Development of Decision Support Formulas for the Prediction of Bladder Outlet Obstruction and Prostatic Surgery in Patients With Lower Urinary Tract Symptom/Benign Prostatic Hyperplasia: Part II, External Validation and Usability Testing of a Smartphone App

Min Soo Choo[1], Seong Jin Jeong[2], Sung Yong Cho[3], Changwon Yoo[4], Chang Wook Jeong[5], Ja Hyeon Ku[5], Seung-June Oh[5]

[1]Department of Urology, Hallym University Dongtan Sacred Heart Hospital, Hwaseong, Korea
[2]Department of Urology, Seoul National University Bundang Hospital, Seongnam, Korea
[3]Department of Urology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Korea
[4]Department of Biostatistics, Robert-Stempel College of Public Health & Social Work, Florida International University, Miami, FL, USA
[5]Department of Urology, Seoul National University Hospital, Seoul, Korea

Purpose: We aimed to externally validate the prediction model we developed for having bladder outlet obstruction (BOO) and requiring prostatic surgery using 2 independent data sets from tertiary referral centers, and also aimed to validate a mobile app for using this model through usability testing.

Methods: Formulas and nomograms predicting whether a subject has BOO and needs prostatic surgery were validated with an external validation cohort from Seoul National University Bundang Hospital and Seoul Metropolitan Government-Seoul National University Boramae Medical Center between January 2004 and April 2015. A smartphone-based app was developed, and 8 young urologists were enrolled for usability testing to identify any human factor issues of the app.

Results: A total of 642 patients were included in the external validation cohort. No significant differences were found in the baseline characteristics of major parameters between the original (n = 1,179) and the external validation cohort, except for the maximal flow rate. Predictions of requiring prostatic surgery in the validation cohort showed a sensitivity of 80.6%, a specificity of 73.2%, a positive predictive value of 49.7%, and a negative predictive value of 92.0%, and area under receiver operating curve of 0.84. The calibration plot indicated that the predictions have good correspondence. The decision curve showed also a high net benefit. Similar evaluation results using the external validation cohort were seen in the predictions of having BOO. Overall results of the usability test demonstrated that the app was user-friendly with no major human factor issues.

Conclusions: External validation of these newly developed a prediction model demonstrated a moderate level of discrimination, adequate calibration, and high net benefit gains for predicting both having BOO and requiring prostatic surgery. Also a smartphone app implementing the prediction model was user-friendly with no major human factor issue.

Keywords: Decision Support Systems, Clinical; Nomograms; Validation Studies; Predictive Values of Tests; Mobile Apps

Corresponding author: Seung-June Oh http://orcid.org/0000-0002-0322-3539
Department of Urology, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea
E-mail: sjo@snu.ac.kr / Tel +82-2-2072-2406 / Fax: +82-2-742-4665
Submitted: March 22, 2017 / **Accepted after revision:** April 6, 2017

INTRODUCTION

Lower urinary tract symptoms (LUTS)/benign prostatic hyperplasia (BPH) is not a life-threatening malignant disease. Treatment practices are generally not standardized within a center because decision-making may vary according to patients' preferences and the subjective judgment of the surgeon [1,2]. Bladder outlet obstruction (BOO) is one of the most important components to assess in patients with LUTS/BPH. However, it is difficult to implement urodynamic studies to diagnose BOO routinely in all patients [3].

In our previous study, we developed a nomogram to predict having BOO and requiring prostatic surgery based on urodynamically determined BOO using our urodynamics database that we have developed over the last 13 years (since 2004) [4]. These prediction models are expected to help support physicians in decision-making and in informing patients of their risk of requiring prostatic surgery. As a next step, it is necessary to confirm the transportability of the model into different but related populations. The successful implementation of prediction models in clinical practice generally requires validation of their performance [5]. The successful estimation of model performance is needed to make judgments regarding model reproducibility and model transportability [6].

Clinical prediction models are commonly developed to facilitate diagnostic or prognostic probability estimations in daily medical practice [6]. A user-friendly modality is necessary for the use of such formulas and nomograms to become a widespread part of routine practice. However, any such app or software should be validated through usability testing among potential end-users. In this study, we aimed to externally validate the prediction model for both having BOO and requiring prostatic surgery using 2 independent data sets from tertiary referral centers and to validate a mobile app for using this model through usability testing.

MATERIALS AND METHODS

Subjects and Established Prediction Model

A retrospective review of medical and surgical records was performed for 2,544 consecutive patients who had undergone a urodynamic study for LUTS between January 2004 and April 2015 at 2 tertiary referral centers: Seoul National University Bundang Hospital (SNUBH) and Seoul Metropolitan Government-Seoul National University Boramae Medical Center

(SMG-SNU BMC). The study design and the use of patients' information stored in the hospital database were approved by the Institutional Review Board of SNUBH (B-1410-272-404) and SMG-SNU BMC (26-2014-99). This study was also approved by the Seoul National University Hospital (SNUH) (H-1406-119-591).

All datasets of the development cohort at SNUH and the external validation cohort at SNUBH and SMG-SNU BMC were constructed in the same manner. Patients with LUTS due to the following causes were excluded: urethral stricture, bladder stone, genitourinary infection or inflammation, previous genitourinary surgery, genitourinary radiation, urinary diversion, genitourinary malignancy, or a neurologic condition. The original prediction model was developed including age, the voiding and storage subscores of the International Prostate Symptom Score (IPSS) questionnaire, the quality of life (QoL) item of the IPSS, maximal flow rate (Qmax) of free flowmetry, postvoid residual (PVR) volume, total prostate volume (TPV) assessed through transrectal ultrasonography, and the BOO index (BOOI) from a pressure-flow study (PFS). All procedures of the urodynamic study were conducted in accordance with the standardization of the International Continence Society and following the same protocol with the same urodynamic instrument (UD-2000 or Solar, Medical Measurement System, Enschede, the Netherlands) [7]. We excluded any patients missing at least one of the IPSS items and uroflowmetry parameters. We also excluded cases with a voided volume of less than 120 mL in free flowmetry.

