Yu et al. BMC Women's Health (2025) 25:55
https://doi.org/10.1186/s12905-025-03584-8

BMC Women's Health

RESEARCH

Open Access

Risk prediction models for stress urinary incontinence after pelvic organ prolapse (POP) surgery: a systematic review and meta-analysis

Bi Jun Yu[1], Hao Chong He[4], Li Wang[2], Han Mei Shao[3], Ying Min Liu[3], Xiao Ying Yan[1] and Jian Liu[3*]

Abstract

Objective To systematically evaluate existing developed and validated predictive models for stress urinary incontinence after pelvic floor reconstruction.

Methods Relevant literature in PubMed, Embase, Web of Science, Cochrane Library, OVID, China National Knowledge Infrastructure(CNKI), Wan Fang Database, VIP database and Chinese Biomedical Literature Service System (SinoMed) were search from inception to 1 March 2024. Literature screening and data extraction were performed independently by two researchers. The chosen study's statistics included study design, data sources, outcome definitions, sample size, predictors, model development, and performance. The Predictive Modelling Risk of Bias Assessment Tool (PROBAST) checklist was used to assess risk of bias and applicability.

Results A total of 7 studies containing 9 predictive models were included. All studies had a high risk of bias, primarily due to retrospective design, small sample sizes, single-center trials, lack of blinding, and missing data reporting. The meta-analysis revealed moderate heterogeneity ($I^2 = 68.8\%$). The pooled AUC value of the validated models was 0.72 (95% CI: 0.65, 0.79), indicating moderate predictive ability.

Conclusion The prediction models evaluated demonstrated moderate discrimination, but significant bias and methodological flaws. The meta-analysis revealed moderate heterogeneity ($I^2 = 68.8\%$) among the included studies, reflecting differences in study populations, predictors, and methods, which limits the generalizability of the findings. Despite these challenges, these models highlight the potential to identify high-risk patients for targeted interventions to improve surgical outcomes and reduce postoperative complications. The findings suggest that by integrating these models into clinical decision-making, clinicians can better tailor surgical plans and preoperative counseling, thereby improving patient satisfaction and reducing the incidence of postoperative stress urinary incontinence. Future research should follow TRIPOD and PROBAST principles, focus on addressing sources of heterogeneity, improve model development through robust designs, large sample sizes, comprehensive predictors, and novel modelling

*Correspondence:
Jian Liu
liujiantougao@163.com

Full list of author information is available at the end of the article

approaches, and validate tools that can be effectively integrated into clinical decision-making to manage stress urinary incontinence after pelvic floor reconstruction.

Keywords Clinical predictors, Pelvic floor reconstruction, Prediction model, Stress urinary incontinence, Systematical analysis

Background

Pelvic organ prolapse (POP) is one of the pelvic floor dysfunctional diseases (PFD). An epidemiological survey from the International Urogynecological Association showed that the prevalence of POP varied widely, ranging from 1–65% [1]. Surgical repair is the primary management strategy; however, 8–55% of patients develop persistent or de novo stress urinary incontinence (de novo SUI) following surgery [2, 3]. This variability underscores the complexity of predicting postoperative SUI risk, which is influenced by diverse surgical approaches and patient-specific factors [4, 5].

Moreover, it remains controversial whether concomitant anti-incontinence surgery is necessary during pelvic organ prolapse repair. Simultaneous surgery reduces the overall risk of anesthesia, shortens the total recovery time, and facilitates urinary incontinence management. However, it may increase operative time, risk of surgical complications, and hospital stays. Furthermore, several studies have shown that outcomes of SUI surgery performed together with POP repair remain uncertain [2, 6, 7]. There is also concern about the use of mesh repair in stress urinary incontinence surgery, which leads to more difficult surgical decisions. This necessitates individual assessment of patients by a specialist clinician [2, 8]. Inadequate assessment or decision-making may lead to negative patient outcomes, such as anxiety and embarrassment due to persistent or new postoperative incontinence. Patients who have undergone inappropriate mesh repair may also develop a series of postoperative complications, such as overactive bladder, obstructive micturition, mesh erosion, infection, dyspareunia, and chronic pelvic pain [9, 10]. This not only reduces patient satisfaction with surgical treatment but also increases subsequent medical costs and affects patients' quality of life [11]. Therefore, it has become a trend in the field to correctly assess the risk factors for SUI in postoperative POP patients and to construct predictive models with good predictive properties. As we all know, age and BMI have been key factors in this model [12], but have shown inconsistencies in different populations. For example, age-related hormonal changes that affect pelvic floor function, especially around menopause, may vary depending on dietary, genetic, and environmental factors [13, 14]. Similarly, the association between BMI and SUI is complex. Although higher BMI increases abdominal pressure and pelvic floor strain, its effects are attenuated by differences in body composition, fat distribution and even access to weight management interventions in different health care systems [15, 16].

In addition, most previous models are based only on risk factor analysis. They are inadequate for internal validation or external validation, which leads to the risk of overfitting the models while not being generalizable and clinically useful [17]. Moreover, with the increasing use of new surgical approaches and the utilization of new model-building methods (e.g., machine learning), it seems that the previous traditional logistic regression methods have become more difficult to explain the complex nonlinear relationships among the influencing factors [18]. With the increasing use of new surgical techniques and advancements in machine learning, there is a growing need to systematically evaluate existing models and develop more accurate and clinically relevant predictive tools. This study aims to fill this gap by critically assessing the development, validation, and performance of current predictive models for SUI after pelvic floor prolapse surgery.

Accordingly, the objective of this systematic review and meta-analysis is to critically evaluate the development, validation, and performance of predictive models for stress urinary incontinence following pelvic floor prolapse surgery, with the aim of providing insights for future improvements in model construction and clinical application.

Methods

We reported this review according to the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) [19]. This systematic review has been registered on the International Prospective Register of Systematic Reviews (PROSPERO) database (CRD42024519370).

Data sources and searches

Our systematic review and meta-analysis focused on two languages, English and Chinese, which may introduce potential selection bias. We systematically searched five English databases: PubMed, Embase, the Cochrane Library, OVID and Embase; and four Chinese databases: China National Knowledge Infrastructure (CNKI), Wan Fang Database, VIP database and Chinese Biomedical Literature Service System (SinoMed) from inception to March 1, 2024. Three groups of terms were combined according to the syntax rules for each database: [1] pelvic organ prolapse-related terms, including Urogynecology, urogynecological surger, vaginal surgery, pelvic organ prolapse, colporrhaphy, pelvic reconstructive surgery, pelvic reconstruct*; [2] SUI-related terms, including De Novo Stress Urinary Incontinence [2], Stress Urinary Incontinence, SUI, De Novo SUI; and [3] prediction-related terms, including model*, prediction model, prognostic model, prognos*, risk factor*, risk.We also identified additional relevant studies by reviewing the reference lists of the retrieved studies and reviews.

For the systematic review, we used the PICOTS system recommended by the Critical Appraisal and Data Extraction for Systematic Reviews of Predictive Modelling Studies (CHARMS) checklist [20]. This system helps to identify the purpose of the review, the search strategy, and the criteria for study inclusion and exclusion [19]. The key items of our systematic review are described as follows:

P (Population): patients who developed stress urinary incontinence after pelvic floor repair.

