Computational Models for Prediction of Intrauterine Insemination Outcomes

Moshe WALD¹, Amy E.T SPARKS², Bradley J VAN VOORHIS², Craig H SYROP², Criag S NIEDERBERGER³

¹Department of Urology, University of Iowa, Iowa City, Iowa, USA
²Department of Obstetrics and Gynecology, Division of Endocrinology and Infertility, University of Iowa, Iowa City, Iowa, USA
³Department of Urology, University of Illinois at Chicago, Chicago, Illinois, USA

Received 29 December 2006; received in revised form 29 December 2006; accepted 11 January 2007; published online 22 February 2007

Abstract

Objective: Intrauterine insemination (IUI) using ejaculated sperm is a common option in the treatment of infertility of various etiologies. We sought to develop a computational model for the prediction of pregnancy following IUI.

Materials and Methods: A data set of 212 exemplars, derived from patients who underwent a first IUI cycle with ejaculated sperm, was divided into separate modelling and cross-validation sets, and analyzed retrospectively. The data set contained input features of maternal age, type of medication used for ovulation induction, semen volume, sperm concentration, motility and morphology and intra-uterine pregnancy output, and was modelled using various mathematical methods, including linear and radial support vector machines, linear and quadratic discriminant function analysis, logistic regression, and neural computation. Various models were used, in an attempt to achieve the highest model accuracy. A logistic regression model was found to have the highest accuracy, with a test set ROC area of 0.717.

Results: Forward regression of this model showed sperm morphology to be the most significant feature in predicting pregnancy \(p=0.39\), followed by maternal age \(p=0.42\), type of medication used for ovulation induction \(p=0.6\), sperm motility \(p=0.61\), semen volume \(p=0.71\) and sperm concentration \(p=0.9\). Reverse regression of the model revealed sperm motility to be the most significant feature in predicting pregnancy \(p=0.37\), followed by sperm morphology \(p=0.39\), maternal age \(p=0.49\), type of medication used for ovulation induction \(p=0.61\), sperm concentration \(p=0.72\) and semen volume \(p=0.74\).

Discussion: A logistic regression model of clinical relevance was developed, and is deployed on the World Wide Web for clinical use.

Keywords: maternal age, ovulation induction, sperm concentration, motility, morphology, IUI, computational models

Özet

Amaç: Farklı etiyolojilerin infertilite tedavisinde ejaküle edilmiş sperm kullanarak intrauterin inseminasyon (IUI) yaygın kullanılan bir seçenektir. Bu çalışmamızda amacımız IUI sonrası gebelik oluflarının tahmini için bilgisayar destekli bir model geliştirmektir.

Materyal ve Metot: Ejakül edilmiş sperm, was divided into separate modelling and cross-validation sets, and analyzed retrospectively. The data set contained input features of maternal age, type of medication used for ovulation induction, semen volume, sperm concentration, motility and morphology and intra-uterine pregnancy output, and was modelled using various mathematical methods, including linear and radial support vector machines, linear and quadratic discriminant function analysis, logistic regression, and neural computation. Various models were used, in an attempt to achieve the highest model accuracy. A logistic regression model was found to have the highest accuracy, with a test set ROC area of 0.717.

Results: Forward regression of this model showed sperm morphology to be the most significant feature in predicting pregnancy \(p=0.39\), followed by maternal age \(p=0.42\), type of medication used for ovulation induction \(p=0.6\), sperm motility \(p=0.61\), semen volume \(p=0.71\) and sperm concentration \(p=0.9\). Reverse regression of the model revealed sperm motility to be the most significant feature in predicting pregnancy \(p=0.37\), followed by sperm morphology \(p=0.39\), maternal age \(p=0.49\), type of medication used for ovulation induction \(p=0.61\), sperm concentration \(p=0.72\) and semen volume \(p=0.74\).

Discussion: A logistic regression model of clinical relevance was developed, and is deployed on the World Wide Web for clinical use.

