

The Progesterone-to-Oocyte Index Independently Predicts Live Birth in GnRH Antagonist Fresh IVF/ICSI Cycles: Development and Temporal Validation of a Nomogram

Xinxin Li

Third Affiliated Hospital of Sun Yat-sen University

Yushan Huang

Third Affiliated Hospital of Sun Yat-sen University

Xiya Zhang

Third Affiliated Hospital of Sun Yat-sen University

Tian Gan

Third Affiliated Hospital of Sun Yat-sen University

Weijie Xing

Third Affiliated Hospital of Sun Yat-sen University

Yue Meng

`mengy67@mail.sysu.edu.cn`

Third Affiliated Hospital of Sun Yat-sen University

Research Article

Keywords: progesterone-to-oocyte index, live birth prediction, nomogram, GnRH antagonist protocol, in vitro fertilization

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10861663/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

This study developed and temporally validated a nomogram for individualized live-birth prediction in GnRH antagonist fresh IVF/ICSI cycles, systematically comparing absolute trigger-day progesterone (P), the progesterone-to-follicle index (PFI), and the progesterone-to-oocyte index (POI).

Methods

This retrospective cohort included 2,365 fresh GnRH antagonist IVF/ICSI cycles (848 live births, 35.9%), chronologically partitioned into development (2015–2021, $n = 1,644$) and temporal validation (2022–2024, $n = 721$) cohorts. Predictor selection used a three-stage framework: univariable filtering ($P < 0.10$), VIF collinearity management with AIC-based comparison of P, PFI, and POI, and LASSO plus AIC stepwise selection.

Results

POI was selected as the optimal progesterone parameter (AIC 2041.45 vs. 2042.24 for PFI and 2046.44 for absolute P). The eight-variable nomogram incorporated age, infertility duration, endometrial thickness, Gn dose, oocytes retrieved, POI, available embryos, and embryos transferred. POI exhibited the strongest inverse association with live birth (adjusted OR 0.027, 95% CI 0.002–0.305; $P = 0.004$); the oocyte-count coefficient reversed from positive (OR 1.059) to negative (OR 0.951) after multivariable adjustment, reflecting a suppression effect. The model showed consistent discrimination (AUC 0.663 training, 0.666 optimism-corrected, 0.675 temporal validation), satisfactory calibration (Hosmer–Lemeshow $P \geq 0.129$), and enhanced performance in poor responders (AUC 0.791 for AFC < 5).

Conclusions

POI independently predicts live birth in GnRH antagonist fresh IVF/ICSI cycles and, when integrated into an eight-variable nomogram, enables trigger-day risk stratification to inform the fresh-transfer versus elective cryopreservation decision, with particular utility in poor ovarian responders.

Background

Compared with frozen-thawed embryo transfer, the fresh transfer strategy affords several distinct benefits, notably a shorter time to live birth (TTLB), lower economic burden, and a diminished likelihood of certain adverse perinatal outcomes [1, 2]. Within this context, the gonadotropin-releasing hormone (GnRH) antagonist protocol—characterized by a briefer stimulation duration—further abbreviates TTLB while circumventing several inherent limitations of cryopreservation cycles. As antagonist-based

regimens continue to be refined in routine clinical practice, optimizing fresh embryo transfer outcomes has become a central priority in assisted reproductive technology (ART), particularly in in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI) programs.

During the late follicular phase, progesterone orchestrates the endometrial transition from a proliferative to a secretory phenotype, a prerequisite for successful embryo implantation. Whether a premature elevation in serum progesterone compromises live birth probabilities remains a matter of ongoing debate, and no universally accepted threshold has been established. Furthermore, the ovarian response of an individual patient may modulate the progesterone level at which reproductive outcomes begin to deteriorate[3, 4]. Since absolute progesterone concentrations are inherently coupled to the number of hormonally active follicles, the progesterone-to-follicle index (PFI)—calculated as trigger-day serum progesterone (ng/mL) divided by the count of follicles measuring ≥ 14 mm—has been proposed as a normalized alternative. Accumulating evidence indicates that PFI outperforms raw progesterone values in forecasting assisted reproduction outcomes[5–8]. Nevertheless, because ultrasound-based follicle enumeration is susceptible to inter-observer variability, some investigators have advocated substituting the number of retrieved oocytes—a more reproducible endpoint—to derive the progesterone-to-oocyte index (POI)[9, 10].

The primary objective of the present study was to systematically compare the prognostic utility of serum progesterone, PFI, and POI for live birth prediction in GnRH antagonist fresh IVF/ICSI cycles. Beyond progesterone-related markers, the multifactorial determination of IVF success underscores the need to identify additional robust predictors and consolidate them into clinically translatable prediction frameworks—a priority in contemporary reproductive medicine.

Although numerous prognostic models and scoring instruments for IVF outcomes have been published over the past decade, a substantial proportion are constrained by limited cohort sizes[11], restricted applicability to specific subpopulations such as polycystic ovary syndrome (PCOS) patients[12], or an absence of external validation[13]. Accordingly, the present investigation sought to construct and validate a multivariable prediction model for live birth specifically tailored to GnRH antagonist fresh IVF/ICSI cycles. By leveraging clinical and embryological parameters routinely recorded in standard practice, we aimed to deliver a pragmatic bedside tool. Subgroup analyses were additionally conducted to evaluate model robustness across clinically heterogeneous patient strata. Such an instrument could facilitate individualized risk counseling, inform evidence-based treatment decisions, and advance the goal of personalized IVF care.

Methods

Study design and population

This retrospective cohort study was conducted at the Third Affiliated Hospital of Sun Yat-sen University and included patients undergoing fresh IVF/ICSI cycles with GnRH antagonist protocols. The study

period extended from January 2015 through December 2024 inclusive. The study was approved by the institutional review board, with a waiver of individual informed consent applicable to this de-identified retrospective design (Ethic no. II2026-284-01).

The inclusion criteria were: (i) fixed or flexible GnRH antagonist regimen was applied for ovarian stimulation and (ii) fresh cleavage-stage (Day 3) or blastocyst-stage (Day 5) embryo transfer was performed.

The exclusion criteria were: (i) women with untreated hydrosalpinx; (ii) women with congenital or acquired uterine malformations; (iii) cycles involving surgical sperm extraction; (iv) cycles with loss to follow-up; and (v) essential predictor variables missing in excess of 20% of cases.

From an initial pool of 2,686 screened cycles, 2,405 proceeded to analysis after application of the above criteria. A further 40 cycles (1.7% of the screened sample) were excluded during quality control of the derived progesterone indices. Both the PFI and the POI are ratios whose denominators approach zero in extremely poor responders (one or two follicles or oocytes), producing spuriously extreme values. The 0.5th and 99.5th centiles were therefore adopted as trimming boundaries (PFI retained range: 0.014–0.461; POI retained range: 0.012–0.470), consistent with recommendations for handling outliers in clinical prediction model development[14].

Finally, 2,365 cycles were included in the study. The cohort was partitioned chronologically, rather than by random sampling, to permit an evaluation of model performance under temporally shifted clinical conditions. Treatment cycles initiated between January 2015 and December 2021 ($n = 1,644$) served as the development sample; those initiated between January 2022 and December 2024 ($n = 721$) constituted an independent temporal holdout for external validation. Partitioning by calendar period, as opposed to random allocation, precludes information leakage between training and testing sets and more faithfully mirrors the deployment logic of a prediction model – namely, the application of historical data to future patients [15]. The flowchart tracing cohort assembly from initial screening through exclusions is provided (Fig. 1).

Clinical trial number: not applicable.

Ovarian stimulation and embryological procedures

Ovarian stimulation was initiated on cycle Day 2–3 using recombinant follicle-stimulating hormone (FSH) (Gonal-F, Merck Serono, Switzerland or Puregon, MSD, USA) at starting doses individually calibrated at 100–300 IU/day according to age, body mass index (BMI), antral follicle count (AFC), anti-Müllerian hormone (AMH), and previous ovarian response history. In the fixed antagonist protocol, the antagonist ganirelix (Orgalutran, MSD, USA) at 0.25 mg/day was administered starting on stimulation Day 5. In the flexible protocol, the antagonist was initiated based on specific criteria: the dominant follicle reached a mean diameter of ≥ 14 mm, serum estradiol (E_2) levels of ≥ 400 pg/mL or serum luteinizing hormone (LH) levels of ≥ 10 IU/L, and continued until the trigger day. Final oocyte maturation

was triggered using 250 µg of recombinant human chorionic gonadotropin (hCG) (Ovidrel, Merck Serono, Switzerland) when at least two follicles reached a diameter of ≥ 18 mm or three follicles reached a diameter of ≥ 17 mm.

