

ARTICLE



Special Edition on Neuro-urology

# Reliability of the Urinary Symptom Questionnaires for people with neurogenic bladder (USQNB) who void or use indwelling catheters

Rochelle E. Tractenberg<sup>1,2,3</sup>✉, Jamie K. Frost<sup>1</sup>, Futoshi Yumoto<sup>1</sup>, Amanda K. Rounds<sup>4,5</sup>, Inger H. Ljungberg<sup>4,5</sup> and Suzanne L. Groah<sup>3,4</sup>

© The Author(s), under exclusive licence to International Spinal Cord Society 2021

**STUDY DESIGN:** This is a descriptive psychometrics study.

**OBJECTIVES:** Neurogenic lower urinary tract dysfunction (NLUTD), also called Neurogenic Bladder (NB), is a common and disruptive condition in a variety of neurologic diagnoses. Our team developed patient-centered instruments, Urinary Symptom Questionnaires for people with neurogenic bladder (USQNB), specific to people with NLUTD who manage their bladders with intermittent catheterization (IC), indwelling catheters (IDC), or who void (V). This article reports evidence of reliability of the IDC and V instruments.

**SETTING:** Online surveys completed by individuals in the United States with NLUTD due to spinal cord injury (SCI), or multiple sclerosis (MS) who manage their bladder with IDC (SCI,  $n = 306$ ), or by voiding (SCI,  $n = 103$ ; MS,  $n = 383$ ).

**METHODS:** Reliability estimates were based on endorsement of the items on the USQNB-IDC and USQNB-V. Reliability evidence was representativeness of these symptoms for a national sample (by determining if endorsement  $> 10\%$ ); internal consistency estimates (by Cronbach's alpha and item correlation coefficient, ICC); and interrelatedness of the items (by inferred Bayesian network, BN). We also tested whether a one-factor conceptualization of "urinary symptoms in NLUTD" was supportable for either instrument.

**RESULTS:** All items were endorsed by  $> 20\%$  of our samples. Urine quality symptoms tended to be the most commonly endorsed on both instruments. Cronbach's alpha and ICC estimates were high ( $> 0.74$ ), but not suggestive of redundancy. BNs showed interpretable associations among the items, and did not discover uninterpretable or unexpected associations. Neither instrument fit a one-factor model, as expected.

**CONCLUSIONS:** The USQNB-IDC and USQNB-V instruments show sufficient, multidimensional reliability for implementation and further study.

*Spinal Cord* (2021) 59:939–947; <https://doi.org/10.1038/s41393-021-00665-x>

## INTRODUCTION

Neurogenic lower urinary tract dysfunction (NLUTD, also referred to as neurogenic bladder, NB) is defined as "lower urinary tract dysfunction due to disturbance of the neurological control mechanism", and is a common and disruptive condition for individuals with spinal cord injury and disease (SCI/D; including multiple sclerosis, MS and spina bifida, SB) [1–3]. While NLUTD/ or NB can result from any trauma or disease of the brain or spinal cord, people with SCI/D are highly affected: in the United States, 70–84% prevalence of NLUTD is estimated for those with SCI [4]. In Canada, up to 90% of SB patients are estimated to have "an active urological issue", such as NLUTD [5]. NLUTD is associated with a disproportionately high risk of

genitourinary complications, including urinary tract infections (UTIs [1–3, 6–12]).

Although UTIs are the most common cause of emergency department visits and rehospitalization among people with NLUTD [1, 3, 9], the most prevalent, bothersome, urinary symptoms as reported by patients with SCI/D and NLUTD in a recent US-based national sample of individuals who utilize intermittent catheterization (IC) for bladder management, cloudier- and more foul-smelling urine than normal [13], are excluded from discussions of clinical management for NLUTD [14], and are also typically excluded from UTI diagnostic criteria [15–18]. Further, while clinical classifications effectively describe urinary symptoms that are consistent with the pathology and

<sup>1</sup> Collaborative for Research on Outcomes and—Metrics, Washington, DC, USA. <sup>2</sup> Departments of Neurology, and Biostatistics, Bioinformatics & Biomathematics, Georgetown University Medical Center, Washington, DC, USA. <sup>3</sup> Department of Rehabilitation Medicine, Georgetown University Medical Center, Washington, DC, USA. <sup>4</sup> MedStar National Rehabilitation Hospital, Washington, DC, USA. <sup>5</sup> MedStar Health Research Health Institute, Hyattsville, MD, USA. ✉email: rochelle.tractenberg@gmail.com

Received: 3 June 2020 Revised: 25 June 2021 Accepted: 28 June 2021  
Published online: 4 August 2021

physiology of NLUTD (see, e.g., Dorsher and McIntosh [3]), these do not incorporate or contemplate the patient's experience of the disorder. Thus, urinary symptoms significantly affect consumer life and function, but instrument(s) specific to UTI (and related) symptoms in the context of NLUTD, that reflect the patient experience, have been lacking (see, e.g., Tsang et al.; Patel et al.; Clark and Welk [19–21]).

To address this gap, and to promote and support research into new interventions, our research team has developed a set of Urinary Symptom Questionnaires for people with NLUTD (USQNBs) that are targeted for people with NLUTD specific to bladder management method (IC/no voiding [13]; indwelling catheter (IDC); and voiding with/without some intermittent catheterization, i.e., primarily voiding to empty the bladder (not catheterizing more than once daily; V)). Our focus has been on capturing and prioritizing the consumer's perspective, to both represent, and facilitate self-management of, urinary signs and symptoms. The USQNB instruments are intended to strengthen UTI diagnosis for individuals with SCI/D, as well as supporting research and patient self-management, with respect to urinary symptoms potentially attributable to UTI for NLUTD, depending on bladder management. By allowing clinicians and researchers to focus on symptoms specific to NLUTD that may be suggestive of UTI, we hope to enable better antimicrobial stewardship, as well as facilitating diagnosis, patient self-management, and research. We previously published our results on reliability and validity evidence for the USQNB for intermittent catheter users [13] and for general scoring considerations for the instruments [22]. In this paper, we explore the psychometric reliability of consumers' responses to the USQNBs for indwelling catheters (IDC) and voiders (V) from these specific populations, following the definitions and recommendations for reliability estimation and reporting of the international Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) featuring reliability specifically [23]. The COSMIN criteria for reliability include estimates of internal consistency, intra-rater reliability, and measurement error. Since it is impossible to estimate measurement error for a patient's experience, no such evidence was sought. We estimated internal consistency and intra-rater reliability as described in the "Methods" section ("Psychometric reliability").

