Prediction of sperm retrieval in men with non-obstructive azoospermia using artificial neural networks: leptin is a good assistant diagnostic marker

Yi Ma1, Bin Chen2,*, HongXiang Wang2, Kai Hu2, and YiRan Huang2

1Department of Science and Study, Renji Hospital, Shanghai JiaoTong University School of Medicine, 7th floor, Building No.1, No 145 Middle Shandong Road, Shanghai, China
2Department of Urology, Renji Hospital, Shanghai JiaoTong University School of Medicine, 7th floor, Building No.1, No 145 Middle Shandong Road, Shanghai, China

*Correspondence address. Tel: +86-21-53882063; Fax: +86-21-63730455; E-mail: dr_binchenn@126.com

Submitted on June 29, 2010; resubmitted on October 31, 2010; accepted on November 10, 2010

BACKGROUND: At present, non-invasive methods are not comprehensive enough to enable urologists to predict sperm retrieval results accurately in patients with non-obstructive azoospermia (NOA). Our aim was to improve the prediction accuracy of sperm retrieval by using leptin and artificial neural networks (ANNs).

METHODS: Data from May 2004 to July 2010 for 280 patients with NOA were reviewed and assigned into the training and testing set for ANNs. All patients underwent standard diagnostic infertility evaluation and testicular sperm extraction (TESE). Twelve factors were recorded as the input variables for ANNs: testicular volume, semen volume, seminal pH, seminal alpha-glucosidase and fructose, serum hormones including FSH, LH, total testosterone (TT), prolactin, estradiol, serum and seminal leptin. Three ANN models were constructed with the following input variables: ANN1-1234, ANN2-123 and ANN3-124. The prediction accuracy for FSH, leptin and ANN models was compared by receiver operating characteristic (ROC) curve analysis.

RESULTS: All ANN models were better than FSH. ANN1 had the largest area under the curve (AUC = 0.83) and demonstrated significant improvement compared with FSH (AUC = 0.63, P < 0.01) and leptin (AUC = 0.59, P < 0.01).

CONCLUSIONS: ANNs improve the prediction accuracy of sperm retrieval. Although the leptin AUC is low, combined use of leptin and FSH can significantly improve the prediction accuracy for sperm recovery in NOA patients. Leptin may be a good assistant marker for diagnosing NOA. However, studies with larger numbers of patients are required to confirm the improved predictive performance of ANNs.

Key words: artificial neural networks / leptin / azoospermia / prediction / testicular sperm extraction

Introduction

Since ICSI was introduced in 1992 (Palermo et al., 1992), testicular sperm extraction (TESE), microsurgical epididymal sperm aspiration or percutaneous epididymal sperm aspiration (MESA/PESA) and some other surgical procedures have created more opportunities for enhancing fertility in patients with azoospermia (Devroey et al., 1994, 1995; Silber et al., 1996). However, complications such as vascularization and fibrosis (Schlegel and Su, 1997) also raised the emotional and financial burden on the patients (Tournaye et al., 1996). To find solutions for this problem, some non-invasive methods, including assessment of testicular volumes, FSH, testicular size, inhibin B and other markers, were used in the prediction of sperm retrieval results, but so far none has proved to be completely reliable (Jezek et al., 1998; Vernaee et al., 2002) and the estimation of success for ICSI remains elusive. Therefore, it is of great importance to develop a more accurate method of predicting TESE outcome, which will help urologists to select patients who are may benefit from TESE or donor insemination procedures.

Leptin, a 167-amino acid protein and the product of the obese gene, has a number of roles linked to the regulation of reproduction. Several studies have reported that leptin was related to spermatogenesis (Banks et al., 1999; Ishikawa et al., 2007). Leptin played an important role in controlling gonadotrophin (FSH, LH) (Yu et al., 1997) and gonadal hormone secretion (total and free TT) (Mah and Wittert, 2010). In addition, leptin might direct the spermatocytes to full development to spermatids (El-Hefnawy et al., 2000). Our recent study also found in rat testis that varicocele-related spermatogenic dysfunction...
was related to an increased expression of leptin and its receptor (Chen et al., 2009).

FSH, testis volume and other markers were not completely accurate for diagnosing non-obstructive azoospermia (NOA), while leptin might reflect spermatogenesis in another way. So what if we combine them? In this study, we examined leptin as an assistant predictive marker in patients with NOA and compared several models using a new multivariate statistical method: artificial neural networks (ANNs).

