human reproduction

ORIGINAL ARTICLE *Andrology*

A risk prediction model of sperm retrieval failure with fine needle aspiration in males with non-obstructive azoospermia

Yue Ma[1], Fuping Li[2,3], Li Wang[4], Wenrui Zhao[2,3], Dingming Li[2,3], Yang Xian[2,3], and Xiaohui Jiang[2,3,*]

[1]West China School of Public Health and Healthy Food Evaluation Research Center, Sichuan University, 610041, Chengdu, Sichuan, China
[2]Human Sperm Bank, West China Second University Hospital of Sichuan University, 610041, Chengdu, Sichuan, China [3]Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, 610041, Chengdu, Sichuan, China
[4]Jinjiang Maternity and Child Health Hospital, 610065, Chengdu, Sichuan, China

*Correspondence address. Human Sperm Bank, West China Second University Hospital of Sichuan University, Chengdu, Sichuan, China. E-mail: jxh424@126.com

STUDY QUESTION: Can we predict the risk of sperm retrieval failure among men with non-obstructive azoospermia (NOA) before they undergo fine needle aspiration (FNA)?

SUMMARY ANSWER: Our model, which includes FSH level, age and testicular volume as variables, can predict the risk of sperm retrieval failure with FNA.

WHAT IS KNOWN ALREADY: Combined with ICSI, testicular sperm aspiration (TESA) can enable patients with NOA to have their own genetic offspring. Nearly all reproductive medicine centres in China have applied FNA, but approximately half of patients with NOA experience testicular sperm retrieval failure. Nevertheless, the models developed to predict the likelihood of obtaining spermatozoa with testicular sperm extraction (TESE) cannot accurately predict sperm retrieval, and few of these models have been sufficiently validated.

STUDY DESIGN, SIZE, DURATION: This study involved three cohorts including 597 men with NOA. From 1 January 2015 to 31 July 2017, a retrospective cohort of 317 males with NOA who underwent FNA procedures at a university affiliated hospital were included to build a risk prediction model of sperm retrieval failure with FNA. Then, from 25 October 2017 to 31 March 2018, two prospective cohorts of 61 and 219 males with NOA from the same hospital and one other reproductive specialist hospital respectively, were recruited to validate the risk prediction model.

PARTICIPANTS/MATERIALS, SETTING, METHODS: All men with NOA undergoing their first TESE procedure as part of a fertility treatment were included. The primary end-point was the presence of one or more spermatozoa (regardless of their motility) obtained with FNA. A binary multivariable logistic model was built to predict the risk of sperm retrieval failure after TESA using the dataset from the retrospective cohort. A cut-off value for risk was calculated with receiver operating characteristic (ROC) curve analysis. Two validation sets from the prospective cohort were used to validate the risk prediction model by measures including prediction accuracy and the true positive rate.

MAIN RESULTS AND THE ROLE OF CHANCE: A total of 327 (54.8%) males with NOA experienced sperm retrieval failure with FNA. FSH level, age and testicular volume were included in the prediction model for sperm retrieval failure risk. The model had an AUC of 82.3% (95% CI: 77.6–87.1%) and a cut-off value of 64.61% with a sensitivity of 0.677 and specificity of 0.863 for predicted risk. The predictive accuracies were 85.25 and 83.56% in the external validation sets from two centres. Specifically, 85.71 and 85.15% of NOA patients from two centres that experienced sperm retrieval failure were correctly identified using our model.

LIMITATIONS, REASONS FOR CAUTION: A small proportion of males with NOA in whom sperm were successfully retrieved with FNA were misclassified; therefore, TESA techniques with higher sperm retrieval rates may be attempted in patients with high predicted risks

Introduction

The incidence of azoospermia, which is divided into obstructive azoospermia (OA) and non-obstructive azoospermia (NOA), ranges from 1 to 15% in male infertility patients (Dabaja and Schlegel, 2013). NOA is diagnosed clinically in 70% of azoospermic men and is caused by testicular spermatogenesis dysfunction, sex chromosome malformations, Y chromosome microdeletions, etc. (Gudeman et al., 2015; Ezeh., 2000; Raman and Schlegel, 2003). Combined with the development of ICSI, testicular sperm aspiration (TESA) can enable NOA patients to have their own genetic offspring (Niederberger, 2012; Bernie et al., 2015). TESA techniques, including fine needle aspiration (FNA), testicular sperm extraction (TESE) and micro-testicular sperm extraction (micro-TESE), have different success rates in terms of sperm acquisition (Bernie et al., 2015). FNA is a simple surgical procedure and causes relatively little damage compared to the other two techniques (Lewin et al., 1999). However, the major limitation of FNA is the uncertainty of obtaining spermatozoa owing to the acquisition of less testicular tissue than the other two surgical methods and the blind puncture technique (Beliveau et al., 2011; Bernie et al., 2015). The FNA mapping technique uses testicular tissue samples collected at different puncture points of the testicle, and once sperm is found, small-incision TESE can be performed locally (Turek et al., 1997). However, multi-incision TESE on the testicular tunica cannot efficiently improve sperm retrieval after FNA failure (Fahmy, et al., 2000, Li, et al., 2001). Currently, almost all reproductive medicine centres in China are applying a revised FNA technology at a single puncture point using a larger side hole puncture needle and increased negative pressure, which can obtain the same quantity of testicular tissue as that obtained by TESE for patients with NOA, but the sperm retrieval rate of FNA in NOA patients is <50% (Ma, et al., 2012, Mao et al., 2018). Therefore, predicting the risk of sperm retrieval failure with FNA for patients with NOA could help when deciding whether FNA techniques should be recommended to reduce unnecessary punctures, which can lead to potential complications. If a patient suffering from NOA is predicted to have a high risk of sperm retrieval failure, a TESA procedure, such as micro-TESE, with a higher sperm retrieval success rate than that of FNA should be recommended (Tsujimura, 2007).

