International Journal of Mathematics Trends and Technology
ISSN: 2231-5373/ https://doi.org/10.14445/22315373/IJMTT-V69I4P503

Volume 69 Issue 4, 17-28, April 2023
© 2023 Seventh Sense Research Group®

Original Article

Study on the Number of Oocytes Retrieved Based on Multinomial Logistic Regression Model

Xi Wang

School of Mathematics and Information Science, Henan Polytechnic University, Henan, China

Received: 26 February 2023 Revised: 01 April 2023 Accepted: 13 April 2023 Published: 28 April 2023

Abstract - *This paper selects the clinical data of patients receiving assisted reproductive therapy from 2019 to 2021 in the department of Obstetrics and Gynecology of a hospital in Jiaozuo City. We screen 1386 patients eligible for single blastocyst transplantation. According to the method of fertilization, patients are divided into natural fertilization group consisting of 1167 patients and an artificial fertilization group consisting of 219 patients. Based on existing studies, we preliminarily identify 5 covariables to explain the number of oocytes retrieved. Then the likelihood ratio test is used to screen covariates, and the multinomial logistic regression model is used to fit the two groups of data respectively. By comparison, we find that the covariates affecting the number of retrieved oocytes in the two groups are not the same. The results show that the age, duration of infertility, Anti-Müllerian Hormone, and Gonadotropins dosage of natural fertilization patients all have a significant impact on the number of oocytes retrieved, while the number of oocytes retrieved in artificial fertilization patients is only affected by Anti-Müllerian Hormone. The doctor can predict the number of oocytes retrieved by the patient based on the regression model and can optimize the number of oocytes retrieved by adjusting the patient's covariables.*

Keywords - *Multinomial logistic regression, the number of oocytes retrieved, Likelihood ratio test, Assisted Reproductive Technology.*

1. Introduction

Since 1988, when the first test-tube baby was born in China, assisted reproductive medical technology has been developed in China for nearly 40 years. In recent years, many researchers have continued to try new methods and new technologies in clinical practice and gradually break through many technical difficulties, which greatly improve the reproductive outcome of patients. China has currently become the country with the highest cycle of assisted reproductive technology services. In 2016, the total number of assisted reproductive technology service cycles in China exceeded 1 million; by 2020, the total number had exceeded 1.3 million.

Although assisted reproductive technology has solved the problems for many families suffering from infertility, the development of this technology is also facing bottlenecks. For example, the clinical pregnancy rate has hovered around 40 percent, and the live birth rate is only 30 percent. Factors such as an aging population and younger families delaying childbearing will continue to reduce the success rate of assisted reproductive technology. Therefore, covariates such as the patient's own age are also increasingly important subjects for researchers to study when developing diagnosis and treatment technologies. This article takes the number of oocytes retrieved during in vitro blastocyst transfer as the research object and analyzes the relationship between the number of oocytes retrieved by patients and covariates. In order to prove that the number of oocytes retrieved is important to blastocyst transplantation, scholars have carried out a large number of theoretical and empirical studies from different perspectives. Next, we will introduce relevant research results from the aspects of medical techniques and statistical methods.

With the development of frozen embryo transfer technology, cryopreservation of all available embryos for future freeze-thaw embryo transfer has become a mainstream treatment. Controlled superovulation and whole embryo freezing are becoming indispensable techniques in assisted reproductive therapy[1]. It can cause multiple follicles to mature at the same time to increase the number of oocytes retrieved. In this way, more potential embryos can be obtained for later treatment, thus increasing the clinical pregnancy rate[2]. However, there are drawbacks to the technique, such as over-stimulation of the ovaries during treatment[3]. And excessive estrogen can have adverse effects on oocyte quality and endometrium[4]. Sunkara's research suggests that excessive oocytes retrieved may lead to a lower pregnancy rate[5]. In addition, Pia studies have shown that low ovulation in women of the same age can lead to increased cycle cancellation rates and poor pregnancy outcomes[6]. Lu et al. Show that the live birth rate and the cumulative live birth rate of fresh embryo transfer decrease to varying degrees in

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

patients of all ages with fewer than 5 oocytes retrieved[7]. Verberg's findings suggest that low oocytes retrieved lead to a lower success rate of IVF treatment[8]. Too many or too few oocytes retrieved will lead to adverse pregnancy outcomes. Too many oocytes retrieved means that controlled hyperovulation during treatment will cause abnormal oocyte division[9], and it leads to apoptosis of granulosa cells[10, 11]. Clinicians are also increasingly looking to the optimal range of oocytes retrieved during an IVF treatment cycle.

Although most studies focus on the relationship between oocytes retrieved and clinical pregnancy rate, in recent years, researchers have gradually shifted their attention to the relationship between oocytes retrieved and covariates. And increasingly, statistical methods and models are used to analyze results in clinical trials. Many studies have shown that oocytes and their quality gradually decline as women age, which affects the number of retrieved oocytes during treatment. Adva et al. establish a linear regression analysis table and find higher oocytes retrieved in younger females and lower oocytes retrieved in the older female during treatment[12]. In addition to female age, Anti-Müllerian Hormone is an active factor secreted by follicular granule cells[13]. Its indicators have also attracted the attention of many researchers. Xu et al. use Spearman correlation analysis and find that Anti-Müllerian Hormone is an important indicator of individualized ovulation induction therapy, and the number of oocytes retrieved is closely related to the Anti-Müllerian Hormone level[14]. In 2022, Liu et al. used the generalized additive model and found that the Anti-Müllerian Hormone level is positively correlated with the number of oocytes retrieved, while the Follicle-stimulating Hormone level is negatively correlated with the number of oocytes retrieved.

