Models for Predicting Recurrence, Complications, and Health Status in Women After Pelvic Organ Prolapse Surgery

J. Eric Jelovsek, MD, MMEd, Kevin Chagin, MS, Emily S. Lukacz, MD, MAS, Tracy L. Nolen, DrPh, Jonathan P. Shepherd, MD, MSc, Matthew D. Barber, MD, MHS, Vivian Sung, MD, MPH, Linda Brubaker, MD, MS, Peggy A. Norton, MD, David D. Rahn, MD, Ariana L. Smith, MD, Alicia Ballard, MD, Peter Jeppson, MD, Susan F. Meikle, MD, MSPH, and Michael W. Kattan, PhD, for the NICHD Pelvic Floor Disorders Network

OBJECTIVE: To develop statistical models predicting recurrent pelvic organ prolapse, surgical complications, and change in health status 12 months after apical prolapse surgery.

METHODS: Logistic regression models were developed using a combined cohort from three randomized trials and two prospective cohort studies from 1,301 participants enrolled in surgical studies conducted by the Pelvic Floor Disorders Network. Composite recurrent prolapse was defined as prolapse beyond the hymen; the presence of bothersome bulge symptoms; or prolapse reoperation or retreatment within 12 months after surgery. Complications were defined as any serious adverse event or Dindo grade III complication within 12 months of surgery. Significant change in health status was defined as a minimum important change of SF-6D utility score (±0.035 points) from baseline. Thirty-two candidate risk factors were considered for each model and model accuracy was measured using concordance indices. All indices were internally validated using 1,000 bootstrap resamples to correct for bias.

RESULTS: The models accurately predicted composite recurrent prolapse (concordance index=0.72, 95% CI 0.69–0.76), bothersome vaginal bulge (concordance index=0.73, 95% CI 0.68–0.77), prolapse beyond the hymen (concordance index=0.74, 95% CI 0.70–0.77), serious adverse event (concordance index=0.60, 95% CI 0.56–0.64), Dindo grade III or greater complication (concordance index=0.62, 95% CI 0.58–0.66), and health status improvement (concordance index=0.64, 95% CI 0.62–0.67) or worsening (concordance index=0.63, 95% CI 0.60–0.67). Calibration curves demonstrated all models were accurate through clinically useful predicted probabilities.

CONCLUSION: These prediction models are able to provide accurate and discriminating estimates of pro-

From the Obstetrics, Gynecology & Women's Health Institute and Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio; the Department of Obstetrics & Gynecology, Duke University, Durham, North Carolina; the Department of Reproductive Medicine, UC San Diego Health System, San Diego, California; RTI International, Research Triangle Park, North Carolina; the Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; the Department of Obstetrics and Gynecology Women and Infants Hospital, Providence, Rhode Island; the Department of Obstetrics & Gynecology and Urology, Stritch School of Medicine, Loyola University Chicago, Chicago, Illinois; the Department of Obstetrics and Gynecology University of Utah, Salt Lake City, Utah; the Department of Obstetrics and Gynecology, University of Texas Southwestern, Dallas, Texas; the Division of Urology, University of Pennsylvania, Philadelphia, Pennsylvania; the Department of Obstetrics and Gynecology, University of Alabama at Birmingham, Birmingham, Alabama; Northwest Physician Group, Amarillo, Texas; and the Department of Medicine, University of Texas Health Science Center at Houston, Houston, Texas.

Supported by grants from the Eunice Kennedy Shriver *National Institute of Child Health and Human Development (2U01HD41249, 2U10 HD41250, 2U10 HD41261, 2U10 HD41267, 1U10 HD54136, 1U10 HD54214, 1U10 HD54215, 1U10 HD54241, U10 HD069013, U10 HD069025, U10 HD069010, U10 HD069010, and U01 HD069031) and the National Institutes of Health Office of Research on Women's Health.*

Presented at the Joint American Urogynecologic Society–International Urogynecologic Society Annual Meeting, July 21–26, 2014, Washington, DC.

Each author has indicated that he or she has met the journal's requirements for authorship.

Received February 27, 2018. Received in revised form April 25, 2018. Accepted May 22, 2018.

Corresponding author: J. Eric Jelovsek, MD, MMEd, 209 Baker House, DUMC 3084, Durham, NC 27710; email: eric.jelovsek@duke.edu.

Financial Disclosure

Dr. Brubaker has received honorarium from UpToDate. Emily S. Lukacz has received research grants or support from Boston Scientific (research funding), Uroplasty (research funding), and Pfizer (study drug donation). She has been a consultant to AMS, Axonics, and Renew Medical and has received royalties from UpToDate. The other authors did not report any potential conflicts of interest.

ISSN: 0029-7844/18

lapse recurrence, complications, and health status 12 months after prolapse surgery.

(Obstet Gynecol 2018;132:298–309)
DOI: 10.1097/AOG.0000000000002750

Surgical counseling before reconstructive surgery for pelvic organ prolapse (POP) includes an estimate of benefits and risks. These estimates are influenced by surgeon and patient preferences, surgical approach, and patient risk factors. For example, although current evidence shows that abdominal sacrocolpopexy has superior short-term anatomic outcomes when compared with vaginal colpopexy, this benefit must be balanced against other aspects of the procedure such as longer operating times, increased complications, increased cost, and less well-known long-term outcomes.[1] Typically, surgeons use average rates from large, randomized studies to counsel a woman about her specific risks. However, there are no large clinical trials comparing vaginal and abdominal approaches that report both subjective and objective recurrence and complication rates. Even when level I evidence exists, estimates are traditionally provided during the counseling process without accurately accounting for all unique patient characteristics such as her vaginal topography, medical comorbidities, preferences, and surgical goals. Refined prediction of an individual's perioperative risk weighed against her probable surgical outcome would enhance the presurgical counseling process and aid clinical decision-making.

The Pelvic Floor Disorders Network has conducted four large surgical studies of prolapse surgical treatment with the collection of standardized and validated measures across all clinical sites. Data collected from women enrolled in these trials were used to test the hypotheses that baseline characteristics can predict outcomes after prolapse surgery. We constructed and validated statistical prediction models to calculate patient-specific probabilities of recurrent prolapse, complications, and overall health status 12 months after prolapse surgery.

