specificity = 1) (Zweig *et al*., 1983). Quantitatively, the area under the curve is used to determine the accuracy of prediction. Area under the ROC curve for presence of testicular spermatids = 0.77, plasma FSH concentration = 0.76, Johnsen score index = 0.72 and testicular volume for both testes = 0.72. 'Testicular spermatids on histology' was the most predictive.

sperm extraction 479 (± 12.1) mg ($P > 0.05$). Testicular sperm extraction was successful in 28 out of the 40 patients studied, giving a sperm retrieval rate of 70%.

Sensitivity, specificity and accuracy

The sensitivity and specificity were obtained at various cut-off levels for each modality (plasma FSH, combined testicular volume, presence of testicular spermatids and Johnsen score) and the corresponding ROC plots are shown in Figure 1. The outcome of testicular extraction in relation to the presence of testicular spermatids is shown in Table II. The sensitivity was 71%, specificity 92%, positive predictive value 95% and negative predictive value 58% for predicting the likelihood of successful testicular sperm extraction using the visualization of testicular spermatids on testicular histology as the predicting variable.

Table III describes the sensitivity, specificity and accuracy of prediction of these variables at their best cut-off levels determined from the ROC in Figure 1. Using the best cut-off point of 27 ml, 75% of patients with testicular spermatozoa (true positives) were predicted accurately but only 67% (true failures) of those without testicular spermatozoa were predicted as failures. Hence, 25% of the patients with testicular spermatozoa would have been denied testicular biopsies and 33% of patients with testicular spermatozoa would have had unnecessary testicular biopsies. The best cut-off point for plasma FSH concentration was 12.7 IU/l. Using the physiological range of 1–12 IU/l, only 50% of the patients with testicular spermatozoa were predicted correctly on the basis of plasma FSH concentration. However, 92% of the true failures were predicted correctly. The best cut-off point for the Johnsen score was ~4.8. At this cut-off point, only 61% of true successes were correctly predicted in contrast to 83% of true failures. Only 17% of patients therefore would have had unnecessary testicular biopsies. However, due to the low sensitivity, 39% of couples would have been denied the chance of undergoing ICSI. The presence of testicular spermatids on histological examination provided the highest prediction rates with an area under the curve of 77%, in addition to high sensitivity (72%) and specificity (92%).

Table II. Testicular sperm extraction in relation to the presence of testicular spermatids

Testicular spermatids	Testicular sperm extraction	
	Success (28)	Failure (12)
Not visualized (n = 19)	8 (42)	11 (58)
Visualized (n = 21)	20 (95)	1 (5)

Percentage values in parentheses are calculated separately for each row (sensitivity 71%, specificity 92%, positive predictive value 95% and negative predictive value 58%).

Table III. The sensitivity, specificity and diagnostic accuracy (the area under the receiver operating characteristic curve) of each variable at the best cut-off point

Variable	Best cut-off points	Sensitivity (%)	Specificity (%)	Accuracy (%)
Testicular volume (ml)	27	75	67	72
Plasma FSH (IU/l)	12.7	50	92	76
Johnsen score	4.8	71	75	72
Testicular spermatids (n)	1	71	92	77

FSH = follicle stimulating hormone.

Multivariate analysis for the independent predictors

Logistic regression analysis showed that only five modalities (testosterone, FSH, testicular volume, Johnsen score and testicular spermatids) were associated with the likelihood of successful testicular sperm extraction with the level of significance set at 5% (Table IV). Five other modalities (maternal and paternal age at birth, age, BMI, LH) were not predictive. The model, however, identified only two independent predictors (testicular spermatid and Johnsen's score) of testicular sperm extraction (Tables IV and V). The other variables failed to improve the power of the discriminating equation. The variable that made the most effective prediction was the presence of testicular spermatid. The odds ratios and 95% CI for each of the variables derived from the logistic model are shown in Table V. The odds of successful sperm extraction if spermatids were visualized on testicular histology were 168 times higher than when spermatids were absent having taken the Johnsen score into consideration. For each increase of 1 point in the

Table IV. Outcome of stepwise logistic regression analysis to identify potentially predictive variables for successful testicular sperm retrieval

Variables	*P* value
Patient's age	0.09
Father's age at birth	0.46
Mother's age at birth	0.81
Body mass index	0.54
Luteinizing hormone	0.10
Testicular volume	0.028
Follicular stimulating hormone	0.006
Testosterone	0.01
Johnsen score	0.008
Testicular spermatids	0.0003

NS = not significant.

