From Prostate Health to Overactive Bladder: Developing a Crosswalk for the IPSS to OAB-V8

Robert Trafford Crump, Anika Sehgal, Ian Wright, Kevin Carlson, and Richard Baverstock

OBJECTIVES To develop a statistical model to facilitate the comparison of 2 common patient-reported outcome (PRO) instruments in male lower urinary tract symptoms.

METHODS Two PROs used by urologists are the International Prostate Symptoms Severity (IPSS) and the Overactive Bladder-Validated 8 questions (OAB-V8). The former measures symptoms related to prostate cancer, the latter measures the severity of symptoms related to OAB. Ordinary least squares regression was used to develop 3 models for translating responses to the IPSS into OAB-V8 scores. The root mean square error was used to compare the models.

RESULTS The sample consisted of 493 participants, ranging from 18 to 93 years of age. The recommended model included the individual responses to the IPSS' items and participants' age. Due to the low root mean square error (0.7606), indicating low variation and high precision, we can explain about 63% (R2 = 0.6260) relationship between IPSS and OAB-v8.

CONCLUSION This study successfully modeled global OAB-V8 scores from IPSS responses. This model performed comparably well to others developed using similar methods. UROLOGY 125: 73–78, 2019. © 2018 Elsevier Inc.

Patient-reported outcomes (PROs) are increasingly being used in a clinical setting to help inform medical decisions.[1] PROs are a measurement of any aspect of a patient's health status that comes directly from the patient without any interpretation by the clinician or any other professional.[2] A more practical definition of PROs is that they are survey instruments that allow patients to self-report their symptom severity, functional impairments, and quality-of-life.[3] Clinicians can use PROs to systematically collect information from patients with which they can monitor and track symptoms, assess the effects of an intervention, or inform a treatment decision.[4] Using PROs in a clinical setting has also demonstrated to engage patients, improve patient–clinician communication, and have patients' perspectives included in the clinical decision-making process.[3]

There are 2 types of PROs, generic and condition-specific. Generic PROs are used to measure general states of health through several domains thought to be relevant to one's overall health such as mobility, pain, and mental health.[5] The EuroQoL's EQ-5D is an example of a generic PRO.[6] Generic PROs are generally used to assess the health of a population, compare outcomes across disparate interventions or conditions, or in economic evaluations where quality-adjusted life years are needed as a measure.[5]

Condition-specific PROs are used to measure the health of a specific organ system or the severity of a specific ailment.[7] From a clinical perspective, condition-specific PROs are more useful than generic PROs because they are more sensitive to change in particular symptoms or functions.[8] Some condition-specific PROs have been found to be as accurate as laboratory tests at diagnosing a problem.[9]

There are however limitations to condition-specific PROs. For example, there are often multiple PRO instruments developed for a specific condition. These instruments will undoubtedly use different questions, response formats, and scoring algorithms; they may even measure different domains of health related to the condition all together. This can make it difficult to compare responses from different PROs, even those that are aimed at the same condition. Such is the case in overactive bladder (OAB), where we estimate there are no fewer than 5-10 relevant PRO instruments.

OAB is a prevalent condition defined by urinary urgency and frequency with or without urgency incontinence.[10] This condition can cause individuals to severely limit their daily activities.[10] The direct and indirect consequences of this condition can lead to reduced independence, quality of life, and confidence.[10] Furthermore, OAB is highly prevalent

Financial Disclosure: *The authors declare that they have no relevant financial interests.*

From the Department of Surgery, University of Calgary, Calgary, Alberta, Canada; and the VESIA [Alberta Bladder Centre], Calgary, Alberta, Canada

Address correspondence to: *R. Trafford Crump, Ph.D., Department of Surgery, University of Calgary, 6601 7007 14 Street SW, Calgary, Alberta, Canada. E-mail: tcrump@ucalgary.ca*

Submitted: June 7, 2018, accepted (with revisions): October 18, 2018

amongst older patients, making them more susceptible to injuries such as falls and fractures due to the frequent need to move to a bathroom.[11]

OAB is not always identified as a primary diagnosis. Rather, it is a diagnosis that is often arrived at after ruling-out other, potentially more severe, diagnoses. For example, in men clinicians may first consider prostate-related symptoms such as benign prostatic hyperplasia (BPH), or even prostate cancer. Consequently, when evaluating these patients, they may be more likely to be asked to complete a prostate-specific PRO rather than an OAB-specific PRO. If—after further testing and evaluation—the patient is found to have OAB, the scores generated from the prostate-specific PRO are usually irrelevant, rendering the data and its collection effort wasted.