## METHODS

All USQNBs were developed following our model for patient-centered patient-reported outcomes [24], namely, we conducted focus groups and interviews with consumers who manage their bladders with the specific method (voiding; IDC, the latter inclusive of both urethral and suprapubic catheters). This paper describes the reliability evidence obtained from national samples responding to the resulting surveys, while a subsequent manuscript describes their respective validity evidence [25], following international consensus guidance on health-related patient-reported outcomes [23]. Approval for all parts of the studies was received from the MedStar National Rehabilitation Hospital Institutional Review Board (IDC: IRB# 2016-088; V: IRB# 2016-212).

### Target patient population participants and recruitment

Participants in the groups these instruments were developed to assess (i.e., the target populations) were recruited in the United States by English-language advertising through Facebook, via email, and with the assistance of the national (US) advocacy networks in SCI and in MS. All of these outreach and recruitment efforts were advertisements seeking respondents with NLUTD who use the specified management method to visit the URL we established for data collection. The target populations for these instruments are individuals with NLUTD and the specific bladder management; since SCI/D are major causes of NLUTD, our instrument development and these national samples focused on SCI/D—specifically traumatic SCI and MS. No personal or identifying data were collected from

any respondent, to protect anonymity, we required that participants complete the survey in one, single session. No identifiable information was collected. All responses remained anonymous, and data were transmitted from the SurveyMonkey form to the analyst anonymously (i.e., the data analyst did not have access to the original data file). We focused recruitment efforts on individuals with SCI/D who have NLUTD or NB, but apart from including questions within the survey about time since diagnosis (of SCI or MS) and whether they received a diagnosis of NB or NLUTD, or their experience with UTI and we did not confirm any of this self-reported information.

Participants followed the link to a SurveyMonkey page and chose their bladder management type. Respondents were specifically directed to consider their "normal" experience with each and endorse an item if it represented a change from "normal". Each item (urinary symptom) was presented as a query about whether the respondent had experienced it during the past year (yes/no). Thus, each endorsement was intended to reflect a difference from usual, and the instructions reflect this difference. If an item was endorsed, there were three additional responses required: average frequency (0–365); average severity (usually not at all severe; usually somewhat severe; usually severe; always very severe); and average impact on, or importance in, daily life (rarely affects my actions or decisions to go about my daily life; sometimes affects my actions or decisions to go about my daily life; usually affects my actions or decisions to go about my daily life; always affects my actions or decisions to go about my daily life). Analyses reported here were focused solely on the endorsement (yes/no) of the items "for the past year", although we intend for the instruments to be used in a much shorter time frame. This yearlong time frame was featured to ensure that the preliminary evidence we collected from our national sample on each instrument was as inclusive as possible. Basic psychometric reliability was focused on endorsement; frequency and the other aspects of the symptoms could not be meaningfully analyzed without this preliminary evidence of reliability.

### Psychometric reliability

These analyses were planned to generate estimates of instrument reliability: specifically, minimization of measurement error; internal consistency or interrelatedness of the items; and maximization of variability that is due to "true" difference between levels of the symptoms across patients. Psychometric reliability is essentially the description of the consistency of an instrument (or rather, the scores it produces) across independent administrations of the instrument: either to the same person over time, or to different people at the same time [26]. However, there can be no diagnostic "gold standard" for what the urinary signs and symptoms represent, so "measurement error" cannot be effectively estimated. Instead, to approximate inter-rater reliability, we sought to understand how "representative" the item pool, that was assembled following our model, is for the lived experience of NLUTD through our national sample. In a previous study, we defined a minimum endorsement rate of 10% [13], such that any item on our instrument that was endorsed by <10% of our target consumer group would be reconsidered for inclusion in the final instrument (suggesting it contributed low inter-rater reliability). With a national endorsement level <10% within the previous year, an item might be only minimally representative of our targeted lived experience and as such, might lower the representativeness of that item for the target consumer group. Such items would be reviewed for potential duplication with other items, and also by clinicians and the investigative team to ensure that rare (infrequently endorsed) items are, in fact, consistent with the "construct on which the proposed ...score interpretations are based" [26], namely, experience of urinary symptoms associated with NLUTD. Thus, observing a >10% endorsement rate for all items on the new instruments was included as a component of our reliability evidence. Our analyses focused on endorsement of the items on these two new instruments, because the impact, frequency, and severity ratings increase the difficulty of generalizing results, and are difficult to reconcile with the representativeness of these items for the patients' lived experience.

