



Original Research

# Optimizing the High-Quality Embryo Rate in Couples With Male Factor Infertility: Insights From Predictive Modeling and Causal Effect Estimations

Qingxin Wang<sup>1,\*</sup>, Xiaoli Sun<sup>1</sup>, Lingna Peng<sup>1</sup>, Xiaoling Gu<sup>1</sup>, Yannan Chen<sup>1</sup>, Minyan Yu<sup>1</sup>, Jian Song<sup>1</sup>, Quan Yuan<sup>2,\*</sup>

<sup>1</sup>Center for Reproductive Medicine, Affiliated Hospital of Nantong University, 226001 Nantong, Jiangsu, China

<sup>2</sup>Department of Urology, QingPu Branch of Zhongshan Hospital Affiliated to Fudan University, 201700 Shanghai, China

\*Correspondence: [cuixiself@163.com](mailto:cuixiself@163.com) (Qingxin Wang); [klz602251294@126.com](mailto:klz602251294@126.com) (Quan Yuan)

Academic Editor: Michael H. Dahan

Submitted: 1 March 2025 Revised: 7 April 2025 Accepted: 25 April 2025 Published: 18 June 2025

## Abstract

**Background:** Male infertility is a critical factor in the success of *in vitro* fertilization (IVF), yet a comprehensive predictive model for assessing its risk remains lacking. This study aimed to explore the factors affecting the high-quality embryo rate in male infertility patients, as well as to develop targeted intervention measures. **Methods:** A retrospective analysis was performed using clinical data from 373 infertility couples who underwent IVF treatment, and the couples were grouped based on a high-quality embryo rate of  $\geq 45\%$  and  $< 45\%$  for statistical analysis. We developed an outcome prediction model and a causal effect estimation model to evaluate the impact of different intervention measures, based on the results of univariate logistic regression analysis. **Results:** The results demonstrated significant differences in Antral Follicle Count (AFC), sperm DNA fragmentation index (DFI), male height, male weight, and mycoplasma infection. Further univariate logistic regression analysis identified that AFC, basal luteinizing hormone (LH), sperm DFI, male height, male weight, and mycoplasma infection significantly affected the high-quality embryo rate, all showing negative correlations. The dataset was divided into a training set (80%) and a test set (20%) for the construction and validation of the outcome prediction model and the causal effect estimation model. The causal effect estimation model for mycoplasma infection demonstrated that treating mycoplasma infection could increase the high-quality embryo rate. The causal effect estimation model for sperm DFI revealed that reducing sperm DFI could increase the high-quality embryo rate. The causal effect estimation model for male weight demonstrated that being overweight can reduce the high-quality embryo rate. **Conclusions:** Reducing sperm DFI levels, weight loss, and treating mycoplasma infection are effective methods for improving the high-quality embryo rate in male infertility.

**Keywords:** male infertility; high-quality embryo rate; intervention measures; predictive model; sperm DNA fragmentation index; mycoplasma infection; overweight

## 1. Introduction

Male infertility is a significant global factor affecting couples' fertility, with its incidence on the rise due to changes in lifestyle and environmental influences. Approximately 10% to 20% of couples face infertility issues, with male factors accounting for up to 50% of these cases [1]. With the aid of Assisted Reproductive Technology (ART), many infertile couples have been able to fulfill their desire to have children. Despite remarkable advancements in *in vitro* fertilization (IVF) techniques, the rate of high-quality embryos remains a critical limiting factor, directly impacting the success rate of IVF and the health of the resulting infants. Therefore, exploring the factors that influence the rate of high-quality embryos is not only scientifically meaningful for understanding and improving male infertility but also holds significant clinical value for enhancing the success rates of assisted reproductive technologies.

In recent years, research has found that in addition to the quality of a woman's eggs, the quality of a man's sperm is also one of the key factors determining the rate of high-quality embryos [2,3]. Sperm DNA Fragmentation Index (DFI), as an important indicator for assessing the integrity of sperm DNA, has been found to be closely related to the rate of high-quality embryos [4,5]. However, due to the influence of experimental design and detection methods, its predictive value for pregnancy outcomes is limited [5]. Ziouziou *et al.* [2] conducted a narrative review and meta-analysis of literature on sperm DNA fragmentation published over the past 5 years, indicating that an increase in sperm DFI has a negative impact on natural pregnancy and ART outcomes. DFI testing may be particularly important in the assessment of male infertility in cases of varicocele, idiopathic or unexplained infertility, recurrent miscarriage, or previous ART failure. Additionally, factors such as the male partner's body weight [6,7] and mycoplasma infections [8] may also affect the formation of high-quality em-



bryos. Lifestyle factors such as smoking, alcohol consumption, drug use, environmental exposure, and genetic factors are all significant determinants of male fertility [9–11].

Previous studies have covered multiple key factors of male infertility, such as lifestyle, environmental exposure, and genetic factors [2–11]. These studies all indicate that male infertility is a complex multifactorial issue that requires a comprehensive consideration of various potential influencing factors to develop effective prevention and treatment strategies. However, these studies also have some shortcomings. Firstly, most studies are observational, making it difficult to establish causal relationships. Secondly, there are differences in the design, samples, and methods of different studies, which limit the comparability and reproducibility of the results. Additionally, few studies have considered the interactive effects of multiple factors, which may limit our understanding of the complex pathophysiological mechanisms of male infertility. Currently, there is a lack of a comprehensive predictive model to assess the risk of male infertility, which restricts our ability to conduct personalized diagnosis and treatment in clinical practice. Therefore, this study aims to construct a comprehensive predictive model to assess the risk of male infertility by systematically analyzing multiple potential influencing factors and their interactive effects. This will help to more accurately identify high-risk populations and provide a scientific basis for personalized diagnosis and treatment, thereby improving the overall success rate of assisted reproductive technology. This not only helps to promote the reproductive health of patients but also provides new perspectives and research directions for related basic scientific research.

## 2. Material and Methods

### 2.1 Patient Selection

A retrospective analysis was conducted on the clinical data of 373 couples who underwent IVF treatment at the Reproductive Medicine Center of Affiliated Hospital of Nantong University from January 1, 2021, to June 30, 2024. The data were collected from medical records. Inclusion criteria: (1) Failure to conceive after 1 year or more of regular, unprotected intercourse with the same partner; (2) Male with normal or abnormal sperm parameters (according to World Health Organization [WHO], 2010 [12]: concentration, motility, and morphology, exclusion of azoospermia) and no andrological history of concern (cryptorchidism, hypogonadotropic hypogonadism, genetic abnormalities such as Klinefelter's syndrome or Y-chromosome microdeletion, drug abuse, cancer treatment, or other iatrogenic factors); Exclusion criteria: Factors affecting egg quality in women were excluded, such as age (selecting women under 38 years old), polycystic ovary syndrome, diminished ovarian reserve (selecting patients with Follicle-Stimulating Hormone [FSH] <10 mIU/mL, Anti-Mullerian Hormone [AMH] >0.12 ng/mL, Antral Follicle Count [AFC] >5 ng/mL), endometriosis, chromosomal abnormalities, and

other factors affecting egg quality. The study protocol was approved by the Ethics Committee of Affiliated Hospital of Nantong University (2025-K037-01).