However, there is still a lack of detailed research on predicting the likelihood of sperm acquisition preoperatively, and predictive factors remain controversial for patients with NOA. A number of factors have been suggested to be predictors of sperm extraction, including testicular volume, serum FSH, inhibin B, and anti-Müllerian hormone levels, BMI, age and seminiferous tubule diameter (Chen et al., 2004; Toulis et al., 2010; Bryson et al., 2014; Yildirim et al., 2014; Yang et al., 2015). Since there are many indicators associated with the sperm acquisition success rate, existing models have not yet shown satisfactory predictive capacity (Samli and Dogan, 2004; Tsujimura et al., 2004; Boitrelle et al., 2011; Ramasamy et al., 2013). In addition, a reliable prediction model needs to perform well not only in the population in which it has been constructed but also in a crucial external validation set so that it can be put into practice in different centres (Cissen et al., 2016). In this study, a multivariable risk prediction model of sperm retrieval failure with FNA for individuals with NOA was developed and validated in both the same and one other reproductive medicine centre before FNA procedures were performed.

Materials and Methods

Ethics approval

The protocol for this multicentre study was approved by the Ethics Commie of West China Second University Hospital of Sichuan University (Project number: 2018028). All couples signed informed consent for treatment and follow-up before participating in this study.

Population and FNA procedure

Data from 317 males with NOA who underwent FNA at West China Second University Hospital between 1 January 2015 and 31 July 2017 were collected to build the prediction model. In this study, NOA was defined as azoospermia without evidence of obstruction in the vas deferens or congenital absence of bilateral vas deferens or azoospermia with a history of scrotal swelling or chronic epididymitis with epididymal enlargement. Additionally, patients diagnosed with Klinefelter syndrome were excluded due to the extremely low sperm retrieval rate by FNA. Each patient with NOA underwent a complete male fertility assessment by an andrologist before surgical sperm extraction. The model was then validated using two datasets from two prospective cohorts collected from 25 October 2017 to 31 March 2018. One cohort was recruited from the same centre as the modelling cohort, and the other was recruited from Jinjiang Maternity and Child Health Hospital.

We performed blind testicular puncture through the skin with FNA. After routine disinfection and draping, the patient was positioned in a supine position, and the procedure was performed under local anaesthesia. The testicle with the relatively larger testicular volume or the right testicle in cases of equal testicular volume was selected as the puncture site to avoid a left invisible varicocele. Then, a spermatic nerve block with 5 ml of 2% lidocaine hydrochloride was performed on the surgical side, infusing anaesthesia into the testicular tunica albuginea for ~3 min. The operator fixed the testicle of the patient with the left hand and held a 20-ml side-hole needle with the right hand to puncture the testicular albuginea after anaesthesia was effective. When the needle reached the appropriate depth to sustain negative pressure for aspiration of testicular tissue and resistance was felt, the needle was slowly withdrawn. The testicular tissue was aspirated through the scrotal skin puncture point, and the needle was completely recovered. Then, haemostasis was carefully evaluated to confirm no visible bleeding at the puncture site. The seminiferous tubules were separated with a needle in the culture fluid drop (Tyrode's fluid, CAF 2.5 mM PTX 7.5 mM pH 7.4) on a slide to observe their appearance, thickness and fullness under a dissecting microscope (×25). Punctured seminiferous tubules were removed, and the cell suspension was covered with a coverslip. The morphology and activity of mature sperm were observed under an inverted microscope (×400). If there were mature spermatozoa, ~100 sperm stained with eosin were counted to calculate the sperm survival rate.

Study design

The study consisted of three stages. In stage I, with respect to data accessibility the demographic and clinical characteristics (including age, testicular volume, infection history, testosterone, FSH, LH, and oestrogen levels and fertility history) were tested as potential predictors using univariate analysis between NOA patients with successful and failed sperm retrieval with FNA. Distributions of potential predictors were also analysed in order to determine the optimal predictors to include.

