https://doi.org/10.34883/PI.2022.12.3.008

Ismayilova M.
The Central Clinic, Baku, Azerbaijan

Binary Logistical Regression as an Optimal Model for Prognosis of the Result of in Vitro Fertilization

Conflict of interest: nothing to declare.

Submitted: 23.05.2022
Accepted: 13.06.2022
Contacts: mic_amu@mail.ru

Abstract

Introduction. Development of the predictors of effectivity of treatment of infertility by means of extracorporeal fertilization is one of the most important tasks of the modern medicine. Despite of the numerous researches there is not a unique point of view regarding significance of impact of any factors on the ECO outcomes yet.
Materials and methods. We have examined 518 patients applying on account of infertility to the Central Clinic of Baku city. All patients were divided into two groups: the patients with onset pregnancy i.e. positive ECO ICSI results – main group (n=234) and patients with negative ECO ICSI efforts – control group (n=284) which did not get pregnant after the ECO ICSI procedure.
Results. The criteria precisely differing in the compared groups were chosen among the big quantity of signs at the women with various ECO outcomes. These parameters were used at the factor analysis. Basing upon the result of the factor analysis there were assigned the main factors for forecasting the ECO outcomes of infertile women by means of logistic regression.
Conclusion. The developed model of forecasting the logistic regression allows precisely forecasting the ECO outcome upon certain parameters in 86.93% cases.
Keywords: supplementary reproductive technologies, infertility, logistic regression, extracorporeal fertilization

Исмайлова М.К.
Центральная клиника, Баку, Азербайджан

Бинарная логистическая регрессия как оптимальная модель для прогноза экстракорпорального оплодотворения

Конфликт интересов: не заявлен.

Подана: 23.05.2022
Принята: 13.06.2022
Контакты: mic_amu@mail.ru

Резюме

Введение. Разработка предикторов эффективности для лечения бесплодия с помощью экстракорпорального оплодотворения (ЭКО) является одной из важнейших задач современной медицины. Несмотря на многочисленные исследования, единой точки зрения относительно значимости влияния того или иного фактора на результаты ЭКО пока нет.

Материалы и методы. Обследовано 518 пациенток, обратившихся по поводу бесплодия в Центральную клинику города Баку, которые были разделены на две группы: пациентки с наступившей беременностью, т. е. с положительными результатами ЭКО ИКСИ – основная группа (n=234), и пациентки с отрицательными попытками ЭКО ИКСИ – группа пациенток (n=284), у которых беременность после процедуры ЭКО ИКСИ не наступила.

Результаты. Среди большого количества признаков у женщин с различными исходами ЭКО были выбраны критерии, четко отличающиеся в сравниваемых группах. Эти параметры были использованы при факторном анализе. На основании результатов факторного анализа были выделены основные параметры для прогнозирования исходов ЭКО у бесплодных женщин с помощью логистической регрессии.

Заключение. Разработанная модель прогнозирования логистической регрессии позволяет точно спрогнозировать исход ЭКО по определенным параметрам в 86,93% случаев.

Ключевые слова: дополнительные репродуктивные технологии, бесплодие, логистическая регрессия, экстракорпоральное оплодотворение

The achieved definite border of efficiency of using the extracorporeal fertilization method does not satisfy the reproductologists' requirements at achievement of further reduction of frequency of infertility at marriage and, it determines the researchers' further scientific interest in search of further prognostic criteria of effectiveness of the program of supplementary reproductive technologies (SRT).

There are some determined factors connected to interim and final ECO outcomes. First of all, it is Anti-Müllerian Hormone (AMH) in the follicular fluid.

Presently the most researchers consider the patient's ages the key predictor of effectiveness of treatment of infertility by means of SRT [Yermolenko K.S., Radzinsky V.Y.,

Rapoport S.I., 2016]. The results of analysis of the observations inserted to the SRT register of Great Britain (HFEA) submitted by S.M. Nelson and D.A. Lawlor in 2011 evidently demonstrate the dynamics of progressing reduction of possibility of birth of a living child subsequently to ECO and PE of the patients older than 35 years old [Nelson N.S., Lawlor D.A., 2011].