Materials and Methods

Patients

We reviewed 280 medical records of men with NOA who underwent TESE between May 2004 and July 2010 in our ambulatory care. This group included those patients showing hypospermatogenesis, complete or incomplete maturation arrest, complete or incomplete germ-cell aplasia (Sertoli cell-only) and tubular sclerosis and atrophy. All patients signed informed consent and underwent a standard diagnostic infertility evaluation, including medical and reproductive history, physical examination, BMI, semen analysis, hormonal status, scrotal duplex ultrasound, genetic testing and TESE. Multiple testicular biopsies were obtained from both testes and wet preparations were made at biopsy to detect spermatozoa in the tissue. Patients with Y chromosome microdeletions (AZFa or AZFb), azoospermia caused by acute parotitis and hypogonadotropic hypogonadism were not included in the study.

Cell-free seminal plasma was obtained after centrifugation (3000g for 10 min) and stored at ≈80°C until the assay. Blood samples were collected between 8 and 10 a.m., centrifuged, and serum was separated and stored at ≈80°C until analysis.

Twelve factors were recorded as the input variables for ANNs: average testicular volume, semen volume, seminal pH, seminal alpha-glucosidase and fructose, serum hormones [FSH, LH, total TT, prolactin, estradiol (E2)], serum and seminal leptin.

Methods

The method of calculation for testicular volume is described in Paltiel et al. (2002). Semen analysis, seminal pH, seminal alpha-glucosidase and fructose were assessed according to World Health Organization recommendations (World Health Organization, 1999). Serum levels of FSH, LH, TT, prolactin and E2 were determined by LIAISON chemiluminescence immunoassay (DiaSorin Deutschland GmbH, Dietzenbach, Germany) and the normal ranges were 1.3–11.8 mIU/l, 2.8–6.8 mIU/l, 2.6–7.4 ng/l, 4.1–18.5 μg/l and 0–56 pg/ml, respectively. Serum and seminal leptin levels were measured using a solid-phase sandwich enzyme-linked immunosorbent assay (RECI Diagnostics GmbH, Inc., Herrenberg, Germany), with a lower detection limit of 0.5 ng/ml and detection range of 0–100 ng/ml. We repeatedly measured leptin in the serum and seminal specimens from 28 (10%) patients, with n = 5 for assessing both inter- and intra-assay variation. The inter-assay and intra-assay coefficients of variation were 6.32 and 6.11%, respectively.

ANN

One hundred and thirty four patients were assigned to a training set, while 146 patients were assigned to a testing set. ANN models were constructed with the MATLAB Neural Network Toolbox (The Mathworks, Natick, MA, USA). The training method was Fuzzy C means. Different back-propagation ANN models with inclusion of 12 input parameters ([1]testicular volume, [2]semen volume, seminal pH, seminal alpha-glucosidase and fructose, [3]serum hormone (FSH, LH, TT, prolactin, E2), [4]serum and seminal leptin] were compared. Three different ANN models were constructed and named as: ANN1: (1)+(2)+(3)+(4); ANN2: (1)+(2)+(3); ANN3: (1)+(2)+(4). The results of TESE were adopted as the gold standard to test the ANN output.

The backbone of this ANN is the feed forward-back propagation architecture. The network consisted of 1 input layer with the above 12 variables, 1 hidden layer with 7 neurons (confirmed to be optimal, 10 000 epochs), and 1 output layer with 1 neuron representing the output value of the predictor, which is a measure of the probability of spermatozoa presence. We used 1 to represent the presence of spermatozoa and 0 to represent no spermatozoa. In the testing phase, the ANN model output value was transformed into the range 0 to 1. An estimated probability of >0 indicated the presence of spermatozoa, whereas an estimated probability of <0 suggested no spermatozoa in the tests.

Statistical analysis

Data were analysed using the Statistical Analysis System (SAS, version 6.12; SAS Institute, Inc., Cary, NC, USA). Values were given as mean ± SD. Statistical analysis was performed using the t-test or Mann–Whitney U test to detect differences between different groups. The diagnostic accuracy for FSH, leptin and ANN models was evaluated using the receiver operating characteristic (ROC) curves, which were generated by plotting sensitivity versus 1-specificity using MedCalc 8.1.1.0 (MedCalc Software, Mariakerke, Belgium). Area under the curve (AUC) and the best discriminatory value were calculated by the ‘ROC curve analysis’ tool of the software, while AUCs for parameters were compared by ‘comparison of ROC curves’ tool. A value of P < 0.05 was considered to be statistically significant.

