

Predicting Risk of Urinary Incontinence and Adverse Events After Midurethral Slings Surgery in Women

J. Eric Jelovsek, MD, MMEd, Audra Jolyn Hill, MD, Kevin M. Chagin, MS, Michael W. Kattan, MBA, PhD, and Matthew D. Barber, MD, MHS

OBJECTIVE: To construct and validate models that predict a patient's risk of developing stress and urgency urinary incontinence and adverse events 12 months after sling surgery.

METHODS: This was a secondary analysis of four randomized trials. Twenty-five candidate predictors (patient characteristics and urodynamic variables) were identified from the National Institute of Diabetes and Digestive and Kidney Diseases Trial of Mid-Urethral Slings (N=597). Multiple logistic models were fit to predict four different outcomes: 1) bothersome stress urinary incontinence; 2) a positive stress test; 3) bothersome urgency urinary incontinence; and 4) any adverse event up to 12 months after sling surgery. Model discrimination was measured using a concordance index. Each model's concordance index was internally validated using 1,000 bootstrap samples and calibration curves were plotted. Final models

were externally validated on a separate data set (n=902) from a combination of three different multicenter randomized trials.

RESULTS: Four best models discriminated on internal validation between women with bothersome stress urinary incontinence (concordance index 0.728, 95% confidence interval [CI] 0.683–0.773), a positive stress test (concordance index 0.712, 95% CI 0.669–0.758), bothersome urgency urinary incontinence (concordance index 0.722, 95% CI 0.680–0.764), and any adverse event (concordance index 0.640, 95% CI 0.595–0.681) after sling surgery. Each model's concordance index was reduced as expected when important variables were removed for external validation, but model discrimination remained stable with bothersome stress urinary incontinence (concordance index 0.548), a positive stress test (concordance index 0.656), bothersome urgency urinary incontinence (concordance index 0.621), and any adverse event (concordance index 0.567). Predicted probabilities are closest to actual probabilities when predictions are less than 50%.

CONCLUSION: Four best and modified models discriminate between women who will and will not develop urinary incontinence and adverse events 12 months after midurethral sling surgery 64–73% and 55–66% of the time, respectively.

(*Obstet Gynecol* 2016;127:330–40)

DOI: 10.1097/AOG.0000000000001269

From the Obstetrics, Gynecology & Women's Health Institute, Urogynecology and Reconstructive Pelvic Surgery, and Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio.

Presented at the Pelvic Floor Disorders Week–36th Annual American Urogynecologic Society meeting, October 13–17, 2015, Seattle, Washington.

The Trial of Mid-Urethral Slings and Value of Urodynamic Evaluation trial were conducted by the Urinary Incontinence Treatment Network Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The data from the Trial of Mid-Urethral Slings and Value of Urodynamic Evaluation trial reported here were supplied by the NIDDK Central Repositories.

This article was not prepared in collaboration with Investigators of the Trial of Mid-Urethral Slings or Value of Urodynamic Evaluation trial and does not necessarily reflect the opinions or views of the Trial of Mid-Urethral Slings or Value of Urodynamic Evaluation trial study, the NIDDK Central Repositories, or the NIDDK.

Corresponding author: J. Eric Jelovsek, MD, MMEd, 9500 Euclid Avenue, Desk A81, Cleveland, OH 44195; e-mail: jelovsj@ccf.org.

Financial Disclosure

The authors did not report any potential conflicts of interest.

© 2016 by The American College of Obstetricians and Gynecologists. Published by Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0029-7844/16

A recent Cochrane review of 81 trials involving surgery for stress urinary incontinence (SUI) in more than 12,000 women concluded that inserting a midurethral sling to support the urethra using a retropubic or transobturator approach results in similar cure rates but differences in complications and need for retreatment.¹ These data support the use of synthetic midurethral slings as effective and safe treatments for women with SUI. However, given the increased media attention and litigation associated with transvaginal mesh,



including slings, there is an urgent need for surgeons to understand and more accurately convey specific risks and benefits of these surgeries at the individual patient level rather than providing the average rates seen across patient populations from the literature. This will allow a woman to carefully balance a number of factors for her specific situation when choosing this commonly performed operation.²

Predictive analytic techniques incorporate data from surgical trials and systematic reviews into prognostic models that estimate the likelihood of these important events occurring for an individual patient.² Moreover, there have been recent proposals that predictive tools should become part of essential recommendations for clinical care and included into guideline recommendations.³⁻⁶ Despite their increased use in many areas involving nonsurgical decisions, the availability of predictive analytic tools and their use during the surgical consent process are less well characterized.

The aims of this study were to 1) develop; 2) internally validate; and 3) externally validate clinical prediction models to predict a patient's risk of bothersome SUI, a positive urinary stress test, bothersome urgency urinary incontinence (UUI), and developing an adverse event 12 months during or after midurethral sling surgery for SUI. Our hypotheses were that prediction models could be developed to predict these outcomes using patient characteristics and urodynamic measures.⁷ Although we specifically tested the hypotheses that these models would discriminate better than chance (concordance index 0.5), our aim was to make the most accurate models possible and report how well they perform. Because there is large variability in the predicted rates given to patients and clinical prediction models show consistent superiority over expert clinicians as a result of physician cognitive biases, we assume that there is added value in using prediction models to provide more accurate estimates of risk to patients.³⁻⁵

MATERIALS AND METHODS

This study was a secondary analysis of four randomized trials. The study received institutional review board exemption status because two trials were analyses of publicly available data sets and two were from previously approved institutional review board-approved trials at our institution. This study was conducted using methods set forth in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement.⁶ The source of data for developing and internally validating all models came from