### 2.2 IVF/ICSI Procedures

We used controlled ovarian stimulation, using a gonadotropin-releasing hormone antagonist and recombinant follicle-stimulating hormone/human menopausal gonadotropin. Once the 3 dominant follicles reached a mean diameter of at least 17 mm, we injected 250 µg of recombinant human chorionic gonadotropin (hCG) version SJ20130091 (Ovidrel, Serono, Geneva, Switzerland) and retrieved oocytes using transvaginal ultrasound guidance 36 hours later. We then incubated the oocytes in G-IVF™ version 10136 (Vitrolife, Gothenburg, Sweden) and performed IVF or intracytoplasmic sperm injection (ICSI) 4–6 hours after retrieval. Normal fertilization was determined by the presence of 2 pronuclei 16–20 hours after insemination. Embryonic development was monitored at 48 and 72 hours after retrieval and graded at 72 hours based on the number of cells, level of fragmentation, and cell size variation [13]. High-quality embryos were defined as grade I and II and reserved for later embryo transfer. The study examined the number of 2 pronuclei (2PN) (which indicated normal fertilized zygote or embryo), identified the day after IVF. The number of cleaved embryos with 2 or more blastomeres, occurred 2 days after IVF. The embryo quality grading was determined 3 days (D3) after IVF and divided into good quality (grade 1 and 2) and poor quality (grade 3 and 4) [14].

### 2.3 Semen Analysis and DNA Fragmentation Assay

On the day of oocyte retrieval, semen samples were collected by masturbation after 2 to 7 days of abstinence and each sample was incubated at 37 °C to liquefy for a maximum of 60 min. After the liquefaction, conventional semen analysis (sperm concentration and motility) was carried out according to the WHO guidelines, 2010 [12] by using a Computer-Assisted Sperm Analyzer (WLJY-9000, Weili, Beijing, China). Semen samples for fertilization were treated with density-gradient centrifugation (DGC) version 10102 (SpermGrad, Vitrolife, Gothenburg, Sweden) and swim up. This discontinuous density gradient centrifugation consisted of 2 (90% and 45%) 1 mL layers of SpermGrad, and 2 mL of semen was deposited on the 45% layer. The gradient was then centrifuged at 300 g for 20 minutes. After centrifugation, the seminal plasma supernatant was discarded, and the sperm pellet was washed with 1 mL of embryo culture medium version 10136 (Vitrolife, Gothenburg, Sweden). After centrifugation for 10 minutes, the pellet was suspended with 1 mL culture medium for sperm to swim up.

All men underwent a routine sperm chromatin dispersion (SCD) test 1 month prior to IVF procedure. SCD test version 20152401259 (ShenZhen Huakang Co., Ltd.,

Shenzhen, Guangdong, China) was carried out according to the manufacturer's instructions as described in our previous study [15]. In brief, semen samples were diluted with phosphate-buffered saline to a concentration of  $5\text{--}10 \times 10^6/\text{mL}$ , the prepared spermatozoa were mixed with melted agarose, then pipetted onto pre-coated slides and covered with a coverslip ( $22 \text{ mm} \times 22 \text{ mm}$ ). The slide was placed in the refrigerator at  $4^\circ\text{C}$  for 4 min to allow the agarose to solidify with sperm cells embedded. The coverslip was gently removed and the slide was incubated in an acid solution for 7 min, and then in lysis buffer for 20 min. After washing the slide for 3 minutes with washing buffer, the slide was dehydrated in increasing concentrations of ethanol (70%, 90%, 100%, each for 2 min) and air-dried. After Wright-Giemsa staining, at least 200 spermatozoa per sample were scored under a 400 objective of the microscope for holes. Sperm with a small halo (similar to/smaller than a third of the minor diameter of the nucleus) or no halo was considered to have significant DNA fragmentation. DFI = sperms with DNA fragmentation/total sperms counted  $\times 100\%$ .

#### 2.4 Model Construction and Validation

The dataset was divided into 80% for training and 20% for testing, resulting in the construction of a predictive model for outcomes and a causal effect estimation model. The outcome prediction model utilized the Random Forest algorithm, which was selected after parameter optimization. The causal effect estimation model employed a double machine learning (DML) approach to analyze the rate of high-quality embryos. Model validation included assessments using Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), as well as SHapley Additive exPlanations (SHAP, a unified framework for interpreting machine learning models) methods for analyzing feature importance and creating marginal causal effect plots.

#### 2.5 Intervention Strategy Development

Based on the potential outcomes framework and causal effect estimation, personalized intervention strategies have been formulated. These strategies take into account modifiable male factors that significantly impact the rate of high-quality embryos in univariate logistic regression, such as pharmacological treatment for mycoplasma positivity, improving sperm quality, and weight reduction, all being evaluated in conjunction with clinical practice.

#### 2.6 Statistical Methods, Software, and Tools

Statistical analysis was performed using the SPSS statistical package version 26.0 (IBM Corp., Armonk, NY, USA). Prior to the analysis, all continuous data underwent normality tests, including the Shapiro-Wilk test and the Kolmogorov-Smirnov test. Data that conform to a normal distribution were expressed as the mean plus or minus the standard deviation, while data that do not conform to a

normal distribution were represented using quartiles. Inter-group comparisons of continuous data were performed using the Wilcoxon rank-sum test. For categorical data, the chi-square test was used without continuity correction, as all theoretical frequencies in our dataset were greater than 5. The significance level was set at  $\alpha = 0.05$ . Variable selection was carried out using univariate and multivariate logistic regression analysis. Model development was based on Python version 3.9 (Python Software Foundation, Wilmington, DE, USA), with the predictive model utilizing the scikit-learn package version 1.0.2 (scikit-learn community, Paris, France) and causal effect estimation employing the EconML package version 0.15.1 (Microsoft Research, Redmond, WA, USA).

### 3. Results

#### 3.1 The Comparison of Male and Female Characteristics Between the 2 Groups

A total of 373 infertile couples were initially enrolled in this study. Participants were categorized into 2 groups based on their D3 high-quality embryo rate: a control group (high-quality embryo rate  $<45\%$ ,  $n = 106$ ) and an observational group (high-quality embryo rate  $\geq 45\%$ ,  $n = 267$ ). The grouping criterion is established based on the expert consensus on key quality control indicators in human ART embryology laboratories in China, which is also consistent with Italian Society of Fertility and Sterility [16]. Significant differences were observed between the 2 groups in terms of several characteristics, including female AFC, sperm DFI, male height, male weight and male mycoplasma levels. In male characteristics, sperm DFI and male weight were found to be lower in the observational group compared to the control group. In the group with a higher rate of high-quality embryos, the proportion of mycoplasma-negative individuals is higher (Table 1). This suggests that the presence of mycoplasma infection may have a negative impact on embryo quality, which is an important consideration in the field of assisted reproductive technology.

#### 3.2 Factors Affecting the Rate of High-Quality Embryos

Univariate logistic regression analysis was first performed on the baseline data presented in Table 1 to identify potential factors associated with the high-quality embryo rate. Table 2 shows the results of factors with  $p < 0.5$  in univariate logistic regression. It can be seen that female AFC, basal LH levels, sperm DFI, male height, male body weight and mycoplasma infection all have significant impacts on the high-quality embryo rate, and all are negatively correlated (Table 2). Taking the high-quality embryo rate as the dependent variable, and incorporating variables with  $p < 0.1$  from the univariate analysis into the multivariate model, a bidirectional stepwise logistic regression was performed. The results revealed that factors significantly affecting the rate of high-quality embryos included mycoplasma infection, basal LH levels and male body weight (Table 3).