Oocyte retrieval was performed 34–36 hours after triggering. The insemination method – conventional IVF or ICSI – was determined by semen parameters. Embryos were graded according to the Istanbul Consensus (cleavage stage) and the Gardner system (blastocyst stage). Fresh embryo transfer occurred 3 days (cleavage stage) or 5 days (blastocyst stage) after oocyte retrieval. The luteal phase was supported with vaginal sustained-release progesterone gel 90 mg/day (Crinone, Merck Serono, Switzerland) combined with oral or Intramuscular injection Progesterone, initiated the day of oocyte retrieval and continued through the 10th gestational week if conception occurred.

Data sources, variable definitions, and candidate predictors

Retrospective case data were abstracted from the center's electronic reproductive database; all records were de-identified before extraction.

Seventeen variables, selected through literature review and clinical judgment and spanning four conceptual domains, were specified as the initial candidate set:

- Patient baseline characteristics: female age (years), body mass index (BMI, kg/m²), infertility duration (years)
- Ovarian reserve indicators: serum AMH (ng/mL), antral follicle count (AFC), basal FSH (IU/L)
- Stimulation and response parameters: total gonadotropin (Gn) dose (IU), trigger-day serum E₂ (pg/mL), trigger-day serum LH (IU/L), trigger-day serum P (ng/mL), follicles ≥ 14 mm on trigger day (n), oocytes retrieved (n), endometrial thickness on trigger day (mm)
- Embryological factors: available transferable embryos (n), embryos transferred (n)
- Progesterone-derived indices: PFI (ng/mL) and POI (ng/mL)

The primary study endpoint was live birth, defined as the delivery of one or more live neonates ≥ 28 weeks of gestation. Outcomes were established by reconciling hospital electronic delivery records with post-treatment telephone follow-up. For deliveries occurring at external institutions, verification was obtained through medical record review or direct patient contact.

Sample size adequacy

Sample size adequacy was assessed using the events per variable (EPV) criterion[16, 17]. With 602 live births in the development cohort and 8 predictors retained, $EPV = 602 / 8 = 75.25$, exceeding both the conventional threshold of 10 and the more stringent recommendation of 20.

Statistical analyses

Statistical computing was performed in R (v4.2.1, R Foundation for Statistical Computing) and Python (v3.13) with the statsmodels, scikit-learn, and SciPy libraries. All tests were two-sided with $\alpha = 0.05$ unless otherwise noted. Continuous variables were summarized as mean $\pm$ standard deviation (SD) or median (interquartile range, IQR) depending on normality (Shapiro–Wilk test) and compared between live birth groups using the t-test or Mann–Whitney U test as appropriate. Categorical variables were presented as frequencies (%) and compared by χ^2 or Fisher’s exact test.

Predictor selection: a three-stage framework

Predictor selection was structured as a sequential three-stage process (Heinze et al., Stat Med 2018).

Stage 1 — Univariable Filter. Each of the 17 candidate predictors was screened via univariable logistic regression; those with $P < 0.10$ were carried forward.

Stage 2 — Collinearity Management and Progesterone Parameter Selection. Variance inflation factors (VIF) were computed for the retained variables. Given the algebraic interdependence of serum P, PFI, and POI, three parallel multivariable models were fitted, each containing one of the three terms alongside the remaining stage-1 covariates. The model yielding the lowest AIC and highest AUC was selected.

Stage 3 — Dual Criterion Selection (LASSO + AIC). Surviving variables were subjected to two complementary procedures: least absolute shrinkage and selection operator (LASSO) logistic regression with 10-fold cross-validation, retaining predictors with non-zero coefficients at $\lambda = \lambda_{\min}$ and bidirectional AIC-based stepwise selection. The final model comprised predictors selected by both methods.

Model specification and visual presentation

A multivariable logistic regression model was fitted, with adjusted odds ratios (OR) and 95% confidence intervals (CI) reported for each predictor. Nagelkerke's R^2 summarized overall model fit, and the model was presented as a nomogram (rms package) [18].

Model validation and performance assessment

Three complementary dimensions of model performance — discrimination, calibration, and clinical net benefit — were quantified.

Discrimination — The area under the curve (AUC) was reported with DeLong 95% CIs[19]. Three estimates were obtained: apparent AUC (training set), optimism-corrected AUC via 1,000 bootstrap resamples, and temporal validation AUC (2022–2024 cohort).

Calibration — Calibration was evaluated by locally estimated scatterplot smoothing (LOESS)-smoothed calibration plots (decile-binned), the Hosmer–Lemeshow test (10 groups; $P > 0.05$ indicating acceptable fit), and the GiViTI calibration belt[20]. The Brier score was also reported.

Clinical Utility — Decision curve analysis [21] and clinical impact curves [22] were used to assess clinical net benefit across a range of threshold probabilities.

Incremental prognostic value of POI

To isolate the unique contribution of POI beyond the information carried by the remaining predictors, the net reclassification improvement (NRI) and integrated discrimination improvement (IDI) statistics were computed [23]. These indices were derived by contrasting the full model with a reference model from which all progesterone-related variables (P, PFI, and POI) were excluded. The associated change in AUC was also documented.

Subgroup analyses

Model robustness was examined across five clinically informative stratifying dimensions: patient age (< 35 versus ≥ 35 years), serum AMH (< 1.2 versus ≥ 1.2 ng/mL), antral follicle count (< 5 versus ≥ 5), embryos transferred (one versus two), and total gonadotropin exposure (< 1,500 versus $\geq 1,500$ IU). For each subgroup, the AUC was estimated, and a cross-product interaction term between the stratifying variable and POI was tested within the full model to assess whether POI’s effect was modified by subgroup membership. Result summaries were presented as a forest plot.

Results

Study population and baseline characteristics

From an initial pool of 2,405 GnRH antagonist protocol fresh IVF/ICSI cycles, 2,365 cycles with 848 live-birth events (35.9%) were retained for analysis after quality-control exclusion of 40 cases (1.7%) with extreme progesterone-index values. The development cohort comprised 1,644 cycles treated between 2015 and 2021, with 602 live births (36.6%); the temporal validation cohort comprised 721 cycles treated between 2022 and 2024, with 246 live births (34.1%).

Table 1 summarizes the baseline characteristics of the two cohorts. The overall cohort had a median age of 33 years (IQR 29–37), a median AMH of 2.4 ng/mL (IQR 1.2–4.2), and a median AFC of 11 (IQR 7–15). The validation cohort was slightly older (median 34 vs. 32 years, $P < 0.001$), had lower ovarian reserve (median AMH 2.0 vs. 2.6 ng/mL, $P < 0.001$; median AFC 10 vs. 11, $P < 0.001$), received a higher Gn dose (median 2,100 vs. 1,800 IU, $P < 0.001$), and had marginally fewer retrieved oocytes (median 8 vs. 9, $P = 0.013$). Trigger-day E_2 , LH, and progesterone levels were also significantly lower in the validation cohort (all $P < 0.001$). Despite these temporal differences, live-birth rates (34.1% vs. 36.6%), endometrial thickness ($P = 0.059$), available embryos ($P = 0.104$), and the derived progesterone indices (PFI, $P =$

0.364; POI, P = 0.518) remained broadly comparable, supporting the use of the later cohort as an independent external validation sample.