I (Intervention model): developed and validated risk prediction model for stress urinary incontinence after pelvic floor repair (predictor ≥ 2).

C (Comparator): no competing model.

O (Outcome): outcome focused on the incidence of postoperative stress urinary incontinence.

T (Time): stress urinary incontinence in the postoperative period (≥ 3 months postoperatively).

S (Setting): The intended use of the risk prediction model is to individualize the prediction of the presence of postoperative stress urinary incontinence, thereby assisting in the clinical medical diagnosis and reducing the incidence of adverse events.

Inclusion and exclusion criteria

The following criteria were used for inclusion: (a) Women aged ≥ 18 years; (b) Types of studies: cohort studies, case-control studies and cross-sectional studies; (c) Preoperative with or without urinary incontinence; (d) POP-Q staging > stage II; (e) Postoperative symptoms of SUI (meeting the ICS definition of SUI); (f) Performance of at least one model (e.g., Receiver Operating Characteristic Curve(ROC), sensitivity, specificity); (g)Studies in Chinese and English.

Studies were excluded if (a) only risk factors were analyzed without validation of the predictive model; (b) Complete datacould not be obtained, and the original text could not be retrieved by contacting the authors; (c) Review studies, conference papers and animal experiments; (d) Models containing < 2 predictor variables (To prevent an increase in MSE and greater bias in parameter estimates [22]); (e) Repeated publications; (g) Inability to contract pelvic floor muscles (PFM) and prior pelvic floor physical therapy were not considered exclusion criteria.

Data collection progress and items

Two researchers initially screened the studies in electronic databases based on keywords and removed duplicate studies. They independently screened titles and abstracts based on the inclusion and exclusion criteria to identify studies on SUI prediction models after POP. Disagreements were resolved in consultation with a third researcher after reading the full text.

Following the selection of all relevant studies, three researchers independently extracted the data using a Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies [20]. In case of disagreement, the review team discussed whether to include these data. The finalized excerpts included: (i) Basic information (e.g., first author, year of publication, country, study population, outcome indicators, data source); and (ii) Model information (e.g., number of predictors, model type, sample size, predictor screening method, final predictors, missing data handling, modeling method, model validation method, calibration method, performance metrics, model presentation type).

Risk of bias and transparent reporting

Three researchers used the Prediction Model Risk of Bias Assessment Tool (PROBAST) independently to evaluate the risk of bias and make the final decision for evaluated difference [23], the assessment of risk of bias in the Prediction Model Risk of Bias Assessment Tool includes four domains (participants, predictors, outcome and analysis), whereas applicability includes three domains (participants, predictors and outcome). The answer to each question is “yes”, “probably yes”, “no information”, “no” or “probably no”, and the risk of bias and applicability of each domain is judged as low, high, or unclear. When all domains in a study were considered to be low, the overall risk of bias was low; when one or more domains were judged to be high, the entire risk of bias was high. If one or more domains were unclear and all other domains were low, the study was considered an unclear risk of bia. Lack of external validation, improper handling of missing data, insufficient sample size (sample size/candidate variable < 20), improper selection and handling of predictors can all affect the reliability and generalizability of the model. For example, if a study has not been externally validated, this limits their applicability in different populations. The lack of external validation means that these models may be over-fit to the original study population and may therefore perform poorly in other clinical settings. This shortcoming significantly affected the overall assessment of the quality of these studies and increased the risk of bias [24].

Data synthesis and statistical analysis

A meta-analysis of the area under the curve (AUC) values from the validated models was conducted using Stata software (version 16.0; Stata Corporation, College Station, Texas, USA). Heterogeneity was tested using the I^2 index. The I^2 index provides a measure of heterogeneity, with values of 25%, 50%, and 75% indicating low, moderate, and high heterogeneity, respectively [25]. A fixed-effects or random-effects model was used according to the heterogeneity of the analysed results. Sensitivity analyses and Egger's test were respectively used to analyse the sources of heterogeneity and to determine publication bias [26, 27], with $P > 0.05$ indicating a low likelihood of publication bias.

Results

Research result

The search resulted in 5243 relevant documents. After eliminating duplicates, 1978 papers remained. After screening, 1,922 papers were excluded based on titles and abstracts: 1,387 papers did not correspond to the topic, 197 only examined risk factors, 64 were informal papers, 94 were case reports, 121 were animal experiments, and 79 were reviews; after further full text screening, 14 papers were only modelled, 4 did not correspond to the topic of the study, and 3 were unavailable for full text. Ultimately, seven studies were included. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flowchart depicts the comprehensive search process and results, as shown in Fig. 1.

Fundamental characteristics of the included literature

Seven papers were finally included [28–34], of which four were published in China, one in the United States, one in the Netherlands, and one in Korea. All seven studies were all retrospective and single-center. Four studies included patients with preoperative SUI diagnosis. Two used SUI incidence at three months post-operation as the outcome indicator, two at six months, and three at one year. Postoperative SUI incidence ranged from 10 to 33.75%. Detailed characteristics are provided in Table 1.

Modeling development and performance

For modelling methods of the seven pelvic floor reconstruction studies included [28–34], all used logistic regression to construct their models, with one study [32] additionally used machine learning methods (Random Forest and Extreme Gradient Boosting [XGBoost]).

In term of predictors, the most common predictor among the nine predictive models was age, which appeared in seven predictive models; other common predictors were body mass index (BMI), POP-Q anterior vaginal wall points (Aa, Ba) and anti-incontinence surgery, each appearing in four models, and the number of vaginal deliveries, appearing in three models.

Regarding performance, all models were internally validated, six of them were externally validated, and three models lacked external validation. Receiver Operating Characteristic Curve (ROC) was the most commonly used assessment method, with reported AUC values ranging from 0.595 to 0.806. Three models reported calibration curves, showing good fit between the model-predicted values and actual values, and only three models reported the Youden's index, sensitivity, and specificity. Details of the models are shown in Table 2.

Risk of quality assessment

In the risk of bias evaluation, all of the seven studies [28–34] were rated as high risk due to the higher risk of bias in the model building and validation process of the retrospective studies, and four studies [29, 31, 32, 34] enrolled patients who had already had SUI preoperatively, which may have led to overestimation of the predictive performance of the model. Two studies [31, 32] reported a lack of index measurements in some subjects, which may lead to a less representative sample as well as an increased risk of bias.

In terms of predictors, only two studies [29, 32] used blinding of researchers during predictor selection, which avoided the influence of subjective factors of researchers on the predictive ability of the model and reduced the generation of inclusion bias.

Regarding statistical analysis, the risk of bias was high in all seven studies due to the following reasons: (i) Insufficient sample size. Three studies [30, 33, 34] had small sample sizes (sample size/candidate variable < 20). (ii) Treatment of missing values: two studies [29, 32] reported the exclusion of subjects with missing values, and four studies [28, 30, 33, 34] performed multiple interpolation for missing values. Four studies [28, 29, 31, 32] did not explicitly report missing rates, and three of them [30, 33, 34] did not report the treatment of missing values or complete data; (iii) Treatment of variables: two studies [29, 32] dichotomized continuous variables. (iv) Regarding the selection of variables, four studies [30, 32–34] reported a screening strategy using univariate variables. (v) Regarding model performance assessment. Three studies [29, 30, 33] did not perform external validation and four studies [28, 29, 31, 33] did not report specific calibration values. In terms of model applicability, all studies had good applicability across domains and overall. See Table 3 for details.