597 participants in the Trial of Mid-Urethral Slings data set (ClinicalTrials.gov number NCT00325039) available from the National Institute of Diabetes and Digestive and Kidney Diseases Central Repositories (<https://www.niddkrepository.org>).⁷ The Trial of Mid-Urethral Slings was a multicenter, randomized equivalence trial comparing outcomes with retropubic and transobturator midurethral slings in women with SUI. Details about study participants, setting, locations, eligibility criteria, and treatments received during the study are available elsewhere.^{7,8} The primary outcome of Trial of Mid-Urethral Slings was treatment success at 12 months after surgery according to both objective criteria (a negative stress test, a negative pad test, and no retreatment) and subjective criteria (self-reported absence of symptoms, no leakage episodes recorded, and no retreatment).^{7,8}

As a result of limitations on how to best define successful and unsuccessful treatment of urinary incontinence, we elected to model both subjective SUI and UUI outcomes and objective SUI outcomes. Subjective SUI and UUI outcomes were considered positive if they were considered by the patient as at least "bothersome." Bothersome symptoms are most meaningful from the patient's perspective and capture overall benefits and harms of managing a condition. Bothersome SUI at 12 months after surgery was defined by the self-reported Urinary Distress Inventory stress item: Do you currently experience urine leakage related to physical activity, coughing, or sneezing? If yes, how bothersome?⁹ Similarly, bothersome UUI at 12 months was defined by Urinary Distress Inventory urge item: Do you currently experience urine leakage related to the feeling of urgency? If yes, how bothersome? In each case, a response of no or not at all bothersome was considered negative and slightly bothersome, moderately bothersome, or greatly bothersome was considered a positive response. The 19-item Urinary Distress Inventory is a valid and reliable urinary incontinence symptom-specific questionnaire that has demonstrated discriminatory validity among known subtypes of urinary incontinence (stress, urge, and mixed) and is responsive to change.^{9,10} Objective SUI was defined as the presence of leaking (positive) visualized during a urinary stress test at a bladder volume of 300 mL or less at 12 months after midurethral sling placement. Urodynamic testing in the Trial of Mid-Urethral Slings was performed in accordance with International Continence Society guidelines.^{11,12} The reporting of adverse events and the classification of the severity of events were also standardized across sites with the use of specific prespecified definitions of events and



a modified version of the classification system of Dindo et al.^{7,13} For this study, the same definition of an adverse event was used as in the Trial of Mid-Urethral Slings and was considered to have occurred if an adverse event was recorded in the data set at the time of surgery or any time up to and including 12 months after surgery (footnote in Table 1).

Multiple logistic models were fit to predict four different outcomes: 1) bothersome SUI; 2) a positive stress test; 3) bothersome UUI; and 4) any adverse event up to 12 months after sling surgery. All prediction models were developed and internally validated using Trial of Mid-Urethral Slings data from participants with 1-year outcomes. A priori, investigators identified 82 candidate risk factors that could be considered during clinical counseling to estimate a patient's risk of having SUI, UUI, or an adverse event after surgery. Candidate variable parity (zero, one, two, three, four or more), pelvic organ prolapse quantification (POP-Q) point Ba (less than +1 compared with +1 or greater), and POP-Q point C (less than +1 compared with +1 or greater) were modeled as discrete ordered values and as continuous values. Each symptom severity scale score or quality-of-life question response was modeled as a candidate predictor with total scale scores and subscale scores

considered as continuous variables and question responses as categorical and ordered discrete variables.

Multiple logistic regression models were fit onto the data set using the identified candidate risk factors and outcomes from the Trial of Mid-Urethral Slings. Missing risk factor values were assessed for missing at random, and multiple imputation using chained equations were used to calculate missing values.¹⁸ The outcomes for all models were based only on actual and not imputed events, whereas imputed values for risk factors may have been used to predict outcomes in each model. Variable selection process was performed using Harrell's "model approximation" process of backward elimination to rank the variables in order of importance starting from the full model using a bootstrap bias-corrected concordance index as the stopping criteria to find the best parsimonious model.¹⁹ As recommended, variables with individual *P* values that were $>.05$ were left in the model if they offered information to improve the overall model accuracy. The removal of each variable was evaluated by determining which variable had the smallest effect on the R^2 and was stopped when the bootstrap concordance index had a change <0.01 . This process provides the best fit parsimonious model that was compared with the full model to determine

Table 1. Models to Predict Urinary Incontinence Outcomes and Adverse Events 12 Months After Midurethral Sling Surgery

Outcome	Best Parsimonious Models			Modified Parsimonious Models					
	Internal Validation* (n=597)			Internal Validation (n=597)			External Validation [†] (n=902)		
	No. of Predictors	Concordance Index	95% CI	No. of Predictors	Concordance Index	95% CI	<i>P</i> [‡]	Concordance Index	<i>P</i> [§]
Bothersome SUI	10	0.728	0.683–0.773	15	0.718	0.674–0.765	.503	0.548	<.001
Stress test	10	0.712	0.669–0.758	13	0.697	0.651–0.739	.834	0.656	.486
Bothersome UUI	22	0.722	0.680–0.764	12	0.697	0.651–0.740	.547	0.621	.034
Any adverse event	10	0.64	0.595–0.681	11	0.636	0.563–0.681	.843	0.567	.114

SUI, stress urinary incontinence; UUI, urgency urinary incontinence.

* Model building and internal validation were performed using the Trial of Mid-Urethral Slings data set.^{7,8}

[†] External validation was performed using a data set from three combined clinical trials.^{14–17}

[‡] Best compared with modified models.

[§] Modified models compared with external validation.