The productive anamnesis data of a married couple are also traditionally considered as potential predictors of effectiveness of ECO and PE. The prognostic significance of existence in anamnesis of pregnancy and viviparity is proved with results of researches [Bhattacharya S., Maheshwari A., Mollison J., 2013].

A lot of authors also consider the indicators of the ovarian reserve as factors determining the forecast of the stage of embryos' implantation and viviparity subsequently to induced pregnancy [Iliodromiti S. et al., 2014, Yao L. et al., 2015, Fang T., Su Z., Wang L., 2015].

Despite of the varying information existing in literature regarding the methods of forecasting the ECO results, all of them have low informational character and are formed not considering a lot of parameters characteristic for infertile women among which the immunological and reproductive factors have important place.

Its proof is many reproductologists' dissatisfaction by the results of the extracorporeal fertilization because frequency of onset of pregnancy after application of ECO is not more than 30–35% even in the best clinics of the world. In this regard, search of new way for improving the effectiveness of the ECO programs is actual.

Basing upon aforesaid issues, the target of this work is creation of ECO outcomes on the basis of the women's forecasting model indicators.

■ MATERIALS AND METHODS

The researching object was infertile women (initial and secondary infertility) directed to the Central Clinical Hospital of Baku city of the Ministry of Health of Azerbaijan for treatment of infertility with ECO/ICSI method. The criterion for patients' inclusion to the research was existence of indication for implementation of the ECO program.

An individual map and algorithm of patients' examination were developed for forecasting the ECO results and complication of the ECO (OHS) procedures (formalized protocol). 160 parameters are exposed to analysis in the analytical map: anamnestic data, anthropometric indicators, the results of the patients' laboratory (biochemical, hormonal, inflectional and hemostasiological) and instrumental examination, the data of gynecological and somatic anamnesis, the results of treatment with extracorporeal fertilization method as well as immunological, immunogenetic and genetic examinations.

All qualitative signs are shown in this map in quantitative gradations and, the quantitative indicators are shown in their absolute values.

The acquired data of the clinical and laboratory researches are processed by the variation statistics method in the Statistica 10 statistical analysis system (Statistica USA program package, version 10 for Windows 8). All options contain the character allocation for normality upon criteria of Kolmogorov-Smirnov, Shapiro – Wilkey and Leven. The Student's parametric criterion is used for comparative analysis if allocation is normal or the number of the researching objects is very high (over 100). Manna – Whitney's method was used at allocation of the indicators differing from the normal one.

The impact of patients' separate signs and their weighting coefficients on the ECO results and its complications (for existence of OHS) was researched by means of logistic regression analysis in the Statistica 10 (USA) program.

The logistic regression is used when the dependent value is binary (i.e. accept value yes/no, has / has not, for instance, positive ECO outcome or negative ECO outcome) and dependent variables of different character (qualitative and/or quantitative) influence its outcome. Possibility of accepting one of the set of assertions under impact of the reviewed signs is factually assessed. Logit of this possibility is the natural logarithm of the relation of the possibility of «positive effect» (p) to the possibility of «negative effect» (1 p).

$$\text{Logit}(p) = \ln\frac{p}{1-p}$$

The value of Logit (p) is permanent and accept value in interval from 0to1 (from negative effect to positive one).

The procedure of logistic regression is creation and assessment of the equation of the following type:

$$\text{Logit}(p) = Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \ldots$$

X_1, x_2, x_3 – independent variables;

b_0, b_1, b_2, b_3 – permanent coefficients.

Whereas, the possibility of positive effect is

$$p = Y = \frac{1}{1 + e^{-Y}} = \frac{1}{1 + 2{,}72^{-Y}} = A$$

Possibility of negative ECO outcome is equal to «A», and positive ECO outcome – 1-A. Or, possibility of negative ECO outcome is equal to (A×100)%, and positive ECO outcome is [(1-A) × 100]%.

The average values of the parameters acquired during research (at numerous values) or figures 1 (in presence of sign) and 0 (in absence of sign) are compared and Y value is acquired.

Self-descriptiveness of the independent signs is initially discovered for inclusion of the necessary signs characterizing the infertile women to the regression analysis. Informative signs are selected for formation of the classification functions basing upon discovery of precision of the compared signs of patients with positive and negative ECO results.