Previous studies established a linear regression model or a nonlinear regression model to analyze the correlation of covariables and reveal the relationship between the number of oocytes retrieved and covariables[15]. On the basis of previous studies, combined with the patient's own covariates, this paper identifies five influencing factors: female age, infertility years, Anti-Müllerian Hormone, Follicle-stimulating Hormone and Gonadotropins dosage. According to the method of fertilization, the patients were divided into a natural fertilization group and an artificial fertilization group. According to the mode of fertilization, the data set is divided into a natural fertilization group of 1167 patients and an artificial fertilization group of 219 patients. We establish the multinomial logistic regression model for the two groups, respectively. Then, we conduct a likelihood test on covariables and establish a multinomial logistic regression model with a better fitting effect after the screening.

The remainder of the paper is organized as follows. Section 2 describes the number of oocytes retrieved by the patient. To describe the relationship between the number of oocytes retrieved and covariates, we introduce the relevant covariates in Section 3 and establish a multinomial logistic regression model. Subsequently, we optimize the model based on likelihood ratio testing. Section 4 concludes with some remarks and prospective research topics.

2. An Overview of the Research Object

We collected clinical data of patients receiving assisted reproductive therapy in the Department of Obstetrics and Gynecology of a hospital in Jiaozuo City from 2019 to 2021. To avoid model bias, we screen data from 1386 patients undergoing their first assist reproduction treatment. We divide these patients according to fertilization mode into a natural fertilization group consisting of 1167 patients and an artificial fertilization group consisting of 219 patients. In order to have an intuitive and clear understanding of the number of oocytes retrieved, we plot scatter plots for the two data sets separately. It can be seen from Figure 1 that the number of oocytes retrieved for patients in the natural fertilization group is mostly concentrated at about 10. But in Figure 2, the number of oocytes retrieved from the artificial fertilization group patients is almost evenly distributed between 5-20, and only a few patients have more than 20 oocytes retrieved.

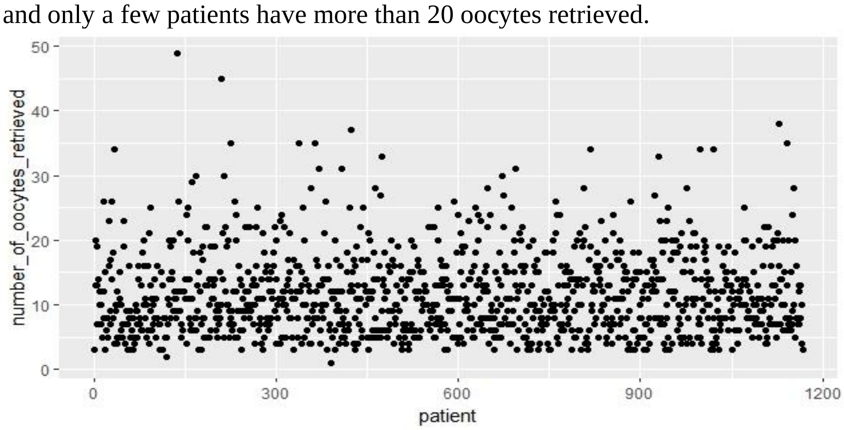

Fig. 1 Scatter plot of patients in the natural fertilization group.

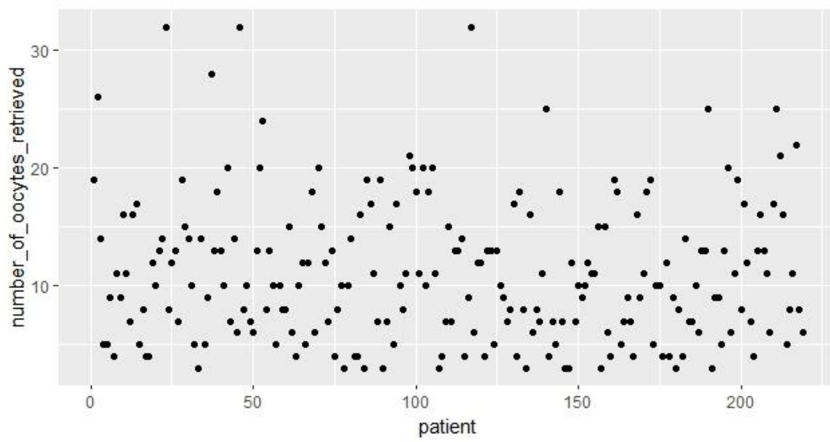

Fig. 2 Scatter plot of patients in the artificial fertilization group.

In order to promote the development of assisted reproductive technologies, it is necessary to apply appropriate statistical methods for the relevant evaluation of the number of oocytes retrieved. In particular, it should shift the research focus to the effect of the patient's own covariates on the treatment outcome, knowing how to control the patient's covariates to achieve the expected number of oocytes retrieved. Only a physician's clarity about these relationships will allow better use of medical resources for patients to achieve optimal outcomes.

3. Discovering the Relationship between the Number of Oocytes Retrieved and Covariates

The number of oocytes retrieved is the key to the success of assisted reproductive technology. Its level has attracted the attention of numerous researchers in clinical practice. Many scholars have described the impact of covariates on the number of oocytes retrieved. Of course, we are also interested in how covariates affect the number of oocytes retrieved. In this section, we will establish a multinomial logistic regression model based on actual clinical data to explain the changes in the number of oocytes retrieved.

3.1 Determination of Covariates

Logistic regression is a generalized linear regression model that models the probability of the dependent variable Y belonging to a category. It is commonly used for disease diagnosis and to explore the risk factors that cause the disease. It connects the dependent variable to one or several independent covariates and quantifies the strength of association between them. In order to ensure the rationality of the model, we select the physical examination index of patients as the covariable of the model. The selection criteria are based on existing literature findings[16] and clinical knowledge. Possible covariables included in this model are female age, infertility years, Follicle-stimulating Hormone (FSH), Anti-Müllerian Hormone (AMH) and gonadotropin (Gn) dose. The specific definition is as follows:

3.1.1. Number of Oocytes Retrieved

At the beginning of a woman's period, doctors will use a B-scan to see how many basal follicles she has. Sufficient oocytes retrieved is a prerequisite for obtaining transplantable embryos. In this paper, the number of oocytes retrieved is taken as the dependent variable, and the relationship between it and other covariables is studied. In clinical practice, the ideal number of oocytes retrieved can be achieved by adjusting the patients' covariables.