Additional reliability evidence was sought by estimating internal consistency for the items on each of the two new instruments using Principal axis factoring (PAF) on tetrachoric correlations (between endorsements of each item, 0 = no and 1 = yes); Cronbach's alpha; and inferred Bayesian networks (BNs) to uncover patterns of associations between signs and symptoms. Because the time period we asked about in our national samples was an entire year, internal consistency and interrelatedness are not likely to yield interpretable estimates. Instead we utilized these modeling methods to test whether (a) a one-factor model

**Table 1.** Descriptive statistics.

	SCI with IDC	MS with IDC	SCI voiders	MS voiders
# of respondents	306	8	103	405
Avg age $\pm$ SD (range)	54.2 $\pm$ 13.2 (8–80)	49.4 $\pm$ 14.2 (21–65)	55.6 $\pm$ 13.8 (20–82)	52.8 $\pm$ 14.6 (10–80)
Avg time living with injury $\pm$ SD (range)	19.9 $\pm$ 14.5 (1–64)	13.6 18.9 (1–50)	16.3 $\pm$ 16.9 (1–65)	14.9 $\pm$ 10.3 (0–53)
% of women	38.5	37.5	50.5	88.0

SCI spinal cord injury, MS multiple sclerosis, IDC using indwelling catheter, Avg average, SD standard deviation.

would fit the items (i.e., “NLUTD causes these symptoms”); and (b) expected patterns were observed in terms of symptoms that should logically associate, and also in those that logically should not associate [27, 28]. In this manner, the BN particularly can contribute evidence on intra-rater reliability, because symptoms that are generally related (e.g., about urine quality) should be associated within-rater, while those that are not related (e.g., cloudy urine and leg pain) should not be associated within-rater. We did not expect a one-factor model to fit the data because items were derived from the patients’ perspectives, rather than a measurement model [29]; and also because items were endorsed “in the previous year”, potentially compromising interpretability of latent factors influencing endorsement.

For the present report, our national samples can provide preliminary estimates of internal consistency of the items on each instrument (Cronbach’s alpha) but again, because of the time frame of our survey, we expected this to be an underestimate (lower bound) for consistency, and we plan to revisit this estimate in future with data from a shorter time interval in an independent sample.

### Data analysis

Statistical analyses were carried out using SPSS v. 24 [30] for descriptive statistics, item correlation coefficient (ICC) and Cronbach’s alpha computation, and to estimate the variance explained by the exploratory factor analysis; these analyses were based on the endorsement (binary, yes = 1, no = 0) of all items for the samples, by instrument. For each of the new instruments (USQNB-IDC; USQNB-V), exploratory factor analysis with tetrachoric correlations was carried out in MPlus v. 8 (Statsoft, Los Angeles, CA). Since our prior work demonstrated that an eight-factor model was the best fit to the national sample data on the Urinary Symptom Questionnaire for neurogenic bladder- Intermittent Catheter (USQNB-IC) [13], we fit models with one to eight factors to the voiders and IDC data. Exploratory factor analysis was used to determine the quality of this range of factor models of the data (from one to eight factors). The extraction method we used was PAF, to ensure an accurate estimation of the total variance that could be extracted from any solution, together with oblimin rotation, which permits correlated factors. This approach was used to extract from one to eight factors from the tetrachoric correlation matrix of endorsement (0 = “no”/1 = “yes”) of items, separately for each instrument. The fit of model(s) to the data was formally qualified using five different aspects of fit, assessed for each model, reflecting general data-model fit (Satorra-Bentler model chi square,  $\chi^2$ —lower is better), assessment of the fit of the model to data in hypothetical replications (Akaike’s information criterion, AIC—lower is better), incremental model fit relative to an independence model (comparative fit index, CFI—between 0.95 and 1.0 is desirable), error in approximation of the data by the model (root mean-squared error of approximation, RMSEA—<0.08 is ideal), and the mean absolute value of the covariance residuals (standardized root mean square residual, SRMR—smaller and <0.09 is best). We chose the model with the best (lowest) AIC that also had acceptable other-fit indicators as “the best model” in each case. If any model had the best AIC but unacceptable other-fit indicators, it “did not fit” the data (i.e., was not an interpretable solution). Our objective with these exploratory factor analyses was to explore: (a) whether the same solution held for all bladder management groups; and (b) whether a one-factor solution was ever the “best” model. Variance explained by the best fitting model was also computed for each group.

BayesiaLab v. 6 (Bayesia S.A.S., Laval, France) was used for the BN modeling: unsupervised structural learning (searching for the overall data structure) using a maximum weight spanning tree learning algorithm to constrain the structure so that only one “parent” per variable is identified. Minimum description length (information, not probability) was used for scoring because linearity in relationships cannot be safely assumed for these items. The resulting tree was visualized in automatic layout for

interpretability (features discussed in Ch. 7 of Conrady and Jouffe) [31]. The BN was inferred using unsupervised learning and the maximum spanning tree option, an approach that is similar to the employment of rotation in factor analysis solutions for “simple structure”, but has no model fit indices to characterize the results. The results are, therefore, associative and not specifically causal. We planned the BN analyses to identify whether each instrument yielded evidence of interpretable associations among the items, such as like items associating (vomiting–nausea), and did not generate uninterpretable or unexpected associations, such as unlike items associating (vomiting–leg pain).

These were done on all patient respondents on each instrument separately; voiders were modeled separately by etiology (MS, SCI) because the sample sizes were large, whereas for those using IDC, there were so few with MS, all were analyzed together.

## RESULTS

### Demographics

Table 1 describes the respondents in the national samples for the specific target groups the instruments were intended for.

We recruited for respondents for 6 months or until our minimum level of response (300 per group (IDC, V) was achieved), and found it difficult to find individuals with SCI who were also voiders; we had to close the survey at the 6-month time point and had only obtained 103 responses from SCI voiders. Similarly, we were only able to find eight individuals with MS who used IDC for bladder management. Since the estimation procedures we planned to use all require substantial sample sizes, we grouped these eight MS IDC respondents in with the SCI respondents.

### Reliability evidence, endorsement rates

Endorsement rates across items are given in Tables 2 (IDC) and 3 (V), and plotted in Fig. 1A (IDC) and Fig. 1B (V).

Tables 2 and 3 show the symptoms on the two instruments. Each bladder-management-specific instrument contained items derived from focus groups, subject matter expert consultations, and the removal of duplicates following our previously published model [24]. There is some overlap in symptoms on the instruments (e.g., cloudy urine appears on both), but many differences. One important difference is that one item (autonomic dysreflexia) was identified by SCI voiders, and not by MS voider participants during instrument development, while another item was only relevant for the MS voiders (MS flare-up). The endorsement rates shown in Tables 2 and 3 range from 16.3%, pain in the suprapubic region to 87.3%, cloudy urine for IDC and from 27.3%, vomiting to 83.2%, fatigue for SCI voiders, and 15.1%, sediment discharge to 95.5%, fatigue for MS voiders. Fatigue was the highest endorsed symptom on the voiders instrument by both groups (82% by SCI and 95% by MS). The endorsement rates by group are visualized in Fig. 1A, B.