In stage II, risk prediction models were built and evaluated using statistical measures, such as goodness of fit and clinical interpretability, to select an optimal model. Receiver operating characteristic (ROC) curve analysis was performed to determine the best cut-off point for the risk of sperm retrieval failure and to calculate the AUC of the prediction model.

In stage III, the prediction accuracy of the optimal model was validated using two validation sets. Validation set 1 was collected from the same centre as the modelling set. Validation set 2 was collected from another centre for external validation. Moreover, the overall accuracy, sensitivity and specificity of the two validation sets were compared to evaluate the consistency of our prediction model.

Statistical analysis

In stage I, univariate tests were performed to evaluate potential predictors. Between males with NOA who had a successful sperm retrieval and those who had a failed sperm retrieval with FNA, the normally distributed variables were tested using the Student's *t*-test, and categorical variables were tested using the chi-square test. Non-parametric tests were used to test continuous variables that were not normally distributed.

In stage II, multivariable logistic regression models were built in a forward stepwise manner, and the criteria for inclusion and exclusion were 0.10 and 0.05, respectively. First-order interactions of significant variables in the model were included to test for effect modifiers. ROC curve analysis was then performed to evaluate the final risk prediction model using the AUC and to calculate the cut-off value for the prediction of sperm retrieval failure.

In stage III, for each individual with NOA in the validation sets, the predicted risks and outcomes of sperm retrieval failure were calculated using the final risk prediction model and the cut-off value for predicted risk. Accuracy was then calculated for all patients and for each validation set. Specifically, to estimate the potential benefit of the prediction model for males undergoing FNA, the true positive rate (TPR) was calculated to describe the proportion of males with NOA who are at high risk of sperm retrieval failure and should thus avoid FNA procedures.

Table I Baseline characteristics of men with non-obstructive azoospermia in the three datasets.

	Modelling set (n = 317)	Validation set 1 (n = 61)	Validation set 2 (n = 219)	Total (n = 597)
Age (years, SD)*	29.66 ± 4.53	29.20 ± 4.98	32.27 ± 5.22	30.50 ± 4.98
Positive fertility history (%, n)	6% (19)			
Sperm retrieval failure (%, n)	58.68% (186)	55.7% (34)	48.9% (107)	54.8% (327)
Infection (%, n)	4.10% (13)			
Testicular texture abnormality (%, n)	18.6% (59)			
Chromosome abnormality (%, n)	5.7% (5 and 229 not tested)			
AZFc deletion (%, n)	12.1% (21 and 173 not tested)			
Testicular volume** (ml, mean ± SD)	13.47 ± 3.86	11.52 ± 4.36	12.13 ± 3.78	12.89 ± 3.79
Testosterone (ng/ml, mean ± SD)	4.20 ± 2.44			
FSH (IU/l, mean ± SD)	12.44 ± 8.60	11.57 ± 6.34	14.16 ± 11.95	12.91 ± 9.84
LH (IU/l, mean ± SD)	5.49 ± 2.84			
Oestrogen (pg/ml, mean ± SD)	34.09 ± 11.93			

*Chi-square = 34.609, $P < 0.001$.
**Chi-square = 7.211, $P = 0.027$.
AZF = azoospermia factor.

Modelling set and validation set 1 are a retrospective cohort and prospective cohort of men with non-obstructive azoospermia (NOA) collected at a university affiliated hospital, while validation set 2 was collected from another centre for external validation. Only variables included in the prediction model were shown in validation datasets.

Results

Overall characteristics of the modelling and validation sets

Our study contained three datasets: one for modelling the risk prediction model and two for external validation. The demographic and clinical characteristics of the different sets are summarized as the means and SDs for continuous variables and as frequencies with percentages for categorical variables. Only variables included in the final model are presented for validation sets 1 and 2.

The sperm retrieval failure rate and FSH levels were slightly lower in the validation sets than in the modelling set, but the difference was not statistically significant. Age and testicular volume were significantly different among the different sets, as assessed by the Kruskal–Wallis test. The significance of age and testicular volume and the non-significant difference in FSH levels (Table I) suggest potential differences in the selection criteria between centres; therefore, the validation sets presented adequate heterogeneity to evaluate the generalizability of the risk prediction model.

Stage I: predictors in the development set

The Mann–Whitney test suggested that age, testicular volume, testosterone, FSH, LH and oestrogen levels were significantly different between NOA patients with and without failed sperm retrieval with FNA (Table II) and therefore could be potential predicators included in stage II.

