



Factors associated with livebirth in couples undergoing their first in vitro fertilization cycle: An internally validated prediction model

İlk IVFsiklusunda canlı doğumla ilişkili faktörler: İnternal doğrulanmış bir prediksyon modeli

Erkan Kalafat¹, Can Benlioğlu², Ali Gökçe³, Yavuz Emre Şükür¹, Batuhan Özmen¹, Murat Sönmezer¹, Cem Somer Atabekoğlu¹, Ruşen Aytaç¹, Bülent Berker¹

¹Ankara University Faculty of Medicine, Department of Obstetrics and Gynecology, Ankara, Turkey

²Doğubeyazıt State Hospital, Ministry of Health, Ağrı, Turkey

³Yıldırım Beyazıt University, Yenimahalle Training and Research Hospital, Ankara, Turkey

Abstract

Objective: The aim of the study is to create a new model to predict successful outcome in assisted reproductive techniques.

Materials and Methods: A retrospective cohort study was conducted in tertiary fertility center between 2010 and 2017. Nulliparous women younger than 45 years-old undergoing in vitro fertilization/intracytoplasmic sperm injection (IVF/ICSI) for the first time were included; frozen embryo transfers, canceled induction cycles, freeze-all cycles were excluded. Two prediction models were built using multivariate logistic regression with a subset of the dataset and then were internally validated using bootstrapping methods.

Results: Four hundred eighty eight women were included with 136 (27.9%) live births. The basal model was built using variable age, antral follicle count (AFC), and basal luteinizing hormone (LH) levels. Age over 37 years [odds ratio (OR): 0.07, 95% confidence interval (CI): 0.00-0.36] and AFC below 5 (OR: 0.15, 95% CI: 0.02-0.53) was associated with poorer outcomes whereas an LH level above 6 mIU/mL (OR: 2.24, 95% CI: 1.27-3.94) was associated with better outcomes. Optimism adjusted area under the curve (AUC) of this model was 0.68 (95% CI: 0.62-0.74). Combined model in addition to basal model variables included the length of induction cycle, the endometrial thickness at the day of transfer, grade and count of the transferred embryo. Cycles lasting more than ten days (OR: 2.23, 95% CI: 1.17-4.42), an endometrial thickness greater than 9 mm (OR: 2.07, 95% CI: 1.00-4.53) were associated with better outcomes. Optimism adjusted AUC of this model was 0.76 (95% CI: 0.70-0.81). Calibration of both models was good according to Hosmer Lemeshow test (p=0.979 and p=0.848, respectively).

Conclusion: This internally validated prediction model has good calibration and can be used predicting outcomes in first time IVF/ICSI cycles with modest sensitivity.

Keywords: Prediction models, assisted reproductive techniques, live birth, in-vitro fertilization

Öz

Amaç: Çalışmanın amacı yardımcı üreme tekniklerinde başarılı sonucu tahmin edebilmek için yeni bir model oluşturmak.

Gereç ve Yöntemler: 2010-2017 yılları arasında üçüncü basamak infertilite merkezinde retrospektif bir kohort çalışması yapılmıştır. İlk kez tüp bebek/intrasitoplazmik sperm enjeksiyonu (IVF/ICSI) uygulanan 45 yaş altı nullipar kadınlar dahil edilmiştir; dondurulmuş embriyo transferleri, iptal edilen indüksiyon siklusları, freeze-all siklusları hariç tutulmuştur. Veri kümesinin bir alt kümesiyle çok değişkenli lojistik regresyon modeli kullanılarak iki tahmin modeli oluşturuldu ve ardından bootstrapping yöntemleri kullanılarak internal olarak doğrulandı.

Bulgular: Yüz otuz altı (%27,9) canlı doğum yapan 488 kadın dahil edildi. Bazal model, yaş, antral folikül sayısı (AFC) ve bazal lüteinize edici hormon (LH) seviyeleri değişkenleri kullanılarak oluşturulmuştur. Otuz yedi yaş üstü [olasılık oranı (OO): 0,07, %95 güven aralığı (GA): 0,00-0,36] ve 5'in altındaki AFC (OO: 0,15, %95 GA: 0,02-0,53) daha kötü sonuçlarla ilişkilendirilirken, LH seviyesinin 6 mIU/mL'nin üzerinde olması (OO: 2,24, %95 GA: 1,27-3,94) daha iyi sonuçlarla ilişkilendirildi. Bu modelin iyimserliğe göre ayarlanmış eğrinin altındaki alanı (AUC) 0,68 (%95 GA: 0,62-0,74) idi. Bazal

PRECIS: Prediction of IVF success can be estimated using baseline characteristics and cycle-specific variable with better precision and calibration compared to traditional models such as templetion.