Table 1. Baseline characteristics of the development and temporal validation cohorts

Variable	Train (≤ 2021) (N=1644)	Validation (≥ 2022) (N=721)	Overall (N=2365)	P-value
Age (years)	32.0 [29.0, 36.0]	34.0 [31.0, 38.0]	33.0 [29.0, 37.0]	<0.001
BMI (kg/m ²)	21.3 [19.6, 23.4]	21.6 [19.8, 23.7]	21.4 [19.7, 23.5]	0.043
AMH (ng/mL)	2.6 [1.3, 4.7]	2.0 [1.1, 3.5]	2.4 [1.2, 4.2]	<0.001
AFC (n)	11.0 [7.0, 15.0]	10.0 [7.0, 14.0]	11.0 [7.0, 15.0]	<0.001
bFSH (IU/L)	6.8 [5.9, 8.2]	6.6 [5.7, 8.0]	6.8 [5.8, 8.2]	0.053
Infertility duration (years)	3.0 [2.0, 5.0]	3.0 [2.0, 5.0]	3.0 [2.0, 5.0]	0.665
Total Gn dose (IU)	1800.0 [1475.0, 2250.0]	2100.0 [1800.0, 2475.0]	1850.0 [1500.0, 2400.0]	<0.001
Trigger-day E ₂ (pg/mL)	1890.0 [1279.8, 2706.0]	1780.8 [1174.3, 2411.6]	1858.4 [1243.0, 2629.0]	<0.001
Trigger-day LH (IU/L)	2.9 [1.9, 4.4]	2.0 [1.2, 3.2]	2.6 [1.7, 4.0]	<0.001
Trigger-day P (ng/mL)	0.6 [0.4, 0.8]	0.5 [0.3, 0.7]	0.5 [0.3, 0.8]	<0.001
≥ 14 mm follicles (n)	7.0 [5.0, 9.0]	6.0 [4.0, 9.0]	7.0 [4.0, 9.0]	<0.001
Oocytes retrieved (n)	9.0 [6.0, 12.0]	8.0 [5.0, 12.0]	8.0 [5.0, 12.0]	0.013
Endometrial thickness (mm)	10.6 [9.3, 11.9]	10.7 [9.5, 12.1]	10.6 [9.4, 12.0]	0.059
Embryos transferred (n)	2.0 [1.0, 2.0]	2.0 [1.0, 2.0]	2.0 [1.0, 2.0]	<0.001
Available embryos (n)	4.0 [3.0, 6.0]	4.0 [3.0, 6.0]	4.0 [3.0, 6.0]	0.104
PFI (P/ ≥ 14 mm follicles)	0.1 [0.1, 0.1]	0.1 [0.1, 0.1]	0.1 [0.1, 0.1]	0.364
POI (P/oocytes)	0.1 [0.0, 0.1]	0.1 [0.0, 0.1]	0.1 [0.0, 0.1]	0.518

Data are presented as median [interquartile range (IQR)]. Normality was assessed using the Shapiro-Wilk test; all variables were non-normally distributed (P < 0.05).

P-values were calculated using the Mann-Whitney U test (Wilcoxon rank-sum test), comparing the development cohort (≤ 2021) with the temporal validation cohort (≥ 2022).

Abbreviations: AFC, antral follicle count; AMH, anti-Müllerian hormone; BMI, body mass index; bFSH, basal follicle-stimulating hormone; E₂, estradiol; Gn, gonadotropin; IU, international units; IQR, interquartile range; LH, luteinizing hormone; P, progesterone; PFI, progesterone-to-follicle index; POI, progesterone-to-oocyte index.

Model development

After univariable screening ($P < 0.10$) and collinearity management among absolute progesterone, PFI, and POI, POI emerged as the optimal progesterone parameter (lowest AIC, 2041.45). The univariable associations of all candidate predictors are provided in **Supplementary Table 1**, and the variance-inflation-factor assessment confirming absence of substantial multicollinearity is summarized in **Supplementary Table 2**. The three-way comparison among absolute P, PFI, and POI showed POI to be the best-supported index (AIC 2041.45 vs. 2042.24 for PFI and 2046.44 for absolute P; **Supplementary Table 3**). Dual LASSO and AIC stepwise selection identified eight independent predictors for the final multivariable logistic regression model, whose adjusted associations with live birth are reported in **Table 2**; the LASSO coefficient path and ten-fold cross-validation curve are shown in **Supplementary Figure 1**.

Table 2. Multivariable logistic regression analysis for live birth prediction

Variable	β	OR	95% CI	P-value
Female age (years)	-0.072	0.930	0.908–0.953	<0.001
Endometrial thickness (mm)	0.068	1.070	1.011–1.133	0.020
Oocytes retrieved (n)	-0.050	0.951	0.915–0.988	0.009
Infertility duration (years)	-0.032	0.968	0.929–1.009	0.124
POI (trigger-day P / oocytes)	-3.621	0.027	0.002–0.305	0.004
Total Gn dose (IU)	0.000	1.000	1.000–1.000	0.123
Available embryos (n)	0.173	1.189	1.118–1.264	<0.001
Embryos transferred (n)	0.333	1.396	1.128–1.726	0.002

Model fitted on the development cohort ($n = 1,644$). All variables entered as continuous predictors. AIC = 2,041.45. The coefficient for oocytes retrieved reversed sign from univariable (OR = 1.059) to multivariable analysis, reflecting a suppression effect mediated by POI and embryo-quality variables.

Abbreviations: AIC, Akaike information criterion; CI, confidence interval; Gn, gonadotropin; IU, international units; OR, odds ratio; P, progesterone; POI, progesterone-to-oocyte index; β , regression coefficient.

Fig. 2 presents the resulting nomogram for bedside risk estimation. Higher age, fewer retrieved oocytes, longer infertility duration, and higher POI each reduced the predicted probability of live birth, whereas greater endometrial thickness, higher Gn dose, more available embryos, and a greater number of

embryos transferred increased it. The most pronounced effect was observed for POI: each unit increase in POI was associated with a marked reduction in the odds of live birth (adjusted OR 0.027, 95% CI 0.002–0.305; P = 0.0035). Notably, the coefficient for oocytes retrieved reversed from positive in univariable analysis (OR 1.059) to negative after multivariable adjustment (OR 0.951, 95% CI 0.915–0.988; P = 0.0094), reflecting the suppression effect of POI and embryo-quality variables on the total progesterone burden pathway.

Model discrimination and calibration

The final model demonstrated acceptable discrimination in both the development and validation cohorts (**Table 3**). The apparent AUC was 0.663 (95% CI 0.636–0.689) in the training cohort, and the optimism-corrected AUC after 1,000 bootstrap resamples was 0.666. In the temporal validation cohort, the AUC was 0.675 (95% CI 0.634–0.715), indicating consistent discrimination across time periods (**Fig. 3**).

Table 3. Discriminative performance of the prediction model across validation methods

Validation method	AUC (95% CI)
Apparent (training, ≤2021)	0.663 (0.636–0.689)
Optimism-corrected (1,000 bootstrap)	0.666 (0.638–0.684)
Temporal validation (≥2022)	0.675 (0.634–0.715)

The apparent AUC was computed on the training cohort (n = 1,644). The optimism-corrected AUC was estimated via 1,000 bootstrap resamples of the training data. The temporal validation AUC was evaluated on the independent validation cohort (n = 721) comprising cycles from 2022 onwards. 95% confidence intervals were estimated using the DeLong method.

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval.

Calibration was satisfactory in both datasets (**Table 4**). The Hosmer–Lemeshow goodness-of-fit test yielded P = 0.129 in the training cohort and P = 0.302 in the validation cohort, and calibration slopes were close to unity (0.981 and 0.986, respectively). Brier scores were 0.215 and 0.208, respectively. The GiViTI calibration belt confirmed no meaningful localized miscalibration (**Supplementary Figure 2**).

Table 4. Calibration performance of the prediction model

Dataset	Brier score	H-L χ^2	H-L P-value	Calibration slope	Calibration intercept
Training (n = 1,644)	0.215	12.525	0.129	0.981	0.007
Validation (n = 721)	0.208	9.495	0.302	0.986	-0.019

The Hosmer–Lemeshow goodness-of-fit test was performed with 10 decile groups; $P > 0.05$ indicates acceptable calibration. Calibration slope close to 1 and intercept close to 0 indicate excellent calibration. Brier scores range from 0 to 1, with lower values indicating better predictive accuracy.

Abbreviations: H-L, Hosmer–Lemeshow; χ^2 , chi-square.

Clinical utility

Decision curve analysis showed that the model provided positive net benefit above both the treat-all and treat-none strategies across a clinically relevant range of threshold probabilities in both the training and validation cohorts (**Fig. 4A**). The corresponding clinical impact curves indicated that, at a threshold probability of approximately 0.30, the model would flag roughly 600–700 high-risk patients per 1,000 screened, of whom approximately 350–400 would be true positives, while the number of false positives remained comparatively low (**Fig. 4B**). These findings were consistent between the development and validation samples, supporting the potential clinical utility of the model for risk stratification.