Table V. The independent predictive variables associated with successful testicular sperm extraction

Variable	Coefficient	SE	*P* value	Odds ratio	95% CI
Testicular spermatid	5.12	1.72	0.003	167.9	5.8–48.72
Johnsen score	1.01	0.43	0.02	2.76	1.19–6.41
Constant	–5.14	2.19	–	–	–

SE = standard error; CI = confidence interval.

Johnsen score, the odds of successful sperm extraction were increased by a factor of 2.76. Using the regression coefficients for spermatids, Johnsen score and a constant, a formula (diagnostic index) was developed for the prediction of testicular sperm retrieval (Appendix 2).

Validation of the diagnostic index

Tables VI and VII describe the predicted probability of testicular sperm extraction using the diagnostic index in the derived population obtained using the Jackknife technique (Efron and Gong, 1983). This population is identical to the same 40 patients involved in this study. Using a cut-off point of 50%, three patients would have been denied ICSI treatment while two patients would have had unnecessary testicular biopsy, giving a misclassification rate of 13%. The addition of Johnsen score to testicular spermatids improved the sensitivity from 71% (presence of testicular spermatids alone) to 93% (testicular spermatids combined with Johnsen score), negative predictive value from 58 to 82% and accuracy of prediction from 77% (presence of spermatids alone) and 72% (Johnsen score alone) to 87% (Johnsen score combined with presence of spermatids; Tables II and VII).

Discussion

Our results are consistent with those from other centres in showing both the limitations of testicular volume and plasma FSH concentration and the advantages of histopathology as predictors of sucessful testicular sperm extraction (Hauser *et al*., 1995; Martin-du-Pan and Bischof, 1995; Tournaye *et al*., 1996, 1997). Tournaye *et al*. (1997) found that the accuracy of predicting testicular sperm extraction was 87% if at least one spermatozoon was observed during histopathological examination of a randomly taken testicular biopsy compared

Table VI. Jackknife predictions from logistic regression models with two independent predictor variables (cut-off point set at 50%)

	Predicted testicular sperm retrieval		
	Success	Failure	Total
Observed testicular sperm retrieval			
Yes	26	2	28
No	3	9	12
Total	29	11	40

The data refer to numbers of patients.

Table VII. Predictive value of the diagnostic index in the derived population

Assessment criteria	Derived population	95% CI
Sensitivity	26/28 (93)	77–99
Specificity	9/12 (75)	43–95
Positive predictive value	26/29 (90)	73–98
Negative predictive value	9/11 (82)	48–98
Misclassification rate	5/40 (13)	4–27

Values in parentheses are percentages.
Derived population refers to the population from which the diagnostic index was derived.
CI = confidence interval.

