Journal of Pharmaceutical and Biomedical Analysis 54 (2011) 106–113

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

^{1}H NMR spectroscopic studies on human seminal plasma: A probative discriminant function analysis classification model

Ashish Gupta[a], Abbas Ali Mahdi[a,*], Mohammad Kaleem Ahmad[a], Kamla Kant Shukla[a], Shyam Pyari Jaiswer[b], Satya Narain Shankhwar[c]

[a] Department of Biochemistry, Chhatrapati Shahuji Maharaj Medical University, Lucknow, India
[b] Department of Obstetric & Gynecology, Chhatrapati Shahuji Maharaj Medical University, Lucknow, India
[c] Department of Urology, Chhatrapati Shahuji Maharaj Medical University, Lucknow, India

ARTICLE INFO

Article history:
Received 24 April 2010
Received in revised form 6 July 2010
Accepted 18 July 2010
Available online 16 August 2010

Keywords:
Seminal plasma
Infertility
NMR spectroscopy
Discriminant function analysis

ABSTRACT

Traditional seminal fluid-based clinical descriptors used to predict infertility and sub-fertility have shortcomings, including lack of insight into the underlying pathology. These methods are also time-consuming and labor-intensive. To address these limitations, ^{1}H nuclear magnetic resonance (NMR) spectroscopy was used to identify and classify signature biomarkers. Semen samples collected from 60 healthy, fertile men and from 125 infertile (normozoospermic and oligozoospermic) patients. Lactate, alanine, choline, citrate, glycerophosphocholine (GPC), glutamine, tyrosine, histidine, phenylalanine, and uridine were measured by ^{1}H NMR spectroscopy. The sperm concentration, motility, lipid peroxidation, and total protein were evaluated with standard laboratory methods in the same samples. NMR-quantified metabolites and clinical laboratory data were analyzed, separately, through linear multivariate discriminant function analysis (DFA) to determine the signature descriptors for each group. DFA reveals that alanine, citrate, GPC, tyrosine, and phenylalanine can be used to determine infertility. DFA-based classification demonstrated high accuracy (92.4% by NMR and 94.1% by clinical laboratory method) in differentiating healthy controls from infertile patients. This statistical analysis was also able to accurately classify normozoospermic to oligozoospermic samples (92.9% by NMR and 92.6% by clinical laboratory method). In conclusion, ^{1}H NMR-based metabolic screening appears to be a promising, rapid, and non-invasive approach to probing infertility that has similar sensitivity and specificity to the tedious laboratory method.

1. Introduction

Infertility, defined as the inability of a couple to conceive after 1 year of regular unprotected intercourse, is a global health issue, affecting 12–15% of couples worldwide [1]. It affects up to one in five couples during reproductive age [2]. The high rates of infertility in the developing part of the world have a variety of causative factor, but in 10–20% of the cases, no definitive cause can be identified, and the infertility remains unexplained [3].

Although rarely socially acknowledged, male infertility contributes to nearly 50% of all cases worldwide and is often the most difficult form of infertility to treat [4].

The assessment of male infertility has traditionally been based on the sperm morphology, motility, and concentration [5]. Despite the availability of advanced diagnostic methods for the diagnosis and treatment of infertility, the cause of infertility is still unknown in 10–20% cases. Therefore, there is a need for new signature biomarkers that offer a rapid, non-invasive, sensitive, accurate, and unambiguous method for the diagnosis and treatment of infertility [6–8].

Seminal plasma is a highly complex biological fluid that contains inorganic ions, low molecular weight organic compounds, peptides, and hormones [9,10]. These organic and inorganic complex mixtures are secreted from different endocrine glands and are the functional reflective index of the corresponding glands [11,12]. The presence of many of these constituents in seminal plasma is well established, and few have proven clinical relevance [12,13].

High-resolution nuclear magnetic resonance (NMR) spectroscopy has emerged as an important diagnostic technique in addressing many health problems using different biological fluids and tissue extracts [13,14]. This technique has made it possible not only to identify the constituents of seminal fluid but also to provide qualitative and quantitative metabolic information with reasonable

Abbreviations: TSP, tri-methylsilylpropionic acid; CZ, control group; NZ, normozoospermic; OZ, Oligozoospermic.

* Corresponding author at: Medical Elementology and Free Radical Biology Lab, Department of Biochemistry, Chhatrapati Shahuji Maharaj Medical University, Lucknow 226003, India. Fax: +91 522 2257539.
E-mail address: mahdiaa@rediffmail.com (A.A. Mahdi).

0731-7085/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2010.07.021

Table 1
Absolute concentrations (mg/dl) of NMR-derived metabolites quantified in seminal plasma using homemade computer program with respect to a known concentration of tri-methylsilylpropionic acid. All results are shown as mean ± S.D. GPC, glycerophosphorylcholine.

Metabolite name	Control ($n = 60$)	Normozoospermic ($n = 65$)	Oligozoospermic ($n = 60$)
Lactate	114 ± 37	119 ± 67	107 ± 37
Alanine	72 ± 45	49 ± 27*	65 ± 25$
Citrate	480 ± 74	361 ± 84**	315 ± 20***
Choline	207 ± 64	184 ± 78	199 ± 59
GPC	217 ± 77	107 ± 50***	114 ± 47***
Glutamine	42 ± 12	39 ± 11	44 ± 15
Tyrosine	134 ± 48	145 ± 68	153 ± 55
Histidine	110 ± 73	85 ± 40	84 ± 32
Phenylalanine	28 ± 14	43 ± 8***	37 ± 12**
Uridine	34 ± 16	35 ± 20	41 ± 20

* $p < 0.05$ as compared to the control.
** $p < 0.01$ as compared to the control.
*** $p < 0.001$ as compared to the control.
$ $p < 0.05$ as compared to the normozoospermic.