Figure 1A, B shows that all items were endorsed at greater than the 10% cutoff we proposed; most items were endorsed by >30% of each sample. The items with the lowest endorsement rates for voiders with MS were sediment discharge in the urine (15%) and blood in urine (20%)—also among the lowest endorsed items by voiders with SCI; and vomiting (20% for MS, 27% for SCI). For the IDC (SCI and MS analyzed together because of the small MS sample), blood clots (35%) and blood in the urine (50%) were more common than for the voiders. The two least-commonly

**Table 2.** Endorsement rates for IDC.

USQNB-IDC symptom	Endorsement rate (%)
	SCI + MS (N = 314)
Bad-smelling, stronger or fouler urine	83.5
Dark urine	85.4
Blood in urine	50.3
Blood clots in urine	35.0
Cloudy urine	87.3
Sediments/debris in urine/clogged catheter	78.3
Bladder spasms	62.6
Sensation of urinary urgency	34.8
Urine leakage that isn't normal	49.0
Discharge, redness, or sensitivity around catheter site	51.1
Increase or decrease from your usual blood pressure	47.2
Increased spasticity (not bladder spasms)	46.4
Sweating	46.4
Flushing of the chest, neck, or face	31.2
Diarrhea	51.9
Numbness, tingling in your extremities	37.5
Sense of burning in the extremities or throughout the body	27.8
Autonomic dysreflexia (AD)	46.0
Felt feverish	42.7
Fatigue/lethargy	60.3
Irritability, mental slowing/confusion	31.5
Bladder discomfort	37.5
Pain in the suprapubic region	16.3
Generalized abdominal discomfort or pain: bloating, pressure, or cramping	48.2
Back pain	40.5
Pain more than usual below the level of injury	25.1

USQNB-IDC Urinary Symptom Questionnaire for neurogenic bladder or neurogenic lower urinary tract dysfunction—for indwelling catheter, SCI spinal cord injury, MS multiple sclerosis were analyzed together.

endorsed items for the IDC sample were pain related: pain in the suprapubic region (16.3%) and pain more than usual below the level of injury (25.1%); however, back pain was more common than these two pain items (40.5%), and two other items (numbness/tingling, 37.5% and sense of burning, 27.8%) reflecting sensations were also more commonly endorsed.

Table 4 summarizes the results for COSMIN criteria for reliability.

### Exploratory factor modeling

PAF with oblimin rotation (accommodating correlated factors) was applied to the tetrachoric correlation matrix of endorsement for each instrument separately, and for MS voiders separate from SCI voiders. The variance explained by any one-factor solution is more accurate when PAF is used to extract factors, so the values reported in Table 4 ranging from 46.1% (for MS voiders) to 50.9% (for the full IDC sample) suggest that only about half of the variability in the responses can be explained by any model. Moreover, the best solution for MS voiders had six factors, while the six-factor solution for SCI voiders data did not converge (possibly due to too small of a sample). The IDC solutions had a variety of fit issues depending on number of factors, with the best

**Table 3.** Endorsement rates for voiders with MS or SCI.

USQNB-V symptom	Endorsement rate (%)	
	SCI (n = 103)	MS (n = 405)
Cloudiness	72.8	58.8
Dark urine	72.5	52.8
Bad, foul, or stronger smelling urine	72.3	58.8
Sediment discharge in urine	41.6	15.1
Blood in urine	34.7	19.5
Increased frequency of urination	69.3	89.9
Urinary urgency	55.4	72.3
Urge incontinence	72.3	80.2
Sensation of urgency/empty bladder	35.6	38.8
Small release of urine	69.3	82.7
Inability to fully empty bladder	62.4	62.4
Weak urine stream	50.5	58.2
Bladder spasms	49.5	38.1
Malaise, feeling unwell	74.3	68.1
Fatigue	83.2	95.0
Muscle weakness	60.4	81.4
Difficulty ambulating (walking)	48.5	72.1
Worsening of motor function	41.0	46.8
Increase in spasticity (not bladder spasms)	46.0	46.8
Chills	60.0	44.3
Sweating	57.0	56.4
Fever	43.0	37.1
Nausea	48.5	43.1
Vomiting	27.3	21.3
Difficulty thinking, mental foginess, forgetfulness	51.5	72.3
Autonomic dysreflexia "SCI ONLY"	43.9	
Multiple sclerosis flare-up (relapse, attack) "MS ONLY"		65.8

USQNB-V Urinary Symptom Questionnaire for neurogenic bladder or neurogenic lower urinary tract dysfunction—for voiding, SCI spinal cord injury, MS multiple sclerosis were analyzed separately.