Details regarding the development of the prediction formulas and nomograms have been presented in our previous study [4]. Briefly, using multivariate logistic regression analysis, 2 formulas, including age, Qmax, and TPV, were developed for calculating the probability of having BOO, with 1 formula for when TPV is available and 1 for when TPV is not available. A total of 4 formulas, including age, IPSS scores, the IPSS QoL score, TPV, and BOOI, were generated for calculating the probability of prostatic surgery, with 1 formula for when all the above variables are available, and the other 3 formulas for cases where TPV and/or BOOI are missing. These formulas and nomograms were validated in the external validation cohort.

External Validation

In the demographic comparison between the development and the external validation cohorts, the Student t-test and the chi-square test were used. We assessed the sensitivity, specificity, positive predictive value (PPV), and negative predictive value

(NPV) of the prediction model on the external validation cohort. In addition, the predictive accuracy of this model was quantified using a receiver operating characteristic curve, and summarized using the area under the curve (AUC) and the Hosmer-Lemeshow goodness-of-fit test. A calibration plot was used to obtain a graphical representation of the extent of overestimation or underestimation of those who actually required surgery versus those for whom the nomogram predicted that prostatic surgery would be necessary. A decision curve analysis was performed to assess the value of the prediction model in terms of net benefit according to different probability thresholds.

Mobile App

An app named BPH Probability Calculator was developed for smartphones and tablet devices for the Android (Google, Mountain View, CA, USA) and iOS (Apple, Cupertino, CA, USA) platforms, as well as a web-based PC version (accessible at bph.snu.ac.kr) with the assistance of Neozensoft, Inc. (Seoul, Korea). Only the smartphone app was validated in this study.

When the app is started, the startup screen appears. It automatically switches to the disclaimer screen. There are a few warnings about using the app; after agreeing, the user can go to the main screen. This app provides 2 interrelated probabilities for having BOO and requiring prostatic surgery. The user is expected to provide the patient's information to use the main functions of the app. The probability of having BOO is calculated from the clinical parameters of age, Qmax, and PVR, with or without TPV. The probability of prostatic surgery is predicted using age, Qmax, and PVR, with or without TPV and/or BOOI. When BOOI is not available, it can still be calculated by the predicted BOO probability obtained from the first part. Age, Qmax, and PVR are mandatory input items for a prediction to be generated, so these are marked with '*' to indicate that they are required input fields. If any of these items are missing, the calculate button is not activated. If an excessively large number out of the normal range is entered, an error massage is displayed. The clear button is at the bottom of the screen and resets all the values at once. There is an 'OK' button in the input window, making it easier to move to the next item. After the user enters the required values and presses the calculate button, this app calculates the probabilities of having BOO or requiring prostatic surgery through the use of a formula. The answer is displayed as a percentage of predicted probability. The 95% confidence interval (CI) of the predicted probability is also provided for reference.

For the convenience of conversion between the 2 interrelated calculators, a button labeled 'move to BPH surgery' was made to jump directly to the calculator for prostatic surgery for a given value of probability of having BOO. If this button is used, the information previously entered in the calculator for having BOO is also automatically entered into the calculator for prostatic surgery. A mailing button at the top of the page was created so that the results can be received directly via e-mail. The smartphone app does not collect any protected health information. The provided information is not saved; it is used only for calculation.

Usability Testing

After we developed this app, usability testing was conducted between November 2015 and February 2016 at the Medical Device Usability Testing Center of the Biomedical Research Institute, SNUH. The usability testing was conducted in an environment similar to a typical clinic office: 150 lux lighting, 32 dB noise, 25°C temperature, 36.4% relative humidity, and a flat floor. To further simulate the environment of a real outpatient clinic, participants were free to communicate verbally while using the smartphone device. According to a predetermined scenario, the usability test comprised 3 phases: (1) identification of usability issues, (2) a usability questionnaire, and (3) face-to-face interviews by interviewers.

Eight young urologists were invited for usability testing to evaluate the BOO probability calculator and the prostatic surgery probability calculator, and to identify if there were any human factor issues. All subjects provided informed consent for inclusion before they participated in the study. Since all participants were very familiar with Android or iOS smartphones, there was no need for prior education on the smartphone operating system. Based on the usability test plan, 1 examiner and 1 observer led the test according to the procedure that had been developed prior to the test. Participants were asked to use both the Korean and English versions. Along with allowing a free exploration of the app, participants were asked to complete several semistructured tasks, such as calculating the probabilities, entering an excessively large number outside of the normal range, and checking the response of the app to a missing value. The participants were encouraged to communicate with staff via a think-aloud approach during the study.

