Am J Cancer Res 2026;16(10):4621-4636
www.ajcr.us /ISSN:2156-6976/ajcr0175041

Original Article

Ultrasound morphological predictors of spontaneous regression of benign adnexal masses in pregnancy and maternal and perinatal outcomes of expectant versus surgical management: a retrospective cohort study

Zhongmei Shi[1], Yuan Qiao[2], Xianglin Lv[1], Lu Zhao[1], Juandi Liu[3], Jie Dang[4], Shan Li[5]

[1]Department of Obstetrics, Northwest Women's and Children's Hospital, No. 1616 Yanxiang Road, Yanta District, Xi'an 710000, Shaanxi, China; [2]Obstetric and Gynecologic Intensive Care Unit, Northwest Women's and Children's Hospital, No. 1616 Yanxiang Road, Yanta District, Xi'an 710000, Shaanxi, China; [3]Department of Medical Genetics, Northwest Women's and Children's Hospital, No. 1616 Yanxiang Road, Yanta District, Xi'an 710000, Shaanxi, China; [4]Department of Gynaecology, Shaanxi Provincial Nuclear Industry 215 Hospital, No. 35 Weiyang West Road, Qindu District, Xianyang 712000, Shaanxi, China; [5]Gynecology Ward I, Xi'an People's Hospital (Xi'an Fourth Hospital), No. 155 Hangtian East Road, Chang'an District, Xi'an 710000, Shaanxi, China

Received June 29, 2026; Accepted August 23, 2026; Epub October 15, 2026; Published October 30, 2026

Abstract: This single-center retrospective cohort study aimed to identify ultrasound morphological predictors of spontaneous regression of benign adnexal masses during pregnancy and compared maternal and perinatal outcomes after expectant versus surgical management. We included 549 patients with singleton pregnancies and benign adnexal masses detected by obstetric ultrasound between January 2016 and December 2025. Spontaneous regression occurred in 367 patients (66.8%), while 182 (33.2%) had persistent masses. Multivariable logistic regression identified higher cyst-diameter category (OR = 4.44), irregular or thickened cyst wall (OR = 3.12), higher septation category (OR = 3.17), higher papillary-projection category (OR = 2.01), elevated CA125 (OR = 2.37), and non-functional pathological type (OR = 2.44) as independent predictors of persistence; higher gravidity (OR = 0.47) and nulliparous status (OR = 0.24) were inversely associated with persistence. These eight variables were incorporated into the Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy (URS-AMP; reported range, 0-28 points). Persistence rates were 3.1%, 26.7%, and 89.9% in the low-risk (0-9), intermediate-risk (10-15), and high-risk (16-28) groups, respectively. The integer score yielded an AUC of 0.933 (95% CI 0.911-0.956), and the bootstrap optimism-corrected AUC was 0.933. Among patients with persistent masses, 94 received expectant management and 88 underwent surgery. After propensity-score matching (76 pairs), hypertensive disorders of pregnancy occurred in 0.0% of the expectant group and 9.2% of the surgical group (P = 0.014); other maternal and perinatal outcomes did not differ significantly. URS-AMP may support individualized assessment of persistence risk, and expectant management may be reasonable in appropriately selected patients; external validation is required before routine clinical use.

Keywords: Adnexal mass in pregnancy, spontaneous regression, ultrasound morphology, risk score, expectant management

Introduction

Adnexal masses are a common incidental finding on obstetric ultrasound. Their reported detection rate ranges from 0.19% to 8.8%, and the most detected masses are benign, with a malignancy rate of 1%-6% [1]. In a prospective single-center cohort of 274 patients, Gaughran et al. [2] reported a detection rate of approximately 1%, with simple cysts accounting for 75.9% of lesions and an overall spontaneous regression rate of approximately 74%. A recent review reported that approximately 68%-72% of benign adnexal masses resolve during pregnancy or the postpartum period and that simple cysts <5 cm regress in 71%-89% of cases [3]. Common benign diagnoses include simple cyst, corpus luteum cyst, mature teratoma, endome-

https://doi.org/10.62347/CASF1322

trioma, serous cystadenoma, and mucinous cystadenoma. In a retrospective series of 63 persistent masses, 90% were benign and mature teratoma was the most common diagnosis (32%); masses with benign sonographic features were generally managed conservatively until delivery [4]. Dur et al. [5] similarly reported that only 3.5% of 194 patients required emergency surgery. As prenatal ultrasound detects more adnexal masses, clinicians need reliable methods to estimate the likelihood of spontaneous regression and avoid unnecessary intervention.

Ultrasound is the first-line imaging modality for adnexal masses in pregnancy and provides information on size, wall morphology, septation, papillary projections, internal echogenicity, and vascularity. In a multicenter retrospective study, Lee et al. [6] reported AUROCs of 0.831 for the Sassone score and 0.883 for a combined model that incorporated six ultrasound features to predict malignancy. The International Ovarian Tumor Analysis (IOTA) Simple Rules [7] and the Ovarian-Adnexal Reporting and Data System (O-RADS) [8] are also used to assess malignancy risk. The IOTA Simple Rules have reported sensitivity and specificity of 93% and 90%, respectively, for distinguishing benign from malignant masses [9]. However, these tools were developed primarily to estimate malignancy risk, not to predict whether a benign mass will regress spontaneously during pregnancy. Evidence on predictors of spontaneous regression remains limited, and no internally validated quantitative tool is available for this specific clinical question.

Management options for persistent benign adnexal masses include observation and surgery. Expectant management avoids operative and anesthetic risks but may be complicated by torsion, rupture, or interval concern for malignancy. Surgery provides a definitive pathological diagnosis and may relieve symptoms or mass effect, but it can also expose the patient and fetus to perioperative risk. In a systematic review of 11 comparative studies involving 803 patients, Cagino et al. [10] found no significant difference between laparoscopy and laparotomy in miscarriage or preterm-delivery rates; elective surgery was associated with a lower risk of preterm delivery than emergency surgery. In a meta-analysis of 67,111 pregnant patients who underwent non-obstetric abdominopelvic surgery, fetal loss was more frequent after emergency than after elective procedures [11]. Non-obstetric surgery during pregnancy has also been associated with an approximately twofold increase in preterm birth [12]. Diagnosing torsion is challenging: in a multicenter IOTA study of 315 torsion cases, the whirlpool sign had a sensitivity of 90.8%, whereas normal Doppler flow did not exclude torsion because more than half of cases retained detectable flow [13]. These considerations make patient selection and surgical timing central to clinical decision-making, yet direct comparisons of expectant and surgical management remain limited.

Against this background, we conducted a single-center retrospective cohort study of patients with singleton pregnancies and benign adnexal masses detected by obstetric ultrasound. We examined clinical and sonographic predictors of spontaneous regression, developed an internally validated quantitative score for persistence, and compared maternal and perinatal outcomes after expectant versus surgical management. URS-AMP is intended to complement, not replace, malignancy-focused systems: it estimates persistence risk, whereas IOTA Simple Rules and O-RADS assess malignant potential. This distinction may help reduce unnecessary surgery for masses likely to resolve while directing persistent or morphologically concerning lesions to closer surveillance and formal malignancy assessment.

Materials and methods

Study design

This retrospective cohort study enrolled consecutive patients with singleton pregnancies and adnexal masses detected by obstetric ultrasound at Northwest Women's and Children's Hospital, Shaanxi Provincial Nuclear Industry 215 Hospital and Xi'an People's Hospital between January 2016 and December 2025. The study was approved by the Ethics Committee of Xi'an People's Hospital and was conducted in accordance with the Declaration of Helsinki [14]. Reporting followed the STROBE statement [15], and prediction-model development and reporting followed TRIPOD [16].

Sample size calculation

The minimum sample size was estimated using the three-criterion framework for clinical pre-

diction-model development proposed by Riley et al. [17] and implemented in the pmsampsize R package (version 1.1.0). We also applied a two-independent-proportions formula to the management-strategy comparison. For prediction-model development, we prespecified P = 10 as the anticipated effective number of predictor parameters in the final model. Other assumptions were an outcome proportion (φ) of 0.30, an expected adjusted Cox-Snell R^2 of 0.15, a target shrinkage factor (S) of 0.90, a two-sided α of 0.05, and a permissible error (δ) of 0.05 for the intercept estimate. For the management comparison, the assumed preterm-birth rates were $p_1 = 0.07$ for expectant management and $p_2 = 0.02$ for surgery, with a two-sided α of 0.05 and 80% power. Criterion 1 (minimizing overfitting) yielded $n_1 = p/[(S - 1) \times \ln(1 - R^2CS, adj/S)] = 549$, corresponding to 165 events and 16.5 events per prespecified parameter. Criterion 2 (limiting optimism in Nagelkerke R^2) yielded $n_2 \approx 420$ in *pmsampsize*. Criterion 3 (precision of the intercept) yielded $n_3 = Z^2_{1-\alpha/2} \times \varphi \times (1 - \varphi)/\delta^2 = 323$. The largest requirement was therefore 549. The conventional EPV calculation (EPV = 15) yielded 500 participants. For the management comparison, the two-proportion formula yielded 266 participants per group (532 total). The final cohort of 549 met the largest calculated requirement.

Inclusion and exclusion criteria

Patients were eligible if (1) an adnexal mass was first detected by ultrasound during pregnancy; (2) the ultrasound record contained complete information on cyst size, wall characteristics, septation, papillary projections, internal echogenicity, and blood-flow signals; (3) the mass was considered benign on ultrasound or was confirmed as benign on postoperative histopathology (simple cyst, corpus luteum cyst, mature teratoma, endometrioma, serous cystadenoma, or mucinous cystadenoma); (4) pregnancy management and delivery occurred at our hospital and postpartum follow-up data were available; and (5) the pregnancy was singleton.

Patients were excluded if (1) ultrasound was highly suspicious for a malignant or borderline tumor, or postoperative pathology confirmed a malignant or borderline tumor; (2) two or more key ultrasound features were missing; (3) transfer or loss to follow-up resulted in incomplete outcome data; (4) a severe pre-existing maternal condition was present (e.g., pregestational diabetes, chronic hypertension, or autoimmune disease); (5) the pregnancy was multiple; or (6) conception occurred through assisted reproductive technology.

Data collection

We retrospectively extracted the following variables from the electronic medical record: (1) clinical characteristics - maternal age, gravidity, pre-pregnancy body mass index (BMI), gestational age at detection, and nulliparous status; (2) laboratory data - cancer antigen 125 (CA125), categorized as elevated when >35 U/mL; (3) ultrasound morphology - maximum cyst diameter (<5, 5-9.9, or ≥10 cm), cyst wall (smooth or thickened/irregular), septation (none, thin and regular, or thick and irregular), papillary projections (none, smooth-surfaced, or irregular-surfaced), internal echogenicity (anechoic, ground-glass, or mixed), color Doppler flow score (1-4) and blood-flow abundance (categorized as not abundant or abundant); (4) cyst-related characteristics - bilaterality, ascites, and pathological type; (5) management - expectant management or surgery (laparoscopy/laparotomy; emergency/elective); (6) maternal outcomes - miscarriage, preterm birth (<37 weeks), premature rupture of membranes (PROM), hypertensive disorders of pregnancy (HDP), and postpartum hemorrhage (PPH); and (7) perinatal outcomes - low birth weight (LBW; <2500 g), small for gestational age (SGA), neonatal intensive care unit (NICU) admission, 5-minute Apgar score <7, and perinatal death. For score construction, lesion types were operationally grouped according to their expected physiologic origin and tendency to resolve during pregnancy [1-3]. Simple cysts and corpus luteum cysts were classified as functional, whereas endometriomas, mature teratomas, serous cystadenomas, and mucinous cystadenomas were classified as non-functional.