**Table 1. Baseline characteristics comparison between control group (D3 high-quality embryo rate <45%) and observation group (D3 high quality embryo rate ≥45%).**

Variable	Overall (n = 373)	Control group (n = 106)	Observation group (n = 267)	<i>p</i>
<b>Female</b>				
Female age (years)	30.08 ± 3.29	30.00 ± 3.40	30.00 ± 3.24	0.396
Infertility duration (years)	2.00 [1.00, 5.00]	2.30 [1.00, 4.00]	2.00 [1.00, 5.00]	0.860
Female BMI (kg/m <sup>2</sup> )	22.31 [20.50, 24.22]	22.35 [20.78, 24.02]	22.31 [20.30, 24.43]	0.838
AFC	18.00 [13.00, 27.00]	19.00 [15.00, 30.00]	18.00 [13.00, 26.00]	0.034
AMH	4.07 [2.88, 5.95]	4.24 [3.10, 6.78]	4.03 [2.80, 5.70]	0.237
Basal FSH levels (IU/L)	6.89 [5.95, 8.02]	6.78 [6.02, 8.02]	6.92 [5.90, 7.99]	0.668
Basal LH levels (IU/L)	4.25 [3.10, 5.66]	4.49 [3.32, 5.66]	4.19 [3.01, 5.57]	0.116
Basal E2 levels (pg/mL)	31.00 [24.00, 40.00]	32.00 [24.00, 38.00]	31.00 [24.00, 41.00]	0.767
LH of hCG day (IU/L)	2.40 [1.48, 3.71]	2.22 [1.56, 3.54]	2.43 [1.48, 3.77]	0.621
E2 of hCG day (pg/mL)	2588.00 [1736.00, 4362.00]	2945.00 [1766.25, 4727.50]	2547.00 [1695.50, 4088.00]	0.186
P of hCG day (ng/mL)	1.21 [0.87, 1.66]	1.20 [0.83, 1.71]	1.21 [0.88, 1.63]	0.848
No. of oocytes retrieved	8.00 [7.00, 12.00]	9.00 [6.25, 12.00]	8.00 [7.00, 12.00]	0.518
MII oocytes retrieved	8.00 [6.00, 11.00]	9.00 [6.00, 11.00]	8.00 [6.00, 11.00]	0.845
Total gonadotropin dose (IU)	1800.00 [1575.00, 2100.00]	1800.00 [1537.50, 2025.00]	1800.00 [1575.00, 2100.00]	0.409
No. of days of stimulation (days)	8.00 [8.00, 9.00]	8.00 [8.00, 9.00]	8.00 [7.00, 9.00]	0.847
<b>Male</b>				
Male age (years)	30.81 ± 4.02	30.00 ± 4.05	30.00 ± 4.01	0.507
Sperm DFI	12.00 [6.00, 22.00]	16.25 [10.25, 25.75]	10.00 [5.00, 20.00]	<0.001
Male height (cm)	173.51 ± 4.92	174.59 ± 4.87	173.08 ± 4.89	0.007
Male weight (kg)	74.60 ± 10.70	76.85 ± 10.76	73.70 ± 10.57	0.007
Male BMI	24.75 ± 3.00	25.17 ± 2.97	24.58 ± 3.01	0.087
Sperm volume before DGC (mL)	2.50 [2.00, 3.50]	2.55 [2.00, 3.50]	2.50 [2.00, 3.50]	0.796
Sperm concentration before DGC (×10 <sup>6</sup> /mL)	75.00 [48.00, 108.00]	72.00 [48.25, 106.75]	76.00 [48.00, 108.00]	0.634
Progressive motility before DGC (%)	44.00 [36.00, 52.00]	44.00 [35.00, 53.50]	44.00 [36.00, 52.00]	0.923
Total sperm before DGC (×10 <sup>6</sup> )	182.00 [110.00, 285.00]	184.00 [118.50, 286.00]	176.00 [105.00, 278.00]	0.670
Total motile sperm count before DGC (×10 <sup>6</sup> )	78.94 [45.00, 125.19]	82.72 [52.26, 120.96]	76.14 [42.28, 125.42]	0.558
Sperm volume after DGC (mL)	1.00 [0.80, 1.00]	1.00 [0.85, 1.00]	1.00 [0.80, 1.00]	0.960
Sperm concentration after DGC (×10 <sup>6</sup> /mL)	10.65 [10.00, 11.60]	10.55 [9.90, 11.66]	10.70 [10.10, 11.57]	0.475
Progressive motility after DGC (%)	98.00 [98.00, 98.00]	98.00 [98.00, 98.00]	98.00 [98.00, 98.00]	0.600
Total sperm after DGC (×10 <sup>6</sup> )	10.85 [7.76, 12.60]	10.88 [7.76, 12.79]	10.80 [8.16, 12.53]	0.900
Total motile sperm count after DGC (×10 <sup>6</sup> )	10.63 [7.53, 12.35]	10.66 [7.53, 12.53]	10.58 [8.00, 12.28]	0.871
Sperm extraction yield (%)	11.91 [8.00, 17.89]	12.74 [6.82, 17.68]	11.69 [8.15, 18.20]	0.237
Male uric Acid (μmol/L)	410.00 [351.00, 456.00]	414.50 [365.45, 449.00]	410.00 [350.50, 459.00]	0.491
Male GGT (U/L)	31.00 [21.00, 47.00]	32.50 [23.25, 47.75]	30.00 [21.00, 47.00]	0.303
Male cholinesterase (U/L)	9.67 [8.45, 10.84]	9.68 [8.54, 10.83]	9.64 [8.41, 10.82]	0.915
Male FBG (mmol/L)	5.38 [5.20, 5.40]	5.38 [5.10, 5.40]	5.38 [5.20, 5.40]	0.810
Infertility Types				0.633
Primary Infertility	218.00 (58.45)	64.00 (60.38)	154.00 (57.68)	
Secondary Infertility	155.00 (41.55)	42.00 (39.62)	113.00 (42.32)	
Fertilization Technique				0.464
IVF	267.00 (71.58)	73.00 (68.87)	194.00 (72.66)	
ICSI	106.00 (28.42)	33.00 (31.13)	73.00 (27.34)	
Mycoplasma				<0.001
Negative	275.00 (73.73)	52.00 (49.06)	223.00 (83.52)	
Positive	98.00 (26.27)	54.00 (50.94)	44.00 (16.48)	

BMI, Body Mass Index; AFC, Antral Follicle Count; AMH, Anti-Mullerian Hormone; FSH, Follicle-Stimulating Hormone; LH, Luteinizing Hormone; E2, Estradiol; hCG, Human Chorionic Gonadotropin; P, Progesterone; MII, Metaphase II; DFI, DNA Fragmentation Index; DGC, Density-Gradient Centrifugation; GGT, Gamma-glutamyl transferase; FBG, Fasting Blood Glucose; IVF, *in vitro* fertilization; ICSI, intracytoplasmic sperm injection.

**Table 2. Factors affecting the rate of high-quality embryos (univariate logistic regression analysis).**

Variable	$\beta$	$p$	OR (95% CI)
Female age (years)	0.03	0.430	1.03 (0.96, 1.10)
Male age (years)	-0.03	0.381	0.98 (0.92, 1.03)
AFC	-0.02	0.039	0.98 (0.96, 1.00)
AMH	-0.06	0.164	0.94 (0.87, 1.03)
Basal LH levels (IU/L)	-0.08	0.048	0.93 (0.86, 1.00)
E2 of hCG day (pg/mL)	0.00	0.124	1.00 (1.00, 1.00)
No. of oocytes retrieved	-0.03	0.249	0.97 (0.92, 1.02)
MII oocytes retrieved	-0.02	0.422	0.98 (0.92, 1.04)
Total gonadotropin dose (IU)	0.00	0.322	1.00 (1.00, 1.00)
Sperm DFI	-0.02	0.027	0.98 (0.97, 1.00)
Male height (cm)	-0.05	0.020	0.95 (0.90, 0.99)
Male weight (kg)	-0.02	0.020	0.98 (0.96, 1.00)
Male BMI	-0.06	0.116	0.94 (0.88, 1.01)
Sperm concentration before DGC ( $\times 10^6$ /mL)	0.00	0.293	1.00 (0.99, 1.00)
Total sperm before DGC ( $\times 10^6$ )	0.00	0.263	1.00 (1.00, 1.00)
Sperm volume after DGC (mL)	0.20	0.353	1.22 (0.81, 1.89)
Total sperm after DGC ( $\times 10^6$ )	0.01	0.422	1.01 (0.98, 1.05)
Total motile sperm count after DGC ( $\times 10^6$ )	0.02	0.400	1.02 (0.98, 1.05)
Sperm extraction yield (%)	0.02	0.169	1.02 (1.00, 1.05)
Fertilization Technique	-0.18	0.464	0.83 (0.51, 1.37)
Mycoplasma	-1.66	0.000	0.19 (0.11, 0.31)