After evaluating the app, each participant was asked to fill out a questionnaire, which included Likert-scale response options to assess the app, with the following 4 options: 1, very dif-

Table 1. Results of the questionnaire administered in the usability testing

Detailed evaluation item	No. of participants responding (score)				Mean ± SD
	1	2	3	4	
BOO probability calculator					
1. Starting and stopping the app	0	0	3	5	3.6 ± 0.5
2. Recognition of the pop-up guide window	0	1	4	3	3.2 ± 0.7
3. Conversion between Korean and English	0	1	3	4	3.3 ± 0.7
4. Ease of viewing references	0	1	6	1	3.0 ± 0.5
5. Entering numbers for each indicator (age, Qmax, PVR, and TPV)	0	0	5	3	3.3 ± 0.5
6. Entering the number for each indicator (IPSS)	0	0	3	5	3.6 ± 0.5
7. Recognition of the BOO probabilities	0	0	5	3	3.3 ± 0.5
8. Switching to the surgery probability calculator	0	0	3	5	3.6 ± 0.5
9. Ease of initialization	0	0	2	6	3.7 ± 0.4
10. Ease of transferring the results externally	0	1	6	1	3.0 ± 0.5
Probability of prostatic surgery					
11. Entering the number for each indicator (IPSS)	0	0	1	7	3.8 ± 0.3
12. Entering numbers for each indicator (age, Qmax, PVR, TPV)	0	0	6	2	3.2 ± 0.4
13. Recognition of BPH surgery probabilities	0	0	4	4	3.5 ± 0.5
14. Ease of transferring the results externally	0	0	6	2	3.2 ± 0.4
Other general usability					
15. App buttons and icon configuration	0	0	3	5	3.6 ± 0.5
16. Visual design of app	0	0	4	4	3.5 ± 0.5
17. Warning about missing required items	0	1	5	2	3.1 ± 0.6
18. Warning about data entry range	0	1	4	3	3.2 ± 0.7
19. Knowing if the indicators are reflected in the probability output	2	3	2	1	2.2 ± 1.0
20. Ease of input data modification	0	2	3	3	3.1 ± 0.8

SD, standard deviation; BOO, bladder outlet obstruction; Qmax, maximal flow rate; PVR, postvoid residual volume; TPV, total prostate volume; IPSS, International Prostate Symptom Score.

ficult; 2, difficult; 3, easy; and 4, very easy. The questionnaire consisted of a total of 20 questions and subjective opinions, of which 5 were on the BOO probability calculator, 8 on the prostatic surgery calculator, and 7 on the overall convenience of use (Table 1). In the subjective evaluation section, participants described their subjective opinions narratively and made proposals for improving the ease of use.

Statistical Analysis

Numerical or categorical variables are presented as the mean ± standard deviation or as absolute numbers with percentages. Two-sided P-values <0.05 were considered to indicate statistical significance, and 95% CIs were calculated. The statistical analysis was performed with IBM SPSS Statistics ver. 23.0 (IBM

Co., Armonk, NY, USA) and the R statistical package system (R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org).

The criterion for passing the usability test was more than 90% of items having an average value of 3 points or more. The results of each item in the usability test were reported as median and range.

RESULTS

Original Development Cohort Versus External Validation Cohort

A total of 642 patients (417 from SNUBH, 225 from SMG-SNU BMC) were included in this external validation study. There were

Table 2. Demographics of the overall patients and the patients who were determined to require prostatic surgery in the original group versus the validation group

Variable	Total			Surgery decision		
	Original (n = 1,179)	Validation (n = 642)	P-value	Original (n = 744)	Validation (n = 158)	P-value
Age (yr)	66.1 ± 7.2	66.1 ± 8.5	0.959	66.8 ± 6.8	67.1 ± 6.2	0.651
IPSS total	18.0 ± 7.7	18.8 ± 8.1	0.036	18.6 ± 7.4	20.0 ± 7.2	0.089
IPSS voiding subscore	10.9 ± 5.4	11.3 ± 5.6	0.112	11.5 ± 5.2	12.4 ± 5.4	0.119
IPSS storage subscore	7.1 ± 3.4	7.5 ± 3.7	0.032	7.1 ± 3.4	7.6 ± 3.4	0.186
IPSS QoL score	4.0 ± 1.2	3.9 ± 1.2	0.392	4.1 ± 1.1	4.0 ± 1.1	0.772
TPV	47.5 ± 23.2	39.0 ± 22.9	<0.001	54.0 ± 24.3	52.5 ± 27.6	0.552
TZV	23.1 ± 18.1	18.9 ± 18.8	<0.001	27.9 ± 19.6	28.8 ± 25.6	0.752
PSA	2.7 ± 3.5	3.9 ± 5.5	<0.001	3.2 ± 3.7	5.4 ± 6.4	0.001
Qmax	11.8 ± 4.8	14.0 ± 8.5	<0.001	10.9 ± 4.3	12.9 ± 5.0	<0.001
VV	229.1 ± 99.0	240.2 ± 104.6	0.026	222.8 ± 89.6	238.6 ± 109.7	0.165
PVR	52.2 ± 64.3	62.5 ± 84.7	0.008	59.2 ± 66.7	76.4 ± 111.3	0.129
First desire	194.5 ± 77.5	202.9 ± 86.7	0.045	192.6 ± 75.1	191.2 ± 89.8	0.883
Normal desire	276.5 ± 96.4	264.0 ± 111.0	0.02	274.1 ± 96.9	251.2 ± 104.6	0.027
Strong desire	374.1 ± 107.1	325.4 ± 99.2	<0.001	366.0 ± 106.6	303.5 ± 102.8	<0.001
MCC	378.3 ± 113.3	375.7 ± 104.3	0.626	372.3 ± 111.1	359.8 ± 106.1	0.283
IDC	289 (24.5)	204 (31.8)	<0.001	184 (24.7)	32 (31.1)	0.018
PdetQmax	52.0 ± 20.2	47.5 ± 27.5	<0.001	57.2 ± 21.6	63.1 ± 26.4	0.033
PFS Qmax	9.1 ± 4.7	11.1 ± 6.7	<0.001	8.4 ± 4.6	10.0 ± 6.2	0.016
BOOI	33.2 ± 23.2	25.8 ± 30.7	<0.001	39.7 ± 24.1	43.1 ± 28.1	0.252
BCI	97.6 ± 26.8	103.2 ± 40.0	0.002	99.3 ± 27.3	113.1 ± 39.9	0.001