Measurement methods

All ultrasound examinations were performed with a GE Voluson E10 color Doppler system (GE Healthcare, Chicago, IL, USA) using a C1-6-D transabdominal convex probe (1-6 MHz) and an RIC5-9-D transvaginal probe (5-9 MHz). Maximum cyst diameter was measured in the

largest plane. A smooth wall was defined as having uniform thickness and clearly defined inner and outer margins; a thickened or irregular wall had focal or diffuse thickness >3 mm or irregular margins. Septa <3 mm with a regular course were classified as thin and regular, whereas septa ≥3 mm or with an irregular course were classified as thick and irregular. Papillary projections were solid structures protruding ≥3 mm from the cyst wall into the cavity and were categorized by surface contour as smooth or irregular. Terminology followed the IOTA consensus [18]. Color Doppler vascularity was graded with the IOTA semiquantitative score [18]: 1, no flow; 2, minimal flow; 3, moderate flow; and 4, abundant flow. Scores ≥3 were categorized as abundant blood flow. Two senior sonographers independently reviewed all images and reports, and disagreements were resolved by consensus. CA125 was measured by electrochemiluminescence immunoassay on a Roche Cobas e601 analyzer (Roche Diagnostics GmbH, Mannheim, Germany) using the manufacturer's reagents and procedures; values >35 U/mL were categorized as elevated. Pathological type was established by postoperative histopathology in surgical patients. In patients managed without surgery or whose masses regressed, the type was assigned from characteristic sonographic patterns according to IOTA terminology [18], supplemented by clinical and laboratory information (e.g., a simple anechoic cyst, ground-glass echogenicity for endometrioma, or a dermoid mesh pattern for mature teratoma). Postpartum imaging was used as additional corroboration when available.

Outcome measures

The primary outcome was spontaneous cyst regression, defined as complete disappearance of the adnexal mass on ultrasound at the last available follow-up examination. The ascertainment time point depended on the clinical course: (1) for patients managed expectantly throughout pregnancy, the later of the final antenatal ultrasound or the routine postpartum examination (approximately 42 days after delivery); (2) for patients who underwent surgery, the ultrasound immediately before surgery; and (3) for patients who experienced miscarriage or perinatal death, the last ultrasound before the event. Except for patients with miscarriage or perinatal death, follow-up extended through delivery and the postpartum period. A mass that was no longer visible at the final examination was classified as regressed; a mass that remained visible was classified as persistent. The outcome was analyzed as a binary variable.

Ultrasound follow-up in this retrospective cohort was scheduled according to clinical indications rather than a standardized research protocol, and examination intervals varied substantially. The exact time of regression was therefore known only to fall between two consecutive examinations, with markedly different interval widths among patients. Under these conditions, a median time to regression would be strongly influenced by follow-up frequency. We therefore did not perform time-to-event analysis or report a median time to regression.

Secondary outcomes comprised maternal outcomes (miscarriage, preterm birth before 37 weeks, PROM, HDP, and PPH) and perinatal outcomes (LBW <2500 g, SGA, NICU admission, 5-minute Apgar score <7, and perinatal death).

Statistical analysis

Statistical analyses were performed in R version 4.5.1 (R Foundation for Statistical Computing, Vienna, Austria) and SPSS version 27.0 (IBM Corp., Armonk, NY, USA). All tests were two-sided, and $P<0.05$ was considered statistically significant. Normally distributed continuous variables were summarized as mean ± standard deviation and compared with Student's t-test or Welch's t-test. Non-normally distributed continuous variables were summarized as median (interquartile range) and compared with the Wilcoxon rank-sum test. Categorical variables were summarized as number (percentage) and compared with Pearson's χ^2 test, the continuity-corrected χ^2 test, or Fisher's exact test, as appropriate. Prediction-model development proceeded as follows: (1) generalized variance inflation factors were used to evaluate multicollinearity, with VIF >5 indicating potential concern; (2) restricted cubic splines (RCS) were used to evaluate possible nonlinear associations between continuous variables and persistence; (3) univariable and multivariable logistic regression models estimated odds ratios (ORs) and

95% confidence intervals (CIs), with cyst persistence as the event; (4) least absolute shrinkage and selection operator (LASSO) regression, using 10-fold cross-validation to select λ, served as a supplementary variable-selection analysis; and (5) an integer-based URS-AMP score was derived from the multivariable β coefficients. Internal validation used 1,000 bootstrap resamples to estimate optimism and the optimism-corrected AUC. Discrimination was assessed with receiver operating characteristic (ROC) curves; the Youden index defined the optimal cutoff, and sensitivity, specificity, positive predictive value, and negative predictive value were calculated. Calibration was examined with calibration curves, and clinical utility was explored with decision-curve analysis (DCA). Pairwise comparisons of correlated AUCs were performed using paired DeLong tests. Comparisons involving the prespecified reference risk stratification were considered exploratory because all models and classifications were evaluated in the same development cohort. For management comparisons, Fisher's exact test was used to estimate ORs and 95% CIs for maternal and perinatal outcomes, with surgery as the reference group. Propensity-score matching (PSM) was performed as a sensitivity analysis using age, pre-pregnancy BMI, gestational age at detection, nulliparous status, elevated CA125, cyst diameter, wall characteristics, septation, papillary projections, internal echogenicity, blood-flow score, and blood-flow abundance. We used 1:1 nearest-neighbor matching with a caliper of 0.2. Standardized mean differences (SMDs) assessed covariate balance, with SMD <0.15 considered acceptable. Surgical subgroup analyses compared laparoscopy with laparotomy and emergency with elective surgery.

Results

Study population and baseline characteristics

After applying the eligibility criteria, 549 patients with benign adnexal masses were included (**Figure 1**). Spontaneous regression occurred in 367 patients (66.8%), while 182 (33.2%) had persistent masses. Among patients with persistent masses, 94 were managed expectantly and 88 underwent surgery (70 laparoscopies and 18 laparotomies; 11 emergency and 77 elective procedures). The cohort had a mean age of 30.32 ± 4.80 years, a mean pre-pregnancy BMI of 23.08 ± 3.34 kg/m^2, and a mean gestational age at detection of 12.48 ± 5.00 weeks (**Table 1**). Overall, 240 patients (43.7%) were nulliparous, and 247 (45.0%) had elevated CA125. Cyst diameter was <5 cm in 332 patients (60.5%), the wall was smooth in 421 (76.7%), and no septation was present in 349 (63.6%). Simple cyst (41.0%) and corpus luteum cyst (24.0%) were the most frequent diagnoses.

Table 2 compares the regression and persistence groups. Age (P = 0.047), pre-pregnancy BMI (P = 0.048), and elevated CA125 (P = 0.001) differed significantly between groups, whereas gravidity, gestational age at detection, and nulliparous status did not (all P>0.05). All major sonographic variables - cyst diameter, wall characteristics, septation, papillary projections, internal echogenicity, color Doppler flow score, and blood-flow abundance - differed significantly (all P<0.001). Bilaterality also differed (P = 0.003), whereas ascites did not (P = 0.079).

Within the persistence group, no statistically significant baseline differences were detected between the expectant and surgical groups in age, pre-pregnancy BMI, gestational age at detection, nulliparous status, elevated CA125, or the major sonographic variables (all P>0.05; **Table 3**). This absence of statistically significant differences does not establish equivalence and does not exclude confounding by indication. In the surgical subgroup, the laparotomy group had higher proportions of thick/irregular septation and mixed internal echogenicity than the laparoscopy group (both P<0.05; Table S1); other measured variables did not differ significantly.

Spontaneous cyst regression rate and ultrasound morphological features

RCS analyses found no significant nonlinear associations of age, gravidity, or pre-pregnancy BMI with regression. Gestational age at detection showed evidence of nonlinearity (P for nonlinearity = 0.034; Figure S1). Among sonographic variables, pathological type (χ^2 = 190.928) and cyst diameter (χ^2 = 151.769) had the strongest unadjusted associations with regression status, followed by septation, papillary projections, wall characteristics, internal

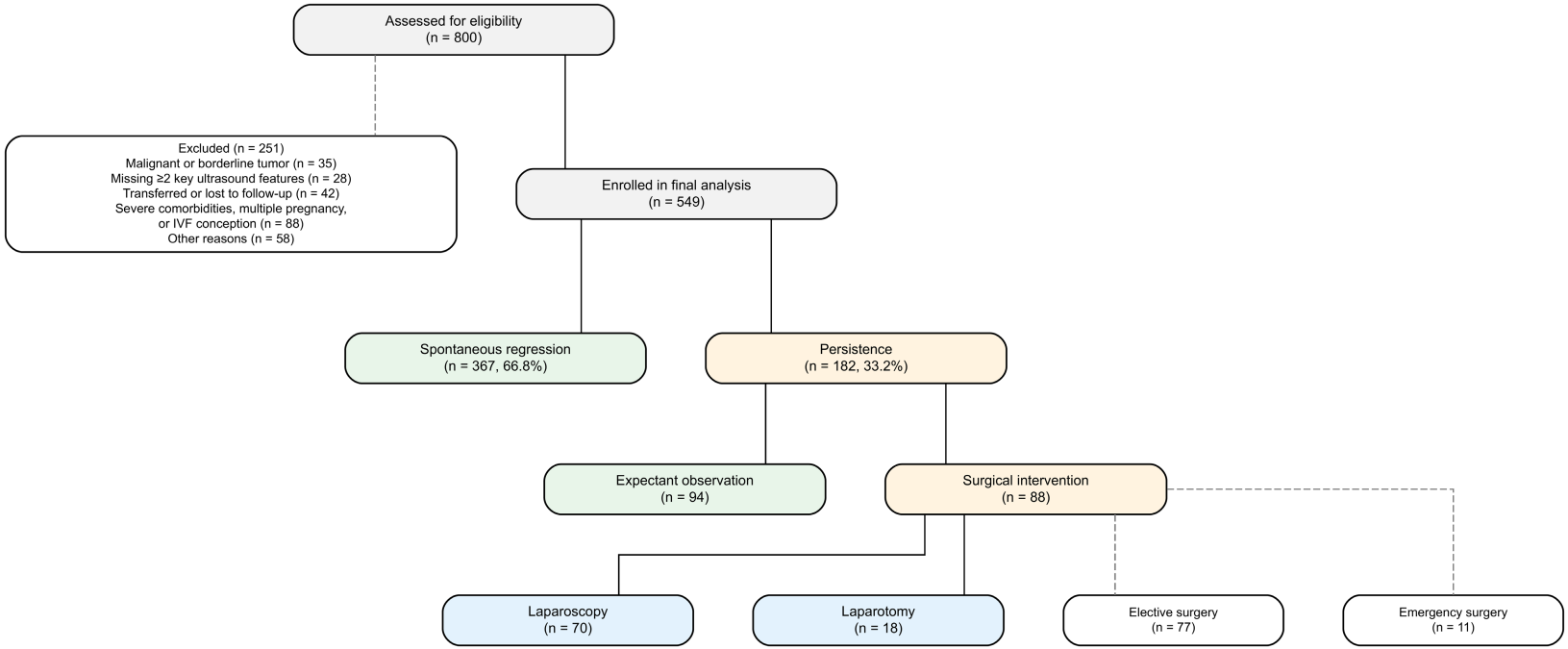


Figure 1. Study flow diagram.

Table 1. Baseline characteristics of the entire cohort (n = 549)

Variable	Category	Entire cohort (n = 549)
Age (years)		30.32 ± 4.80
Gravidity		1.74 ± 0.78
Pre-pregnancy BMI (kg/m^2)		23.08 ± 3.34
GA at detection (weeks)		12.48 ± 5.00
Nulliparous	No	309 (56.3%)
	Yes	240 (43.7%)
Elevated CA125	No	302 (55.0%)
	Yes	247 (45.0%)
Cyst diameter	<5 cm	332 (60.5%)
	5-9.9 cm	161 (29.3%)
	≥10 cm	56 (10.2%)
Cyst wall characteristics	Smooth	421 (76.7%)
	Thickened/irregular	128 (23.3%)
Septation	None	349 (63.6%)
	Thin regular	119 (21.7%)
	Thick irregular	81 (14.8%)
Papillary projections	None	421 (76.7%)
	Smooth surface	83 (15.1%)
	Irregular surface	45 (8.2%)
Internal echogenicity	Anechoic	367 (66.8%)
	Ground-glass	118 (21.5%)
	Mixed echogenicity	64 (11.7%)
Color Doppler Blood flow score	1 point	220 (40.1%)
	2 points	156 (28.4%)
	3 points	119 (21.7%)
	4 points	54 (9.8%)
Rich blood flow	No	376 (68.5%)
	Yes	173 (31.5%)
Bilaterality	Unilateral	476 (86.7%)
	Bilateral	73 (13.3%)
Ascites	None	544 (99.1%)
	Present	5 (0.9%)
Pathological type	Simple cyst	225 (41.0%)
	Corpus luteum cyst	132 (24.0%)
	Mature teratoma	71 (12.9%)
	Endometrioma	60 (10.9%)
	Serous cystadenoma	36 (6.6%)
	Mucinous cystadenoma	25 (4.6%)
Reference URS-AMP risk category	Low risk	302 (55.0%)
	Intermediate risk	137 (25.0%)
	High risk	110 (20.0%)

Note: Continuous variables are presented as mean ± standard deviation; categorical variables are presented as number (percentage). BMI, body mass index; CA125, cancer antigen 125; URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy.

echogenicity, color Doppler flow score, and blood-flow abundance (all P<0.001; Figure S2).