Given the prevalence of both OAB and prostate-related diseases, we suspect that the administration of off-target PROs happens quite frequently. Therefore, the purpose of this study was to develop a method whereby the responses provided to a common prostate-specific PRO could be used in assessing OAB. This method is commonly referred to as a "crosswalk" which is essentially a statistical model that predicts scores for 1 PRO instrument using response data from items or scales from another, separate PRO instrument.[12] Previously, crosswalks have demonstrated success in areas such as quality of life,[13] health assessment,[14] and disability scales.[15] A meta-analysis conducted by Brazier et al[16] investigated several mapping studies and found over 119 models being used to develop these crosswalks. The results from our study could be applicable to both clinicians and health services researchers who find themselves with data collected from related but different PRO instruments. In the field of OAB, no such crosswalk has been created between any PROs.

METHODS

This study conducted secondary analyses of PRO data collected from patients being treated at a large community-based urology clinic located in Calgary, Canada. The clinic is comprised of a multidisciplinary team focused on the treatment of lower urinary tract symptoms (LUTS), including: 2 urologists, an internist, 2 urology fellows, 4 general practitioners, physiotherapists, and nurses. As part of the clinic's routine care, male patients at this clinic were asked to complete both the international prostate symptom score (IPSS) and the overactive bladder validated-eight (OAB-V8).

To be included in this study, patients had to have completed both the IPSS and the OAB-V8 in-full at the same time. Patients were excluded if they were female (because the IPSS was designed for prostate-specific conditions), were under the age of 18, could not read English or have a translator available (because reliable translations of the PRO instruments have not been developed), were missing 1 of the 2 PRO instruments of interest, or had an incomplete PRO instrument. Given the minimum risk associated with this study, and the use of anonymous data for analysis, the Conjoint Health Research Ethics Board at the University of Calgary granted a waiver of consent.

Patient-Reported Outcomes Data

The IPSS is an 8-item scale with 7 symptom items and one quality of life item.[17] The 7 symptom items are measured using a 6-point Likert scale ranging from "not at all" (ie, scored as 0) to "almost always" (ie, scored as 5). Three of the 7 items evaluate storage symptoms, while the other 4 measure emptying symptoms.[18] The scores from these 7 items are added together for a global score that can range between 0 and 35. Mild symptoms are indicated by a score of 7 or less, moderate symptoms are indicated by a score of 8-19, and severe symptoms are indicated by a score of 20-35. The quality of life item is measured on a 7-point Likert scale. The IPSS is commonly used to assess men with LUTS or BPH.[19] It is particularly useful in longitudinal follow-up assessments of changes in symptom severity and the evaluation of treatment efficacy.[19]

The OAB-V8 is an 8-item symptom bother scale which is commonly used as an awareness tool.[20] This scale was adapted from the Symptom Bother Scale of OAB-q.[20] The OAB-V8 asks how bothered someone is by the 4 main symptoms of OAB: urinary frequency, urgency, nocturia, and urge incontinence.[21] These 8 items are measured on a 6-point Likert scale ranging from "not at all" (ie, scored as 0) to "a very great deal" (ie, scored as 5). Responses to all 8 items are scored and summed, and 2 points are added if the respondent is male, resulting in a global score that can range between 0 and 40 (42 for males). An OAB-V8 global score greater than 8 indicates possible OAB.[20] The OAB-V8 has been validated in male and female populations.[21] Using item response theory, a recent study found the OAB-V8 to have strong psychometric properties for measuring overactive bladder.[22]

Both the IPSS and OAB-V8 instruments were scanned into patients' electronic health records at the time they were completed. Completed surveys between 2012 and 2017, along with relevant demographic and diagnostic data were abstracted from patients' records, de-identified, and entered into a research database.