Keywords: maternal age, ovulation induction, sperm concentration, motility, morphology, IUI, computational models

 Anahtar sözcükler: maternal age, ovulasyon indüksiyonu, sperm konsantrasyonu, motilite, morfoloji, intrauterin inseminasyon (IUI), bilgisayar destekli modeller

Corresponding Author: Dr. Moshe Wald
200 Hawkins Drive, 3 RCP 52242 Iowa City, Iowa, USA
Phone : +01 319 356 89 22
E-mail : moshe-wald@uiowa.edu
Introduction

Intrauterine inseminations (IUI) with partner’s spermatozoa is commonly used as first-line treatment for subfertile couples with ovulatory disorders, cervical or male factor infertility, and unexplained infertility (1-2). While cervical female infertility is considered to have the best prognosis with IUI (3), controversy still remains regarding the outcomes of this procedure for male and unexplained infertility (1-3).

Several published studies have investigated the prognostic importance of various factors with regard to IUI success (1,4-7). Interestingly, the findings of these studies were partially conflicting. For example, while sperm motility was reported in one study to be one of the four most predictive factors of IUI (with follicle number, endometrial thickness, and duration of infertility) (1), another study has suggested that sperm morphology rather than motility is a more sensitive guide to IUI outcome (5). A meta-analysis performed to assess the clinical value of the postwash total motile sperm count (TMC) as a test to predict IUI outcome concluded that postwash TMC at insemination could potentially be used in counseling patients for either IUI or IVF (7). The different studies also did not evaluate the same parameters. Some of these previous studies have used a single method for ovulation induction prior to IUI, and most did not include the type of ovulation induction as one of the investigated parameters. In addition, previous studies have not compared the accuracy of different models in predicting IUI outcomes. In this study, we have aimed to assess the importance of selected key parameters for the achievement of intrauterine pregnancy through IUI by developing computational models to investigate the relationship of maternal age, type of medication used for ovulation induction and various semen parameters (sperm head shapes) with intrauterine pregnancy achieved by IUI with high goodness-of-fit.

In previous data sets, we had developed computational models that addressed various reproductive clinical problems. We developed separate models using neural computation, a mathematical method that simulates the physiology of the biologic neuron (8), to predict intra-uterine pregnancies achieved by IVF/ICSI, the presence of prostate cancer and the existence of erectile dysfunction, given a set of relevant input features (9-11). We deployed these models using the javascript language for ready availability on the World Wide Web and in PalmOS for physicians using handheld computers. The main goal of this study was to develop a computational model that could serve as a useful tool for clinicians who are considering IUI for treatment of subfertile couples.

Materials and Methods

Upon Institutional Review Board (IRB) approval was not obtained for this study because data were collected via retrospective chart review without disclosure of patient identification. In addition, the study analyzes a well-established clinical therapeutic procedure that is not experimental and is not under IRB guidance. Clinical data was collected from the first IUI cycle of 212 women (age range 23-45-years, mean 32.2-years), done primarily for unilateral tubal (n=24), ovulatory (n=70), and unexplained (n=84) factors (Table 1). Input features used for this study included maternal age, type of medication used for ovulation induction prior to the IUI cycle (Table 2), and four semen parameters recorded on the day of the insemination procedure: semen volume, sperm concentration (million/ml), motility (%) and morphology (% normal sperm head shapes). Intra-uterine pregnancies achieved through the first IUI cycle were defined as outcomes. The various medications used for ovulation induction prior to the IUI cycle and intrauterine pregnancies achieved through the first IUI cycle were assigned either 1, if present for each exemplar, or 0, if not. Two hundred and twelve intra-uterine pregnancy scores as output and corresponding input variates of maternal age, type of medication used for ovulation induction prior to the IUI cycle, as well as semen volume, sperm concentration, motility and morphology on the day of insemination, comprised the dataset for computational modeling using neUROn++, a set of C++ programs we developed using the Cygwin (Red Hat) GNU C++ port for Windows (Microsoft) distributed across Pentium (Intel) platforms.