Regression rates by pathological type are shown in Table S2. Functional cysts had higher

Table 2. Comparison of baseline characteristics between the regression and persistence groups

Variable	Category	Persistence group (n = 182)	Regression group (n = 367)	Statistic	*P* value
Age (years)		30.90 ± 5.07	30.03 ± 4.64	t = -2.000	0.047
Gravidity		1.65 ± 0.68	1.78 ± 0.83	t = -1.846	0.065
Pre-pregnancy BMI (kg/m^2)		23.48 ± 3.50	22.88 ± 3.24	t = 1.985	0.048
GA at detection (weeks)		12.95 ± 5.50	12.24 ± 4.72	t = 1.559	0.121
Nulliparous	No	98 (53.8%)	211 (57.5%)	χ^2 = 0.519	0.472
	Yes	84 (46.2%)	156 (42.5%)		
Elevated CA125	No	82 (45.1%)	220 (59.9%)	χ^2 = 10.900	0.001
	Yes	100 (54.9%)	147 (40.1%)		
Cyst diameter	<5 cm	46 (25.3%)	286 (77.9%)	χ^2 = 151.769	<0.001
	5-9.9 cm	91 (50.0%)	70 (19.1%)		
	≥10 cm	45 (24.7%)	11 (3.0%)		
Cyst wall characteristics	Smooth	91 (50.0%)	330 (89.9%)	χ^2 = 106.212	<0.001
	Thickened/irregular	91 (50.0%)	37 (10.1%)		
Septation	None	55 (30.2%)	294 (80.1%)	χ^2 = 143.280	<0.001
	Thin regular	64 (35.2%)	55 (15.0%)		
	Thick irregular	63 (34.6%)	18 (4.9%)		
Papillary projections	None	91 (50.0%)	330 (89.9%)	χ^2 = 110.917	<0.001
	Smooth surface	55 (30.2%)	28 (7.6%)		
	Irregular surface	36 (19.8%)	9 (2.5%)		
Internal echogenicity	Anechoic	73 (40.1%)	294 (80.1%)	χ^2 = 94.234	<0.001
	Ground-glass	63 (34.6%)	55 (15.0%)		
	Mixed echogenicity	46 (25.3%)	18 (4.9%)		
Color Doppler flow score	1 point	36 (19.8%)	184 (50.1%)	χ^2 = 79.148	<0.001
	2 points	46 (25.3%)	110 (30.0%)		
	3 points	64 (35.2%)	55 (15.0%)		
	4 points	36 (19.8%)	18 (4.9%)		
Rich blood flow	No	82 (45.1%)	294 (80.1%)	χ^2 = 67.656	<0.001
	Yes	100 (54.9%)	73 (19.9%)		
Bilaterality	Unilateral	146 (80.2%)	330 (89.9%)	χ^2 = 8.946	0.003
	Bilateral	36 (19.8%)	37 (10.1%)		
Ascites	None	178 (97.8%)	366 (99.7%)	χ^2 = 3.092	0.079
	Present	4 (2.2%)	1 (0.3%)		
Pathological type	Simple cyst	34 (18.7%)	191 (52.0%)	χ^2 = 190.928	<0.001
	Corpus luteum cyst	13 (7.1%)	119 (32.4%)		
	Mature teratoma	50 (27.5%)	21 (5.7%)		
	Endometrioma	36 (19.8%)	24 (6.5%)		
	Serous cystadenoma	28 (15.4%)	8 (2.2%)		
	Mucinous cystadenoma	21 (11.5%)	4 (1.1%)		
Reference URS-AMP risk category	Low risk	27 (14.8%)	275 (74.9%)	χ^2 = 213.249	<0.001
	Intermediate risk	64 (35.2%)	73 (19.9%)		
	High risk	91 (50.0%)	19 (5.2%)		

Note: Continuous variables are presented as mean ± standard deviation and were compared with Student's t-test or Welch's t-test. Categorical variables are presented as number (percentage) and were compared with Pearson's χ^2 test, the continuity-corrected χ^2 test, or Fisher's exact test, as appropriate. BMI, body mass index; CA125, cancer antigen 125; URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy. All *P* values are two-sided.

Table 3. Comparison of baseline characteristics between the expectant-management and surgical groups

Variable	Category	Expectant-management group (n = 94)	Surgical group (n = 88)	Statistic	*P* value
Age (years)		30.53 ± 4.69	31.28 ± 5.44	t = -1.002	0.318
Gravidity		1.67 ± 0.69	1.62 ± 0.67	t = 0.447	0.655
Pre-pregnancy BMI (kg/m^2)		23.47 ± 3.31	23.49 ± 3.71	t = -0.039	0.969
GA at detection (weeks)		13.13 ± 5.65	12.75 ± 5.35	t = 0.462	0.644
Nulliparous	No	51 (54.3%)	47 (53.4%)	χ^2 = 0.000	1.000
	Yes	43 (45.7%)	41 (46.6%)		
Elevated CA125	No	44 (46.8%)	38 (43.2%)	χ^2 = 0.117	0.732
	Yes	50 (53.2%)	50 (56.8%)		
Cyst diameter	<5 cm	23 (24.5%)	23 (26.1%)	χ^2 = 4.079	0.130
	5-9.9 cm	53 (56.4%)	38 (43.2%)		
	≥10 cm	18 (19.1%)	27 (30.7%)		
Cyst wall characteristics	Smooth	47 (50.0%)	44 (50.0%)	χ^2 = 0.000	1.000
	Thickened/irregular	47 (50.0%)	44 (50.0%)		
Septation	None	27 (28.7%)	28 (31.8%)	χ^2 = 3.069	0.216
	Thin regular	29 (30.9%)	35 (39.8%)		
	Thick irregular	38 (40.4%)	25 (28.4%)		
Papillary projections	None	50 (53.2%)	41 (46.6%)	χ^2 = 4.366	0.113
	Smooth surface	31 (33.0%)	24 (27.3%)		
	Irregular surface	13 (13.8%)	23 (26.1%)		
Internal echogenicity	Anechoic	40 (42.6%)	33 (37.5%)	χ^2 = 1.655	0.437
	Ground-glass	34 (36.2%)	29 (33.0%)		
	Mixed echogenicity	20 (21.3%)	26 (29.5%)		
Color Doppler flow score	1 point	22 (23.4%)	14 (15.9%)	χ^2 = 3.932	0.269
	2 points	21 (22.3%)	25 (28.4%)		
	3 points	36 (38.3%)	28 (31.8%)		
	4 points	15 (16.0%)	21 (23.9%)		
Rich blood flow	No	43 (45.7%)	39 (44.3%)	χ^2 = 0.002	0.965
	Yes	51 (54.3%)	49 (55.7%)		
Bilaterality	Unilateral	75 (79.8%)	71 (80.7%)	χ^2 = 0.000	1.000
	Bilateral	19 (20.2%)	17 (19.3%)		
Ascites	None	92 (97.9%)	86 (97.7%)	χ^2 = 0.000	1.000
	Present	2 (2.1%)	2 (2.3%)		
Pathological type	Simple cyst	18 (19.1%)	16 (18.2%)	χ^2 = 4.175	0.524
	Corpus luteum cyst	9 (9.6%)	4 (4.5%)		
	Mature teratoma	21 (22.3%)	29 (33.0%)		
	Endometrioma	21 (22.3%)	15 (17.0%)		
	Serous cystadenoma	14 (14.9%)	14 (15.9%)		
	Mucinous cystadenoma	11 (11.7%)	10 (11.4%)		
Reference URS-AMP risk category	Low risk	14 (14.9%)	13 (14.8%)	χ^2 = 0.413	0.813
	Intermediate risk	35 (37.2%)	29 (33.0%)		
	High risk	45 (47.9%)	46 (52.3%)		

Note: Continuous variables are presented as mean ± standard deviation and were compared with Student's t-test or Welch's t-test. Categorical variables are presented as number (percentage) and were compared with Pearson's χ^2 test, the continuity-corrected χ^2 test, or Fisher's exact test, as appropriate. BMI, body mass index; CA125, cancer antigen 125; URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy. All *P* values are two-sided.

Table 4. URS-AMP scoring system: point-assignment rules

Predictor	Category/value	Points	Direction
Cyst diameter	<5/5-9.9/≥10 cm	+3	Increases persistence risk
Cyst wall characteristics	Thickened/irregular vs. smooth	+2	Increases persistence risk
Septation	None/thin regular/thick irregular	+2	Increases persistence risk
Papillary projections	None/smooth surface/irregular surface	+1	Increases persistence risk
Elevated CA125	Yes vs. no	+2	Increases persistence risk
Pathological type	Functional vs. non-functional	+2	Increases persistence risk
Gravidity	Continuous variable	-2	Decreases persistence risk
Nulliparous	Yes vs. no	-3	Decreases persistence risk

URS-AMP risk stratification

Risk level	Total score range	n	Persistence rate	Regression rate
Low risk	0-9	223	3.1% (7/223)	96.9% (216/223)
Intermediate risk	10-15	187	26.7% (50/187)	73.3% (137/187)
High risk	16-28	139	89.9% (125/139)	10.1% (14/139)

Note: URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy; CA125, cancer antigen 125. Points were derived from rounded multivariable regression coefficients. Higher scores indicate a greater risk of persistence. The reported score range is 0-28 points. Patients were categorized as low risk (0-9 points), intermediate risk (10-15 points), or high risk (16-28 points).

regression rates (corpus luteum cyst, 90.2%; simple cyst, 84.9%) than non-functional lesions (endometrioma, 40.0%; mature teratoma, 29.6%; serous cystadenoma, 22.2%; mucinous cystadenoma, 16.0%).

Predictor selection and logistic regression analysis

Fourteen candidate variables were evaluated: age, gravidity, pre-pregnancy BMI, gestational age at detection, nulliparous status, elevated CA125, cyst diameter, wall characteristics, septation, papillary projections, internal echogenicity, color Doppler flow score, blood-flow abundance, and pathological type. Bilaterality and ascites were not included in the multivariable model because of sparse or imbalanced category distributions. All evaluated variables had VIFs <5.0; the highest values were 4.898 for color Doppler flow score and 4.763 for blood-flow abundance, suggesting no severe multicollinearity.

Multivariable logistic regression identified eight independent predictors of persistence: higher cyst-diameter category (OR = 4.443, 95% CI 2.741-7.202; P<0.001), higher septation category (OR = 3.168, 95% CI 2.062-4.866; P<0.001), thickened/irregular wall (OR = 3.119, 95% CI 1.493-6.515; P = 0.002), non-functional pathological type (OR = 2.441, 95% CI 1.957-3.045; P<0.001), elevated CA125 (OR = 2.365, 95% CI 1.287-4.346; P = 0.006), higher papillary-projection category (OR = 2.009, 95% CI 1.211-3.334; P = 0.007), gravidity (OR = 0.466, 95% CI 0.228-0.954; P = 0.037), and nulliparous status (OR = 0.240, 95% CI 0.078-0.736; P = 0.013). Age, pre-pregnancy BMI, gestational age at detection, internal echogenicity, color Doppler flow score, and blood-flow abundance were not independently associated with persistence (all P>0.05). LASSO results were broadly consistent with the multivariable model (Figure S3), and univariable and multivariable estimates are shown in Figure S4.

Construction and validation of the URS-AMP scoring system

The URS-AMP score was derived from the multivariable β coefficients and included cyst diameter, wall characteristics, septation, papillary projections, elevated CA125, pathological type, gravidity, and nulliparous status. The reported point allocations and risk strata are shown in **Table 4**. Higher scores indicated a greater probability of persistence; the reported total range was 0-28 points, with low-risk (0-9), intermediate-risk (10-15), and high-risk (16-28) categories.