Address for Correspondence/Yazışma Adresi: Asst. Prof. Erkan Kalafat, Koc University Hospital, Reproductive Endocrinology and Infertility Division, Topkapı, Istanbul

Phone: +90 850 250 82 50 **E-mail:** mail@erkankalafat.com **ORCID ID:** orcid.org/0000-0003-0658-138X

Received/Geliş Tarihi: 17.06.2021 **Accepted/Kabul Tarihi:** 21.08.2021

model değişkenlerine ek olarak birleşik model, indüksiyon döngüsünün uzunluğunu, transfer gününde endometriyal kalınlığı, transfer edilen embriyonun derecesini ve sayısını içermiştir. On günden fazla süren sikluslar (OO: 2,23, %95 GA: 1,17-4,42), endometriyal kalınlığın 9 mm'den büyük (OO: 2,07, %95 GA: 1,00-4,53) olması daha iyi sonuçlarla ilişkilendirildi. Bu modelin iyimserliğe göre ayarlanmış AUC'si 0,76'dır (%95 GA: 0,70-0,81). Hosmer-Lemeshow testine göre her iki modelin de kalibrasyonu iyiydi (sırasıyla $p=0,979$ ve $p=0,848$).

Sonuç: Bu internal olarak doğrulanmış tahmin modeli iyi bir kalibrasyona sahip olmakla birlikte ilk kez IVF/ICSI yapılan hastalarda sonuçları tahmin etmek için iyi bir hassasiyetle kullanılabilir.

Anahtar Kelimeler: Tahmin modeli, yardımcı üreme teknikleri, canlı doğum, in vitro fertilizasyon

Introduction

Subfertile couple counseling is one of the most important parts of assisted reproductive technology (ART) treatment. Several prediction models in the literature provided a data-driven perspective to both clinicians and patients⁽¹⁻³⁾. Live birth is the ultimate goal of ART treatment, and the most common and recognized models by Nelson and Lawlor⁽⁴⁾ and Templeton et al.⁽⁵⁾ used a live birth as the primary outcome. However, both models underestimate the live birth rate based on external validation studies⁽⁶⁾. Another external validation study concluded that a better calibration was achieved for both models after adjustments based on current trends of in vitro fertilization (IVF) success; however, the Templeton model underestimated and the Nelson model overestimated the chances of live birth⁽⁷⁾.

Delaying conception attempts and pregnancy until the later ages of a childbearing period is one of the most common causes of increased IVF/intracytoplasmic sperm injection (ICSI) uptake⁽⁸⁾. The success rates of ART treatments for women at advanced age were remarkable. Based on human fertilization and embryology authority reports, even women aged 40-42 years have a higher chance of live birth in 2018 than those aged under 35 in 1991 (11% vs 9%, respectively). The same authority reports concluded that the average age for an IVF cycle was older in 2017 than in 1991 (35.5 vs 33.5, respectively)⁽⁹⁾. These increasing success rates with an older patient population are explained with the individualized cycle management, novel techniques for embryo transfer protocols, and ART laboratory evaluations⁽⁹⁾. With improvements, prediction models are updated based on newer approaches to rationalize the usage of these tools.

This study primarily aimed to establish a well-calibrated model, which combined both patient demographics, cycle management, and embryo transfer day characteristics. The secondary aim is to estimate the live birth rates and compare the Templeton model with the present prediction model.

Materials and Methods

This was a retrospective cohort study conducted in a single tertiary infertility clinic in the Department of Obstetrics and Gynecology at Ankara University. Dataset was gathered from patients evaluated between January 2010 and January 2017. The study was approved by the Institutional Clinical Research

Ethics Committee (date: April 25, 2016; number: 08-341-16). Women under 45 years old with fresh embryo cycles were included. All included cycles underwent ICSI. The frozen embryo cycles, women with prior IVF/ICSI cycles, patients with secondary infertility, canceled cycles due to a nonviable sperm during testicular sperm extraction, cycles with >2 transferred embryos, and patients with donor sperm or egg, were excluded. Hospital records from patient files were used to create an anonymous dataset for internal validation. These records were searched manually by E.K. and A.G. Age, infertility duration, hysterosalpingography evaluation notes, body mass index (BMI), infertility indication, and ovarian reserve assessment at day 3 of menstrual cycle were used as patient demographics, and total gonadotropin dose, cycle duration, drugs used for ovarian induction, and sonographic assessments of the follicles and endometrium were used as cycle characteristics. Endometrial thickness, embryo quality, and embryo age were used as transfer characteristics.

After an initial assessment of patients with a detailed historical examination, semen analysis based on the World Health Organization criteria, ovarian reserve, and tubal patency assessment; ovarian stimulation was started during days 3 and 5 of the menstrual cycle. The starting dose was individualized based on patient age, ovarian reserve, and BMI. Further adjustment was also individualized based on the ovarian response assessment. The planned antagonist protocols (Cetrotide, Merck-Serono) were started after 5 days of gonadotropin usage or at least a 12 mm diameter of follicles were seen. Patients with a high risk of ovarian hyperstimulation syndrome were triggered with a dual trigger method or Gonadotrophin-releasing hormone (GnRH) agonists. Vaginal progesterone at 90 mg/day (Crinone 8% gel; Merck-Serono, Istanbul, Turkey) was used for luteal phase support from the day of embryo transfer to 12 weeks of gestational age. The ongoing pregnancy was defined as a pregnancy completed >20 weeks of gestational age. Antenatal follow-ups were organized based on the Ministry of Health guidelines.