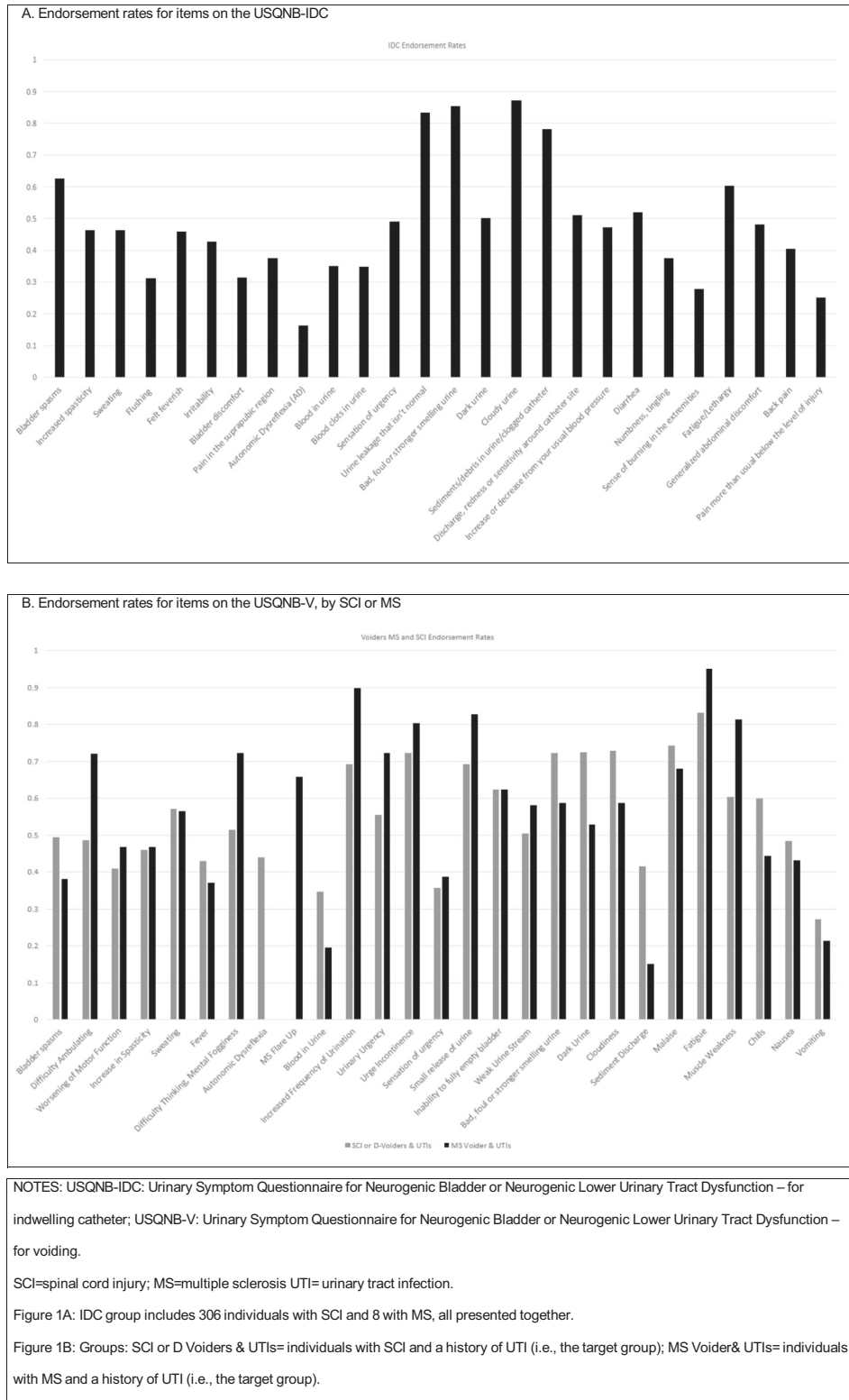
fitting model (seven factors) exhibiting better characteristics than any others, but not actually fitting significantly better than more parsimonious models (with four to six factors). Thus, as can be seen in Table 4, evidence for reliability based on the internal structure of these instruments was not consistent. We did observe that the one-factor model fit in each sample very poorly (data not shown), with a minimum of four factors required to achieve even a marginally good fit for each instrument.

### Inferred causal network

The BNs inferred from the endorsed items by all respondents on the USQNB-IDC (Fig. 2A, SCI and MS together), and by SCI (Fig. 2B) and MS (Fig. 2C) respondents to the USQNB-V, are represented in Fig. 2 below.

The networks shown in Fig. 2 highlight the complexity of the lived experience of urinary signs and symptoms in NLUTD for those managing their bladders with IDC (Fig. 2A, mostly SCI, eight MS) and by voiding (Fig. 2B, C). As can be seen in Fig. 2B (SCI) and Fig. 2C (MS), the inferred network of symptomatology over 1 year differs for MS and SCI patients who void; additionally, the instruments and networks include two etiology-specific symptoms





**Fig. 1** Item-level endorsement rates by instrument. **A** Endorsement rates for items on the USQNB-IDC. IDC group includes 306 individuals with SCI and 8 with MS, all presented together. **B** Endorsement rates for items on the USQNB-V, by SCI or MS. Groups: SCI or D-voiders and UTIs = individuals with SCI and a history of UTI (i.e., the target group); MS voider and UTIs = individuals with MS and a history of UTI (i.e., the target group). Notes: USQNB-IDC Urinary Symptom Questionnaire for neurogenic bladder or neurogenic lower urinary tract dysfunction—for voiding, SCI spinal cord injury, MS multiple sclerosis, UTI urinary tract infection.

**Table 4.** IDC and voiders COSMIN internal reliability evidence.

COSMIN construct	Group	Idc whole group (N = 314, including 8 with MS)	Voiders: SCI (N = 103)	Voiders: MS (N = 405)
Reliability-internal				
Principal axis factoring/oblimin		8f <sup>a</sup> would not converge. 7f best fit but 4f better than 5f, 5–7f not better than 4f ( $p > 0.18$ ). PAF <b>4</b> correlated factors explain <b>50.9%</b> of the variance	6–8f would not converge. 4f had the best fit. PAF <b>4</b> correlated factors explain <b>47.6%</b> of the variance	Neither 7f nor 8f converged. 6f best fit and better than 5f ( $p < 0.01$ ). Good fit on 4 of 5 indices. PAF <b>6</b> correlated factors explain <b>46.1%</b> of the variance
Cronbach's alpha		0.784 (26 items; based on 296 complete cases)	0.851 (26 items) <sup>c</sup> (based on 67 complete cases)	0.785 (26 items) <sup>c</sup> (based on 372 complete cases)
ICC <sup>a</sup>		ICC = 0.784 (95% CI: 0.746–0.818)	ICC = 0.851 (95% CI: 0.792–0.899)	ICC = 0.785 (95% CI: 0.752–0.815)

All available data was used, no imputations were done.