We used neUROn++ to model the data set using the linear mathematical modeling methods linear and quadratic discriminant function analysis (LDFA, QDFA) and logistic regression (LR), and the nonlinear method of neural computation (NNET). Other modeling methods included linear and radial support vector machines (LSVM, RSVM). The data set was randomized into a larger modeling “training” set and a smaller, separate, completely independent cross-validation, “test” set.

### Table 1. Infertility factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexplained</td>
<td>84 (39.6)</td>
</tr>
<tr>
<td>Ovulatory dysfunction</td>
<td>70 (33)</td>
</tr>
<tr>
<td>(ovulating on medications)</td>
<td></td>
</tr>
<tr>
<td>Tubal</td>
<td>24 (11.3)</td>
</tr>
<tr>
<td>Endometriosis (mild)</td>
<td>13 (6.1)</td>
</tr>
<tr>
<td>Uterine</td>
<td>9 (4.2)</td>
</tr>
<tr>
<td>Recurrent loss</td>
<td>5 (2.4)</td>
</tr>
<tr>
<td>Cervical</td>
<td>4 (1.9)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (1.4)</td>
</tr>
</tbody>
</table>

### Table 2. Medications used for hormonal stimulation prior to IUI

<table>
<thead>
<tr>
<th>Hormonal stimulation method</th>
<th>Number of IUI cycles (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural cycle</td>
<td>5 (2.36)</td>
</tr>
<tr>
<td>Clomiphene/tamoxifen</td>
<td>71 (33.49)</td>
</tr>
<tr>
<td>Letrozole/anastrozole</td>
<td>33 (15.57)</td>
</tr>
<tr>
<td>Clomiphene/tamoxifen with hMG injection</td>
<td>42 (19.81)</td>
</tr>
<tr>
<td>Letrozole/anastrozole with hMG injection</td>
<td>56 (26.41)</td>
</tr>
<tr>
<td>hMG injection only</td>
<td>5 (2.36)</td>
</tr>
</tbody>
</table>
Model training was considered to be completed when the error was observed to be oscillating at a local error minimum. We investigated different architectures with varying hidden nodes and varying numbers of exemplars in the training and test sets, to identify the model with the greatest goodness-of-fit. Receiver operator characteristic curve (ROC) area served to assess the model’s accuracy, and was computed using the statistical method described by Wickens (12) and by the traditional trapezoidal method. The best ROC area achieved for each model was then recorded (Table 3). Of note, for the Wickens statistical method a smaller $p$-value is worse (closer to 1 is more linear) in goodness-of-fit. We used Wilk’s Generalized Likelihood Ratio test to determine which input features were significant to the model’s outcome in a forward regression analysis (13).

<table>
<thead>
<tr>
<th>Method</th>
<th>ROC area training set</th>
<th>ROC area test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>0.627</td>
<td>0.717</td>
</tr>
<tr>
<td>LSVM</td>
<td>0.552</td>
<td>0.677</td>
</tr>
<tr>
<td>RSVM</td>
<td>0.679</td>
<td>0.653</td>
</tr>
<tr>
<td>4 HN Neural network</td>
<td>0.923</td>
<td>0.625</td>
</tr>
<tr>
<td>QDFA</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>LDFA</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

ROC: receiver operator characteristic curve  
LDFA: linear discriminant function analysis  
QDFA: quadratic discriminant function analysis  
HN: hidden nodes  
LSVM: linear support vector machine  
RSVM: radial support vector machine

We deployed the model in the javascript language for ready availability on the World Wide Web (Figure 1), and in PalmOS for physicians using handheld computers. The clinician enters the patient’s data using the forms-based interface, and the model reports the odds ratio for intra-uterine pregnancy. Currently, the model may be found at www.urocomp.net

**Semen analysis, sperm preparation and insemination procedure**

All couples were requested to abstain from intercourse for 2-7 days before IUI. Semen samples were produced by masturbation into sterile containers. Volume of the ejaculate was recorded. Sperm concentration and motility were determined by analyzing 5 micro-liters of semen on a Standard Count Chamber (Spectrum Technologies; Heraldsburg, CA, USA). Sperm morphology was determined by Tygerberg strict criteria (14). All inseminates were prepared by centrifuging the ejaculated semen through a discontinuous saline-coated silica density gradient, as previously described (15). Inseminations were performed with the use of a Wallace insemination catheter (SIMS Portex Ltd., Hythe, Kent, U.K.).