Statistical Analysis

Descriptive statistics of all variables used in the study were investigated. The distribution properties of variables were evaluated with the Shapiro-Wilk test and were assumed with a normal distribution feature if the p value was >0.05. Theoretical quantile-quantile graphs of parameters with normal distribution properties (Shapiro-Wilk test $p>0.05$) were created

and the distribution assumption was visually tested. The variable presentation was made in the form of median value and interquartile range, and specific presentation types were not used for distribution assumption.

T-test or Wilcoxon rank-sum test was used following the distribution assumption of the examined variable for binary group comparisons. Logistic regression analysis was used to create the prediction model. First, all examined parameters were modeled alone and relative probability ratios, confidence interval (CI), and p-values were found. A variable selection model was applied to create the multiparameter model. Step models are prone to produce biased or incompatible clinical reality models, thus all combinations of clinically important parameters or parameters that are important in the univariate regression analysis ($p < 0.25$) were tested. The Akaike criterion was used as an aid in parameter selection, and the accuracy and calibration of the model were tested in each step^(10,11). The accuracy of the created models was tested with receiver operating characteristic curves. Model calibration was tested with the Hosmer-Lemeshow test and calibration curves⁽¹²⁾. A certain part of dataset was used to create the model in the study (60%) and all dataset was included in the validation stage (60% + 40%). The internal validation of the model was done using 10,000 different datasets created using the bootstrapping method. The deviation corrected CI of the parameters used for internal validation were found and operating characteristic curves were created with corrected optimism. All statistical analysis R for Windows: Software language for statistical computing (Version 3.1.3) and packages of the same program "pROC," "ModelGood," "rms," "caret," "boot," "ggplot2," and "ROC632" were used. Unless otherwise stated, 0.05 was accepted as the statistically significant p-value limit^(13,14).

Results

A total of 488 women who started controlled hyperstimulation for their first embryo transfer were included in the present study. The missing values other than basic patient characteristics were below 1% in the whole dataset. No imputation was applied to the dataset.

Among 488 cycles, 136 (27.9%) resulted in an ongoing pregnancy. The model was based on the comparison of the main characteristics of 305 patients [live birth (number): 85, failed cycles (number): 220], which were presented in Table 1. Age, antral follicle count, day 3 serum luteinizing hormone (LH) level, gonadotropin induction duration, mature oocyte total count, fertilization rate, endometrial thickness on the embryo transfer day, and transferred embryo grade were significantly different between patients with and without live birth.

Selection of Parameters for Univariate Regression Analysis and Prediction Model

Patient demographics, cycle, and transfer day characteristics were put into the one-way regression analysis. All parameters with plausible associations ($p < 0.250$) were tested. The cut-

off values for age, infertility duration, induction duration, day 3 serum LH level, endometrial thickness, and the total number of retrieved oocytes were visually determined based on the probability distribution graphs. Significant changes were observed in over 37 years of age, over 10 years of infertility duration, over 6 mIU/L for serum LH level, over 9 mm of endometrial thickness, and below four retrieved oocytes. Limit values were used for further regression analysis. Among the parameters examined are age above 37 years ($p=0.032$), low antral follicle number ($p=0.004$), basal LH levels above 6 mIU/mL ($p=0.001$), stimulation cycle lasting longer than 10 days ($p=0.003$), <500 pg/mL estradiol level ($p=0.077$) on the triggering day, <4 collected oocytes ($p=0.013$), and grade B embryo transfer ($p=0.005$). The significance levels were not obtained in other parameters; however, parameters with known clinical effects on live birth were prioritized with the prediction model creation.

Multivariate Regression Analysis and Creation and Calibration of the Predictive Model

Two separate prediction models were created. In the basal model, only the patient demographics were used. In the combined model, cycle and embryo transfer characteristics were included in the patient demographics. One parameter was added or subtracted at a time. The model accuracy and calibration were tested at each stage. Parameters without a significant model consistency increase or that impair its calibration were excluded. The patient age, basal antral follicle count, and day 3 serum LH level were used in the basal model (Table 2). The probability of success decreased (odds ratio: 0.07, 95% CI: 0.00-0.36) in patients aged over 37 years, success rate decreased (odds ratio: 0.15, 95% CI: 0.02-0.53) in <5 antral follicles, and LH greater than 6 mIU/mL (odds ratio: 2.24, 95% CI: 1.27-3.94) was associated with success. The accuracy of the tested basal model revealed an area under the curve (AUC) of 0.68 (95% CI: 0.62-0.74) and model sensitivity of 0.28 (95% CI: 0.17-0.39) for a 10% fixed false-positive rate (Figure 1). The calibration curve of the basal model revealed that the observed probabilities were consistent with the predicted probabilities. The Hosmer-Lemeshow test revealed that the model calibration was good under this observation ($p=0.979$).