3.1.2. Female Age

Because age can directly affect the patient's body function. And as women age, ovarian reserve function declines, and the number of sinus follicles decreases[17]. The frequency of normal ovulation will gradually decrease, leading to a decrease in fertility. So age is a non-negligible covariable. All the statistical ages in this paper are calculated by calendar after birth.

3.1.3. Duration of Infertility

It is calculated from the date when the sexual activity pattern of the couple after marriage is normal and contraceptive measures are not used. Our statistical deadline is until patients seek medical attention.

3.1.4. Follicle-Stimulating Hormone

Its abbreviation is FSH. It is a hormone secreted by basophils in the anterior pituitary gland, mainly composed of glycoproteins, and its main function is to promote follicle maturation[18]. Its content value is related to various diseases and is a factor that cannot be ignored during the treatment period. Moreover, due to significant physiological changes in follicle-stimulating hormone levels in the serum, it is important to repeat the measurement multiple times during clinical treatment to avoid erroneous judgments. The data in this paper are accurate after multiple measurements. The normal values of FSH content are 1-9U/L during the follicular phase, 6-260U/L during the ovulation phase, 1-9U/L during the luteal phase, and 30-118U/L during menopause.

3.1.5. Anti-Müllerian Hormone

Its abbreviation is FSH. Its content is mainly used to evaluate the ovarian reserve function of women. It can provide a basis for the auxiliary diagnosis of reproductive system-related diseases and the evaluation of female reproductive function in the treatment of other diseases[19]. AMH examination is an essential step during treatment. Doctors can quickly determine ovarian reserve function based on its content value, thereby determining a woman's reproductive ability. The normal range of AMH is 2-6.8ng/ml.

3.1.6. Gonadotropins

Its abbreviation is Gn. It can regulate the development of vertebrate gonads and is a glycoprotein hormone that promotes the production and secretion of sex hormones. It has the function of promoting the growth and development of follicles. The normal value of its content will vary with the physiological changes of women[20]. Reasonable regulation of gonadotropin levels for patients is an essential step in the clinical treatment stage.

Table 1. Data of natural fertilization group and artificial fertilization group.

Index	Natural fertilization group (n=1167)	Artificial fertilization group (n=219)
woman's age (year)	30.40±4.26	29.95±4.71
20-34years [(%)]	1000(85.68)	184(84.02)
35-39years [(%)]	122(10.45)	25(11.42)
>40years [(%)]	45(3.87)	10(4.56)
Duration of infertility (year)	3.49±2.69	3.98±2.75
0-4.25 [(%)]	852(73.01)	138(63.01)
4.26-10.25 [(%)]	280(23.99)	75(34.25)
>10.25 [(%)]	35(3.00)	6(2.74)
FSH (U/L)	6.50±1.83	6.64±2.23
0-10 [(%)]	1117(95.72)	208(94.98)
10 [(%)]	50(4.28)	11(5.02)
AMH (ng/ml)	4.50±3.56	4.15±3.02
0-2 [(%)]	260(21.28)	51(23.29)
2-6.8 [(%)]	712(61.01)	139(63.47)
>6.8 [(%)]	195(17.71)	29(13.24)
Total dose of Gn (IU)	2738.33±1066.68	2772.26±1104.63
0-1200 [(%)]	24(2.06)	4(1.83)
1200-2400 [(%)]	525(44.99)	101(46.12)
2400-3600 [(%)]	403(34.53)	69(31.51)
>3600 [(%)]	215(18.42)	45(20.54)
Number of oocytes retrieved	11.66±6.42	11.05±6.00
0-10 [(%)]	509(43.62)	99(45.21)
10-15 [(%)]	393(33.68)	73(33.33)
>15 [(%)]	265(22.70)	47(21.46)

Next, we will use these covariates to establish the multinomial logistic regression model. The covariates of 1167 natural fertilization patients and 219 artificial fertilization patients are statistically analyzed, as shown in Table 1.

To facilitate the description and analysis of covariates, we use mean ± standard deviation in the table to represent intra-group data characteristics. Prepare for the next modeling work.

3.2. Establishment of Multinomial Logistic Regression Model

To discover the relationship between the dependent variable and the covariates, we first attempt to establish a multinomial logistic regression model for five covariates. The multinomial logistic regression model is an extension of the logistic regression model with a binary dependent variable. Logistic regression models the probability of the dependent variable belonging to a category rather than directly modeling it[21]. When the dependent variable has a classification of J categories ($J > 2$) and uses $p_1, p_2, \ldots, p_J$ represents the probability of the dependent variable Y taking the value. So there must be $p_1 + p_2 + \ldots + p_J = 1$. Firstly, select any category as the reference category, and here we assume that the first category is the reference category. So other categories can be defined by logit:

$$\ln\frac{p_j}{p_1} = \beta_{j0} + \beta_{j1}x_1 + \ldots + \beta_{jm}x_m. \quad (1)$$

Where $J = 2, 3, \ldots, J$, m is the number of covariates. Simultaneously using $(J-1)$ equations to estimate regression coefficients, the predicted values can be calculated after obtaining the coefficients:

$$p_j = p_1^{\beta_{j0}+\beta_{j1}x_1+\ldots+\beta_{jm}x_m}. \quad (2)$$

Because the sum of all predicted values p_j is 1, so:

$$p_j = \frac{1}{1+\sum_{j=2}^{J} e^{\beta_{j0}+\beta_{j1}x_1+\ldots+\beta_{jm}x_m}}. \quad (3)$$

Transforming equations (1) and (2) yields:

$$p_j = \frac{e^{\beta_{j0}+\beta_{j1}x_1+\ldots+\beta_{jm}x_m}}{1+\sum_{j=2}^{J} e^{\beta_{j0}+\beta_{j1}x_1+\ldots+\beta_{jm}x_m}}. \quad (4)$$

For other values of the dependent variable, j_0 can be represented based on equation (4):

$$\ln\frac{p_{j_0}}{p_1} = \beta_{j0} + \beta_{j1}x_1 + \ldots + \beta_{jm}x_m. \quad (5)$$

Next, we will establish multinomial logistic regression models for each of the two data groups.