Values are presented as mean ± standard deviation.
IPSS, International Prostate Symptom Score; voiding subscore, sum of questions 1, 3, 5, and 6 of the IPSS; storage subscore, sum of questions 2, 4, and 7 of the IPSS; TPV, total prostate volume; TZV, transitional zone volume; PSA, prostate-specific antigen; Qmax, maximal flow rate; VV, voiding volume; PVR; postvoid residual volume; MCC, maximal cystometric capacity; IDC, involuntary detrusor contraction; PdetQmax, detrusor pressure at the maximal flow rate; PFS, pressure flow study; BOOI, bladder outlet obstruction index; BCI, bladder contractility index.

18 patients (2.8%) without BOOI, 22 patients (3.4%) without TPV, and 2 patients (0.3%) without either. No significant difference in age was found between the patients in the original group and the validation group. However, the patients in the validation group had smaller prostate volumes (47.5 ± 23.2 mL vs. 39.0 ± 22.9 mL, P<0.001) and lower BOOI scores (33.2 ± 23.2 vs. 25.8 ± 30.7, P<0.001) than the patients in original group (Table 2).

Decision Versus Actual Surgery
In original development cohort, 744 of the 1,179 patients (63.1%) were judged as requiring prostatic surgery, and 627 patients (84.3%) actually underwent surgery. In the external validation cohort, 158 patients (25.2%) were judged as requiring prostatic surgery, and 111 patients (70.3%) actually underwent surgery.

Among the patients determined to require prostatic surgery, no significant difference was found in the baseline characteristics of major parameters between the original and validation groups, except Qmax (10.9 ± 4.3 vs. 12.9 ± 5.0, P<0.001). The TPV was likewise not significantly different between the 2 cohorts among the patients determined to require prostatic surgery (54.0 ± 24.3 mL vs. 52.5 ± 27.6 mL). The BOOI was slightly higher in the external validation cohort, but this difference was not statistically significant (39.7 ± 24.1 vs. 43.1 ± 28.1).

External Validation of the Accuracy of the Nomograms
The prediction model of BOO in the external validation cohort showed a sensitivity of 24.0%, a specificity of 97.8%, a PPV of 72.0%, and an NPV of 84.4%. In the BOO prediction model
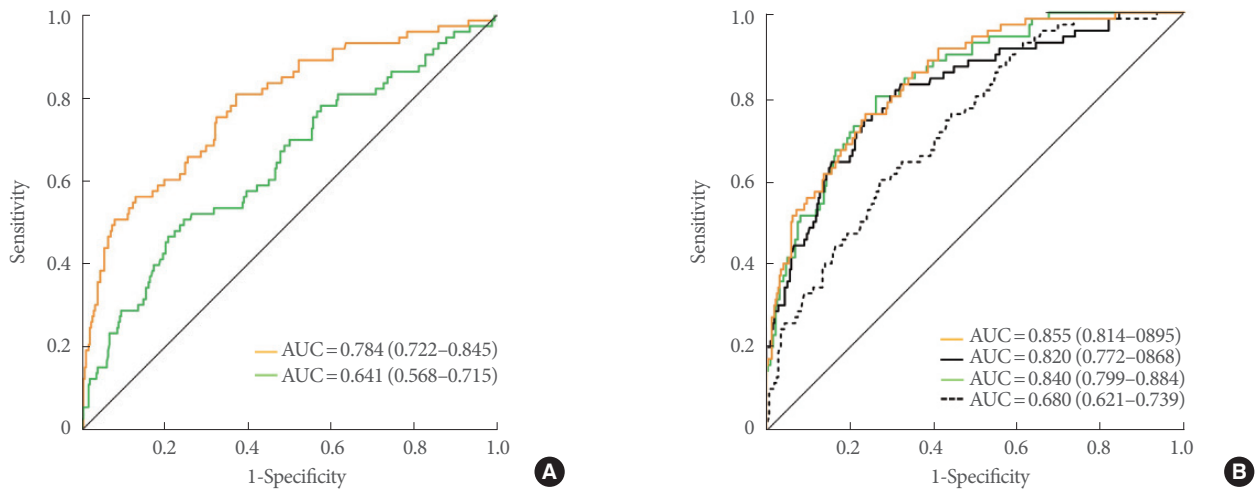
Fig. 1. The area under the curve (AUC) of receiver operating characteristic curves for nomograms predicting (A) urodynamic bladder outlet obstruction; the orange line indicates the nomogram with all parameters, and the green line indicates the nomogram without the bladder outlet obstruction (BOO) index; and (B) the need for prostatic surgery; the orange line indicates the nomogram with all parameters, the black line indicates the nomogram without the BOO index, the green line indicates the nomogram without total prostate volume, and the black dotted line indicates the nomogram without the BOO index and total prostate volume.
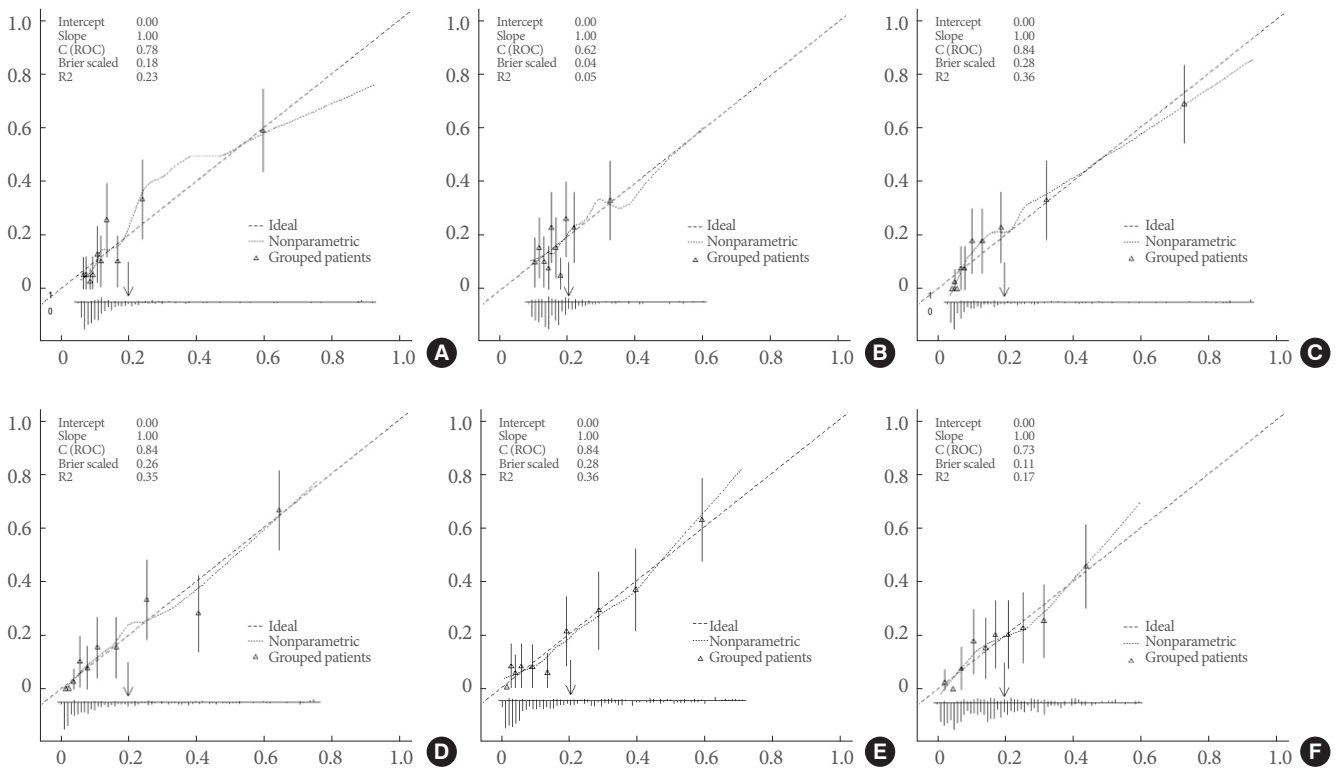