In addition to the basal model parameters, the duration of the stimulation cycle, the endometrial thickness, and the number and grade of the embryo transferred were used in the combined model creation. A clinical and statistical interaction was found between the number and grade of embryos, thus, it was adapted to the model considering this feature. Cycles with 10 days or longer duration (odds ratio: 2.23, 95% CI: 1.17-4.42) and endometrial thickness wider than 9 mm (probability ratio: 2.07, 95% CI 1.00-4.53) were more successfully observed. The transferred embryo characteristics revealed a significantly increased successful single grade B embryo transfer, which was found as a negative effect (odds ratio: 0.07, 95% CI 0.00-0.39). The accuracy of the combined model revealed an AUC of 0.76

Table 1. Comparison of basic variables used to establish prediction model

Patient demographics	Live birth (n=85)	No live birth (n=220)	P†
Age (years), median (IQ)	29.0 (26.00-32.0)	31.0 (27.0-36.0)	0.004
Body Mass Index (kg/m ²), median (IQR)	22.4 (21.1-26.2)	23.8 (21.6-26.6)	0.132
Infertility duration (years), median (IQR)	5.0 (3.0-6.0)	5.0 (3.0-8.0)	0.664
Tubal occlusion			
- Unilateral, n (%)	2 (2.4)	7 (3.2)	0.999
- Bilateral, n (%)	4 (4.7)	8 (3.6)	0.743
Infertility etiology			
- Male factor, n (%)	28 (32.9)	93 (42.3)	0.152
- Female factor, n (%)	14 (16.5)	43 (19.5)	0.624
internal validation.	43 (50.6)	84 (38.2)	0.052
Antral follicle count, n (%)			
- ≥5 follicle	62 (72.9)	130 (59.1)	0.025
- <5 follicle	2 (2.4)	38 (17.3)	<0.001
- Polycystic ovaries	21 (24.7)	52 (23.6)	0.881
Day 3 serum FSH level (mIU/mL), median (IQR)	7.0 (6.0-9.0)	7.2 (6.0-9.1)	0.864
Day 3 serum estradiol level (pg/mL), median (IQR)	44.0 (29.5-56.0)	40.5 (30.0-54.0)	0.973
Day 3 serum LH level (mIU/mL), median (IQR)	5.3 (3.7-7.0)	5.0 (3.2-6.0)	0.045
Controlled ovarian hyperstimulation (COH) characteristics			
Initial gonadotropin dosage (IU), median (IQR)	225.0 (225.0-300.0)	225.0 (225.0-300.0)	0.894
Down-regulation method			
- Antagonist, n (%)	82 (96.5)	200 (90.9)	0.144
- Agonist, n (%)	3 (3.5)	20 (9.1)	
Type of gonadotropin for ovarian stimulation			
Only HMG, n (%)	23 (27.1)	73 (33.2)	0.337
- Only recombinant FSH, n (%)	12 (14.1)	25 (11.4)	0.558
- Recombinant FSH and HMG, n (%)	50 (58.8)	122 (55.4)	0.609
Ovarian trigger agent			
Only GnRH agonist, n (%)	7 (8.2)	21 (9.6)	0.827
- Only hCG, n (%)	63 (74.1)	167 (75.9)	0.767
GnRH agonist and hCG, n (%)	15 (17.7)	32 (14.5)	0.485
Total gonadotropin dosage (IU), median (IQR)	2400 (1800-3000)	2400 (2025-2925)	0.566
The number of >17 mm oocytes at trigger day, median (IQR)	3.0 (2.0-4.0)	3.0 (2.0-4.0)	0.242
COH duration (day), median (IQR)	12.0 (11.0-12.0)	11.0 (10.0-12.0)	0.003
Estradiol level at trigger day (pg/mL), median (IQR)	2018 (1370-3196)	1729 (993-2868)	0.071
Progesterone level at trigger day (ng/mL), median (IQR)	0.54 (0.41-1.00)	0.67 (0.47-1.05)	0.331
LH level at trigger day (mIU/mL), median (IQR)	2.0 (1.2-3.5)	2.0 (1.0-3.5)	0.571
OPU and ET characteristics			
Total picked up oocytes, median (IQR)	10.0 (6.0-14.0)	8.0 (5.0-13.0)	0.007
Mature (MII) oocyte number, median (IQR)	9.0 (4.0-12.0)	5.0 (3.0-10.0)	<0.001

Table 1. Continued

	Live birth (n=85)	No live birth (n=220)	P†
Fertilization rate (%), median (IQR)	69.2 (46.4-89.9)	50.0 (33.3-80.0)	0.007
Endometrial thickness at transfer day mm, median (IQR)	11.0 (10.0-12.0)	10.0 (9.0-12.0)	0.001
Embryo transfer number			
- Single, n (%)	65 (76.5)	151 (68.6)	0.206
- Double, n (%)	20 (23.5)	69 (31.4)	
Transferred embryo grade			
- Grade A, n (%)	83 (97.6)	185 (84.1)	<0.001
- Grade B, n (%)	2 (3.4)	35 (15.9)	
Transferred embryo age (day)			
Day 3 embryo, n (%)	66 (77.6)	184 (83.6)	0.246
Day 5 embryo, n (%)	19 (22.4)	36 (16.4)	

IQR: Interquartile range, HMG: Human menopausal gonadotropin, FSH: Follicle-stimulating hormone, GnRH: Gonadotropin-releasing hormone, hCG: Human chorionic gonadotropin, OPU: Oocyte-pick up, ET: Embryo transfer
 †Wilcoxon t-test or Fisher's exact test