3.2.1. Multinomial Logistic Regression Model for Natural Fertilization Group

Due to the fact that the data are all continuous variables, we establish a multinomial logistic regression model with the number of oocytes retrieved as the dependent variable and female age, infertility years, FSH, AMH, and Gn dose as covariates. The established regression model is as follows:

$$\begin{aligned} G = {} & \alpha_0 + \alpha_1 z_{age20-34} + \alpha_2 z_{age35-39} + \alpha_3 z_{doi0-4.25} + \alpha_4 z_{doi4.26-10.25} + \alpha_5 z_{FSH0-10} + \alpha_6 z_{AMH0-2} \\ & + \alpha_7 z_{AMH2-6.8} + \alpha_8 z_{Gn0-1200} + \alpha_9 z_{Gn1200-2400} + \alpha_{10} z_{Gn2400-3600}. \end{aligned} \quad (6)$$

where α_0 is a constant, $\alpha_1, \alpha_2, \ldots, \alpha_{10}$ is the regression coefficient corresponding to each covariates, z_{age} is the female age, z_{doi} is the duration of infertility, z_{FSH} is the FSH, z_{AMH} is AMH, and z_{Gn} is the dosage of Gn. By using R software, we can establish a multinomial logistic regression model (6) based on natural fertilization group data and obtain regression equation as follows:

$$\begin{aligned} G_1 = \ln\frac{P(\text{oocytes retrieved}<10)}{P(\text{oocytes retrieved}>15)} = {} & 0.7964 - 1.2428 z_{age20-34} - 0.6630 z_{age35-39} + 1.1357 z_{doi0-4.25} \\ & + 1.0673 z_{doi4.26-10.25} - 1.1149 z_{FSH0-10} + 6.1253 z_{AMH0-2} + 1.2810 z_{AMH2-6.8} - 2.0273 z_{Gn0-1200} \\ & - 0.7329 z_{Gn1200-2400} - 0.3961 z_{Gn2400-3600}. \end{aligned} \quad (7)$$

$$\begin{aligned} G_2 = \ln\frac{P(10\le\text{oocytes retrieved}\le15)}{P(\text{oocytes retrieved}>15)} = {} & -1.0703 + 0.0715 z_{age20-34} + 0.6229 z_{age35-39} + 2.0333 z_{doi0-4.25} \\ & + 1.6893 z_{doi4.26-10.25} - 0.4054 z_{FSH0-10} + 3.5313 z_{AMH0-2} + 0.5440 z_{AMH2-6.8} - 1.3345 z_{Gn0-1200} \\ & - 0.9305 z_{Gn1200-2400} - 0.3658 z_{Gn2400-3600}. \end{aligned} \quad (8)$$

$$G_3 = 0\text{(Reference)}. \quad (9)$$

Some other regression results, like coefficient standard deviation and p-value, are displayed in Table 2.

Table 2. Multinomial logistic regression of natural fertilization group.

Index	Regression coefficient	Standard deviation	p-value	$exp(\mathbf{B})$	Significance
oocytes retrieved < 10					
α_0	0.7964	1.0904	0.4652		
α_1	-1.2428	0.6393	0.0519	0.2886	
α_2	-0.6330	0.7004	0.3661	0.5310	
α_3	1.1357	0.5609	0.0429	3.1135	*
α_4	1.0673	0.5749	0.0634	2.9076	
α_5	-1.1149	0.7776	0.1516	0.3280	
α_6	6.1253	1.0292	<0.0001	457.2773	***
α_7	1.2810	0.2445	<0.0001	3.6004	***
α_8	-2.0273	0.8626	0.0188	0.1317	*
α_9	-0.7329	0.2757	0.0079	0.4805	**
α_{10}	-0.3961	0.2887	0.1700	0.6729	
$10 \leq$oocytes retrieved≤ 15					
α_0	-1.0703	1.2215	0.3809		
α_1	0.0715	0.7301	0.9220	1.0741	
α_2	0.6229	0.7877	0.4291	1.8643	
α_3	2.0333	0.6597	0.0021	7.6394	**
α_4	1.6893	0.6713	0.0119	5.4155	*
α_5	-0.4054	0.8117	0.6175	0.6667	
α_6	3.5313	1.0330	0.0006	4.1681	***
α_7	0.5440	0.2007	0.0067	1.7229	**
α_8	-1.3345	0.5467	0.0146	0.2633	*
α_9	-0.9305	0.2686	0.0005	0.3944	***
α_{10}	-0.3658	0.2805	0.1921	0.6936	

From the table, we find that most of the selected covariates have relatively small p-values, but the p-values of covariates, such as FSH, are not significant. Further screening is required for the selected covariates.

In order to obtain better results, we will perform likelihood ratio tests on the multinomial logistic regression equation, and the results are shown in Table 3.

Table 3. Likelihood ratio test of multi-classification logistic model in natural fertilization group.