Analysis

Participants were characterized based on age, diagnostic code, and PRO scores. To assess the relationship between the IPSS and the OAB-V8, participants' responses were first compared using a correlation matrix. A scatterplot of the global scores were compared to assess the appropriateness of using linear regression.

Ordinary least squares (OLS) regression was used to model the OAB-V8 global score as the dependent variables using responses to the IPSS as the independent variable(s). Three models were developed. The first used the IPSS global score:

$$\textbf{Model } 1: \ \gamma = \alpha + \beta_1(\textit{IPSS total score})$$

The second used the individual responses to the IPSS' items:

$$\textbf{Model } 2: \ \gamma = \alpha + \beta_{1\text{–}7}(\textit{IPSS item score})$$

The third model included the individual responses to the IPSS' items and participants' age.

$$\textbf{Model } 3: \ \gamma = \alpha + \beta_{1\text{–}7}(\textit{IPSS item score}) + \beta_8(\textit{age})$$

where for all 3 models γ is the OAB-V8 global score, α is the intercept, and β are the independent coefficient values.

For all 3 of the models, the Shapiro-Wilk test indicated that the residuals were significantly skewed. Therefore, a Box-Cox transformation was used to normalize the dependent variable. Because 0 was a possible score for the IPSS and the OAB-V8, we added 0.5 to all the scores. Using the transformed data, we conducted bootstrap estimates for the standard errors of the coefficients used in the OLS regression models using a random number seed of 1 with 100 replications. To compare the models, we used the root mean square error (RMSE), where smaller values indicate better the performance. Stata 14.2 (StataCorp LLC; College Station) was used to perform all the analyses.

RESULTS

The sample consisted of 493 participants, ranging from 18 to 93 years of age. A large majority (61%) of participants were diagnosed with urinary symptoms (ie, ICD-9 range 788-788.9). The mean IPSS global score was 15.0. The mean OAB-V8 global score was 17.2. See Table 1 for a summary of the sample characteristics.

All of the items from the IPSS and OAB-V8 were positively correlated. IPSS item #2 (ie, "How often have you had to urinate less than every 2 hours?") and item #4 (ie, "How often have you found it difficult to postpone urination?") had the strongest correlation with the OAB-V8. IPSS item #5 (ie, "How often have you had a weak urinary stream?") and item #6 (ie, "How often have you had to strain to start urination?") had the weakest correlation with the OAB-V8. Table 2 provides the full correlation matrix.

The scatter plot of IPSS and OAB-V8 global scores (Fig. 1), demonstrated that linear regression was appropriate. Box-Cox analysis was run independently for each of the study's 3 OLS models. For each model, the estimated transformation was approximately 0.44, so we took the square root of the dependent variable to normalize the data. In each model, the skewness was significantly improved after the transformation.

Results from the 3 OLS models are provided in Table 3. The adjusted R^2 value ranged from 0.44 in model 1 to 0.62 in models 2 and 3. However, the R^2 value does not necessarily imply predictability, therefore, RMSE needed to be taken into consideration. Model 3 was the best model, with the lowest RMSE value of 0.760. It is important to note that the β estimates and standard errors are based on the transformed values; interpreting the results from these models requires back-transforming by squaring the independent and dependent variables.

Table 1. Characteristics of participants completing both the IPSS and the OAB-V8

n	493
Age	
Range	18-93
Mean	59.2
Standard deviation	16.7
Diagnostics categories	
Urinary symptoms	302 (61%)
Bladder or urinary tract dysfunctions	72 (15%)
Hyperplasia or prostatitis	71 (14%)
Other diagnosis	48 (10%)
IPSS score	
Range	0-35
Mean	15.0
Standard deviation	8.3
OAB-V8 score	
Range	2-42
Mean	17.2
Standard deviation	10.1

DISCUSSION

The purpose of this study was to develop a crosswalk method that could be used to translate scores from the IPSS to the OAB-V8. This is useful when trying to compare PRO data between different instruments—as may be the case when something like BPH is suspected (and the IPSS administered) instead of OAB. This study demonstrates that the crosswalk approach is especially useful when clinicians want to limit patient burden and thus make use of existing data.