time and clinically relevant means of measuring concentrations (μM range) [15,16]. Importantly, the information obtained can be used to identify infertility [12,13,17]. It has been used to distinguish between spermatogenic failure and obstructive azoospermia as well as different forms of spermatogenic failure on the basis of glycerophosphorylcholine (GPC)/choline ratio [12]. The relative amount of GPC, choline and other metabolites are typically reported in experimental and clinical NMR studies because it is measured in a single NMR acquisition, provides important metabolic profile of seminal plasma in healthy men, and is perturbed by pathological conditions [12,13,17]. However, the ratio GPC, choline and other metabolites does not provide a complete picture of the metabolic profile of infertility. The aim of this study was to quantitatively measure the absolute concentration of metabolites using NMR spectroscopy and to determine the signature biomarkers and descriptors of infertility using linear discriminant function analysis (DFA). This approach may provide a rapid, precise, simple, sensitive, and comprehensive diagnostic method for the differential diagnosis of different classes of infertility. Further, use of this technique may contribute toward better understanding of the vital biological causes of infertility, an essential step for improving its clinical management.

2. Materials and methods

2.1. Patients and sample collection

The institutional review board and ethical committee of Chhatrapati Shahuji Maharaj Medical University, Lucknow, approved this study. Participants were recruited from couples attending the infertility clinic of the Department of Obstetrics and Gynecology, and the Department of Urology, Chhatrapati Shahuji Maharaj Medical University. The study population included 22–45 years-old men who were healthy fertile male ($n = 60$) and infertile ($n = 125$). Infertile men, were either normozoospermic ($n = 65$) or oligozoospermic ($n = 60$). Written informed consent was obtained from all participants before enrollment in the study. Exclusion criteria included: infections of accessory grand, smoking, stress, diabetes, hypertension, arthritis, malignancies, tuberculosis, HIV, urinary tract infections, endocrine disorders, drug abuse, and other conditions known to influence sperm physiology and male fertility.

The control group comprised healthy males who had previously initiated at least one pregnancy and exhibited a normal semen profile (>20 × 10^6 spermatozoa/ml; >40% motility and >40% normal morphology). The patient group consisted of two subgroups. The first included normozoospermic infertile males who had a normal semen profile (>20 × 10^6 spermatozoa/ml; >40% motility; and >40% normal morphology) and fertile female partners who had undergone extensive infertility evaluation without showing any detectable gynecological abnormality. The second subgroup included oligozoospermic males who had a sperm concentration <20 × 10^6 ml^{-1}, >40% motility, and >40% normal morphology.

A semen sample profile was constructed according to the procedures described by WHO [18]. Semen samples were collected by masturbation after 3–4 days of abstinence and divided into two parts. In one part, the semen was liquefied, and an aliquot was taken to assess sperm concentration and motility. In second part, after liquefaction semen sample was centrifuged at 1200 × g for 20 min for separation of seminal plasma. The supernatant (seminal plasma) was centrifuged at 10,000 × g for 30 min to remove all possible contaminating cells and rapidly frozen in liquid nitrogen.

2.2. NMR experiments

The frozen seminal plasma samples were thawed and 500 μl of each sample was taken separately in 5 mm NMR tubes. For field-frequency-lock and quantitative measurements of metabolites, a sealed reusable capillary containing 25 μl of 0.375% sodium salt of tri-methylsilylpropionic acid (TSP) in deuterium oxide was inserted into the tube, and the spectra were obtained. For all samples, one-dimensional ^{1}H NMR experiments were performed on a Bruker Avance 400 MHz spectrometer at 22 °C using one-pulse sequence

Table 2
The clinical variables; sperm concentration (million/ml), sperm motility (%), LPO (nmolMDA/ml) and total protein (mg/dl) as observed in control and infertile patient groups. All results are shown as mean ± S.D.

Clinical parameters	Control ($n = 60$)	Normozoospermic ($n = 65$)	Oligozoospermic ($n = 60$)
Sperm concentration	76 ± 16	63 ± 30	36 ± 29***,$$$
Motility	73 ± 9	72 ± 12	38 ± 15***,$$$
LPO	1.7 ± 0.2	2.6 ± 0.3***	2.3 ± 0.4***,$$$
Protein	29 ± 3	26 ± 2***	25 ± 4***

*** $p < 0.001$ as compared to the control.
$$$ $p < 0.001$ as compared to the normozoospermic.

with suppression of the water resonance by pre-saturation. The acquisition parameters were spectral width, 8000 Hz; time domain points, 32K; relaxation delay, 20 s; pulse angle, 90°; number of scans, 128; spectrum size, 32K. The spectra were processed with line broadening: 0.3 Hz. Metabolites were quantified with an in-house custom program based on the integration area of identified metabolites relative to the area of the reference signal from TSP [17].

2.3. Statistical analysis

The results obtained for the NMR-based quantified metabolites (mg/dl) are expressed as mean ± S.D. (Table 1). The clinical laboratory variables comprising the sperm concentration (million/ml), sperm motility (in percentage), lipid peroxides (LPO) (nmolMDA/ml), and total protein (mg/dl) content information (Table 2) are also given.

2.3.1. Univariate analysis

The statistical significance for the quantified metabolites and clinical variables were determined by one-way ANOVA followed by a *post hoc* Student–Newman–Keuls multiple comparisons test, carried out with the Graph Pad INSTAT 3.0 software. A probability *p*-value of less than 0.05 was taken to indicate statistical significance.