The contingency table 2×2 and some indicators like sensitiveness and specificity were applied for assessing the efficiency of the linear forecasting equations.

Sensitiveness is a share of persons with positive ECO result of examined patients. This indicator characterizes the possibility of the true positive ECO result. Its values were calculated upon the Sensitiveness TP/TP+FN formula.

Specificity is a share of persons with negative ECO result of examined patients. This indicator was calculated upon the Specificity TP/TP+FN formula.

Table 1
Efficiency of the acquired regression equities

Decision upon the tested method	The objects' factual state	
	positive	negative
Positive	TP	FP
Negative	FN	TN

Notes: TP – true positive; FN – false negative; TN – true negative; FP – false positive.

The parameters for researching in the regression analysis were selected close to the forecasting predictors or with the factor analysis method in the Statistica 10 (USA) method.

■ RESULTS

The women passing ECO were characterized with a lot of signs like the anthropometric data and anamnestic indicators as well as gynecological, clinical, immunological, immunogenetic and reproductive anamneses, etc.

The factor analysis i.e. method of the main components was used for definition of the women's main signs and determination of their weighting coefficients.

Using this method, all existing signs were joined into several groups – 4 factors (table 2).

Each factor is characterized by set of signs with relatively high weighting coefficients upon significance. The signs with low weighting coefficient are excluded from the acquired groups.

The analysis of the acquired data discovered 4 main grouping factors upon the principal signs: 1st factor – gynecological anamnesis; 2nd factor – reproductive anamnesis; 3rd factor –embryological analysis and 4th factor – combined analysis (table 2).

Passage from the higher number of signs to the lower one is implemented for the purpose of exclusion of the few informative signs, existence of the doubling information and rationality of consolidation of the frequently combining signs into one group.

As we see from the table, the symptomatic complex is named as «Gynecological anamnesis» because it has two signs with high weighting coefficient: «outcome of the previous pregnancies at secondary infertility» (0.787), «loss of spontaneous uterine pregnancy» (0.788). This factor explains 14.23% dispersion of the system.

The 2nd factor is named as «Reproductive anamnesis» because this symptomatic complex contains an indicator with high weighting coefficient belonging to the patients' reproductive anamnesis, in particular, «pregnancy subsequently to ECO with viviparity» (–0.87). However, it is noticeable that these factors contains also «Age» (0.72). The load weight bearing of other parameters on this factor is not high. This grouping factor explains 7.85% dispersion of the system. The 3rd grouping factor contains these parameters having high weighing parameters: «Number of mature oocytes preceding SRT» (0.73) and «Number of high-quality embryos preceding SRT on the transfer day» (0.72). Consequently, this grouping factor is conditionally named as «Embryologic» and explains 7.36% of dispersion of the system. The 4th grouping factor named as «Combined» explaining 6.53% of dispersion of the system contains a lot of signs with low weighing coefficients (mycoplasmosis, age, syndrome of polycystic ovaries (SPO), clamidiosis). Thus, this factor is excluded from the further research.

These parameters are combined in four groups (grouping factors) (table 2), symptomocomplexes explaining 35.94% of the used dispersion of system. Whereas, the greatest power is of 1st factor – grouping factor – «Gynecological anamnesis» determining 14.23% of variabilities of the established dispersion model, the 2nd grouping factor – Reproductive anamnesis + age with 7.85% contribution to the variability of the system, 3rd grouping factor – Embryological, responsible for 7.36% of dispersion of the system, 4th factor – combined – 6.53%. The weighing coefficients of the signs in the content of the grouping factors are also determined.

Table2
The main components of the examined women' ECO outcomes

Factor 1	**Gynecological anamnesis14.23%**			
	Outcome of the previous pregnancies at secondary infertility: miscarriage1 – one; 2 – 2 and more; 0 – no	0.787	Loss of spontaneous uterine pregnancy: 1 – yes; 0 – no	0.788
Factor 2	Reproductive anamnesis 7.85%			
	Age	0.72	ECO cycle with viviparity preceding SRT	–0.87
Factor 3	Embryological 7.36%			
	Number of the mature oocytes preceding SRT	–0.73	Number of high-quality embryos preceding SRT on the transfer day	–0.72
Factor 4	Combined – 6.53%			
	Mycoplasmosis	–0.609	Age	0.30
	SPO	0.27	Clamidiosis	0.36

Thus, among the entire massive of the researched signs precisely differing between women with positive and negative EWCO results with method of factor analysis, upon deduction of some factors, there were acquired the parameters characterizing the women preparing to ECO and deemed as significant for forecasts. The parameters of women having high weighing coefficients acquired subsequently to the factor analysis are used for forecasting the ECO results.