Fig. 2. Calibration plot for nomograms predicting (A) urodynamically confirmed bladder outlet obstruction (BOO) with all parameters, (B) urodynamically confirmed BOO without total prostate volume, (C) the need for prostatic surgery with all parameters, (D) the need for prostatic surgery without the BOO index, (E) the need for prostatic surgery without total prostate volume, and (F) the need for prostatic surgery without the BOO index and total prostate volume. ROC, receiver operating characteristic.
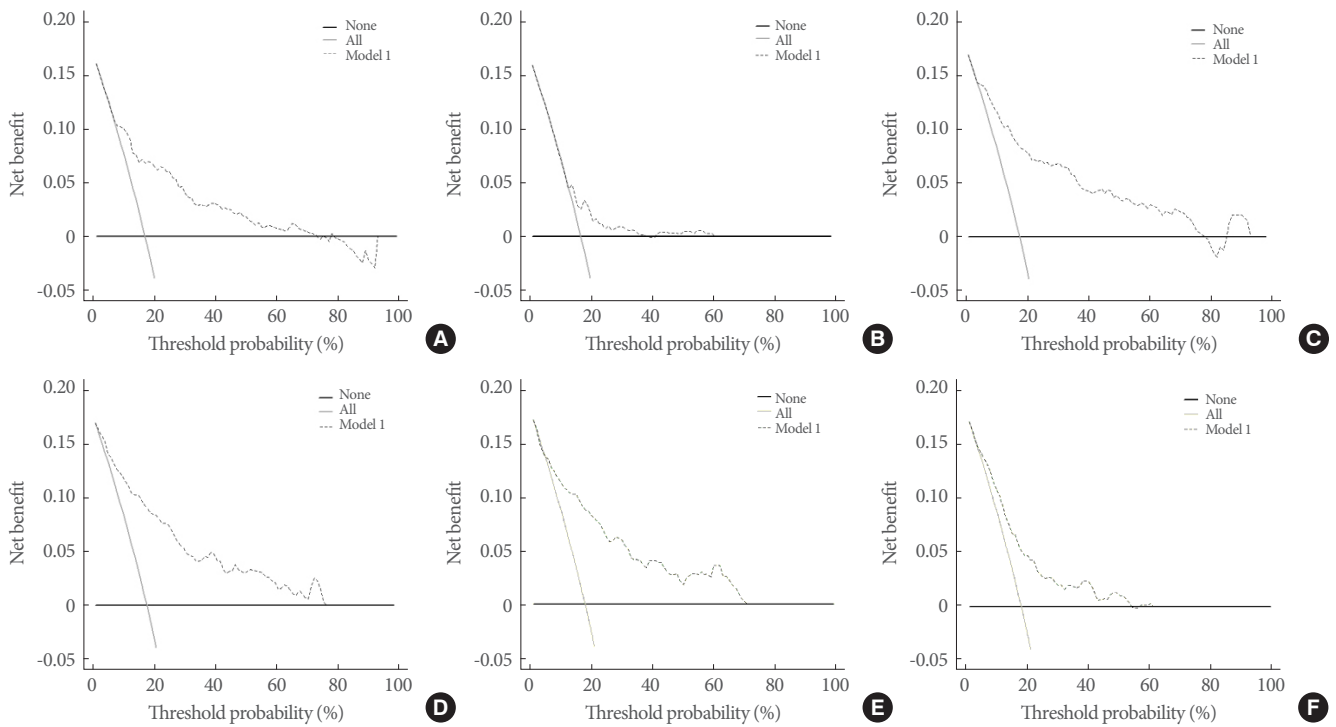
Fig. 3. Decision curve analysis for nomograms predicting (A) urodynamically confirmed BOO with all parameters, (B) urodynamically confirmed bladder outlet obstruction (BOO) without total prostate volume, (C) the need for prostatic surgery with all parameters, (D) the need for prostatic surgery without the BOO index, (E) the need for prostatic surgery without total prostate volume, and (F) the need for prostatic surgery without the BOO index and total prostate volume.