Stage II: modelling and evaluation

Multivariable logistic regression modelling

In Stage II, to identify patients who would experience sperm retrieval failure with FNA, logistic regression models were built to predict whether sperm extraction would fail. Outcome 1 indicates that sperm extraction failed. Multicollinearity was insignificant in the final model. The included variables were defined as those used in former studies. Specifically, variables that were not normally distributed, such as age and volume of the testicle in which FNA was performed, were categorized into groups and included as dummy variables. Dummy variables with close regression coefficients were merged to improve the simplicity of the model. After comparison and merging, testicular volume was categorized into three groups (<12, 12–14 and 15 ml or greater), and age was categorized into two groups (<35 years and 35 years or older). Additionally, natural cubic splines and cubic splines were used to fit the non-linearity exhibited by age and testicular volume but produced only slight improvement in discrimination. Therefore, age and testicular volume were included as dummy variables regarding the clinical interpretability and feasibility of the model. After performing stepwise variable selection (Wald forward) and assessing clinical significance, the final risk prediction model was generated as below.

The multivariable logistic regression model demonstrated that low FSH levels, large testicular volume and age 35 years or older were associated with a low risk of sperm retrieval failure (Table III).

ROC analysis: cut-off value and discrimination evaluation

Following the creation of the final prediction model, the predicted risk of sperm retrieval failure for each male with NOA was calculated as follows:

$$B = 0.003 + (0.124 \times \text{FSH}) + (-1.448 \times \text{Age 35 years or older})$$
$$+ (-0.651 \times \text{Testicular volume } 12 - 14\text{ml})$$
$$+ (-1.448 \times \text{Testicular volume } 15 \text{ ml or greater})$$

Table II Univariable tests for potential predictors.

	Mann–Whitney *U*	Wilcoxon *W*	*Z*	*P*
Age (years)	9635	27 026	−3.18	0.001
Testicular volume (ml)	6880.5	24 271.5	−6.806	<0.001
Testosterone (ng/ml)	4747.5	13 393.5	−9.254	<0.001
FSH (IU/l)	8809	17 455	−4.199	<0.001
LH (IU/l)	9777	27 168	−2.994	0.003
Oestrogen (pg/ml)	10 232	27 623	−2.428	0.015

Non-parametric tests showed that age, testicular volume, testosterone, FSH, LH and oestrogen were different between men with NOA having failure and successful sperm retrieval. $N = 317$

Table III Binary multivariable logistic risk prediction model of sperm retrieval failure with fine needle aspiration.

Variables	*B*	*P*-value	Odds ratio (95% CI)
FSH	0.124	<0.001	1.132 (1.083–1.183)
Testicular volume group 2:12–14 ml	−0.651	0.095	0.521 (0.234–1.120)
Testicular volume group 3: > 15 ml	−1.382	0.001	0.251 (0.118–0.535)
Age 35 years or older	−1.448	0.002	0.235 (0.106–0.519)
Constant	0.003	0.994	1.003

B = regression coefficient. $N = 317$.

Predicted Risk of Sperm Retrieval Failure = 1/(1 + exp(−B))

The corresponding ROC (Fig. 1) had an AUC of 0.823 (95% CI: 0.776–0.871) and an AUC of 0.816 with optimism correction using bootstrapping, suggesting excellent discrimination capacity. Youden's index showed that the best cut-off value for predicting the risk of sperm retrieval failure was 64.61%, with a sensitivity of 0.677 and a specificity of 0.863.

Calibration analysis
To assess the similarity between the predicted risk and observed risk of sperm retrieval failure, individuals were categorized into four groups of equal size according to the level of predicted risk (Fig. 2 and Table IV). The mean predicted risks for each group were then compared to the observed failure rates. Linear regression with observed failure as the outcome and predicted risk as the independent variable showed good calibration in both the modelling sets (*B* = 1.04(0.87–1.21), constant = 0.02(0.13–0.09)) and validation sets (*B* = 1.13-(0.98–1.28), constant = 0.14(−0.24–0.04)).

Stage III: prediction capacity and benefit evaluation

Predictive results and benefit
To evaluate the predictive accuracy and benefit of the risk prediction model if it was put into practice, we recruited 280 males with NOA who had not yet undergone FNA from two centres and included them in validation sets 1 and 2. For each individual in both validation sets, the predicted risk of sperm retrieval failure was calculated, and the predicted outcome was then determined using the cut-off value for predicted risk from stage II. The predicted outcome and observed

Figure 1 Receiver operating characteristic curve analysis of the prediction model. The area under the receiver operating characteristic curve indicates the prediction capacity of the model.

outcome of FNA were compared for external validation of the predictive model (Table V).

Overall, the observed outcomes of 83.93% (235 out of 280) of males with NOA were the same as the predicted outcomes. This finding suggested that the prediction model generally worked well in the external validation sets from either the same or different centre, and both the prediction accuracy and generalizability of the model were validated.

The objective of this study was to reduce the unnecessary complications of FNA and focus on those patients who have sperm retrieval failure. Our prediction model accurately identified 85.29% of individuals who experienced sperm retrieval failure (116 out of 136) in the two validation sets. This result presented an impressive improvement in the accuracy of clinical advice provided to men with NOA who were likely to have failed FNA outcomes. Considering the heterogeneity between different centres, the prediction results in validation set 1 showed an overall accuracy of 85.25% and a TPR of 85.71%; validation set 2 also showed a good accuracy of 83.56% and a TPR of 85.15%. The accuracy of the prediction model showed no significant differences among patients from different centres.