Meta-analysis of validation models included in the review

For all studies, a meta-analysis was performed. Notably, Fu's study [32] involved multiple model development methods, all based on the same sample, hence only the

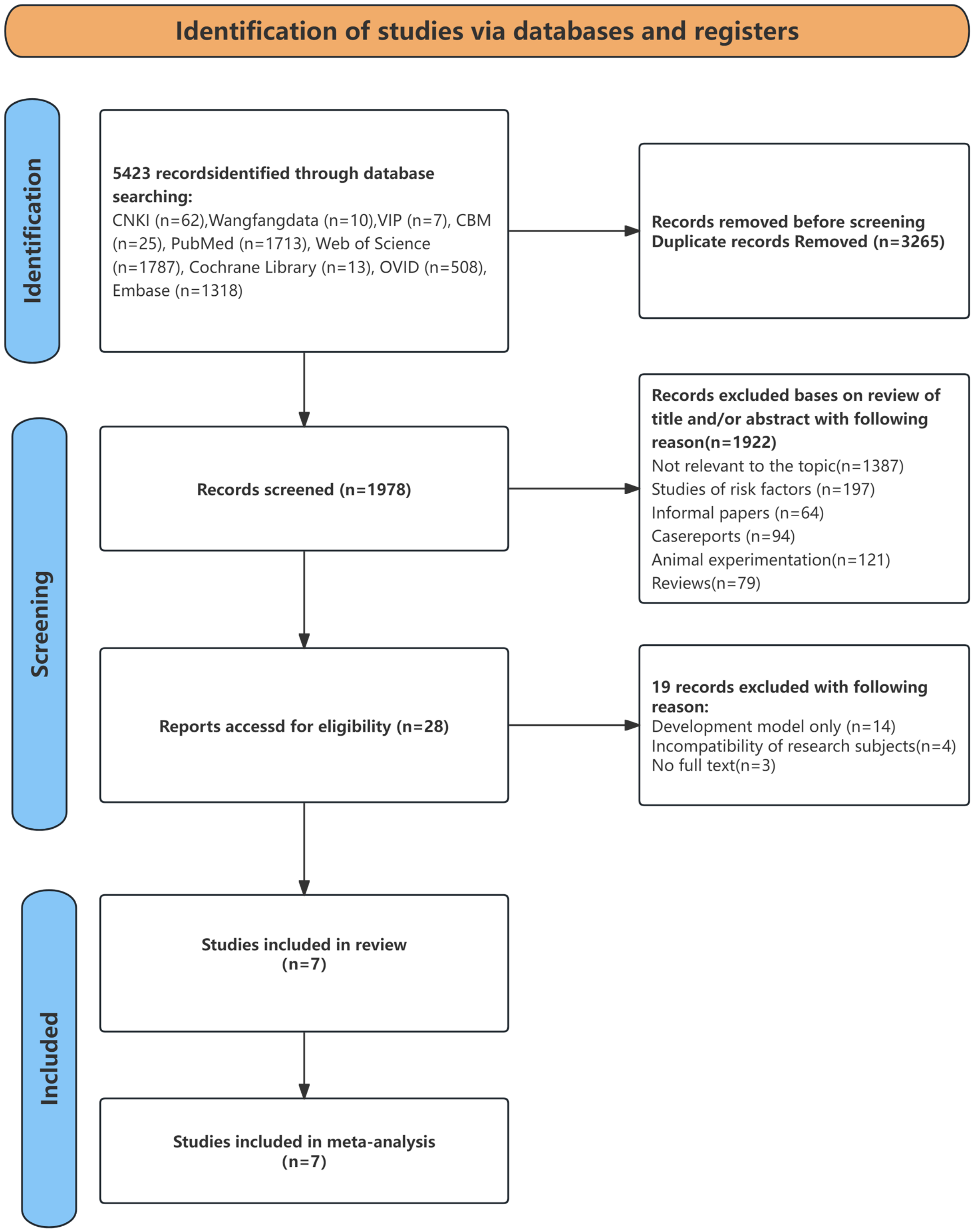

Fig. 1 Search flow diagram

model developed using logistic regression was included. The pooled AUC value calculated using the random-effects model was 0.72 (95%CI: 0.65–0.79) (Fig. 2). The I^2 value was 68.8% ($p < 0.05$), indicating moderate heterogeneity. Subsequently, sensitivity analyses showed that the exclusion of individual studies had minimal impact, indicating stable results (Fig. 3). The Egger's test showed no significant publication bias ($P = 0.71$).

Discussion

In this study, we summarized seven studies [15–21] comprising nine models. The modeling AUC ranged from 0.595 to 0.842, with seven models exceeding an AUC of 0.7, indicating good predictive performance. The pooled AUC value was 0.72 (95% CI: 0.65–0.79), showing that the models could predict SUI occurrence relatively well. However, the moderate to high heterogeneity observed

Table 1 Basic characteristics of included studies

Study	Country	research design	Participants	preoperative SUI included	Outcome index	Data sources	SUI cases/sample size (%)
Jelovsek 2014	USA	Retrospective	Patients undergoing transvaginal repair of pelvic organ prolapse	No	3 months	Clinical trial-related data on outcomes after vaginal prolapse repair and midurethral sling surgery	115/457(25.16%)
Ploeg 2019	Netherlands	Retrospective	All patients undergoing vaginal prolapse repair	Yes	1 year	Clinically relevant data from two randomized controlled trials (CUPIDO-1 and CUPIDO-2)	61/356(17.13%)
Ding 2021	China	Historical Cohort	Transvaginal total pelvic suspension in patients with organ prolapse	Yes	3 months	Shanghai First People's Hospital, China	40/129(31%)
Oh 2022	Korea	Historical Cohort	All patients undergoing prolapse surgery	Yes	1year	Seoul National University Hospital (SNUH) Seoul National University Bundang Hospital (SNUBH), Korea	114/1142(10%)
Fu 2023	China	Historical Cohort	Patients undergoing various pelvic floor surgeries	Yes	6 months	Peking Union Medical College Hospital (PUMC)	116/555(20.90%)
Zhu 2023	China	Retrospective	Transvaginal mesh repair patients	No	6 months	Henan Yellow River Sanmenxia hospital, China	68/224(30%)
Li 2024	China	Historical Cohort	Patients undergoing laparoscopic pelvic floor reconstruction surgery	Yes	6 months	The Second Hospital of Shandong University, China	53/157(33.75%)

3months: SUI symptoms at 3 months postoperatively; 6months: SUI symptoms at 6 months postoperatively;1years: SUI symptoms at 1 years postoperatively

among the studies (I^2 = 68.8%) highlights significant variability in model performance, which may be attributed to differences in study populations, predictors used, and methodological approaches, as confirmed by sensitivity analyses and publication bias tests.