Bold values identify the final model/final results.

COSMIN Consensus-based Standards for the selection of health Measurement Instruments [23], PAF principal axis factoring, ICC item correlation coefficient.

<sup>a</sup>ICC was estimated from a two-way mixed effects (random rater, fixed measurement) model.

<sup>b</sup>f = factors; 8f = the model with eight factors extracted. Oblimin = rotation in factor analysis where simple structure is favored in the model but factors in the solution are allowed to be correlated. Models with one to eight factors (1f–8f) were tested.

<sup>c</sup>Etiology-specific items (MS flare-up; autonomic dysreflexia) were excluded from these analyses.

(autonomic dysreflexia, only for SCI; and MS flare ups, only for MS). None of these structures show a single causal or originating factor for all or even some of the symptoms; and although symptoms tend to group together in interpretable ways, (e.g., weak stream and small release group together for MS voiders; muscle weakness and cloudy urine do not group together in any model), the overall associations among these symptoms are not clearly interpretable.

For the SCI subset of voiders respondents, USQNB-V items reflecting bladder function tend to associate together, but it is the urine changes that tend to associate with feverish and chills (more commonly associated with infection). By contrast, for the MS group, motor symptoms tend to associate more with bladder symptoms. Arrows in the figures indicate whether there is an association between two variables given the modeling constraints (e.g., similar to “simple structure” in exploratory factor analysis, the modeling seeks to limit associations to one variable.) Attention is paid to whether any two variables that are shown to be associated (linked with an arrow) are plausibly associated (e.g., felt feverish-chills), and that variables that are not plausibly associated (e.g., felt feverish-leg pain) do not appear to have an association. No estimation is included in the figure because the exploration is meant to be qualitative, not quantitative. While associations differ depending on etiology, the items that are associated within each group are interpretable, and differences between the networks are, as well. This suggests that the voider instrument, in particular, has the potential to respond to subtle differences in respondent characteristics without impairing its potential reliability for either type of respondent (i.e., one with SCI or MS).

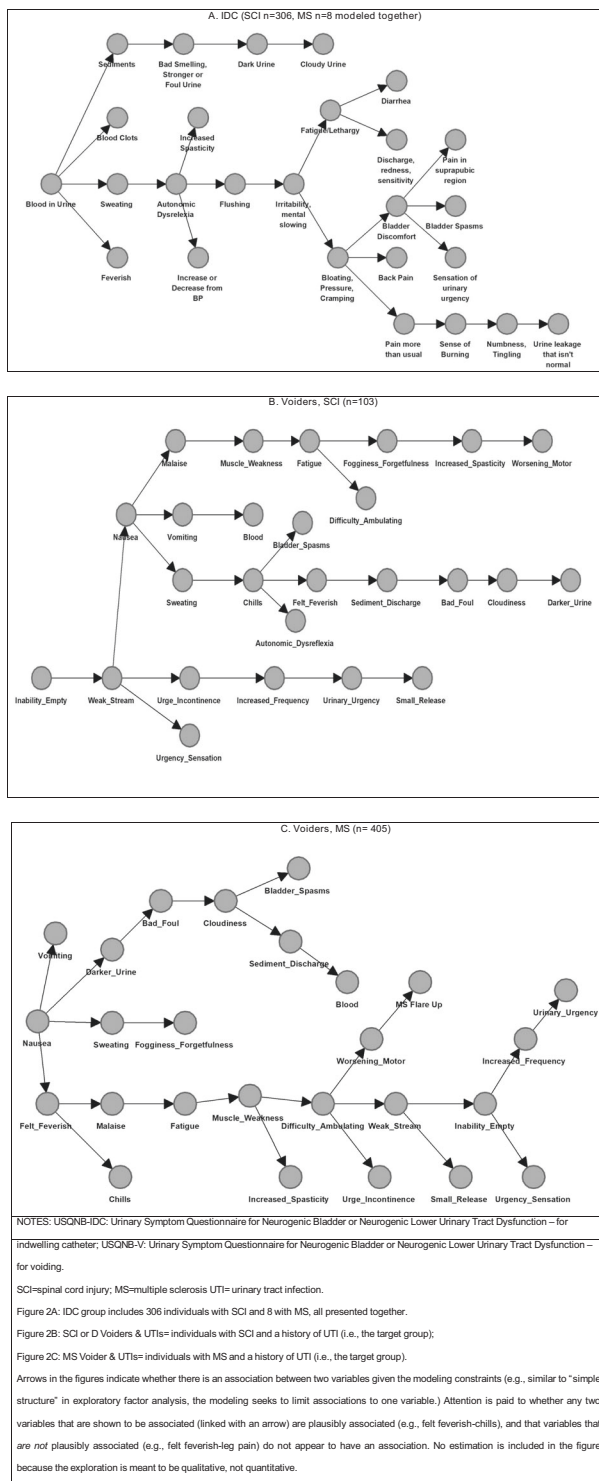
## DISCUSSION

The purpose of this study was to describe the evidence for psychometric reliability of two new patient-centered patient-reported outcome instruments for urinary signs and symptoms in people with NLUTD, and who use either indwelling catheterization to manage their bladder (USQNB-IDC), or who void (USQNB-V). These instruments were all developed using explicitly patient-centric methods, so as to be specific to the lived experience of the NLUTD population, and so it is not possible to estimate the error with which these symptoms are assessed, which is a hallmark of the estimation of psychometric reliability. Instead, we sought to define one type of “error” as the inclusion of items on our instruments that <10% of our national samples endorsed. We observed that >10% of our national sample endorsed all items on their respective instrument, suggesting that the items on each instrument do represent recognizable urinary symptoms according to our national samples. This captures at least one level of “consistency” in terms of

describing the experience of independent individuals with NLUTD who manage their bladders according to either instrument’s target group: all groups on both instruments generated sufficiently high levels of endorsement and internal consistency to support conclusions of reliability. Moreover, with acceptable levels of internal consistency [27] (all Cronbach’s alpha > 0.78) that did not suggest redundancy (i.e., upper bounds on 95% CIs for our internal consistency coefficients < 0.90), we compiled additional evidence of the reliability of these instruments, even though “error” cannot be meaningfully defined in this context because the only “true” or gold standard response comes from the respondent’s endorsement of the item.