2.3.2. Multivariate analysis

The data were subjected to multivariate linear DFA to define important variables for differentiation of infertile patients from controls, followed by discrimination of the two types of infertility. For DFA, 10 NMR-based observed metabolites (Table 1) and all four clinical laboratory variables (Table 2) were used, and the analysis were carried out by Statistical Package for the Social Sciences (SPSS version 11.0, SPSS Inc., Chicago, IL). The discriminant functional scores were obtained by summing the independent variable value (from either the NMR data or the clinical variables such as citrate, sperm concentration etc.) and its corresponding discriminant function-based coefficient/weights value.

To develop a comprehensive DFA for infertility, a stepwise forward variable selection procedure was adapted to determine the best set of predictors of infertility using DFA-based discriminant scores. The jack-knife cross-validation (leave-one-out) method of

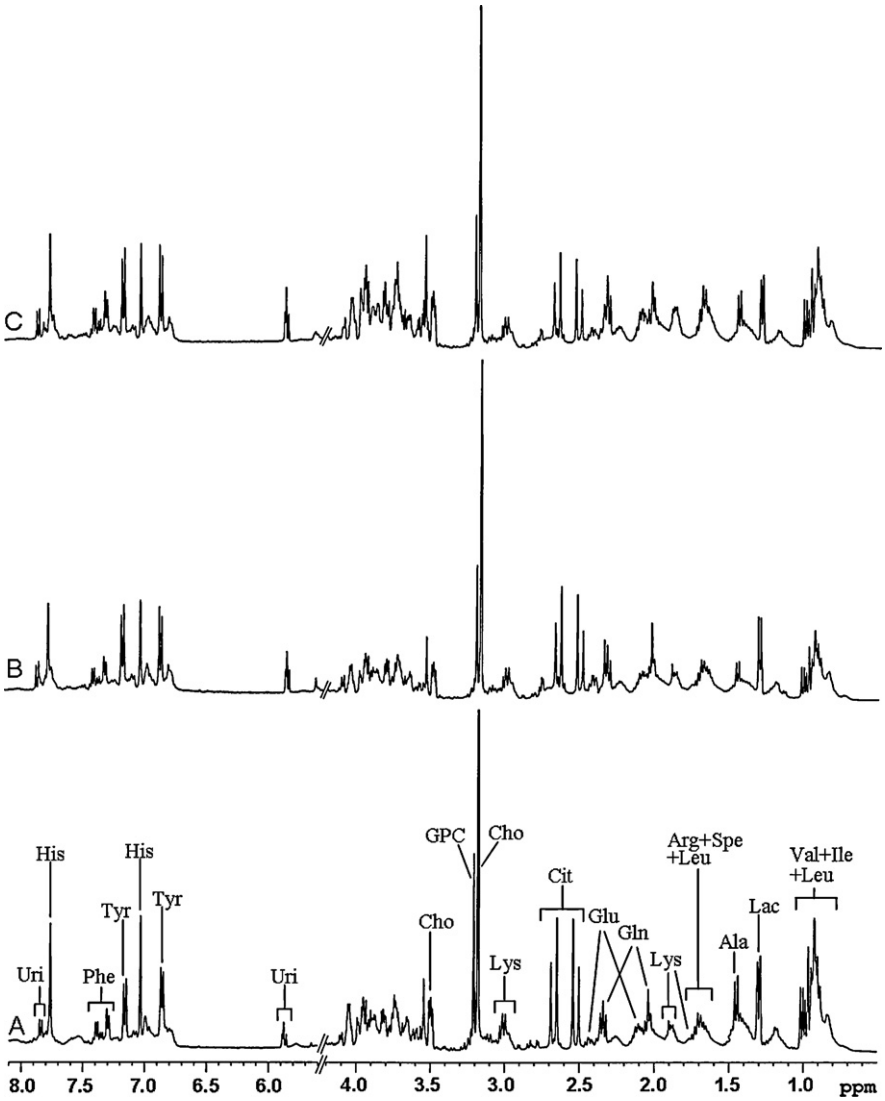

Fig. 1. Typical ^{1}H NMR spectra of human seminal plasma from different groups viz: (A) control subjects, (B) normozoospermic, (C) oligozoospermic. *Key*: Val, valine; Ile, isoleucine; Leu, leucine; Lac, lactate; Lys, lysine; Ala, alanine; Arg, arginine; Spe, spermine; Glu, glutamate; Gln, glutamine; Cit, citrate; Cho, choline; Gly, glycine; Tyr, tyrosine; Uri, uridine; His, histidine; Phe, phenylalanine.

Lachenbruch [19] was used in all calculations to derive optimized discriminant scores. In this method, one subject was removed, and DFA was carried out on the remaining subjects. The resulting discriminant function equation was then applied to the data from the removed subject to predict the group to which the subject belongs. This analysis was performed sequentially on each subject, and the prediction was compared with the actual group. This method provides a relatively unbiased estimate of the error rate of DFA [19]. The receiver-operating characteristic (ROC) plots were also constructed by using the discriminant function score to determine the cutoff values between the groups. Statistical significance of the differences between the ROC curves was evaluated by comparing the area under the curves.

The observed predictors were appraised further by re-substitution and prospective test data methods that provide accuracy of the performance of the DFA model between the chosen categories of infertility. The re-substitution method, which was carried out as part of the external validation process, involved feeding the same infertile and control data into the DFA model to ensure the accurate classifications based on the discriminant function weights of each variable. The prospective test data methods comprised a re-assessment of the prediction possibility by treating 75% of each group of infertility data as training sets and examining the remaining 25% of the data as test sets, which were obtained based on the Random Numbers Table by Fisher and Yates. This method provides a validation of the model built based on DFA comprising four separate training sets, viz. (1) CZ vs. NZ + OZ, (2) CZ vs. NZ, (3) CZ vs. OZ, and (4) NZ vs. OZ (where CZ is a control subject, NZ is normozoospermic, and OZ is oligozoospermic) and evaluating the classifications of the remaining 25% of the infertile patient's data treated as test sets.