^{||} Adverse events in the Trial of Mid-Urethral Slings were considered positive if investigators responded yes to "Did any adverse events or complication occur during surgery" after surgery and at hospital discharge or "Did any adverse events or complications other than voiding dysfunction, urge incontinence, or pain occur since the last study visit?" Adverse events included: bladder or urethral perforation, acute renal failure, anesthetic complication, device malfunction, deep vein thrombosis, pulmonary embolus, myocardial infarction, cerebrovascular accident, death, intraoperative or postoperative bleeding, injury (bowel, rectal, vascular), mesh erosion, mesh exposure, surgical site infection (superficial, deep, organ or space), recurrent urinary tract infection, fistula (vesicovaginal, urethrovaginal, enterovesical, rectovaginal), neurologic symptoms, granulation tissue, or other. Specific adverse events were not provided in the public data set.^{7,8}



which was more accurate. Overall, 82 candidate risk factors were identified in the Trial of Mid-Urethral Slings data set; 50 were eliminated during the backward elimination process leaving 32 risk factors ultimately used to build and internally validate the best models. Twenty-five of 82 candidate variables were identified in the Trial of Mid-Urethral Slings data set that were also present in the external validation data set and 23 of these variables were used to build the modified models after variable reduction.

Each logistic model's discriminative ability was measured using the concordance statistic or *c*-statistic. The concordance index is a measurement of the model's ability to discriminate between high- and low-risk patients. Given a pair of individuals, in which one has the outcome and the other does not, the concordance index is the proportion of pairs in which the model will yield a higher predicted risk for the first patient than for the second. Values closer to 1 indicate better discrimination between patients with and without the outcome and a value of 0.5 indicates that the model is no better than chance. All concordance indices were internally validated using 1,000 bootstrap samples to correct for bias and overfitting within the model. The bootstrapping method of validation has been shown to be superior to other approaches to estimate internal validity such as split-sample methods.²⁰ The concordance index measures the ability of the model to rank patients from high to low risk but does not assess the ability of a model to assign accurate probabilities of an event occurring (that is measured by the model's calibration curve). Calibration curves were plotted to depict the relationship between the model's predicted outcomes against the cohort's observed outcome, where a perfectly calibrated model follows a 45° line. Finally, Brier scores are also reported as a measure of overall performance and most similar to a model's R^2 .²¹

The external validation data set consisted of three combined data sets ($n=902$) from previously performed multicenter randomized trials involving 15 different centers in the United States: 1) the Value of Urodynamic Evaluation trial ($n=596$) obtained from the National Institute of Diabetes and Digestive and Kidney Diseases Central Repository^{14,15}; 2) a trial of transobturator tape compared with tension-free vaginal tape for the treatment of SUI¹⁶; and 3) a multicenter trial comparing single-incision minisling compared with tension-free vaginal tape for the treatment of SUI.¹⁷ A summary of each trial is displayed in Appendix 1, available online at <http://links.lww.com/AOG/A768>, and details about each study are available elsewhere.¹⁴⁻¹⁷ Participants in the Value of

Urodynamic Evaluation study who received a traditional bladder neck sling (approximately 4%, $n=34$) were excluded from the data set. We were unable to exclude participants in the Value of Urodynamic Evaluation study who received a minisling (only 10 participants, less than 2%) because these patients were indistinguishable from those in the transobturator group in the public use data set. Participants who received the minisling in the Barber et al 2012 study were also excluded ($n=127$) from the external validation data set.

For this study, it was known that of the 45 predictors identified while building the best models using the Trial of Mid-Urethral Slings data set, only 25 were present in the external validation data sets. Therefore, we chose to build two sets of models for each outcome: 1) models with the best available predictors from the Trial of Mid-Urethral Slings resulting in four "best" parsimonious models; and 2) a second set of models using predictors present in all data sets resulting in four "modified" parsimonious models, which could be externally validated on the available data sets. This study reports both sets of models so users could decide which models to use for their own situation. We hypothesized that each best model's accuracy would drop as these important variables were removed for external validation but that each modified parsimonious model would continue to discriminate and have predictions close to actual probabilities of each outcome on the external data set making them ready for clinical use. All analyses were performed using R 2.15.2.

RESULTS

There were a total of 1,499 women with data available 12 months after surgery (Trial of Mid-Urethral Slings= 597 , Value of Urodynamic Evaluation= 596 , a trial of transobturator tape compared with tension-free vaginal tape= 170 , and a multicenter trial comparing single-incision minisling compared with tension-free vaginal tape= 136). Four prediction models were fit to the Trial of Mid-Urethral Slings data set including: one to predict bothersome SUI, one to predict a positive stress test, one to predict bothersome UUI, and one to predict any adverse event within or up to 12 months after midurethral sling placement. The concordance indices of all models are shown in Table 1. The unadjusted comparisons of the Trial of Mid-Urethral Slings and external validation data set by outcome are provided in Table 2 and Appendices 2-5, available online at <http://links.lww.com/AOG/A769>, <http://links.lww.com/AOG/A770>, <http://links.lww.com/AOG/A771>, and <http://links.lww.com/AOG/A772>, respectively.



Table 2. Candidate Predictors of Bothersome Stress Urinary Incontinence in the Trial of Mid-Urethral Slings and External Validation Data Set