**Ovulation induction and monitoring**

All but five IUI cycles were performed in conjunction with some form of ovulation induction. Our typical practice is to start with oral ovulation induction agents, including clomiphene citrate, tamoxifen and letrozole, given on cycle days 3-7. The choice of the oral medication was arbitrary, and made at the discretion of the attending physician. Clomiphene citrate was typically prescribed at a dose of 100 mg/d, tamoxifen at 40 mg/d and letrozole at 2.5 mg/d. For spontaneous cycles and cycles involving stimulation with oral agents, women performed daily urine LH testing starting on cycle day 12 in the evening and had an IUI in the morning after the LH surge. If a woman did not ovulate in response to oral medications, 1 of 2 “next step” induction protocols were selected, at the discretion of the attending physician and in consultation with the patient.

The hMG protocol consisted of daily hMG injections, which were started on cycle day 3 after baseline ultrasound. hMG injections were most commonly used at a starting dose of
two ampules a day (150 IU/d), with adjustments made based on ovarian response. Alternatively, a combination protocol was used, which consisted of oral ovulation induction agents (clomiphene citrate, tamoxifen or letrozole, given on cycle days 3-7), followed by hMG injections, beginning on cycle day 7, usually at a dose of 1 ampule per day (75 IU/d). For both of the latter protocols, hMG injections were continued until at least one mature ovarian follicle (mean diameter of 18 mm) was seen on ultrasound. Subsequently, 10 000 IU of hCG were administered to induce ovulation, and a single IUI was performed 36 hours later. If 3 or more mature follicles were identified, risks of multiple gestations and the option of multi-fetal reduction were discussed. If the couple was not comfortable with the possibility of multiple gestations and/or would not consider multi-fetal reduction, the IUI cycle was cancelled. Cycle outcome was determined by hCG measurement performed 2 weeks after the insemination. After a positive pregnancy test, ultrasound examinations were scheduled 3 weeks later to confirm fetal viability.

Results

Data obtained from the first IUI cycle of 212 women were included in this study. The women’s mean age was 32.2 years (range 23-45). Fresh ejaculated semen samples were collected, analyzed, processed and inseminated on the day of the IUI procedure in all cycles. Semen analysis parameters regarding the ejaculated semen samples used for IUI are summarized in Table 4. Oral ovulation induction agents were used in 49.06% of all cycles. Ovulation was induced by gonadotropin injections (either alone or combined with oral agents) in 48.58% of the cycles. Natural cycles comprised only 2.36% of all cycles performed. Intra-uterine pregnancies have been achieved in 37 cycles (17.45%).

Randomization of the dataset into a modeling “training” set of 150 exemplars and a cross-validation “test” set of 62 exemplars was found to consistently provide the most representative data sets for the various models that we have investigated. A logistic regression (LR) model was found to predict most accurately the intrauterine pregnancy, with ROC area of 0.717 for the test set. Results of LDFA, QDFA, SVM and neural network are shown in Table 3.

**Table 4. Semen analysis parameters**

<table>
<thead>
<tr>
<th>Semen volume (ml)</th>
<th>Sperm concentration (million/ml)</th>
<th>Sperm motility (%)</th>
<th>Sperm morphology (% strictly normal head shapes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>3.29</td>
<td>63.78</td>
<td>52.82</td>
</tr>
<tr>
<td>Range</td>
<td>0.8-11.8</td>
<td>2.78-285.2</td>
<td>9.8-91.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.13</td>
<td>0-18.5</td>
</tr>
</tbody>
</table>

Discriminant analysis (LDFA, QDFA)

This is a multivariate statistical procedure that mathematically defines a special discriminant function to separate a study population by one classification variable. The discriminant function can use several quantitative variables, each of which makes an independent contribution to the overall discrimination. Taking into consideration the effect of all quantitative variables, this discriminant function produces the statistical decision for guessing to which subgroup of classification variable each subject belongs. Assuming a multivariate normal distribution of quantitative variables within each level of classification variable, a parametric method generates either a linear discriminant function (equal within-class covariance) or a quadratic discriminant function (unequal within-class covariance). In either case, the discriminant function is a weighted combination of all quantitative variables. The performance of discriminant analysis can be evaluated by estimating the error rate (probability of misclassification).