The PAF results clearly show that the instruments do not fit a single-factor model, but the variance explained by any of the multi-factor models was low, ranging from 46.1% (for MS voiders) to 50.9% (for the full IDC sample). Thus, only roughly half of the variability in the endorsement responses can be explained by any model for either instrument, but none of the factor models fit well. This was expected given the origins of these items (i.e., from focus groups, and not with a specific measurement model in mind); and also because the survey asked about experience in the past year—just because a patient experienced two symptoms does not mean they were associated in time. This could be part of the reason for the inconsistent factor analysis results. We did observe that the one-factor model fit in each sample very poorly, as was expected.

Like the PAF, the BN was based on the endorsement of all items for each instrument. This method was used to uncover basic structure, but limited the associations that appeared to a single neighbor (i.e., a Markov Blanket approach, similar to “simple structure” in factor analysis). This method is not parametric, does not use or require inferences, and uses information instead of probability for estimating associations. Our PAF directly, and the BN indirectly, support a conclusion that there is no single factor or dimension underlying all the urinary symptoms in these respondents with NLUTD. This evidence is not unexpected given that the development of the instrument did not formally follow measurement model-driven methods [29]. We observed a similar multidimensional result, with poor fit of the one-factor model, in our previously published national sample for the intermittent catheter-specific instrument [13]. Although we have discussed the use of symptom-type classifications, rather than total score summarization, of individual levels of symptom burden in the intermittent catheter bladder management context [14, 22], the overall ICC values for the instruments specific to voider and to indwelling catheter management are high. The PAF did not support a one-factor model, as expected, and the BNs suggest



**Fig. 2 Inferred Bayesian networks for USQNB-IDC (A) and USQNB-V (B, C).** A IDC (SCI  $n = 306$ , MS  $n = 8$  modeled together), IDC group includes 306 individuals with SCI and 8 with MS, all presented together. B Voiders, SCI ( $n = 103$ ), SCI or D-Voiders and UTIs = individuals with SCI and a history of UTI (i.e., the target group); C Voiders, MS ( $n = 405$ ), MS voider and UTIs = individuals with MS and a history of UTI (i.e., the target group). Arrows in the figures indicate whether there is an association between two variables given the modeling constraints (e.g., similar to “simple structure” in exploratory factor analysis, the modeling seeks to limit associations to one variable). Attention is paid to whether any two variables that are shown to be associated (linked with an arrow) are plausibly associated (e.g., felt feverish-chills), and that variables that are not plausibly associated (e.g., felt feverish-leg pain) do not appear to have an association. No estimation is included in the figure because the exploration is meant to be qualitative, not quantitative. Notes: USQNB-IDC Urinary Symptom Questionnaire for neurogenic bladder or neurogenic lower urinary tract dysfunction—for indwelling catheter, USQNB-V Urinary Symptom Questionnaire for neurogenic bladder or neurogenic lower urinary tract dysfunction—for voiding, SCI spinal cord injury, MS multiple sclerosis, UTI urinary tract infection.

for either instrument was not our purpose in the studies, we did not estimate fit, factors, or strengths of associations between any of the items on either instrument—not with the exploratory factor analyses (Table 2) nor with the inferred networks (Fig. 2A–C). Instead, the BN represents inferred, and not estimated, structural relationships among symptoms as they are experienced (simple endorsement (yes/no) over 1 year). The observation of expected associations (e.g., cloudy with foul-smelling urine for all models) and not unexpected ones (e.g., cloudy urine with nausea) within all three of the networks shown in Fig. 2 is evidence that the USQNB-IDC and USQNB-V have sufficient structural interpretability for continuing the evaluation of their measurement properties. These methods were also applied to the previously published instrument, for intermittent catheter users (USQNB-IC) [13] and these results are similarly supportive of a conclusion of reliability. Importantly, all three instruments were developed for the same “urinary signs and symptoms” independently (following the same method) with focus groups, representing each bladder management method, leading to overlap (both instruments include the same items), or similarities (both items include similar items with wording differences). We are currently completing an analysis of the similarities and differences between the three instruments—and identification of redundancies—to clarify our understanding of urinary symptoms associated with NLUTD that vary according to bladder management method.

There are some similarities among items on all of the instruments, but crucial differences as well. We are in the process of examining the similarities and differences in the symptoms included on these instruments, in terms of how they may improve our understanding of NLUTD and patients’ experience of it, including an evaluation of concordance between patients’ reports of their symptoms and the International SCI Basic Data Set for lower urinary tract function [32]. The USQNB-IC has already been utilized in a clinical trial for treating bothersome symptoms associated with NLUTD, and we are hopeful that the two new instruments described here will be similarly useful for those with NLUTD who use other bladder management methods. Crucially, while a basic data set may support the uniform documentation of patient status across contexts and countries, the USQNB instruments are specifically intended to support decision making, so all items reflect something that has changed—i.e., something that might be targeted for change (improvement or prevention) with an intervention or management decision.

Of interest, the levels of endorsement for one item that appears on both of the instruments by individuals with SCI were quite different: fatigue was among the most commonly endorsed (83%)

expected (structural) associations are present in the data, and unexpected ones are not.