Based on our results, we recommend using response values to the individual IPSS items plus the respondent's age (ie, model 3). This not only had the lowest RMSE, indicating its greater predictive power, but also includes data that will likely be available to end-users. As OAB is an age-related condition,[23] it also makes clinical sense to include as an explanatory variable.

The results of our crosswalk are comparable to other studies undertaking similar analysis. In terms of how our results compare to previous studies that have used the crosswalk method, the systematic review by Brazier et al[16] determined that most crosswalk studies had an R^2 value of at least 0.5. In this study, the R^2 value is 0.626, well above the 0.5 value identified in the systematic review.

To demonstrate the clinical application of the crosswalk model, consider several examples. In the first example, a 50-year-old respondent is severely burdened by his symptoms on the IPSS and scores all 7 items "5", the highest possible score. Plugging these values into model 3 (making sure to back-transform the dependent variable by squaring it) we get an equivalent OAB-V8 score of 31.8. In the second example, a 50-year-old respondent experiences no symptom burden whatsoever. Consequently, he scores all items on the IPSS "0". Plugging these values into model 3 results in an equivalent OAB-V8 score of 3.5. In the final example, a 50-year-old respondent had moderate symptoms, scoring "3" on all items of the IPSS. The resulting OAB-V8 score would be 17.1, using model 3.

While the crosswalk method is typically used to translate values from condition-specific PROs to preference-based indices, in order to calculate quality-adjusted life years, the use of this technique is not exclusive to this purpose. The growing use of PROs and the number of instruments available have necessitated that crosswalks be developed to facilitate converting scores between instruments for a variety of purposes.[15,24-25] Provided that the 2 instruments share sufficient conceptual overlap,[16] the crosswalk technique can be as useful for comparing

Table 2. Correlation matrix between IPSS and OAB-V8 items in regression sample

	OAB-V8							
IPSS	*Frequent Urination During the Daytime Hours?*	*An Uncomfortable Urge to Urinate?*	*A Sudden Urge to Urinate With Little or no Warning?*	*Accidental Loss of Small Amounts of Urine?*	*Nighttime Urination?*	*Waking Up at Night Because You had to Urinate?*	*An Uncontrollable Urge to Urinate?*	*Urine loss Associated With A Strong Desire to Urinate?*
How often have you had the sensation of not emptying your bladder?	0.3891	0.4307	0.3430	0.2720	0.2765	0.3085	0.3553	0.3258
How often have you had to urinate less than every 2 hours?	**0.5689**	**0.5451**	0.4727	0.3225	0.4967	**0.5490**	**0.5255**	0.4194
How often have you found you stopped and started again several times when you urinated?	0.3303	0.3596	0.3532	0.2404	0.2655	0.2852	0.3652	0.3584
How often have you found it difficult to postpone urination?	**0.5374**	**0.5741**	**0.6408**	0.4340	0.4193	0.4513	**0.6508**	**0.5699**
How often have you had a weak urinary stream?	0.2204	0.3144	0.2760	0.1294	0.2315	0.2411	0.2589	0.2585
How often have you had to strain to start urination?	0.2227	0.3406	0.2442	0.1408	0.1801	0.2026	0.2830	0.2484
How many times did you typically get up at night to urinate?	0.3805	0.3299	0.2836	0.2392	**0.5424**	**0.6375**	0.3944	0.2921

Bold values indicate correlations greater than 0.5.

