**ORIGINAL RESEARCH ARTICLE**

# Identification of risk factors and construction of a predictive model for postoperative new-onset stress urinary incontinence in patients with pelvic organ prolapse: A single-center retrospective study

DOI: 10.29063/ajrh2026/v30i2.11

*Yan Yue*[*]*, Wei Wu, Jianshan Zhou and Liming Wang*

Department of Gynecology, Maternal and Child Health Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430070, Hubei Province, China

***For Correspondence:** Email: *yueyan7710@126.com*

## Abstract

This study was to identify independent risk factors for new-onset stress urinary incontinence (SUI) after pelvic organ prolapse (POP) surgery and to develop and validate optimized prediction models, thereby providing an evidence-based tool for clinical decision-making. A single-center retrospective cohort study was conducted, including 213 patients who underwent POP surgery. Data on baseline characteristics, preoperative assessments, surgical details, and 12-month postoperative follow-up were collected via the electronic medical record system. Multiple imputation was used to handle missing data, least absolute shrinkage and selection operator (LASSO) regression for feature selection, and both logistic regression and classification and regression tree (CART) decision tree models were developed. LASSO regression identified preoperative POP-Q stage, abnormal urodynamics with prolapse reduction, postmenopausal duration >10 years, and parity as core risk factors. The logistic regression model achieved an internally validated area under curve (AUC) of 0.78 and accuracy of 83.0%, with external validation AUC of 0.75. After calibration, Hosmer-Lemeshow test yielded *P*=0.65. The CART decision tree model demonstrated an internally validated AUC of 0.84 and accuracy of 87.4%, with external validation AUC of 0.81 and accuracy of 85.2%. Preoperative POP-Q stage and abnormal urodynamics with prolapse reduction are the strongest predictors of new-onset SUI after POP surgery. *(Afr J Reprod Health 2026; 30 [2]: 130-144).*

**Keywords:** POP; postoperative new-onset SUI; predictive model; logistic regression; decision tree

## Résumé

Cette étude visait à identifier les facteurs de risque indépendants d'incontinence urinaire d'effort (IUE) d'apparition récente après une chirurgie pour prolapsus des organes pelviens (POP) et à développer et valider des modèles de prédiction optimisés, fournissant ainsi un outil factuel pour la prise de décision clinique. Une étude de cohorte rétrospective monocentrique a été menée, incluant 213 patientes ayant subi une chirurgie pour prolapsus des organes pelviens (POP). Les données sur les caractéristiques initiales, les évaluations préopératoires, les détails chirurgicaux et le suivi postopératoire à 12 mois ont été collectées via le système de dossiers médicaux électroniques. L'imputation multiple a été utilisée pour gérer les données manquantes, la régression LASSO (Least Absolute Shrinkage and Selection Operator) pour la sélection des caractéristiques, et des modèles de régression logistique et d'arbre de décision CART (Classification and Regression Tree) ont été développés. La régression LASSO a identifié le stade POP-Q préopératoire, une urodynamique anormale avec réduction du prolapsus, une durée postménopausique > 10 ans et la parité comme principaux facteurs de risque. Français Le modèle de régression logistique a atteint une aire sous la courbe (ASC) validée en interne de 0,78 et une précision de 83,0 %, avec une ASC de validation externe de 0,75. Après étalonnage, le test de Hosmer-Lemeshow a donné un p = 0,65. Le modèle d'arbre de décision CART a démontré une ASC validée en interne de 0,84 et une précision de 87,4 %, avec une ASC de validation externe de 0,81 et une précision de 85,2 %. Le stade POP-Q préopératoire et une urodynamique anormale avec réduction du prolapsus sont les meilleurs prédicteurs d'une IUE d'apparition récente après une chirurgie POP. *(Afr J Reprod Health 2026; 30 [2]: 130-144).*

**Mots-clés**: IUE postopératoire d'apparition récente ; modèle prédictif ; régression logistique ; arbre de décision

## Introduction

Pelvic organ prolapse (POP), a prevalent pelvic floor dysfunction disorder among middle-aged and elderly women, is fundamentally characterized by the impairment or weakening of pelvic support structures due to factors such as pregnancy, childbirth, and aging. This leads to abnormal

displacement of pelvic organs including the uterus, vagina, bladder, and rectum.[1-3] Clinically, patients often present with symptoms such as pelvic heaviness, urinary incontinence, and defecatory dysfunction. These manifestations not only significantly disrupt normal physiological functions but also adversely impact psychological well-being and social participation, markedly diminishing quality of life.[4-6] Notably, multiparous and postmenopausal women exhibit particularly high incidence rates of POP, primarily attributable to long-term pelvic floor tissue damage and hormonal changes. This has emerged as a critical public health concern requiring urgent attention.

Currently, surgical reconstruction of pelvic support structures remains the primary treatment for moderate-to-severe POP. However, the clinical outcomes are not entirely satisfactory, with postoperative new-onset stress urinary incontinence (SUI) emerging as a common complication occurring in 15%-30% of cases, representing a major challenge affecting surgical prognosis. This postoperative new-onset SUI not only prolongs patient recovery but may also necessitate secondary surgical interventions, thereby increasing healthcare burdens.[7,8] Although existing studies have recognized its clinical significance, there remains a notable lack of accurate risk assessment tools in clinical practice. Current management predominantly relies on physicians' empirical judgment, which may lead to either under-identification of high-risk patients or overtreatment of low-risk cases.[9-11]

A comprehensive investigation of risk factors for postoperative new-onset SUI and the development of reliable predictive models hold significant clinical implications. First, such models would enable clinicians to identify high-risk populations for targeted concomitant anti-incontinence procedures. Second, they could prevent low-risk patients from undergoing unnecessary invasive interventions, thereby optimizing the risk-benefit balance of treatment strategies.[12,13] However, existing studies present several limitations: i) ambiguous definitions of "new-onset SUI," with some erroneously including aggravated preexisting occult SUI in their analyses; ii) restricted predictive power due to inadequate sample sizes or suboptimal handling of missing data; and most critically, iii) lack of standardization in urodynamic study (UDS) protocols-particularly regarding whether testing was performed under prolapse reduction-compromising result comparability.[14] To address these gaps, this retrospective cohort study strictly enrolled POP patients without preoperative observable SUI symptoms. Through systematic analysis of clinical data, this study aimed to identify independent risk factors for postoperative new-onset SUI, develop and optimize an accurate predictive model, and provide robust decision-support tools for clinical practice.