Supplementary Table 5 reports the categorical net reclassification improvement (NRI = 0.033) and integrated discrimination improvement (IDI = 0.005) achieved by adding POI to a model containing the remaining seven predictors, indicating modest but consistent incremental value of the progesterone index.

Subgroup performance

Fig. 5 displays the model's discriminative performance across clinically relevant subgroups. The AUC was highest in women with AFC < 5 (0.791, 95% CI 0.673–0.908) and in women with AMH < 1.2 ng/mL (0.730, 95% CI 0.666–0.794), suggesting particular value in poor ovarian responders. Performance was also higher in women aged ≥ 35 years (0.691, 95% CI 0.637–0.744) than in those aged < 35 years (0.608, 95% CI 0.574–0.642), and in cycles with two or more embryos transferred (0.675, 95% CI 0.639–0.711) than in single-embryo transfer cycles (0.642, 95% CI 0.598–0.686). In all subgroups, the AUC exceeded 0.60, indicating stable predictive ability across diverse clinical scenarios.

Formal interaction testing was performed to evaluate potential effect modification of the number of embryos transferred across pre-specified subgroups. A significant interaction was observed between embryo number and age group (P for interaction = 0.012), indicating that the prognostic contribution of transferred embryo counts to live birth differed between women aged < 35 and ≥ 35 years. No significant interactions were detected with AMH (P = 0.534), AFC (P = 0.432), or Gn dose (P = 0.538), suggesting that the effect of embryo number remained consistent across these stratifications.

Discussion

In this retrospective cohort of 2,365 fresh GnRH antagonist IVF/ICSI cycles, we developed and temporally validated an eight-variable nomogram for individualized live-birth prediction. POI was systematically selected over both absolute trigger-day P and PFI through AIC comparison (AIC 2041.45

vs 2042.24 vs 2046.44) and VIF screening. The final model – incorporating age, infertility duration, endometrial thickness, Gn dose, oocytes retrieved, POI, available embryos, and embryos transferred – showed stable discrimination (AUC 0.663, 0.666 after bootstrap correction, 0.675 in the 2022–2024 holdout), satisfactory calibration (Hosmer–Lemeshow $P \geq 0.129$; GiViTI belt non-significant), and positive net benefit on decision curve analysis. Performance was notably enhanced in poor ovarian responders (AUC 0.791 for AFC < 5; 0.730 for AMH < 1.2 ng/mL), enabling cycle-specific risk stratification to inform the fresh-transfer versus freeze-all decision.

This selection of POI extends a conceptual evolution driven by the recognition that raw progesterone confounds physiologically appropriate output from a large follicular cohort with pathologically excessive per-unit secretion in diminished responders. PFI (P per ≥ 14 -mm follicle) was the first attempt to resolve this ambiguity, demonstrating that per-follicle normalization linearly predicts outcome across its full range whereas absolute P is detrimental only at extreme percentiles[5]. POI (P per retrieved oocyte) further refines this approach by normalizing to the functional output of the follicular cohort, capturing the progesterone burden per developmental unit more directly and predicting implantation and live birth without affecting fertilization[10, 24]. In our model, POI independently predicted live birth (adjusted OR 0.027, $P = 0.0035$). The detrimental mechanism is endometrial rather than embryonic: elevated P alters endometrial gene expression [25, 26] while leaving embryo ploidy unaffected [27], rationally supporting a freeze-all strategy when POI is elevated. The disproportionate per-follicle P output in poor responders – the subgroup with the greatest model enhancement – is explained by FSH-driven LH-receptor upregulation on granulosa cells [28] and disrupted oocyte–granulosa gatekeeper signaling [29], accounting for superior discrimination precisely where absolute P thresholds are least informative. The FSH-driven physiology of P elevation in antagonist cycles further identifies stimulation-dose modulation as a testable strategy to reduce per-unit progesterone burden in subsequent cycles.

A particularly noteworthy finding is the reversal of the oocyte-count coefficient from positive in univariable analysis (OR 1.059) to negative after multivariable adjustment (OR 0.951) – a classical suppression effect, i.e., a reversal paradox wherein the marginal association between two variables reverses sign upon conditional adjustment for a third correlated variable[30, 31], decomposable here into three pathways. First, the embryonic advantage pathway (positive): more oocytes yield more available embryos and thus higher live-birth probability, an effect captured by embryo variables (available embryos, embryos transferred) in the multivariable model[32, 33]. Second, the per-oocyte P burden pathway (negative): the POI variable independently captures the detrimental endometrial effect of elevated progesterone per oocyte[10, 24]. Third, the residual total-P load pathway (negative): at a fixed POI, more oocytes necessarily increase total circulating progesterone (total P = POI $\times$ oocyte count), which independently advances endometrial secretory maturation and disrupts embryo–endometrial synchrony[26, 27, 34] – an effect captured by the oocyte-count coefficient itself. Nested model comparisons confirmed that this sign reversal requires both POI and embryo variables to be simultaneously present, consistent with a suppression effect in which collinear predictors unmask a directional effect otherwise hidden by confounding. This three-path decomposition, unreported in prior

PFI/POI literature, provides a mechanistic framework for understanding how normalized and absolute progesterone metrics interact within a multivariable prediction model.

Several nomograms for live-birth prediction in IVF/ICSI have recently been published, yet none has systematically evaluated and selected a normalized progesterone index through formal statistical comparison. The most directly comparable model — Zhong et al.[11], a six-variable antagonist-protocol nomogram (AUC 0.72/0.81) — simultaneously entered PFI, absolute trigger-day P, and P/total follicles without VIF screening, introducing coefficient instability from three highly collinear progesterone metrics; our AIC-guided three-way selection explicitly resolved this collinearity by retaining only POI. Moreover, Zhong's validation cohort (168 cycles, EPV $\approx$ 13.2) may partly explain the unusually high validation AUC, whereas our 735-cycle temporal holdout (EPV = 32.9) yielded a more conservative AUC increment (0.663 $\rightarrow$ 0.675). Two PCOS-specific antagonist nomograms[12, 35] achieved higher AUCs (0.804 and 0.711), but both used absolute trigger P rather than a normalized index and likely benefit from the constrained prognostic heterogeneity of PCOS populations (mean oocytes 16.65); Si et al. further included trigger-day LH, which lacked predictive value in our non-PCOS cohort (AUC 0.506, P = 0.084). General-population models omitting cycle-specific hormonal dynamics — such as the four-variable Zubeyde Hanim nomogram (AUC 0.64)[13] — achieve discrimination no higher than ours despite spanning mixed protocols, underscoring the incremental value of POI (NRI 0.033; IDI 0.005). Finally, all five prior nomograms relied on random data partitioning, which cannot detect temporal drift in predictor-outcome relationships [14]; our temporal split ($\leq$ 2021 vs $\geq$ 2022) more faithfully mimics prospective deployment and provides stronger evidence of transportability.

This study possesses both methodological and clinical strengths. Methodologically, the formal AIC-plus-VIF three-way selection among absolute P, PFI, and POI resolves the collinearity inconsistency of prior models, and temporal validation ($\leq$ 2021 vs $\geq$ 2022) provides a more realistic estimate of prospective performance than random splitting. Clinically, the nomogram translates POI into a bedside scoring instrument that enables trigger-day risk stratification between fresh transfer and elective cryopreservation, precisely when clinicians can still alter cycle management; and the mechanistic evidence that elevated P impairs endometrial receptivity but not embryo ploidy [27] rationally supports freeze-all rather than cancel-and-restart when POI is elevated, preserving the cycle's reproductive investment. The model's enhanced discrimination in poor ovarian responders (AUC 0.791 for AFC < 5; 0.730 for AMH < 1.2 ng/mL) further identifies the subgroup in which absolute P thresholds are least informative and intervention yield is highest. Several limitations warrant acknowledgment: the retrospective single-center design precludes causal inference and may introduce selection bias; temporal validation remains internal rather than independent multicenter external validation, and transportability to different populations or protocols requires confirmation; the progesterone range (0.01–1.72 ng/mL) lacked cycles with $P > 2.0$ ng/mL, limiting applicability to high-responder populations where threshold dynamics may differ; and embryo quality was represented only by available-embryo counts without morphological grading. Prospective multicenter validation, incorporation of embryo scoring, and a POI-guided stimulation-dose randomized trial constitute the most pressing next steps.