During the model evaluation, several articles were found to not fully comply with the Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement [35], leading transparency issues and potential risk of bias. All the studies included were retrospective, and the predictors incorporated during the model-building process may not be comprehensive, which may lead to incomplete data and potentially biased results [32]. Additionally, retrospective studies rely on existing data, and researchers may not be able to control for potential confounding variables—factors that might influence the outcome but are not measured or accounted for in the analysis. Some historical data may be incomplete or missing, leading to gaps in the analysis or bias due to nonrandom missing data, which may affect the robustness and reliability of the findings [36]. Three studies [30, 33, 34] reported small sample sizes, which may result in fewer participants and greater variability and non-representativeness of results due to sampling error, making it difficult to generalize findings to larger populations. Small sample studies also run the risk that the models or statistical methods used may overfit the data [37]. All seven studies [28–34] were single-center studies, which limit the diversity of the samples and the generalizability of the findings. They may not be representative of the wider patient population [38, 39]. Only two studies [29, 32] employed blinding in the predictor selection process, which could increase the risk of overfitting and reduce the credibility of the results. Four studies [28, 30, 33, 34] did not clearly report the rate of missing data, and three of these [30, 33, 34] did not report methods for handling missing values or complete data. Four studies [30, 32–34] reported using univariable variable selection strategies. Such strategies are relatively simple but may be less accurate in capturing the more complex non-linear relationships between predictor variables and predicted outcomes, thus missing important risk factors. Two studies [31, 32] reported the use of randomized split validation, which reduces the actual sample size for model development and increases the risk of overfitting. One study [33] reported only internal validation discrimination and did not report modelling ROC, casting doubt on the transparency of their model in the building process and having selective reporting, which introduces some bias.

Notably, three studies [29, 30, 33] did not perform external validation. External validation is particularly crucial for predictive models in the clinical domain due to the inherent variability in patient populations, surgical techniques, and clinical settings. The absence of external validation significantly undermines the generalizability of these models, as it limits their ability to be reliably applied across different healthcare environments and patient cohorts [40]. Without external validation, models may overfit to the specific characteristics of the original

Table 2 Incorporating developments in the predictive model

Study	Continuous variable handling methods	Methods of screening Variables	Variables	Missing data(%) handling	Develop methods	Validation Methods	Calibration method	Performance indicators	Other calibrated metrics	Presentation
Jelovsek 2014	Maintain continuity	Stepwise	Age, number of vaginal deliveries, BMI, PPDT, anti-incontinence surgery, urinary urgency, diabetes	Multiple interpolation (-)	LR	Bootstrap/Domain Validation	ROC	A:0.73 (0.65,0.80) B:0.73/0.62(0.56, 0.69)	Calibration curve	Forest plot and model formula
Ploeg 2019	Biclassification	Stepwise	Age (< 55 years), number of vaginal deliveries (≥ 3), Ba, subjective UI, anti-incontinence surgery	Multiple interpolation (26%)	LR	Bootstrap/-	ROC	B:0.77(0.71,0.83)/-	-	Scoring tables and risk formulas
Ding 2021	Maintain continuity	Univariate and multivariate	Ultrasound residual urine, luteinizing hormone, triglycerides	-	LR	-/-	ROC	A:0.738 (0.625,0.850) B:0.75(0.64,0.86)/-	Youden's index 0.475 Sensitivity 0.775 specificity is 0.7	Forest plot and risk score presentation
Oh 2022	Maintain continuity	LR, exhaustive, stepwise	Age (≥ 55 years), diabetes, subjective UI, decompression test after prolapse repositioning, type of prolapse surgery, anti-incontinence surgery	Multiple interpolation (26%)	LR	5-fold cross validation, leave-one-out cross validation, Randomized Split External Verification	ROC	A:0.78 (0.67,0.89) B:0.76 /0.74 (0.62,0.86)	Calibration curve	Forest plot
Fu 2023	Biclassification	LR, RF	BMI, C, Age, Aa, TVM	Multiple interpolation(10%)	XGBoost	50% off nested cross validation, Randomized Split External Verification	Spiegelhalter Z($P > 0.05$), MAE, ROC	A:0.714 (0.658,0.770) B:0.721/0.704 (0.588,0.820)	Youden's index 0.207 Sensitivity: 0.783 Specificity: 0.598 Accuracy: 0.636	Total Gain
Study	Continuous variable handling methods	Methods of screening Variables	Variables	Missing data(%) handling	Develop methods	Validation methods	Calibration method	Performance indicators	Other calibrated metrics	Presentation
Fu 2023	Biclassification	LR, univariate and multivariate	Age (Jordon's Index Split), Total/Half Vaginal Closure Surgery and TVM	Multiple interpolation (10%)	LR	50% off nested cross validation, Randomized Split External Verification	Spiegelhalter Z($P > 0.05$), MAE, ROC	A:0.5950 (0.532,0.657) B:0.631 /0.593 (0.472,0.715)	-	-
	Biclassification	LR, RF	BMI, age, C, residual urine output, Ba	Multiple interpolation (10%)	RF	50% off nested cross validation, Randomized Split External Verification	Spiegelhalter Z($P < 0.001$), MAE, ROC	A:0.842 (0.798,0.887) B:0.648/0.603 (0.485,0.721)	-	Gini Index

Table 2 (continued)

Study	Continuous variable handling methods	Methods of screening Variables	Variables	Missing data(%) handling	Develop methods	Validation Methods	Calibration method	Performance indicators	Other calibrated metrics	Presentation
Zhu 2023	Maintain continuity	LR, univariate and multivariate	Age ≥ 50y, ≥3 pregnancies, ≥3 births, history of major labor, history of chronic respiratory disease, vaginal delivery, perineal laceration, biofeedback stimulation	-	LR	Bootstrap/-	ROC, HL	B:0.841 (0.703,0.979) /-	Calibration curve	Scoring model based on regression coefficient β
Li 2024	Maintain continuity	LR, univariate and multivariate	BMI, age, Ba	-	LR	CV, Bootstrap/time verification	ROC	A:0.806 (0.731,0.881) B:0.807 /0.757 (0.606,0.908)	Youden's index 0.534 Sensitivity 0.736 Specificity 0.798	Forest plot

Unreported; "-/-":The front and back are internal and external validation respectively; A: modeling AUC; B: validation AUC; Spiegelhalter Z: This metric is a measure of the model's calibration and predictive performance, often used to assess whether the predicted risk closely aligns with the observed risk. MAE: This metric used to measure the accuracy of a predictive model. Youden's Index: a statistic used to evaluate the performance of binary classification models, calculated as the maximum of the difference between the true-positive rate and the false-positive rate, used to select the optimal classification threshold.Sensitivity (True Positive Rate): the ability of the model to correctly identify Positive cases, such as disease.Specificity (True Negative Rate): Specificity, or the True Negative Rate, indicates the model's ability to correctly identify Negative cases (such as the absence of disease)

study population, leading to optimistic performance estimates that do not hold true in real-world clinical practice. This is especially concerning given the diverse etiologies and risk factors associated with SUI after POP surgery, which can vary widely between populations [40]. Four studies [28, 29, 31, 33] did not report specific performance metrics, which may not allow models to be fully evaluated and used for clinical decision making.