# Methods

## *Research design and source of research object*

This study is a single-center retrospective cohort study. The study subjects were sourced from a database of patients who underwent POP surgery at Hubei Maternal and Child Health Hospital between June 2022 and June 2024. Cases meeting the inclusion and exclusion criteria were extracted through the hospital's electronic medical record system. The database initially recorded 328 POP surgery patients. After stepwise screening according to the inclusion and exclusion criteria, 213 patients were ultimately included. The screening process was as follows: total cases (n=328); excluded 52 cases with preoperative SUI (positive cough stress test or self-reported SUI symptoms); excluded 42 cases with postoperative follow-up <12 months; excluded 19 cases with >20% missing baseline data; final inclusion: 213 cases.

## *Inclusion and exclusion criteria*

### *Inclusion criteria*

Age ≥ 40 years; diagnosed with POP confirmed by clinical examination and POP quantification (POP-Q) staging, with POP-Q stage ≥ II, and scheduled for surgical treatment; absence of preoperative SUI manifestations: negative cough stress test without prolapse reduction and no self-reported symptoms of stress urinary leakage, including leakage during coughing, sneezing, or heavy lifting; complete baseline clinical data: including demographic

information, preoperative pelvic floor function assessment (POP-Q stage, urodynamic results), surgical records, and other key data without missing values; postoperative follow-up data traceable, with completion of at least 12 months of postoperative follow-up.

### *Exclusion criteria*

Preoperative SUI (positive cough stress test with or without prolapse reduction) or other types of urinary incontinence (*e.g.*, urge incontinence, mixed incontinence); comorbid neurological disorders (*e.g.*, spinal cord injury, multiple sclerosis), urethral anatomical abnormalities (*e.g.*, urethral diverticulum), or urinary system diseases (*e.g.*, bladder stones, urinary tract infection) that may confound postoperative SUI assessment; history of anti-incontinence surgery, pelvic radiotherapy, or major pelvic surgery (*e.g.*, radical hysterectomy) that may affect pelvic floor structure and function evaluation; interruption of postoperative follow-up due to psychiatric disorders, cognitive impairment, or loss to follow-up, preventing completion of 12-month follow-up; baseline data missingness >20%, or critical variables (*e.g.*, POP-Q stage, urodynamic results) missing and unable to be reasonably imputed.

### *Follow-up plan*

(1) Follow-up methods and time points
Postoperative follow-up was conducted through a combination of outpatient reviews and supplementary telephone interviews at three time points: 3 months, 6 months, and 12 months after surgery.
Outpatient review: patients were required to return to the hospital for gynecological examinations, including POP-Q staging reassessment, cough stress test, and UDS.
Telephone interview: for patients unable to attend the hospital on schedule, telephone interviews were conducted to confirm urinary incontinence symptoms, such as the occurrence of stress urinary leakage and leakage frequency, and to trace recent outpatient examination records.
(2) Follow-up completion criteria and handling of loss to follow-up
Definition of "completed follow-up": At twelve months postoperatively, patients must have completed at least one outpatient review including cough stress test and UDS examination; or, via telephone confirmation, reported no SUI symptoms and provided outpatient review records from at least 6 months postoperatively.

Assessment of loss to follow-up: for patients with follow-up <12 months due to loss of contact or personal reasons for refusal, the time of loss to follow-up (*e.g.*, 6 months, 10 months postoperatively) was recorded. Baseline characteristics (age, POP-Q stage, menopausal status, BMI) of lost-to-follow-up patients were compared with those of included patients using chi-square tests (categorical variables) or independent samples t-tests (continuous variables) to analyze intergroup differences and evaluate attrition bias.

### *Data collection and variable definition*

(1) Data collection personnel and tools
Data were independently extracted from the electronic medical record system by two uniformly trained gynecologists. Inter-rater reliability was assessed prior to extraction (Kappa coefficient > 0.85). Discrepancies were resolved by a third senior physician. A standardized data extraction form covering four variable categories was used.
(2) Variable classification and definitions
a. Demographic and baseline characteristics
Age: continuous variable (years).
Parity: categorical variable (parous: ≥1 delivery; nulliparous: 0 deliveries).
Menopausal status and duration: categorical variables (menopausal status: menopausal/premenopausal; menopausal duration: ≤10 years/>10 years; menopause defined as ≥12 months of amenorrhea since last menstrual period).
Body mass index (BMI): continuous variable (kg/m², calculated as weight/height²).
Comorbidities: categorical variables (history of hypertension or diabetes, defined as preoperatively diagnosed and medically managed).
b. Preoperative pelvic floor function assessment
POP-Q stage: continuous variable (recorded as specific values for points Aa, Ba, and C, or quantified scores: stage II = 2, stage III = 3, stage IV = 4).

UDS results: categorical variable (abnormal/normal; abnormal defined as reduced urethral closure pressure, detrusor over activity, urodynamic stress incontinence, *etc.*, assessed with prolapse reduction).
ICIQ-SF score: continuous variable (range 0-21, assessing preoperative urinary symptom severity).[15]
Pelvic floor muscle strength: categorical variable (assessed via Oxford Grading Scale, 0-5).[16]
c. Surgery-related variables
Surgical type: categorical variable (*e.g.*, transvaginal pelvic reconstruction + sacrospinous ligament fixation, laparoscopic sacrocolpopexy, vaginal hysterectomy + adnexal surgery, transvaginal pelvic reconstruction + posterior colporrhaphy + cystoscopy, and other combined procedures).
Operative duration: continuous variable (minutes, from incision to closure).
Intraoperative blood loss: continuous variable (mL, estimated via suction volume + gauze weight).
Mesh use: categorical variable (yes/no).
d. Study outcome variable
Postoperative new-onset SUI: defined as stress urinary leakage during activities increasing abdominal pressure (*e.g.*, coughing, sneezing, heavy lifting) within 12 months postoperatively, confirmed by urodynamic stress incontinence on UDS with prolapse reduction, and excluding other causes (*e.g.*, urinary tract infection, bladder stones, detrusor over activity).

## *Data preprocessing*

### *Missing data handling*

a. Assessment and reporting of missing data
The missingness of all variables was first assessed, including the number of missing cases and missing rates. The specific handling principles were as follows:
Variables with a missing rate >30% were excluded due to excessive information loss.
Variables with a missing rate ≤30% were imputed using multiple imputation by chained equations (MICE), with the imputation model and parameters specified prior to imputation.
Patients with a missing rate >20% were excluded directly.
The missingness of key variables is summarized in Table 1
b. MICE imputation details
Imputation model: included all baseline variables (age, parity, menopausal duration, BMI, POP-Q stage, UDS results, *etc.*) and the outcome variable (postoperative new-onset SUI) to ensure associations between variables were considered during imputation.
Imputation method: methods were selected based on variable types. Continuous variables were imputed using predictive mean matching, and categorical variables were imputed using multinomial logistic regression.
Number of imputations and pooling: the number of imputations was set to 5, with one complete dataset generated per imputation. For final analysis, results from the 5 datasets were pooled using the method of combining predicted probabilities, and the mean of the 5 model-predicted probabilities was used as the final predicted value.
c. Independence of imputation and validation
During subsequent internal validation (10-fold cross-validation), MICE imputation was performed separately for each fold. Specifically, data were first split into training and validation sets, and imputation was applied only to the training set to avoid data leakage that would occur if imputation were performed on the entire dataset prior to splitting. This ensured unbiased validation results.