Sensitivity analysis

To evaluate the uncertainty of the prediction model, sensitivity analyses were performed (Table VI). All 597 males with NOA were included to model the risk of extraction failure. The model was built with the same method as the prediction model.

Compared to the prediction model, the regression coefficients of the groupings of FSH levels, age and testicular volume were consistent after including the validation sets, suggesting that the model had good consistency even for individuals from different centres.

Discussion

In this study, a risk prediction model was built for individuals undergoing FNA. The FNA technique has advantages in obtaining adequate testicular tissue (the same amount as obtained by TESE) while causing relatively little trauma. This simple surgery can be performed with economical equipment and is associated with fewer complications than other techniques. However, failure of FNA to obtain sperm is a key issue when determining the preoperative indications of patients. Prediction using a single factor has been somewhat clinically effective in the past, but an effective prediction model based on multi-factor clinical data may have a high reference value when choosing surgical methods for NOA patients. In our study, three of the independent predictors included in the final model had also been identified previously, i.e. serum FSH level, testicular volume and age, each having a correlation with the testicular sperm retrieval rate (Carpi, *et al.*, 2009, Turunc, *et al.*, 2010, Cissen *et al.*, 2016). However, it is notable that FSH was found not to be correlated with the sperm acquisition rate in a retrospective study including 1275 patients with NOA using micro-TESE combined with ICIS (Ishikawa, *et al.*, 2015). It is possible that in most micro-TESE series with fully detailed dissections, small, focal areas of sperm production can be found that may be missed with FNA. As predictors of the sperm acquisition rate may vary in different procedures, our model predicts only the sperm acquisition rate for FNA and should not be applied to other procedures.
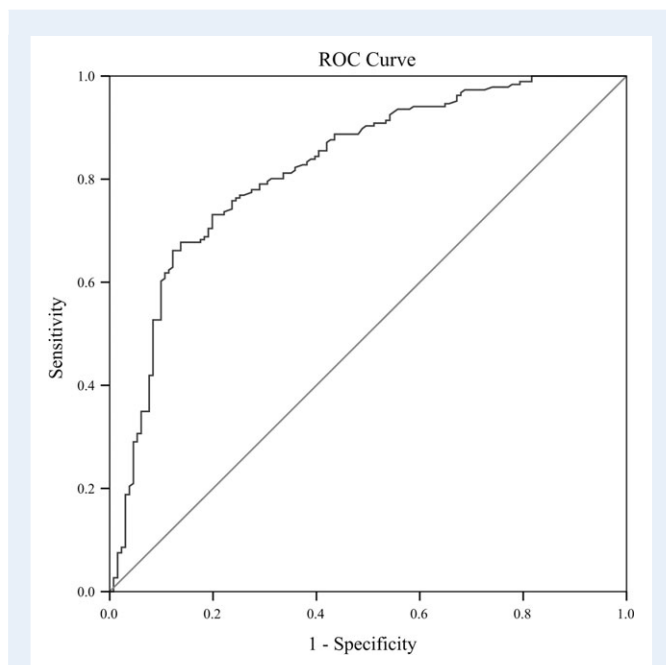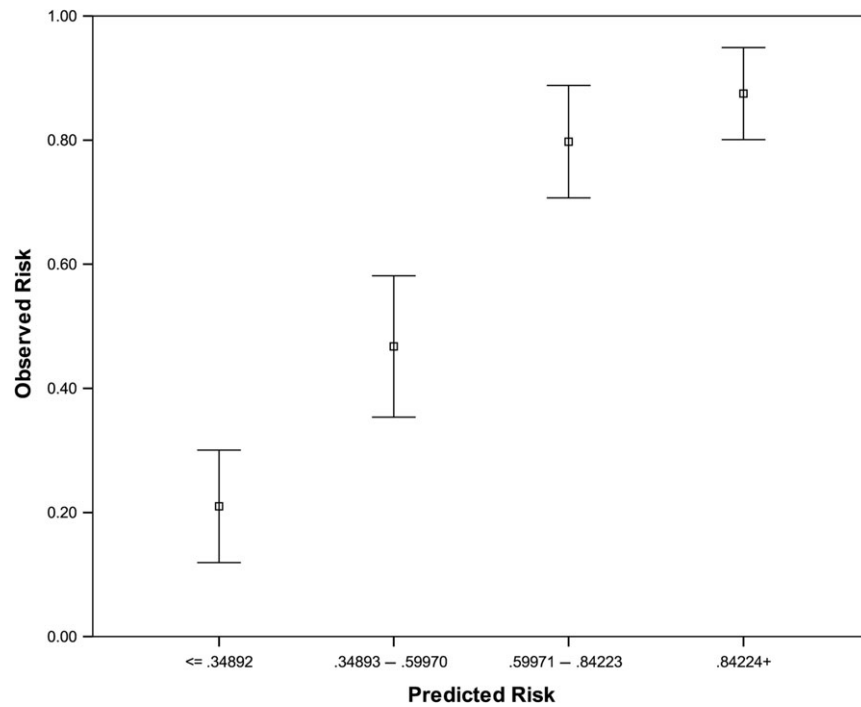
Figure 2 Association of predicted and observed risk of sperm retrieval failure. The four groups with same size represent the predicted risk and 95% CI.