The full logistic regression model yielded an AUC of 0.943 (95% CI 0.924-0.962). The integer URS-AMP score yielded an AUC of 0.933 (95% CI 0.911-0.956), a reduction of 0.010

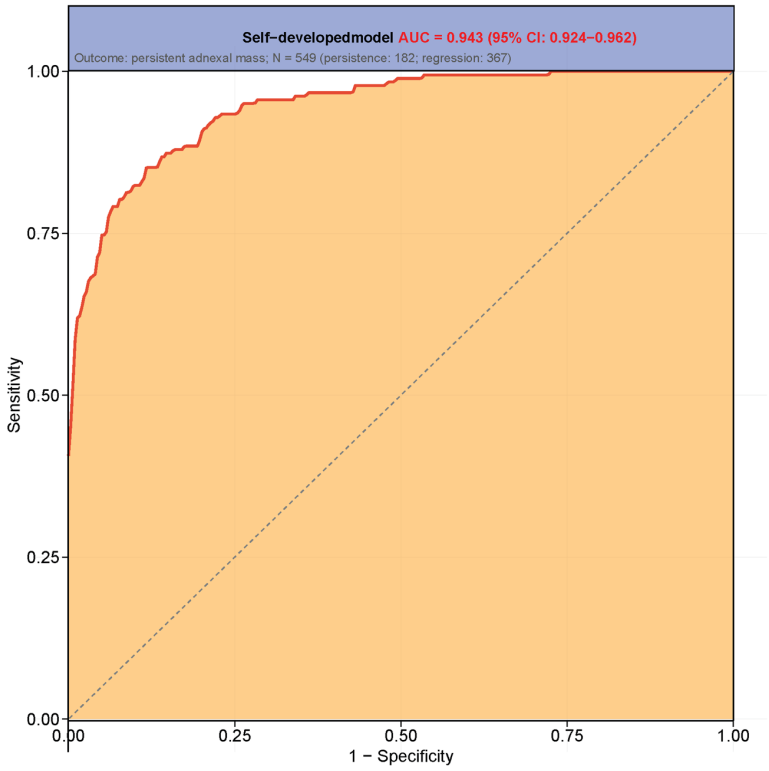


Figure 2. ROC curves for predicting cyst persistence. ROC curves for the full logistic regression model (AUC = 0.943), the URS-AMP integer score (AUC = 0.933), and the prespecified reference risk stratification (AUC = 0.841). The comparison with the reference stratification is descriptive and exploratory. Note: ROC, receiver operating characteristic; AUC, area under the curve; URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy.

(**Figure 2**). The Youden-optimal cutoff was 12.5 points, with 86.8% sensitivity and 84.7% specificity (Table S3). Persistence rates increased across the low-, intermediate-, and high-risk groups (3.1%, 26.7%, and 89.9%, respectively; **Figure 3**). Individual sonographic features had AUCs of 0.675-0.787, all lower than the composite score (Figure S5). The prespecified reference risk stratification had an AUC of 0.841 (95% CI, 0.807-0.875). Paired DeLong tests showed that the URS-AMP integer score had a significantly higher AUC than the prespecified reference risk stratification (AUC difference, 0.092; Z = 4.990; P<0.001; Figure S6). The full logistic model also had a slightly higher AUC than the integer score (AUC difference, 0.010; Z = 2.270; P = 0.023) and a higher AUC than the reference risk stratification (AUC difference, 0.102; Z = 5.907; P<0.001; **Figure 2**). These within-cohort comparisons were considered exploratory. Bootstrap validation with 1,000 valid resamples yielded a mean optimism of 0.010 and an optimism-corrected AUC of 0.933 (Table S3). Calibration and DCA results are shown in Figure S7.

Impact of expectant management versus surgical intervention on maternal and perinatal outcomes

Among the 182 patients with persistent masses, 94 received expectant management and 88 underwent surgery (**Table 5**). In the unadjusted comparison, maternal and perinatal outcomes did not differ significantly, except for HDP. HDP occurred in 0 of 94 patients (0.0%) in the expectant group and 7 of 88 (8.0%) in the surgical group (P = 0.005). Miscarriage, preterm birth, PROM, PPH, LBW, SGA, NICU admission, 5-minute Apgar score <7, and perinatal death did not differ significantly (all P>0.05).

Sensitivity analysis and subgroup analysis

PSM was used to reduce measured baseline differences between the expectant and surgical groups. The propensity model included age, pre-pregnancy BMI, gestational age at detection, nulliparous status, elevated CA125, cyst diameter, wall characteristics, septation, papillary projections, internal echogenicity, color Doppler flow score, and blood-flow abundance. After 1:1 nearest-neighbor matching with a 0.2 caliper, 76 pairs remained, and all post-matching SMDs were <0.15 (Figure S8). The matched analysis was consistent with the unadjusted comparison: HDP occurred in 0 of 76 patients (0.0%) in the expectant group and 7 of 76 (9.2%) in the surgical group (P = 0.014), whereas the other maternal and perinatal outcomes did not differ significantly (Table S4 and **Figure 4**). Because PSM cannot account for unmeasured confounding, these findings should not be interpreted causally.

Maternal and perinatal outcomes did not differ significantly between laparoscopy (n = 70) and laparotomy (n = 18) (all P>0.05; Table S5), or

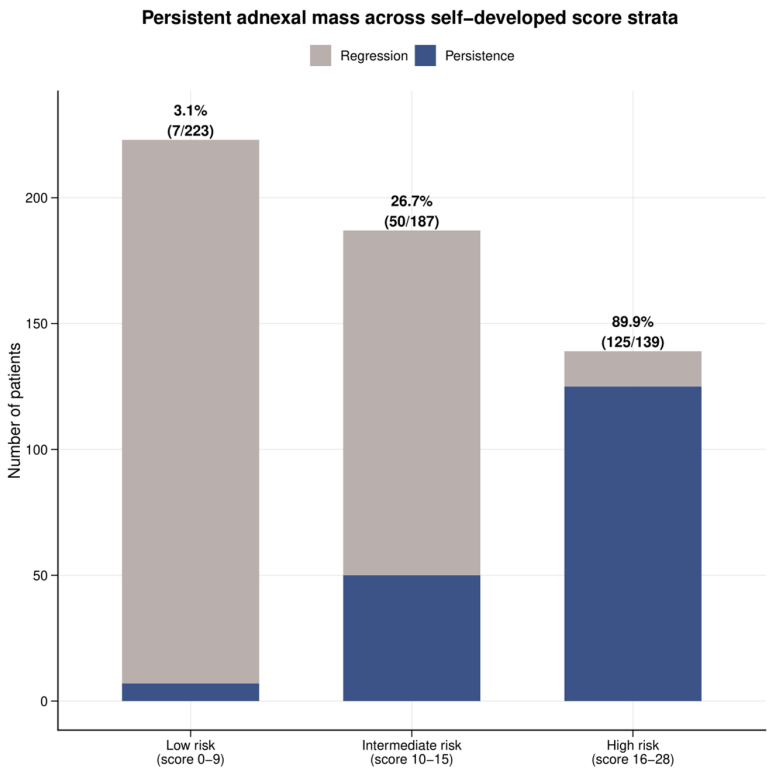


Figure 3. Persistence rates across URS-AMP risk tiers. Persistence rates were 3.1% in the low-risk group (0-9 points), 26.7% in the intermediate-risk group (10-15 points), and 89.9% in the high-risk group (16-28 points). Note: URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy.

between emergency (n = 11) and elective surgery (n = 77) (all P>0.05; Table S6). These comparisons were underpowered because the laparotomy and emergency-surgery subgroups were small and outcome events were uncommon. Figure S9 shows the distribution of the new URS-AMP score across the reference risk tiers; Figure S10 summarizes baseline imbalance between the regression and persistence groups.

Discussion

In this single-center cohort of 549 patients with benign adnexal masses during pregnancy, 66.8% of masses regressed spontaneously. We identified eight independent predictors of persistence, developed and internally validated the URS-AMP score, and compared expectant with surgical management using both unadjusted analyses and PSM. The observed regression rate was broadly consistent with previous reports [1, 2, 5], despite differences in case selection, follow-up schedules, and outcome definitions. Higher cyst-diameter, septation, and papillary-projection categories; thickened or irregular walls; elevated CA125; and non-functional pathological type were associated with persistence, whereas gravidity and nulliparous status were inversely associated. The integer score retained most of the full model's discrimination (AUC, 0.933 vs. 0.943). Most maternal and perinatal outcomes did not differ significantly by management strategy, although residual confounding limits interpretation.

Cyst-diameter category had the strongest adjusted association with persistence (OR = 4.443), consistent with previous evidence that larger masses are less likely to resolve [1, 2]. Larger size may reflect a non-functional lesion or more established cyst architecture, although the underlying biological mechanism cannot be determined from this study. Higher septation (OR = 3.168), higher papillary-projection category (OR = 2.009), and thickened or irregular walls (OR = 3.119) indicate greater structural complexity and may be more common in lesions such as cystadenomas or mature teratomas. These features are also important in malignancy assessment [19], but their role here was to predict persistence among masses classified as benign. Elevated CA125 was independently associated with persistence (OR = 2.365), but interpretation requires caution because CA125 may rise physiologically during pregnancy, especially in the first trimester [3]. Using the conventional >35 U/mL cutoff may reduce specificity and misclassify physiological elevations. Future studies should assess gestational-age-specific thresholds and evaluate whether CA125 adds incremental predictive value beyond ultrasound morphology.

Papillary projections require careful interpretation during pregnancy. In a series of 34 surgically removed ovarian masses, Mascilini et al.

Table 5. Maternal and perinatal outcomes in the expectant-management and surgical groups

Outcome	Expectant-management group (n = 94)	Surgical group (n = 88)	*P* value
Maternal outcomes			
Miscarriage	1 (1.1%)	3 (3.4%)	0.355
Preterm birth (<37 weeks)	4 (4.3%)	5 (5.7%)	0.741
PROM	5 (5.3%)	3 (3.4%)	0.722
HDP	0 (0.0%)	7 (8.0%)	**0.005**
PPH	5 (5.3%)	3 (3.4%)	0.722
Perinatal outcomes			
LBW (<2500 g)	7 (7.4%)	9 (10.2%)	0.604
SGA	6 (6.4%)	8 (9.1%)	0.583
NICU admission	2 (2.1%)	8 (9.1%)	0.052
5-min Apgar <7	2 (2.1%)	2 (2.3%)	1.000
Perinatal death	1 (1.1%)	0 (0.0%)	1.000

Note: Categorical variables are presented as number (percentage) and were compared with the χ^2 test, continuity-corrected χ^2 test, or Fisher's exact test, as appropriate. NICU, neonatal intensive care unit. All *P* values are two-sided; P<0.05 is shown in bold.

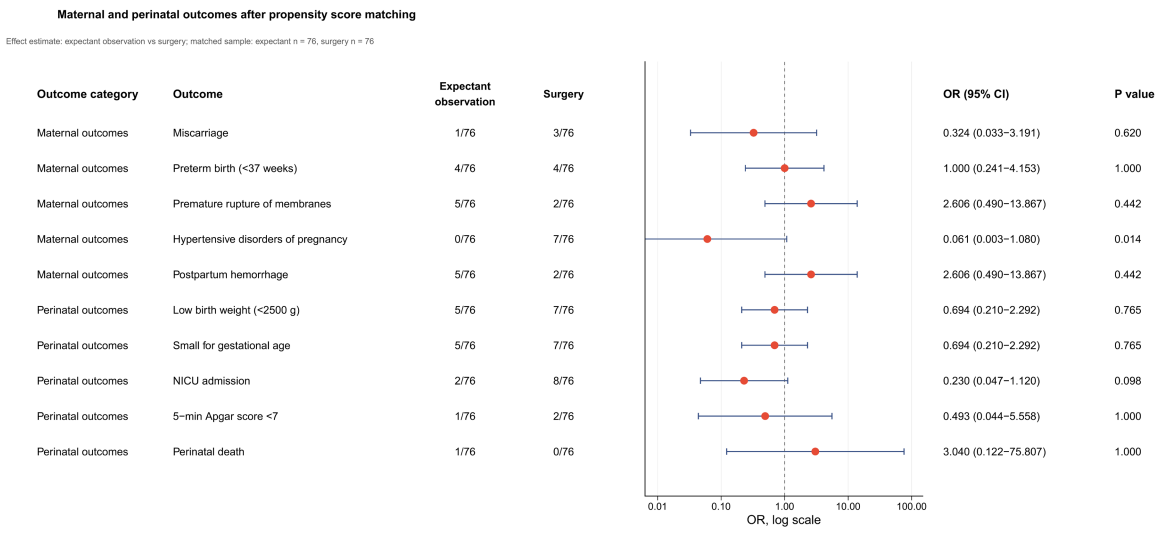


Figure 4. Maternal and perinatal outcomes after propensity-score matching. Odds ratios (ORs) and 95% confidence intervals (CIs) compare the expectant-management group with the surgical group after PSM (76 matched pairs). OR<1 favors expectant management, OR>1 favors surgery, and a 95% CI that crosses 1 indicates no statistically significant difference. Note: PSM, propensity-score matching; OR, odds ratio; CI, confidence interval; PROM, premature rupture of membranes; HDP, hypertensive disorders of pregnancy; PPH, postpartum hemorrhage; LBW, low birth weight; SGA, small for gestational age; NICU, neonatal intensive care unit.