### *Outlier handling*

Extreme values were verified against original medical records (*e.g.*, surgical notes, examination reports) to confirm whether they were recording errors. Extreme values confirmed not to be recording errors were retained, and no cases were excluded.

### *Variable transformation and encoding*

a. Continuous variable processing
Continuous variables (*e.g.*, age, POP-Q score, BMI) were standardized using Z-score normalization (equation: $Z = (X - \mu)/\sigma$, where μ is the mean and σ is the standard deviation) to eliminate scale differences.

**Table 1:** Missingness of key variables

| Variable | Missing cases | Missing rate | Handling method |
|---|---|---|---|
| Intraoperative blood loss | 72 | 33.80% | Variable excluded |
| Preoperative ICIQ-SF score | 18 | 8.50% | MICE imputation |
| Menopausal duration | 11 | 5.20% | MICE imputation |
| Pelvic floor muscle strength | 9 | 4.20% | MICE imputation |
| Total patients (missing >20%) | 19 | 8.90% | Patient excluded |

Some continuous variables were discretized for subsequent categorical variable analysis.
b. Categorical variable encoding
Binary variables (*e.g.,* UDS results: abnormal = 1, normal = 0; parity: parous = 1, nulliparous = 0) were encoded using 0-1 coding.
Multicategorical variables (*e.g.*, surgical type, POP-Q stage) were encoded using one-hot encoding to avoid false ordinal relationships.
Ordinal categorical variables (*e.g.,* pelvic floor muscle strength: grades 0-5) were encoded ordinally (0 = grade 0, 1 = grade 1, *etc*.).

### *Sample size estimation*

The sample size was estimated using modern clinical prediction model (CPM) methodology with the following parameters: [17]
Outcome event: postoperative new-onset SUI
Significance level (α): 0.05 (two-tailed test)
Power (1-β): 0.8 (β=0.2)
Expected outcome incidence rate (p): 30%
Number of initial candidate variables (p): 12 (age, parity, menopausal status, menopausal duration, BMI, POP-Q stage, UDS results, surgical type, preoperative ICIQ-SF score, pelvic floor muscle strength, history of hypertension, history of diabetes)
Expected model explained variance ($R^2$): 0.2
The calculated minimum required sample size was 189 cases. Accounting for a potential 10% loss to follow-up, this study ultimately included 213 patients, meeting the sample size requirement.

### *Feature selection method*

A two-step feature selection method combining clinical rationality screening and LASSO regression was employed, with the following specific steps:

1.Clinical rationality screening
Based on existing clinical evidence and guidelines, variables without a clear association with postoperative new-onset SUI were excluded.
a. Exclusion of hypertension history and diabetes history: Insufficient evidence currently exists to directly link these comorbidities with new-onset SUI after POP surgery.
b. Retention of ten candidate variables: age, parity, menopausal status, menopausal duration, BMI, POP-Q stage, UDS results, surgical type, preoperative ICIQ-SF score, and pelvic floor muscle strength.
2. Dimensionality reduction via LASSO regression
LASSO regression was applied to further screen key variables and control model overfitting.
a. Model specification: with postoperative new-onset SUI as the dependent variable and the aforementioned 10 candidate variables as independent variables, 5-fold cross-validation was used to select the optimal regularization parameter λ, corresponding to the minimum AIC value.
b. Variable retention principle: variables whose coefficients were not compressed to zero at the optimal λ value were retained as key variables.
c. Independence assurance: LASSO regression was performed within the nested cross-validation framework of internal validation, with feature selection conducted separately on each training set fold to avoid validation bias introduced by feature selection

### *Prediction model construction*

Logistic regression model and classification and regression tree (CART)decision tree model were developed separately, with distinct roles and parameter settings defined for each model.
1. Logistic regression model

a. Role: provides quantitative risk prediction probabilities, suitable for clinicians to perform individualized risk assessments for patients.
b. Model construction: key variables selected by LASSO regression were used as independent variables, with postoperative new-onset SUI as the dependent variable. The model was fitted using maximum likelihood estimation to calculate regression coefficients (β), odds ratios (OR), and 95% confidence intervals (CI) for each variable.
c. Risk probability calculation: the predicted probability of postoperative new-onset SUI was derived through Logit function transformation, with the formula:
$\text{Logit}(P) = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \ldots + \beta_n \times X_n$,
where $\beta_0$ is the intercept, $X_1$-$X_n$ are the key variables, and $\beta_1$-$\beta_n$ are the corresponding regression coefficients.

### *CART decision tree model*

a. Role: provides intuitive classification rules, suitable for rapid screening of patients at high risk of postoperative new-onset SUI.
b. Base algorithm and parameters: the CART algorithm was employed, using the Gini index as the node splitting criterion. Initial parameter settings: max_depth=10 (maximum tree depth), min_samples_leaf=5 (minimum samples per leaf node).
c.Hyperparameter optimization: key hyperparameters were optimized via GridSearchCV grid search, with the following search ranges:
max_depth: 8–15 (tree depth)
min_samples_split: 10–20 (minimum samples required to split a node)
min_samples_leaf: 3–8 (minimum samples per leaf node)
max_features: 3–8 (maximum number of features considered for splitting at each node)
The parameter combination corresponding to the minimum cross-validation error was selected as the optimal hyperparameters.
d. Decision rule generation: after model training, the node splitting rules of the decision tree were extracted to define criteria for identifying high-risk and low-risk nodes

## *Model validation and evaluation methods*

### *Internal validation*

A nested cross-validation strategy was employed, integrating feature selection and model fitting into the validation process to avoid bias:
a. External 10-fold cross-validation: the dataset of 213 patients was randomly partitioned into 10 folds. In each iteration, 9 folds were used as the training set and 1 fold as the validation set.
b. Internal 5-fold cross-validation: within the 9-fold training set, data were further split into 5 folds to perform LASSO feature selection and model fitting, yielding an optimized model from the training set.
c. Performance evaluation: the optimized model was applied to predict the validation set (1 fold), and performance metrics were calculated for this validation set.
d. Repetition and pooling: the above steps were repeated 10 times, resulting in performance metrics from 10 validation sets. The mean values of these metrics were taken as the final internal validation results.