Conclusions

In GnRH antagonist fresh IVF/ICSI cycles, the progesterone-to-oocyte index (POI) — selected over absolute trigger-day progesterone and the progesterone-to-follicle index through formal AIC-based comparison — independently predicts live birth and, when integrated into an eight-variable nomogram, enables trigger-day risk stratification between fresh transfer and elective cryopreservation. The model's enhanced performance in poor ovarian responders, coupled with mechanistic evidence that elevated progesterone impairs endometrial receptivity rather than embryo ploidy, supports a freeze-all strategy when POI is elevated and identifies stimulation-dose modulation as a testable intervention to reduce per-unit progesterone burden in subsequent cycles.

Abbreviations

GnRH — gonadotropin-releasing hormone; IVF — in vitro fertilization; ICSI — intracytoplasmic sperm injection; P — progesterone; PFI — progesterone-to-follicle index; POI — progesterone-to-oocyte index; VIF — variance inflation factor; AIC — Akaike information criterion; LASSO — least absolute shrinkage and selection operator; AUC — area under the curve; AMH — anti-Müllerian hormone; AFC — antral follicle count; Gn — gonadotropin; OR — odds ratio; CI — confidence interval; TTLB — time to live birth; ART — assisted reproductive technology; PCOS — polycystic ovary syndrome; FSH — follicle-stimulating hormone; BMI — body mass index; E₂ — estradiol; LH — luteinizing hormone; hCG — human chorionic gonadotropin; EPV — events per variable; SD — standard deviation; IQR — interquartile range; LOESS — locally estimated scatterplot smoothing; NRI — net reclassification improvement; IDI — integrated discrimination improvement

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the Third Affiliated Hospital of Sun Yat-sen University (protocol code II2026-284-01, date of approval: 29 July 2026). Informed consent was waived by the Institutional Review Board due to the retrospective nature of the study, which exclusively used de-identified data extracted from the center's electronic reproductive database. Clinical trial number: not applicable.

Consent for publication

Not applicable. This retrospective study used de-identified data, and no individual patient details, images, or videos are presented.

Availability of data and materials

The datasets used and/or analysed during the current study are not publicly available due to privacy and ethical constraints but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Clinical Research Program of the Third Affiliated Hospital of Sun Yat-sen University, China (grant number YHJH202209), the MOE Key Laboratory of Gene Function and Regulation, and the Guangzhou Science and Technology Planning Project (grant number 202201020411).

Authors' contributions

XL and YM conceived and designed the study. XL, YH, XZ, and TG collected the data. XL and YH performed the statistical analysis. XZ validated the analysis. TG and WX provided resources. XL and YH drafted the manuscript. WX and YM revised the manuscript critically. XL and XZ created the visualizations. WX and YM supervised the project. YM administered the project. WX and YM acquired funding. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Maheshwari A, Bell JL, Bhide P, Brison D, Child T, Chong HY, Cheong Y, Cole C, Coomarasamy A, Cutting R, et al. Elective freezing of embryos versus fresh embryo transfer in IVF: a multicentre randomized controlled trial in the UK (E-Freeze). *Hum Reprod.* 2022;37(3):476–87.
2. Zaat T, Zagers M, Mol F, Goddijn M, van Wely M, Mastenbroek S. Fresh versus frozen embryo transfers in assisted reproduction. *Cochrane Database Syst Rev.* 2021;2(2):CD011184.
3. Xu B, Li Z, Zhang H, Jin L, Li Y, Ai J, Zhu G. Serum progesterone level effects on the outcome of in vitro fertilization in patients with different ovarian response: an analysis of more than 10,000 cycles. *Fertil Steril.* 2012;97(6):1321–e13271321.
4. Griesinger G, Mannaerts B, Andersen CY, Witjes H, Kolibianakis EM, Gordon K. Progesterone elevation does not compromise pregnancy rates in high responders: a pooled analysis of in vitro fertilization patients treated with recombinant follicle-stimulating hormone/gonadotropin-releasing hormone antagonist in six trials. *Fertil Steril.* 2013;100(6):1622–e16281621.
5. Shufaro Y, Sapir O, Oron G, Ben Haroush A, Garor R, Pinkas H, Shochat T, Fisch B. Progesterone-to-follicle index is better correlated with in vitro fertilization cycle outcome than blood progesterone level. *Fertil Steril.* 2015;103(3):669–74. e663.
6. Roque M, Valle M, Sampaio M, Geber S, Checa MA. Ratio of progesterone-to-number of follicles as a prognostic tool for in vitro fertilization cycles. *J Assist Reprod Genet.* 2015;32(6):951–7.

7. Tokgoz VY, Ekici GC, Tekin AB. The efficiency of progesterone/estradiol and progesterone/follicle ratio without elevated trigger-day progesterone levels on the reproductive outcomes of GnRH antagonist IVF/ICSI cycles. *Gynecol Endocrinol*. 2021;37(10):885–90.
8. Kadhim AA, Abdulla TH, Alkawaz UM. Using progesterone to follicular index ratio is better correlated with intracytoplasmic sperm injection outcome than using serum progesterone level alone. *Ginekol Pol*. 2024;95(7):525–30.
9. Singh N, Malik N, Malhotra N, Vanamail P, Gupta M. Impact of progesterone (on hCG day)/oocyte ratio on pregnancy outcome in long agonist non donor fresh IVF/ICSI cycles. *Taiwan J Obstet Gynecol*. 2016;55(4):503–6.
10. Grin L, Mizrachi Y, Cohen O, Lazer T, Liberty G, Meltcer S, Friedler S. Does progesterone to oocyte index have a predictive value for IVF outcome? A retrospective cohort and review of the literature. *Gynecol Endocrinol*. 2018;34(8):638–43.
11. Zhong Y, Kang Q, Pang X, Wang N. Nomogram-Based Prediction of Live Birth in GnRH Antagonist Protocol Fresh IVF/ICSI Cycles. *Int J Womens Health*. 2025;17:4143–64.
12. Zhu S, Chen X, Li R, Jiang W, Zheng B, Sun Y. Constructing a predictive model for live birth following fresh embryo transfer in antagonist protocol for polycystic ovary syndrome. *J Assist Reprod Genet*. 2024;41(10):2709–19.
13. Karacin P, Ozelci R, Kumcu E, Kaya Kaplanoglu D, Dilbaz S, Ustun Y. A Clinically Applicable Nomogram for Live Birth Prediction After IVF: The Zubeyde Hanim Model. *J Clin Med*. 2026;15(3).
14. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
15. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York, NY: Springer; 2009.
16. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49(12):1373–9.
17. Riley RD, Ensor J, Snell KIE, Harrell FE Jr., Martin GP, Reitsma JB, Moons KGM, Collins G, van Smeden M. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441.
18. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.
19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–45.
20. Nattino G, Finazzi S, Bertolini G. A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. *Stat Med*. 2014;33(14):2390–407.
21. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26(6):565–74.

22. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *J Clin Oncol*. 2016;34(21):2534–40.
23. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27(2):157–72. discussion 207 – 112.
24. Aflatoonian A, Davar R, Hojjat F. Elevated serum progesterone/ MII oocyte ratio on the day of human chorionic gonadotropin administration can predict impaired endometrial receptivity. *Iran J Reprod Med*. 2014;12(6):427–34.
25. Van Vaerenbergh I, Fatemi HM, Blockeel C, Van Lommel L, In't Veld P, Schuit F, Kolibianakis EM, Devroey P, Bourgain C. Progesterone rise on HCG day in GnRH antagonist/rFSH stimulated cycles affects endometrial gene expression. *Reprod Biomed Online*. 2011;22(3):263–71.
26. Labarta E, Martinez-Conejero JA, Alama P, Horcajadas JA, Pellicer A, Simon C, Bosch E. Endometrial receptivity is affected in women with high circulating progesterone levels at the end of the follicular phase: a functional genomics analysis. *Hum Reprod*. 2011;26(7):1813–25.
27. Kofinas JD, Mehr H, Ganguly N, Biley Y, Bochkovsky S, McCulloh D, Grifo J. Is it the egg or the endometrium? Elevated progesterone on day of trigger is not associated with embryo ploidy nor decreased success rates in subsequent embryo transfer cycles. *J Assist Reprod Genet*. 2016;33(9):1169–74.
28. Friis Wang N, Skouby SO, Humaidan P, Andersen CY. Response to ovulation trigger is correlated to late follicular phase progesterone levels: A hypothesis explaining reduced reproductive outcomes caused by increased late follicular progesterone rise. *Hum Reprod*. 2019;34(5):942–8.
29. Younis JS. The role of progesterone/estradiol ratio in exploring the mechanism of late follicular progesterone elevation in low ovarian reserve women. *Med Hypotheses*. 2019;125:126–8.
30. Tu YK, Gunnell D, Gilthorpe MS. Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon—the reversal paradox. *Emerg Themes Epidemiol*. 2008;5:2.
31. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. *Prev Sci*. 2000;1(4):173–81.
32. Sunkara SK, Rittenberg V, Raine-Fenning N, Bhattacharya S, Zamora J, Coomarasamy A. Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles. *Hum Reprod*. 2011;26(7):1768–74.
33. Vermey BG, Chua SJ, Zafarmand MH, Wang R, Longobardi S, Cottell E, Beckers F, Mol BW, Venetis CA, D'Hooghe T. Is there an association between oocyte number and embryo quality? A systematic review and meta-analysis. *Reprod Biomed Online*. 2019;39(5):751–63.
34. Kalakota NR, George LC, Morelli SS, Douglas NC, Babwah AV. Towards an Improved Understanding of the Effects of Elevated Progesterone Levels on Human Endometrial Receptivity and Oocyte/Embryo Quality during Assisted Reproductive Technologies. *Cells*. 2022;11(9).

35. Si M, Jiang H, Zhao Y, Qi X, Li R, Long X, Qiao J. Nomogram for Predicting Live Birth after the First Fresh Embryo Transfer in Patients with PCOS Undergoing IVF/ICSI Treatment with the GnRH-Ant Protocol. *Diagnostics (Basel)*. 2023;13(11).

Figures

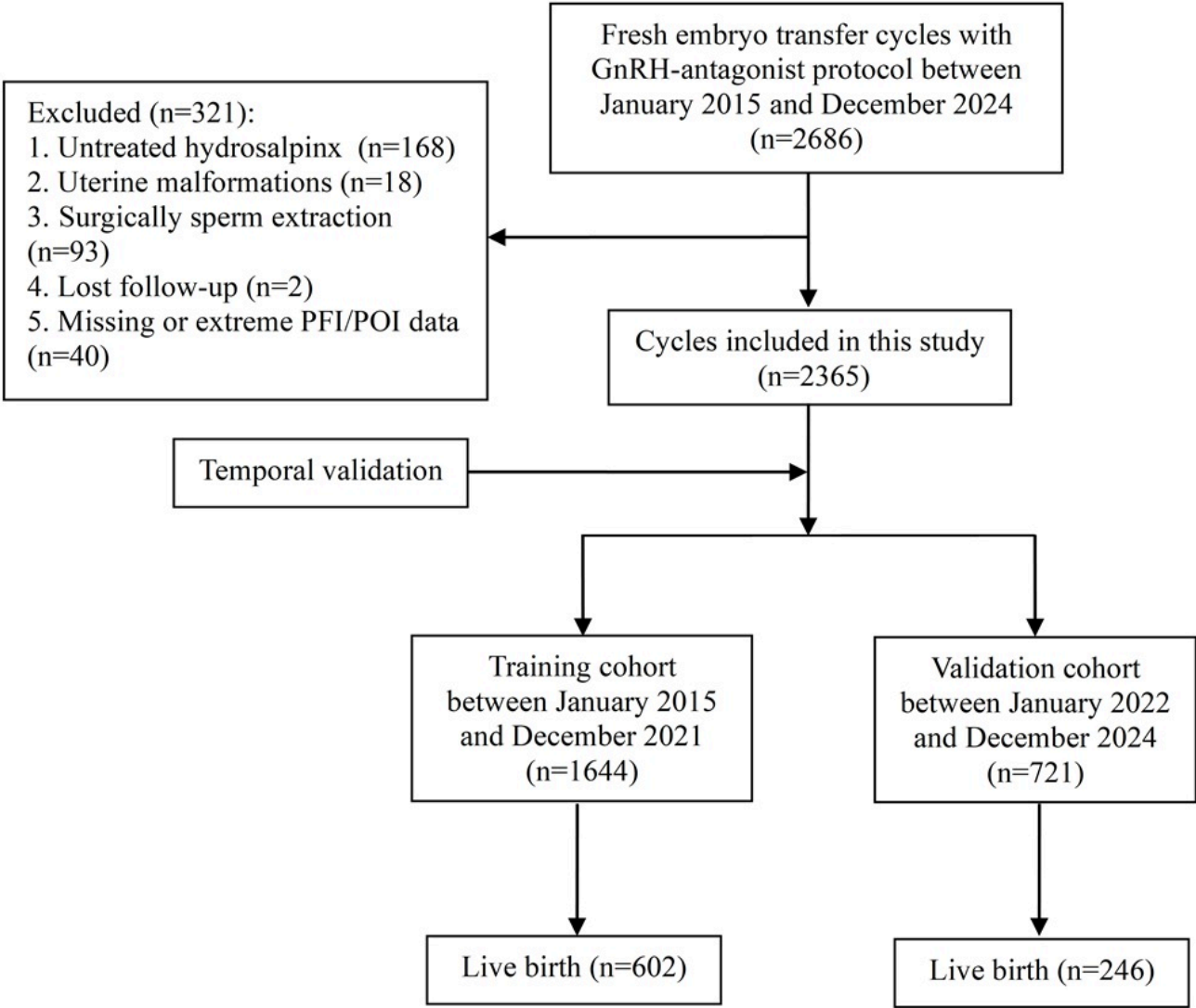


Figure 1

Flowchart of participant selection. From an initial pool of 2,686 screened GnRH antagonist protocol fresh IVF/ICSI cycles, 2,365 cycles were included in the analysis after applying inclusion and exclusion criteria and quality-control exclusions. The cohort was chronologically partitioned into a development set (n = 1,644, January 2015–December 2021) and a temporal validation set (n = 721, January 2022–December 2024). GnRH, gonadotropin-releasing hormone; ICSI, intracytoplasmic sperm injection; IVF, in vitro fertilization.