Index	Chi-square	p-value	Significance
female age	12.0694	0.0168	*
Duration of infertility	13.8649	0.0077	**
FSH	4.4023	0.1107	
AMH	256.5386	0.0001	***
Total dose of Gn	21.1865	0.0017	**

After likelihood ratio testing, we find that only the female age, duration of infertility, AMH, and Gn do have a significant impact on the number of oocytes retrieved. Next, we use these four covariates to establish the multinomial logistic regression model. The established regression model is as follows:

$$G = \alpha_0 + \alpha_1 z_{age20-34} + \alpha_2 z_{age35-39} + \alpha_3 z_{doi0-4.25} + \alpha_4 z_{doi4.26-10.25} + \alpha_5 z_{AMH0-2} + \alpha_6 z_{AMH2-6.8} + \alpha_7 z_{Gn0-1200} + \alpha_8 z_{Gn1200-2400} + \alpha_9 z_{Gn2400-3600}. \quad (10)$$

where α_0 is a constant, $\alpha_1,\alpha_2, \ldots ,\alpha_9$ is the regression coefficient corresponding to each covariates, z_{age} is the female age, z_{doi} is the duration of infertility, z_{AMH} is AMH, and z_{Gn} is the dosage of Gn. We can reestablish a multinomial logistic regression model (8) based on natural fertilization group data and obtain regression equation as follows:

$$G_1 = \ln\frac{P(\text{oocytes retrieved} < 10)}{P(\text{oocytes retrieved} > 15)} = -0.3204 - 1.2810 z_{age20-34} - 0.6489 z_{age35-39} + 1.1819 z_{doi0-4.25} + 1.1072 z_{doi4.26-10.25} + 6.2089 z_{AMH0-2} + 1.3040 z_{AMH2-6.8} - 2.0496 z_{Gn0-1200} - 0.7368 z_{Gn1200-2400} - 0.4030 z_{Gn2400-3600}. \quad (11)$$

$$G_2 = \ln\frac{P(10 \le \text{oocytes retrieved} \le 15)}{P(\text{oocytes retrieved} > 15)} = -1.4521 + 0.0546 z_{age20-34} + 0.6033 z_{age35-39} + 2.0294 z_{doi0-4.25} + 1.6900 z_{doi4.26-10.25} + 3.5458 z_{AMH0-2} + 0.5518 z_{AMH2-6.8} - 1.3368 z_{Gn0-1200} - 0.9335 z_{Gn1200-2400} - 0.3762 z_{Gn2400-3600}. \quad (12)$$

$$G_3 = 0(\text{Reference}) \quad (13)$$

The detailed results of multinomial logistic regression for the modified natural fertilization group are shown in Table 4.

Table 4. Modified multinomial logistic regression of natural fertilization group.

Index	Regression coefficient	Standard deviation	p-value	$exp(\mathbf{B})$	Significance
oocytes retrieved< 10					
α_0	-0.3204	0.7669	0.6761		
α_1	-1.2810	0.6372	0.0444	0.2777	*
α_2	-0.6489	0.6985	0.3529	0.5226	
α_3	1.1819	0.5620	0.0355	3.2606	*
α_4	1.1072	0.5760	0.0546	3.0260	
α_5	6.2089	1.0285	<0.0001	497.1402	***
α_6	1.3040	0.2440	<0.0001	3.6842	***
α_7	-2.0496	0.8682	0.0182	0.1288	*
α_8	-0.7368	0.2754	0.0075	0.4786	**
α_9	-0.4030	0.2883	0.1622	0.6683	
10 ≤oocytes retrieved≤ 15					
α_0	-1.4521	0.9143	0.1122		
α_1	0.0546	0.7289	0.9403	1.0561	
α_2	0.6033	0.7867	0.4432	1.8281	
α_3	2.0294	0.6592	0.0021	7.6096	**
α_4	1.6900	0.6708	0.0118	5.4195	*
α_5	3.5458	1.0323	0.0006	4.6673	***
α_6	0.5518	0.2004	0.0059	1.7364	**
α_7	-1.3368	0.5464	0.0144	0.2627	*
α_8	-0.9335	0.2683	0.0005	0.3932	***
α_9	-0.3762	0.2802	0.1794	0.6865	

Based on the results in the table and regression equation, we can draw the following conclusions:

- From Table 4, we find that the p-values of all selected covariates are relatively small, indicating that the selected covariates have a certain significance for the number of oocytes retrieved. After calculation by R software, it is found that the model's $p = 0.0010$ indicates a good fit for our established model.
- When the number of oocytes retrieved is greater than 15 as the reference standard, the probability of females aged 20-34 years old among females with 0-10 oocytes retrieved is 0.2777 times that of females aged over 39 years old. And the probability of females aged 35-39 years old is 0.5226 times that of females aged over 39 years old. Compared to females over 40 years old, younger females are more likely to have more than 15 oocytes retrieved; Compared to females aged 40, the number of oocytes retrieved by females aged 20-39 tends to be 10-15.
- When the number of oocytes retrieved is greater than 15 as the reference standard, the probability of patients with 0-10 oocytes retrieved having a duration of infertility of 0-4.5 years is 3.2606 times that of patients with a duration of infertility greater than 10.25 years. And the probability of patients with a duration of infertility of 4.25-10.25 years is 3.0260 times that of patients with a duration of infertility greater than 10.25 years. Compared to patients with a duration of infertility greater than 10.25 years, patients with a duration of infertility of 0-10.25 years tend to have 10-15 oocytes retrieved.
- When the number of oocytes retrieved is greater than 15 as the reference standard, the probability of AMH being 0-2ng/ml in patients with 0-10 oocytes retrieved is 497.1402 times higher than that of patients with AMH greater than 6.8ng/ml, the probability of AMH being 2-6.8ng/ml is 3.6848 times higher than that of patients with AMH greater than 6.8ng/ml. And the probability of AMH in patients with 10-15 oocytes retrieved is 4.6673 times higher than that of patients with AMH greater than 6.8ng/ml, and the probability of AMH in patients with 2-6.8ng/ml is 1.7364 times higher than that of patients with AMH greater than 6.8ng/ml. Compared to patients with AMH levels greater than 6.8ng/ml, patients with lower AMH levels tend to have lower oocytes retrieved.
- When the number of oocytes retrieved is greater than 15 as the reference standard, the probability of patients with 0-1200IU Gn doses among patients with 0-10 oocytes retrieved is 0.1288 times higher than that of patients with Gn doses greater than 3600IU, the probability of patients with 1200-2400IU Gn doses is 0.4786 times higher than that of patients with Gn doses greater than 3600IU, the probability of patients with 2400-3600IU Gn doses is 0.6683 times higher than that of patients with Gn doses greater than 3600IU. And the probability of patients with 0-1200IU Gn doses among patients with 10-15 oocytes retrieved is 0.2627 times higher than that of patients with Gn doses greater than 3600IU, the probability of patients with 1200-2400IU Gn doses is 0.3932 times higher than that of patients with Gn doses greater than 3600IU, the probability of patients with 2400-3600IU Gn doses is 0.6865 times higher than that of patients with Gn doses greater than 3600IU. Compared to patients with Gn greater than 3600IU, patients have 0-3600IU Gn tend to have more than 15 oocytes retrieved.