[20] found that smooth papillary projections with ground-glass echogenicity were characteristic of decidualized endometriomas, whereas irregular projections with anechoic or low-level echogenicity were more concerning for borderline tumors. Decidualized endometriomas can also show marked vascularity and mimic malignancy [21]. Prospective observations indicate that many endometriomas and other endometriotic lesions regress during late pregnancy or postpartum [22, 23]. Thus, a smooth projection in a lesion with otherwise typical endometrioma features may justify short-interval expert ultrasound follow-up rather than immediate surgery, provided that formal malignancy assessment remains reassuring. Progressive or suspicious imaging findings should prompt timely referral to gynecologic oncology [24].

Gravidity and nulliparous status were inversely associated with persistence. The biological explanation is uncertain, and these findings

may reflect residual confounding or cohort-specific reproductive characteristics. They should be validated before being used to guide care.

URS-AMP uses routinely available clinical and ultrasound variables and therefore has potential practical value. Converting the full model to an integer score reduced the AUC by only 0.010. However, URS-AMP predicts persistence, not malignancy; its AUC should not be compared directly with malignancy models developed for different outcomes and populations [6, 25, 26]. Persistence rates differed substantially across the three reported risk groups. A low score may support expectant management, whereas a high score may justify closer surveillance and individualized consideration of elective surgery. Because irregular walls, thick septa, and papillary projections also overlap with established markers of ovarian malignancy, a high URS-AMP score should trigger a separate malignancy assessment with validated tools such as IOTA Simple Rules or O-RADS. The URS-AMP integer score had a significantly higher apparent AUC than the prespecified reference risk stratification according to the paired DeLong test. Nevertheless, this comparison remained exploratory because both classifications were evaluated in the same development dataset.

After PSM, most measured maternal and perinatal outcomes did not differ significantly between expectant and surgical management. This finding suggests that observation may be reasonable for carefully selected patients, particularly when sonographic features are reassuring, but it does not establish equivalence or safety in all patients. Previous studies support individualized management but mainly compare operative approaches or evaluate non-obstetric surgery more broadly [4, 5]. HDP occurred less often in the expectant group in both the unadjusted and matched analyses. Because only seven HDP events occurred and treatment selection may have depended on unmeasured symptoms, cyst growth, patient preference, or perceived torsion risk, this association should not be interpreted as a causal effect of surgery. Naing et al. [27] reported emergency surgery for torsion in 17.9% of conservatively managed masses ≥5 cm, illustrating the need to balance the risks of observation against operative and anesthetic risks. Larger prospective studies with detailed longitudinal and treatment-selection data are required.

The surgical subgroup analyses should also be interpreted with caution. Only 18 patients underwent laparotomy and 11 underwent emergency surgery, so the comparisons had limited statistical power. Patients selected for laparotomy had more complex sonographic features, consistent with confounding by indication. Zhang et al. [28] similarly found no major differences in pregnancy outcomes between laparoscopy and laparotomy for adnexal torsion, although laparoscopy was associated with less blood loss, a shorter hospital stays, and fewer wound-healing complications. When torsion or rupture risk is substantial, planned surgery may be preferable to waiting for an emergency; however, the present data cannot define the optimal threshold or timing.

This study has several limitations. First, the retrospective design is susceptible to selection and information bias. Management was not randomized. Although PSM balanced 12 measured covariates, unmeasured factors - including symptom severity, cyst growth, patient preference, and the surgeon's assessment of torsion risk - may have influenced treatment selection. The HDP finding should therefore be interpreted as an association rather than a causal effect. Second, ultrasound follow-up followed clinical practice rather than a fixed research schedule. Regression times were interval-censored with variable interval widths, so the study could not estimate a reliable median time to regression. The binary endpoint is less sensitive to examination timing but does not describe the trajectory of regression. Third, URS-AMP was validated only internally using bootstrap resampling. The small estimated optimism does not establish transportability, and external validation in independent centers is required before clinical implementation. Fourth, ultrasound interpretation depends on equipment quality and operator experience. Two senior sonographers reviewed the examinations, but formal interobserver and intraobserver agreement were not assessed. Fifth, pathological type was histologically confirmed only in the 88 surgical patients and was assigned from imaging patterns in the other 461 patients. This approach is clinically unavoidable for regressed lesions but introduces diagnostic uncertainty and possible incorporation bias. A low VIF does not eliminate this concern. Sixth, several maternal and perinatal outcomes were rare, resulting in imprecise esti-

mates and limited power, particularly in the laparotomy and emergency-surgery subgroups. Seventh, the dataset did not capture longitudinal changes in cyst size or morphology. Dynamic features may improve prediction and should be evaluated prospectively. Eighth, although paired DeLong testing showed a higher apparent AUC for the new score than for the reference risk stratification, both were evaluated in the same development dataset, and the reference categorization may not be independent of the model-development process. Therefore, this comparison should be interpreted as exploratory and requires external validation.

In conclusion, clinical and ultrasound morphological features were associated with persistence of benign adnexal masses during pregnancy. The URS-AMP score, which combines cyst diameter, wall morphology, septation, papillary projections, CA125, pathological type, gravidity, and nulliparous status, showed good apparent discrimination (AUC = 0.933) after internal bootstrap validation. In this cohort, persistence occurred in 3.1%, 26.7%, and 89.9% of the low-, intermediate-, and high-risk groups, respectively. Pending external validation, low-risk patients may be considered for routine expectant follow-up; intermediate-risk patients, for shorter-interval ultrasound surveillance; and high-risk patients, for individualized multidisciplinary assessment that considers gestational age, symptoms, cyst size, torsion risk, malignancy assessment, and patient preference. URS-AMP should complement rather than replace malignancy-focused systems such as IOTA Simple Rules or O-RADS.

Disclosure of conflict of interest

None.

Address correspondence to: Shan Li, Gynecology Ward I, Xi'an People's Hospital (Xi'an Fourth Hospital), No. 155 Hangtian East Road, Chang'an District, Xi'an 710000, Shaanxi, China. E-mail: 199301lishan@sina.com

References

[1] Cathcart AM, Nezhat FR, Emerson J, Pejovic T, Nezhat CH and Nezhat CR. Adnexal masses during pregnancy: diagnosis, treatment, and prognosis. Am J Obstet Gynecol 2023; 228: 601-612.

[2] Gaughran J, Magee C, Mitchell S, Knight CL and Sayasneh A. Adnexal masses in pregnancy: a single-centre prospective observational cohort study. Diagnostics (Basel) 2024; 14: 2182.

[3] Djokovic D, Pinto P and Reis I. Maternal adnexal masses in pregnancy. Best Pract Res Clin Obstet Gynaecol 2025; 102: 102643.

[4] Vatsa R, Singhal S, Goel R, Sharma AK, Kulshrestha V, Bharti J, Dhiman S, Dadhwal V and Malhotra N. Clinicopathological profile and obstetric outcome of pregnant women with persistent adnexal masses. J Obstet Gynaecol India 2025; 75 Suppl 1: 227-234.

[5] Dur R, Demirdag E, Yucel Celik O, Karahanoglu E, Dur G, Ozdemir CY, Yucel A and Moraloglu Tekin O. Adnexal masses and pregnancy: a single-center experience of 9 years. Arch Gynecol Obstet 2024; 310: 387-394.

[6] Lee SJ, Kim YH, Lee MY, Ko HS, Oh SY, Seol HJ, Kim JW, Ahn KH, Na S, Seong WJ, Kim HS, Park CW, Park JS, Jun JK, Won HS, Kim MY, Hwang HS and Lee SM. Ultrasonographic evaluation of ovarian mass for predicting malignancy in pregnant women. Gynecol Oncol 2021; 163: 385-391.

[7] Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, Paladini D, Van Calster B, Vergote I, Van Huffel S and Valentin L. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol 2008; 31: 681-690.

[8] Andreotti RF, Timmerman D, Strachowski LM, Froyman W, Benacerraf BR, Bennett GL, Bourne T, Brown DL, Coleman BG, Frates MC, Goldstein SR, Hamper UM, Horrow MM, Hernanz-Schulman M, Reinhold C, Rose SL, Whitcomb BP, Wolfman WL and Glanc P. O-RADS US risk stratification and management system: a consensus guideline from the ACR Ovarian-Adnexal Reporting and Data System Committee. Radiology 2020; 294: 168-185.

[9] Timmerman D, Van Calster B, Testa A, Savelli L, Fischerova D, Froyman W, Wynants L, Van Holsbeke C, Epstein E, Franchi D, Kaijser J, Czekierdowski A, Guerriero S, Fruscio R, Leone FPG, Rossi A, Landolfo C, Vergote I, Bourne T and Valentin L. Predicting the risk of malignancy in adnexal masses based on the Simple Rules from the International Ovarian Tumor Analysis group. Am J Obstet Gynecol 2016; 214: 424-437.

[10] Cagino K, Li X, Thomas C, Delgado D, Christos P and Acholonu U Jr. Surgical management of adnexal masses in pregnancy: a systematic review and meta-analysis. J Minim Invasive Gynecol 2021; 28: 1171-1182, e1172.

[11] Cusimano MC, Liu J, Azizi P, Zipursky J, Sajewycz K, Sussman J, Kishibe T, Wong E,

Ferguson SE, D'Souza R and Baxter NN. Adverse fetal outcomes and maternal mortality following nonobstetric abdominopelvic surgery in pregnancy: a systematic review and meta-analysis. Ann Surg 2023; 278: e147-e157.

[12] Haataja A, Kokki H, Uimari O and Kokki M. Non-obstetric surgery during pregnancy and the effects on maternal and fetal outcomes: a systematic review. Scand J Surg 2023; 112: 187-205.

[13] Moro F, Bolomini G, Sibal M, Vijayaraghavan SB, Venkatesh P, Nardelli F, Pasciuto T, Mascilini F, Pozzati F, Leone FPG, Josefsson H, Epstein E, Guerriero S, Scambia G, Valentin L and Testa AC. Imaging in gynecological disease (20): clinical and ultrasound characteristics of adnexal torsion. Ultrasound Obstet Gynecol 2020; 56: 934-943.

[14] World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human participants. JAMA 2025; 333: 71-74.

[15] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC and Vandenbroucke JP; STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007; 370: 1453-1457.

[16] Collins GS, Reitsma JB, Altman DG and Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015; 162: 55-63.

[17] Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG and Collins GS. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019; 38: 1276-1296.

[18] Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H and Vergote I; International Ovarian Tumor Analysis (IOTA) Group. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. Ultrasound Obstet Gynecol 2000; 16: 500-505.

[19] Lee SJ, Oh HR, Na S, Hwang HS and Lee SM. Ultrasonographic ovarian mass scoring system for predicting malignancy in pregnant women with ovarian mass. Obstet Gynecol Sci 2022; 65: 1-13.

[20] Mascilini F, Savelli L, Scifo MC, Exacoustos C, Timor-Tritsch IE, De Blasis I, Moruzzi MC, Pasciuto T, Scambia G, Valentin L and Testa AC. Ovarian masses with papillary projections diagnosed and removed during pregnancy: ultrasound features and histological diagnosis. Ultrasound Obstet Gynecol 2017; 50: 116-123.

[21] Yin M, Wang T, Li S, Zhang X and Yang J. Decidualized ovarian endometrioma mimicking malignancy in pregnancy: a case report and literature review. J Ovarian Res 2022; 15: 33.

[22] Bean E, Knez J, Setty T, Tetteh A, Casagrandi D, Naftalin J and Jurkovic D. Natural history of endometriosis in pregnancy: ultrasound study of morphology of deep endometriosis and ovarian endometrioma. Ultrasound Obstet Gynecol 2023; 62: 585-593.

[23] Zajicek M, Berkowitz E, Yulzari V, Kassif E, Burke Y, Elizur S, Inbar Y, Zolti M, Weisz B and Soriano D. Ultrasound appearance of decidualized non-ovarian endometriotic lesions during pregnancy and after delivery. Ultrasound Obstet Gynecol 2024; 63: 258-262.

[24] Causa Andrieu PI, Wahab SA, Nougaret S and Petkovska I. Ovarian cancer during pregnancy. Abdom Radiol (NY) 2023; 48: 1694-1708.