Results

Baseline characteristics and laboratory results

The characteristics and mean parameter values of 134 patients in training set and 146 patients in testing set were shown in Table I according to the presence or absence of spermatozoa. In training set, the sperm negative group differed significantly in FSH (P = 0.02) and seminal leptin (P = 0.04) compared with the sperm positive group. While in the testing set, only the average testicular volume was significantly lower in the sperm negative group (P = 0.04).

Overall, testicular sperm were successfully recovered in 110 of the 280 patients (39.3%). No sperm were found in 170 patients (60.7%). Sperm retrieval rates were 38.8 and 39.7% in the training and testing set, respectively.

ANN and ROC analysis

ROC curves for two parameters (FSH, seminal leptin) and three different ANN models were made to check the diagnostic usefulness of these variables in their effects on differentiating between sperm negative and sperm positive groups (Table II, Fig. 1).

Data from Table II indicated that all the ANN models were better than FSH or seminal leptin, although some differences were not significant. In addition to that, ANN1 had the highest AUC and demonstrated a significantly better improvement (over FSH or seminal leptin) than ANN2 (P = 0.04), ANN3 (P = 0.01), FSH (P < 0.01) and seminal leptin (P < 0.01). ANN2 had the second largest AUC,
and comparisons showed only leptin was significantly lower than ANN2 ($P = 0.02$).

The AUCs for FSH and seminal leptin were 0.62 and 0.59, respectively, with the best discriminative concentration of 14.32 mIU/l (sensitivity 70.7%, specificity 68.2%) and 2.9 ng/ml (sensitivity 43.1%, specificity 75.0%), respectively. Seminal leptin had a smaller AUC than FSH in comparisons. Furthermore, ANN3 with leptin as input variable also had a smaller AUC than ANN2, with FSH. However, ANN1 with both FSH and leptin had the highest AUC, and ANN3 with leptin as input variable improved its performance compared with FSH.

### Table I Clinical characteristics, laboratory data and examination findings for 280 patients with non-obstructive azoospermia randomly assigned to training and testing groups according to spermatozoa on testicular sperm extraction.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Spermatozoa (training set)</th>
<th>P-value</th>
<th>Spermatozoa (testing set)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative</td>
<td>Positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>82</td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>33.6 ± 12.1</td>
<td>34.8 ± 11.4</td>
<td>0.57</td>
<td>30.4 ± 18.6</td>
</tr>
<tr>
<td>Infertility duration (years)</td>
<td>4.1 ± 3.4</td>
<td>3.2 ± 5.1</td>
<td>0.22</td>
<td>4.8 ± 3.8</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>22.5 ± 5.2</td>
<td>23.8 ± 4.7</td>
<td>0.15</td>
<td>24.1 ± 5.5</td>
</tr>
<tr>
<td>Average testicular volume (ml)</td>
<td>11.8 ± 8.6</td>
<td>14.5 ± 7.1</td>
<td>0.06</td>
<td>11.8 ± 5.2</td>
</tr>
<tr>
<td>Semen volume (ml)</td>
<td>2.8 ± 1.0</td>
<td>3.0 ± 1.7</td>
<td>0.39</td>
<td>2.3 ± 1.8</td>
</tr>
<tr>
<td>Seminal pH</td>
<td>7.1 ± 1.2</td>
<td>6.9 ± 1.1</td>
<td>0.33</td>
<td>7.5 ± 1.9</td>
</tr>
<tr>
<td>alpha-glucosidase (U/ml)</td>
<td>48.7 ± 30.1</td>
<td>39.4 ± 45.5</td>
<td>0.16</td>
<td>34.8 ± 38.4</td>
</tr>
<tr>
<td>Fructose (µmol)</td>
<td>35.6 ± 50.3</td>
<td>30.3 ± 42.6</td>
<td>0.53</td>
<td>35.8 ± 43.5</td>
</tr>
<tr>
<td>FSH (mIU/l)</td>
<td>16.2 ± 5.8</td>
<td>13.7 ± 6.8</td>
<td>0.02</td>
<td>15.68 ± 6.1</td>
</tr>
<tr>
<td>LH (mIU/l)</td>
<td>5.8 ± 3.5</td>
<td>6.8 ± 4.9</td>
<td>0.17</td>
<td>6.8 ± 6.2</td>
</tr>
<tr>
<td>TT (ng/l)</td>
<td>6.3 ± 3.7</td>
<td>6.9 ± 3.4</td>
<td>0.53</td>
<td>5.2 ± 5.5</td>
</tr>
<tr>
<td>Prolactin (ng/ml)</td>
<td>12.5 ± 19.8</td>
<td>14.1 ± 12.2</td>
<td>0.60</td>
<td>16.1 ± 13.2</td>
</tr>
<tr>
<td>Estradiol (ng/ml)</td>
<td>38.8 ± 30.0</td>
<td>34.9 ± 29.4</td>
<td>0.46</td>
<td>30.5 ± 19.5</td>
</tr>
<tr>
<td>Serum leptin (ng/ml)</td>
<td>7.4 ± 3.1</td>
<td>6.7 ± 4.2</td>
<td>0.27</td>
<td>7.2 ± 4.0</td>
</tr>
<tr>
<td>Seminal leptin (ng/ml)</td>
<td>3.4 ± 2.0</td>
<td>2.5 ± 3.2</td>
<td>0.04</td>
<td>3.3 ± 2.3</td>
</tr>
</tbody>
</table>