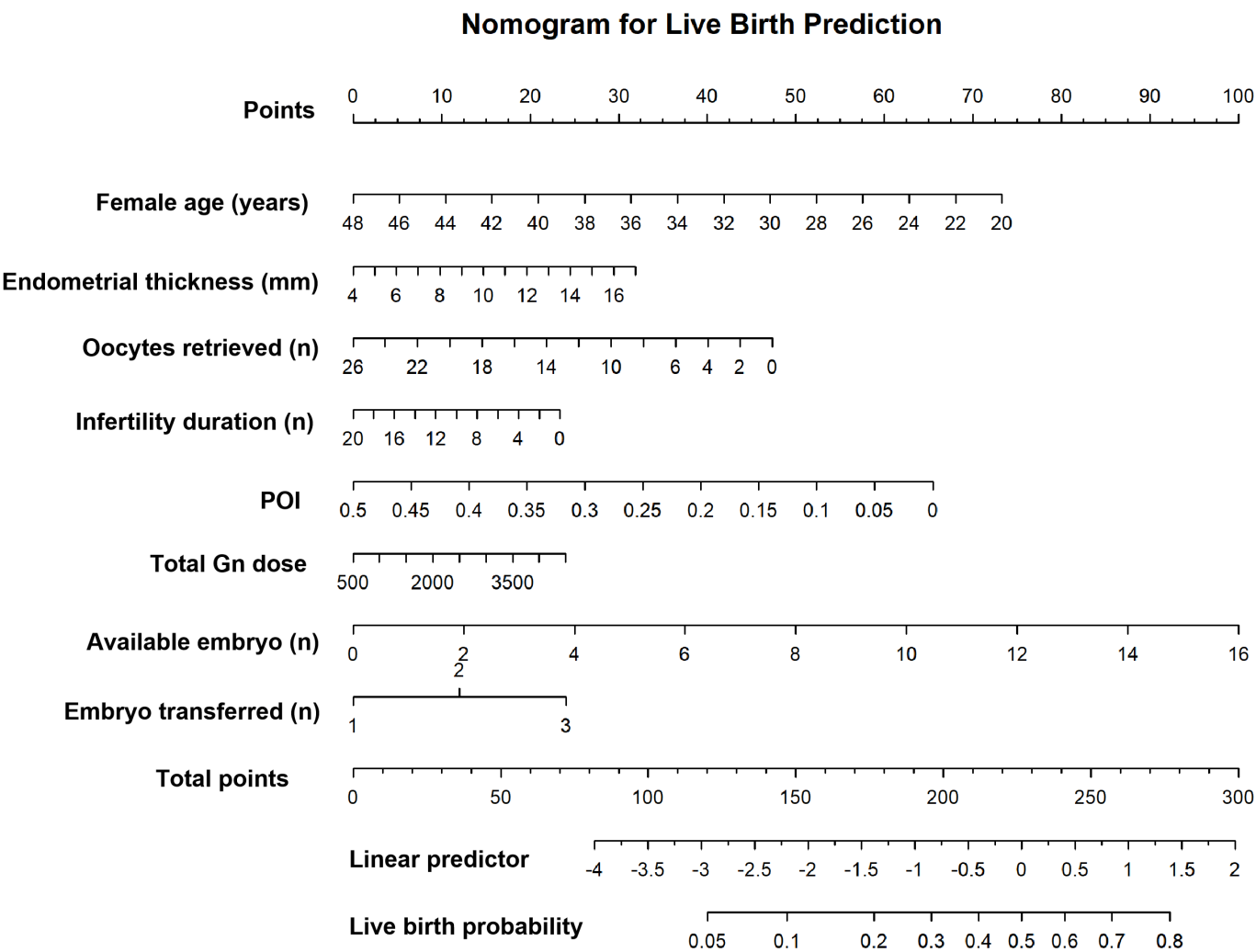


Figure 2

Nomogram for predicting the probability of live birth in GnRH antagonist protocol fresh IVF/ICSI cycles. The nomogram integrates eight predictors selected through a three-stage framework: female age, endometrial thickness on trigger day, oocytes retrieved, infertility duration, progesterone-to-oocyte index (POI), total gonadotropin dose, available embryos, and embryos transferred. To use the nomogram, locate each predictor value on its corresponding axis, draw a vertical line upward to the Points scale to obtain the individual score, sum the scores across all eight predictors, locate the total on the Total Points

axis, and draw a vertical line downward to the Probability of live birth axis to obtain the estimated probability. Gn, gonadotropin; IU, international unit; POI, progesterone-to-oocyte index.

ROC curve - Tripple validation
Bootstrap-corrected AUC = 0.666

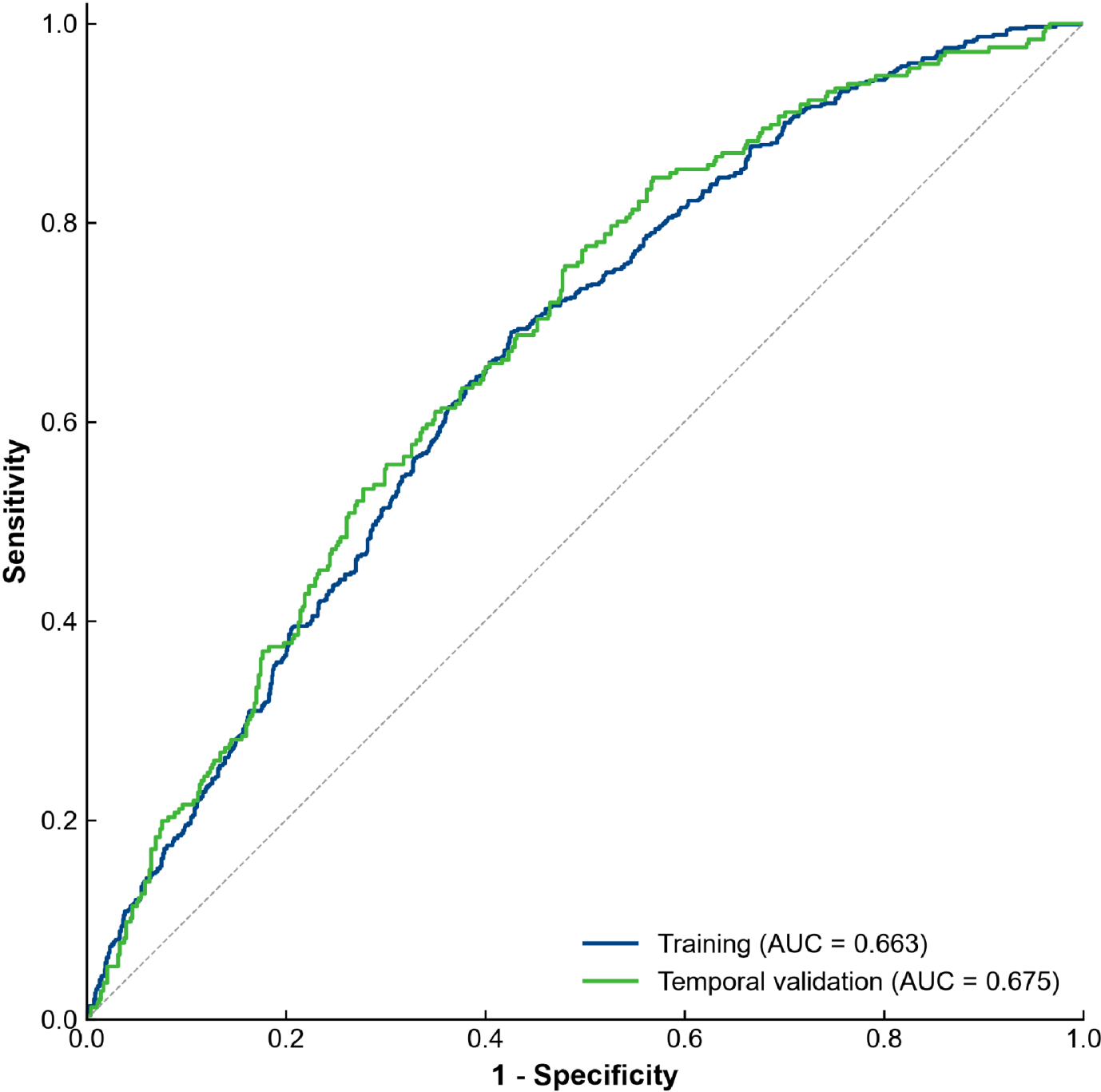


Figure 3

Receiver operating characteristic (ROC) curves for the live-birth prediction model across three validation methods. The apparent area under the curve (AUC) in the training cohort (2015–2021, n = 1,644) was

0.663 (95% CI 0.636–0.689), the optimism-corrected AUC after 1,000 bootstrap resamples was 0.666, and the temporal validation AUC (2022–2024, n = 721) was 0.675 (95% CI 0.634–0.715), indicating consistent discrimination across time periods. AUC, area under the curve; CI, confidence interval; ROC, receiver operating characteristic.