without TPV, the sensitivity was 11.5%, the specificity was 96.9%, the PPV was 47.4%, and the NPV was 82.1%. The model that incorporated all the variables had an AUC of 0.785 (95% CI, 0.725–0.845), while the model without TPV had an AUC of 0.648 (95% CI, 0.576–0.721) (Fig. 1). The calibration plot showed a good agreement between the expected and observed rates in both models (Fig. 2). In addition, the decision curve showed also a high net benefit across the entire spectrum of probability thresholds in both models (Fig. 3).

Similar results were seen in the prediction model of prostatic surgery. In the external validation cohort, the prediction model of prostatic surgery showed a sensitivity of 80.6%, a specificity of 73.2%, a PPV of 49.7%, and an NPV of 92.0%. The model showed an AUC of 0.84 (95% CI, 0.838–0.913) (Fig. 1B). The calibration plot demonstrated that this prediction model showed a good correspondence between the predicted probability of prostatic surgery and the actual rate of requiring surgery (Fig. 2C). The overall model performance was excellent, with a scaled Brier score of 0.28 and an explained variance (R^2) of 0.36. In addition, the decision curve showed also a high net benefit across

the entire spectrum of probability thresholds (Fig. 3C).

Usability Testing

Fig. 4 shows the user interface of the app. The mean age of the participants was 35.0 ± 4.2 years (range, 28–41 years), and they were male urology residents or fellows with average work experience of 4.6 ± 2.0 years (range, 2–8 years) in urological patient care. None had hearing or visual impairments.

On the questionnaire survey, all 20 items (100%) exceeded 2 points (difficult) on average. The average score of the items on the questionnaire was 3.3, with a standard deviation of 0.4 (Table 1). The questionnaire satisfied the previously established passing criterion that 90% or more items had an average value of 3 points or more.

The score of question 19 (Q19) was the lowest, with an average of 2.2 points. Q19 was a question about recognizing whether the results corresponded to having all parameters present or were based on missing information. The overall data of this usability test demonstrated that the app was user-friendly, and no major human errors occurred in its use.
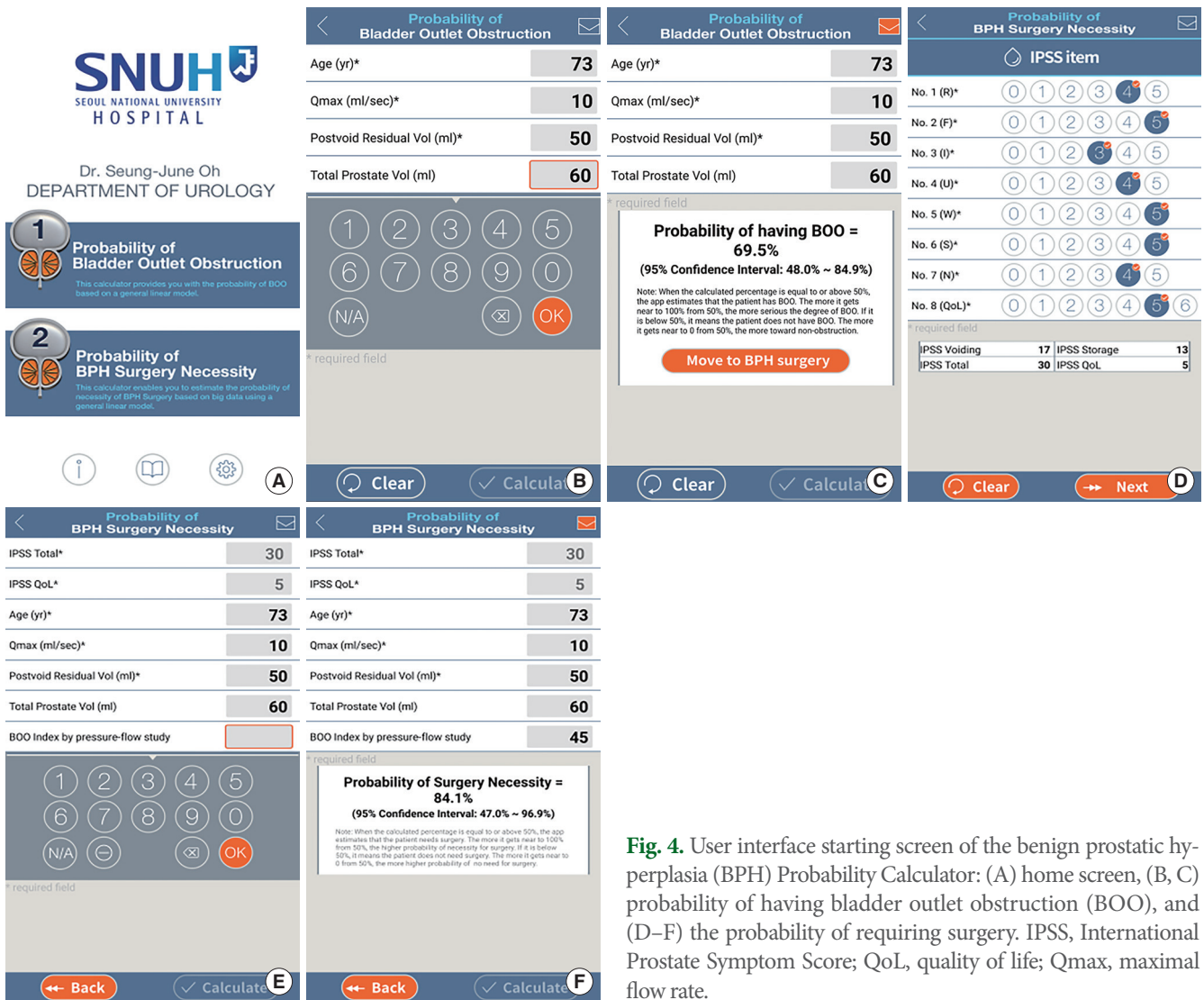
Fig. 4. User interface starting screen of the benign prostatic hyperplasia (BPH) Probability Calculator: (A) home screen, (B, C) probability of having bladder outlet obstruction (BOO), and (D–F) the probability of requiring surgery. IPSS, International Prostate Symptom Score; QoL, quality of life; Qmax, maximal flow rate.