3. Results

Fig. 1 shows the typical ^{1}H NMR spectra of a complete metabolic profile and chemical shift assignments of different resonances in seminal plasma from the three different groups. Resonance signals of various metabolites in seminal plasma were assigned by using known chemical shift and coupling constant parameters [10,12,13]. Several metabolite resonances emerged in the complex pattern. Valine, leucine, isoleucine, arginine, glutamate, and lysine were not quantified because of overlapping of resonance assignments. Ten metabolites (lactate, alanine, glutamine, citrate, choline, GPC, tyrosine, histidine, phenylalanine, and uridine) were unequivocally and ubiquitously present in all samples, and their quantities were estimated from their respective resonances (viz., 1.33, 1.47, 2.37, 2.55, 3.21, 3.23, 6.88, 7.01, 7.30–7.44 and 7.84 ppm) and subjected to univariate and multivariate statistical analysis.

3.1. Univariate statistical analysis

Table 1 reveals that the concentrations of citrate, GPC, and phenylalanine were significantly different in the infertile groups than in the control group. The concentration of alanine also was able to differentiate between NZ and OZ as well as NZ and CZ. The other measured metabolites were not significantly different among the three groups. As shown in Table 2 the clinical variables also revealed significant differences between the infertile and control groups. However, compared to multivariate statistical analysis, univariate analysis lacked the ability to provide expedient, sensitive, specific, and discriminant predictive probability for differential diagnosis of each class of infertility.

Table 3
DFA-based classification of individual category of patient from respective controls.

Variables taken for DFA	CZ vs. NZ + OZ				CZ vs. NZ				CZ vs. OZ				NZ vs. OZ			
	*Sen.	*Spe.	*F* ratio	*p*-Value	*Sen.	*Spe.	*F* ratio	*p*-Value	*Sen.	*Spe.	*F* ratio	*p*-Value	*Sen.	*Spe.	*F* ratio	*p*-Value
NMR variables	99%	96%	2.31	<0.0001	99%	96%	2.58	<0.0001	95%	98%	2.09	<0.001	98%	93%	2.29	<0.0001
Selected MR variables[a]	95%	93%	2.29	<0.0003	99%	98%	4.08	<0.0001	95%	98%	3.41	<0.0001	96%	90%	3.78	<0.0001
Clinical variables	95%	95%	3.67	<0.0001	95%	96%	3.42	0.002	95%	93%	3.28	<0.0001	94%	95%	3.12	0.025

CZ, control subject; NZ, normozoospermic; OZ, oligozoospermic; *Sen., sensitivity; *Spe., specificity.
[a] For control vs. all infertile patients, alanine, citrate, GPC, tyrosine, and phenylalanine; For control vs. normozoospermic patients, GPC, tyrosine, alanine, lactate and phenylalanine; For control vs. oligozoospermic patients, GPC, citrate, tyrosine, alanine, histidine and phenylalanine; For normozoospermic vs. oligozoospermic patients, lactate, alanine, phenylalanine, histidine, citrate, and uridine.

3.2. Multivariate statistical analysis

In the array of multivariate analysis, data shown in Tables 1 and 2 were considered for evaluating their effect and revealing the key descriptor variables for the classification of each group. The correlation matrix observations suggest the intra-relationship of spectroscopic and clinical variables separately, and that allowed selection of a set of few variables for correct classification of each group. Table 3 elucidates the classification of infertility as compared to the respective control groups. Step 1: on the basis of DFA, when NZ + OZ (all infertile) cases were compared with CZ, 10 metabolites could successfully classify 96.6% of cases with 99% sensitivity and 96% specificity (F ratio, 2.31; $p < 0.0001$). Step 2: of these 10 metabolites, when only alanine, citrate, GPC, tyrosine, and phenylalanine were chosen based on their discriminant function coefficients, and DFA was performed again, overall 93.2% of infertile cases (NZ + OZ) were effectively classified with 95% sensitivity and 93% specificity (F ratio 2.29; $p < 0.0003$). Similarly, for clinical variables (step 3), DFA revealed that overall 95% of NZ + OZ cases were successfully classified with a sensitivity of 95% and a specificity of 95% with CZ cases (F ratio, 3.67; $p < 0.0001$). Based on the observed correlation matrix of clinical variables, all variables were considered for building the DFA model. We observed that NMR spectroscopic data and clinical variables provide almost the same grade of classification.

When the NZ cases were compared to CZ, 10 metabolites could successfully classify 97% of NZ cases with 99% sensitivity and 96% specificity (F ratio, 2.58; $p < 0.0001$). Of these 10 metabolites, GPC, tyrosine, alanine, lactate, and phenylalanine were chosen based on their discriminant function coefficients. DFA was performed again with the chosen metabolites, and the result showed 98.7% classification of NZ cases with 99% sensitivity and 98.2% specificity (F ratio, 4.08; $p < 0.0001$). For clinical variables, overall 96% of NZ cases were successfully classified with 95% sensitivity and 96.4% specificity (F ratio, 3.42; $p < 0.002$) when compared to CZ. The observations suggest that only few NMR-observed metabolites were accurately differentiated between NZ and CZ cases.

Similarly, when OZ cases were compared with CZ, all measured metabolites could successfully classify 96.8% of OZ cases with a sensitivity of 95% and a specificity of 97.6% (F ratio, 2.09; $p < 0.001$). GPC, citrate, tyrosine, alanine, histidine, and phenylalanine were chosen based on their discriminant function coefficients. DFA was performed again, and the results showed that 96.8% of OZ cases were classified with 95.0% sensitivity and 97.6% specificity (F ratio, 3.41; $p < 0.0001$). Thus, the selected metabolites played crucial roles in segregating OZ cases from CZ. Analysis of clinical variables revealed that 93.5% of OZ cases were successfully classified with 95% sensitivity and 92.9% specificity (F ratio, 3.28; $p < 0.0001$). Therefore, the NMR-based variables were better in distinguishing OZ cases from CZ.