Wilk’s Generalized Likelihood Ratio test (GLRT) was used to assess the significance of the neural network’s input parameters. This statistical test was shown to be useful for deciding which of several subsets of artificial neural network system architectures is most appropriate for a certain statistical environment, given that the full model provides a good fit to the observed data. In forward regression, subnetworks are designed by starting with a baseline network with no input nodes, and adding input variables of the original model according to a predetermined algorithm. The subnetworks so generated are trained to completion and their errors recorded. The GLRT provides a formula by which the fitness of a reduced model to the data can be compared to that of the full model, at a certain significance level (13).

Statistical analysis of neural computational models is based on the observation that two fully-trained neural networks with final errors $E(y_{\text{reduced}})$ and $E(y_{\text{full}})$ that are sufficiently close may be discriminated by a $\chi^2$ probability distribution. In stepwise regression, the two networks $y_{\text{reduced}}$ and $y_{\text{full}}$ are chosen to differ by one input variable. Specifically, stepwise forward regression begins with training a baseline network with no input nodes, and recording that network’s final error. Input variables are added one at a time, the subnetworks so generated are trained to completion and their errors recorded, and the variable with the smallest $p$-value (most significant) is then recorded with its $p$-value. Using the final error of the network which is the one node with the smallest $p$-value as a starting point, input variables are then added two at a time, with each pair containing the one variable with the smallest $p$-value and one of the remaining input variables. The pair of variables with the smallest $p$-value (most significant) is then recorded with its $p$-value, and the process repeats until a chosen threshold $p$-value is reached.
Forward regression of the LR model based on Wilk’s GLRT revealed sperm morphology to be the most significant feature in predicting pregnancy ($p=0.39$), followed by maternal age ($p=0.42$), type of medication used for ovulation induction ($p=0.6$), sperm motility ($p=0.61$), semen volume ($p=0.71$) and sperm concentration ($p=0.9$). Reverse regression of the model showed sperm motility to be the most significant feature in predicting pregnancy ($p=0.37$), followed by sperm morphology ($p=0.39$), maternal age ($p=0.49$), type of medication used for ovulation induction ($0.61$), sperm concentration ($p=0.72$) and semen volume ($p=0.74$). This analysis supports the presence of redundant or surrogate variates.

**Discussion**

The beneficial effects of IUI in cases of male and unexplained infertility are still controversial (1-3). As such, the identification of clinical and laboratory prognostic factors that could predict the chances of success for IUI in these clinical circumstances would be helpful for appropriate counseling of these couples and designing the best treatment plan for them.

Several studies were conducted to investigate the prognostic importance of various factors with regard to IUI success (1,4-7,15). However, these studies differ in their design and results. The investigated input parameters vary among the different studies, to include male and female age, duration of infertility, number of follicles (>18 mm diameter), endometrial thickness, sperm concentration, motility and ideal forms, in different combinations. Interestingly, the type of medication used for ovulation induction prior to IUI was included as an investigated parameter in only one of these studies (15). Additionally, none of these previous studies included a comparison of various modelling methods, in an attempt to identify the most accurate model for a certain group of investigated parameters. In fact, all studies have used a single modelling method. Comparison of the previous studies also revealed discrepancy regarding the reported prognostic importance of certain parameters. While sperm motility was reported in one study to be one of the four most predictive factors for IUI success (1), another study has suggested that sperm morphology rather than motility is a more sensitive guide to IUI outcome (5).