MATERIALS AND METHODS

Data from 1,301 women who underwent surgical treatment for POP in three randomized trials, including one trial that included a patient-preference cohort, and a second prospective cohort study conducted by the Pelvic Floor Disorders Network were analyzed for this study. Data were combined into a single longitudinal cohort and used to develop and internally validate prediction models. Trials analyzed included the Colpopexy and Urinary Reduction Efforts, Outcomes Following Vaginal Prolapse Repair and Midurethral Sling, and Operations and Pelvic Muscle Training in the Management of Apical Support Loss trials. The two cohort studies included women from the Outcomes Following Vaginal Prolapse Repair and Midurethral Sling trial who chose not to undergo randomization but participated in a patient-preference cohort and a prospective cohort of women undergoing colpocleisis for advanced prolapse (Colpocleisis).[2–5] Each model used all four data sets and was designed to separately predict patient-specific probabilities of developing recurrent prolapse, perioperative and postoperative complications, and change in overall health status 12 months after prolapse surgery. The Colpopexy and Urinary Reduction Efforts (n=322), Outcomes Following Vaginal Prolapse Repair and Midurethral Sling (n=460), Operations and Pelvic Muscle Training in the Management of Apical Support Loss (n=372), and Colpocleisis (n=147) studies conducted by the Pelvic Floor Disorders Network included multiple geographically diverse clinical sites sponsored by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development and the National Institutes of Health Office of Research on Women's Health. Each study received institutional review board approval at all sites and all participants signed informed research consent.

Briefly, the Colpopexy and Urinary Reduction Efforts trial was designed to estimate the rates of de novo stress urinary incontinence (SUI) for stress-continent women undergoing abdominal sacrocolpopexy surgery for prolapse. The intervention evaluated the effectiveness of prophylactic Burch cystourethropexy continence surgery compared with no Burch.[2] The Outcomes Following Vaginal Prolapse Repair and Midurethral Sling trial compared rates of de novo SUI for women who underwent concomitant vaginal prolapse surgery with a retropubic midurethral sling compared with a sham procedure.[3] The Operations and Pelvic Muscle Training in the Management of Apical Support Loss trial compared surgical outcomes after sacrospinous ligament fixation compared with uterosacral vaginal vault suspension in women undergoing vaginal surgery for apical or uterine prolapse with SUI (participants underwent concomitant midurethral sling). The Operations and Pelvic Muscle Training in the Management of Apical Support Loss trial also examined the effects of a structured perioperative program consisting of behavioral techniques and pelvic floor muscle training compared with usual care.[4] The final study, Colpocleisis, was a prospective cohort study designed to determine the effect of

Table 1. Discrimination of the Statistical Models to Predict Recurrent Pelvic Organ Prolapse, Complications, and Overall Health Status After Pelvic Organ Prolapse Surgery

Prediction Model	Outcome of Model	Concordance Index	95% CI
Recurrent POP	Composite definition*	0.72	0.69–0.76
	Bothersome feeling of vaginal bulge	0.73	0.68–0.77
	Prolapse beyond the vaginal hymen	0.74	0.70–0.77
Complications	1 or more serious adverse events	0.60	0.56–0.64
	1 or more Dindo grade III or higher complications[8]	0.62	0.58–0.66
Overall health status	Overall health status improves†	0.64	0.62–0.67
	Overall health status worsens‡	0.63	0.60–0.67

POP, pelvic organ prolapse.

* Composite definition includes: any POP quantification points (Ba, C, or Bp) beyond the hymen 12 months after surgery; the presence of "somewhat," "moderately," or "quite a bit" bothersome bulge symptoms (Pelvic Floor Distress Inventory question 5) 12 months after surgery; or any POP reoperation (or retreatment with pessary) any time up to and including 12 months after surgery.[7]

† Overall health status improves is defined as an increase of at least the minimum clinically important difference (0.035 points or more) in the SF-6D Health Utility Score.[9–11]

‡ Overall health status worsens is defined as a decrease of at least the minimum clinically important difference (0.035 points or less) in the SF-6D Health Utility Score.[9–11]

colpocleisis on pelvic organ support, pelvic symptoms, quality of life, and patient satisfaction. The cohort also aimed to describe the morbidity associated with colpocleisis, patient sexual function and body image, and outcomes with and without concomitant incontinence surgery.[5]

For this report, prediction models were developed using data from participants with 1-year outcomes from the studies. A group of 10 experienced surgeons within the Pelvic Floor Disorders Network identified 32 candidate risk factors that clinicians commonly use to counsel women about the risk of recurrence of prolapse and development of complications during or after surgery. Based on consensus among these surgeons, the following 32 risk factors were selected from all data sets: 1) age; 2) vaginal parity; 3) race (African American, Caucasian, other); 4) cardiac disorder; 5) upper gastrointestinal diagnosis; 6) lower gastrointestinal diagnosis; 7) vascular disorder; 8) history of connective tissue disease (eg, Ehlers-Danlos, Marfan's); 9) smoking history (current, former, never); 10) menopausal status (pre-, post-); 11) current estrogen replacement therapy; 12) anticoagulant use; 13) number of comorbid conditions; 14) prior hysterectomy; 15) prior incontinence surgery; 16) prior surgery for prolapse; 17) body mass index (BMI, calculated as weight (kg)/[height (m)]^2); 18) overall POP quantification (POP-Q) stage; POP-Q points: 19) Ba, 20) Bp, 21) C, and 22) GH; 23) type of anesthesia (general or regional); 24) concurrent anterior colporrhaphy; 25) concurrent posterior colporrhaphy; 26) concurrent total hysterectomy or oophorectomy; 27) concurrent continence procedure; and 28) and type of apical suspension. Additional factors included 29) baseline limitation in activity assessed using the SF-36 question 3, "Does your health now limit you in these activities? If so, how much? Running, lifting heavy objects, or participating in strenuous sports?"[6]; 30) heavy lifting using the question, "During the past month, how often did you perform physical activities that required a major effort such as lifting heavy furniture, shoveling snow, or lifting people or objects weighing more than 25 pounds?" We also measured a 31) history of strenuous exercise defined as "During the past month, on average, on how many days in each week did you do strenuous or very hard exercise; that is, exercise that caused you to work up a sweat and made your heart beat fast. For example, aerobics, dancing, jogging, or tennis?" For the purposes of the model, 32) strenuous physical activity was defined as any positive response to either of the heavy lifting or strenuous exercise questions. A natural log transformation was applied to BMI to improve normality.