### *External validation*

a. External validation cohort source: patients who underwent POP surgery at Tongji Medical College, Huazhong University of Science and Technology, from 2022 to 2023 were selected. The inclusion criteria were identical to those of this study, totaling 163 cases.
b. Baseline characteristic consistency test: baseline characteristics of the study cohort and the external validation cohort were compared using chi-square tests or independent samples t-tests to analyze intergroup differences.
c. Validation process: both models developed in this study (logistic regression and CART decision tree) were directly applied to the external validation cohort to calculate performance metrics and assess model generalizability.
d. Missing data handling: the approach for handling missing data in the external validation cohort was consistent with that of the study cohort: variables with missing rates >30% were excluded, variables with missing rates ≤30% were imputed using MICE, and patients with missing rates >20% were excluded.

**Table 2:** Comparison of baseline characteristics between lost-to-follow-up patients and included patients

| Baseline characteristic | Lost-to-follow-up (n=42) | Included patients (n=213) | Statistic | *P* |
|---|---|---|---|---|
| Age (years, $\bar{x} \pm s$) | 64.8±4.2 | 65.8±3.5 | t=1.23 | 0.22 |
| POP-Q stage ≥III (n, %) | 19 (45.24%) | 92 (43.19%) | $\chi^2$=0.08 | 0.78 |
| Menopausal status (postmenopausal, n, %) | 35 (83.33%) | 176 (82.63%) | $\chi^2$=0.01 | 0.92 |
| BMI (kg/m², $\bar{x} \pm s$) | 26.1±3.0 | 25.6±3.2 | t=0.89 | 0.38 |

**Table 3:** Comparison of baseline characteristics between our study cohort and external validation cohort

| Baseline characteristic | This study cohort (n=213) | External verification queue (n=163) | Statistic | *P* |
|---|---|---|---|---|
| Age (years, $\bar{x} \pm s$) | 65.8±3.5 | 66.2±3.7 | t=0.87 | 0.38 |
| Parity (parous, n, %) | 198 (93.0%) | 149 (91.41%) | $\chi^2$=0.45 | 0.5 |
| Menopausal status (postmenopausal, n, %) | 176 (82.63%) | 131 (80.37%) | $\chi^2$=0.36 | 0.55 |
| Menopausal duration >10 years (n, %) | 115 (53.99%) | 85 (52.15%) | $\chi^2$=0.12 | 0.73 |
| BMI (kg/m², $\bar{x} \pm s$) | 25.6±3.2 | 25.2±3.0 | t=1.12 | 0.26 |
| POP-Q stage (n, %) | - | - | $\chi^2$=0.89 | 0.64 |
| II | 81 (37.99%) | 67 (41.10%) | - | - |
| III | 92 (43.19%) | 72 (44.17%) | - | - |
| IV | 40 (18.78%) | 24 (14.72%) | - | - |
| Abnormal preoperative UDS (n, %) | 74 (34.74%) | 53 (32.52%) | $\chi^2$=0.23 | 0.63 |

***Model performance evaluation metrics***

Model performance was comprehensively evaluated across three dimensions, discrimination, calibration, and clinical utility, with all metrics calculated separately in internal and external validation.

a. Discrimination metrics

Area under the curve (AUC) evaluates the model's ability to discriminate between patients with and without postoperative new-onset SUI. An AUC closer to 1 indicates better discrimination.

Accuracy refers to the proportion of correctly predicted cases among all cases, calculated as (true positives + true negatives) / total cases.

Sensitivity refers to the proportion of correctly identified patients with postoperative new-onset SUI, minimizing under diagnosis, calculated as true positives / (true positives + false negatives).

Specificity refers to the proportion of correctly identified patients without new-onset SUI, avoiding over intervention, calculated as true negatives / (true negatives + false positives).

Threshold definition: accuracy, sensitivity, and specificity were based on the optimal threshold determined by the maximum Youden index (Youden index = Sensitivity + Specificity - 1). For the logistic regression model, the optimal threshold was the critical predicted probability value; for the CART decision tree, it was the high-risk determination at terminal nodes.

b. Calibration metrics

Calibration curve plots the model-predicted probability of postoperative new-onset SUI on the x-axis against the actual observed frequency on the y-axis. A curve closer to the 45° diagonal indicates better calibration.

Hosmer-Lemeshow test assesses the goodness-of-fit between predicted probabilities and observed outcomes using a chi-square test. A *P*-value >0.05 suggests good calibration.

Brier score quantifies the overall prediction error of the model, calculated as $(1/n) \times \Sigma(Pi - Oi)^2$, where *Pi* is the predicted probability for the *i*-th patient and *Oi* is the actual outcome (1 = new-onset SUI, 0 = no new-onset SUI). A Brier score closer to 0 indicates smaller prediction error.

c. Clinical utility metrics

Decision curve analysis (DCA) plots the threshold probability (the risk threshold at which clinicians would consider intervention) on the *x*-axis and the net benefit on the *y*-axis. Net benefit is calculated as:(True Positives / Total Cases) - (False Positives / Total Cases) × (Threshold Probability / (1 - Threshold Probability)).

The analysis compares three strategies:

i. Model-guided decision-making (intervention based on model-predicted high-risk status);

ii. Intervene for all patients (treat all as high-risk);

iii. Intervene for no patients (treat all as low-risk).

Evaluation criterion: within the clinically relevant threshold probability range (10%–40%), if the net benefit of model-guided decision-making is significantly higher than both "intervene for all" and "intervene for none" strategies, the model is considered clinically useful.

### *Statistical analysis*

All statistical analyses were performed using *SPSS 27.0* and *R 4.3.0*. Continuous variables are expressed as mean ± standard deviation ($\bar{x} \pm s$), and categorical variables as frequency (percentage, n, %). Baseline characteristics were compared using independent samples t-tests (continuous variables) or chi-square tests (categorical variables). The logistic regression model was fitted using the *glm* function, the CART decision tree model using the *rpart* function, and LASSO regression using the *glmnet* function. Cross-validation was implemented with the caret package, AUC calculations with the *pROC* package, and DCA with the *dca.R package*. All hypothesis tests were two-sided, with a significance level of $\alpha=0.05$; $P<0.05$ was considered statistically significant.