Our study comes to address these problems, in an attempt to determine the individual prognostic importance of certain key clinical and laboratory parameters regarding the success rate of IUI. We further sought to develop a user-friendly and accurate computational model, which can be used by clinicians for deciding whether IUI should be used in a given clinical setting.

The data set used in our study was derived from the first IUI cycles of 212 different couples, as the highest pregnancy rates have been reported in the first cycle of IUI (5,16-18), with significant decrease in the second and third cycles (5). In fact, it has been suggested that inclusion of additional IUI cycles in the data analysis in studies of IUI outcomes is a probable cause for selection, because not all patients undergo the same number of cycles (19). Input features were carefully selected, to include a balanced combination of male and female factors, and to consist of basic parameters that are usually available to the clinician in the office at the time of initial evaluation of the infertile couple, excluding parameters that require more advanced studies. Thus, we have used the data obtained from ejaculated semen, and not from the processed samples, as the latter would not be available at the time of the initial consultation. While many studies report on the effects of single variables on the success rate of IUI, such as the influence of age (19-21), our study integrates a number of key input features to create an accurate model that can be used to predict the outcomes of future IUI cycles. Moreover, further statistical analysis of our model allowed for the determination of the significance of the individual input features to the model. Forward regression of our model showed sperm morphology to be the most significant feature in predicting pregnancy ($p=0.39$). Interestingly, reverse regression of the model revealed a slightly better $p$-value for sperm motility ($p=0.37$), followed closely by sperm morphology ($p=0.39$). These findings are in concordance with the results of other studies, which suggested sperm motility (1) and morphology (5) to be sensitive guides to IUI outcome.

Interestingly, forward and reverse regression of our model showed that sperm parameters, semen volume and maternal age were not found to be statistically significant predictors of outcome. Some authors have reported that the number of motile spermatozoa influences the chance of pregnancy achieved through IUI (19,22). However, while the number of inseminated spermatozoa was reported by Van Der Westerlaken et al. (19) to significantly affect the pregnancy rate (<2 million, 4.6%; ≥2 to <10 million, 3.9%; and ≥10 million, 11.3%), stratification of pregnancy rates according to the indication for IUI revealed no statistically significant difference in IUI outcomes between male-related and non-male related indications (that included tubal, hormonal and idiopathic indications). However, it is possible that the smaller effect of sperm parameters on IUI pregnancies demonstrated in our study is related to the relative good quality of the semen samples that were used for insemination, in terms of sperm concentration and motility. In our practice, couples with total motile sperm counts that are lower than 10 million are often counselled against IUI and are guided to IVF more quickly. Thus, the most severe male factor patients may not be included in the data set.

Studies which evaluated the impact of maternal age on pregnancy rate after IUI have reported varying results. Some authors have reported that women 40 years or older did not become pregnant. Interestingly, in women younger than 40 who participated in one study, maternal age did not have a significant effect on IUI outcomes (20). Others found a decrease in pregnancy rate after IUI already at 35 years of age (19,23). Corsan et al. reported that the pregnancy rate per IUI cycle in women 40 years or older dropped to less than half.
that in women younger than 40-years of age (6.69 versus 17.95%) (24). While women 40-years or older enrolled in that study were still able to become pregnant, this ability rapidly declined at 42-years of age.

Computational models have been used to address various clinical problems in different areas of medicine. We have recently developed a neural network to predict intrauterine pregnancies achieved by IVF/ICSI using surgically retrieved sperm (9). This model, shown to have high predictive accuracy, was deployed in the javascript language for ready availability on the World Wide Web and in PalmOS for handheld computers, allowing clinicians to simply enter the input data, consisting of maternal age, sperm retrieval technique, type of sperm used and type of male factor, and receive an immediate outcome prediction. In this study, although the ROC area in the training set was high for a neural computational model, the significant decay in ROC area for the test set in a topology with only two hidden nodes indicates overfitting, and highlights that in certain medical datasets such as the one for IUI modeled in the current study, linear tools such as logistic regression are sufficient. We developed a user-friendly computational tool designed to assist clinicians in the decision making process regarding the option of intrauterine insemination for infertile couples.

References