with 69% for plasma FSH concentration, 69% for volume of the larger testis and 52% if at least one spermatozoon was observed in the semen. These authors, however, used testicular spermatozoa rather than spermatids as the histopathology variable. The problem with this approach is the difficulty of identifying spermatozoa on testicular histology in patients with azoospermia due to defective spermatogenesis (Silber and Rodriguez-Rigau, 1981; Silber *et al*., 1996). Spermatids are the germ cells most easily identified morphologically (Silber and Rodriguez-Rigau, 1981), hence facilitating the analysis of histology. Our finding that visualization of testicular spermatids correlates with the probability of successful testicular sperm extraction is consistent with other reported findings (Salzbrunn *et al*., 1996; Mulhall *et al*., 1997; Silber *et al*., 1997). Salzbrunn *et al*. (1996) used the histological identification of mature spermatids to determine which cryopreserved testicular tissues to process for sperm extraction of patients (aged 68–76 years) with obstructive and non-obstructive azoospermia. Although all the patients showed mature spermatids on histological sections and yielded spermatozoa during testicular sperm extraction, the sample size comprised only five patients. Other reports have evaluated patients with only non-obstructive azoospermia. Silber *et al*. (1997) reported successful sperm extraction in 22 out of 26 (85%) azoospermic men with unsuitable mature spermatids in their testicular histology and in one out of 19 (5%) patients in whom no spermatids were seen (sensitivity 96%, specificity 82%, positive predictive value 85% and negative predictive value 95%). While their study resembles ours in terms of sperm retrieval rate in cases where testicular spermatids were visualized on testicular histology, its negative predictive values differed. However, our prediction results with spermatids alone (sensitivity 71%, specificity 92%, positive predictive value 95% and negative

predictive value 58%) are similar to those of another study that reported successful sperm extraction from 13 out of 22 patients (59%) who had no identifiable spermatids on testicular histology and from all eight patients (100%) with spermatids present (sensitivity 38%, specificity 100%, positive predictive value 100% and negative predictive value 41%) (Mulhall *et al*., 1997). The reasons for discordance in the false negative results remain unexplained.

The poorer predictive values of testicular volume and plasma FSH concentration compared to the presence of testicular spermatids are not surprising. Variations in plasma FSH concentration can arise for reasons unrelated to spermatogenesis and many patients with maturation arrest have normal plasma FSH concentrations and testicular volume (Martin-du-Pan and Bischof, 1995). Testicular volume shows a wide variation as a result of many factors, including racial variation (Takihara *et al*., 1983). Although 79% of our patients were caucasians, 21% came from other ethnic backgrounds. The epididymis or scrotum can obscure the true measurement of testicular volume made with an orchidometer. Estimation of testicular volume using ultrasonography may improve its accuracy (Patel and Pareek, 1989). Plasma FSH is directly under the influence of the Sertoli cell via inhibin and Leydig cells via negative feedback from testosterone. However, the activities of the germ cells themselves, rather than the Sertoli or Leydig cells, are assessed directly when the presence of spermatids is assessed. That spermatids can predict successful testicular sperm retrieval is not surprising, as maturation arrest is usually rare (Soderstrom and Suominen, 1980; Silber and Rodriguez-Rigau, 1981), apoptosis has decreased and cell division has ceased at the level of spermiogenesis, so that some testicular spermatozoa would invariably be produced somewhere in the testes once spermatogenesis has progressed to this stage.

To investigate the independent variables that are predictive of successful testicular sperm extraction, logistic regression analysis was undertaken to avoid the related effects of different co-variates. The model identified only two independent predictors (testicular spermatid and Johnsen score). The odds of successful sperm extraction if spermatids were present were 168 times higher than if they were absent having taken the Johnsen score into consideration. Although the 95% confidence intervals are very wide, there is a clinically important relationship between testicular spermatids and testicular sperm extraction because the true odds ratio is $\geqslant 6$. The reason for the relatively low predictive value of the Johnsen score is unknown but may reflect the fact that the score was originally designed to correlate sperm count with testicular pathology in men with oligozoospermia. The improved performance of the Johnsen score in the logistic regression analysis compared to analysis with ROC may be due to the removal of the effects of other covariates associated with spermatogenesis. The diagnostic index correctly classified 87% of the patients in the derived sample. The misclassification observed was due to the focal nature of spermatogenesis with the result that the tissues taken for histology was different in terms of the status of spermatogenesis from those used for testicular sperm extraction (Tournaye *et al*., 1997). However, the addition of Johnsen score to testicular spermatids improved its sensitivity, the negative predictive value and the accuracy of prediction (Tables II and VII). The quality of the histopathological evaluation of diagnostic testicular biopsies is critical to the predictive power of the diagnostic index. A number of factors, including the method of handling of the biopsy sample, tissue fixation, embedding and sectioning, as well as the experience of the assessor are important (Holstein *et al*., 1994). Seminiferous tubules are very delicate and must be collected by an atraumatic technique. Bouin's solution has been reported to be better than formalin for tissue fixation and epoxy resin better than paraffin wax for tissue embedding in terms of minimizing the distortion of tissue architecture (Holstein *et al*., 1994). Despite the use of paraffin for tissue embedding, all our histological slides were suitable for analysis and a detailed cytological evaluation of the various stages of spermatogenesis was possible. The combined opinion of different assessors reduced the possibility of inter- and intra-observation errors.