Table IV Predicted and observed risk of sperm retrieval failure.

	Group	Group (n)	Mean of predicted risk (%)	Failure (n)	Mean of observed risk (%)
Modelling set (internal validation) (n = 317)	<35	81	23.83	17	20.99
	35–60	78	46.95	37	47.44
	65–85	83	73.53	67	80.72
	>85	75	92.06	65	86.67
Validation set 1 (n = 61)	<35	13	23.10	2	15.38
	35–60	12	50.14	1	8.33
	65–85	28	74.90	24	85.71
	>85	8	89.86	7	87.50
Validation set 2 (n = 219)	<35	78	21.25	8	10.26
	35–60	34	47.27	11	32.35
	65–85	44	75.39	33	75.00
	>85	63	92.64	55	87.30
Validation set total (external validation) (n = 280)	<35	91	23.10	10	10.99
	35–60	46	50.14	12	26.09
	65–85	72	74.90	57	79.17
	>85	71	89.86	62	87.32

Modelling set was categorized into four groups of equal size according to the level of predicted risk. The mean predicted risks for each group were then compared to the observed failure rates to assess the similarity between the predicted risk and observed risk of sperm retrieval failure in different sets.

Other models (Samli and Dogan, 2004; Tsujimura et al., 2004; Boitrelle et al., 2011; Ramasamy et al., 2013) have been developed previously to predict the spermatozoa acquisition success rate in patients with NOA; one study reported a sensitivity of 71.0% and a specificity of 71.4% (Tsujimura et al., 2004), and another study reported a sensitivity of 68.0% and a specificity of 87.5% (Samli and Dogan, 2004). Two other studies performed AUC analyses and found AUC values of 0.64 and 0.66 (Boitrelle et al., 2011; Ramasamy et al.,

Table V Predicted and observed outcomes of fine needle aspiration in the validation sets.

Observed outcome	Predicted outcome		
	Success	Failure	Total
Validation set 1 (from the same centre, $n = 61$)			
Success	22	5	27
Failure	4	30	34
Total	26	35	61
Validation set 2 (from another centre, $n = 219$)			
Success	97	15	112
Failure	21	86	107
Total	118	101	219

Predicted outcomes came from the comparisons between predicted risk and cut-off value.

Table VI Sensitivity analysis: risk prediction model built using all 597 males with non-obstructive azoospermia.

	B	P-value	Odds ratio (95% CI)
FSH	0.133	<0.001	1.142(1.107–1.179)
Testicular volume Group 2: 12–14 ml	−0.651	0.02	0.522(0.302–0.902)
Testicular volume Group 3: 15 ml~	−1.424	<0.001	0.241(0.143–0.405)
Age 35 years or older	−1.256	<0.001	0.285(0.163–0.496)
Constant	−0.315	0.286	0.730

2013). These models were not validated using external data, except for the model from Cissen's study (Cissen *et al.*, 2016). Our model showed advantages and clinical application value. According to the results of the external validation sets, our model showed good performance in terms of calibration, consistency, generalizability and, most importantly, prediction accuracy for FNA failure. When this risk prediction model was applied to patients with NOA prior to FNA, we could accurately identify 86.4% of those who were likely to experience sperm retrieval failure. In addition, the overall sperm retrieval failure rate was considerably reduced to 26.81% (85 in 317) compared to 54.8% in all 597 patients. The improvement in accuracy by identifying subgroups of NOA patients may also lead to considerable relief in terms of physical and mental burdens, as well as economic aspects. In addition, by identifying those most likely to have successful FNA outcomes, hospitals and doctors can focus their attention on those not likely to have successful outcomes with FNA and recommend further tests and more efficient TESA procedures. This focus may consequently lead to potential improvements in the success rate of sperm retrieval for all NOA patients. In addition, we prospectively recruited NOA patients from two different centres to build the validation sets. This method helped to maximize not only data quality but also generalizability. Although recommending FNA for specific patients might be quite arbitrary and probably differs in different centres, our model not only performed well in a validation set from the same centre where the modelling set was collected but also performed even better in a validation set from another centre. This demonstrates the generalizability of our model and further studies incorporating centres in different regions, with patients of different races are about to begin.