Moreover, Only one study [32] utilized logistic regression combined with machine learning to potentially enhance model accuracy [41]. Hence, there is currently little application of machine learning in this field, and logistic regression is still predominantly used to build models. Logistic regression and machine learning each have their own pros and cons. Logistic regression is easy to implement and interpret, providing clear insight into the relationship between the predictor variables and the outcome through odds ratios. Challenges include small sample sizes and the handling of continuous variables [42]. Nevertheless, logistic regression assumes a linear relationship between the predictor variables and the log odds of the outcome, and cannot capture more complex nonlinear relationships. When the number of predictor variables is large relative to the sample size, logistic regression may overfit the model [43]. Machine learning can effectively capture complex nonlinear relationships between input variables and outcomes, and is generally superior to traditional models in terms of prediction accuracy [44]. It can improve generalization ability by aggregating the results of multiple models, thereby reducing the possibility of overfitting. However, machine learning models also lack interpretability, making it difficult to interpret clinical significance, and require a lot of computing resources and time to train [24, 45].

Despite these limitations, all models showed moderate to good performance but were highly biased. The meta-analysis revealed a moderate level of heterogeneity (I^2 = 68.8%), this heterogeneity and bias underscore the complexity of predicting SUI after POP surgery and highlights several potential sources of variation. Firstly, the included studies involved diverse patient populations, with variations in age, BMI, preoperative SUI status, and surgical procedures. Secondly, the choice of predictors varied widely across the studies.Thirdly, the studies included in this review were predominantly retrospective and single-centered, with some relying on small sample sizes. Therefore, improvements are essential in data sourcing, predictor selection, and model calibration.

The reviewed predictive models hold clinical significance, notably for nursing practice and future studies. Age is widely recognized as a risk factor for SUI after POP surgery, with most studies highlighting increased risk in older women, particularly post-menopause due to reduced oestrogen levels affecting pelvic floor tissues

Table 3 Quality assessment for risk of bias and applicability concern of the included studies

Study	ROB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Jelovsek 2014	-	?	+	?	+	+	+	?	+
Ploeg 2019	-	+	+	-	+	+	+	-	+
Ding 2020	-	?	+	-	+	+	+	-	+
Oh 2022	-	?	+	-	+	+	+	-	+
Fu 2023	-	+	+	-	+	+	+	-	+
Zhu 2023	-	?	?	-	+	+	+	-	+
Li 2024	-	+	+	?	+	+	+	-	+

+:Low bias risk/high applicability; -:indicates high bias risk/low applicability;?: indicates the risk of bias or unclear applicability

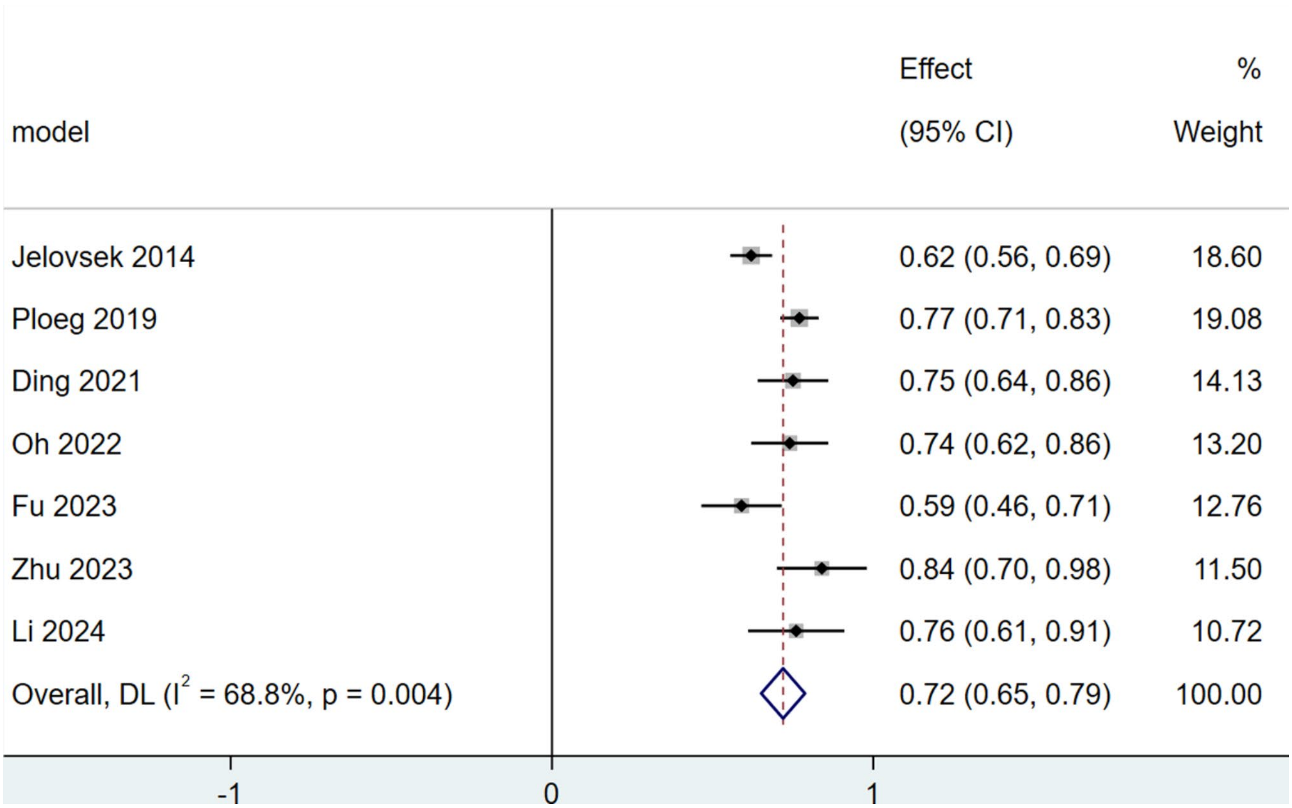

Fig. 2 Forest plot of the random effects meta-analysis of pooled AUC estimates for 7 validation models

[29, 31–34, 46–50]. BMI is also an independent risk factor for the development of SUI after POP [7, 28, 32, 34, 51]. Scholars believe that excess weight and increased abdominal pressure will cause some pressure on the pelvic floor muscles, leading to an increased risk of urinary leakage [52]. However, the influence of BMI, with some studies suggesting its predictive value is impacted by ethnic, cultural, and systemic differences [29, 53]. The POP-Q system's Aa and Ba points, indicators of anterior vaginal wall prolapse, have been shown to predict the risk of postoperative SUI effectively [54]. However, accuracy varies between studies due to individual anatomical and muscle activity differences, impacting the reliability of these measurements [29, 32, 34, 53, 55, 56]. Notably, intraoperative conditions like anesthesia and traction can also alter these measurements, suggesting a need for more precise intraoperative assessments to enhance predictive accuracy [57]. Combining POP surgery with anti-incontinence procedures has been recommended to decrease postoperative SUI risk, although this approach is debated due to potential increases in complications and patient burden [28, 29, 31, 58]. Current practices favor segmented surgery to minimize unnecessary interventions, with tailored surgical plans necessary for optimal patient outcomes [59]. In additions, challenges also arise from the underlying mechanisms of pelvic floor dysfunction, often linked to vaginal delivery. These include mechanical damage to tissues and nerves, leading to structural changes and increased likelihood of long-term incontinence [60]. Some suggest that POP