Exploration of the genital tract was not used to exclude obstructive azoospermia and patients with primary gonadal failure were selected on clinical criteria. Testicular histology confirmed this diagnosis in all the patients studied. In spite of this we achieved a sperm retrieval rate of 70%, similar to the value of 62% reported by Schlegel *et al*. (1997) from a single large biopsy but higher than the value of 50% reported by Tournaye *et al*. (1996) from up to 20 small multiple biopsies. This finding suggests that the quantity of tissue used for sperm extraction may be as important as the site of the biopsy. The efficacy of sperm extraction may be further increased by the use of a combination of methods including in-vitro culture of testicular tissue, the use of erythrocyte lysing buffer and mechanical shredding of the testicular tissues (Zhu *et al*., 1995; Silber *et al*., 1996; Nagy *et al*., 1997). Moreover, the testicular sperm retrieval rate also was not correlated with sperm vitality or with the outcome of ICSI in our study.

It is important to state the benefits and limitations of this diagnostic index. (i) Although the process of deriving the diagnostic index is complicated, it is very simple to use since it involves only two variables—spermatids and Johnsen score. Most pathologists can assess Johnsen score and the presence of spermatids from a diagnostic testicular biopsy and the natural logorithm values of the predicted scores can be calculated with or without a computer to obtain the predicted probability for each individual patient. (ii) Because more than one oocyte is often retrieved during ICSI–IVF treatment, the single testicular spermatozoon used as the main outcome measure in this analysis is less than the 6–10 spermatozoa which are usually required for successful ICSI (Devroey *et al*., 1996). Nevertheless, the exact number of testicular spermatozoa which should be retrieved in a prior diagnostic biopsy before a couple proceeds to ICSI is unresolved. Although some patients are ready to undergo ICSI treatment in the presence of only a single testicular spermatozoon, this is unrealistic and has a poor prognosis. (iii) The predictive power of our diagnostic index contains a margin of error indicated by the confidence interval. Hence it is necessary to continue data collection in order to improve this. (iv) This model should ideally be validated prospectively in a larger cohort. However, it has