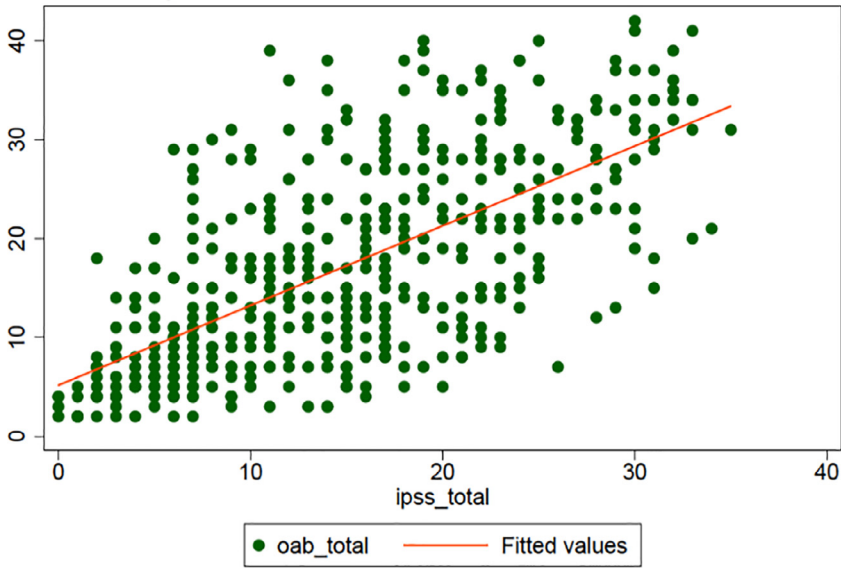

Figure 1. Scatter plot of IPSS and OAB-V8 total scores in regression sample. (Color version available online.)

Table 3. Ordinary least squares for transformed (squared) estimates

	Model 1			Model 2			Model 3		
Adjusted R^2	0.4458			0.6253			0.6260		
RMSE	0.9258			0.7612			0.7606		
	β Estimate	β Standard errors	*P* Value	β Estimate	β Standard Errors	*P* Value	β Estimate	β Standard Errors	*P* Value
Intercept	2.4715	0.7741	<.0001	1.8633	0.0856	<.0001	1.7103	0.1287	<.0001
IPSS total score	0.1003	0.0042	<.001						
IPSS Q1				0.0803	0.0254	.002	0.0823	0.0256	.001
IPSS Q2				0.1913	0.0282	<.0001	0.1937	0.0342	<.0001
IPSS Q3				0.0296	0.0231	.236	0.0256	0.0254	.313
IPSS Q4				0.3186	0.0225	<.0001	0.3138	0.0285	<.0001
IPSS Q5				−0.0183	0.0274	.458	−0.0269	0.0252	.347
IPSS Q6				−0.0196	0.0239	.494	−0.0121	0.0289	.676
IPSS Q7				0.1844	0.0231	<.0001	0.1800	0.0261	<.0001
Age							0.0030	0.0020	.147

patients' symptoms in clinical practice, registries, and studies as it is for measuring quality-adjusted life years.

There are several limitations to the generalizability of these results. First, the IPSS—being a prostate-specific instrument—is designed exclusively for use by males. The OAB-V8, on the other hand, can be used with both males and females. We limited the response data used in this study to males; thus, the results may not be generalizable to the female population. Second, the data were collected from patients seeking treatment from a single urology clinic. This patient population may be different in some unmeasured way from those seeking treatment elsewhere. However, we observed a range of age, initial diagnoses, and PRO scores that lead us to believe we have a well-distributed sample that is representative of what other community-based clinics would see.

In conclusion, this study successfully modeled global OAB-V8 scores from IPSS responses. This model performed comparably well to others developed using similar methods. The results from this study could be used by clinicians wanting to compare scores from the IPSS to the OAB-V8, which may be needed if the wrong instrument was initially administered or if clinicians want to limit the response burden on patients.

References

1. Zandbelt LC, Smets EM, Oort FJ, Godfried MH, de Haes HC. Medical specialists' patient-centered communication and patient-reported outcomes. *Med Care*. 2007;45:330–339.
2. U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health, U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U. S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006;4:79.