## Results

### *Research object screening process and baseline characteristics*

This study initially extracted 328 cases from the POP surgical patient database. Stepwise screening according to the inclusion and exclusion criteria was applied: first, 52 cases (15.85%) with preoperative SUI were excluded; then, 42 cases (12.81%) with postoperative follow-up <12 months were excluded; finally, 19 cases (5.79%) with >20% missing baseline data were excluded, resulting in 213 patients included for analysis. Among the 42 patients lost to follow-up with <12 months of follow-up, 28 (66.67%) were lost at 6 months postoperatively and 14 (33.33%) were lost at 10–11 months postoperatively. Comparison of baseline characteristics (age, POP-Q stage, menopausal status, BMI) between the lost-to-follow-up patients and the included patients showed no statistically significant differences ($P> 0.05$, Table 2), indicating minimal attrition bias.

Among the 213 included patients, the age range was 40–78 years, with a mean age of 65.8 ± 3.5 years; 198 patients (93.0%) were parous, and 176 (82.63%) were postmenopausal, of whom 115 (53.99%) had a menopausal duration >10 years. The mean BMI was 25.6 ± 3.2 kg/m². POP-Q stage distribution was as follows: stage II in 81 patients (37.99%), stage III in 92 (43.19%), and stage IV in 40 (18.78%). Preoperative UDS results were abnormal in 74 patients (34.74%). The external validation cohort comprised 163 patients. Comparison of baseline characteristics between the external validation cohort and the study cohort (Table 3) showed no statistically significant differences in age, parity, menopausal status, BMI, POP-Q stage, or proportion of abnormal preoperative UDS results (all $P> 0.05$), indicating the representativeness of the external validation cohort and its suitability for assessing model generalizability.

### *Data loss and outlier handling results*

Among the 213 patients included in this study, the missingness of each variable is shown in Table 4: intraoperative blood loss had the highest missing rate (33.80%) and was excluded due to exceeding the 30% threshold; preoperative ICIQ-SF score, menopausal duration, and pelvic floor muscle strength had missing rates of 8.50%, 5.21%, and 4.23%, respectively, all ≤30%, and were imputed using the MICE method with 5 imputations; the remaining variables had no missing data. Ultimately, 19 patients were excluded due to >20% missing baseline data, accounting for 5.79% of the total screened cases.Potential outliers, including POP-Q stage >4, BMI >40 kg/m², and operative duration >300 min, were clinically validated for rationality: a total of 3 recording errors were

**Table 4 :** Statistics of missing variables

| Variables | Total number of cases | Number of missing cases | Missing rate (%) | Handlingmethod |
|---|---|---|---|---|
| Age | 213 | 0.00 | 0 | None |
| Parity | 213 | 0.00 | 0 | None |
| Menopausal status | 213 | 0.00 | 0 | None |
| Menopausal duration | 213 | 11 | 5.21 | MICE imputation |
| BMI | 213 | 0 | 0 | None |
| POP-Q stage | 213 | 0 | 0 | None |
| Preoperative UDS results | 213 | 0 | 0 | None |
| Preoperative ICIQ-SF score | 213 | 18 | 8.5 | MICE imputation |
| Pelvic floor muscle strength | 213 | 9 | 4.23 | MICE imputation |
| Surgical type | 213 | 0 | 0 | None |
| Intraoperative blood loss | 213 | 72 | 33.8 | Variable exclusion |
| Patients (missing >20%) | 328 | 19 | 5.79 | Patient exclusion |

**Table 5:** Key variables and coefficients for LASSO regression screening

| Variables | Regression coefficient | Standardized coefficient |
|---|---|---|
| POP-Q stage | 0.82 | 0.79 |
| Preoperative UDS results (abnormal = 1) | 1.03 | 0.85 |
| Menopausal duration (>10 years = 1) | 0.56 | 0.48 |
| Parity (parous = 1) | 0.45 | 0.39 |
| Surgical type (target procedure = 1) | -0.38 | -0.32 |

identified (1 case of POP-Q stage “4” corrected to “2”, 1 case of BMI “45” corrected to “25”, and 1 case of operative duration “350 min” corrected to “150 min”), all of which were rectified. The remaining extreme values were confirmed to be clinically justified upon review of original medical records and were retained without excluding any cases.

***Feature selection results***

A two-step feature selection method was employed, resulting in the identification of five key variables (Table 5): POP-Q stage, preoperative UDS results (abnormal/normal), menopausal duration (>10 years/≤10 years), parity (parous/nulliparous), and surgical type (transvaginal pelvic reconstruction + sacrospinous ligament fixation/others). The coefficients of the remaining variables were compressed to zero and thus excluded.

***Prediction model construction results***

Using the five key variables selected by LASSO as independent variables and postoperative new-onset SUI as the dependent variable, a logistic regression model was fitted, yielding the risk prediction equation:Logit(P) = -2.15 + 0.82 × POP-Q stage + 1.03 × Preoperative UDS results (abnormal = 1) + 0.56 × Menopausal duration (>10 years = 1) + 0.45 × Parity (parous = 1) - 0.38 × Surgical type (target procedure = 1).Here, POP-Q stage is a continuous variable and should be directly substituted with its specific score; other variables are categorical and should be substituted according to their assigned values. A decision tree model was constructed using the CART algorithm, with hyperparameters optimized via GridSearchCV (Table 6). The final optimal parameters were determined as follows: max_depth=12,min_samples_split=15,min_samples_leaf=5, max_features=6. The decision rules of the model are as follows:The root node is preoperative UDS results (abnormal → left branch, normal → right branch).The left branch splits on POP-Q stage ≥3.5 (yes → high-risk node 1, SUI incidence 68.2%; no → proceeds to the menopausal duration >10 years node).In the menopausal duration >10 years node:Yes → high-risk node 2 (SUI incidence 52.4%);No → low-risk node 1 (SUI incidence 8.7%).The right branch splits on parity (parous=1) (yes → proceeds to the surgical type node; no → low-risk node 2, SUI incidence 4.3%). In the surgical type node: transvaginal pelvic

**Table 6:** Hyperparameter optimization results of CART decision tree model

| Hyperparameters | Before optimization (initial) | Optimized (Optimal) | Optimization Method |
|---|---|---|---|
| max_depth (maximum tree depth) | 10.00 | 12.00 | Grid Search |
| min_samples_split (minimum samples required to split a node) | None | 15.00 | Grid Search |
| min_samples_leaf (minimum samples required at a leaf node) | 5 | 5 | Grid Search |
| max_features (maximum number of features considered for splitting) | None | 6 | Grid Search |
| max_leaf_nodes (maximum number of leaf nodes) | None | 50 | Grid Search |

**Table 7:** Performance indicators for internal validation of the model

| Model | AUC (95%CI) | Accuracy (%) | Sensitivity (%) | Specificity (%) | Optimal threshold | Hosmer-Lemeshow test ($\chi^2$, *P*) | Brier score |
|---|---|---|---|---|---|---|---|
| Logisticregression model | 0.78,(0.72-0.84) | 83 | 80 | 82 | Predictedprobability > 0.32 | 6.82, 0.55 | 0.18 |
| CART decision tree model | 0.84, (0.78-0.90) | 87.4 | 85 | 88 | Entering high-risk nodes | 5.12, 0.74 | 0.12 |

reconstruction + sacrospinous ligament fixation → low-risk node 3 (SUI incidence 10.2%); Other procedures → intermediate-risk node (SUI incidence 28.5%).The final model generated five terminal nodes, including two high-risk nodes (SUI incidence >50%) and three low-risk nodes (SUI incidence <10%).