However, our study has some limitations. With respect to other studies focused on the sperm retrieval rate with FNA, some clinical features, such as azoospermia factor c (AZFc) deletion, fertility history, and LH levels, were not included in our final model. There might be two different situations in which these potential predictors could be used. Although some of these predictors, such as AZFc deletions and chromosomal abnormalities, might play important roles and be identified as significant in studies on the mechanisms of infertility, patients incur additional costs and must undergo further tests. Moreover, these potential risk factors are not very common in the overall cohort of males with NOA. For example, 72.24 and 54.57% of NOA patients in the modelling set did not undergo AZFc deletion and chromosome testing, respectively. Therefore, very few patients were willing to take and pay for these tests. Predictors collected by routine tests for NOA patients undergoing TESA, such as LH levels, were excluded from the prediction model because of the lack of association with outcomes. As the objective of this study was to reduce unnecessary FNA procedures for NOA patients rather than to explain the mechanisms of infertility, we mainly focused on the accessibility of predictors and the prediction accuracy of the model. The results of external validation using two validation sets showed that the final risk prediction model might be a good solution for the study objective. Another potential limitation of our model is that we might have misclassified 14.23% patients as having sperm retrieval failure with FNA. Although they were a minority of the overall NOA patients, we still suggest that these patients undergo further tests and other sperm retrieval procedures with higher success rates than FNA, such as micro-TESE, rather than ending the pursuit of having their own genetic offspring. Also, as only the sperm retrieval

was investigated in this study, good pregnancy outcomes are not guaranteed to follow successful FNA procedures. Therefore, studies of, for example, the fertilization, pregnancy (hCG test) and live birth rates after successful FNA should be performed in the future.

Compared to TESE, micro-TESE can achieve a higher sperm concentration (63 versus 45%) while obtaining smaller amounts of tissue (9.4 mg versus 720 mg); the sperm retrieval rate of micro-TESE in 563 NOA patients was 61% (Schlegel, 1999). Franco *et al.* (2016) used stepwise micro-TESE for 64 NOA patients who previously underwent unsuccessful TESE and achieved a sperm retrieval rate of 28.1% (18/64) (Franco *et al.*, 2016). A study of 435 NOA patients also showed that micro-TESE had a higher rate of sperm acquisition (81% than conventional TESE (50%)) in this patient population and that the incidence of complications was significantly reduced (Ramasamy *et al.*, 2005). Thus, we recommend using micro-TESE rather than TESE owing to its higher sperm retrieval success rate in patients with NOA and a high predicted risk of FNA failure. According to the results, we believe that our model can be used in clinical practice to predict sperm retrieval failure risk for individuals with NOA hoping for their own genetic offspring before they undergo the FNA procedures. Those with a predicted risk less than the cut-off value would undergo the relatively easy FNA procedure (and with less damage), while micro-TESE would be recommended to give a better chance of sperm retrieval in those men with a high predicted risk of failure with FNA.

Authors' roles

Funding

Conflict of interest

There are no competing interests related to this study.

References

Beliveau ME, Turek PJ. The value of testicular 'mapping' in men with non-obstructive azoospermia. *Asian J Androl* 2011;**13**:225–230.

Bernie AM, Mata DA, Ramasamy R, Schlegel PN. Comparison of microdissection testicular sperm extraction, conventional testicular sperm extraction, and testicular sperm aspiration for nonobstructive azoospermia: a systematic review and meta-analysis. *Fertil Steril* 2015;**104**:1099–1103.

Boitrelle F, Robin G, Marcelli F, Albert M, Leroy-Martin B, Dewailly D, Rigot JM, Mitchell V. A predictive score for testicular sperm extraction

quality and surgical ICSI outcome in non-obstructive azoospermia: a retrospective study. *Hum Reprod* 2011;**26**:3215–3221.

Bryson CF, Ramasamy R, Sheehan M, Palermo GD, Rosenwaks Z, Schlegel PN. Severe testicular atrophy does not affect the success of microdissection testicular sperm extraction. *The Journal of urology*. **2014**;**191**(1):175–178.

Carpi A, Sabanegh E, Mechanick J. Controversies in the management of nonobstructive azoospermia. *Fertil Steril* 2009;**91**:963–970.

Chen SC, Hsieh JT, Yu HJ, Chang HC. Appropriate cut-off value for follicle-stimulating hormone in azoospermia to predict spermatogenesis. *Reprod Biol Endocrinol* 2010;**8**:108.

Cissen M, Meijerink AM, D'Hauwers KW, Meissner A, van der Weide N, Mochtar MH, de Melker AA, Ramos L, Repping S, Braat DDM *et al.* Prediction model for obtaining spermatozoa with testicular sperm extraction in men with non-obstructive azoospermia. *Hum Reprod* 2016;**31**:1934–1941.

Dabaja AA, Schlegel PN. Microdissection testicular sperm extraction: an update. *Asian J Androl* 2013;**15**:35–39.

Ezeh UI. Beyond the clinical classification of azoospermia: opinion. *Hum Reprod* 2000;**15**:2356–2359.

Fahmy I, Kamal A, Mansour RT, Serour GI, Aboulghar M. Single large testicular biopsy is comparable to multiple biopsies to retrieve spermatozoa in patients with non-obstructive azoospermia. *Fertil Steril* 2000;**74**:S86–S87.