been suggested that model testing in a derived population using the Jackknife technique is nearly as good as its evaluation in a prospective population (Efron and Gong, 1983). (v) It can be argued that a testicular biopsy must be obtained to provide the diagnostic index and hence that any spermatozoa obtained may be cryopreserved for future ICSI. Moreover, it has been demonstrated that both diagnostic and therapeutic testicular biopsies may be associated with potential vascular complications (Jarow, 1991; Harrington *et al*., 1996; Schlegel and Su, 1997). Nevertheless, a prior diagnostic testicular biopsy may enable detection of carcinoma *in situ* (CIS) of the testis, which occurs in 1–3% of patients with severe male factor infertility (Skakkebaek, 1978; Devroey, 1996). Half (50%) of such cases will progress to testicular cancer (Nevero *et al*., 1996). Both CIS and seminoma of the testis have been reported by patients undergoing TESE (Devroey, 1996; Nevero *et al*., 1996; Tournaye *et al*., 1996). In addition, there is increasing evidence that a preliminary diagnostic testicular biopsy can ascertain that complete spermatogenesis is present in at least in one of the seminiferous tubules (Silber *et al*., 1997; Tournaye *et al*., 1997; Vanderzwalmen *et al*., 1997). Percutaneous needle aspiration does not suffice for sperm extraction in men with azoospermia due to defective spermatogenesis (Frielder *et al*., 1996; Ezeh *et al*., 1998), and a single needle biopsy may cause less vascular injury than a single open biopsy (Harrington *et al*., 1997). Only ~4 μg of testicular tissue is required for histopathological examination and the assessments of Johnsen score and testicular spermatids (Craft *et al*., 1997) causing minimal vascular injury to the testis. (vi) It has recently been found that the outcome of oocyte injection with spermatids is better in patients with incomplete spermiogenic arrest (who have round, elongating or elongated spermatids) than in patients with complete spermiogenic arrest (no development beyond the round spermatid stage) (Vanderzwalmen *et al*., 1997; Amer *et al*., 1998). Since microspermatid injection of oocytes is currently banned in the UK, our therapeutic biopsies were not classified, but future studies should relate the stage of spermatid development to the outcome of microinjections. (vii) The presence of spermatids is a categorical variable with a binary outcome, and hence can only generate one reference point on the ROC plot. This type of analysis may be limited for a categorical variable without many reference points, relative to continuous variables such as plasma FSH, testicular volume, Johnsen score or a combination of categorical variables with multiple outcomes. However, a subsequent analysis with logistic regression analysis confirmed the value of this assessment campared with other variables in predicting the likelihood of successful testicular sperm extraction. Moreover, the area under the ROC curve was calculated from a Mann–Whitney test rather than the trapezoidal rule, which is more sensitive to the location and the spread of points defining the ROC curve (Hanley and McNeil, 1982).

This study suggests that the presence of testicular spermatids and the Johnsen score as measures of germ cell status are the most reliable variables associated with successful testicular sperm extraction. Further research on the prediction of testicular sperm retrieval should therefore focus on germ cell assessment. The results also suggest that the predictive power is improved by the inclusion of the Johnsen score. These results provide further support for the use of a diagnostic testicular biopsy prior to a therapeutic testicular biopsy for testicular sperm extraction (Tournaye *et al*., 1996; Silber *et al*., 1997; Vanderzwalmen *et al*., 1997). To gain maximum information and optimize results for men undergoing an unpleasant invasive procedure, a combination of histology, trial testicular extraction and cryopreservation of testicular tissue or spermatozoa for use at a later date should be the procedure of choice.

References

Amer M., Soliman, E., El-Sadek, M. *et al*. (1998) Is complete spermiogenesis failure a good indication for spermatid conception? *Lancet*, **350**, 116.

Chillion, M., Casals, T., Gimenez, J. *et al*. (1994) Analysis of the CFTR gene confirms the high genetic heterogenicity of the Spanish population: 43 mutations account for only 78% of CF chromosomes. *Hum. Genet*., **93**, 447–451.

Craft, I., Tsirigotis, M., Courtauld, E. and Farrer-Brown, G. (1997) Testicular needle aspiration as an alternative to biopsy for the assessment of spermatogenesis. *Hum. Reprod*., **12**, 1483–1487.

Devroey, P. (1996) Microsurgical epididymal aspiration (MESA) and testicular sperm extraction (TESE) — indications and results. ESHRE Campus symposium, Tel Aviv, Israel, March 27–28, 1996.

Devroey, P., Liu, J., Nagy, Z., Goossens, A. *et al*. (1995) Pregnancy after testicular sperm extraction and intracytoplasmic injection in non-obstructive azoospermia. *Hum. Reprod*., **10**, 1457–1460.

Efron, B. and Gong, G. (1983) A leisurely look at the bootstrap, the Jackknife and crossvalidation. *Amer. Statistician*, **37**, 36–48.

Ezeh, U.I.O., Moore, H.D.M. and Cooke, I.D. (1998) A prospective study of multiple needle versus versus a single open testicular biopsy for testicular sperm extraction in men with non–obstructive azoospermia. *Hum. Reprod*., **13**, 3075–3080.