Because reported recurrence rates vary widely based on how recurrent prolapse is defined,[7] we planned to build four different recurrence models. First, we modeled outcomes for recurrent prolapse using a composite definition that included individual components of anatomic, symptomatic. and retreatment criteria. The outcome for the composite recurrent prolapse model was considered affirmative if a participant met any one of the three criteria for recurrent prolapse. Then we separately modeled the three individual components of the composite definition. The composite definition of recurrent prolapse was defined as any POP-Q points (Ba, C, or Bp) beyond the hymen 12 months after surgery, the

presence of bothersome bulge symptoms 12 months after surgery (response to Pelvic Floor Distress Inventory question 5 of "somewhat," "moderately," or "quite a bit" of bother), or any prolapse retreatment with surgery or a pessary any time up to and including 12 months after surgery.[7] The three individual components of the composite definition were modeled separately as individual outcomes, because data suggest that absence of vaginal bulge symptoms has a significant relationship with a patient's self-assessment of improvement, whereas anatomic success alone does not.[7]

Because there is no consensus on what constitutes a clinically relevant complication to patients, we built two complication models using two different but commonly used definitions: a participant developing one or more serious adverse events during or any time up to and including 12 months after surgery or a participant developing one or more complications classified as grade III or greater using the Clavien-Dindo scale.[8] A serious adverse event for all studies was considered anything that is fatal, life-threatening, results in initial or prolonged hospitalization, disability or permanent damage, required intervention to prevent permanent impairment or damage, or is otherwise considered by the investigator to be a serious important medical event (ie, an event that may jeopardize the patient and may require treatment to prevent one of the other previously mentioned outcomes). A Clavien-Dindo grade III complication is any surgical complication requiring surgical, endoscopic, or radiologic intervention not under general anesthesia (IIIa) and under general anesthesia (IIIb); a grade IV complication is life-threatening requiring management in an intensive care unit, which can be further subdivided into single organ (IVa) or multiorgan dysfunction (IVb); and a grade V complication is death.[8] The outcomes for the serious adverse event and Dindo grade III or greater models were cumulative in the sense that, if a participant experienced a perioperative serious adverse event or reported one or more serious adverse events at any postoperative visit through 12 months (ie, 3-, 6-, or 12-month visit), she was considered to have an affirmative value for this outcome variable.

Two additional models were built to predict significant improvement and significant worsening of overall health status. Change in overall health status was defined by the minimally important change in utility measures. Utility measures are preference values that patients attach to their overall health status.[9] Because utility values summarize positive and negative effects of an intervention into one value between 0 (equal to death) and 1 (equal to perfect health), health state utilities are commonly used to compare overall health status across different diseases states.[9,10] Because the purpose of our models was to provide estimates of recurrent prolapse (a health state for which patients receiving prolapse surgery are familiar and with which they may identify) and complications (a health state for which patients may not be familiar with and may not appropriately understand how this experience may affect them), we felt that providing an individual's predicted overall health state change may allow a patient to better understand and weigh the risks of their upcoming surgery during the decision-making process. The SF-6D Health State Classification provides a useful tool for estimating a preference-based single index from the SF-12 generic health-related quality-of-life index.[9] Prior studies using more than 20 different patient groups with different disease states have estimated the minimally important clinical difference (improvement or worsening) for the SF-6D utility score to be ± 0.033, ± 0.035, and ± 0.041.[9–11] For the utility models in this study, we chose a cutoff of ± 0.035 (dichotomous) because this was the median of the reported minimally important difference. One model was built to predict the probability of the SF-6D score improving 0.035 or more from baseline and a second was built to predict the probability of the SF-6D score worsening -0.035 or less from baseline.

Multiple models were fit using logistic regression to the risk factors identified by consensus and outcomes from the Pelvic Floor Disorders Network prolapse studies. Multiple imputation using chained equations was performed for missing risk factors.[12] The outcomes for all models were based only on actual and not imputed events. All 32 risk factors were considered as candidate variables for each model. Variable selection was done using Harrell's[13] "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. As a result, 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 adjusted R^2 and was stopped when the bootstrap concordance index had a change of less than 0.01.

Each logistic model's discriminative ability was measured by the area under the curve (AUC) for the receiver operating characteristic curve based on the sensitivity and specificity of the model. An AUC

Table 2. Risk Factors and Their Odds Ratios (95% CIs) in the Statistical Models to Predict Recurrent Pelvic Organ Prolapse, Complications, and Overall Health Status After Pelvic Organ Prolapse Surgery

	Models to Predict Recurrent POP		
Variable	Composite Definition (n=1,059)	Bothersome Vaginal Bulge (n=982)	Prolapse Beyond the Vaginal Hymen (n=1,061)
Age	–	0.98 (0.95–1.01)	–
Vaginal parity	1.117† (1.03–1.21)	1.11 (0.98–1.25)	–
Race	+	+	–
African American	Referent	Referent	–
Caucasian	1.68 (0.78–3.66)	1.07 (0.38–2.98)	–
Other	2.61 (0.95–7.17)	2.71 (0.75–9.82)	–
Cardiac disorder	–	0.53 (0.25–1.12)	–
Upper gastrointestinal disorder	0.594‡ (0.37–0.95)	0.55 (0.29–1.04)	0.66 (0.43–1.02)
Lower gastrointestinal disorder	–	–	–
Vascular disorder	–	–	–
Connective tissue disorder	–	–	–
Current health limits vigorous activities such as running, lifting heavy objects, participating in strenuous sports	–	+	–
Not limited at all	–	Referent	–
Limited a little	–	0.82 (0.41–1.64)	–
Limited a lot	–	1.38 (0.72–2.65)	–
Heavy lifting frequency	+	+	–
Never	Referent	Referent	–
Once/wk	0.91 (0.52–1.59)	1.17 (0.53–2.62)	
More than once/wk	1.18 (0.74–1.88)	1.35 (0.70–2.60)	–
Less than once/mo	1.49 (0.56–3.98)	0.64 (0.07–2.60)	–
Once/mo	1.34 (0.79–2.28)	1.38 (0.65–2.90)	–
2 to 3 times/mo	1.56 (0.94–2.59)	1.52 (0.76–3.05)	–
Smoking status	–	–	–
Current	–	–	–
Former	–	–	–
Never	–	–	–
Estrogen therapy	–	–	–
Anticoagulant use	–	0.37 (0.05–2.99)	–
No. of comorbid conditions	1.08 (0.97–1.21)	1.279‡ (1.10–1.49)	–
Prior hysterectomy	–	–	–
BMI§	–	–	–
POP-Q stage	–	+	+
II	–	Referent	Referent
III	–	0.61 (0.30–1.22)	2.119‡ (1.19–3.78)
IV	–	0.5 (0.12–2.14)	1.8 (0.68–4.73)
POP-Q C	–	–	–
POP-Q Ba	1.183† (1.09–1.28)	1.204‡ (1.02–1.43)	1.177† (1.05–1.31)
POP-Q Bp	–	–	–
POP-Q GH	1.199† (1.05–1.37)	1.16 (0.96–1.42)	1.219† (1.07–1.39)
Anesthesia type	–	2.18 (0.68–7.03)	–
Concurrent anterior repair	0.648‡ (0.43–0.94)	0.563‡ (0.33–0.98)	0.575† (0.38–0.86)
Concurrent posterior repair	0.618‡ (0.43–0.90)	0.523‡ (0.31–0.89)	0.622‡ (0.43–0.91)
Concurrent hysterectomy or oophorectomy	–	0.67 (0.38–1.19)	–
Concurrent continence procedure	–	–	–
Burch	–	–	–
Sling or TVT	–	–	–
Other	–	–	–
None	–	–	–
Vault suspension repair type‖	+	+	+
Abdominal sacrocolpopexy	Referent	Referent	Referent
Uterosacral ligament suspension	9.443† (5.28–16.90)	10.283† (4.18–25.28)	11.111† (6.19–19.94)
Sacrospinous ligament suspension	9.443† (5.28–16.90)	10.283† (4.18–25.28)	11.111† (6.19–19.94)
Other	9.443† (5.28–16.90)	10.283† (4.18–25.28)	11.111† (6.19–19.94)
None	10.160† (3.42–30.17)	7.027‡ (1.57–31.53)	6.893† (1.69–28.05)
Colpocleisis	1.13 (0.51–2.53)	0.49 (0.09–2.64)	1.77 (0.88–3.55)