Variable	Bothersome SUI					
	TOMUS ^{7,8}			External ¹⁴⁻¹⁷		
	Yes [112 (21)]	No [414 (79)]	P	Yes [131 (17)]	No [620 (83)]	P
Age	53.1 (48.0–62.1)	51.6 (45.9–60.5)	.128	55.6 (48.3–63.6)	50.5 (44.3–58.4)	<.001
Race						
Caucasian	98 (87)	326 (79)	.038	113 (86)	519 (85)	.73
Other	14 (13)	88 (21)		18 (14)	91 (15)	
Parity						
0	18 (16)	46 (72)	.045	14 (11)	61 (10)	.303
1	25 (22)	63 (15)		21 (16)	104 (17)	
2	23 (21)	141 (34)		56 (43)	226 (37)	
3	29 (26)	99 (24)		31 (24)	142 (23)	
4 or greater	17 (15)	65 (16)		9 (7)	82 (13)	
Menopausal state						
Yes	79 (75)	277 (69)	.232	56 (43)	293 (48)	.344
No	26 (25)	123 (31)		74 (57)	322 (52)	
POP-Q Ba						
Less than +1	105 (94)	377 (91)	.363	120 (67)	569 (96)	.563
+1 or greater	7 (6)	37 (9)		4 (3)	26 (4)	
POP-Q C						
Less than +1	106 (95)	393 (95)	.904	123 (99)	586 (98)	.484
+1 or greater	6 (5)	21 (5)		1 (1)	13 (2)	
Hormonal therapy						
Yes	30 (27)	127 (31)	.448	40 (31)	171 (28)	.496
No	81 (73)	286 (69)		90 (69)	444 (72)	
Prior pelvic surgery						
Yes	78 (70)	297 (72)	.734	82 (63)	375 (61)	.73
No	33 (30)	116 (28)		49 (37)	240 (39)	
Pretreatment with a medication						
Yes	36 (32)	120 (29)	.517	32 (24)	123 (20)	.253
No	76 (68)	294 (71)		99 (76)	493 (80)	
UDI stress score	83.3 (66.7–83.3)	83.3 (50.0–95.8)	.076	83.3 (66.7–100)	83.3 (50–100)	.25
Feeling of incomplete bladder emptying?						
Yes	52 (46)	230 (56)	.077	63 (49)	273 (45)	.373
No	60 (54)	182 (44)		65 (51)	335 (55)	
Urine leakage related to a feeling of urgency?						
Yes	72 (64)	282 (69)	.386	88 (69)	387 (64)	.274
No	40 (36)	129 (31)		40 (31)	221 (36)	
Pain in the lower abdominal or genital area?						
Yes	23 (20)	74 (18)	.52	27 (21)	117 (19)	.638
No	89 (80)	340 (82)		101 (79)	490 (81)	
Heaviness or dullness in the pelvic area?						
Yes	32 (29)	102 (25)	.405	36 (28)	167 (27)	.868
No	80 (71)	311 (75)		94 (72)	452 (73)	
Frequent urination?						
Yes	85 (76)	271 (65)	.037	82 (64)	383 (63)	.837
No	27 (24)	143 (35)		46 (36)	224 (37)	
Entertainment activity limited by urinary leakage?						
Yes	80 (71)	228 (55)	.002	77 (61)	306 (50)	.037
No	32 (29)	185 (45)		50 (39)	301 (50)	

(continued)



Table 2. Candidate Predictors of Bothersome Stress Urinary Incontinence in the Trial of Mid-Urethral Slings and External Validation Data Set (continued)

Variable	Bothersome SUI					
	TOMUS ^{7,8}			External ¹⁴⁻¹⁷		
	Yes [112 (21)]	No [414 (79)]	<i>P</i>	Yes [131 (17)]	No [620 (83)]	<i>P</i>
Recreational activity limited by urinary leakage?						
Yes	95 (85)	349 (85)	.935	114 (89)	530 (87)	.586
No	17 (15)	64 (25)		14 (11)	77 (13)	
Emotional health limited by urinary leakage?						
Yes	87 (78)	277 (67)	.031	91 (72)	387 (64)	.094
No	25 (22)	136 (33)		36 (28)	219 (36)	
Frustration resulting from urinary leakage?						
Yes	102 (91)	359 (87)	.214	120 (94)	544 (90)	.151
No	10 (9)	55 (13)		8 (6)	63 (10)	
Maximum flow	24 (15.3–31.5)	23.3 (16.8–30.0)	.838	22 (15.2–30.3)	21 (14–29.8)	.388
Bladder volume at MCC	334 (261.8–396)	338.5 (273–403.2)	.595	364 (299–458)	367 (300–466.2)	.362
Void volume	267 (216–344)	293 (214–351.8)	.451	183 (97.5–267.5)	187 (90.8–302.8)	.727
Lowest Valsalva leak point pressure	84.8 (66–116.2)	91 (65–118.2)	.386	103 (69–181)	108 (74–151.2)	.998
Concomitant surgical procedures						
Yes	28 (25)	104 (25)	.971	37 (28)	176 (29)	.932
No	84 (75)	310 (75)		94 (72)	439 (71)	
Type of midurethral sling						
Retropubic	51 (45)	213 (51)	.267	95 (72)	443 (72)	.806
Transobturator	61 (55)	201 (49)		36 (28)	177 (28)	

SUI, stress urinary incontinence; TOMUS, Trial of Mid-Urethral Slings; POP-Q, pelvic organ prolapse quantification; UDI, Urinary Distress Inventory; MCC, maximum cystometric capacity.

Data are n (%) or median (interquartile range) unless otherwise specified.

Users can enter predictor values using the online calculator (<http://rcalc.ccf.org>) to view predicted probabilities of each outcome.

Data from 526 patients were used to predict bothersome SUI, with 112 (21%; 71 missing) patients reporting bothersome SUI at 12 months (Table 2). Data from all 597 Trial of Mid-Urethral Slings participants were used to predict a positive stress test, and 128 (21%) experienced a positive stress test at 12 months (Appendix 2, <http://links.lww.com/AOG/A769>). The best parsimonious model accurately discriminated between women with and without bothersome SUI 72.8% of the time (concordance index 0.728, 95% confidence interval [CI] 0.683–0.773, Brier score 0.152, 95% CI 0.134–0.173) and between women with a positive stress test 71.2% of the time (concordance index 0.712, 95% CI 0.669–0.758, Brier score 0.155, 95% CI 0.137–0.174). Appendix 6, available online at <http://links.lww.com/AOG/A773>, lists the predictors included in the best and modified models along with their parameter estimates and the direction each variable contributes to the predicted risk. For example,

age has a parameter estimate of 0.0592 in the modified bothersome SUI model; this indicates that a one-unit increase in age is associated with an increase in the log odds of bothersome SUI by 0.0592 units. For categorical variables, Caucasian race has a parameter estimate of –0.538 in the modified any adverse event model; this indicates that being Caucasian is associated with a decrease in the log odds of developing an adverse event by 0.538 units.