These indicators are «Age», «Outcome of the previous pregnancies at the secondary infertility-miscarriage» (OP/S); «Loss of spontaneous uterine pregnancy» (LSP); «Pregnancy subsequently to ECO with viviparity» (P/ECO); «Number of the mature oocytes preceding SRT» (NMO); «Number of high-quality embryos preceding SRT on the transfer day» (NHQE).

The model of regression analysis is – positive outcome (1) – Existing pregnancy (EP) at existence of pregnancy and negative outcome (0), at absence of pregnancy – Absent pregnancy (AP). EP (1) and AP (0) are dependents variables and the aforesaid signs are independent ones.

The equation coefficients are significant at the 5% level (p-level <0.05) for indicators of «Outcome of the previous pregnancies at the secondary infertility-miscarriage» with coefficient of 4.702, «Loss of spontaneous uterine pregnancy» – with coefficient of 5.84, «Number of the mature oocytes preceding SRT» with coefficient of 0.22 and «Number of high-quality embryos preceding SRT on the transfer day» with coefficient of 2.28 and «Age» with coefficient of 0.72 (table 3).

In this case the p-level of significance of the established regression model was lower than 5% – the value of statistics of Wald's Chi-square for difference between the current model and the model containing only free member is highly significant.

Thus, it may be concluded that the aforesaid selected variables (predicting signs) impact on the ECO outcome. The result of the regression analysis for forecasting the ECO outcomes is statistically precise (Chi^2=117.80 p=0.0000) (table 3).

The results of the implemented calculation discovered the following equations of regression (Equation 1):

Table 3
Coefficients of equation of regression for forecasting the ECO outcomes

	Model: Logistic regression (logit) N of 0's: 72 1's: 81 (Spreadsheet ECO3) Dep. var: ECO Positive income 1, Negative income 0 Loss: Max likelihood (MS-err. scaled to 1) Final loss: 46,885199634 Chi?(6)=117,80 p=0,0000 Include cases: 1:285						
	Const. B0	**Outcome of the previous pregnancies at the secondary infertility-miscarriage**	**Age**	**Loss of spontaneous uterine pregnancy**	**ECO cycle with viviparity preceding SRT**	**Number of mature oocytes preceding SRT**	**Number of high-quality embryos preceding SRT on the transfer day**
Estimate	5.6116	4.702045	0.72659	5.489602	–0.4128	– 0.2284	– 2.288715
Standard Error	1.0137	2.453776	1.457591	2.859676	286.4096	0.088713	0.405077
t(229)	5.5356	1.916249	–0.49848	–1.91965	–0.00144	–2.57492	–5.650073
p-value	0.0000	0.0508761	0.041835	0.050685	0.008851	0.011020	0.0000000
–95%CL	3.60814	–0.1474647	–2.59019	–11.1413	–566.457	–0.40375	–3.089287
+95%CL	7.61509	9.551555	9.137018	0.162105	541.631	–0.05310	–1.488142
Wald's Chi-square	30.6432	3.672009	0.248489	3.685091	0.365322	6.63022	31.92332
p-value	0.00000	0.05534235	0.41835	0.054909	0.008851	0.010030	0.0000000
Odds ratio (unit ch)	273.587	110.1722	3.783554	0.004129	0.661774	0.795781	0.1013967
–95%CL	36.8975	0.8628929	0.027592	0.000014	0	0.667805	0.0455344
+95%CL	2028.58	14066.55	8.474132	1.175984		0.948283	0.2257917
Odds ratio (range)		12137.92	0.483554	0.004129	0.661774	0.040842	0.0000001
–95% CL		0.7445841	0.02759	0.000014	0	0.003508	0.0000000
+95% CL		197867700	8.474132	1.175984		0.475482	0.0000299

Notes: Estimate – free equation member; Standard Error – standard error; p-value – significance level of the equation coefficients; –95% CL +95% CL – confidence interval; Wald's Chi-square – Chi-square by Wald; p-value – Wald's significance level.