POP, pelvic organ prolapse; BMI, body mass index; POP-Q, pelvic organ prolapse quantification; TVT, tension-free vaginal tape; +, factor present in the final model; –, factor not present in the final model.

* The outcomes were dichotomous. The improving model outcome was the probability of the SF-6D score improving 0.035 or more from baseline vs less than 0.035 from baseline and the worsening outcome model was the probability of the SF-6D score worsening −0.035 or less vs more than −0.035 from baseline.

† Significant at .01.

‡ Significant at .05.

§ Natural log transformation was performed.

‖ One participant received a uterosacral ligament suspension and sacrospinous ligament suspension and was classified as receiving a uterosacral ligament suspension for this analysis.

Models to Predict Complications		Models to Predict Overall Health Status Using Health Utilities*	
1 or More Serious Adverse Events (n=1,301)	1 or More Dindo Grade 3 or Higher Complications (n=1,301)	Overall Health Status Improves (n=1,118)	Overall Health Status Worsens (n=1,118)
1.02 (0.99–1.03)	–	0.974† (0.96–0.99)	1.021‡ (1.00–1.04)
–	–	–	–
+	–	–	+
Referent	–	–	Referent
0.69 (0.39–1.23)	–	–	0.527‡ (0.29–0.95)
0.57 (0.22–1.48)	–	–	0.46 (0.29–0.95)
–	1.37 (0.87–2.16)	–	–
–	–	–	–
0.598‡ (0.37–0.97)	0.57 (0.32–1.01)	–	0.75 (0.46–1.23)
0.78 (0.55–1.12)	–	–	–
–	2.79 (0.83–9.41)	0.35 (0.11–1.11)	–
–	–	+	+
–	–	Referent	Referent
–	–	2.415† (1.68–3.47)	0.333† (0.22–0.50)
–	–	3.030† (2.11–4.36)	0.219† (0.14–0.33)
–	+	–	+
–	Referent	–	Referent
	0.96 (0.55–1.70)		1.07 (0.65–1.76)
–	1.07 (0.65–1.75)	–	1.10 (0.69–1.73)
–	0.69 (0.29–1.68)	–	0.82 (0.37–1.84)
–	0.53 (0.25–1.12)	–	0.92 (0.52–1.60)
–	0.91 (0.51–1.61)	–	1.608‡ (1.00–2.58)
–	+	+	+
–	Referent	Referent	Referent
–	0.68 (0.36–1.30)	1.68 (0.96–2.92)	0.57 (0.29–1.12)
–	0.467‡ (0.25–0.88)	1.47 (0.86–2.51)	0.71 (0.37–1.35)
0.74 (0.53–1.02)	0.73 (0.50–1.09)	–	–
–	–	–	–
1.1 (0.99–1.22)	1.10 (0.98–1.23)	0.919† (0.85–0.99)	1.09 (0.99–1.21)
0.74 (0.52–1.06)	0.63 (0.36–1.11)	–	–
–	–	0.438‡ (0.21–0.93)	0.45 (0.18–1.15)
+	–	–	–
Referent	–	–	–
1.18 (0.85–1.63)	–	–	–
0.88 (0.56–1.40)	–	–	–
–	–	0.97 (0.93–1.01)	1.03 (0.98–1.08)
–	–	–	0.95 (0.87–1.04)
–	–	1.058‡ (1.01–1.11)	–
–	–	1.09 (0.99–1.19)	–
1.59 (0.84–2.99)	–	1.815‡ (1.03–3.20)	–
–	1.28 (0.82–2.00)	–	–
–	–	–	–
0.72 (0.52–1.06)	0.4509† (0.25–0.80)	–	1.31 (0.90–1.90)
–	+	+	–
–	Referent	Referent	–
–	1.14 (0.52–2.50)	0.99 (0.56–1.77)	–
–	3.61 (0.77–16.85)	0.177‡ (0.03–0.95)	–
–	1.40 (0.76–2.58)	0.81 (0.50–1.31)	–
+	+	+	+
Referent	Referent	Referent	Referent
0.428† (0.29–0.62)	0.378† (0.19–0.74)	1.684‡ (1.07–2.65)	0.94 (0.63–1.41)
0.473† (0.30–0.74)	0.443† (0.21–0.95)	1.805‡ (1.06–3.09)	0.83 (0.50–1.38)
0.89 (0.31–2.58)	0.74 (0.21–2.65)	1.7 (0.61–4.78)	0.84 (0.22–3.17)
0.212† (0.07–0.63)	0.147† (0.04–0.59)	1.97 (0.89–4.37)	0.81 (0.32–2.04)
0.326† (0.18–0.59)	0.417† (0.19–0.92)	1.08 (0.61–1.92)	1.22 (0.66–2.25)

value closer to 1 indicates a better prediction of the outcome and an AUC value of 0.5 indicates that the model predicts no better than chance. The AUC is also a representation of the concordance index and measures the model's ability to generate a higher predicted probability of the outcome occurring in a patient who has a worse outcome. For example, if we have a pair of patients, in which one patient has recurrent prolapse and the other does not, the concordance index measures the model's ability to assign a higher risk to the patient with recurrent prolapse. All concordance indices and receiver operating