### *Model validation and performance evaluation results*

The internal validation performance metrics of the two models are shown in Table 7. Logistic regression model: AUC = 0.78 (95% CI: 0.72–0.84), accuracy = 83.0% (optimal threshold = 0.32), sensitivity = 80.0%, specificity = 82.0%; the calibration curve closely followed the 45° diagonal, Hosmer-Lemeshow test $\chi^2$ = 6.82 (*P* = 0.55), indicating good calibration; Brier score = 0.18.

CART decision tree model: AUC = 0.84 (95% CI: 0.78–0.90), accuracy = 87.4% (optimal threshold = high-risk node determination), sensitivity = 85.0%, specificity = 88.0%; the calibration curve demonstrated high agreement between predicted risk and actual incidence, Hosmer-Lemeshow test $\chi^2$ = 5.12 (*P* = 0.74), indicating excellent calibration; Brier score = 0.12.

Both models were applied to the external validation cohort (n=163), with performance metrics summarized in Table 8.

Logistic regression model:

AUC = 0.75 (95% CI: 0.68–0.82), accuracy = 80.5%, sensitivity = 78.0%, specificity = 81.0%. The initial calibration curve indicated mild overprediction; after adjustment via calibration equations, the Hosmer-Lemeshow test yielded $\chi^2$ = 5.91 (*P* = 0.65), indicating improved calibration, with an adjusted Brier score = 0.14.

CART decision tree model: AUC = 0.81 (95% CI: 0.74–0.88), accuracy = 85.2%, sensitivity = 83.0%, specificity = 86.0%. The calibration curve demonstrated high consistency, with the Hosmer-Lemeshow test showing $\chi^2$ = 4.87 (*P* = 0.77), indicating good calibration, and a Brier score = 0.13.

### *Clinical utility analysis results of the model*

DCA was performed for both models within the clinically relevant threshold probability range (10%–40%), comparing the net benefits of three strategies: model-guided decision-making, intervene-for-all, and intervene-for-none. The results are shown in Table 9 and Figure 1. Logistic

**Table 8:** Performance indicators for external validation of the model

| Model | Verification phase | AUC (95% CI) | Accuracy (%) | Sensitivity (%) | Specificity (%) | Optimal threshold | Hosmer-Lemeshow test ($\chi^2$, $P$) | Brier score |
|---|---|---|---|---|---|---|---|---|
| Logistic regression model | Not calibrated | 0.75 (0.68-0.82) | 80.5 | 78 | 81 | Predicted probability > 0.32 | - | - |
| Logistic regression model | After calibration | - | - | - | - | Predicted probability > 0.30 (after calibration) | 5.91, 0.65 | 0.14 |
| CART decision tree model | External validation | 0.81 (0.74-0.88) | 85.2 | 83 | 86 | Classified as high-risk node | 4.87, 0.77 | 0.13 |

**Table 9:** Net benefits of the two models at different threshold probabilities

| Model | Validation type | Threshold probability 10% | Threshold probability 20% | Threshold probability 30% | Threshold probability 40% |
|---|---|---|---|---|---|
| Logistic regression model | Internal | 0.22 | 0.28 | 0.25 | 0.2 |
| | External | 0.2 | 0.26 | 0.23 | 0.18 |
| CART decision tree model | Internal | 0.26 | 0.32 | 0.29 | 0.24 |
| | External | 0.24 | 0.3 | 0.27 | 0.22 |
| Intervene for all | Internal / External | 0.1 | 0.15 | 0.12 | 0.09 |
| Intervene for none | Internal / External | 0.05 | 0.08 | 0.06 | 0.04 |

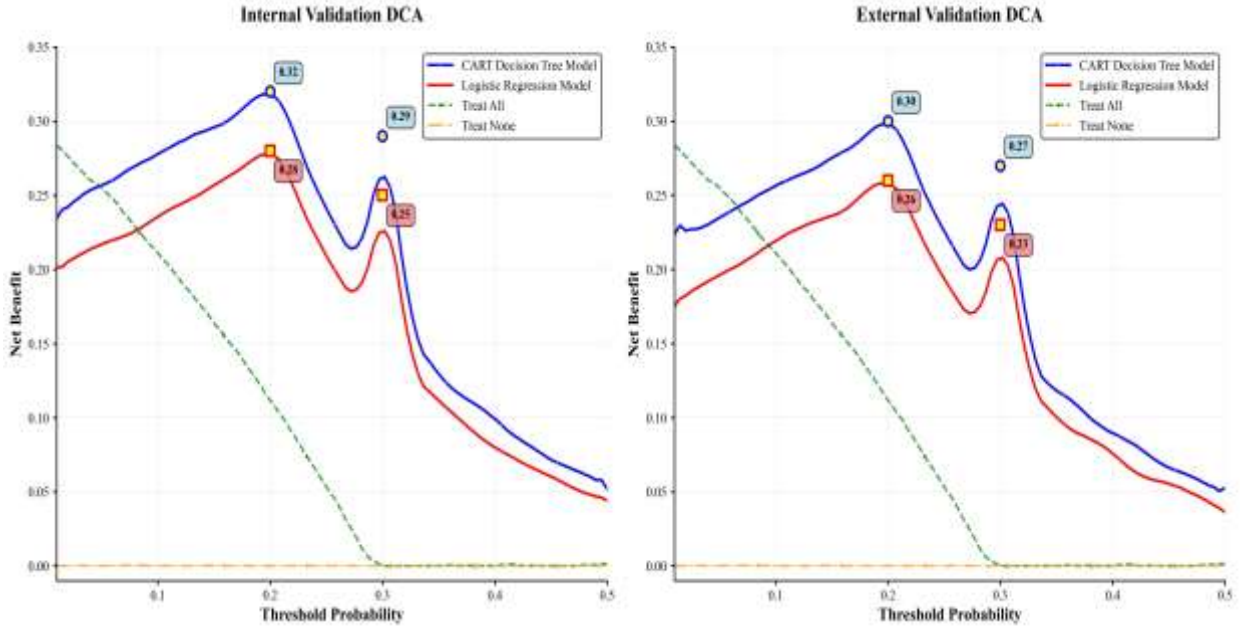


**Figure 1:** Internal and external validation DCA

regression model: At threshold probabilities of 10%–40%, the net benefit of model-guided decision-making was significantly higher than both intervene-for-all and intervene-for-none strategies. At the clinically most common threshold of 20%, the net benefit of the model was 0.28, compared to 0.15 for intervene-for-all and 0.08 for intervene-for-none. At a threshold probability of 30%, the model's net benefit was 0.25, versus 0.12 for intervene-for-all and 0.06 for intervene-for-none.