Frielder, S., Raziel, A., Strassburger, D. *et al*. (1997) Testicular sperm retrieval by fine needle sperm aspiration compared with testicular sperm extraction by open biopsy in men with non-obstructive azoospermia. *Hum Reprod*., **12**, 1488–1493.

Hanley, J.A. and McNeil, B.J. (1982) The meaning and use of the areas under the receiver operating characteristic (ROC) curve. *Radiology*, **142**, 29–36.

Harrington, T., Schauer, D. and Gilbert, B. (1996) Percutaneous testis biopsy: an alternative to open testicular biopsy in the evaluation of subfertile man. *J. Urol*., **156**, 1647–1651.

Hauser, R., Temple-Smith, P.D., Southwick, G.J. and de Kretser, D. (1995) Fertility in cases of hypergonadotropic azoospermia. *Fertil. Steril*., **63**, 631–636.

Holstein, A.F., Schultze, W. and Breuker, H. (1994) Histopathology of human testicular and epididymal tissues. In Hargreave, T.B. (ed.), *Male Infertility*, 2nd edn. Springer, London, pp 104–148.

Jarow, J.P. (1991) Clinical significance of inter-testicular anatomy. *J. Urol*., **145**, 777–779.

Johnsen, S.G. (1970) Testicular biopsy score count—a method for registration of spermatogenesis in human testes: normal values and results in 335 hypogonadal males. *Hormones*, **1**, 2–25.

Jow, W.W., Steckel, J., Schlegel, P. *et al*. (1993) Motile sperm in human testis biopsy specimens. *J. Androl*., **14**, 194–198.

Lindsay, K.S., Floyd, I. and Swan, R. (1995) Classification of azoospermic samples. *Lancet*, **345**, 1642–1643.

Martin-du-Pan, R.C. and Bischof, P. (1995) Increased follicle stimulating hormone in infertile men. Is increased plasma FSH always due to damaged germinal epithelium? *Hum. Reprod*., **10**, 1940–1945.

Mulhall, J.P., Burgesss, C.M., Cunningham, D. *et al*. (1997) The presence of mature sperm in testicular parenchyma of men with non obstructive azoospermia: prevalence and predictive factors. *Urology*, **49**, 91–95.

Nagy, Z.P., Verheyen, G., Tournaye, H. *et al*. (1997) An improved treatment procedure for testicular biopsy specimens offers more efficient recovery: case series. *Fertil. Steril*., **68**, 376–379.

Novero, V., Goosens, A., Tournaye, H. *et al*. (1996) Seminoma discovered in two males undergoing successful sperm extraction for intracytoplasmic sperm injection. *Fertil. Steril*., **65**, 1051–154.

Patel, P.J. and Pareek, S.S. (1989) Scrotal ultrasound in male infertility. *Eur. Urol*., **16**, 423–425.

Ron-El, R., Strassburger, D., Friedler, S. *et al.* (1997) Extended sperm preparation: an alternative to testicular sperm extraction in non-obstructive azoospermia. *Hum. Reprod.*, **12**, 1222–1226.

Salzbrunn, A., Benson, D.M., Holstein, A.F. and Schulze, W. (1996) A new concept for the extraction of testicular spermatozoa as a tool for assisted fertilization (ICSI). *Hum. Reprod.*, **11**, 752–755.

Schlegel, P.N. and Su, L. (1997) Physiological consequences of testicular sperm extraction. *Hum. Reprod.*, **12**, 1688–1992.

Silber, S.J. and Rodriguez-Rigau, L.Y. (1981) Quantitative analysis of testicular biopsy: determination of partial obstruction and prediction of sperm count after surgery for obstruction. *Fertil. Steril.*, **36**, 480–485.

Silber, S.J., Van Steirteghem, A.C., Nagy, Z. *et al.* (1996) Normal pregnancies resulting from testicular sperm extraction and intracytoplasmic sperm injection for azoospermia due to maturation arrest. *Fertil. Steril.*, **66**, 110–117.