3. Jenkinson C, Morley D. Patient reported outcomes. *Eur J Cardiovasc Nurs*. 2016;15. 112-112.
4. Wiebe ER, Kaczorowski J, MacKay J. Why are response rates in clinician surveys declining? *Can Family Phys*. 2012;58:225–228.
5. Wells GA, Russell AS, Haraoui B, Bissonnette R, Ware CF. Validity of quality of life measurement tools—from generic to disease-specific. *J Rheumatol Suppl*. 2011;88. 2-6.
6. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199–208.
7. Fairclough DL. Patient reported outcomes as endpoints in medical research. *Stat Methods Med Res*. 2004;13:115–138.
8. Haywood KL, Garratt AM, Lall R, Smith JF, Lamb SE. EuroQol EQ-5D and condition-specific measures of health outcome in women with urinary incontinence: reliability, validity and responsiveness. *Qual Life Res*. 2008;17:475–483.
9. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346. f167.
10. Abrams P, Kelleher CJ, Kerr LA, Rogers RG. Overactive bladder significantly affects quality of life. *Am J Manage Care*. 2000;6. S580-590.
11. Henderson E, Drake M. Overactive bladder. *Maturitas*. 2010;66:257–262.
12. Williams BC, Li Y, Fries BE, Warren RL. Predicting patient scores between the functional independence measure and the minimum data set: development and performance of a FIM-MDS "crosswalk. *Arch Phys Med Rehabil*. 1997;78:48–54.
13. Edelen MO, Stucky BD, Sherbourne C, Eberhart N, Lara M. Correspondence between the RAND-negative impact of asthma on quality of life item bank and the marks asthma quality of life questionnaire. *Clin Ther*. 2014;36:680–688.
14. Pennington B, Davis S. Mapping from the health assessment questionnaire to the EQ-5D: the impact of different algorithms on cost-effectiveness results. *Value Health*. 2014;17:762–771.
15. ten Klooster P, Oude Voshaar MAH, Gandek B, et al. Development and evaluation of a crosswalk between the SF-36 physical functioning scale and health assessment questionnaire disability index in rheumatoid arthritis. *Health Qual Life Outcomes*. 2013;11:199.
16. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010;11:215–225.
17. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: report from the standardisation sub-committee of the international continence society. *Am J Obstet Gynecol*. 2002;187:116–126.
18. Barry MJ, Fowler JFJ, O'Leary MP, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol*. 1992;148:154.
19. Liao C, Chung S, Kuo H. Diagnostic value of International Prostate Symptom Score voiding—to—storage subscore ratio in male lower urinary tract symptoms. *Int J Clin Pract*. 2011;65:552–558.
20. Coyne KS, Zyczynski T, Margolis MK, Elinoff V, Roberts RG. Validation of an overactive bladder awareness tool for use in primary care settings. *Adv Ther*. 2005;22:381–394.
21. Cheung WW, Khan NH, Choi KK, Bluth MH, Vincent MT. Prevalence, evaluation and management of overactive bladder in primary care. *BMC Family Pract*. 2009;10:8.
22. Peterson A, Sehgal A, Crump T, Baverstock R, Sutherland J, Carlson K. Evaluating the 8-item overactive bladder questionnaire (OAB-v8) using item response theory. *Neurourol Urodyn*. 2018;37:1095–1100.
23. Coyne KS, Sexton CC, Bell JA, et al. The prevalence of lower urinary tract symptoms (LUTS) and overactive bladder (OAB) by racial/ethnic group and age: Results from OAB—POLL. *Neurourol Urodyn*. 2013;32:230–237.
24. Ghomrawi HMK, Lee Yyu, Herrero C, et al. A Crosswalk Between UCLA and Lower Extremity Activity Scales. *Clin Orthop Rel Res*. 2017;475:542–548.
25. Askew RL, Kim J, Chung H, Cook KF, Johnson KL, Amtmann D. Development of a crosswalk for pain interference measured by the BPI and PROMIS pain interference short form. *Qual Life Res*. 2013;22:2769–2776.