Equation 1

Y = 5,61 + 4,70 x OPP/M + 0,72 x B + 5,84 x LSP – 0,41 PECO / PV – 0,22 NMO – 2,28 x NHQE

Note: OPP/M – Outcome of the previous pregnancies at the secondary infertility-miscarriage;

LSP – Loss of spontaneous pregnancy;

PECO / PV – ECO cycle with viviparity preceding SRT;

NMO – Number of mature oocytes;

NHQE – Number of high-quality embryos;

A – Age.

The contingency table 2×2 and some indicators like sensitiveness and specificity were applied for assessing the efficiency of the linear forecasting equations.

Sensitiveness is a share of persons with positive ECO result of examined patients. This indicator characterizes the possibility of the true positive ECO result.

Specificity is a share of persons with negative ECO result of examined patients.

Table 4
Sensitiveness and specificity of equation of regression for forecasting the ECO outcomes

Decision upon the tested method	The objects' factual state		
	positive	negative	% of precise predictions
Positive	61	9	84.72222
Negative	11	72	88.88889

Notes: Classification of Cases (Spreadsheet ECO3)
Odds ratio: 44,364. Perc. correct: 86,93%

The ECO results of the patients with positive and negative ECO results were additionally checked for determination of sensitivity and specificity of the established equations of regression. The results are shown in the table 4.

The equation of regression is statistically significant, $p<0.001$. The aforesaid issue shows that the established forecasting model is able to forecast the ECO outcome upon the aforesaid ECO outcome for 86.93% (table 4). This combination of signs may predict the ECO outcome and it is precise; furthermore, the forecast is quite high. The sensitivity of this forecast is equal to 84.7% and the specificity of equations is 88.8%. These indicators are able to predict the positive and negative ECO outcomes of infertile women before implementation of ECO.

■ DISCUSSION

For the purpose of researching the impact of various factors on formation of the initial and secondary infertility there was implemented a research basing upon investigation of the forecasting factors of the ECO outcomes. The infertile women passing ECO were characterized with a lot of signs like the anthropometric data and anamnestic indicators as well as gynecological, clinical, immunological, immunogenetic and reproductive anamneses, etc. and, their psychological status was checked. The results of our researches showed that the parameters having forecasting significance and impacting on ECO outcome are «Age», «Outcome of the previous pregnancies at the secondary infertility-miscarriage» (OP/S); «Loss of spontaneous uterine pregnancy» (LSP); «Pregnancy subsequently to ECO with viviparity» (P/ECO); «Number of mature oocytes preceding SRT» (NMO); «Number of high-quality embryos preceding SRT on the transfer day» (NHQE).

The criteria precisely differing in the compared groups were chosen among the big quantity of signs at the women with various ECO outcomes. These parameters were used at the factor analysis. Basing upon the result of the factor analysis there were assigned the main factors for forecasting the ECO outcomes of infertile women by means of logistic regression.

These factors are: «Outcome of the previous pregnancies at the secondary infertility-miscarriage», «Loss of spontaneous uterine pregnancy», «age», «pregnancy subsequently to ECO with viviparity», «ECO cycle with viviparity preceding SRT», «Number of mature oocytes preceding SRT», «Number of high-quality embryos preceding SRT on the transfer day».

However, as it was noticed before, the final equation of regression for forecasting the ECO outcomes does not contain all parameters acquired at the factor analysis. The signs with value of $p>0.05$ were excluded from the equation of regression. Consequently, the prognostic parameters for the ECO outcomes owing to the method of logistic regression

used the signs of loss of spontaneous uterine pregnancy, outcome of the previous pregnancies at the secondary infertility-miscarriage, number of mature oocytes preceding SRT, age and number of high-quality embryos preceding SRT on the transfer day.

The aforesaid parameters of the examined women had statistical significance and various equation coefficients in the established model at independent variables i.e. at the aforesaid parameters.