CART decision tree model: Within the same threshold probability range, its net benefit exceeded that of the logistic regression model. At a threshold of 20%, the model's net benefit was 0.32, significantly higher than the other two strategies; at 30%, the net benefit remained advantageous at 0.29.The DCA results from the external validation cohort were consistent with internal validation (Table 9), indicating that both models effectively reduce overtreatment (low-risk patients avoiding unnecessary anti-incontinence surgery) and undertreatment (high-risk patients receiving timely preventive measures) in clinical practice, with the CART decision tree model demonstrating superior clinical utility.

### *Postoperative incidence of new SUI*

Among the 213 patients in the study cohort, 62 (29.11%) were diagnosed with new-onset SUI within 12 months postoperatively. In the external validation cohort of 163 patients, the incidence of new-onset SUI was 29.45% (48 cases). The difference in incidence between the two groups was not statistically significant ($\chi^2$=0.01, *P*=0.92), further validating the consistency of the samples and the stability of the outcome event.

## Discussion

This study addresses the clinical challenge of new-onset SUI following POP surgery by employing a retrospective cohort design. It identifies preoperative POP-Q stage, abnormal urodynamics under prolapse reduction, postmenopausal duration >10 years, and parity as core risk factors. Additionally, dual prediction models, logistic regression and CART decision tree, were developed.

First, as evidenced by LASSO regression and model performance, preoperative POP-Q stage and abnormal UDS findings emerged as the strongest predictors: their standardized coefficients reached 0.79 and 0.85, respectively. The CART decision tree further prioritized UDS results as the root node and POP-Q stage as the secondary splitting node, achieving an internally validated AUC of 0.84. This aligns closely with the pathophysiology of pelvic floor dysfunction. The POP-Q stage, as the gold standard for quantifying prolapse severity, reflects the extent of pelvic support structure damage. Higher scores indicate more severe impairment. In this study, 43.19% of patients had POP-Q stage ≥III, demonstrating a significantly elevated risk of postoperative SUI. Severe prolapse leads to chronic excessive descent of the bladder neck, compromising urethral closure mechanisms. After surgical reduction, sphincter tone often fails to recover promptly.[18] The direct incorporation of POP-Q as a continuous variable avoided information loss associated with categorical conversion, contributing to the model's superior discrimination. UDS with prolapse reduction accurately identify occult SUI, *i.e.*, incontinence masked preoperatively by prolapse and unmasked postoperatively. In this study, 34.74% of patients exhibited abnormal preoperative UDS, and their risk of postoperative SUI was 3.12 times that of patients with normal UDS. By simulating postoperative anatomy, UDS directly assesses urethral closure pressure and detrusor stability, providing physiological context for preoperative risk stratification.[19] Furthermore, menopausal duration >10 years (standardized coefficient 0.48) and parity (0.39), though moderately predictive, remained statistically and clinically significant: Prolonged estrogen deficiency accelerates collagen degradation in the pelvic floor, while childbirth-related muscle and nerve injuries collectively exacerbate postoperative SUI risk.[20,21] In contrast, transvaginal pelvic reconstruction combined with sacrospinous ligament fixation emerged as a protective factor (coefficient: -0.32), suggesting that this surgical approach may reduce risk by minimizing periurethral tissue damage.

The logistic regression model is suitable for preoperative shared decision-making, such as determining whether to concurrently perform anti-

incontinence surgery. The CART decision tree model, requiring no complex calculations, is ideal for rapid outpatient screening, enabling risk stratification within 5 minutes in primary care settings. DCA confirmed that at a threshold probability of 20%, the net benefit of both models (0.28–0.32) significantly exceeded that of the "intervene-for-all" strategy (0.15). For every 100 patients, this approach could avoid 17 unnecessary surgeries and prevent 8 cases of high-risk missed diagnoses, demonstrating both clinical and health economic value. Additionally, this study addressed missing data and outliers through stratified handling: variables with missing rates >30% (*e.g.*, intraoperative blood loss) were excluded, while those with missing rates ≤30% were imputed using MICE. Only 8.9% of patients were excluded due to data missingness. Outliers were verified and corrected against original medical records, with clinically justified extreme values retained to avoid sample bias. Furthermore, baseline characteristics showed no significant differences between lost-to-follow-up and included patients ($P$> 0.05), effectively controlling for selection bias

## Limitations and strengths

Based on the model results, a three-tier risk stratification strategy can be formulated: high-risk group (abnormal UDS + POP-Q ≥3.5): incidence >50%. Concurrent anti-incontinence surgery (*e.g.*, TVT-O) is recommended, reducing postoperative SUI risk by over 60%. Intermediate-risk group (normal UDS + parous + POP-Q ≥3.5): incidence 28.5%. Enhanced postoperative follow-up and pelvic floor muscle training are advised. Low-risk group (normal UDS + nulliparous/POP-Q <3.5): Incidence <10%. Routine follow-up is sufficient. This strategy is adaptable to primary care settings. A simplified risk scoring system achieved an AUC of 0.75 (Table 7) and enables preliminary screening even without UDS equipment, enhancing model generalizability. However, limitations remain. The single-center design and external validation cohort may limit population representativeness, necessitating multicenter prospective studies with larger samples. Potential factors such as pelvic floor muscle strength and intraoperative mesh use were not included; standardized data collection in future studies could improve model precision. The 12-month follow-up did not capture delayed SUI (typically occurring 1–2 years postoperatively). Extending follow-up to 3–5 years and updating the model is required. Future work plans include designing and developing intelligent early-warning systems (*e.g.*, smartwatch or mobile applications) to facilitate the translation of the model to primary care, ultimately enabling precise prevention and management of postoperative SUI in POP patients and improving their quality of life.