**Table 3. Factors affecting the rate of high-quality embryos (multivariate logistic regression analysis).**

Variable	$\beta$	Standard Error	Z	$p$	OR (95% CI)
Intercept	3.60	0.89	4.04	<0.001	36.43 (6.36, 208.57)
Basal LH levels (IU/L)	-0.08	0.04	-1.99	0.046	0.92 (0.85, 0.99)
Male weight (kg)	-0.02	0.01	-2.08	0.046	0.98 (0.96, 0.99)
Mycoplasma	-1.66	0.26	-6.42	<0.001	0.19 (0.11, 0.32)

**Table 4. Comparison of prediction performance for different machine learning models.**

Evaluation Indicators	SVM	Decision Tree	Random Forest
MSE	0.12	0.13	0.10
RMSE	0.35	0.35	0.32
MAE	0.28	0.28	0.26

The average results of randomly splitting the training set and validation set 100 times. SVM, Support Vector Machine; MSE, Mean Square Error; RMSE, Root Mean Square Error; MAE, Mean Absolute Error.

### 3.3 The Process of Formulating Intervention Strategies Based on Causal Effect Estimation

Based on the causal effect estimation, we have developed a personalized intervention strategy for male infertility patients to improve their high-quality embryo rates and determine the most effective intervention. This is a counterfactual prediction problem, where we are interested in knowing the potential increase in high-quality embryo rates if the patient receives a certain intervention. To achieve this, we constructed a predictive model for outcomes and a causal effect estimation model for counterfactual predictions.

First, we preprocessed the collected patient data, including imputing missing values of continuous variables with mean values and excluding outliers. Then, we divided the dataset into a training set (80%) and a test set (20%) for building and validating the outcome prediction model and the causal effect estimation model. We evaluated the predictive model using common regression model evaluation metrics such as MSE, RMSE, and MAE. Each model underwent parameter optimization, testing 3 commonly used machine learning models: Support Vector Machine (SVM), Decision Tree, and Random Forest. Table 4 shows the evaluation results of each model after running 100 times. It can be observed that the Random Forest model outperformed

**Table 5. Performance evaluation of different causal effect estimation models.**

Evaluation Indicators	Causal Forest DML	NonParam DML	Linear DML
MSE	0.07	0.07	0.08
RMSE	0.27	0.27	0.27
MAE	0.23	0.22	0.23

DML, double machine learning.

the other 2 models in predictive performance, hence we selected the Random Forest model as the final outcome prediction model (Table 4). The evaluation of the causal effect model mainly involved analyzing model features through the SHAP method and plotting marginal causal effect diagrams to confirm clinical consistency. By analyzing the high-quality embryo rates through the predictive model and the causal effect estimation model, targeted intervention measures were formulated. The causal effect estimation employed a DML method, which is a method for estimating heterogeneous intervention effects. We compared 3 commonly used DML methods, and the results are shown in Table 5. It can be observed that the non-parametric model (NonParam DML) performed the best. Compared to the outcome prediction model, there was a certain improvement in performance, which also confirms that causal models are more stable and reliable. We selected modifiable male factors significantly affecting the high-quality embryo rate from the univariate logistic regression analysis, namely sperm DFI, male body weight, and mycoplasma infection, and constructed causal effect estimation models with these indicators as intervention variables. First, the high-quality embryo rate was predicted based on these indicators, and the need for intervention was determined according to the magnitude of the high-quality embryo rate. When intervention is needed, the causal effects of potential intervention strategies, that is, the changes in the high-quality embryo rate, are assessed one by one based on the causal effect estimation model, providing a reference for clinicians to formulate targeted intervention measures (Fig. 1).

### 3.4 Test Results for Intervention Variable of Mycoplasma

The rate of high-quality embryos is significantly higher in mycoplasma-negative individuals than in mycoplasma-positive individuals (75.0% vs 33.3%,  $p = 0.000$ , Fig. 2A). When the intervention variable is mycoplasma and the outcome variable is the rate of high-quality embryos, the heterogeneity-adjusted covariates include modifiable indicators among male factors, such as sperm concentration before DGC, progressive motility, sperm volume, total sperm count, total motile sperm count, sperm DFI, sperm extraction yield, fertilization technique, and other female factors, as well as other covariates among male factors that cannot be changed in the short term. In the collected data, 23.3% of patients underwent treatment for mycoplasma positivity, meaning their mycoplasma test results changed from positive to negative after intervention,

and this group was used to validate the accuracy of the model's causal effect estimation. The results showed that sperm DFI, the proportion of PR sperm before DGC, and male weight are the 3 factors that have a greater impact on the intervention effect, with sperm DFI and male weight generally having a negative correlation with the intervention effect of mycoplasma, and the proportion of PR sperm before DGC generally having a positive correlation with the intervention effect of mycoplasma (Fig. 2B). Fig. 2C shows the marginal causal average effect for mycoplasma-positive patients in the test set, which are almost all negative values, indicating that when patients change from mycoplasma-positive to negative, the rate of high-quality embryos increases.

### 3.5 Test Results for Intervention Variable of Sperm DFI

We further analyzed the relationship between sperm DFI and the rate of high-quality embryos and found no clear linear relationship (Fig. 3A). When the intervention variable is sperm DFI, the factors that have a significant impact on the intervention effect include the total sperm before DGC, sperm extraction yield, and male weight (Fig. 3B). Fig. 3C shows the marginal effect for patients in the test set with DFI >25, indicating that there are differences in the marginal effects among different patients, but the vast majority are negative values, meaning that reducing DFI would increase the rate of high-quality embryos. However, the improvement in the high-quality embryo rate from reducing sperm DFI is minimal, with the maximum increase being only 0.3% (Fig. 3B). This finding is consistent with the results of the multivariate logistic regression analysis, which attenuated the impact of sperm DFI on the high-quality embryo rate.

### 3.6 Test Results for Intervention Variable of Male Weight

We analyzed the relationship between male weight and the rate of high-quality embryos, and observed no clear linear relationship. However, it was noted that patients with higher body weights tend to have relatively lower rates of high-quality embryos (Fig. 4A). When the intervention variable is male weight, the factors that significantly impact the intervention effect are sperm extraction yield, sperm DFI, and total motile sperm count before DGC (Fig. 4B). Fig. 4C illustrates the marginal causal effect of body weight for patients who are overweight (>80 kg), showing that for the majority, the effect values are negative, indicating that being overweight reduces the rate of high-quality embryos.