3.2.2. Multinomial Logistic Regression Model for Artificial Fertilization Group

For the artificial fertilization group data, we also used female age, fertility years, FSH, AMH, and Gn dose as covariates to establish a multinomial logistic regression model. The established regression model is as follows:

$$G_1 = \ln\frac{P(\text{oocytes retrieved} < 10)}{P(\text{oocytes retrieved} > 15)} = 0.2517 + 0.2043 z_{age20-34} + 1.0881 z_{age35-39} - 0.3768 z_{doi0-4.25}$$
$$+ 0.1629 z_{doi4.26-10.25} - 0.6428 z_{FSH0-10} + 3.2353 z_{AMH0-2} + 0.3193 z_{AMH2-6.8} - 0.4498 z_{Gn0-1200}$$
$$+0.1364 z_{Gn1200-2400} + 0.3731 z_{Gn2400-3600}. \tag{14}$$

$$G_2 = \ln\frac{P(10 \le \text{oocytes retrieved} \le 15)}{P(\text{oocytes retrieved} > 15)} = -0.0145 + 0.6519 z_{age20-34} + 0.4007 z_{age35-39} + 0.8297 z_{doi0-4.25}$$
$$-1.0143 z_{doi4.26-10.25} + 0.8282 z_{FSH0-10} + 0.2169 z_{AMH0-2} + 0.3025 z_{AMH2-6.8} - 0.7447 z_{Gn0-1200}$$
$$-0.7169 z_{Gn1200-2400} + 0.0025 z_{Gn2400-3600}. \tag{15}$$

$$G_3 = 0(\text{Reference}). \tag{16}$$

Some other regression results, like coefficient standard deviation, and p-value, are displayed in Table 5.

Table 5. multinomial logistic regression of artificial fertilization group.

Index	Regression coefficient	Standard deviation	p-value	$exp(\mathbf{B})$	Significance
oocytes retrieved< 10					
α_0	0.2517	2.1816	0.9082		
α_1	0.2043	1.2830	0.8735	1.2267	
α_2	1.0881	1.4328	0.4476	2.9686	
α_3	-0.3768	1.4887	0.8002	0.6860	
α_4	0.1629	1.4826	0.9125	1.1770	
α_5	-0.6428	1.2595	0.6098	0.5258	
α_6	3.2353	0.9431	0.0006	25.4148	
α_7	0.3193	0.5776	0.5804	1.3762	
α_8	-0.4498	1.3591	0.4872	1.2575	
α_9	0.1364	0.5676	0.8101	1.1461	
α_{10}	0.3731	0.6111	0.5415	1.4523	
$10 \leq$oocytes retrieved≤ 15					
α_0	-0.0145	2.3513	0.9082		
α_1	0.6519	1.4716	0.8735	1.9192	
α_2	0.4007	1.6339	0.4476	1.4929	
α_3	0.8297	1.2718	0.8002	0.4362	
α_4	-1.0143	1.2697	0.9125	0.3628	
α_5	0.8282	1.4646	0.6098	2.2892	
α_6	0.2169	1.0838	0.0006	1.2423	
α_7	0.3025	0.5585	0.5804	1.3532	
α_8	-0.7447	1.2497	0.5682	0.0410	
α_9	-0.7169	0.5325	0.8101	0.1730	
α_{10}	0.0025	0.5733	0.5415	0.3259	

According to the table and regression equation, we find that the p-values of all covariates are not significant, indicating that further screening is needed for the selected covariates. Further screening is required for the selected covariates. And from the model, we find that the number of oocytes retrieved in the natural fertilization group is closely related to the patients' own physical indicators. However, the covariates we select in the artificial fertilization group cannot provide a good explanation for the number of oocytes retrieved. Next, we will perform a likelihood ratio test on the multi-class logistic regression equation, and the test results are shown in Table 6.

Table 6. Likelihood ratio test of multi-classification logistic model in artificial fertilization group.

Index	Chi-square	p-value	Significance
female age	3.9374	0.4149	
Duration of infertility	4.5650	0.3349	
FSH	1.5132	0.4693	
AMH	49.9362	<0.0001	***
Total dose of Gn	7.4136	0.2843	

After likelihood ratio testing, we find that only AMH have a significant impact on the number of oocytes retrieved. Next, we will use AMH as covariates to establish a new multinomial logistic regression equation. The established regression model is as follows:

$$G = \alpha_0 + \alpha_1 z_{AMH0-2} + \alpha_2 z_{AMH2-6.8}. \quad (17)$$

where α_0 is a constant, α_1,α_2 is the regression coefficient corresponding to each covariates, z_{AMH} is AMH. Re-fitting the artificial fertilization group data and obtaining the following results:

$$G_1 = \ln\frac{P(\text{oocytes retrieved}<10)}{P(\text{oocytes retrieved}>15)} = -0.2231 + 3.3586 z_{AMH0-2} + 0.4745 z_{AMH2-6.8}. \quad (18)$$

$$G_2 = \ln\frac{P(10\leq\text{oocytes retrieved}\leq 15)}{P(\text{oocytes retrieved}>15)} = 0.0953 + 0.3102 z_{AMH0-2} + 0.4269 z_{AMH2-6.8}. \quad (19)$$

$$G_3 = 0(\text{Reference}). \quad (20)$$

The detailed results of multinomial logistic regression for the modified artificial fertilization are shown in Table 7.

Table 4. Modified multinomial logistic regression of artificial fertilization group.