Table 2. Odds ratios of models created based on multivariate logistic regression analysis

	Odds ratio (OR) (95% confidence interval (CI))	P†
Basal model (AUC: 0.68, 95% CI: 0.62-0.74)		
Age		
- ≤37 age	Reference	
- >37 age	0.07 (0.00-0.36)	0.011
Basal antral follicle count		
- ≥5 follicle	Reference	
- <5 follicle	0.15 (0.02-0.53)	0.012
- Polycystic ovaries	0.60 (0.32-1.09)	0.102
Day 3 serum LH level >6.0 mIU/mL	2.24 (1.27-3.94)	0.004
Combined model (AUC: 0.76, 95% CI: 0.70-0.81)		
Age		
- ≤37 age	Reference	
- >37 age	0.10 (0.00-0.55)	0.030
Basal antral follicle count		
- ≥5 follicle	Reference	
- <5 folikül	0.23 (0.03-0.96)	0.026
- Polycystic ovaries	0.50 (0.25-0.96)	0.043
Day 3 serum LH level >6.0 mIU/mL	2.36 (1.30-4.32)	0.004
COH duration		
- Shorter than 10 days	Reference	
- 10 days or longer	2.23 (1.17-4.42)	0.055
Endometrial thickness at transfer day >9 mm	2.07 (1.00-4.53)	0.016
Transferred embryo grade and number		
- Single, Grade A	Reference	
- Double, Grade A	0.90 (0.45-1.77)	0.697
- Single, Grade B	0.07 (0.00-0.39)	0.014
- Double, Grade B	0.39 (0.02-2.65)	0.456

AUC: Area under curve, CI: Confidence interval

†Multivariate logistic regression

(95% CI: 0.70-0.81) and model sensitivity of 0.31 (95% CI: 0.20-0.42) for a 10% fixed false-positive rate (Figure 2). The consistency of the combined model was statistically significantly higher than the baseline model (AUC: 0.76 vs AUC: 0.68, $p < 0.001$ De Long test, respectively). The calibration curve of the combined model revealed that the observed probabilities were consistent with the predicted probabilities. The Hosmer-Lemeshow test revealed a good model calibration following this observation ($p = 0.848$).

Nomograms were created for the practical application of the models (Supplementary Figure 1,2). The values from the lines

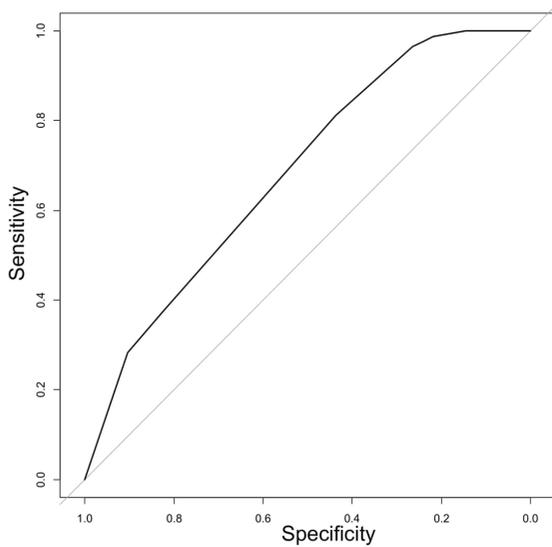


Figure 1. Receiver operating characteristic curve of the basal model using parameters of age, antral follicle count, and luteinizing hormone level

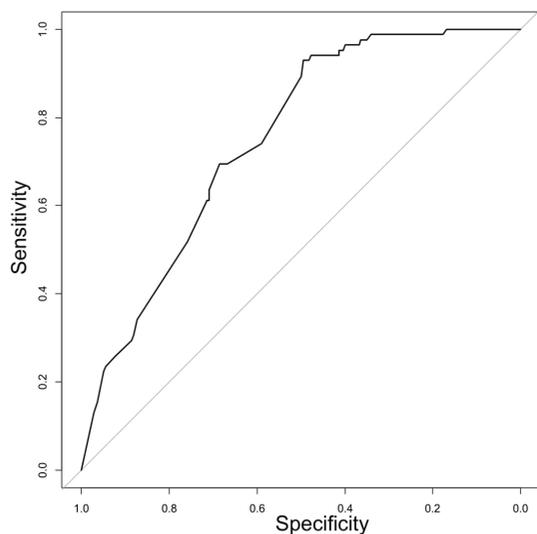
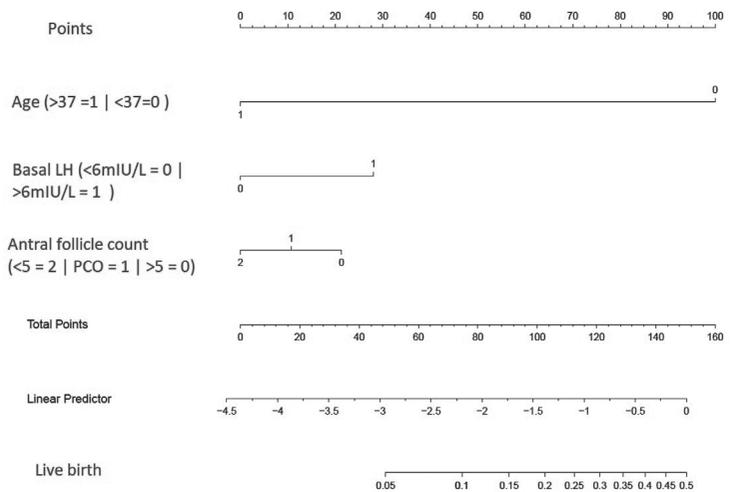