The best bothersome SUI model included 10 predictors including age, race, parity, menopausal status, Urinary Distress Inventory irritative, Urinary Distress Inventory stress, Urinary Distress Inventory obstructive, feeling of incomplete emptying, leaking related to urgency, and limitations in entertainment activities. The best stress test model also included 10 predictors including: age, parity, menopausal status, pretreatment with a device for urinary incontinence, Urinary Distress Inventory irritative symptom score, POP-Q point Ba, maximum flow, pressure flow voiding pattern, any concomitant surgical procedure, and sling type (Appendix 6, <http://links.lww.com/AOG/A773>).



Figures 1 and 2 demonstrate the calibration curves for the best and modified bothersome SUI and stress test models during internal calibration, respectively. Both models assigned accurate probabilities of each event occurring when predictions ranged from 0 to 40–50%. Both bothersome SUI models had a tendency to assign higher than actual probabilities as the predictions started to approach 40% or greater. However, this only occurred in a very small subset of the population.

A total of 526 patients was used to build a model to predict UII because 71 patients were missing result information (Appendix 3, <http://links.lww.com/AOG/A770>). Overall, 135 (26%) reported bothersome UII and 391 (74%) did not report bothersome UII at 12 months. The best parsimonious UII model accurately discriminated between women with and without bothersome UII 72.2% of the time (concordance index 0.722, 95% CI 0.680–0.764, Brier score 0.174, 95% CI 0.155–0.195) after midurethral sling surgery. The best UII model included 22 predictors that are listed in Appendix 6 (<http://links.lww.com/AOG/A773>). Maximum flow rate, maximum cystometric capacity, and lowest Valsalva leak point pressure improved accuracy of predicting bothersome UII, whereas sling type did not.

All 597 patients were used to predict the adverse events model. One hundred eighty-four (31%) experienced an adverse event and 413 (69%) did not experience an adverse by 12 months (Appendix 4, <http://links.lww.com/AOG/A771>). The best parsimonious model accurately discriminated between women with and without any adverse event after sling

surgery 64% of the time (concordance index 0.640, 95% CI 0.595–0.681, Brier score 0.206, 95% CI 0.191–0.225). The best model included 10 predictors including: race, parity, education, urine leakage related to urgency, pain in the abdominal or genital area, frustration resulting from urinary leakage, maximum capacity, Valsalva leak point pressure, concomitant surgical procedures, and sling type (Appendix 6, <http://links.lww.com/AOG/A773>). Sling type increased predictive accuracy in determining who would experience an adverse event with transobturator slings reducing risk compared with retropubic slings.

Figures 3 and 4 demonstrate the calibration curves for the UII and adverse event models, respectively. The best and modified UII models had accurate predictions along a range of predicted probabilities. The best UII model assigned accurate probabilities of developing UII when predicted probabilities were between 0% and 40% but had a tendency to assign higher predicted probabilities than actual as probabilities approached approximately 40–50% or greater. The best and modified adverse event models had a tendency to assign slightly lower than actual probabilities of developing an adverse event when predicted probabilities were less than 20% and assigned higher than actual probabilities when predicted probabilities exceeded 50–60%.

Several urodynamic parameters added value to predicting the outcomes in this study and remained in the best and modified models (Table 1). Maximum flow rate using noninstrumented flowmetry and the type of voiding pattern (pure or predominate abdominal, pure or predominate detrusor, mixed or

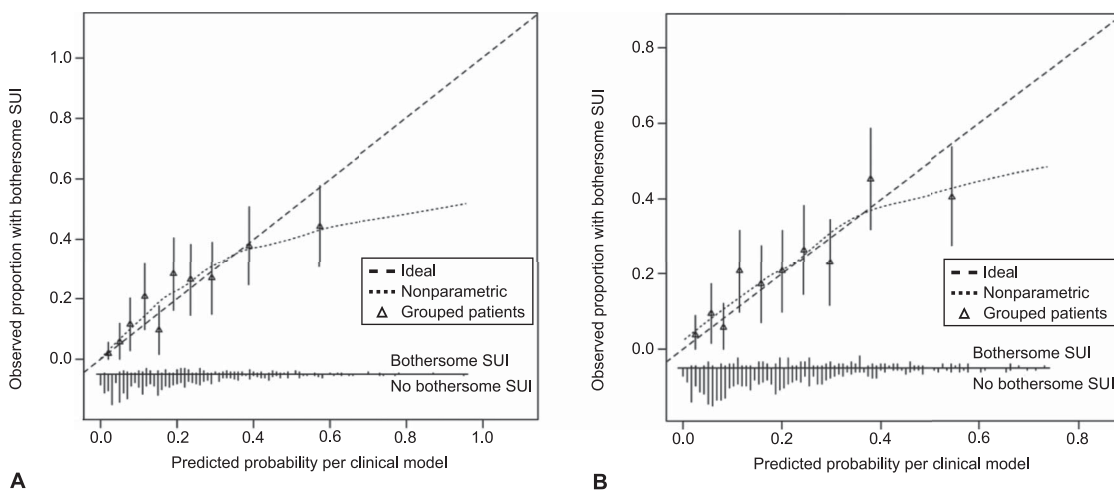


Fig. 1. Calibration curves for the best bothersome stress urinary incontinence (SUI) model (A) and the modified bothersome SUI model (B).

Jelovsek. Incontinence and Adverse Events After Sling Surgery. *Obstet Gynecol* 2016.