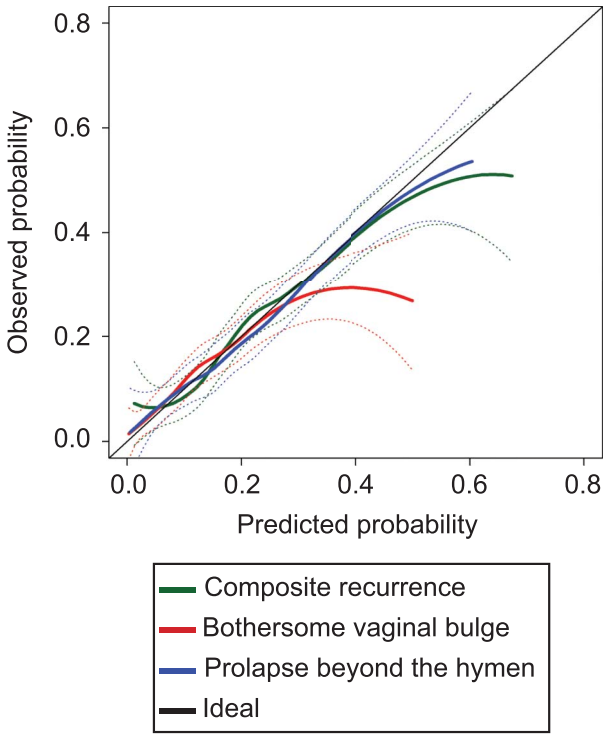


Fig. 1. Calibration curves for three models predicting probability of developing recurrent pelvic organ prolapse (POP) 1 year after surgery for POP.

Jelovsek. Predicting Outcomes After POP Surgery. Obstet Gynecol 2018.

characteristic curves were internally validated using a 1,000 bootstrap resample 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.[14] Calibration curves were also 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.

After the best model was selected and internally validated, the composite recurrence and serious adverse event complication models were compared with the best currently available method of estimating risk, that is, an expert clinician's predictions. To perform these comparisons, a subset of 50 participants was randomly selected from all four data sets for comparing the probability of developing recurrent prolapse and the probability of developing one or more serious adverse events between the model and the panel of experts. These 50 participants were used to compare predictions of the models with experts' predictions and not as a true independent validation subset. Both models were rebuilt using the remaining participants in the four data sets excluding the 50 randomly selected participants. The preoperative candidate risk factors of these 50 participants were given to 20 "expert" surgeons with representation from each of the Pelvic Floor Disorders Network sites for review resulting in 1,000 expert predictions and 50 model predictions for each outcome. All surgeons were considered to be experienced in treating patients with prolapse. Each of the 20 experts was asked to consider each woman's data from all 32 variables among the 50 randomly selected patients and provide their best estimated outcome by answering the following question: "Out of 100 women with these exact characteristics, estimate the number of women with recurrent prolapse 12 months after surgery for prolapse and estimate the number of women with one or more serious adverse events during or any time up to and including 12 months after surgery." Individual clinicians' predictions were not averaged to yield a single value because incorporating each clinician's predictions substantially increased statistical power. Each model's predictions were compared with the experts' predictions, which included all risk factors, to determine which was most accurate. The difference in accuracy was determined by using a bootstrap method from their respective receiver operating characteristic curves. All analyses were performed using R 2.15.2.

RESULTS

Eight prediction models were initially explored: four to predict recurrent prolapse (composite recurrence, anatomic recurrence, symptom recurrence, and retreatment), two to predict complications, and two to predict overall change in health status. Of 1,301 women who underwent prolapse surgery, 1,263 received a vault suspension and 38 did not receive a vault prolapse repair. A total of 1,059 of 1,301 participants were used to predict recurrent prolapse using the composite definition because 242 patients were missing one or more variables (Appendix 1, available online at http://links.lww.com/AOG/B124). A total of 982 participants were used to predict bothersome vaginal bulge because 319 did not report the outcome at 12 months and a total of 1,061 was used to predict prolapse beyond the vaginal hymen because 240 were missing the examination outcome.

Overall, 209 (20%) experienced composite recurrent prolapse (met one or more of the three definitions of recurrence) including 201 with prolapse beyond the vaginal hymen, 95 who reported a bothersome vaginal bulge, and 23 who underwent retreatment for prolapse. Three of the four recurrence models had good discrimination with all bias-corrected concordance indices above 0.70 (Table 1). We were unable

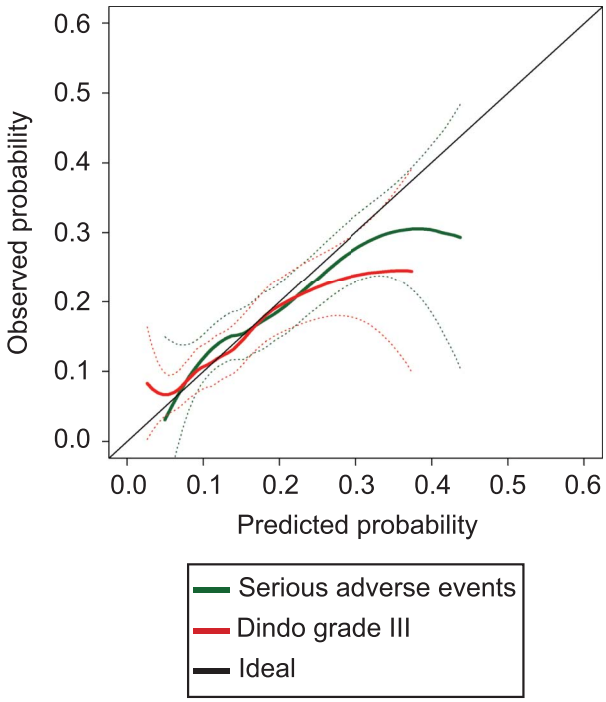


Fig. 2. Calibration curves for two models predicting probability of developing one or more serious adverse events or Dindo grade III or higher complications 1 year after surgery for pelvic organ prolapse.