In the same way, when NZ cases were pooled with OZ, all NMR-measured metabolites could successfully classify 95.9% of cases with 98.2% sensitivity and 92.9% specificity (F ratio, 2.29; $p < 0.0001$). On the basis of discriminant function coefficient values, when lactate, alanine, phenylalanine, histidine, citrate, and uridine were chosen and DFA was performed again, the result showed that 93.8% of NZ cases were classified with 96.4% sensitivity and 90.5% specificity (F ratio, 3.78; $p < 0.0001$). Thus, the selected metabolites are important for distinguishing the NZ cases from OZ cases. With clinical variables, 94.8% of NZ cases were successfully classified with a sensitivity of 94.5% and a specificity of 95.2% (F ratio, 3.12; $p = 0.025$). The results suggest that the NMR-based data and the clinical variables provide similar results for classification of NZ and OZ.

Fig. 2 shows the ROC curves of discriminant predicted probability with sensitivity and 1-specificity of different groups with their respective controls. The area under the ROC curves of discrimi-

Table 4
Comparisons of area under the receiver's operating characteristic curve based on discriminant predicted probability of NMR and clinical variables for different classifications.

Test result classification	Discriminant predicted probability of NMR method					Discriminant predicted probability of clinical method				
	Area	Standard error[a]	Asymptotic significance[b]	Asymptotic 95% confidence interval		Area	Standard error[a]	Asymptotic significance[b]	Asymptotic 95% confidence interval	
				Lower bound	Upper bound				Lower bound	Upper bound
CZ vs. NZ + OZ	0.989	0.007	0.000	0.975	1.003	0.951	0.033	0.000	0.887	1.015
CZ vs. NZ	0.994	0.007	0.000	0.980	1.007	0.943	0.035	0.000	0.874	1.011
CZ vs. OZ	0.993	0.007	0.000	0.980	1.006	0.987	0.010	0.000	0.967	1.007
NZ vs. OZ	0.995	0.004	0.000	0.987	1.003	0.951	0.024	0.000	0.904	0.999

CZ, control subject; NZ, normozoospermic; OZ, oligozoospermic.
[a] Under the nonparametric assumption.
[b] Null hypothesis: true area = 0.5.