Index	Regression coefficient	Standard deviation	p-value	exp(B)	Significance
oocytes retrieved<10					
α_0	-0.2231	0.4743	0.6380		
α_1	3.3586	0.8641	0.0001	28.7500	***
α_2	0.4745	0.5252	0.3663	1.6071	
$10\leq$oocytes retrieved≤ 15					
α_0	0.0953	0.4369	0.8273		
α_1	0.3102	1.0120	0.7593	1.3636	
α_2	0.4269	0.4862	0.3800	1.5325	

Based on the results in the table and regression equation, we can draw the following conclusions:

1. From the results, we find that the model fitted by re-selecting covariates is significantly better than before.

2. When the number of oocytes retrieved is greater than 15 as the reference standard, the probability of AMH being 0-2ng/ml in patients with 0-10 oocytes retrieved is 28.7500 times higher than that of patients with AMH greater than 6.8ng/ml, the probability of AMH being 2-6.8ng/ml is 1.6071 times higher than that of patients with AMH greater than 6.8ng/ml. And the probability of AMH in patients with 10-15 oocytes retrieved is 1.3636 times higher than that of patients with AMH greater than 6.8ng/ml, and the probability of AMH in patients with 2-6.8ng/ml is 1.5325 times higher than that of patients with AMH greater than 6.8ng/ml. Compared to patients with AMH levels greater than 6.8ng/ml, patients with lower AMH levels tend to have lower oocytes retrieved. This conclusion is consistent with the results of the natural fertilization group, but there are differences in multiples.

3.3. Predicting the Number of Oocytes Retrieved using the Multinomial Logistic Regression Model

When using the above multinomial logistic regression model to predict the number of oocytes retrieved by patients, it is necessary to substitute the patient's fertilization method into the corresponding model. Finally, incorporate G_1, G_2, and G_3 into the following formula:

$$P_1 = \frac{exp(G_1)}{exp(G_1)+exp(G_2)+exp(G_3)}. \quad (21)$$

$$P_2 = \frac{exp(G_2)}{exp(G_1)+exp(G_2)+exp(G_3)}. \quad (22)$$

$$P_3 = \frac{exp(G_3)}{exp(G_1)+exp(G_2)+exp(G_3)}. \quad (23)$$

Where P_1, P_2, P_3 is the probabilities of patients receiving 0-10, 10-15, or more than 15 oocytes, respectively.

4. Concluding Remarks

This paper selects the clinical data of patients receiving assisted reproductive therapy from 2019 to 2021 in the department of Obstetrics and Gynecology of a hospital in Jiaozuo City. We select data from 1386 patients for our study. They are divided into a natural fertilization group consisting of 1167 patients and an artificial fertilization group consisting of 219 patients according to the fertilization type. Multinomial logistic regression analysis is conducted on the natural fertilization group data using the number of oocytes retrieved as the dependent variable and female age, duration of infertility, AMH, and Gn dose as covariates. Multinomial logistic regression analysis is conducted on the artificial fertilization group data using the number of oocytes retrieved as the dependent variable and AMH covariates. Doctors can develop corresponding plans for future treatment of patients based on a multinomial logistic regression model, ensuring that the number of oocytes retrieved by the patient is within expectations. And doctors can achieve the ideal number of oocytes retrieved by adjusting the patient's covariates.

From the results, there is a significant difference in the relationship between the number of oocytes retrieved and the covariates in patients with two different fertilization methods. Among them, AMH is a very important indicator for patients with different fertilization types. And the higher the AMH content, the greater the number of oocytes retrieved. Research has shown that there is a certain correlation between AMH and pregnancy outcome, and its indicators are directly related to clinical pregnancy and live birth rate. Xiao et al. find that AMH is positively correlated with the number of oocytes retrieved, fertilization rate, and the number of blastocyst transfers [22]. Moreover, a large amount of clinical trial evidence indicates that AMH levels have strong predictive value for ovarian reserve function and ovarian responsiveness[23, 24] and have moderate predictive value for in vitro fertilization[25]. AMH performs similarly in our model. The number of oocytes retrieved by artificial fertilization patients is more focused on in vitro culture technology, which has little impact on the patient's own physical indicators. However, natural fertilization patients need to have good physical conditions, especially good ovarian reserve function. From the model, we find that significant indicators such as female age and AMH are all related to the number of oocytes retrieved, and they all reflect the patient's ovarian reserve function to a certain extent.

The model of the artificial fertilization group has limitations. Because we can choose fewer significant covariates, it may lead to lower accuracy of the prediction results. Collecting data related to in vitro fertilization may lead to better models. However, an excessive number of oocytes retrieved does not necessarily mean a better pregnancy outcome. Their relationship is still a topic worth exploring in the future.

References

[1] Gobor Kovacs et al., “Can you ever Collect too Many Oocytes?,” *Human Reproduction*, vol. 30, no. 1, pp. 81-87, 2015. [CrossRef] [Google Scholar] [Publisher link]

[2] L. Zhuang, and C.F. Hao, “Effects of Thyroid-Stimulating Hormone and Thyroid Autoantibodies on Pregnancy Outcome After in Vitro Fertilization and Embryo Transfer,” *Journal of Reproductive Medicine*, vol. 62, no. 4, pp. 391-396, 2017.

[3] X.Y. Gao, and T. Li, “The Effect of Repeated Controlled Ovarian Stimulation on Ovarian Functional Status,” *Journal of Sun Yat sen University*, vol. 29, no. 2, pp. 186-189, 2008.