Jelovsek. Predicting Outcomes After POP Surgery. Obstet Gynecol 2018.

to construct an accurate model to predict reoperation or retreatment of prolapse within a year after surgery as a result of a low number of events. Table 2 demonstrates the baseline characteristic factors included in each model along with their parameter estimates. The composite recurrence model included 10 factors, the vaginal bulge model included 17 factors, and prolapse beyond the hymen model included seven factors. Concurrent anterior and posterior repair each substantially reduced the risk of recurrence in all three models. Factors that increased risk of recurrence included vaginal vault suspension (uterosacral or sacrospinous) compared with abdominal sacrocolpopexy and greater POP-Q examination points Ba and GH.

Figure 1 displays the calibration curves for the three recurrence models. The composite recurrence and prolapse beyond the hymen models had accurate predictions along a range of recurrence probabilities. The bulge model had a tendency to overestimate probabilities of approximately 20–30% or more. It is important to note that although the model for predicting bulge overestimates when observed probabilities exceed 30%, this only occurred in a very small subset (5%) of the population.

All 1,301 participants were used in the creation of the two complications models. Overall, 222 (17%) participants experienced any serious adverse event and 147 (11%) experienced one or more Dindo grade III or higher complications with 147 (11%) participants having both. Both models had acceptable discrimination with all bias-corrected concordance indices 0.60 or more (Table 1). The serious adverse event model included 11 factors and the Dindo grade III or higher complication model included 17 factors. Increasing age and number of comorbidities were associated with increased risk; factors that substantially lowered the risk of serious adverse events or Dindo grade III or higher complications in both models included presence of a lower gastrointestinal disorder, presence of a vascular disorder, current estrogen use of any type, having a concurrent hysterectomy or oophorectomy, and vaginal vault suspension or colpocleisis (vs abdominal sacrocolpopexy). Figure 2 displays the calibration curves. Both complication models had accurate predictions along a range of probabilities with a slight tendency to underestimate probabilities less than 10% and overestimate probabilities of approximately 30–40% or more in the serious adverse event model and 20–30% or higher in the Dindo grade III model. These extreme ends for over- and underestimation occur in a small subset (Dindo grade III or higher greater than 30%, 2% and serious adverse event greater than 40%, 0.8%) of the data. Comparisons of each individual risk factor between patients with and without composite recurrent prolapse, serious adverse events, and Dindo grade III or greater complications are presented in Appendices 1–3, available online at http://links.lww.com/AOG/B124.

Of the 1,301 participants, 1,118 were used in the creation of the health status models because 183 participants had missing health status data. Both health status models had discrimination between women who reported that their overall health status did or did not meaningfully improve or worsen with all bias-corrected concordance indices above 0.60 (Table 1). Table 2 displays which risk factors were included in each model along with their parameter estimates. The overall health status improvement model included 12 factors and the overall health status worsening model also included 12 somewhat different factors. Figure 3 displays the calibration curves for both overall health status models. The overall health status improvement model had accurate predictions at probabilities of 40% or higher but had a tendency to underestimate probabilities below 40%. The overall health status worsening model had a tendency to

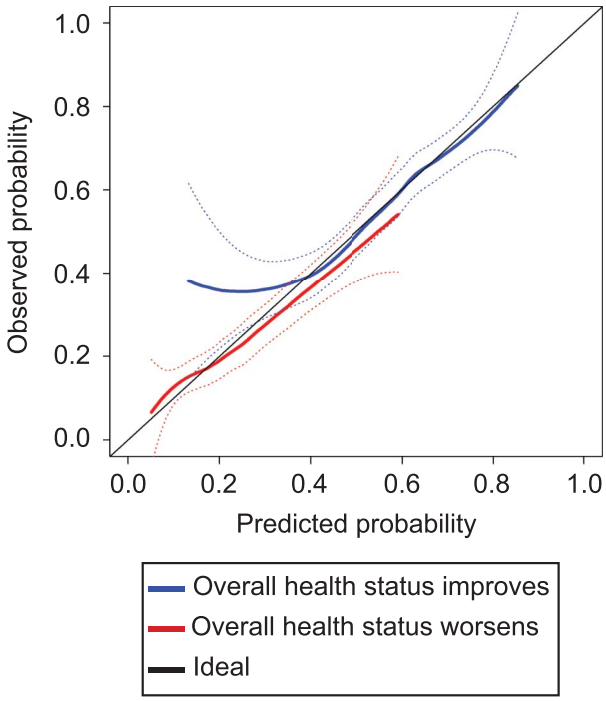


Fig. 3. Calibration curves for two models predicting probability of overall health status improvement and worsening from baseline 1 year after surgery for pelvic organ prolapse.
Jelovsek. Predicting Outcomes After POP Surgery. Obstet Gynecol 2018.

overestimate probabilities of 40–50% or more. Both extreme values only encompassed a small subset of the data set (health status worsening more than 50%, 2.6% and health status improving less than 40%: 13%).

The composite recurrence model was statistically better at predicting the risk of recurrent prolapse (Fig. 4A and B) when compared with predictions by experts (concordance index=0.77 vs 0.58, $P<.001$). The serious adverse event model was also better at predicting the risk of a patient developing any serious adverse event (Fig. 5A and B) than predictions by experts but was not significantly different (concordance index=0.58 vs 0.48, $P=.305$).

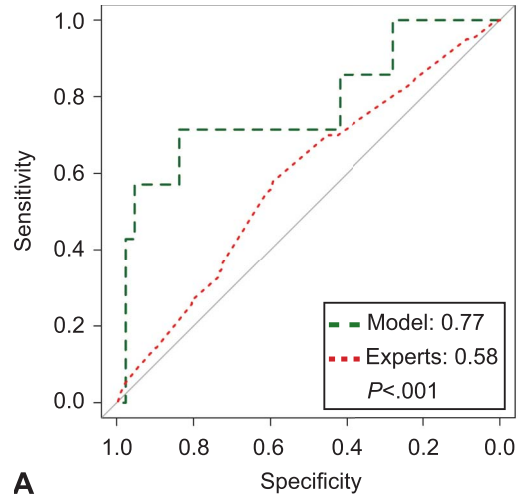
DISCUSSION

Although prolapse surgery success rates are an important factor in determining the best surgical approach, deciding on which type of repair is a complex process that must go beyond just success rates and also take into account the individual patient and surgeon's goals balanced against acceptability of risks and complications. We believe that using predictions from our new models combined with surgeon judgment is more accurate than using crude risk groups or experience alone. We are confident that the level of rigor, size, and multicenter design of the studies used to build these models included data on the most common risk factors for recurrence, complications, and health status change. As such, unmeasured confounders are likely minimized. We hope this tool can augment the surgical consent process for health care providers and their patients (http://riskcalc.org:3838/PRECISE_Models/).