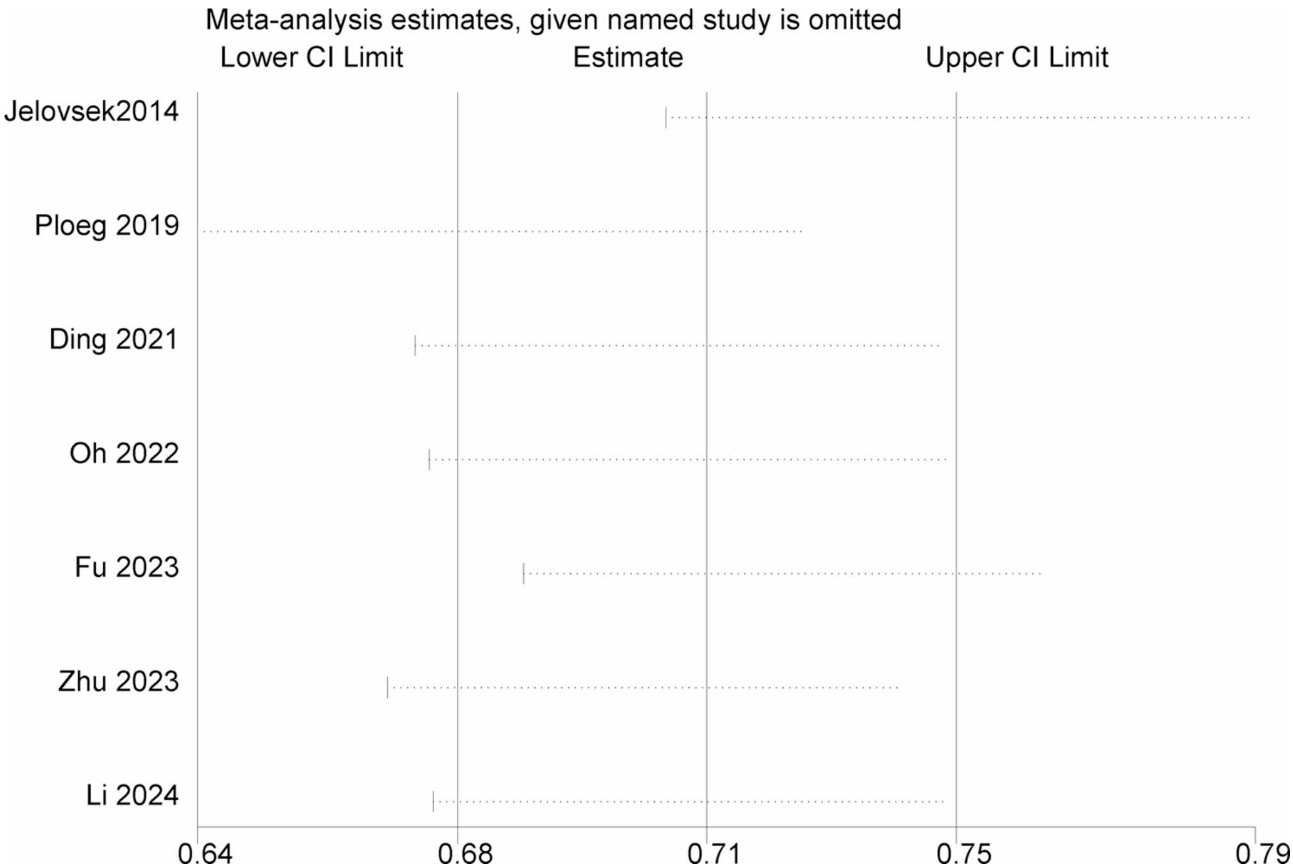

Fig. 3 Sensitivity analysis for 7 models

surgery, while repositioning prolapsed organs, does not comprehensively address pelvic floor defects, potentially exacerbating postoperative incontinence due to altered abdominal pressure dynamics [61]. Therefore, clinicians should pay more attention to patients with this risk factor and perform appropriate anti-incontinence surgery.

In terms of predictors, differences were found between models. Most models included patients' preoperative physiology, disease-related factors, and intraoperative treatment indicators but lacked inclusion of emerging surgeries (e.g., contrast surgery, da Vinci surgery) and postoperative lifestyle factors (e.g., constipation, smoking, alcohol consumption, heavy exercise, blood glucose values, BMI). The impact of these predictors on SUI development needs further exploration.

As a corollary, although there is an urgent need for clinical application of predictive models for the development of stress incontinence after pelvic floor reconstruction, more clinical trials are required to validate the efficacy of the existing predictive models in reducing the incidence of stress incontinence in patients after pelvic floor reconstruction, so as to apply them to a wider population.

Limitation

There are some limitations in this systematic evaluation: (i) Only Chinese and English literature was included, which may introduce language bias, and findings in other major languages were not included in this review; (ii) The literature was only included in this study where the models were developed and internal and/or external validation was completed, which may have a selection bias; (iii) seven validated models from seven studies were included in the meta-analysis of this study, which limits the ability to test for sources of heterogeneity and publication bias discussion. Nevertheless, these issues did not impact the assessment of the models and also in part reflect the methodological and reporting issues we identified.

Conclusion

In this study, we conducted a systematic review and meta-analysis of seven risk prediction model studies, demonstrating that the pooled AUC of the validated models was 0.72 (95% CI: 0.65–0.79), indicating a certain level of discrimination. However, according to PROBAST, all included studies were assessed as being at high risk of bias, and some predictors remain controversial, limiting the clinical applicability of these models.

To improve the clinical relevance of predictive models for stress urinary incontinence (SUI) after pelvic organ prolapse (POP) surgery, these models should be integrated into the clinical workflow for risk stratification, surgical planning, and preoperative counseling. By identifying high-risk patients before surgery, clinicians can tailor interventions, such as concomitant anti-incontinence surgery, to improve postoperative outcomes. Integrating these models into clinical decision support systems will provide clinicians with patient-specific risk assessments, ensuring that the chosen surgical approach is appropriate for each patient. During preoperative consultations, clinicians should show patients the results of the prediction models and explain their individual risk of developing SUI after surgery. This helps patients to actively participate in shared decision-making and have realistic expectations about their postoperative recovery. High-risk patients should be informed that persistent or new urinary incontinence may occur after surgery to help them adjust their expectations and improve patient satisfaction. Incorporating these models into clinical practice will not only help improve decision-making, but also improve the patient experience by promoting clear communication and personalized care.

Furthermore, to further establish good clinical prediction models, future research should follow the Transparent Reporting of Individual Prognosis or Diagnosis (TRIPOD) and PROBAST guidelines. These guidelines provide a comprehensive framework for transparent reporting of the development, validation and performance of prediction models, ensuring that all aspects of the research are clearly documented and reproducible. In addition, multi-centre prospective studies are essential to capture the diversity of patient populations and clinical settings, thereby improving the external validity of these models. In addition, emerging technologies such as artificial intelligence and machine learning offer great potential to overcome current limitations by capturing complex relationships between predictors and outcomes. However, the clinical application of these technologies needs to be rigorously validated using independent datasets and adhere to established guidelines such as TRIPOD to ensure reliability and interpretability. In summary, a combination of reliable methods and innovative technologies is more conducive to supporting clinical decision-making.