[25] Strachowski LM, Jha P, Phillips CH, Blanchette Porter MM, Froyman W, Glanc P, Guo Y, Patel MD, Reinhold C, Suh-Burgmann EJ, Timmerman D and Andreotti RF. O-RADS US v2022: an update from the American College of Radiology's Ovarian-Adnexal Reporting and Data System US Committee. Radiology 2023; 308: e230685.

[26] Han J, Wen J and Hu W. Comparison of O-RADS with the ADNEX model and IOTA SR for risk stratification of adnexal lesions: a systematic review and meta-analysis. Front Oncol 2024; 14: 1354837.

[27] Naing ZH, Ozaki R, Kitade M, Murakami K, Kawasaki Y and Itakura A. Effectiveness of prophylactic laparoscopic surgery for benign adnexal mass during pregnancy: a single-center retrospective study. Gynecol Minim Invasive Ther 2023; 13: 30-36.

[28] Zhang Z, Zhang Y, Fu H and Guo R. Laparotomy versus laparoscopy for the treatment of adnexal torsion during pregnancy. BMC Pregnancy Childbirth 2024; 24: 714.

Table S1. Baseline characteristics of the laparoscopy and laparotomy groups

Variable	Category	Laparoscopy group (n = 70)	Laparotomy group (n = 18)	Statistic	*P* value
Age (years)		31.46 ± 5.28	30.61 ± 6.15	t = 0.586	0.559
Pre-pregnancy BMI (kg/m^2)		23.51 ± 3.76	23.39 ± 3.63	t = 0.127	0.899
GA at detection (weeks)		12.53 ± 5.50	13.61 ± 4.79	t = -0.764	0.447
Nulliparous	No	37 (52.9%)	10 (55.6%)	-	1.000
	Yes	33 (47.1%)	8 (44.4%)		
Cyst diameter	<5 cm	19 (27.1%)	4 (22.2%)	χ^2 = 2.038	0.361
	5-9.9 cm	32 (45.7%)	6 (33.3%)		
	≥10 cm	19 (27.1%)	8 (44.4%)		
Septation	None	26 (37.1%)	2 (11.1%)	χ^2 = 9.155	0.010
	Thin regular	29 (41.4%)	6 (33.3%)		
	Thick irregular	15 (21.4%)	10 (55.6%)		
Internal echogenicity	Anechoic	30 (42.9%)	3 (16.7%)	χ^2 = 7.985	0.018
	Ground-glass	24 (34.3%)	5 (27.8%)		
	Mixed echogenicity	16 (22.9%)	10 (55.6%)		
Ascites	None	70 (100.0%)	16 (88.9%)	-	0.053
	Present	0 (0.0%)	2 (11.1%)		

Note: Continuous variables are presented as mean ± standard deviation and were compared with Student's t-test or Welch's t-test. Categorical variables are presented as number (percentage) and were compared with Pearson's χ^2 test, the continuity-corrected χ^2 test, or Fisher's exact test, as appropriate. An em dash indicates that Fisher's exact test was used and no test statistic is reported. BMI, body mass index. All *P* values are two-sided. Only selected baseline variables are shown.

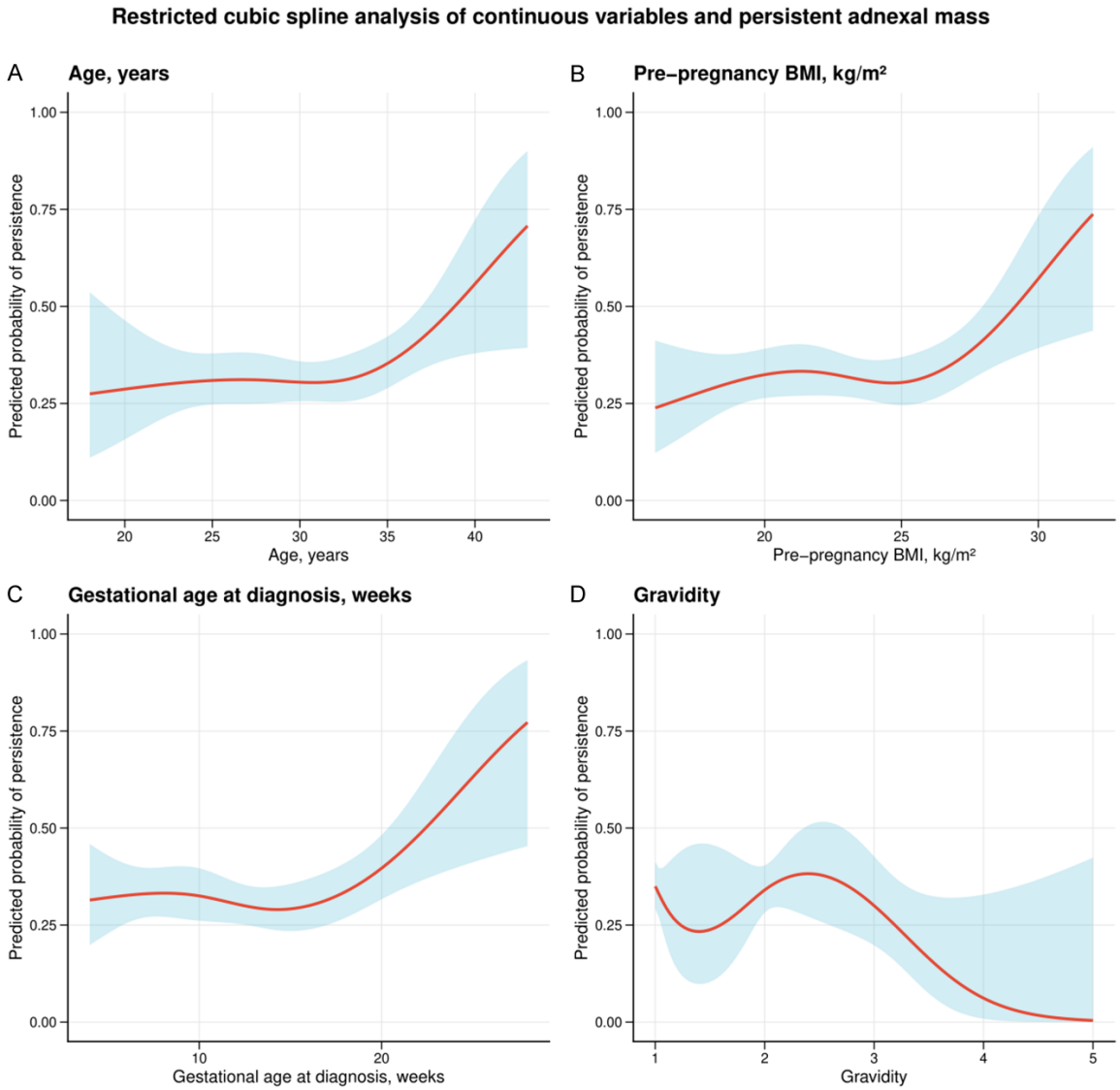


Figure S1. Restricted cubic spline analyses of continuous variables and spontaneous regression. Restricted cubic spline analysis of the relationships between age (A), gravidity (B), pre-pregnancy BMI (C), and gestational age at detection (D) and the probability of spontaneous cyst regression. Shaded areas represent 95% confidence intervals. Age, gravidity, and BMI did not show significant nonlinear trends; the nonlinear test for gestational age at detection yielded P = 0.034, suggesting a significant nonlinear association. Note: RCS, restricted cubic spline; BMI, body mass index.

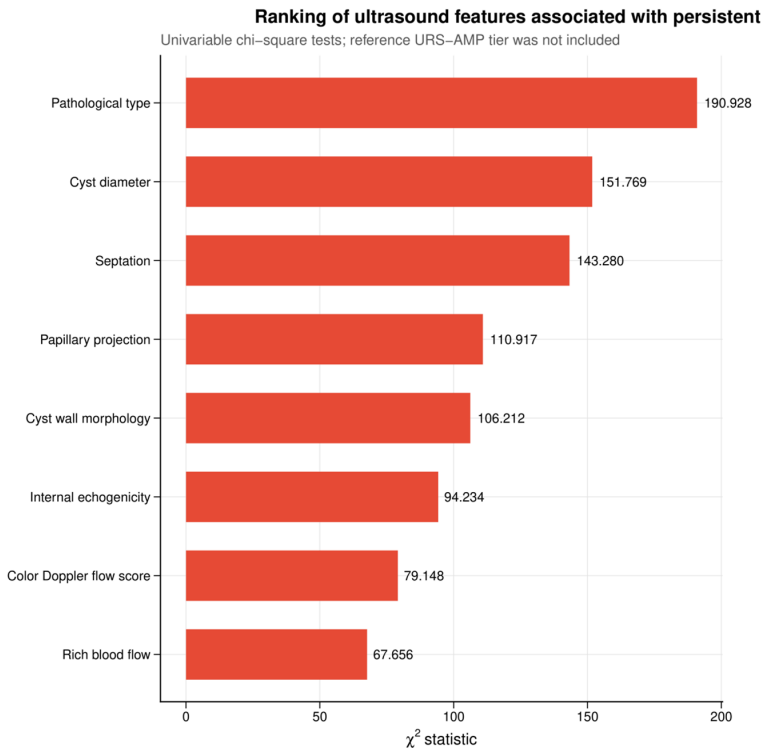


Figure S2. Ranking of associations between ultrasound features and regression status. Ultrasound features are ranked by the χ^2 statistic for their association with regression status. All eight variables had P<0.001. Abbreviations/labels: pathological type; cyst diameter; septation; papillary projections; cyst wall characteristics; internal echogenicity; color Doppler flow score; and blood-flow abundance.

Table S2. Spontaneous regression rates by pathological type

Pathological type	Total, n	Persistence, n	Regression, n	Regression rate, %
Corpus luteum cyst	132	13	119	90.2%
Simple cyst	225	34	191	84.9%
Endometrioma	60	36	24	40.0%
Mature teratoma	71	50	21	29.6%
Serous cystadenoma	36	28	8	22.2%
Mucinous cystadenoma	25	21	4	16.0%

Note: Values are numbers of patients unless otherwise indicated. Regression rate was calculated as the number of spontaneously regressed cases divided by the total number of cases within each pathological type. Functional cysts comprised corpus luteum and simple cysts; non-functional lesions comprised endometrioma, mature teratoma, serous cystadenoma, and mucinous cystadenoma. Rows are ordered by decreasing regression rate.

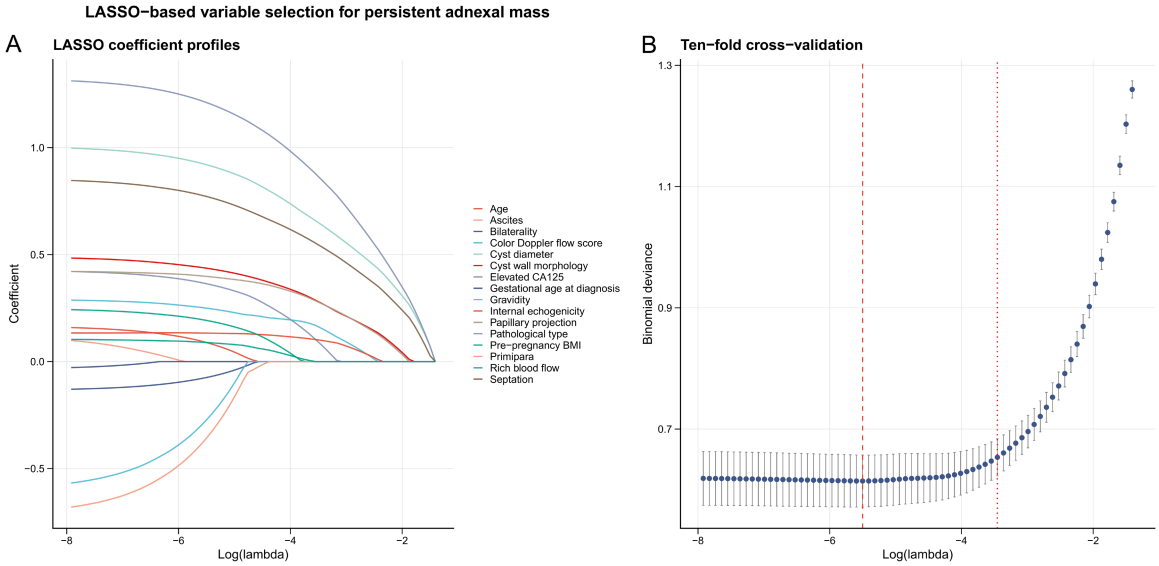
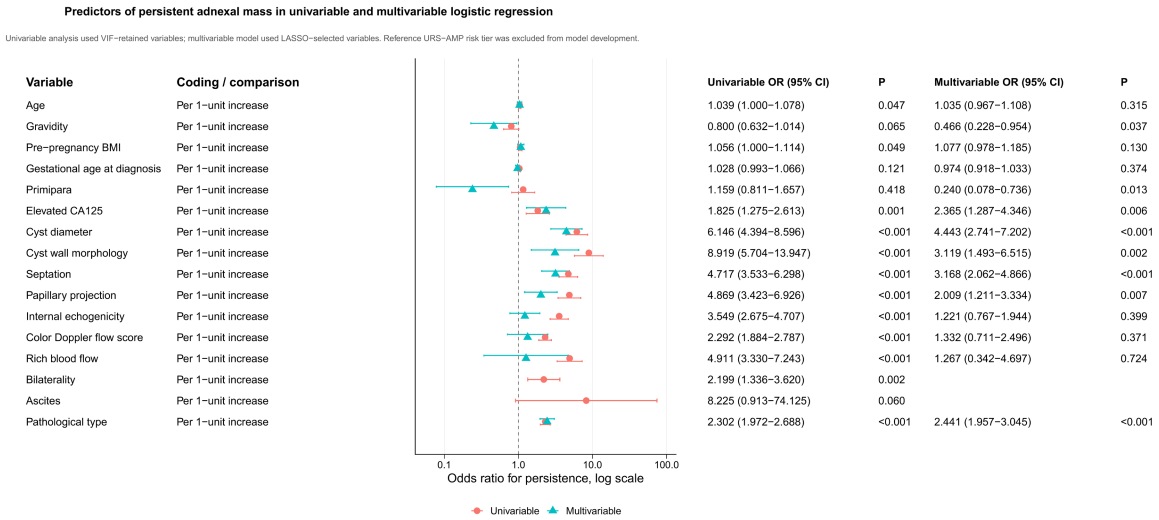