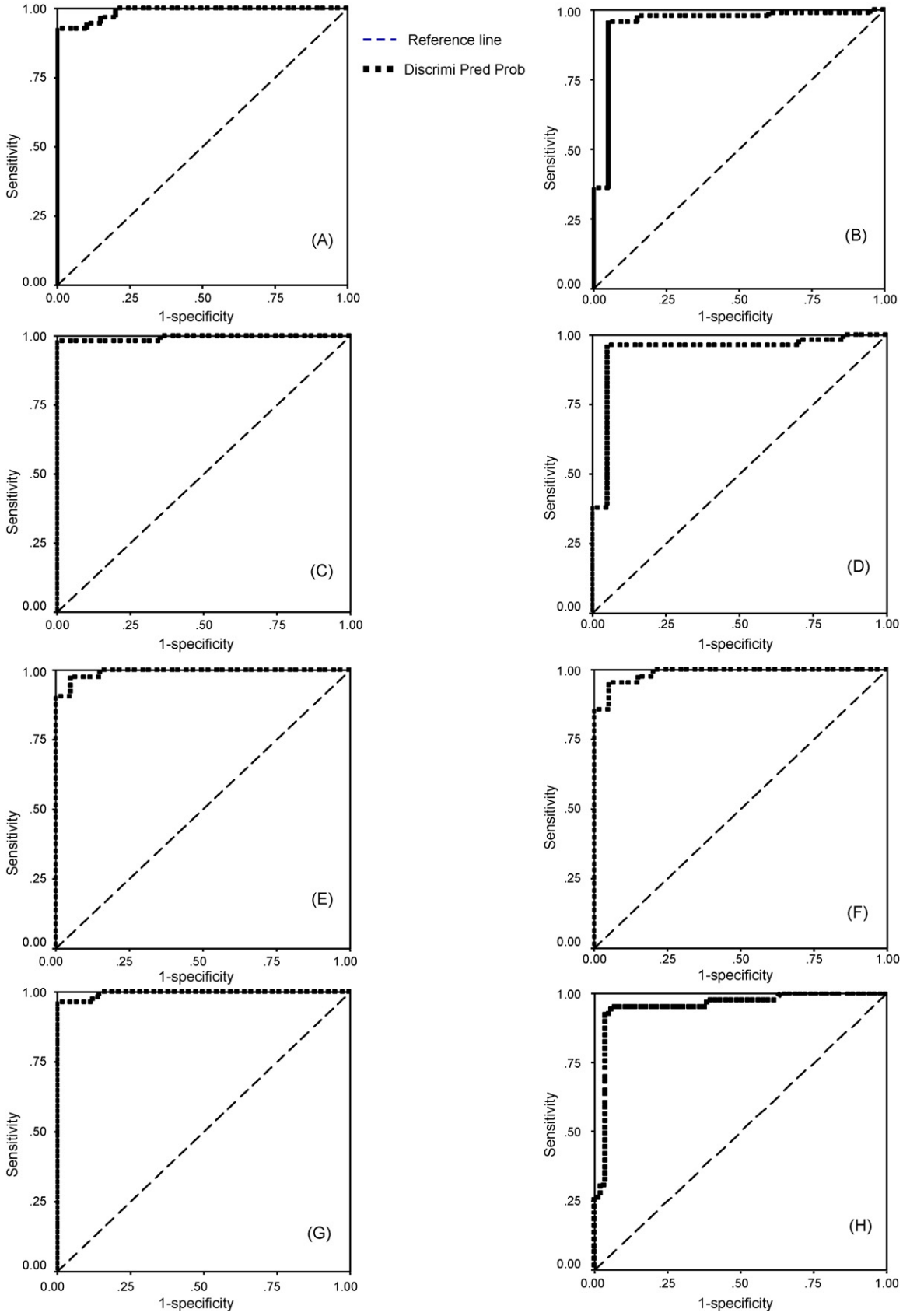

Fig. 2. Comparisons of receiver's operating characteristic (ROC) curves showing the diagnostic accuracy based on discriminant predicted probability of NMR and clinical variables for different groups. (A) ROC curves of CZ vs. NZ + OZ, (C) CZ vs. NZ, (E) CZ vs. OZ, (G) NZ vs. OZ using predicted probability of NMR variables and (B) ROC curves of CZ vs. NZ + OZ, (D) CZ vs. NZ, (F) CZ vs. OZ, (H) NZ vs. OZ using predicted probability of clinical variables (here CZ: control subject; NZ: normozoospermic; and OZ: oligozoospermic).

nant predicted probability of NMR and clinical variables for each classification is compared in Table 4.

4. Discussion

The seminal fluid-based clinical descriptor variables are inefficient for predicting infertility [7]. Sperm morphology appears to be the most important characteristic, but none of the measures alone or together can accurately diagnose infertility and sub-fertility [5,7]. Moreover, clinical variables may not be the sole differentiating markers for infertility and sub-fertility, possibly because of some reported pitfalls [7]. Several studies have demonstrated the potential use of NMR in the identification of seminal plasma constituents and have attempted to use the ratio of different metabolites in the diagnosis of infertility [10,12,13,20]. To circumvent the shortcomings of the traditional clinical method and expand the use of NMR-based quantitative analysis, this study judiciously selected a few absolute-quantified metabolites from seminal plasma and subsequently, analyzed if, they had any effect on the classification of infertility and sub-fertility. Given our results, the NMR spectroscopic information can be used as a predictive model for differential diagnosis of infertility. The data can be used to provide better treatment for infertility, reduce its incidence, and improve its clinical management.

Univariate statistical analysis of the NMR data revealed that citrate, GPC, and phenylalanine metabolites could be used as biomarkers to identify control and the infertile groups. Alanine was able to differentiate normozoospermic from control and oligozoopsermic cases. Similarly, the clinical variables were able to differentiate among the multiple groups.

However, univariate analysis lacks sensitivity, specificity, and predictive probability of the differential diagnosis of the subgroups of infertility. Therefore, we carried out multivariate analysis. This consequential analysis allowed us to construct a sensitive, non-invasive, accurate, and unambiguous identification and classification method of infertility. Hence, a DFA-based systematic approach was performed on the classified data set, followed by a complete cross-validation of data that engaged in comparing each category of infertility with respective controls.

In the linear DFA, the individual infertile category could be efficiently distinguished from the respective control group with excellent sensitivity, specificity, and extremely significant *p*-values (Table 3). This DFA reveals that the control group could be segregated from the infertile group by determining concentrations of alanine, citrate, GPC, tyrosine, and phenylalanine. These spectroscopic metabolites are remarkable descriptors for infertility, which is consistent with an earlier report [12]; however in that study, the ratios of GPC/choline and citrate/lactate were explored, and these ratios provide an inexact description of infertility [12]. The absolute concentration of metabolites is technically more promising to measurement of ratios with other metabolites, as it can uncover perturbations in metabolite pool size and help for correct classification and prediction of infertility, as we observed in this study. The area under the ROC curves (Fig. 2A and B and Table 4) of discriminant predicted probability through these NMR variables is much better when compared to clinical variables. However, the NMR method is fast, sensitive, and non-invasive when compared to the time-consuming, labor-intensive and sample-destructive clinical method.

DFA reveals that GPC, tyrosine, alanine, lactate, and phenylalanine could act as vital descriptors for the segregation of normozoospermic patients from the control group. Furthermore, concentrations of lactate, alanine, phenylalanine, histidine, citrate, and uridine could distinguish normozoospermia from oligozoospermia. Some of these spectroscopic metabolites have been reported previously to identify infertility [12,20], and investigators have attempted to use the different ratios of citrate, glycine, lactate, GPC, and choline to predict different forms of infertility [12]. However, the ratio of metabolites does not provide an accurate measure of the different forms of infertility. Instead of using metabolites ratios, the absolute quantification of metabolites is more rewarding, as it can detect changes in metabolite pool size and support accurate predictions of infertility. Moreover, in addition to other metabolites, the phenylalanine, and tyrosine can be used as key variables to distinguish normozoospermic cases from controls; and phenylalanine and histidine can be used to discriminate normozoospermia and oligozoospermia. Alanine and lactate are products of alanine aminotransferase and lactate dehydrogenase, respectively. The activities of both enzymes were shown to be significantly decreased in the seminal plasma of infertile patients [21], a finding that supports the outcome of our study. Phenylalanine, tyrosine, and histidine or their derivatives are known to be precursors of catecholamines or neurotransmitters [22,23]. Studies indicate that heightened sexual interest may be linked with elevated levels of the brain neurotransmitter, dopamine. Conversely, low dopamine levels correlate with depression and a loss of interest in sex. Hence, dopamine therapy showed improvement in the fertility of oligozoospermic patients [24]. In addition, the neurotransmitter, norepinephrine, directly stimulates the brain's sex center in the hypothalamus and is considered one of the body's natural prosexual chemicals [22,23,25]. Fibrotic thickening of the testicular peritubular region is a common observation in infertile patients and parallels the increased numbers of activated testicular mast cells in these regions. The mast cells, via their major secretory product histamine, are crucially and causally involved in these events and closely associated with male infertility [26]. In mammals, the only source of histamine is histidine, and it is synthesized through histidine decarboxylase (HDC). HDC is localized in the acrosomes of spermatids and spermatozoa. A substantial amount of histamine and HDC activity has been detected in the testis, epididymis, and spermatozoa [27]. W/W^V mice, known to lack most of their germ cells in the seminiferous tubules, were found to lack HDC protein expression as well as HDC activity in the testis [27]. These reports support the decreased amount of histidine in seminal plasma of infertile patients in our study. The area under the ROC curves of CZ vs. NZ (Fig. 2C and D and Table 4) and NZ vs. OZ (Fig. 2G and H and Table 4) of discriminant predicted probability of NMR variables are much better as compared to clinical variables. However, the former method is a fast, sensitive, and non-destructive as compared to the latter time-consuming, labor-intensive, and sample-destructive clinical method.