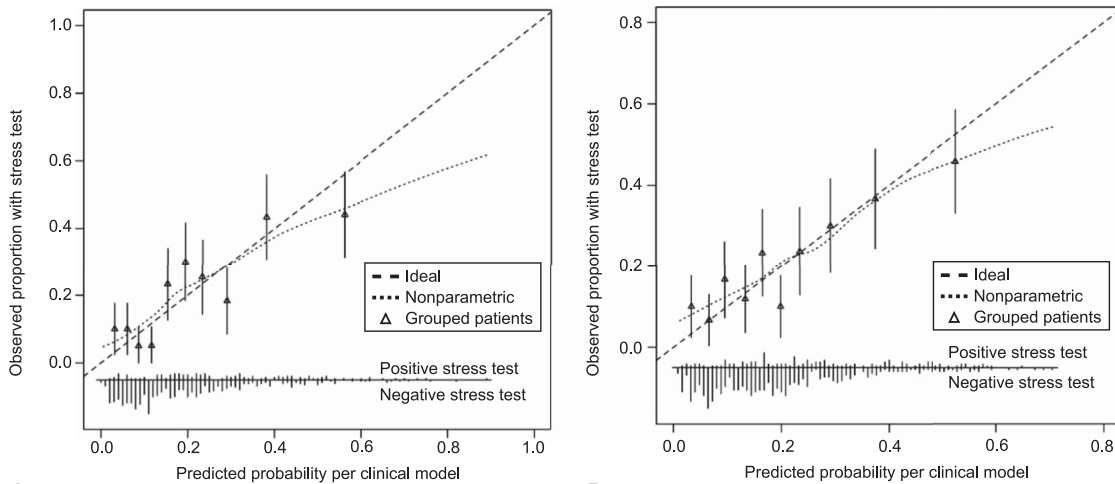


Fig. 2. Calibration curves for the best stress test model (A) and the modified stress test model (B). Jelovsek. *Incontinence and Adverse Events After Sling Surgery*. *Obstet Gynecol* 2016.

indeterminate or uninterpretable) on a pressure flow study were useful in predicting an individual having a positive stress test (increase in concordance index of 0.001) and bothersome UUI (increase in concordance index of 0.002). Increasing bladder volume to achieve maximum cystometric capacity was associated with increased accuracy in discriminating patients with UUI (increase in concordance index of 0.013) and developing an adverse event (increase in concordance index of 0.018), whereas higher Valsalva leak point pressures predicted lower risk of UUI (increase in concordance index of 0.009) and developing an adverse event (increase in concordance index of 0.012). Lowest Valsalva leak point pressure did not increase predictive accuracy for either the stress test or

bothersome SUI models. It is recognized that not all health care providers have access to urodynamics. Therefore, models were also created that did not consider urodynamic parameters. The concordance indices and calibration curves are presented in Appendix 7, available online at <http://links.lww.com/AOG/A774>.

The external validation data set to test the models consisted of 902 patients (Value of Urodynamic Evaluation n=596, Barber et al 2008, n=170, Barber et al 2012, n=136). A comparison among the demographics, predictors, and outcomes from the Trial of Mid-Urethral Slings data set and the external validation data set are provided in Appendix 5, <http://links.lww.com/AOG/A772>. The predictors and their

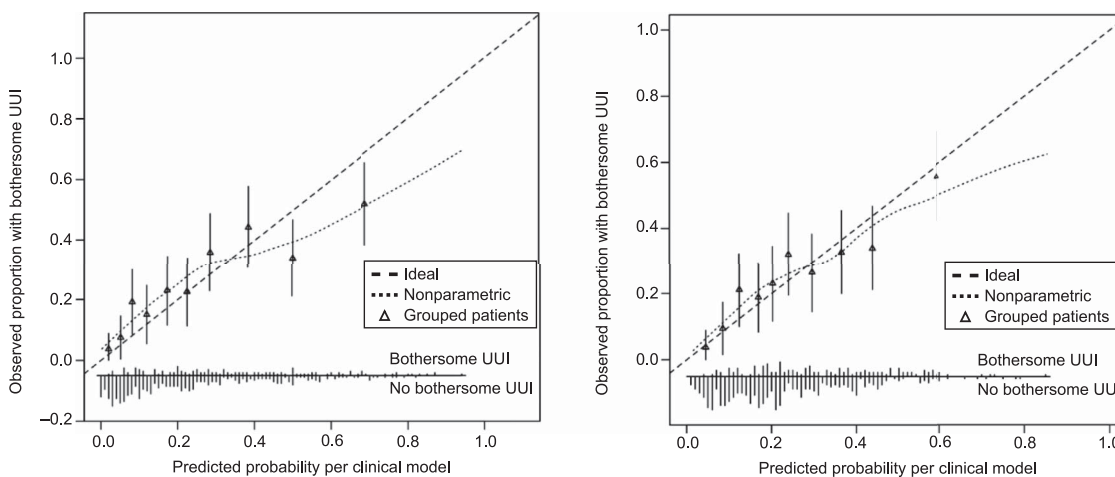


Fig. 3. Calibration curves for best bothersome urgency urinary incontinence (UUI) (A) and modified bothersome UUI (B). Jelovsek. *Incontinence and Adverse Events After Sling Surgery*. *Obstet Gynecol* 2016.