Franco G, Scarselli F, Casciani V, De Nunzio C, Dente D, Leonardo C, Greco PF, Greco A, Minasi MG, Greco E. A novel stepwise micro-TESE approach in non obstructive azoospermia. *BMC Urol* 2016;**16**:20.

Gudeman SR, Townsend B, Fischer K, Walters RC, Crain D. Etiology of azoospermia in a military population. *J Urol* 2015;**193**:1318–1321.

Ishikawa T, Yamaguchi K, Takaya Y, Nishiyama R, Kitaya K, Matsubayashi H. Predictors for sperm retrieval in microdissection sperm extraction for non-obstructive azoospermia. *Fertil Steril* 2015;**104**:e294.

Lewin A, Reubinoff B, Porat-Katz A, Weiss D, Eisenberg V, Arbel R, Bar-el H, Safran A. Testicular fine needle aspiration: the alternative method for sperm retrieval in non-obstructive azoospermia. *Hum Reprod* 1999;**14**:1785–1790.

Li P, Goldstein M, Schlegel PN. Surgical sperm retrieval: which method to use? *Natl J Androl* 2001;**7**:71–78.

Ma M, Ping P, Wang J, Li P, Yang S, Zhu J, Lu H, Hu H, LI Z. Three-step sperm retrieval for 73 non-obstructive azoospermia patients. *Natl J Androl* 2012;**18**:606–610.

Mao JM, Liu DF, Zhao LM, Hong K, Zhang L, Ma LL, Jiang H, Qiao J. Effect of testicular puncture biopsy on the success rate of microdissection testicular sperm extraction for idiopathic non-obstructive azoospermia. *J Peking University (Health Sciences)* 2018;**50**:613–616.

Niederberger C. Re: A predictive score for testicular sperm extraction quality and surgical ICSI outcome in non-obstructive azoospermia: a retrospective study. *J Urol* 2012;**188**:558–559.

Raman JD, Schlegel PN. Testicular sperm extraction with intracytoplasmic sperm injection is successful for the treatment of nonobstructive azoospermia associated with cryptorchidism. *J Urol* 2003;**170**:1287–1290.

Ramasamy R, Padilla WO, Osterberg EC, Srivastava A, Reifsnyder JE, Niederberger C, Schlegel PN. A comparison of models for predicting sperm retrieval before microdissection testicular sperm extraction in men with nonobstructive azoospermia. *J Urol* 2013;**189**:638–642.

Ramasamy R, Yagan N, Schlegel PN. Structural and functional changes to the testis after conventional versus microdissection testicular sperm extraction. *Urology*. 2005;**65**(6):1190–1194.

Samli MM, Dogan I. An artificial neural network for predicting the presence of spermatozoa in the testes of men with nonobstructive azoospermia. *J Urol* 2004;**171**:2354–2357.

Schlegel PN. Testicular sperm extraction: microdissection improves sperm yield with minimal tissue excision. *Hum Reprod* 1999;**14**:131–135.

Tsujimura A, Matsumiya K, Miyagawa Y, Takao T, Fujita K, Koga M, Takeyama M, Fujioka H, Okuyama A. Prediction of successful outcome

of microdissection testicular sperm extraction in men with idiopathic nonobstructive azoospermia. *Journal of Urology.* 2004;**172**(5 l):1944–7.

Tsujimura A. Microdissection testicular sperm extraction: prediction, outcome, and coplications. *Int J Urol* 2007;**14**:883–889.

Toulis KA, Iliadou PK, Venetis CA, Tsametis C, Tarlatzis BC, Papadimas I, Goulis, DG. Inhibin B and anti-Mullerian hormone as markers of persistent spermatogenesis in men with non-obstructive azoospermia: a meta-analysis of diagnostic accuracy studies. *Hum Reprod Update* 2010;**6**:713–724.

Turek PJ, Cha I, Ljung BM. Systematic fine-needle aspiration of the testis: correlation to biopsy and results of organ'mapping'for mature sperm in azoospermic men. *Urology* 1997;**49**:743–748.

Turunc T, Gul U, Haydardedeoglu B, Bal N, Kuzgunbay B, Peskircioglu L, Ozkardes H. Conventional testicular sperm extraction combined with the microdissection technique in nonobstructive azoospermic patients: a prospective comparative study. *Fertil Steril* 2010;**94**:2157–2160.

Yang Q, Huang YP, Wang HX, Hu K, Wang YX, Huang YR, Chen B. Follicle-stimulating hormone as a predictor for sperm retrieval rate in patients with nonobstructive azoospermia: a systematic review and meta-analysis. *Asian J Androl* 2015;**17**:281–284.

Yildirim ME, Koc A, Kaygusuz IC, Badem H, Karatas OF, Cimentepe E, Unal D. The association between serum follicle-stimulating hormone levels and the success of microdissection testicular sperm extraction in patients with azoospermia. *Urol J* 2014;**11**:1825–1828.