Figure S3. LASSO variable selection. A. LASSO coefficient paths showing the trajectories of each variable's coefficient as the penalty parameter λ varies. B. LASSO cross-validation curve; the left dashed line indicates λ.min and the right dashed line indicates λ.1se. Note: LASSO, least absolute shrinkage and selection operator.



Variable	Coding / comparison	Univariable OR (95% CI)	P	Multivariable OR (95% CI)	P
Age	Per 1−unit increase	1.039 (1.000−1.078)	0.047	1.035 (0.967−1.108)	0.315
Gravidity	Per 1−unit increase	0.800 (0.632−1.014)	0.065	0.466 (0.228−0.954)	0.037
Pre−pregnancy BMI	Per 1−unit increase	1.056 (1.000−1.114)	0.049	1.077 (0.978−1.185)	0.130
Gestational age at diagnosis	Per 1−unit increase	1.028 (0.993−1.066)	0.121	0.974 (0.918−1.033)	0.374
Primipara	Per 1−unit increase	1.159 (0.811−1.657)	0.418	0.240 (0.078−0.736)	0.013
Elevated CA125	Per 1−unit increase	1.825 (1.275−2.613)	0.001	2.365 (1.287−4.346)	0.006
Cyst diameter	Per 1−unit increase	6.146 (4.394−8.596)	<0.001	4.443 (2.741−7.202)	<0.001
Cyst wall morphology	Per 1−unit increase	8.919 (5.704−13.947)	<0.001	3.119 (1.493−6.515)	0.002
Septation	Per 1−unit increase	4.717 (3.533−6.298)	<0.001	3.168 (2.062−4.866)	<0.001
Papillary projection	Per 1−unit increase	4.869 (3.423−6.926)	<0.001	2.009 (1.211−3.334)	0.007
Internal echogenicity	Per 1−unit increase	3.549 (2.675−4.707)	<0.001	1.221 (0.767−1.944)	0.399
Color Doppler flow score	Per 1−unit increase	2.292 (1.884−2.787)	<0.001	1.332 (0.711−2.496)	0.371
Rich blood flow	Per 1−unit increase	4.911 (3.330−7.243)	<0.001	1.267 (0.342−4.697)	0.724
Bilaterality	Per 1−unit increase	2.199 (1.336−3.620)	0.002		
Ascites	Per 1−unit increase	8.225 (0.913−74.125)	0.060		
Pathological type	Per 1−unit increase	2.302 (1.972−2.688)	<0.001	2.441 (1.957−3.045)	<0.001

Figure S4. Univariable and multivariable logistic regression for cyst persistence. The left panel shows univariable estimates and the right panel shows multivariable estimates. Points represent ORs, horizontal lines represent 95% CIs, and the dashed vertical line indicates OR = 1. Note: OR, odds ratio; CI, confidence interval; BMI, body mass index; CA125, cancer antigen 125; GA, gestational age.

Table S3. Bootstrap internal validation of the URS-AMP scoring system (1,000 Resamples)

Metric	Value
Apparent AUC	0.943
Mean optimism	0.010
Optimism-corrected AUC	0.933
Valid bootstrap samples	1000/1000
URS-AMP integer score AUC (95% CI)	0.933 (0.911-0.956)
Optimal cutoff (Youden index)	12.5 points
Sensitivity	86.8%
Specificity	84.7%
PPV	73.8%
NPV	92.8%
Accuracy	85.4%

Note: AUC, area under the curve; CI, confidence interval; URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy. Bootstrap validation used 1,000 samples of the same size as the original cohort. The optimal cutoff was selected with the Youden index.

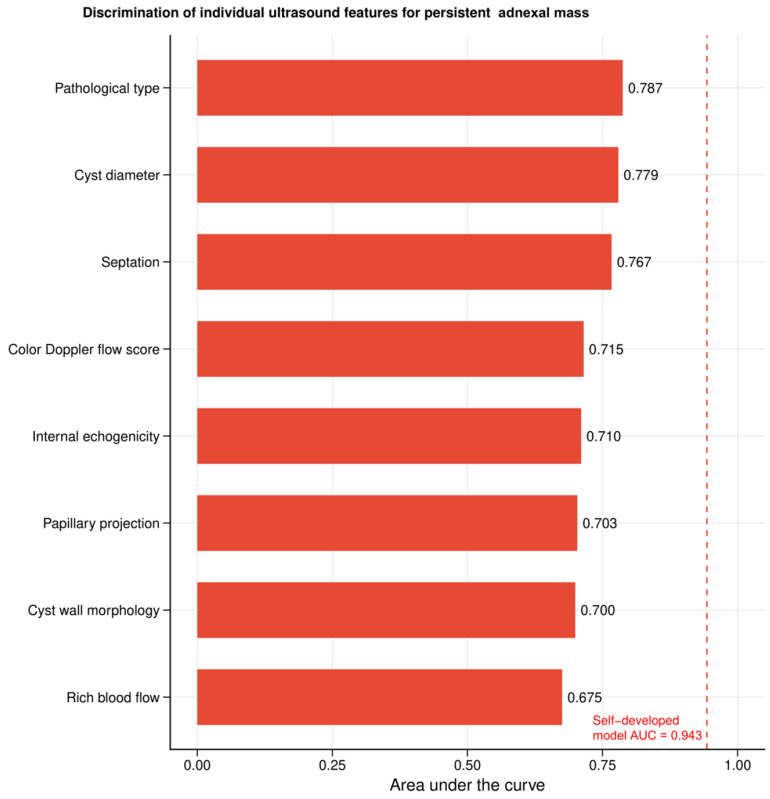


Figure S5. ROC curves for individual ultrasound features. ROC curves for pathological type (AUC = 0.787), cyst diameter (AUC = 0.779), septation (AUC = 0.767), blood flow score (AUC = 0.715), internal echogenicity (AUC = 0.710), papillary projections (AUC = 0.703), cyst wall characteristics (AUC = 0.700), and blood flow abundance (AUC = 0.675). Note: AUC, area under the curve.

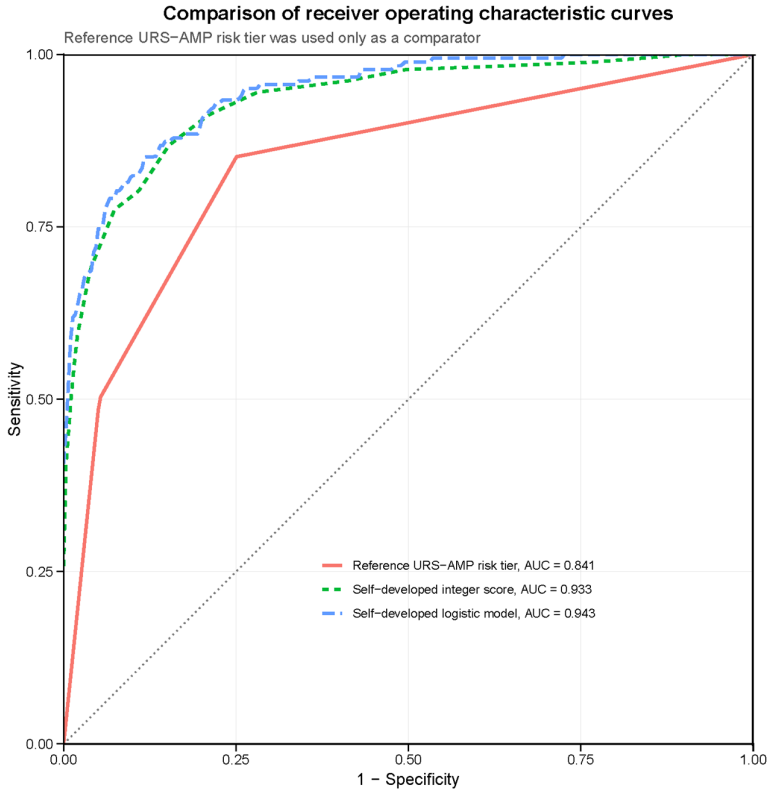


Figure S6. Exploratory ROC comparison of URS-AMP and the reference risk stratification. Descriptive ROC curves for the URS-AMP integer score (AUC = 0.933) and the reference risk stratification (AUC = 0.841). The URS-AMP integer score had a higher AUC than the reference risk stratification in a paired DeLong test (AUC difference, 0.092; Z = 4.990; P<0.001). This within-cohort comparison was exploratory. Note: URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy; ROC, receiver operating characteristic; AUC, area under the curve.

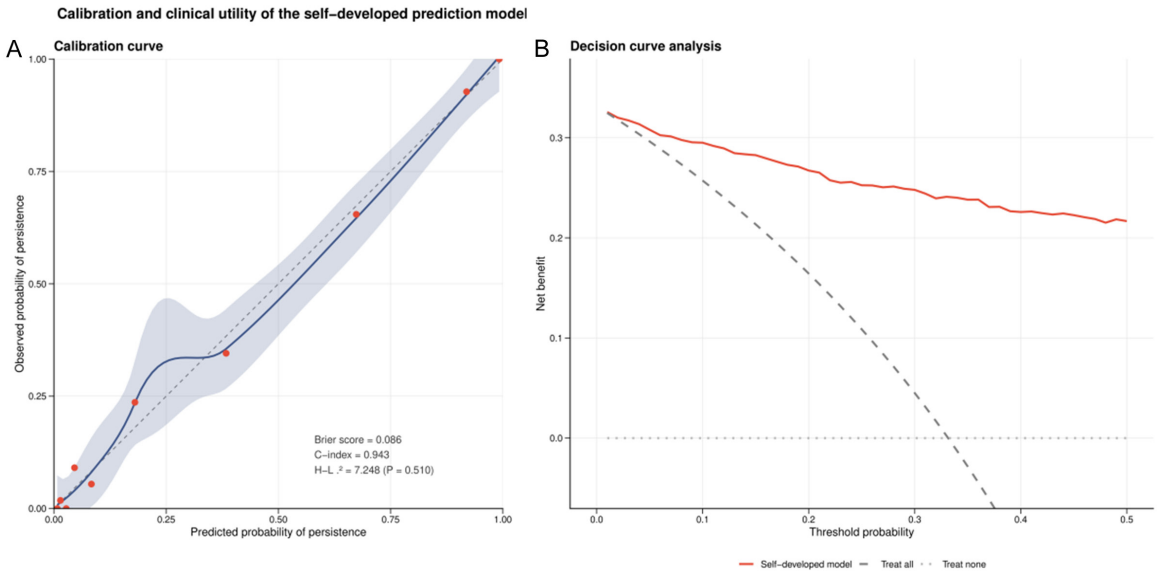


Figure S7. Calibration and decision-curve analysis. A. Calibration curve of the URS-AMP scoring system. The x-axis represents predicted probability, and the y-axis represents observed probability. The diagonal dashed line represents perfect calibration; the solid line is the apparent calibration curve. The plotted curve summarizes apparent calibration in the development cohort; optimism correction was applied to the AUC through 1,000 bootstrap resamples. B. Decision curve analysis (DCA). The x-axis represents threshold probability, and the y-axis represents net benefit. The red solid line represents the URS-AMP score, the gray solid line represents the “treat all” strategy, and the black solid line represents the “treat none” strategy. Note: DCA, decision-curve analysis.