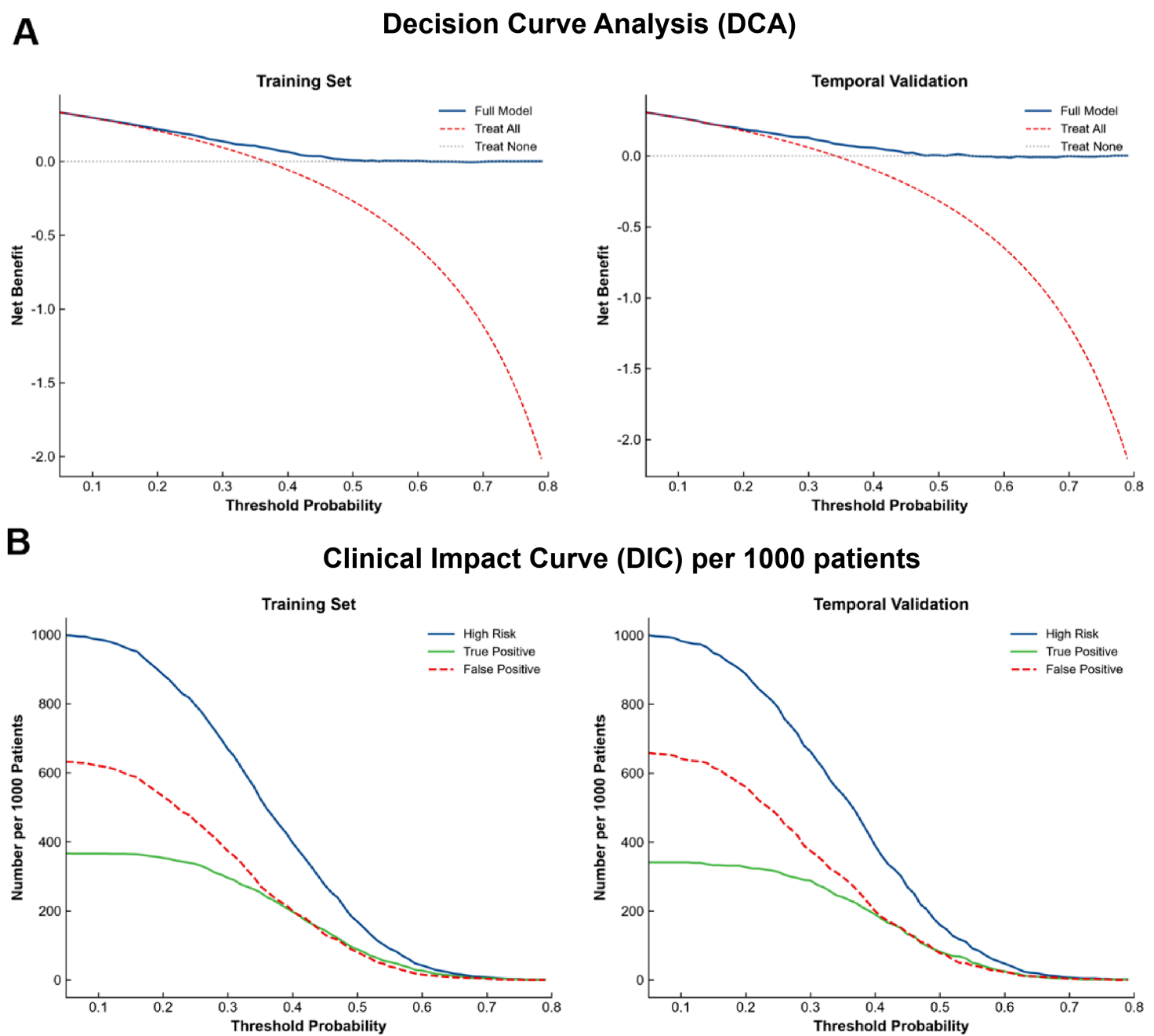


Figure 4

Clinical utility of the prediction model. (A) Decision curve analysis showing the net benefit of using the model for risk stratification across a range of threshold probabilities, compared with the treat-all and treat-none strategies, in both the training (left) and validation (right) cohorts. (B) Clinical impact curves displaying the number of patients classified as high-risk per 1,000 screened (red curve) and the number

of true positives among them (blue curve) across threshold probabilities. DCA, decision curve analysis; CIC, clinical impact curve.

Subgroup Analysis - AUC by Clinical Subgroups

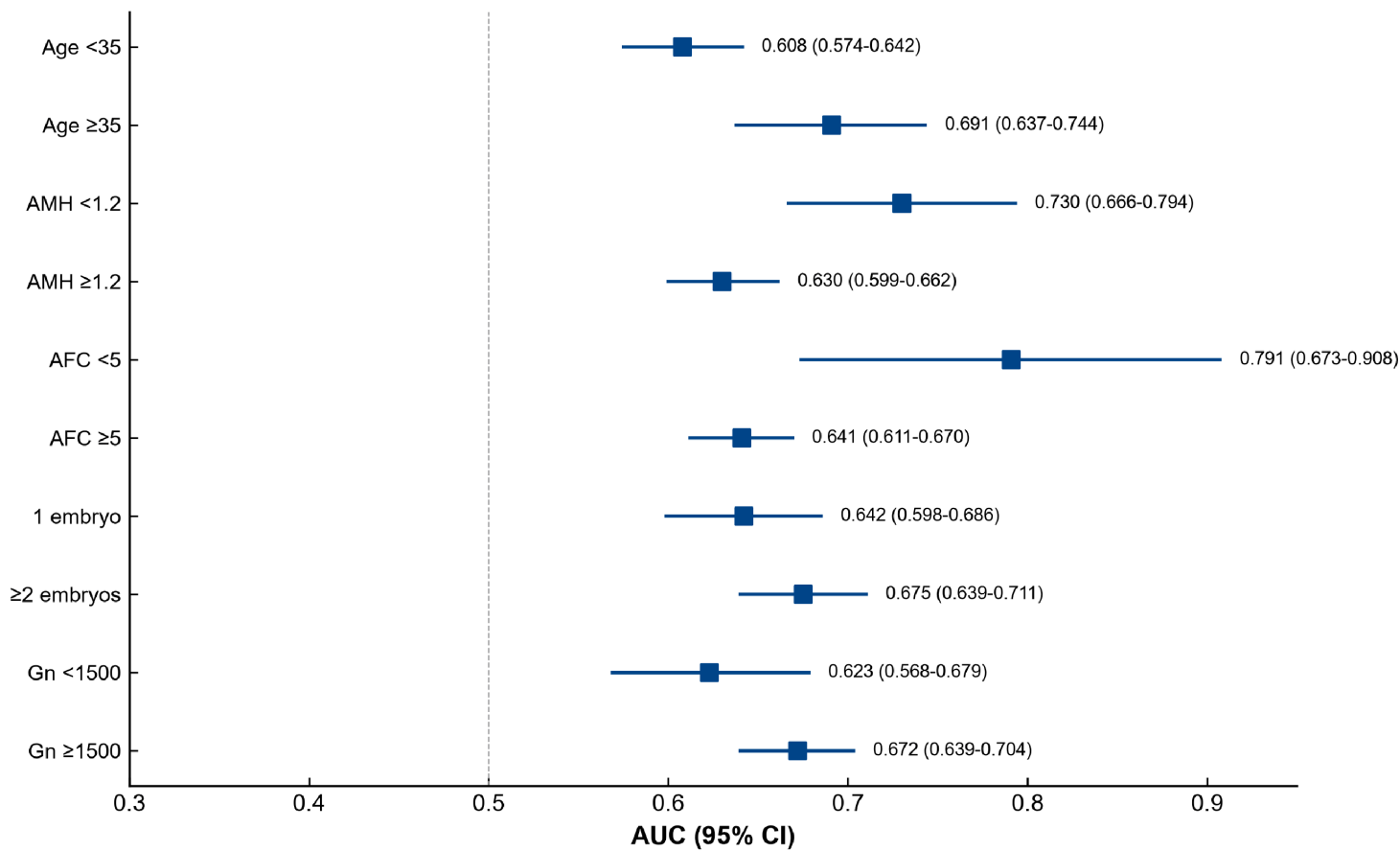


Figure 5

Subgroup analysis of model discrimination. Forest plot showing the area under the curve (AUC) with 95% confidence intervals for each subgroup, stratified by age (< 35 vs. ≥ 35 years), anti-Müllerian hormone (AMH; < 1.2 vs. ≥ 1.2 ng/mL), antral follicle count (AFC; < 5 vs. ≥ 5), gonadotropin dose (< 2,000 vs. ≥ 2,000 IU), and embryos transferred (1 vs. ≥ 2). The AUC exceeded 0.60 in all subgroups, indicating stable predictive ability across diverse clinical scenarios. AFC, antral follicle count; AMH, anti-Müllerian hormone; AUC, area under the curve; CI, confidence interval; Gn, gonadotropin; IU, international unit.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable3threechoice.docx](#)
- [SupplementaryTable4variableselection.docx](#)

- [SupplementaryTable5NRIDI.docx](#)
- [SupplementaryTable2VIF.docx](#)
- [SupplementaryTable1univariateanalysis.docx](#)
- [SupplementaryFigure1LASSO.tif](#)
- [SupplementaryFigure2calibrationGiViTI.tif](#)