$t$-test or Mann–Whitney $U$ test were used to detect differences. TT, total testosterone.

### Table II Analysis of receiver operating characteristic curves for FSH, seminal leptin, and three artificial neural network (ANN) models.

<table>
<thead>
<tr>
<th></th>
<th>AUC for non-obstructive azoospermia</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>SE</td>
<td>95% CI</td>
</tr>
<tr>
<td>ANN1</td>
<td>0.832</td>
<td>0.037</td>
<td>0.761–0.888</td>
</tr>
<tr>
<td>ANN2</td>
<td>0.721*</td>
<td>0.044</td>
<td>0.641–0.792</td>
</tr>
<tr>
<td>ANN3</td>
<td>0.693*</td>
<td>0.046</td>
<td>0.611–0.766</td>
</tr>
<tr>
<td>FSH</td>
<td>0.628**</td>
<td>0.046</td>
<td>0.545–0.707</td>
</tr>
<tr>
<td>Seminal leptin</td>
<td>0.590***</td>
<td>0.047</td>
<td>0.505–0.670</td>
</tr>
</tbody>
</table>

AUC, area under curve.  
CI, confidence interval.  
*Significantly lower than ANN1 ($P < 0.05$).  
**Significantly lower than ANN1 ($P < 0.01$).  
***Significantly lower than ANN2 ($P < 0.05$).

**Figure 1** Receiver operating curves for artificial neural network ANN1, ANN2, ANN3, FSH, and seminal leptin concentrations for discriminating successful and failed testicular sperm extraction in non-obstructive azoospermia patients.

### Discussion

In this study, we revealed for the first time that leptin might be a good assistant marker to increase the prediction accuracy for NOA using
ANNs. The combined use of leptin and some other routine markers can significantly improve the prediction accuracy of sperm retrieval in NOA patients. However, in this study the small number of NOA patients will have limited the ability of ANNs to improve performance. Furthermore, leptin is a multifunctional hormone and the exact role(s) of leptin in male spermatogenesis is still elusive.

Recently, a number of studies have focused on predicting the presence of spermatozoa in the testis using non-invasive methods (Isikoglu et al., 2006; Tunc et al., 2006; Nistal et al., 2007; Gouli et al., 2009); however, there is still a lack of an accurate method. In this study, we tried to improve the prediction accuracy of sperm retrieval using ANNs. Recently leptin has been confirmed as playing an important role in human reproduction. It was reported that leptin and its receptor were over-expressed in the testis of patients with Sertoli cell-only syndrome (Ishikawa et al., 2007) and leptin levels were negatively related to the levels of inhibin B (Banks et al., 1999). In addition, it was supposed that leptin might have different effects on differentiation of germ cells at different stages: leptin directed the spermatocytes to full development to spermatids (El-Hefnawy et al., 2000). These data led us to propose that leptin probably affects spermatogenesis in some way. In order to confirm the hypothesis, we carried out a study and indeed, we found that the expression of leptin and leptin receptor in testis was inversely correlated with spermatogenic function in rats with varicocele (Chen et al., 2009). We conducted this study to further explore the use of leptin as a marker in a clinical application.