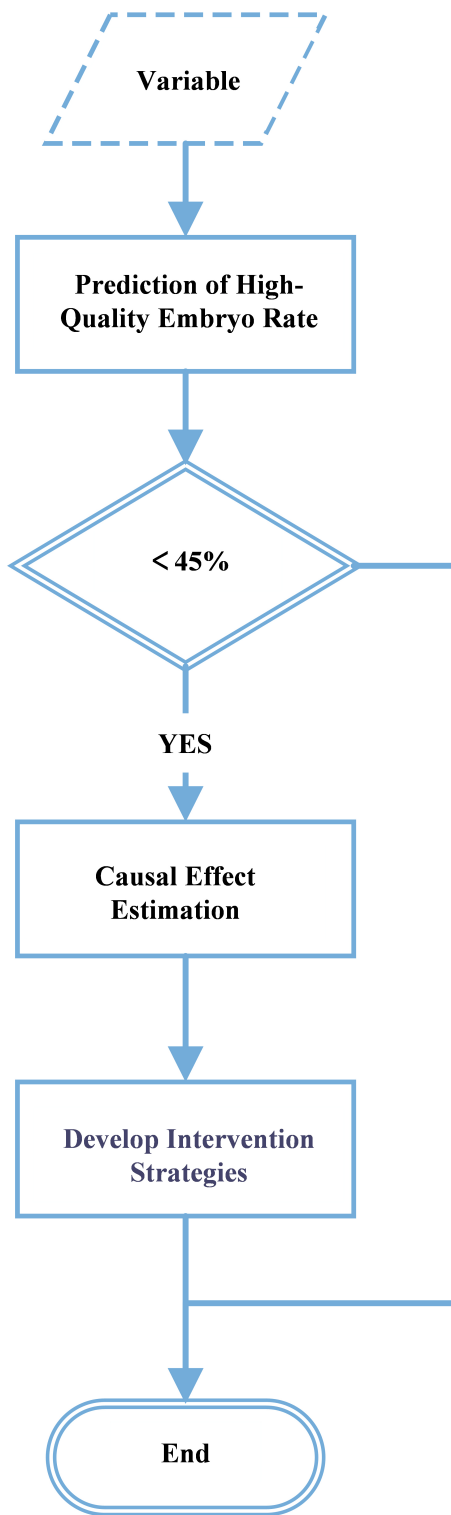


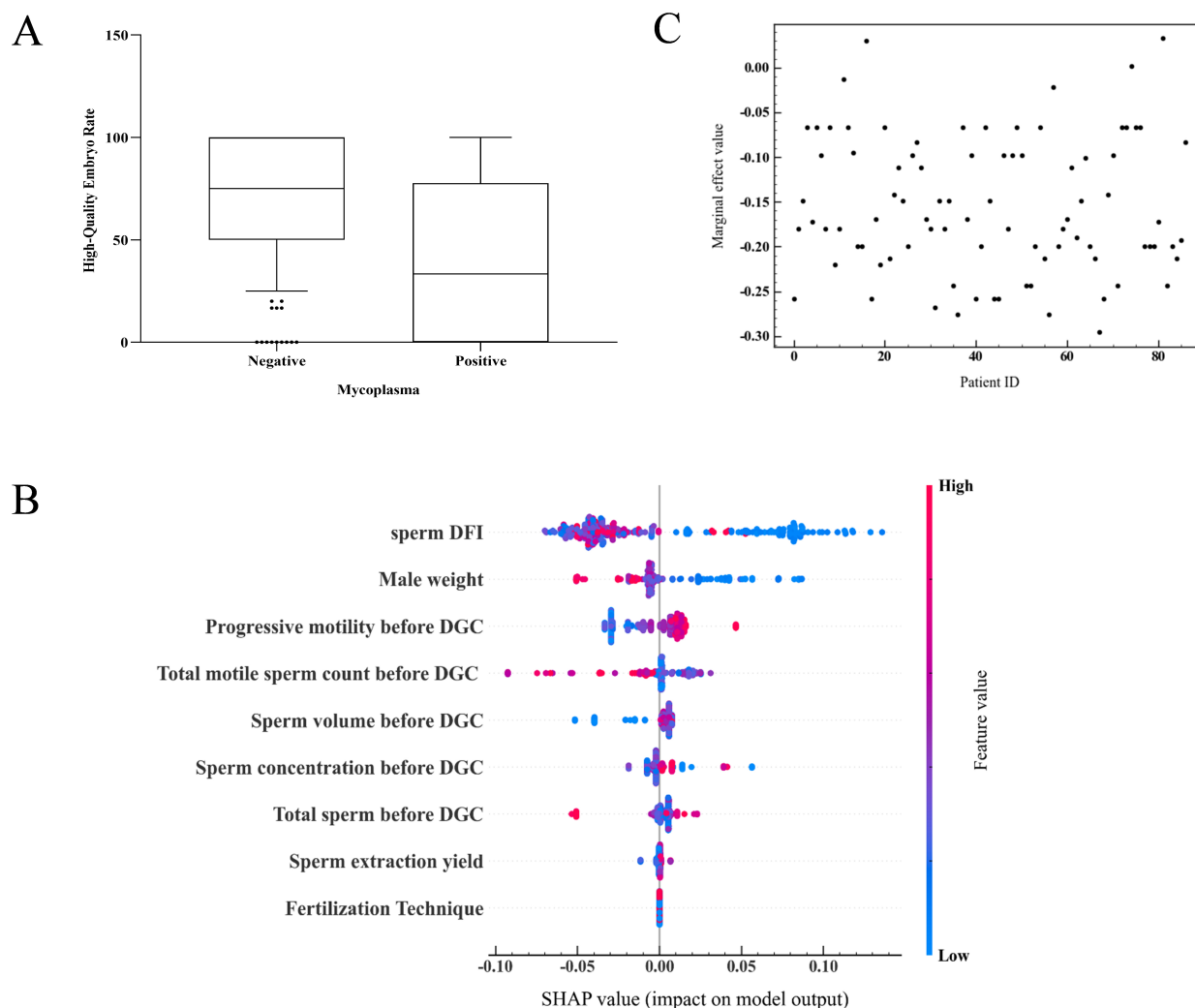
Fig. 1. Intervention strategy formulation process.

#### 4. Discussion

In IVF cycles, female factors play a significant role in embryo development, particularly the impact of egg quality on embryo development. The influence of male factors and sperm quality on embryo development has been less studied. Pellegrini *et al.* [17] found the relationship between

male sperm quality and embryo quality in donor egg IVF is that poor sperm quality is associated with changes in morphokinetic parameters of embryos on day 3 after donor egg IVF, emphasizing the important role of sperm in embryo development. Our study aimed to explore male factors affecting the rate of good embryos and to develop intervention strategies based on causal effect estimation. Through the analysis of data from 373 infertile couples, we found that male factors such as sperm DFI, male height, weight, and mycoplasma infection status are all significantly correlated with the rate of good embryos.