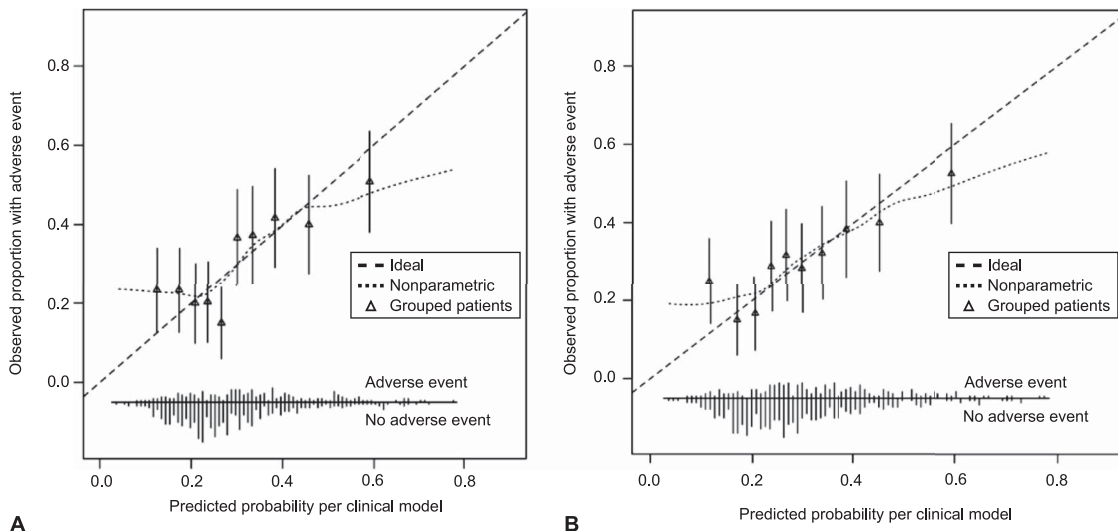


Fig. 4. Calibration curves for the best adverse event model (A) and the modified adverse event model (B). Jelovsek. *Incontinence and Adverse Events After Sling Surgery*. *Obstet Gynecol* 2016.

parameter estimates for these modified models are located in Appendix 6 (<http://links.lww.com/AOG/A773>). All concordance indices in the modified models were slightly lower than the best performing models but were able to discriminate 55–66% of the time when externally validated using the data set from three combined trials as shown in Table 1 (modified bothersome SUI [concordance index 0.548], modified positive stress test [concordance index 0.656], modified bothersome UUI [concordance index 0.621], and modified any adverse event [concordance index 0.567]). Sling type improved accuracy of predictions in the best and modified stress test and adverse event models. Sling type did not improve accuracy of predictions in either UUI model and when the bothersome SUI model was modified for external validation, sling type did improve accuracy.

DISCUSSION

Our study has demonstrated that clinical prediction models can be developed that can discriminate between women with and without recurrent bothersome SUI, having or not having a positive stress test, with or without bothersome UUI, and developing or not developing an adverse event 12 months after midurethral sling surgery for SUI 55–73% of the time. These models are most accurate when predicted probabilities are less than 50% and begin to have higher predicted probabilities than actual probabilities when actual risk is higher than 50%. Consistent with recent systematic reviews, sling type (retropubic or transobturator) is an important predictor of bothersome SUI, stress test, and adverse events 12 months after

midurethral sling placement but not bothersome UUI. Additionally, when specific variables such as urodynamic predictors are available and combined with patient variables, the discriminatory ability of the models improves. This highlights an important distinction for future researchers to make when a study is attempting to predict an outcome compared with when they are investigating whether individual risk factors are associated with an outcome. In prediction, the aim is to build the best predictive model to provide the most accurate estimate to the patient to make a decision. Therefore, less consideration is paid whether one of the risk factors is independently and significantly associated with the outcome; but rather, does the risk factor help improve overall accuracy when all other included factors are known? When making traditional decisions about inference, or cause and effect, one specifically wants to understand the independent relationship between a risk factor and outcome. This is a fundamental difference that explains why some risk factors are or are not present in the models for each outcome.

A recent Cochrane systematic review found moderate quality evidence that up to 1 year after midurethral sling placement for SUI, the rate of subjective cure of transobturator and retropubic approaches is similar with cure rates ranging from 62% to 98% in the transobturator group and from 71% to 97% in the retropubic group.¹ The majority of adverse events were higher using the retropubic approach and included major vascular or visceral injury, postoperative voiding dysfunction, and higher mean operating time, operative blood loss, and length



of hospital stay, whereas rates of groin pain were lower than the transobturator group.¹ Although both routes offer comparable efficacy and many surgeons will choose one type of sling over another depending on the patient's risk profile and tolerance for adverse events, the Cochrane findings highlight the need for a balanced discussion of risks and benefits of each approach and the need to provide more accurate estimates at the individual level. Additionally, most recent American College of Obstetricians and Gynecologists and National Institute for Health and Care Excellence guidelines for the evaluation and treatment of urinary incontinence do not recommend the performance of multichannel filling and voiding cystometry in the preoperative evaluation of women with uncomplicated SUI because additional diagnostic evaluation with this testing compared with basic evaluation does not appear to affect treatment outcomes.^{22,23} However, studies included in these guidelines investigated urodynamic parameters as independent predictors using traditional hypothesis-generating epidemiologic approaches rather than using more powerful predictive analytic methods focused on optimizing predictive accuracy while accounting for important contributing risk factors. Additionally, none of these studies investigated predictive accuracy using combined data sets that have higher statistical power to detect small but important predictive differences. For example, in a separate study using the Trial of Mid-Urethral Slings data set, Nager et al reported that women with low Valsalva leak point pressures are nearly twofold more likely to experience SUI 1 year after surgery but no specific cut point was determined for predicting failure and the study did not use recent recommendations for reporting methods and results of multivariable prediction models for individual prognosis.¹¹

This study makes an important contribution to clinical care and supports existing guidelines by providing tools that surgeons and patients can use to obtain more accurate predictions of important outcomes after surgery for SUI. These prediction models are timely and could aid surgeons given the available evidence and increased scrutiny by health care organizations and national safety advisors. They were developed from and tested in large, multicenter randomized trials with subjective and objective outcome measures in diverse patient populations across the United States. Although the best models could not be externally tested given the lack of consistent predictors collected across the different studies, we were able to rebuild each model using a set of commonly collected predictors and demonstrate

acceptable discrimination 55–66% of the time when actual predictions were than 50% without much loss in accuracy.