Abbreviations

CNKI	China National Knowledge Infrastructure
SinoMed	Chinese Biomedical Literature Service System
POP	Pelvic organ prolapse
PFD	pelvic floor dysfunctional diseases
SUI	Stress urinary incontinence De novo SUI: de novo stress urinary incontinence
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta Analyses
PROSPERO	International Prospective Register of Systematic Reviews
CHARMS	Critical Appraisal and Data Extraction for Systematic Reviews of Predictive Modelling Studies
PFM	Pelvic floor muscles
PROBAST	Prediction Model Risk of Bias Assessment Tool
AUC	Area under the curve
XGBoost	Random Forest and Extreme Gradient Boosting
BMI	body mass index; ROC: reported calibration curve
TRIPOD	Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis

Author contributions

Bijun Yu: study design, methodology and manuscript. Li Wang and Yingmin Liu: Article search, full-text review, and data extraction. Xiaoying Yan and Hanmei Shao: Basic features of the analytical model. Haochong He: Expert review of the methodology, analysis and results. All authors reviewed the manuscript.

Funding

Jiangmen cultivates and grows the team of science and technology specialists and scientific research co-operation projects: 2023760200160008650.

Data availability

All data generated or analysed during this study are included in this published article.

Declarations

Ethics approval and consent to participate

This manuscript is a review article and does not involve a research protocol requiring approval by the relevant institutional review board or ethics committee.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Guangdong Pharmaceutical University, Guangzhou, Guangdong, China
[2]People's Hospital, Jiangmen, Guangdong, China
[3]Jiangmen Central hospital, No. 23, Haibang Street, Pengjiang District, Jiangmen, Guangdong 529030, China
[4]Guangdong Jiangmen Chinese Medicine College, Jiangmen, Guangdong, China

Received: 28 September 2024 / Accepted: 28 January 2025
Published online: 08 February 2025

References

1. Brown HW, Hegde A, Huebner M, Neels H, Barnes HC, Marquini GV, et al. International urogynecology consultation chap. 1 committee 2: epidemiology of pelvic organ prolapse: prevalence, incidence, natural history, and service needs. Int Urogynecol J. 2022;33(2):173–87.
2. Baessler K, Christmann-Schmid C, Maher C, Haya N, Crawford TJ, Brown J. Surgery for women with pelvic organ prolapse with or without stress urinary incontinence. Cochrane Database Syst Rev. 2018;8(8):CD013108.
3. MCPh Slieker-tenH, Pool-Goudzwaard AL, Eijkemans MJC, Steegers-Theunissen RPM, Burger CW, Vierhout ME. The prevalence of pelvic organ prolapse symptoms and signs and their relation with bladder and bowel disorders in a general female population. Int Urogynecol J. 2009;20(9):1037–45.
4. Antosh DD, Kim-Fine S, Meriwether KV, Kanter G, Dieter AA, Mamik MM, et al. Changes in sexual activity and function after pelvic organ prolapse surgery: a systematic review. Obstet Gynecol. 2020;136(5):922.
5. Toyama Y, Suzuki Y, Nakayama S, Endo Y, Kondo Y, Ichikawa M, et al. Outcome of modified laparoscopic sacrocolpopexy and its effect on voiding dysfunction. J Nippon Med Sch. 2022;89(2):222–6.