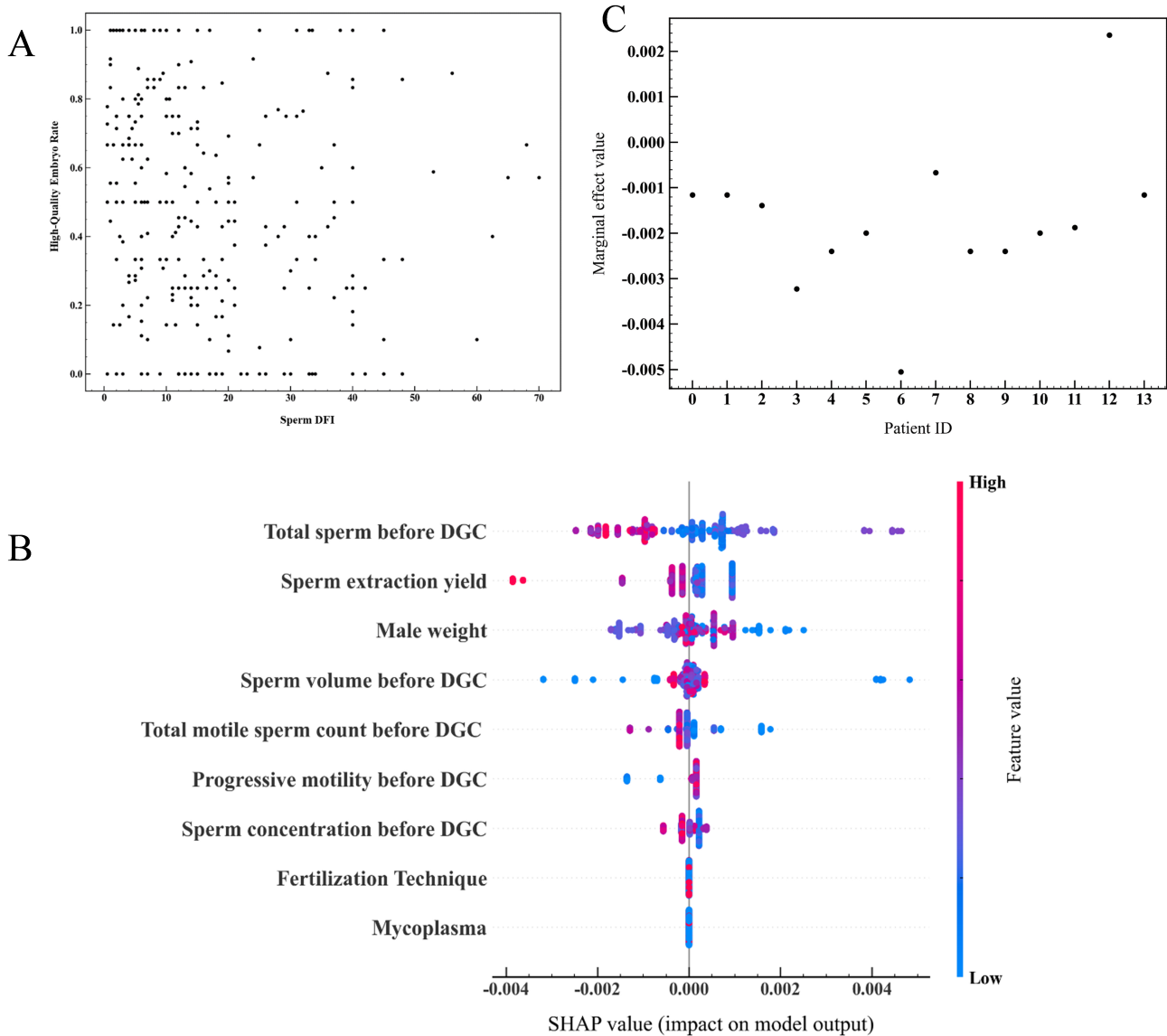
Firstly, we observed that in the group with a higher rate of good embryos, the sperm DFI in males was lower. Sperm DFI, as an important indicator of sperm genetic quality, is associated with improved embryo quality when reduced. This is consistent with previous research findings. Jiang *et al.* [18] found that DFI and sperm morphology affect the blastocyst formation rate and the rate of good blastocyst formation on day 6. A meta-analysis by Kaiyal *et al.* [5] also found that the fertilization rate and embryo cleavage rate in the group with high sperm DFI were significantly lower than those in the low DFI group. Our results show that the weight of males in the high-quality embryo group was lower. Liu *et al.* [19] studied 6569 first fresh *in vitro* fertilization-embryo transfer (IVF-ET) cycles and found that the number of available embryos and high-quality embryos was significantly reduced in the overweight/obese female group and the mixed overweight/obese male and female group compared to the normal weight group. A recent meta-analysis indicated that male obesity has a negative impact on basic sperm parameters, such as sperm count, concentration, and motility, increasing the incidence of infertility and reducing pregnancy rates [6]. Although there is controversy in human studies [6,19] due to significant heterogeneity between studies and the use of body mass index as an indicator of obesity, the relationship between male obesity and sperm DNA damage is controversial, but evidence from rodent models clearly shows that male obesity increases sperm DNA damage [20]. A meta-analysis indicated that being overweight and/or obese is associated with lower sperm quality, such as sperm volume, sperm count and concentration, sperm motility, total motility and normal morphology [21]. Additionally, there is an increased proportion of sperm with low mitochondrial membrane potential (MMP) in obese men, and the proportion of DNA fragments and abnormal morphology has also increased [6]. These all suggest that obesity may affect pregnancy outcomes by affecting sperm DFI. Genital tract mycoplasma can affect the health of fetuses and newborns. However, the sperm optimization process in current ART treatments cannot ensure the eradication of mycoplasma. It is beneficial to avoid using semen contaminated or infected with mycoplasma in ART treatments, and routine screening for mycoplasma before ART treatment is advisable. Our results show that the proportion of positive mycoplasma in



**Fig. 2. Patients with positive mycoplasma have a lower rate of high-quality embryos, and treating mycoplasma infections can improve the rate of high-quality embryos.** (A) High-quality embryo rate in patients with different mycoplasma test results (75.0% vs 33.3%,  $p = 0.000$ ). (B) Causal estimation model: SHAP explanation diagram of the impact of each covariate on the intervention effect. (C) Marginal causal effect of intervention in patients with mycoplasma positive. DFI, DNA Fragmentation Index; DGC, Density-Gradient Centrifugation; SHAP, SHapley Additive exPlanations.

the low-quality embryo group was significantly increased. In a study of 306 couples undergoing IVF treatment, 32% of semen samples were found to have ureaplasma urealyticum infection, with similar infection rates at the prostate and seminal vesicle levels, and a significant decrease in pregnancy rates after embryo transfer in the infected group [22]. In couples infected with mycoplasma who received specific antibiotic treatment to eradicate the infection, improvements in semen quality and better ART outcomes were also observed [23]. Therefore, routine screening for genital tract mycoplasma can improve semen quality, thereby simplifying IVF procedures and increasing the success rates of embryo implantation and pregnancy [8]. Secondly, our univariate logistic regression analysis showed that male sperm DFI, male height, weight, and mycoplasma infection were all negatively correlated with the rate of high-quality em-