Figure 2. Receiver operating characteristic curve of the basal model using the parameters of age, antral follicle number, luteinizing hormone level, cycle duration, endometrial thickness, and embryo quality and number

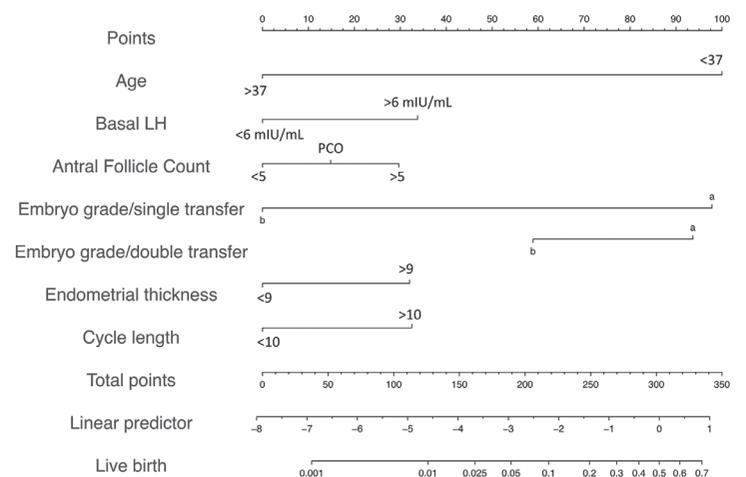
next to the parameters are marked first to use the nomogram. Each parameter score is calculated with the lines drawn perpendicular to the score curve above. After the scores are collected, the total score is marked in the total score line below and the possibility of live birth is read with the perpendicular line drawn below.

The Comparison Between the Templeton and the Present Models

A comparison was made with the Templeton model to show the practical benefit of the basal model. The Templeton model parameters were adapted to our dataset, and receiver operating characteristic curves were created for both the models. The AUC of the Templeton model was 0.60 (95% CI: 0.53-0.67) (Figure



Supplementary Figure 1. The nomogram of the basal model for clinical use. The probability of live births corresponding to the total score observed in the lowest row after the corresponding scores are added on the upper score sheet for each parameter



Supplementary Figure 2. The nomogram of the combined model for clinical use. The probability of live births corresponding to the total score observed in the lowest row after the corresponding scores are added on the upper score sheet for each parameter

3). The accuracy of the Templeton model was lower ($p=0.062$, DeLong test) than that of the basal model (Figure 3). The sensitivity of the Templeton model for a fixed 10% false positivity rate was very low for clinical use (0.10, 95% CI: 0.03-0.19).

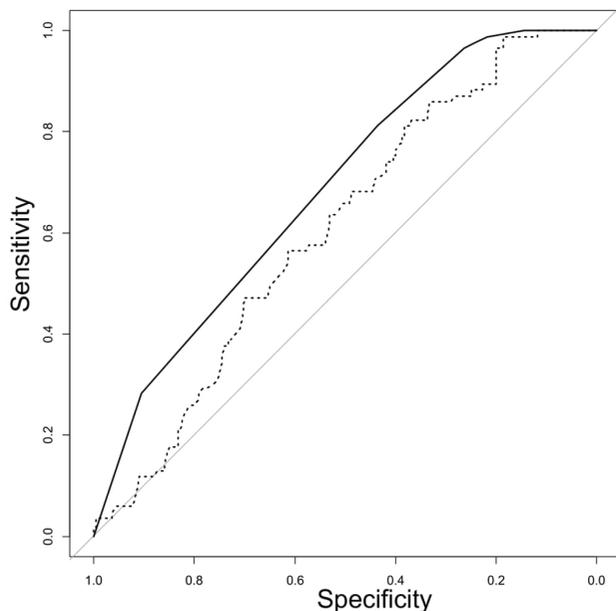


Figure 3. Comparison of the receiver operating characteristic curve of the Templeton model (dashed line) with the curve of our basal model (solid line). The higher predictive accuracy of the present model was demonstrated by the De Long test ($p=0.062$)

Discussion

The prediction models in the present study have acceptable precision and good calibration. The baseline model was used in the pretreatment phase while informing patients or making treatment decisions. In addition, static (Supplementary Figures 1,2) nomograms are available for practical use of the model, especially for clinicians.

More than 30 prediction models were presented in the literature, wherein most used similar parameters for model creation. Infertility duration and infertility type were not found to be associated with the odds of live birth in this model compared with more recognizable models. One of the reasons is the time censored nature of the infertility duration, which was inevitably affected by the patients' age and the unpredictable exact duration. A recent meta-analysis by van Loendersloot et al.⁽³⁾ found a weak association between the infertility duration and live birth (odds ratio: 0.99 95% CI: 15 0.98-1.00), which also concluded that among 21 external validation studies, only the model could be generalized, which includes female age, number of retrieved oocytes, developmental stage score, and morphology score of two best embryos. In the present prediction model, in addition to these parameters, day 3 serum LH level, and endometrial thickness were found significantly associated with the odds of live birth.