[4] P. Blondin et al., “Superovulation can Reduce the Developmental Competence of Bovine Embryos,” *Theriogenology*, vol. 46, no. 7, pp. 1191-1203, 1996. [CrossRef] [Google Scholar] [Publisher link]

[5] Sesh Kamal Sunkara et al., “Association between the Number of Oocytes and Live Birth in IVF Treatment: An Analysis of 400 135 Treatment Cycles,” *Human Reproduction*, vol. 26, no. 7, pp. 1768-1774, 2011. [CrossRef] [Google Scholar] [Publisher link]

[6] Pia Saldeen, Karin Kallen, and Per Sundstrom, “The Probability of Successful IVF Outcome after Poor Ovarian Response,” *Acta Obstetricia Et Gynecologica Scandinavica*, vol. 86, no. 4, pp. 457-461, 2007. [CrossRef] [Google Scholar] [Publisher link]

[7] H.Q. Lu, Y.J. He, “The Impact of Female Age and Number of Oocytes Retrieved on Pregnancy outcomes During the Rectangular IVF/ICSI Cycle,” *Journal of Reproductive Medicine*, vol. 28, no. 9, pp. 1017-1026, 2019.

[8] M.F.G.Verberg et al., “The Clinical Significance of the Retrieval of a Low Number of Oocytes following Mild Ovarian Stimulation for IVF: A Meta-Analysis,” *Human Reproduction Update*, vol. 15, no. 1, pp. 5-12, 2009. [CrossRef] [Google Scholar] [Publisher link]

[9] Alba Cascales et al., “Effect of Ovarian Stimulation on Embryo Aneuploidy and Mosaicism Rate,” *Systems Biology in Reproductive Medicine*, vol. 67, no. 1, pp. 42-49, 2021. [CrossRef] [Google Scholar] [Publisher link]

[10] Jia Huang et al., “Insufficient Histone-3 Lysine-9 Deacetylation in Human Oocytes Matured Invitro is Associated with Aberrant Meiosis,” *Fertility and Sterility*, vol. 97, no. 1, pp. 178-184, 2012. [CrossRef] [Google Scholar] [Publisher link]

[11] Veronica N. Lobach et al., “Caspase-3 Gene Expression in Human Luteinized Granulosa Cells is Inversely Correlated with the Number of Oocytes Retrieved after Controlled Ovarian Stimulation,” *Human Fertility*, vol. 22, no. 1, pp. 33-38, 2019. [CrossRef] [Google Scholar] [Publisher link]

[12] Adva Aizer et al., “Is There Any Association Between the Number of Oocytes Retrieved, Women Age, and Embryo Development?,” *Reproductive Sciences*, vol. 28, no. 7, pp. 1890-1900, 2021. [CrossRef] [Google Scholar] [Publisher link]

[13] Farah Tamirou et al., “The Euro-Lupus Low-Dose Intravenous Cyclophosphamide Regimen Does Not Impact the Ovarian Reserve, as Measured by Serum Levels of Anti–Müllerian HormonE,” *Arthritis and Rheumatology*, vol. 69, no. 6, pp. 1267-1271, 2017. [CrossRef] [Google Scholar] [Publisher link]

[14] H.Y. Xu, and Y. Feng et al., “Study on the Changes of Serum Anti Mullerian Hormone Levels During and Between Menstrual Cycles in Women of Childbearing Age,” *J Pract Obstet Gynecol*, vol. 33, no. 5.

[15] James O. Berger, Xiaojing Wang, and Lei Shen, “A Bayesian Approach to Subgroup Identification,” *Journal of Biopharmaceutical Statistics*, vol. 24, no. 1, pp. 110-129, 2014. [CrossRef] [Google Scholar] [Publisher link]

[16] L.L Van Loendersloot et al., "Predictive Factors in in vitro Fertilization (IVF): A Systematic Review and Meta-Analysis," *Fertility and Sterility*, vol. 16, no. 6, pp. 577-589, 2010. [CrossRef] [Google Scholar] [Publisher link]

[17] Adva Aizer, and J. Haas, "Is There any Association Between the Number of Oocytes Retrieved, Women Age, and Embryo Development?," *Medicine(Baltimore)*, vol. 95, p. e4957, 2021.

[18] W.X. Zhu, "Advances in the Application of Recombinant Follicle-Stimulating Hormone in Assisted Reproduction in Mammals," *Chinese Journal of Animal Husbandry*, vol. 58, no. 7, pp. 42-48, 2022.

[19] X.D. Wu, "Clinical Application of Serum Anti-Müllerian Hormone and Dehydroepiandrosterone Sulfate in the Auxiliary Diagnosis of PCOS," *Chinese Journal of Eugenics and Genetics*, vol. 31, no. 1, pp. 164-168, 2023.

[20] X.H. Zhang, "Application and Clinical Value of Human Chorionic Gonadotropin Detection," *Health for all*, vol. 558, no. 4, pp. 105-107, 2023.

[21] J.S. Cramer, "The Origins of Logistic Regression," *Tinbergen Institute Working Paper*, vol. 119, 2003. [CrossRef] [Google Scholar] [Publisher link]

[22] J.X. Xiao, L.Y. Cai, and J. Chen, "Study on the Correlation between Anti-Müllerian Hormone and Pregnancy Outcome of Controlled Ovarian Hyperstimulation," *Maternal and Child Health Care of China*, vol. 31, no. 19, pp. 4010-4013, 2016.

[23] L. B. Simone, "Added Value of Ovarian Reserve Testing on Patient Characteristics in the Prediction of Ovarian Response and Ongoing Pregnancy: An Individual Patient Data Approach," *Maternal and Child Health Care of China*, vol. 19, no. 1, pp. 26-36, 2013.

[24] Sara S.E. Alson et al., "Anti-müllerian Hormone Levels are Associated with Live Birth rates in ART, but The Predictive Ability of Anti-müllerian Hormone is Modest," *European Journal of Obstetrics, Gynecology and Reproductive Biology*, vol. 225, pp. 199-204, 2018. [CrossRef] [Google Scholar] [Publisher link]

[25] Reshef Tal et al., "Antimüllerian Hormone as a Predictor of Live birth following Assisted Reproduction: An Analysis of 85,062 Fresh and Thawed Cycles from the Society for Assisted Reproductive Technology Clinic Outcome Reporting System database for 2012-2013," *Fertility and Sterility*, vol. 109, no. 2, pp. 258-265, 2018. [CrossRef] [Google Scholar] [Publisher link]