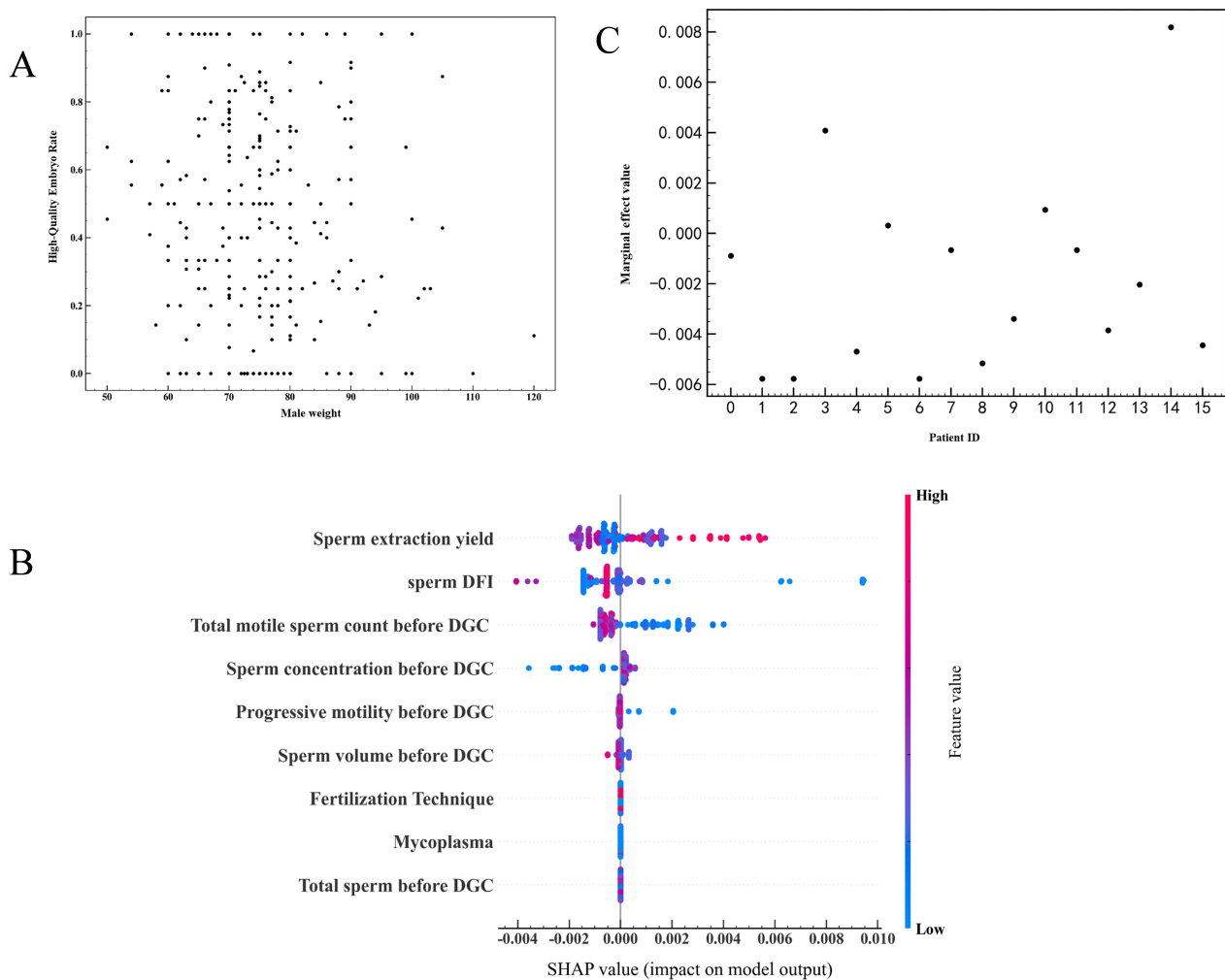
bryos. The identification of these factors provides potential targets for further intervention. However, further multivariate logistic regression analysis found that while sperm DFI was significant in the univariate analysis, it was not significant in the multivariate analysis. This may be due to collinearity between sperm DFI and other variables (such as mycoplasma positivity), or its effect being diminished when considering other variables. In contrast, mycoplasma infection and male weight were significant in both univariate and multivariate analyses, indicating that they have an independent significant impact on the high-quality embryo rate. In particular, a negative status of mycoplasma infection is associated with a higher rate of high-quality embryos, which further suggests the importance of screening and treatment of infection status in infertility treatment.



**Fig. 3. Reducing DFI would increase the rate of high-quality embryos.** (A) Correlation between sperm DFI and high-quality embryo rate. (B) SHAP explanation diagram of the causal effect model with intervention factor as sperm DFI. (C) Marginal causal effect of intervention in patients with sperm DFI >25.

Causal effect estimation is a science based on observational data to perform counterfactual estimation and analyze the causal relationship between interventions and outcomes. It attempts to answer who is the cause and who is the effect, and how much of an effect the cause will have. In practical applications, causal effect estimation can help us understand the extent to which specific interventions affect outcomes [24]. Therefore, based on the above results, we constructed an outcome prediction model and a causal effect estimation model to evaluate the effects of different intervention measures according to Aaron Baum's method [25]. By comparing 3 commonly used machine learning models (SVM, Decision Tree, and Random Forest), we found that the Random Forest model outperformed the other 2 models in predictive performance. Therefore,

we chose the Random Forest model as the final outcome prediction model. At the same time, we also used the two-machine learning method to estimate heterogeneous intervention effects and identified modifiable male factors that have a significant impact on the rate of good embryos. In medical research, the same treatment may have significantly different effects on populations with different characteristics, that is, there is heterogeneity in treatment effects. Causal forest analysis can effectively address the limitations of traditional analysis methods in testing treatment effect heterogeneity and provide more accurate heterogeneity assessment. Through this analysis, researchers can identify the response of different patient groups to specific treatments, thereby providing more customized treatment plans for patients [26]. We found through causal for-



**Fig. 4. Being overweight reduces the rate of high-quality embryos.** (A) Relationship between male weight and high-quality embryo rate. (B) SHAP explanation diagram of the causal effect model with intervention factor as male body weight. (C) Marginal causal effect of intervention in patients over 80 kg.

est analysis that sperm DFI, male weight, and mycoplasma infection are modifiable factors that significantly affect the rate of high-quality embryos. This finding provides a basis for targeted interventions for clinicians. In the intervention analysis of mycoplasma infection, we found that the rate of high-quality embryos in the mycoplasma-negative population was significantly higher than that in the positive population. In addition, sperm DFI, the proportion of PR sperm before DGC, and male weight are factors that have a greater impact on the intervention effect. These results indicate that treating mycoplasma infections can increase the rate of high-quality embryos, which is of great significance for clinical practice. In the intervention analysis of sperm DFI, we found that reducing DFI can increase the rate of high-quality embryos. This finding is consistent with existing literature on the relationship between sperm DFI and embryo quality [27]. However, the effect of reducing sperm DFI on improving the high-quality embryo rate is minimal. This may be because sperm DFI is often tested before ART

and may not truly reflect the sperm DFI levels on the day of fertilization. Further analysis of sperm on the day of fertilization is needed to determine the true impact of sperm DFI on the high-quality embryo rate. Moreover, we found that a decrease in male weight is also associated with an increase in the rate of high-quality embryos, further emphasizing the importance of lifestyle interventions in infertility treatment.

## 5. Conclusion

In summary, our research results emphasize the importance of assessing male factors in infertility treatment. By identifying and intervening in key factors such as sperm DFI, male weight, and mycoplasma infection status, the rate of high-quality embryos can be increased, thereby increasing positive fertility outcomes for infertile couples. However, it should be noted that this study only analyzed a specific group of infertile couples, and further studies are needed in the future to verify the applicability of these findings in other populations, as well as to further explore the

intervention strategies of these factors and how they interact with other factors not considered in this study. Our study was conducted at a single center, which may introduce selection bias and limit the generalizability of our results. A multi-center design should be considered to account for potential variations in patient populations and treatment protocols. In conclusion, this study provides clinicians with a process for formulating intervention strategies based on causal effect estimation, which helps to improve the rate of good embryos in infertile patients.

### Availability of Data and Materials

The datasets that support the findings of this study are available from the corresponding author (QW) upon reasonable written request.

### Author Contributions

QW designed the research study; XS, LP, XG, YC, MY, and JS were responsible for data acquisition; QW and QY performed the research and analyzed the data; XS, LP, XG, YC, MY and JS provided help and advice on the research. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

### Ethics Approval and Consent to Participate

The study was carried out in accordance with the guidelines of the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Affiliated Hospital of Nantong University (2025-K037-01). All patients or their families/legal guardians gave their informed consent for inclusion before they participated in the study.

### Acknowledgment

Not applicable.

### Funding

This work was funded by the scientific research project of Shanghai Qingpu District Health Commission (QWJ2022-06).

### Conflict of Interest

The authors declare no conflict of interest.

### Declaration of AI and AI-Assisted Technologies in the Writing Process

We thank Kimi for the suggestions on the outline design of the article and the help with language polishing. The research design, data analysis, and interpretation of results were all completed by the authors.