The models developed in this study support the idea that selection of abdominal sacrocolpopexy and concurrent use of anterior and posterior colporrhaphies substantially reduced recurrence risk–by all definitions of recurrence–compared with vaginal suspensions. However, all vaginal surgeries were associated with less risk of serious adverse events or Dindo grade III or higher complications. Predictably, larger anterior prolapse (Ba) and genital hiatus measures overall increased risk of recurrence in all models, whereas increased age and comorbidities increased complication risk across both models. Because individual patients may have difficulty interpreting or quantifying the effect of an adverse event on their lives, we predicted whether overall health state would meaningfully improve or worsen after surgery. By a small degree, transvaginal uterosacral and sacrospinous ligament suspensions increased the likelihood of clinically meaningful overall health status improvement compared with abdominal sacrocolpopexy.

The interactive models are simple to use and outperformed subspecialist "experts" in accuracy of predictions for both success and complications. Subspecialist experts predicted the composite recurrence risk for a random sample of 50 participants with a concordance index of 0.58 suggesting that an expert's prediction is limited. The prediction model performed better in this sample with a concordance index of 0.77 ($P<.001$). Interestingly, the experts' ability to predict the risk of developing any serious adverse event was no better than chance, with a concordance index of 0.48, whereas the model's concordance index was 0.58. The concordance indices are similar to other instruments and nomograms commonly used in clinical practice such as those for predicting prostate cancer[15] or pancreatic adenocarcinoma recurrence[16] or the Gail model of breast cancer risk prediction[17] and Framingham coronary heart disease prediction scoring.[18]

There are some limitations to the utility of these predictive models. First, the patients undergoing abdominal sacrocolpopexy underwent laparotomies without the inclusion of minimally invasive approaches, and there was not a surgical cohort



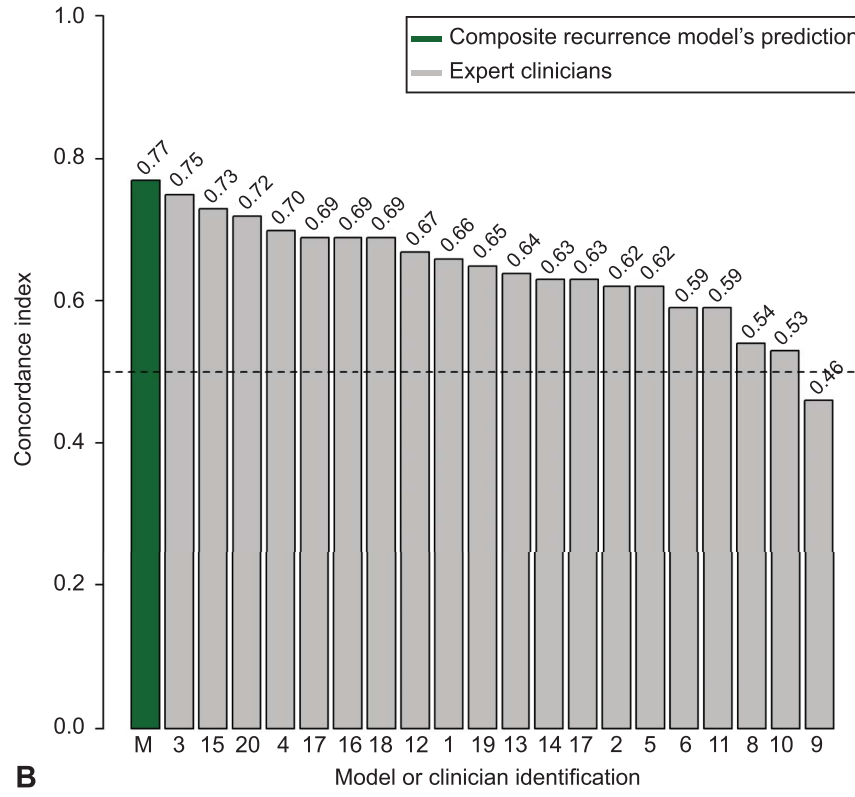
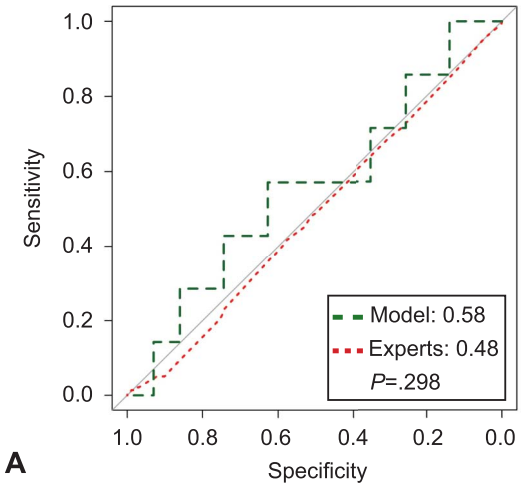


Fig. 4. Comparing accuracy of the composite recurrence model's predictions to all expert predictions (**A**) using receiver operating characteristic curves and each expert (**B**) using the concordance index for 50 random patients selected from the Colpopexy and Urinary Reduction Efforts, Outcomes Following Vaginal Prolapse Repair and Midurethral Sling, and Operations and Pelvic Muscle Training in the Management of Apical Support Loss trials and the prospective cohort study Colpocleisis.