This study has limitations in that our models are not perfect. In theory, a mathematical model could have close to perfect predictions (concordance index of 1) if all possible predictors of the outcome are known.² This is not currently possible because only a portion of the information we collect in surgical trials or in clinical care explains the variation in outcomes resulting in models that are conditional on the predictors included in each data set. However, these models advance our current abilities to predict an individual's rate better than providing average event rates from groups of patients in a clinical trial or a range from systematic reviews. Additionally, we were unable to model important distinct adverse event outcomes such as mesh exposure resulting from infrequent occurrence or inconsistent data collection across the studies. We are also unable to separate patients who had a transobturator sling from those who had a minisling placed in the Value of Urodynamic Evaluation data set although this was a relatively small contribution to the external data set as a whole (10/902 patients in external validation only) and therefore should minimally influence that predictor's estimates.¹⁴

Future considerations include updating the models as new predictors become available and testing the models in new larger data sets from different international settings. As new information becomes available about factors that affect these important outcomes after surgery for SUI, predictive analytics can incorporate this information into existing or new models. As predictive tools gain more widespread use in clinical care, they can be included in guidelines and care pathways for consistent clinical use and future testing. Although these prediction models are available to provide valid individualized estimates of a patient having incontinence and an adverse event 12 months after midurethral sling surgery, caution should be used because these models imperfectly explain clinical outcomes requiring surgeons to understand the limitations of each model and recognize that additional factors not included in each model may alter the likelihood of an event occurring.

REFERENCES

1. Ford AA, Rogerson L, Cody JD, Ogah J. Mid-urethral sling operations for stress urinary incontinence in women. The Cochrane Database of Systematic Reviews 2015, Issue 7. Art. No.: CD006375. DOI: 10.1002/14651858.CD006375.pub3.



2. Sniderman AD, D'Agostino RB Sr, Pencina MJ. The role of physicians in the era of predictive analytics. *JAMA* 2015;314:25–6.
3. McGinn T, Jervis R, Wisnivesky J, Keitz S, Wyer PC Evidence-based Medicine Teaching Tips Working Group. Tips for teachers of evidence-based medicine: clinical prediction rules (CPRs) and estimating pretest probability. *J Gen Intern Med* 2008;23:1261–8.
4. 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;123:279–87.
5. Kattan MW, O'Rourke C, Yu C, Chagin K. The wisdom of crowds of doctors: their average predictions outperform their individual ones. *Med Decis Making* 2015 Apr 15. pii: 0272989X15581615 [Epub ahead of print].
6. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol* 2015;68:134–43.
7. Richter HE, Albo ME, Zyczynski HM, Kenton K, Norton PA, Sirls LT, et al. Retropubic versus transobturator midurethral slings for stress incontinence. *N Engl J Med* 2010;362:2066–76.
8. Urinary Incontinence Treatment Network (UITN). The Trial of Mid-Urethral Slings (TOMUS): design and methodology. *J Appl Res* 2008;8. pii: AlboVol8No1.
9. Shumaker SA, Wyman JF, Uebersax JS, McClish D, Fantl JA. Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program in Women (CPW) Research Group. *Qual Life Res* 1994;3:291–306.
10. Barber MD, Spino C, Janz NK, Brubaker L, Nygaard I, Nager CW, et al. The minimum important differences for the urinary scales of the pelvic floor distress inventory and pelvic floor impact questionnaire. *Am J Obstet Gynecol* 2009;200:580.e1–7.
11. Nager CW, Sirls L, Litman HJ, Richter H, Nygaard I, Chai T, et al. Baseline urodynamic predictors of treatment failure 1 year after mid urethral sling surgery. *J Urol* 2011;186:597–603.
12. Nager CW, Albo ME, Fitzgerald MP, McDermott S, Wruck L, Kraus S, et al. Reference urodynamic values for stress incontinent women. *Neurourol Urodyn* 2007;26:333–40.
13. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004;240:205–13.
14. Nager CW, Brubaker L, Litman HJ, Zyczynski HM, Varner RE, Amundsen C, et al. A randomized trial of urodynamic testing before stress-incontinence surgery. *N Engl J Med* 2012;366:1987–97.
15. Nager CW, Brubaker L, Daneshgari F, Litman HJ, Dandreo KJ, Sirls L, et al. Design of the Value of Urodynamic Evaluation (ValUE) trial: A non-inferiority randomized trial of preoperative urodynamic investigations. *Contemp Clin Trials* 2009;30:531–9.
16. Barber MD, Kleeman S, Karram MM, Paraiso MF, Walters MD, Vasavada S, et al. Transobturator tape compared with tension-free vaginal tape for the treatment of stress urinary incontinence: a randomized controlled trial. *Obstet Gynecol* 2008;111:611–21.
17. Barber MD, Weidner AC, Sokol AI, Amundsen CL, Jelovsek JE, Karram MM, et al. Single-incision mini-sling compared with tension-free vaginal tape for the treatment of stress urinary incontinence: a randomized controlled trial. *Obstet Gynecol* 2012;119:328–37.
18. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;30:377–99.
19. Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York (NY): Springer; 2001.
20. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
21. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol* 2010;63:938–9.
22. Evaluation of uncomplicated stress urinary incontinence in women before surgical treatment. Committee Opinion No. 603. American College of Obstetricians and Gynecologists. *Obstet Gynecol* 2014;123:1403–7.
23. Agency for Healthcare Research and Quality. Urinary incontinence: the management of urinary incontinence in women (clinical guideline; no. 171). Available at: <http://www.guideline.gov/content.aspx?id=47099>. Retrieved August 6, 2015.