## References

- [1] Eisenberg ML, Esteves SC, Lamb DJ, Hotaling JM, Giwercman A, Hwang K, *et al.* Male infertility. *Nature Reviews. Disease Primers.* 2023; 9: 49. <https://doi.org/10.1038/s41572-023-00459-w>.
- [2] Ziouziou I, Rambhatla A, Shah R, Agarwal A. Sperm DNA fragmentation and infertility: a narrative review. *World Journal of Urology.* 2024; 42: 408. <https://doi.org/10.1007/s00345-024-05090-2>.
- [3] Bashiri Z, Amidi F, Amiri I, Zandieh Z, Maki CB, Mohammadi F, *et al.* Male Factors: the Role of Sperm in Preimplantation Embryo Quality. *Reproductive Sciences.* 2021; 28: 1788–1811. <https://doi.org/10.1007/s43032-020-00334-z>.
- [4] Adler A, Roth B, Lundy SD, Takeshima T, Yumura Y, Kuroda S. Sperm DNA fragmentation testing in clinical management of reproductive medicine. *Reproductive Medicine and Biology.* 2023; 22: e12547. <https://doi.org/10.1002/rmb2.12547>.
- [5] Kaiyal RS, Karna KK, Kuroda S, Sgayer I, Shlush E, Vij SC, *et al.* Sperm chromatin dispersion assay reliability and assisted reproductive technology outcomes: Systematic review and meta-analysis. *Andrology.* 2025; 13: 718–730. <https://doi.org/10.1111/andr.13725>.
- [6] Campbell JM, Lane M, Owens JA, Bakos HW. Paternal obesity negatively affects male fertility and assisted reproduction outcomes: a systematic review and meta-analysis. *Reproductive Biomedicine Online.* 2015; 31: 593–604. <https://doi.org/10.1016/j.rbmo.2015.07.012>.
- [7] Hoek A, Wang Z, van Oers AM, Groen H, Cantineau AEP. Effects of preconception weight loss after lifestyle intervention on fertility outcomes and pregnancy complications. *Fertility and Sterility.* 2022; 118: 456–462. <https://doi.org/10.1016/j.fertnstert.2022.07.020>.
- [8] Veiga E, Treviño M, Romay AB, Navarro D, Trastoy R, Macía M. Prevalence of genital Mycoplasma and response to eradication treatment in patients undergoing assisted reproductive techniques. *Revista Espanola De Quimioterapia.* 2019; 32: 327–332.
- [9] Taniguchi R, Hatakeyama S, Ohgi S, Yanaihara A. Effect of Male Cigarette Smoking on In Vitro Fertilization (IVF) Outcomes and Embryo Morphokinetic Parameters. *Cureus.* 2024; 16: e52788. <https://doi.org/10.7759/cureus.52788>.
- [10] Andrea S, Carla DD, Cristina DA, Davide M, Carlotta P, Rosario P, *et al.* Smoke, alcohol and drug addiction and male fertility. *Reproductive Biology and Endocrinology.* 2018; 16: 3. <https://doi.org/10.1186/s12958-018-0320-7>.
- [11] Rotimi DE, Singh SK. Implications of lifestyle factors on male reproductive health. *JBRA Assisted Reproduction.* 2024; 28: 320–330. <https://doi.org/10.5935/1518-0557.20240007>.
- [12] Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HWG, Behre HM, *et al.* World Health Organization reference values for human semen characteristics. *Human Reproduction Update.* 2010; 16: 231–245. <https://doi.org/10.1093/humupd/dmp048>.
- [13] Edwards RG. The history of assisted human conception with especial reference to endocrinology. *Experimental and Clinical Endocrinology & Diabetes.* 1996; 104: 183–204. <https://doi.org/10.1055/s-0029-1211443>.
- [14] Wintner EM, Hershko-Klement A, Tzadikévitch K, Ghetler Y, Gonen O, Wintner O, *et al.* Does the transfer of a poor quality embryo together with a good quality embryo affect the In Vitro Fertilization (IVF) outcome? *Journal of Ovarian Research.* 2017; 10: 2. <https://doi.org/10.1186/s13048-016-0297-9>.
- [15] Wang QX, Wang X, Yu MY, Sun H, Wang D, Zhong SP, *et al.* Random sperm DNA fragmentation index is not associated with clinical outcomes in day-3 frozen embryo transfer. *Asian Journal of Andrology.* 2022; 24: 109–115. <https://doi.org/10.4103/aja.aj>

a\_17\_21.

- [16] Vaiarelli A, Zacà C, Spadoni V, Cimadomo D, Conforti A, Alviggi C, *et al.* Clinical and laboratory key performance indicators in IVF: A consensus between the Italian Society of Fertility and Sterility and Reproductive Medicine (SIFES-MR) and the Italian Society of Embryology, Reproduction and Research (SIERR). *Journal of Assisted Reproduction and Genetics*. 2023; 40: 1479–1494. <https://doi.org/10.1007/s10815-023-02792-1>.
- [17] Pellegrini L, Gatti S, Navarro N, Hervas I, Marcos M, Viviana V, *et al.* The effect of male factors on embryo morphokinetics: a retrospective analysis of 2726 blastocysts. *Journal of Assisted Reproduction and Genetics*. 2024; 41: 3467–3475. <https://doi.org/10.1007/s10815-024-03275-7>.
- [18] Jiang W, Zhu L, Gao L, Wang J, Fang J, Chen L, *et al.* What sperm parameters effect blastocyst formation and quality during ICSI with severe male infertility. *Systems Biology in Reproductive Medicine*. 2024; 70: 218–227. <https://doi.org/10.1080/19396368.2024.2375710>.
- [19] Liu X, Shi S, Sun J, He Y, Zhang Z, Xing J, *et al.* The influence of male and female overweight/obesity on IVF outcomes: a cohort study based on registration in Western China. *Reproductive Health*. 2023; 20: 3. <https://doi.org/10.1186/s12978-022-01558-9>.
- [20] Peel A, Saini A, Deluao JC, McPherson NO. Sperm DNA damage: The possible link between obesity and male infertility, an update of the current literature. *Andrology*. 2023; 11: 1635–1652. <https://doi.org/10.1111/andr.13409>.
- [21] Salas-Huetos A, Maghsoumi-Norouzabad L, James ER, Carrell DT, Aston KI, Jenkins TG, *et al.* Male adiposity, sperm parameters and reproductive hormones: An updated systematic review and collaborative meta-analysis. *Obesity Reviews*. 2021; 22: e13082. <https://doi.org/10.1111/obr.13082>.
- [22] Montagut JM, Leprêtre S, Degoy J, Rousseau M. Ureaplasma in semen and IVF. *Human Reproduction*. 1991; 6: 727–729. <https://doi.org/10.1093/oxfordjournals.humrep.a137416>.
- [23] Toth A, Lesser ML, Brooks C, Labriola D. Subsequent pregnancies among 161 couples treated for T-mycoplasma genital-tract infection. *The New England Journal of Medicine*. 1983; 308: 505–507. <https://doi.org/10.1056/NEJM198303033080907>.
- [24] Chen J, Zhou Y. Causal effect estimation for multivariate continuous treatments. *Biometrical Journal. Biometrische Zeitschrift*. 2023; 65: e2200122. <https://doi.org/10.1002/bimj.202200122>.
- [25] Baum A, Scarpa J, Bruzelius E, Tamler R, Basu S, Faghmous J. Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: a machine learning-based post-hoc analysis of heterogeneous treatment effects in the Look AHEAD trial. *The Lancet. Diabetes & Endocrinology*. 2017; 5: 808–815. [https://doi.org/10.1016/S2213-8587\(17\)30176-6](https://doi.org/10.1016/S2213-8587(17)30176-6).
- [26] Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Briefings in Bioinformatics*. 2023; 24: bbad002. <https://doi.org/10.1093/bib/bbad002>.
- [27] Casanovas A, Ribas-Maynou J, Lara-Cerrillo S, Jimenez-Macedo AR, Hortal O, Benet J, *et al.* Double-stranded sperm DNA damage is a cause of delay in embryo development and can impair implantation rates. *Fertility and Sterility*. 2019; 111: 699–707.e1. <https://doi.org/10.1016/j.fertnstert.2018.11.035>.