DISCUSSION

In this study, we externally validated the overall accuracy of previously proposed prediction models for BOO and prostatic surgery. These prediction models showed an excellent discriminant ability in the independent patient cohort. In addition, the models were well calibrated and had high net benefits. They could be used to support physicians in decision-making and in informing patients of their risk of undergoing prostatic surgery in routine clinical use among the general population.

In this external validation study, the prediction model for requiring prostatic surgery showed high values for accuracy, with a sensitivity of 80.6% and a specificity of 73.2%. The validity of the test is measured by sensitivity and specificity [8]. This is be-

cause the PPV and NPV depend directly on the prevalence of the disease in the population. Assuming all other factors remain constant, NPVs will increase as prevalence decreases [8]. In the present study, the NPVs were very high (92.0%). This means that the rate of patients requiring BPH surgery in the validation group was lower than in the original group among patients who underwent a urodynamic study.

The calibration plot for requiring prostatic surgery showed very high agreement. A decision curve analysis also showed high net benefits at all probability thresholds. Our predictive models were analyzed to determine whether they made accurate predictions of whether prostatic surgery was required, even without TPV or BOOI. These models are expected to be of great help for surgical decision-making in real clinical practice.

However, in the external validation group, there were not many patients with a high probability, i.e., 70%–80% or more, in case of absence of BOOI or prostate volume, we should pay attention to the interpretation of these results. It is necessary to add more cases and revise the nomogram in the near future. The external validity of the prediction model of BOO was high, with an AUC of 79.5%. However, it decreased to 64.8% when TPV was absent. The specificity of the prediction model of BOO was 97.8%, and that of BOO without TPV also had a specificity of 96.7%. In contrast, the sensitivity of the prediction model for BOO was relatively low, with or without TPV, with values of only 24.0% and 11.5%. Several studies have explored the correlation between BOO and TPV. Although TPV can be an important factor in causing BOO, no strict relationship has been found between these parameters [9]. It has been reported that not only prostate volume, but also morphological and/or functional changes, such as intravesical protrusion, the urethral prostatic angle, and the ratio of the transitional zone volume, affect BOO [10-12]. The combination of high specificity and low sensitivity for the prediction of BOO in this study indicates that age, Qmax, PVR, and TPV are the minimum necessary factors that cause BOO, while other factors may act indirectly.

The calibration plot of the prediction model of BOO showed that the overall accuracy was acceptable, although it displayed a tendency towards underestimation at low predicted probabilities and overestimation at high predicted probabilities. This means that conservative judgments are required when using these models to predict BOO in high-probability patients with a very large prostate or a very low Qmax. This was also confirmed by the decision curve analysis. The net benefit changed to a negative value at a probability of approximately 80%. In the BOO prediction model without TPV, although the calibration plot did not deviate much from the ideal line, the decision curve analysis showed that the net benefit was not high for all thresholds. Although the prostate volume is not strictly related to BOO, it can be inferred that predicting BOO without TPV is not accurate. In the case of the absence of a PFS for various reasons, obtaining information on the TPV may have clinical benefits for predicting BOO.