The most recent prediction model in the literature aimed to calculate "the number of mature oocytes required to obtain at least one euploid embryo"⁽¹⁵⁾. This model was externally validated and revealed >80% positive predictive values with all the predicted used possibilities by the authors^(15,16). The primary ending for the IVF/ICSI cycle outcome was different than our model; however, female age, sperm source used for ICSI, and the number of mature oocytes were used as parameters of the predictive model. In this study, testicular sperms were used as a sperm source in all of the included cycles to overcome the negative effects of the malefactors.

Endometrial preparation for successful implantation was another key phase of any cycle's endpoint^(17,18). Several factors revealed the functionality of the endometrium. The optimal endometrial thickness was 10 mm at the Vaegter's prediction model, and the impact of endometrial thickness on the models was also similar to ours⁽¹⁹⁾. Other than the endometrial thickness, the duration of gonadotropin induction showed significance on our model and is a possible indicator of ovarian and endometrial response. Day 3 serum LH level was an important parameter in this study. In two recent studies, basal serum LH level was highly associated with an ovarian response especially agonist protocols^(20,21). Lower LH levels during cycles are also related to a lower ongoing pregnancy rate, and ongoing pregnancy rates are higher at protocols that are supported with a recombinant LH based on a Cochrane review⁽²²⁾.

Stimulation durations longer than 10 days were associated with better outcomes in the first cycles, which is a reflection of the ovarian reserve and its effect on cycle success. Poor responders usually have a short stimulation duration due to already high endogenous follicle-stimulating hormone levels and asynchronous follicle growth. These patients have poorer outcomes compared to normo- and high-responders and our results reflect this mechanism.

Finally, transferred embryo grade and number were associated with ongoing pregnancy rates. This is an expected finding and was established in the literature.

Study Limitations

Several limitations were encountered in this study. Firstly, the number of included patients was below the average from similar studies in the literature. However, considering the patient volume of the clinic where the study was conducted and the included patient group, the number of patients was kept as high as possible and a wide range of years was chosen. In addition, the number of live births ($n=85$) in the cohort in which the model was developed is above the minimum number ($n=10$) per parameter in the logistic models⁽²³⁾. Therefore, the problem is not encountered in terms of statistical power. In addition, a possibility of selection bias is due to its retrospective nature. The possibility of a selection bias is never completely excluded although restrictive exclusion criteria were not set since the data source of the research was the patient files with

complete records. Another limitation was the indicator used for an ovarian reserve. The only parameter was the number of antral follicles and some studies reported that the serum anti-Müllerian hormone (AMH) level reflects the ovarian reserve better. The predictive value of the AMH level was not evaluated since AMH was not a routine parameter recorded in the years in which research records were obtained in our clinic. The internal validation of the developed model was made by developing a mixed-method due to the limited number of patients. The reserved patient population is mixed with the cohort in which the model was developed, and the validation study that performed with the bootstrapping method is more insufficient than the studies using the external cohort. Finally, some interventional procedures were reported with IVF success association, and the relationship of these factors was not studied in our patient population.

The main strength of this study was the patients treated with current IVF protocols and techniques. Given that the prediction models perform best in populations with characteristics similar to the developed cohorts, our model was expected to perform better in external validation studies compared to its historical counterparts. Since the parameters used in the model were easily measured and generally recorded in IVF cycles, no technical problems were expected in external validation studies. In addition, during the creation and testing of the model, the highest standard statistical practices were adhered to, and the model was created with careful attention to technical principles. The value of the area (0.76) remaining in the high curve quotation observed in our study is the result of careful parameter selection and good statistical practice. Finding static and dynamic nomograms for the practical use of our model was another strong aspect.

Conclusion

The present created model was well-calibrated and easily interpretable to routine IVF/ICSI cycles. The combined model aid in the informed decision phase of the fertility-seeking couples; however, external validation is necessary with a large-sized prospective cohort to confirm the clinical usage.

Ethics

Ethics Committee Approval: The study was approved by the Institutional Clinical Research Ethics Committee (date: April 25, 2016; number: 08-341-16).

Informed Consent: Retrospective study.

Peer-review: Externally peer-reviewed.

Authorship Contributions

Surgical and Medical Practices: Y.E.Ş., B.Ö., M.S., C.B., R.A., Concept: E.K., B.B., Design: E.K., B.B., Data Collection or Processing: E.K., C.B., A.G., Analysis or Interpretation: E.K., B.B., Literature Search: E.K., C.B., A.G., Y.E.Ş., Writing: E.K., C.B., B.B.

Conflict of Interest: The authors report no conflict of interest.

Financial Disclosure: The authors have no financial interests about the research.