Jelovsek. Predicting Outcomes After POP Surgery. Obstet Gynecol 2018.

undergoing transvaginal repair augmented by graft material, so one cannot predict recurrence risk or likelihood of complications for these approaches. Another limitation is that the models do not include outcomes more than 12 months after surgery because collection beyond this time point was not performed uniformly across all four studies. Importantly, the models should be externally validated once public data sets become available and can be recalibrated to include minimally invasive sacrocolpopexy and mesh-augmented vaginal repairs if large data sets or longer outcomes become available. We also could not predict likelihood of retreatment or reoperation alone because of low numbers of these events. Additionally, the standardized method of adverse event reporting, whether serious adverse events or Dindo classification, may not have captured important specific adverse outcomes such as chronic pain, dyspareunia,



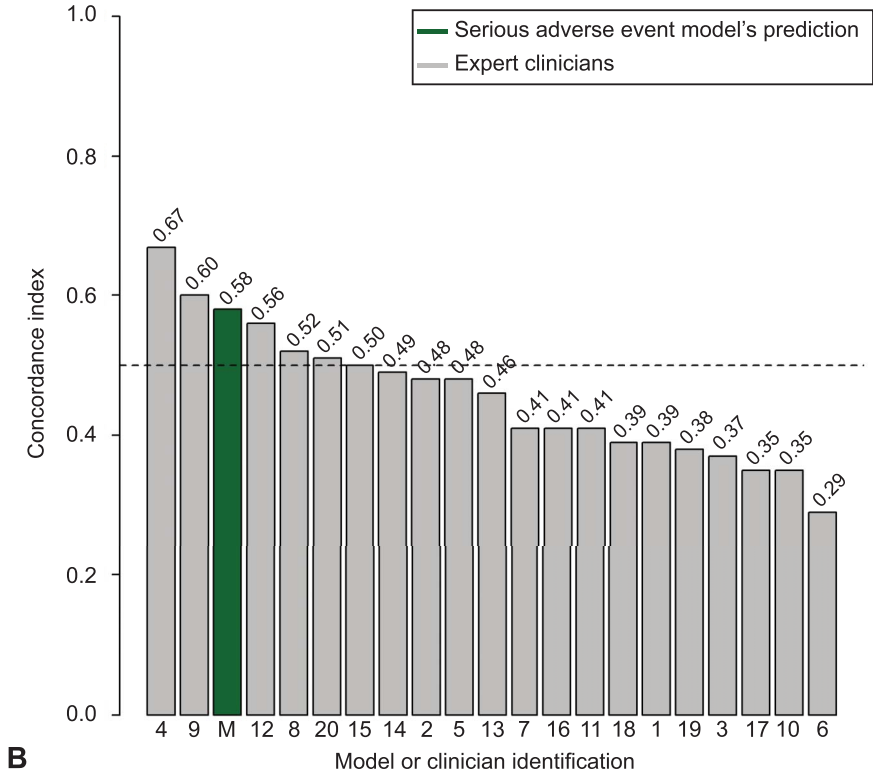


Fig. 5. Comparing accuracy of the serious adverse event model's predictions to all expert predictions (**A**) using receiver operating characteristic curves and each expert (**B**) using the concordance index for 50 random patients selected from the Colpopexy and Urinary Reduction Efforts, Outcomes Following Vaginal Prolapse Repair and Midurethral Sling, and Operations and Pelvic Muscle Training in the Management of Apical Support Loss trials and the prospective cohort study Colpocleisis.

Jelovsek. Predicting Outcomes After POP Surgery. Obstet Gynecol 2018.

or other events that do not require surgical or invasive management. Furthermore, given the broad range of potential adverse events, we were only able to capture system-based categories (eg, upper gastrointestinal disorders included "the presence of any disorder of the esophagus, stomach, duodenum, biliary, or pancreatic tract"). Finally, the models only present an overall probability of complications and some patients may weigh certain specific complications more or less heavily than others. Caution should also be used in assuming that predictors are the same as risk factors. Rather, the variables should be thought of as "predictors" of the outcome when controlling for all other predictors in the model rather than cause-and-effect associations. The primary strengths stem from building models from several large, multicenter, prospective surgical trials with well-characterized populations using standardized, validated postoperative outcomes 12 months after surgery. Predicting overall health status by a minimally important difference adds another

useful dimension beyond prolapse recurrence or complication occurrence.

In summary, these prediction models are able to provide accurate and discriminating estimates of prolapse recurrence, complications, and health status 12 months after prolapse surgery.

REFERENCES

1. Maher C, Feiner B, Baessler K, Schmid C. Surgical management of pelvic organ prolapse in women. The Cochrane Database of Systematic Reviews 2013, Issue 4. Art. No.: CD004014. DOI: 10.1002/14651858.CD004014.pub5.
2. Brubaker L, Cundiff GW, Fine P, Nygaard I, Richter HE, Visco AG, et al. Abdominal sacrocolpopexy with Burch colposuspension to reduce urinary stress incontinence. N Engl J Med 2006; 354:1557–66.
3. Wei JT, Nygaard I, Richter HE, Nager CW, Barber MD, Kenton K, et al. A midurethral sling to reduce incontinence after vaginal prolapse repair. N Engl J Med 2012;366:2358–67.
4. Barber MD, Brubaker L, Burgio KL, Richter HE, Nygaard I, Weidner AC, et al. Comparison of 2 transvaginal surgical approaches and perioperative behavioral therapy for apical vaginal prolapse: the OPTIMAL randomized trial. JAMA 2014; 311:1023–34.
5. Fitzgerald MP, Richter HE, Bradley CS, Ye W, Visco AC, Cundiff GW, et al. Pelvic support, pelvic symptoms, and patient satisfaction after colpocleisis. Int Urogynecol J Pelvic Floor Dysfunct 2008;19:1603–9.
6. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63.
7. Barber MD, Brubaker L, Nygaard I, Wheeler TL 2nd, Schaffer J, Chen Z, et al. Defining success after surgery for pelvic organ prolapse. Obstet Gynecol 2009;114:600–9.
8. 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.
9. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42:851–9.
10. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14:1523–32.
11. Gandhi PK, Ried LD, Bibbey A, Huang IC. SF-6D utility index as measure of minimally important difference in health status change. J Am Pharm Assoc (2003) 2012;52:34–42.
12. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30:377–99.
13. Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York (NY): Springer; 2001.
14. 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.
15. Stephenson AJ, Scardino PT, Eastham JA, Bianco FJ Jr, Dotan ZA, DiBlasio CJ, et al. Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. J Clin Oncol 2005;23:7005–12.
16. Brennan MF, Kattan MW, Klimstra D, Conlon K. Prognostic nomogram for patients undergoing resection for adenocarcinoma of the pancreas. Ann Surg 2004;240: 293–8.
17. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. J Natl Cancer Inst 2001;93:358–66.
18. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P; CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001;286:180–7.