Whereas, these parameters have the highest prognostic features: «Loss of spontaneous uterine pregnancy», with coefficient of (+5.84), and «outcome of the previous pregnancies at the secondary infertility-miscarriage», with coefficient of (+4.70). They are followed by «Number of high-quality embryos preceding SRT on the transfer day» (–2.28), «Age» (+0.72), « ECO cycle with viviparity preceding SRT» (–0.41), «Number of mature oocytes preceding SRT» (–0.22). It is noticeable that the parameters with positive coefficients had negative connection with positive ECO outcome and the indicators having minus before the coefficient had positive association with the positive ECO outcome. In other words, as the loss of spontaneous uterine pregnancy and miscarriage during previous pregnancies at the secondary infertility are high and women are older, as the possibility of positive ECO outcome is lower. On the contrary, as ECO cycle with viviparity preceding SRT, high number of mature oocytes and number of high-quality embryos preceding SRT were frequently observed, as the possibility of positive ECO outcome is higher.

The proof of the fact that the chosen model of forecasting the pregnancy outcome upon the chosen parameters is precise is the level of significance of the established regression model (P) which was lower than 0.05 and, the value of statistics of Wald's Chi-square for difference between the current model and the model containing only free member is highly significant (Chi^2=117.80 p=0.0000).

Thus, it may be concluded that the aforesaid selected variables (predicting signs) impact on the ECO outcome.

The contingency table 2×2 and some indicators like sensitiveness and specificity were applied for assessing the efficiency of the linear forecasting equations.

Sensitiveness is a share of persons with positive ECO result of examined patients. This indicator characterizes the possibility of the true positive ECO result.

Specificity is a share of persons with negative ECO result of examined patients.

The ECO results of the patients with positive and negative ECO results were additionally checked for determination of sensitivity and specificity of the established equations of regression.

The total forecasting value of the established model of forecasting the ECO outcomes upon the shown parameters is equal to 86.93%.

The aforesaid issue shows that the established forecasting model is able to forecast the ECO outcome upon the aforesaid ECO outcome for 86.93%. This combination of signs may predict the ECO outcome and it is precise; furthermore, the forecast is quite high. The sensitivity of this forecast is equal to 84.7% and the specificity of equations is 88.8%. These indicators are able to predict the positive and negative ECO outcomes of infertile women before implementation of ECO.

The relation of chances equal to 44.3 shows that the classification upon model is correct for 44.3 times in comparison with occasional prediction of ECO outcome. The high forecasting ability of the model is connected with precise selection of the predictors of the prognostic model (equation of regression).

REFERENCES

1. Ermolenko K.S. (2016) Sovremennoe sostoyanie problemy realizacii fertil'noj funkcii zhenshchin pozdnego reproduktivnogo agea [Current status of the problem of realizing the fertile function of late reproductive age women]. *Klinicheskaya medicina*, vol. 94, no 1, pp. 10–15.
2. Bhattacharya S. (2013) Factors Associated with Failed Treatment: an Analysisof 121,744 Women Embarking on Their First IVF Cycles. *PLOS ONE,* vol. 8, Issue 12, e82249, pp. 1–13.
3. Iliodromiti S. (2014) The predictive accuracy of anti-Müllerian hormone for live birth after assisted conception: a systematic review and meta-analysis of the literature. *Hum. Reprod. Update,* vol. 20 (4), pp. 560–570.
4. Lingnv Yao, Wei Zhang, Hong Li (2015) The role of serum AMH and FF AMH in predicting pregnancy outcome in the fresh cycle of IVF/ICSI: a meta-analysis. *Int. J. Clin. Exp. Med.,* vol. 8(2), pp. 1755–1767.
5. Nelson S.M. (2011) Predicting Live Birth, Preterm Delivery, and Low Birth Weightin Infants Born from In Vitro Fertilisation: A Prospective Study of 144,018 Treatment Cycles [Electronic resource]. *PLoS Medicine,* vol. 8, Issue 1, e1000386. Access mode: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000386,free. Title screen. (date of access 29.11.15).
6. Wang N. (2015) Luteal-phase ovarian stimulation vs conventional ovarian stimulation in patients with normal ovarian reserve treated for IVF: a large retrospective cohort study. *Clinical Endocrinology,* vol. 0, pp. 1–9.