Many algorithms or formulas have been developed to support medical professionals in the diagnosis or management of certain diseases or conditions [13]. However, directly applying such models in daily practice is not an easy task. A mobile or PC app can be used to support physicians in decision-making about patients with specific conditions. Algorithms or formulas

can support medical professionals in the diagnosis or management of certain diseases or conditions. We first developed formulas for calculating the probabilities of both BOO and requiring prostate surgery, with the goal of creating a smartphone-based mobile health app and PC-based software to facilitate the widespread use of the formulas. Using such technologies is advantageous compared to conventional paper-based methods. First, data entry and computation are very efficient and fast. Second, such methods are easier, simpler, and more user-friendly. Therefore, the number of health care mobile apps predicting risk factors or prognoses is rapidly increasing. However, an app should be tested and validated among potential end-users before it comes into use in clinical practice [14,15]. If the user interface of the software is not well constructed, user errors can occur, which may lead to potentially serious consequences in the ultimate decision-making process. Therefore, usability testing is as important as developing the formulas.

Our study has a few limitations. First, differences were found in the baseline characteristics of patients between the original development cohort and the external validation cohort. However, no significant differences were found in the clinical parameters of patients who required BPH surgery. It is suggested that this reflects the environmental variation across referral centers, which differ in terms of the patient population, disease severity, socioeconomic status, and/or proximity to primary or secondary referral centers. Patients with more severe symptoms requiring surgery may be referred to an institution similar to the institution where the development cohort was created. It was also observed that the percentages of patients who received PFS and required surgery were 71.8% and 24.7%, respectively. However, no differences were found in the major parameters of patients who required surgery, such as BOOI, TPV, total IPSS, and the IPSS voiding subscale score. This means that patients who were referred to the 3 centers differed in their clinical characteristics, but that the decision about prostatic surgery was made on the basis of similar criteria in these 3 independent institutions. Therefore, it seems that there was no problem regarding the criteria for surgical decision-making in the development cohort.

Second, in this study, a usability test was performed among a relatively small number of participants at the end of the user-interface development. Additionally, the PC version of the software was not tested. There are 2 different types of usability tests in medical device development: formative and summative usability tests. A formative test is conducted during the develop-

ment of the medical device to identify product strengths and shortcomings. The present test was performed as a summative test, where there should be at least 25 users to reveal subtle interaction issues that might not be identified in a test involving fewer participants [16]. However, during the development of the apps, the authors circulated and reviewed development prototype versions more than 10 times before the usability test. Therefore, we believe that most user errors and interface errors were corrected before the test.

In conclusion, external validation of the newly developed probability models demonstrated a moderate level of discrimination, adequate calibration, and high net benefits for predicting both having BOO and requiring surgery. Moreover, usability testing showed that the smartphone app was user-friendly and that no major human errors occurred in its use.

ACKNOWLEDGMENTS

REFERENCES

1. Piao S, Choo MS, Kim M, Jeon HJ, Oh SJ. Holmium laser enucleation of the prostate is safe for patients above 80 years: a prospective study. Int Neurourol J 2016;20:143-50.

2. Kim M, Ramirez L, Yoo C, Choo M, Paick JS, Oh SJ. Factors influencing nonabsolute indications for surgery in patients with lower urinary tract symptoms suggestive of benign prostatic hyperplasia: analysis using causal bayesian networks. Int Neurourol J 2014;18: 198-205.

3. Jeon HJ, Choo MS, Oh SJ. The effect of posture and repetition on urodynamic parameters: A prospective randomized study. Investig Clin Urol 2017;58:34-41.

4. Choo MS, Yoo C, Cho SY, Jeong SJ, Jeong CW, Ku JH, et al. Development of decision support formulas for the prediction of bladder outlet obstruction and prostatic surgery in patients with lower urinary tract symptom/benign prostatic hyperplasia: Part I, development of the formula and its internal validation. Int Neurourol J

2017:21(Suppl 1):S55-65.

5. Lindholm B, Nilsson MH, Hansson O, Hagell P. External validation of a 3-step falls prediction model in mild Parkinson's disease. J Neurol 2016;263:2462-9.

6. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015;68:279-89.

7. Abrams P, Cardozo L, Fall M, Griffiths D, Rosier P, Ulmsten U, et al. The standardisation of terminology of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. Neurourol Urodyn 2002;21:167-78.

8. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 2008;56:45-50.

9. Rosier PF, de la Rosette JJ. Is there a correlation between prostate size and bladder-outlet obstruction? World J Urol 1995;13:9-13.

10. Chia SJ, Heng CT, Chan SP, Foo KT. Correlation of intravesical prostatic protrusion with bladder outlet obstruction. BJU Int 2003; 91:371-4.

11. Ku JH, Ko DW, Cho JY, Oh SJ. Correlation between prostatic urethral angle and bladder outlet obstruction index in patients with lower urinary tract symptoms. Urology 2010;75:1467-71.

12. Kaplan SA, Te AE, Pressler LB, Olsson CA. Transition zone index as a method of assessing benign prostatic hyperplasia: correlation with symptoms, urine flow and detrusor pressure. J Urol 1995;154: 1764-9.

13. Vashist SK, Schneider EM, Luong JH. Commercial smartphone-based devices and smart Apps for personalized healthcare monitoring and management. Diagnostics (Basel) 2014;4:104-28.

14. Cruz Zapata B, Hernández Niñirola A, Idri A, Fernández-Alemán JL, Toval A. Mobile PHRs compliance with Android and iOS usability guidelines. J Med Syst 2014;38:81.

15. Wiklund M KJ, Strochlic A. Risk management and usability testing. In: Wiklund M, Kendler J, Strochlic A, editors. Usability testing of medical devices. 2nd ed. Boca Raton (FL): CRC Press; 2016. p. 37-51.

16. Wiklund M KJ, Strochlic A. Types of tests. In: Wiklund M, Kendler J, Strochlic A, editors. Usability testing of medical devices. 2nd ed. Boca Raton (FL): CRC Press; 2016. p. 111-34.