Our recent study (Chen et al., 2009) showed that the seminal leptin level was significantly higher in patients with varicocele (low spermatogenesis) compared with that of the controls; with regard to serum leptin, at comparable BMI, although leptin concentration in patients is higher than controls, the difference is not significant. Besides, it has been reported that there was no correlation between serum leptin level and a history of varicocele (Zom et al., 2007). It was postulated that serum leptin increases only in patients with extremely poor spermatogenic function. In this study, we confirmed this hypothesis in NOA patients. Although the diagnostic accuracy for leptin was lower than FSH (AUCs: 0.590 versus 0.628), combined use of leptin and FSH could significantly improve the prediction accuracy of sperm recovery. Leptin is therefore a good assistant diagnostic marker, and a marker for differentiating between sperm negative and sperm positive patients.

Leptin levels are related to body composition, and it is secreted possibly by the whole genital tract. Why could leptin be a marker for differentiating sperm negative and sperm positive patients? First, we have excluded the influence of body composition by taking BMI into account. Second, over-expression of leptin and its receptor may occur in hypospermatogenesis tests (Ishikawa et al., 2007). Besides, it is known that reproductive function is affected by leptin through actions at the hypothalamic level (Tena-Sempere et al., 1999, 2001). A small amount of circulating leptin is capable of crossing the blood–testis barrier (Banks et al., 1999), which may also lead to a high level of seminal leptin. These may be the reasons why a sperm negative patient has a higher leptin level than that of a sperm positive patient.

We revealed in this study that ANN models were reliable for predicting the sperm retrieval results. ANN models demonstrated a better performance than FSH and seminal leptin in NOA patients. The combined use of all parameters could improve the prediction accuracy for ANN models. We did not compare ANNs with inhibin-B, because the importance of inhibin-B is becoming more and more controversial now: it has been widely considered that inhibin-B might be not superior to FSH, such as in predicting the presence of sperm in testicular fine needle aspirate (Gouli et al., 2008). Furthermore, inhibin-B is not been firmly established as an independent predictive index for the finding of spermatozoa in TESE (Adamopoulos and Koukkou, 2010), testsis biopsy (Nowroozi et al., 2008) and MESA (Smit et al., 2007). To date, FSH is also the hormone marker that is most widely accepted by andrologists.

An ANN is a type of algorithm that can combine several variables and improve prediction accuracy as a whole. Several studies have reported that ANNs are an ideal multivariate statistical method (Stephan et al., 2002; Anderson et al., 2010; Hsieh et al., 2010) and are superior to other predictive models in this area (Samli and Dogan, 2004; el-Mekresh et al., 2009). ANNs are algorithms that offer a number of theoretical advantages, including the ability to detect non-linear relationships implicitly, the ability to combine different variables, and the ability to identify new markers to improve the diagnostic accuracy. In contrast, classical statistical modelling required the explicit assumption of certain relationships, which was also what the validity of the models largely depended upon (Fujikawa et al., 2003).

The data from this study show that leptin is a good choice for use along with the ‘classic’ markers (i.e. FSH, testicular volume) in clinical prediction. Although the precise mechanisms of action of leptin in male reproduction have not yet been fully elucidated, it is convenient marker and easy to detect. ANNs are now being investigated as predictive methods in many areas, such as prediction of bladder outlet obstruction, prostate cancer diagnosis, and bladder cancer. Also, ANNs are being used more and more in clinical decision-making. Clinical institutes could establish their own database for ANN training. Software packages, such as SPSS (Statistical Package for the Social Sciences, SPSS, Inc., Chicago, IL, USA), NeuroSolutions (NeuroDimension, Inc., Gainesville, FL, USA), and MATLAB (The Mathworks, Natick, MA, USA), can provide the ANN module. The advantage of ANN is that the models are used more often in clinical practice, more specific data will be generated which can be used to update the model and improve the training efficiency and prediction accuracy.

Authors’ roles

Y.M. performed experiments, acquired data, and wrote the manuscript; B.C. conceived the project, designed experiments; HX.W. performed experiments and analysed data; K.H. participated in forming experiments and analysed data; K.H. participated in designing experiments; HX.W. performed experiments and writing the manuscript; YR.H. supervised the project and revised manuscript.

Acknowledgements

We thank Yan Jin and Hao Yang for generous donation of the freezing and thawing media.

Funding

This study was supported by a grant of Shanghai Population and Family Planning Commission (No. 2010JG03).
References