References

1. Youssef MA, Van der Veen F, Al-Inany HG, Mochtar MH, Griesinger G, Nagi Mohesen M, et al. Gonadotropin-releasing hormone agonist versus HCG for oocyte triggering in antagonist-assisted reproductive technology. *Cochrane Database Syst Rev* 2014;10:CD008046. doi: 10.1002/14651858.CD008046.pub4.
2. Toftager M, Bogstad J, Løssl K, Prætorius L, Zedeler A, Bryndorf T, et al. Cumulative live birth rates after one ART cycle including all subsequent frozen-thaw cycles in 1050 women: secondary outcome of an RCT comparing GnRH-antagonist and GnRH-agonist protocols. *Hum Reprod* 2017;32:556-67.
3. van Loendersloot LL, van Wely M, Limpens J, Bossuyt PM, Repping S, van der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. *Hum Reprod Update* 2010;16:577-89.
4. Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: a prospective study of 144,018 treatment cycles. *PLoS Med* 2011;8:e1000386.
5. Templeton A, Morris JK, Parslow W. Factors that affect outcome of in-vitro fertilisation treatment. *Lancet* 1996;348:1402-6.
6. Smith AD, Tilling K, Lawlor DA, Nelson SM. External validation and calibration of IVFpredict: a national prospective cohort study of 130,960 in vitro fertilisation cycles. *PLoS One* 2015;10:e0121357. doi: 10.1371/journal.pone.0121357.
7. te Velde ER, Nieboer D, Lintsen AM, Braat DD, Eijkemans MJ, Habbema JD, et al. Comparison of two models predicting IVF success; the effect of time trends on model performance. *Hum Reprod* 2014;29:57-64.
8. Sunkara SK, Seshadri S. Increase in older women presenting as unexplained subfertility may explain overuse of in vitro fertilisation. *BMJ* 2014;348:g1583. doi: 10.1136/bmj.g1583.
9. Human Fertilisation and Embryology Authority. Pilot national fertility patient survey 2018. Available from: <https://www.hfea.gov.uk/media/2702/pilot-national-fertility-patient-survey-2018.pdf>
10. Greenland S. Modeling and variable selection in epidemiologic analysis. *Am J Public Health* 1989;79:340-9.
11. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. 1973 In: Petrov BN, Csaki F, editors. *Proceedings of the 2nd International Symposium on Information Theory*. Budapest: Akademiai Kiado; 1973. p. 267-81.
12. Hosmer W, Lemeshow S. *Applied logistic regression*. New York: Wiley; 2000.
13. Foucher Y. ROC632: Construction of diagnostic or prognostic scoring system and internal validation of its discriminative capacities based on ROC curve and 0.633+ bootstrap resampling. 2013; R package version 0.6. Available from: <http://CRAN.R-project.org/package=ROC632>.
14. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
15. Esteves SC, Carvalho JF, Bento FC, Santos J. A novel predictive model to estimate the number of mature oocytes required for obtaining at least one euploid blastocyst for transfer in couples undergoing in vitro fertilization/intracytoplasmic sperm injection:

- the ART calculator. *Front Endocrinol (Lausanne)* 2019;10:99.
16. Esteves SC, Yarali H, Ubaldi FM, Carvalho JF, Bento FC, Vaiarelli A, et al. Validation of ART calculator for predicting the number of metaphase II oocytes required for obtaining at least one euploid blastocyst for transfer in couples undergoing in vitro fertilization/ intracytoplasmic sperm injection. *Front Endocrinol (Lausanne)* 2019;10:917.
 17. Broekmans FJ, Verweij PJ, Eijkemans MJ, Mannaerts BM, Witjes H. Prognostic models for high and low ovarian responses in controlled ovarian stimulation using a GnRH antagonist protocol. *Hum Reprod* 2014;29:1688-97.
 18. Mackens S, Santos-Ribeiro S, van de Vijver A, Racca A, Van Landuyt L, Tournaye H, et al. Frozen embryo transfer: a review on the optimal endometrial preparation and timing. *Hum Reprod* 2017;32:2234-42.
 19. Vaegter KK, Lakic TG, Olovsson M, Berglund L, Brodin T, Holte J. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. *Fertil Steril* 2017;107:641-8.
 20. Depalo R, Trerotoli P, Chincoli A, Vacca MP, Lamanna G, Cicinelli E. Endogenous luteinizing hormone concentration and IVF outcome during ovarian stimulation in fixed versus flexible GnRH antagonist protocols: an RCT. *Int J Reprod Biomed* 2018;16:175-82.
 21. Gizzo S, Andrisani A, Noventa M, Manfè S, Oliva A, Gangemi M, et al. . Recombinant LH supplementation during IVF cycles with a GnRH-antagonist in estimated poor responders: a cross-matched pilot investigation of the optimal daily dose and timing. *Mol Med Rep* 2015;12:4219-29.
 22. Mochtar MH, Danhof NA, Ayeleke RO, Van der Veen F, van Wely M. Recombinant luteinizing hormone (rLH) and recombinant follicle stimulating hormone (rFSH) for ovarian stimulation in IVF/ICSI cycles. *Cochrane Database Syst Rev* 2017;5:CD005070. doi: 10.1002/14651858.CD005070.pub3.
 23. Harrell FE, Lee KL, Mark DB. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.