Silber, S.J., Nagy, Z., Liu, J. *et al.* (1997) Distribution of spermatogenesis in the testicles of azoospermic men: the presence or absence of spermatids in the testes of men with germinal failure. *Hum. Reprod.*, **12**, 2422–2428.

Skakkebaek, N.E. (1978) Carcinoma *in situ* of the testis: frequency and relationship to invasive germ cell tumours in infertile men. *Histology*, **2**, 157–170.

Soderstrom, K.O. and Suominen, J. (1980) Histopathology and ultrastructure of meiotic arrest in human spermatogenesis. *Arch. Pathol. Lab. Med.*, **104**, 476–82.

Takahira, H., Sakatoku, J., Fujii, M. *et al.* (1983) Significance of testicular size measurement in andrology. 1. A new orchidometer and its clinical application. *Fertil. Steril.*, **39**, 836–840.

Tournaye, H.J., Liu, J., Nagy, Z. *et al.* (1996) Correlation between testicular histology and outcome after intracytoplasmic sperm injection using testicular sperm. *Hum. Reprod.*, **11**, 127–132.

Tournaye, H. J., Verheyen, G., Nagy, Z. *et al.* (1997) Are there any predictive factors for successful testicular sperm recovery in azoospermic patients? *Hum. Reprod.*, **12**, 80–86.

Vanderzwalmen, P., Zech, H., Birkenfeld, A. *et al.* (1997) Intracytoplasmic injection of spermatids retrieved from testicular tissue: influence of testicular pathology, type of selected spermatids and oocyte activation. *Hum. Reprod.*, **12**, 1203–1213.

Verma, R.S. and Babu, A. (1995) *Human Chromosomes: Principles and Techniques*, 2nd edn. McGraw-Hill, New York, Vol. 5, pp. 309–314.

World Health Organization (1992) *WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interaction*, 3rd edn. Cambridge University Press, Cambridge.

Zhu, J., Tsirigotis, M. and Craft, I. (1996) In-vitro maturation of testicular spermatozoa. *Hum. Reprod.*, **11**, 231–232.

Zweig, M.H. and Campbell, G. (1993) Receiver–operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, **39**, 561–577.

Received on June 8, 1998; accepted on December 9, 1998

Appendix 1

Determination of the area under the receiver operating characteristic (ROC) plot. The area under the ROC plot was determined statistically using the Mann–Whitney *U*-test (Hanley and McNeil, 1982; Zweig *et al.*, 1993) using the formula:

$$\text{area under the ROC curve} = 1 - \frac{U}{n_a n_b}$$

where U = (Mann–Whitney value)^2, n_a = number of patients with failed sperm extraction, and n_b = number of patients with successful sperm extraction.

Appendix 2

The formulae for predicting testicular sperm retrieval using the presence of one testicular spermatid and the Johnsen score:

(i) $Y = a + [C2.(J)] + [C1.(S)]$

(ii) $P = e^Y$ divided by $(1 + e^Y)$

where Y = predicted score, P = the probability of finding testicular spermatozoa, a = the constant, $C1$ = the coefficient for testicular spermatid (Table V), $C2$ = the coefficient for Johnsen score (Table V), S = value for testicular spermatid (present = 1 or absent = 0), J = value for Johnsen score (1–10), e^Y = exponential function applied to Y

diagnostic index: $Y = -5.14 + [1.01 \times J] + [5.12 \times 1 \text{ or } 0]$.

Y is then used to determine the probability (P) of sperm extraction.

Example of how to use the diagnostic index

(a) A patient with a Johnsen score of 7 and with spermatids seen on histology:

(i) $Y = -5.14 + [1.01 \times 7] + 5.12 = 7.05$
(ii) $P = e^{7.05} \div (1 + e^{7.05}) = 0.999$ (almost 100% chance)

(b) A patient with a Johnsen score of 2 and no spermatid seen (Sertoli cell-only)

(i) $Y = -5.14 + [1.01 \times 2] + 0 = 7.15$
(ii) $P = e^{3.12} \div (1 + e^{3.12}) = 0.04$ (1 in 25 chance)