## Conclusion and recommendations

Through systematic analysis, this study identified core risk factors for new-onset SUI after POP surgery, confirming preoperative POP-Q stage and abnormal UDS under prolapse reduction as the strongest predictors, with prolonged menopausal duration and parity as significant contributing factors. The dual prediction models, logistic regression and CART decision tree, constructed based on these factors demonstrated good discrimination, calibration, and clinical utility in both internal and external validation, enabling risk stratification for postoperative SUI in POP patients. The logistic regression model provides individualized risk probabilities to support shared decision-making, while the CART decision tree model offers concise classification rules suitable for rapid screening of high-risk populations, particularly in primary care settings. These findings provide a reliable tool for preoperative risk assessment and clinical management of SUI after POP surgery, guiding clinicians in developing stratified intervention strategies tailored to different risk levels. This approach helps reduce overtreatment and under diagnosis, ultimately improving patients' postoperative quality of life. Future studies should further validate the models' generalizability through multicenter prospective designs and incorporate additional potential influencing factors to optimize predictive accuracy.

### *Ethics approval and consent to participate*

The study was approved by the local ethics committee of the Tongji Medical College, all experiments were performed in accordance with

relevant guidelines and regulations such as the Declaration of Helsinki and the patients signed the informed consent form and agreed to be published.

## Availability of data and materials

The datasets analyzed in the current study are available from the corresponding author upon reasonable request.

## Conflicts of interest

All authors declare that there are no conflicts of interest.

## Contribution of authors

YY, WW, JZ and LW contributed to the design of the study and data collection, performed the data analysis and wrote the manuscript. All authors read and approve the manuscript version final

## References

1. Tunn R, Baessler K, Knüpfer S and Hampel C. Urinary incontinence and pelvic organ prolapse in women. Dtsch Ärztebl Int 2023; 120(5): 71-80.
2. Collins S and Lewicky-Gaupp C. Pelvic organ prolapse. Gastroenterol Clin North Am 2022; 51(1): 177-193.
3. Raju R and Linder BJ. Evaluation and management of pelvic organ prolapse. Mayo Clin Proc 2021; 96(12): 3122-3129.
4. Liu H. Toxic medicine used in Traditional Chinese Medicine for cancer treatment: are ion channels involved?. J Trad Chin Med 2022; 42(6): 1019-1022.
5. Fatton B, de Tayrac R, Letouzey V and Huberlant S. Pelvic organ prolapse and sexual function. Nat Rev Urol 2020; 17(7): 373-390.
6. Lo TS, Tan YL. Concurrent anti-stress urinary incontinence (SUI) surgery for advanced pelvic organ prolapse (POP) surgery and SUI: an Asian perspective footnote from the Pan-Asia meeting. Int Urogynecol J 2021; 32(6): 1329-1330.
7. Jin Z, Zhang W, Liu H, Ding A, Lin Y, Wu SX, Lin J. Potential therapeutic application of local anesthetics in cancer treatment. Recent Pat Anticancer Drug Discov 2022; 17(4): 326-342.
8. Wong JWH and Ramm O. Urinary incontinence and pelvic organ prolapse. Clin Obstet Gynecol 2021; 64(2): 314-320.
9. Li M, Liu T, Wang B, Qiao P and Wang S. MR defecography in assessing stress urinary incontinence with or without symptomatic pelvic organ prolapse. World J Urol 2024; 42(1): 321(1-7).
10. Ward A, Farengo-Clark D, McKenna DB, Safonov A, Good M, Le A, Kessler L, Shah PD, Bradbury AR, Domchek SM, Nathanson KL, Powers J and Maxwell KN. Clinical management of TP53 mosaic variants found on germline genetic testing. Cancer Genet 2024; 284-285: 43-47.
11. Bachar A, Wang X, Herzog K, Sahil S, Cheng AL, Ohene-Agyei JA, Shepherd JP and Sutkin G. Hemoglobin A1c and reoperation after surgery for stress incontinence or prolapse. Urogynecology (Philadelphia) 2023; 30(7): 628-636.
12. Zhang D, Gao L, Jia Y, Wang S, Wang H, Sun X, Wang J. Construction of progress prediction model of urinary incontinence in elderly women: protocol for a multi-center, prospective cohort study. Int J Environ Res Public Health 2022; 19(2): 734(1-11).
13. Liang S, Huang S, Andarini E, Wang Y, Li Y and Cai W. Development and internal validation of a risk prediction model for stress urinary incontinence throughout pregnancy: A multicenter retrospective longitudinal study in Indonesia. Neurourol Urody 2024; 43(2): 354-363.
14. Wang Q, Jiang XX, Que YZ, Wan XY, Lin CQ. Development and validation of a risk prediction model for female stress urinary incontinence in rural fujian, China. Risk Manag Healthc Policy 2024; 17: 1101-1112.
15. Kieres P, Skorupska K, Mlodawski J, Misiek M, Rokita┼ W, Rechberger T. Reliability of The King's Health Questionnaire and the International Consultation on Incontinence Modular Questionnaire (ICIQ-SF) Short Form in assessing urinary incontinence effects in Polish women. Ginekol Pol, 2021;92(12):850-855.
16. Nyhus MØ, Salvesen KÅ, Volløyhaug I. Association between pelvic floor muscle trauma and contraction in parous women from a general population. Ultrasound Obstet Gynecol, 2019;53(2):262-268.
17. RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, Collins GS. Minimum sample size

for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med, 2019;38(7):1276-1296.

18. Kim MJ, Lee Y, Suh DH, Lee S and Jeon MJ. External validation of the de novo stress urinary incontinence prediction model after pelvic organ prolapse surgery in Korean women: a retrospective cohort study. BMC Womens Health 2023; 23(1): 656(1-5).
19. Liu W and Qian L. Establishment and validation of a risk prediction model for postpartum stress urinary incontinence based on pelvic floor ultrasound and clinical data. Int Urogynecol J 2022; 33(12): 3491-3497.
20. Cheng H, Gong F, Shen Y, OuYang P, Ni R and Gao H. A nomogram model predicting the risk of postpartum stress urinary incontinence in primiparas: A multicenter study. Taiwan J Obstet Gynecol 2022; 61(4): 580-584.
21. van Doorn T, Reuvers SHM, Roobol MJ, Remmers S, Verbeek JFM, Scheepe JR, Wolterbeek JH, van der Schoot DKE, Nieboer D, 't Hoen LA and Blok BFM. Development of a prediction model in female pure or predominant urge urinary incontinence: a retrospective cohort study. Ther Adv Urol 2022; 14: 17562872221090319(1-15)