This study also comprised and compared the ability of the metabolites to predict cases of oligozoospermia. The result explained that GPC, citrate, tyrosine, alanine, histidine and phenylalanine to be key descriptors for classification of oligozoospermic and control cases. The role of citrate metabolite, with the ratio of citrate/lactate and choline/citrate, was earlier reported for the prediction of different forms of infertility [12]. Rather the measurement of the ratio of citrate with other metabolites, the absolute concentration of citrate provides a better picture to determine profound cause of infertility. The role of alanine is already explained above. Effect of catecholamines precursors; tyrosine, histidine and phenylalanine, on infertility are explained above. Histidine is necessary for purine synthesis and tyrosine formed by hydroxylation of phenylalanine. Interestingly, phenylalanine is the precursor of dopamine, which has been used successfully as a therapeutic agent for oligozoospermia [24]. The outcome of our study and the irrefutable results from the literature reveal that these amino acids can have either a direct or indirect impact on infertility. The area under the ROC curves (Fig. 2E and F and Table 4) of discriminant predicted probability using NMR variables is superior if compared to clinical variables.

In order to check the accuracy of the prediction of this model using the random numbers of Fisher and Yates method, the prediction possibility of the classifications, based on the discriminant functions obtained by NMR variables, was assessed with a 75/25 data split, by using 75% of the data in each category as training sets and the remaining 25% as test sets. The classifications of 84.7, 86.7, 85.9, and 84.8% were obtained for the test set groups containing CZ vs. NZ + OZ, CZ vs. NZ, CZ vs. OZ, and NZ vs. OZ, respectively, followed by classifications of 92.4, 95.2, 96.0, and 92.9% for the training set groups, respectively. Similarly, the clinical variables reveal the classifications of 84.4, 85.6, 86.9, and 85% for the test and 94.1, 94.6, 92.2 and 92.6% for the training set groups. This evaluation supported that the prediction accuracy of the NMR variables is more or less the same with results obtained through clinical variables.

The advantage of measuring the amino acids – alanine, histidine, phenylalanine, and tyrosine by NMR spectroscopy is that they can be quantified specifically and separately with well-resolved spectra. By other means, such as calorimetric method, either aromatic amino acids or others D-amino acids cannot be quantified individually because all aromatic amino acids or other D-amino acids are assessed simultaneously with color formation reaction. Moreover, other methods are tedious with a small but significant probability of error. NMR spectroscopy is a rapid, non-invasive approach that, once established, can be used to significantly improve the therapeutic interventions of infertile patients. The NMR technique may add to the diagnostic accuracy in the differentiation of pathospermias and contribute to the understanding of fertility and sub-fertility caused at a molecular level.

In essence, because alterations of inner metabolites commonly precede the onset of infertility, NMR spectroscopic semen metabolic profile may contribute as a surrogate technique to routine clinical methods and it may serve as the preeminent diagnostic method for the differential diagnosis of infertility and the improvement of its clinical management. Furthermore, the application of semen metabolic profiles in infertility diagnosis is limited by complexity of dataset, the development of multivariate statistical method seems to be very crucial to solve this problem. Adaption of this technique may open new windows for the differential diagnosis of infertility that can be used for making decisions before the recommendation of an assisted reproductive technology as well as for the new drug development because it is clear that NMR spectroscopy has the requisite sensitivity and specificity to detect subtle changes at clinically relevant means of measuring concentrations of the endogenous metabolites.

Acknowledgements

This work was supported by the Indian Council of Medical Research, New Delhi, India (Grant no. 5/10/8/2004-RHN) and the Central Council for Research in Unani Medicine, New Delhi, India (3-94/2005-CCRUM/Tech). This study was conducted at Centre of Biomedical Magnetic Resonance, Lucknow, India.

The authors thank Dr. Saifuddin Ahmed, Department of Population, Family and Reproductive Health, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA, for critical discussion, suggestions, and statistical analysis.

The authors thank Navneeta Bansal and Brian Houston for her and his assistance in preparing this manuscript, figures and tables.