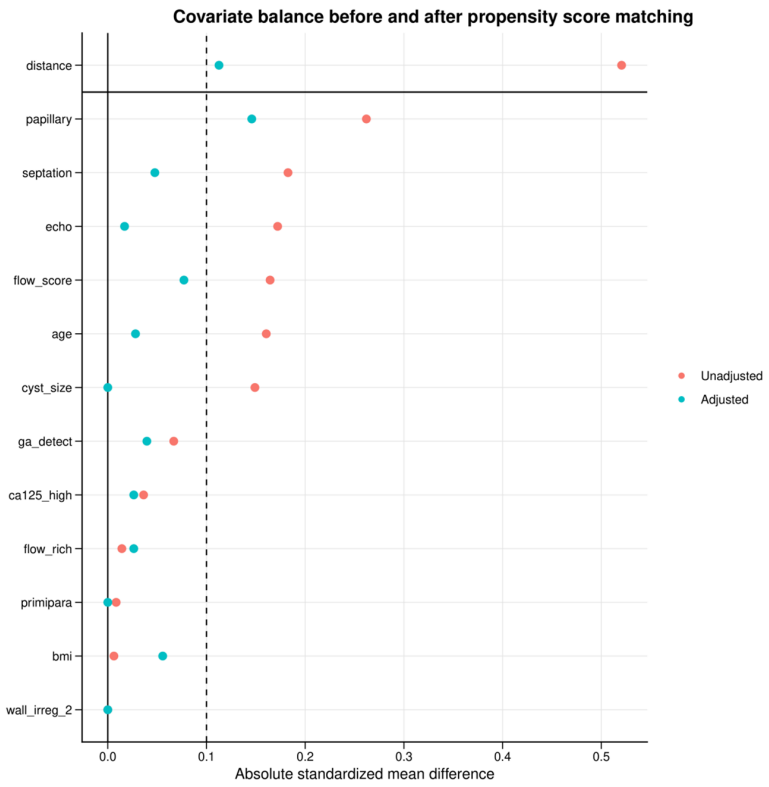


Figure S8. Covariate balance before and after propensity-score matching. Standardized mean differences (SMDs) for each covariate before and after matching. All post-matching SMDs were <0.15. Note: PSM, propensity-score matching; SMD, standardized mean difference; BMI, body mass index; GA, gestational age; CA125, cancer antigen 125.

Table S4. Maternal and perinatal outcomes after propensity-score matching

Outcome	Expectant-management group (n = 76)	Surgical group (n = 76)	*P* value
Maternal outcomes			
Miscarriage	1 (1.3%)	3 (3.9%)	0.620
Preterm birth (<37 weeks)	4 (5.3%)	4 (5.3%)	1.000
PROM	5 (6.6%)	2 (2.6%)	0.442
HDP	0 (0.0%)	7 (9.2%)	**0.014**
PPH	5 (6.6%)	2 (2.6%)	0.442
Perinatal outcomes			
LBW (<2500 g)	5 (6.6%)	7 (9.2%)	0.765
SGA	5 (6.6%)	7 (9.2%)	0.765
NICU admission	2 (2.6%)	8 (10.5%)	0.098
5-min Apgar <7	1 (1.3%)	2 (2.6%)	1.000
Perinatal death	1 (1.3%)	0 (0.0%)	1.000

Note: Categorical variables are presented as number (percentage) and were compared with the χ^2 test, continuity-corrected χ^2 test, or Fisher's exact test, as appropriate. PSM, propensity-score matching; BMI, body mass index; CA125, cancer antigen 125; NICU, neonatal intensive care unit. PSM used 1:1 nearest-neighbor matching with a caliper of 0.2. Covariates were age, pre-pregnancy BMI, gestational age at detection, nulliparous status, elevated CA125, cyst diameter, wall characteristics, septation, papillary projections, internal echogenicity, color Doppler flow score, and blood-flow abundance. The surgical group was the reference. All *P* values are two-sided; P<0.05 is shown in bold.

Table S5. Maternal and perinatal outcomes after laparoscopy and laparotomy

Outcome	Laparoscopy group (n = 70)	Laparotomy group (n = 18)	*P* value
Maternal outcomes			
Miscarriage	3 (4.3%)	0 (0.0%)	1.000
Preterm birth (<37 weeks)	3 (4.3%)	2 (11.1%)	0.270
PROM	2 (2.9%)	1 (5.6%)	0.501
HDP	4 (5.7%)	3 (16.7%)	0.148
PPH	3 (4.3%)	0 (0.0%)	1.000
Perinatal outcomes			
LBW (<2500 g)	7 (10.0%)	2 (11.1%)	1.000
SGA	6 (8.6%)	2 (11.1%)	0.664
NICU admission	7 (10.0%)	1 (5.6%)	1.000
5-min Apgar <7	2 (2.9%)	0 (0.0%)	1.000
Perinatal death	0 (0.0%)	0 (0.0%)	1.000

Note: Categorical variables are presented as number (percentage) and were compared with the χ^2 test, continuity-corrected χ^2 test, or Fisher's exact test, as appropriate. NICU, neonatal intensive care unit. The laparotomy group was the reference. Because the laparotomy subgroup was small (n = 18), estimates should be interpreted cautiously. All *P* values are two-sided.

Table S6. Maternal and perinatal outcomes after emergency and elective surgery

Outcome	Emergency-surgery group (n = 11)	Elective-surgery group (n = 77)	*P* value
Maternal outcomes			
Miscarriage	1 (9.1%)	2 (2.6%)	0.333
Preterm birth (<37 weeks)	0 (0.0%)	5 (6.5%)	1.000
PROM	0 (0.0%)	3 (3.9%)	1.000
HDP	0 (0.0%)	7 (9.1%)	0.589
PPH	1 (9.1%)	2 (2.6%)	0.333
Perinatal outcomes			
LBW (<2500 g)	2 (18.2%)	7 (9.1%)	0.313
SGA	2 (18.2%)	6 (7.8%)	0.261
NICU admission	1 (9.1%)	7 (9.1%)	1.000
5-min Apgar <7	1 (9.1%)	1 (1.3%)	0.236
Perinatal death	0 (0.0%)	0 (0.0%)	1.000

Note: Categorical variables are presented as number (percentage) and were compared with the χ^2 test, continuity-corrected χ^2 test, or Fisher's exact test, as appropriate. NICU, neonatal intensive care unit. The elective-surgery group was the reference. Because the emergency-surgery subgroup was small (n = 11), estimates should be interpreted cautiously. All *P* values are two-sided.

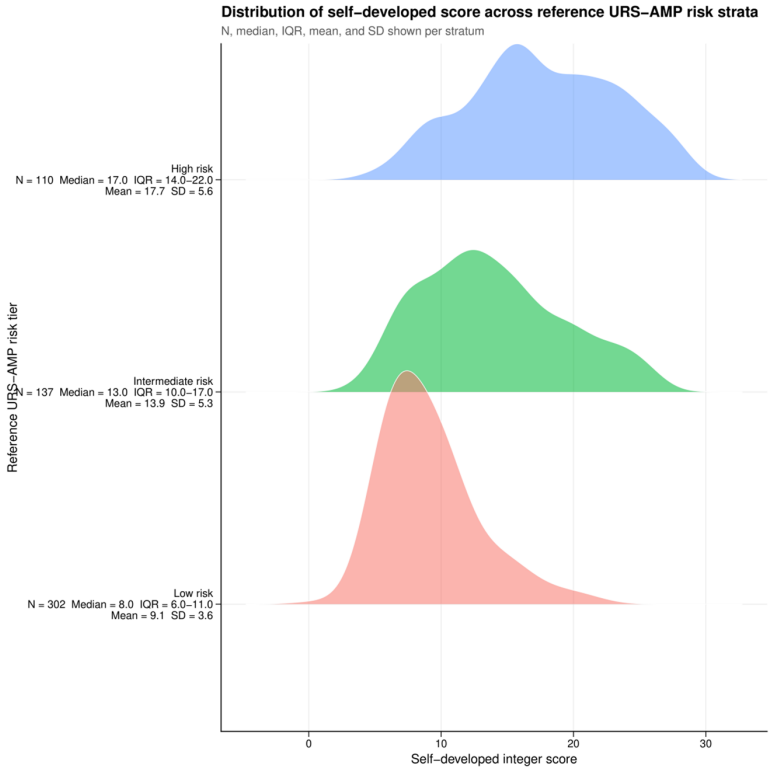


Figure S9. Distribution of the new URS-AMP score across reference risk tiers. Ridgeline plot of the distribution of the new URS-AMP integer score across the reference low-, intermediate-, and high-risk tiers. Note: URS-AMP, Ultrasound Risk Scoring System for Adnexal Mass Persistence in Pregnancy.

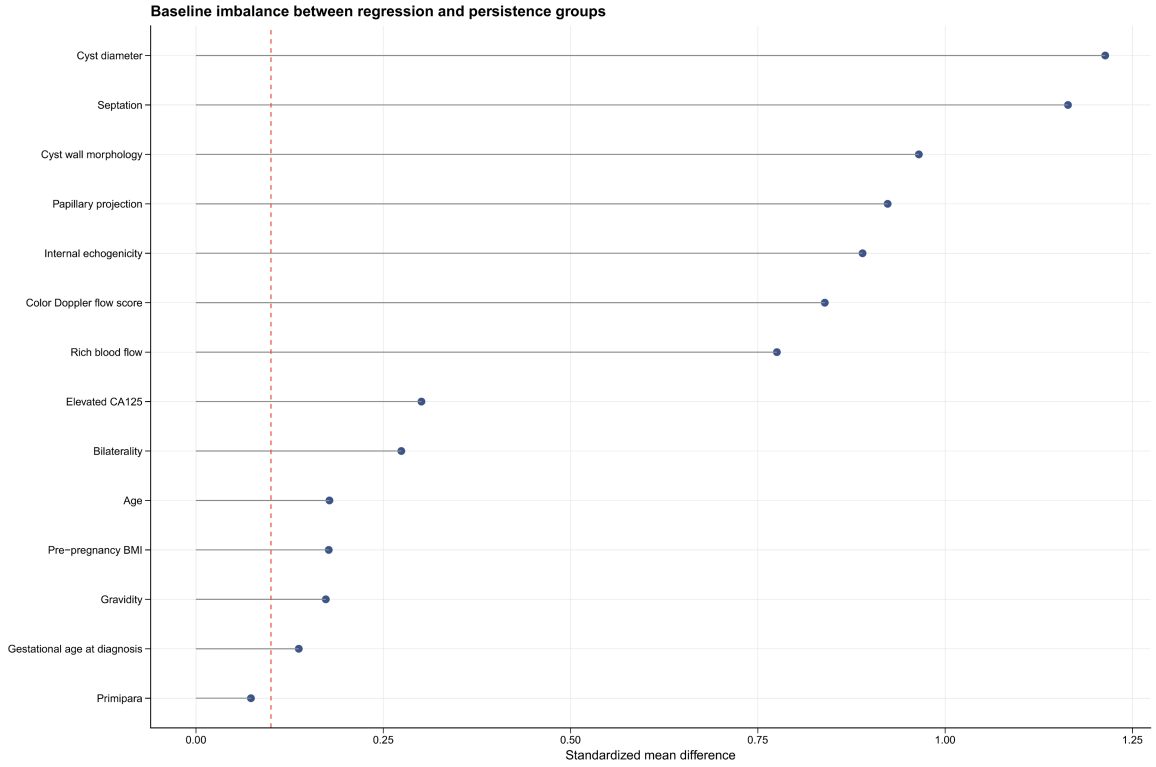


Figure S10. Baseline imbalance between the regression and persistence groups. Standardized mean differences (SMDs) for baseline variables in the regression (n = 367) and persistence (n = 182) groups. Larger SMDs for ultrasound features indicate more pronounced between-group differences. Note: SMD, standardized mean difference.