6. Polanco Pujol L, Caño Velasco J, Rodríguez Fernández E, Cancho Gil MJ, Lledó García E, López-Fando Lavalle L et al. Prevalence and management of urinary incontinece after pelvic organ prolapse surgery (sacrocolpopexy). A literature review. Actas Urol Esp. 2024;S2173-5786(24)00015– 5.
7. Syan R, Dallas KB, Sohlberg E, Rogo-Gupta L, Elliott CS, Enemchukwu EA. Rates and risk factors for future stress urinary incontinence surgery after pelvic organ prolapse repair in a large Population-based Cohort in California. Urology. 2019;123:81–6.
8. van der Ploeg J, van der Steen A, Zwolsman S, van der Vaart C, Roovers J. Prolapse surgery with or without incontinence procedure: a systematic review and meta-analysis. BJOG Int J Obstet Gynaecol. 2018;125(3):289–97.
9. Barski D, Deng DY. Management of Mesh complications after SUI and POP Repair: review and analysis of the current literature. BioMed Res Int. 2015;2015(1):831285.
10. Chughtai B, Mao J, Matheny ME, Mauer E, Banerjee S, Sedrakyan A. Long-term safety with sling mesh implants for stress incontinence. J Urol. 2021;205(1):183–90.
11. Kobashi * Kathleen C, Ward JB, Feinstein L, Abbott KC, Bavendam T, Kirkali Z et al. Mp27-03 economic cost of urinary incontinence for insured adult women in the united states, 2004–2013. J Urol. 2020;203(Supplement 4):e410–e410.
12. Giri A, Hartmann KE, Hellwege JN, Velez Edwards DR, Edwards TL. Obesity and pelvic organ prolapse: a systematic review and meta-analysis of observational studies. Am J Obstet Gynecol. 2017;217(1):11–e263.
13. Tinelli A, Malvasi A, Rahimi S, Negro R, Vergara D, Martignago R, et al. Age-related pelvic floor modifications and prolapse risk factors in postmenopausal women. Menopause N Y N. 2010;17(1):204–12.
14. Li Y, Li Z, Li Y, Gao X, Wang T, Huang Y, et al. Genetics of female pelvic organ prolapse: up to date. Biomolecules. 2024;14(9):1097.
15. Hjorth S, Axelsen SM, Gommesen D, Kjeldsen ACM, Taastrøm KA, Nohr EA. Body mass index, waist circumference, and urinary incontinence in midlife: a follow-up of mothers in the Danish National Birth Cohort. Neurourol Urodyn. 2023;42(5):1111–21.
16. Aune D, Mahamat-Saleh Y, Norat T, Riboli E. Body mass index, abdominal fatness, weight gain and the risk of urinary incontinence: a systematic review and dose-response meta-analysis of prospective studies. BJOG Int J Obstet Gynaecol. 2019;126(12):1424–33.
17. Kernbach JM, Staartjes VE. Foundations of machine learning-based clinical prediction modeling: part II-generalization and overfitting. Acta Neurochir Suppl. 2022;134:15–21.
18. Colmenarejo G. Machine learning models to predict childhood and adolescent obesity: a review. Nutrients. 2020;12(8):2466.
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
20. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data extraction for systematic reviews of Prediction Modelling studies: the CHARMS Checklist. PLOS Med. 2014;11(10):e1001744.
21. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;i6460.
22. Engebretsen S, Bohlin J. Statistical predictions with glmnet. Clin Epigenetics. 2019;11(1):123.
23. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to assess the risk of Bias and Applicability of Prediction Model studies. Ann Intern Med. 2019;170(1):51–8.
24. Fu H, Hou D, Xu R, You Q, Li H, Yang Q, et al. Risk prediction models for deep venous thrombosis in patients with acute stroke: a systematic review and meta-analysis. Int J Nurs Stud. 2024;149:104623.
25. Higgins JPT. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557–60.
26. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629–34.
27. Saltelli A, Ratto M, Tarantola S, Campolongo F. Sensitivity analysis practices: strategies for model-based inference. Reliab Eng Syst Saf. 2006;91(10):1109–25.
28. Jelovsek JE, Chagin K, Brubaker L, Rogers RG, Richter HE, Arya L, et al. A model for Predicting the risk of De Novo Stress Urinary Incontinence in women undergoing pelvic organ prolapse surgery. Obstet Gynecol. 2014;123(2 0 1):279–87.
29. van der Ploeg JM, Steyerberg EW, Zwolsman SE, van der Vaart CH, Roovers JPWR. Stress urinary incontinence after vaginal prolapse repair: development and internal validation of a prediction model with and without the stress test. Neurourol Urodyn. 2019;38(4):1086–92.
30. Ding JX. Construction and validation of a predictive model for the risk of new-onset stress urinary incontinence after pelvic floor repair [Internet] [Master]. Nanjing Medical University; 2021. Available from: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD%26dbname=CMFD202102%26filename=1021573463.nh%26v=.
31. Oh S, Lee S, Hwang WY, Suh DH, Jeon MJ. Development and validation of a prediction model for bothersome stress urinary incontinence after prolapse surgery: a retrospective cohort study. BJOG Int J Obstet Gynaecol. 2022;129(7):1158–64.
32. Fu LR. A machine learning-based predictive model for postoperative stress urinary incontinence after pelvic organ prolapse surgery [Internet] [Master]. Peking Union Medical College; 2023 [cited 2024 Jan 31]. Available from: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD%26dbname=CMFDTEMP%26filename=1023124069.nh%26v=.
33. Zhu L, Guo Y, Wen Y, Yan H, Li Q, Ma J. Risk factors and countermeasures of stress urinary incontinence after Mesh Implantation for patients with pelvic organ prolapse. Arch Esp Urol. 2023;76(3):182–8.
34. Li XC. Study of factors influencing new-onset stress urinary incontinence after pelvic floor reconstruction and risk prediction [Internet] [Master]. Shandong University; 2024 [cited 2024 Jan 31]. Available from: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD%26dbname=CMFD202401%26filename=1023071439.nh%26v=.
35. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
36. Powell JT, Sweeting MJ. Retrospective studies. Eur J Vasc Endovasc Surg. 2015;50(5):675.
37. Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. NeuroImage. 2018;180:68–77.
38. Unverzagt S, Prondzinsky R, Peinemann F. Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. J Clin Epidemiol. 2013;66(11):1271–80.
39. Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single-center trials. Crit Care Med. 2009;37(12):3114–9.
40. Cabitza F, Campagner A, Soares F, García de Guadiana-Romualdo L, Challa F, Sulejmani A, et al. The importance of being external. Methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed. 2021;208:106288.
41. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the wards. Crit Care Med. 2016;44(2):368–74.
42. Lemeshow DWH Jr, Sturdivant S. RX. Applied logistic regression. Wiley; 2013. p. 528.
43. Mood C. Logistic regression: why we cannot do what we think we can do, and what we can do about it. Eur Sociol Rev. 2010;26(1):67–82.
44. Auret L, Aldrich C. Interpretation of nonlinear relationships between process variables by use of random forests. Min Eng. 2012;35:27–42.
45. Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning based prediction models in healthcare [Internet]. 2020 [cited 2025 Jan 19]. Available from: http://arxiv.org/abs/2002.08596
46. Moosavi SY, Samad-Soltani T, Hajebrahimi S, Sadeghi-Ghyassi F, Pashazadeh F, Abolhasanpour N. Determining the risk factors and characteristics of de novo stress urinary incontinence in women undergoing pelvic organ prolapse surgery: a systematic review. Turk J Urol. 2020;46(6):427–35.
47. Gluck O, Rusavy Z, Grinstein E, Abdelkhalek Y, Deval B. Effect of Age on complications Rate and Surgical outcomes in Women undergoing laparoscopic sacrohysteropexy and Sacrocolpopexy. J Minim Invasive Gynecol. 2022;29(6):753–8.
48. Hu P, Lei L, Wei L, Wang Y, Liu L. Investigation of risk factors of de novo urinary stress incontinence after cystocele repair: a retrospective cohort study. Int J Gynecol Obstet. 2022;158(1):213–5.
49. Wu PC, Wu CH, Lin KL, Liu Y, Loo Z, Lee YC, et al. Predictors for de novo stress urinary incontinence following pelvic reconstruction surgery with transvaginal single-incisional mesh. Sci Rep. 2019;9(1):19166.
50. Capobianco G, Madonia M, Morelli S, Dessole F, De Vita D, Cherchi PL, et al. Management of female stress urinary incontinence: a care pathway and update. Maturitas. 2018;109:32–8.

51. Lensen EJM, Withagen MIJ, Kluivers KB, Milani AL, Vierhout ME. Urinary incontinence after surgery for pelvic organ prolapse. Neurourol Urodyn. 2013;32(5):455–9.
52. Cruz RA, Faria CA, Gomes SCS. Predictors for de novo stress urinary incontinence following pelvic reconstructive surgery with mesh. Eur J Obstet Gynecol Reprod Biol. 2020;253:15–20.
53. Wang SY, Cao TT, Wang RZ, Yang X, Sun XL, Wang JL. Incidence and risk factors of De novo stress urinary incontinence after Pelvic Floor Reconstruction: a nested case-control study. Chin Med J (Engl). 2017;130(06):678–83.
54. Liu C, Wu WY, Yang Q, Hu M, Zhao Y, Hong L. Diagnostic and prognostic value of aa and Ba indicator points for occult stress incontinence during pelvic floor reconstruction surgery. Chin J Obstet Gynecol. 2015;50(6):415–9.
55. Sato H, Abe H, Ikeda A, Miyagawa T, Tsukada S. Severity of Cystocele and Risk factors of postoperative stress urinary incontinence after laparoscopic sacrocolpopexy for pelvic organ prolapse. Gynecol Minim Invasive Ther. 2022;11(1):28–35.
56. Ostrzenski A. Pelvic organ prolapse quantification (POP-Q) system needs revision or abandonment: the anatomy study. Eur J Obstet Gynecol Reprod Biol. 2021;267:42–8.
57. Chaudhuri S, Vaitheeswari J. Correlation of preoperative and intraoperative assessment of pelvic organ prolapse by pelvic organ prolapse quantification system. Int J Gynecol Obstet. 2022;157(2):466–70.
58. El Hamamsy D, Fayyad AM. New onset stress urinary incontinence following laparoscopic sacrocolpopexy and its relation to anatomical outcomes. Int Urogynecol J. 2015;26(7):1041–5.
59. Huang KH, Chang YW, Ngo YG, Wu LY, Yang TH, Lee YL, et al. Short-term results of stress urinary incontinence in women undergoing laparoscopic sacrocolpopexy with and without midurethral sling. Low Urin Tract Symptoms. 2021;13(3):341–6.
60. Weintraub AY, Glinter H, Marcus-Braun N. Narrative review of the epidemiology, diagnosis and pathophysiology of pelvic organ prolapse. Int Braz J Urol. 2019;46:5–14.
61. Fang LX. Analysis of influencing factors associated with new-onset stress urinary incontinence after pelvic floor reconstruction surgery in patients with POP [Internet] [Master]. Third Military Medical University; 2018 [cited 2024 Feb 27]. Available from: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD%26dbname=CMFD201801%26filename=1017235692.nh%26v=.

Publisher's note