References

[1] Reproductive Health Outlook, Infertility: overview and lessons learned, available: http://www.rho.org, 2002.
[2] L.W. Chamley, G.N. Clarke, Antisperm antibodies and conception, Semin. Immunopathol. 29 (2007) 169–184.
[3] F. Dondero, A. Lenzi, L. Gandini, F. Lombardo, Immunological infertility in humans, Exp. Clin. Immunogenet. 10 (1993) 65–72.
[4] D.S. Irvine, Epidemiology and aetiology of male infertility, Hum. Reprod. 13 (1998) 33–44.
[5] D.S. Guzick, J.W. Overstreet, P.F. Litvak, C.K. Brazil, S.T. Nakajima, C. Coutifaris, S.A. Carson, P. Cisneros, M.P. Steinkampe, J.A. Hill, D. Xu, D.L. Vogel, Sperm morphology, motility and concentration in fertile and infertile men, N. Engl. J. Med. 345 (2001) 1388–1393.
[6] D.A. Grimes, L.M. Lopez, Oligozoospermia, azoospermia and other semen analysis terminology: the need for better science, Fertil. Steril. 88 (2007) 1491–1494.
[7] R. Menkveld, W.Y. Wong, C.J. Lombard, A.M. Wetzels, C.M. Thomas, H.M. Merkus, R.P. Theunissen, Semen parameters, including WHO and strict criteria morphology, in a fertile and subfertile population: an effort toward standardization of *in vivo* thresholds, Hum. Reprod. 16 (2001) 1165–1171.
[8] W. Ombelet, E. Bosmans, M. Janssen, A. Cox, J. Vlasselaer, W. Gyselaers, H. Vandeput, J. Gielen, H. Pollet, M. Maes, O. Steeno, T. Kruger, Semen parameters in a fertile versus subfertile population: a need for change in the interpretation of semen testing, Hum. Reprod. 12 (1997) 987–993.
[9] M.B. Sorensen, I.A. Bergdahl, N.H. Hjollund, J.P. Bonde, M. Stoltenberg, E. Ernst, Zinc, magnesium and calcium in human seminal fluid: relations to other semen parameters and fertility, Mol. Hum. Reprod. 5 (1999) 331–337.
[10] M.A. Tomlins, J.D.P. Foxall, J.M. Lynch, J. Parkinson, R.J. Everett, J.K. Nicholson, High resolution ^{1}H NMR spectroscopic studies on dynamic biochemical processes in incubated human seminal fluid samples, Biochim. Biophys. Acta 1379 (1998) 367–380.
[11] J.D. Meeker, B.L. Godfrey, R. Hauser, Relationships between serum hormone levels and semen quality among men from an infertility clinic, J. Androl. 28 (2007) 397–406.
[12] S. Hamamah, F. Seguin, L. Bujan, C. Barthelemy, R. Mieusset, J. Lansac, Quantification by magnetic resonance spectroscopy of metabolites in seminal plasma able to differentiate different forms of azoospermia, Hum. Reprod. 13 (1998) 132–135.
[13] M.L. Lynch, J. Masters, J.P. Pryor, J.C. Lindon, M. Spraul, P.J.D. Foxall, J.K. Nicholson, Ultra high field NMR spectroscopic studies on human seminal fluid, seminal vesicle and prostatic secretions, J. Pharm. Biomed. Anal. 12 (1994) 5–19.
[14] M.E. Bollard, E.G. Stanley, J.C. Lindon, J.K. Nicholson, E. Holmes, NMR-based metabonomic approaches for evaluating physiological influences on biofluids composition, NMR Biomed. 18 (2005) 143–162.
[15] A. Gupta, M. Dwivedi, A.A. Mahdi, G.A.N. Gowda, C.L. Khetrapal, M. Bhandari, ^{1}H NMR spectroscopy for identifying and quantifying common uropathogens: a metabolic approach to the urinary tract infection, Brit. J. Uro. Int. 104 (2009) 236–244.
[16] P. Ranjan, A. Gupta, S. Kumar, G.A.N. Gowda, A. Ranjan, A.A. Sonkar, Detection of new amino acid marker of liver injury by proton nuclear magnetic resonance spectroscopy, Liver Int. 26 (2006) 703–707.
[17] U. Sharma, K. Chaudhury, N.R. Jagannathan, S.K. Guha, A proton NMR study of the effect of a new intravasal injectable male contraceptive RISUG on seminal plasma metabolites, Reproduction 122 (2001) 431–436.
[18] World health organization, WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction, 4th ed., Cambridge Univ. Press, Cambridge, UK, 1999.
[19] P.A. Lachenbruch, An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis, Biometrics 23 (1967) 639–645.
[20] D.A. Maher, P. Patki, J.C. Lindon, J.E. Want, E. Holmes, M. Craggs, J.K. Nicholson, Seminal oligouridinosis: low uridine secretion as a biomarker for infertility in spinal neurotrauma, Clin. Chem. 54 (2008) 2063–2066.
[21] L.L. Gershbein, D.R. Thielen, Enzymatic and electrolytic profiles of human semen, Prostate 12 (1988) 263–269.
[22] P. Humphries, E. Pretorius, H. Naude, Direct and indirect cellular effects of aspartame on the brain, Eur. J. Clin. Nutr. 62 (2008) 451–462.
[23] E. Pap, K. Racz, K.J. Kovacs, I. Varga, E. Buzas, B. Madarasz, A. Foldes, C. Szalai, T. Watanabe, H. Ohtsu, A. Ichikawa, A. Nagy, A. Falus, Histidine decarboxylase deficiency in gene knockout mice elevates male sex steroid production, J. Endocr. 175 (2002) 193–199.
[24] J.C. Lavieri, A.A. Pierini, L-Dopa and oligozoospermia, Andrologia 10 (1978) 74–79.
[25] F. Giulianoa, J. Allardb, Dopamine and male sexual function, Eur. Urol. 40 (2001) 601–608.
[26] M. Albrecht, R. Ramsch, F.M. Kohn, J.U. Schwarzer, A. Mayerhofer, Isolation and cultivation of human testicular peritubular cells: a new model for the investigation of fibrotic processes in the human testis and male infertility, J. Clin. Endocr. Metab. 91 (2006) 1956–1960.
[27] F. Safina, S. Tanaka, M. Inagaki, K. Tsuboi, Y. Sugimoto, A. Ichikawa, Expression of L-Histidine decarboxylase in mouse male germ cells, J. Biol. Chem. 277 (2002) 14211–14215.