The results suggest these new instruments have extensive and diverse evidence of their reliability. However, limitations to these analyses must be noted. Firstly, since this was a preliminary evidence gathering study, and we asked participants to endorse items they had experienced at least once in the previous year, the time frame of the questions may partly underlie the failures to find structural coherence in the models. Further, because uncovering a measurement model structure (e.g., path weights, factor makeup)

items among voiders with SCI, but was endorsed by roughly 60% among those with SCI who use IDC to manage their bladders—still one of the most highly endorsed items, but not as often as by voiders with SCI. This item was the most prevalent symptom for all voiders, endorsed by 95% of MS and 83% of SCI voiders, while cloudy urine (87.3%), darker urine (85.4%), and stronger smelling urine (83.5%) were the most prevalent for those using IDC for bladder management. We have reported previously that, among those using IC, urine quality symptoms are the most prevalent in our national sample [13], and while these tend to be excluded from UTI diagnostic consideration (see Tractenberg et al. [14] and Table 1), these are common and bothersome symptoms that can be treated. The reliability evidence presented here, together with validity evidence published elsewhere, will support the use of these instruments in further studies of interventions that can address these symptoms. We have also completed an analysis of the validity evidence [25] supporting these instruments, further demonstrating consistency with Mokkink criteria [23] beyond the reliability evidence discussed here. The clinical relevance of this investigation lies in the identification of patient-centered, patient-reported urinary symptoms, their prevalences among individuals with NLUTD who have experience with UTIs. Our results—the identification of so many items by individuals with NLUTD, and their documentation as reliable (this paper) and valid [25]—go beyond the symptoms identified by authoritative guidelines as indicative of UTI [15, 16]. We previously compared the patient-centered, patient-reported urinary symptoms associated with NLUTD and IC [24], with other NLUTD-related UTI guidelines or recommendations, discovering that there is a disparity between the clinician-determined urinary symptoms indicative of UTI, and those urinary symptoms reported by people with NLUTD and considered by them to be related to UTI. By having an instrument that allows for measurement of both patient- and clinician-determined symptoms that are potentially related to UTI, we can begin work to actually differentiate those symptoms that are definitely, probably, likely, or unlikely related to NLUTD-related UTI. This can improve treatment and research into UTI and bothersome urinary symptoms [14], as well as promote antibiotic stewardship by focusing treatments where antibiotics would be most likely to be effective.

## CONCLUSIONS

The results presented here are based on the collection of preliminary evidence to allow the estimation of measurement properties that are of greatest interest [23], specifically, reliability. The results of our national samples suggest that the instruments have sufficient reliability to continue our work; we continue to develop scoring criteria [22] to complete the documentation of the instruments that describe the patient's lived experience with urinary symptoms specific to NLUTD.

## DATA AVAILABILITY

Data supporting the results reported in the article can be found in Files section of the Open Science Framework Project page, <https://osf.io/pfdje/>.

## REFERENCES

- Haisma JA, van der Woude LH, Stam HJ, Bergen MP, Sluis TA, Post MW, et al. Complications following spinal cord injury: occurrence and risk factors in a longitudinal study during and after inpatient rehabilitation. *J Rehabil Med*. 2007;39:393–8.
- Werhagen L, Gabrielsson H, Westgren N, Borg K. Medical complication in adults with spina bifida. *Clin Neurol Neurosurg*. 2013;115:1226–9.
- Dorsher PT, McIntosh PM. Neurogenic bladder. *Adv Urol*. 2012;2012:1–16.
- Manack A, Motsko SP, Haag-Molkenteller C, Dmochowski RR, Goehring EL, Nguyen-Khoa B-A, et al. Epidemiology and healthcare utilization of neurogenic bladder patients in a us claims database. *Neurourol Urodyn*. 2011;30:395–401.
- Saavedra AA, MacLellan D, Gray GJ. Spina bifida. *Can. Urol. Assoc. J*. 2018;12:S3–9.
- Vaidyanathan S, Soni BM, Dundas S, Krishnan KR. Urethral cytology in spinal cord injury patients performing intermittent catheterisation. *Paraplegia*. 1994; 32:493–500.
- Consortium for Spinal Cord Medicine. Bladder management for adults with spinal cord injury: a clinical practice guideline for health-care providers. *J Spinal Cord Med*. 2006;29:527–73.
- Cardenas DD, Hoffman JM, Kirshblum S, McKinley W. Etiology and incidence of rehospitalization after traumatic spinal cord injury: a multicenter analysis. *Arch Phys Med Rehabil*. 2004;85:1757–63.
- DeJong G, Tian W, Hsieh C-H, Junn C, Karam C, Ballard PH, et al. Rehospitalization in the first year of traumatic spinal cord injury after discharge from medical rehabilitation. *Arch Phys Med Rehabil*. 2013;94:S87–97.
- Disability NI on, Rehabilitation Research Consensus Statement January 27–29 1992. The prevention and management of urinary tract infections among people with spinal cord injuries. *J Am Paraplegia Soc*. 1992;15:194–207.
- Benevento BT, Sipski ML. Neurogenic bladder, neurogenic bowel, and sexual dysfunction in people with spinal cord injury. *Phys Ther*. 2002;82:601–12.
- Stillman MD, Frost KL, Smalley C, Bertocci G, Williams S. Health care utilization and barriers experienced by individuals with spinal cord injury. *Arch Phys Med Rehabil*. 2014;95:1114–26.
- Tractenberg RE, Groah SL, Rounds AK, Ljungberg IH, Schladen MM. Preliminary validation of the Urinary Symptom Questionnaire for individuals with neuropathic bladder using intermittent catheterization (USQNB-IC): a patient-centered patient reported outcome. *PLoS ONE*. 2018;13:e0197568. <https://doi.org/10.1371/journal.pone.0197568>.
- Tractenberg R, Groah SL, Rounds AK, Frost JK, Ljungberg IH, Davis E, et al. Effects of intravesical *Lactobacillus rhamnosus* GG on urinary symptom burden in people with neurogenic lower urinary tract dysfunction. *Phys Rehabilitation Med*. 2021; 13:229–40.
- Hooton TM, Bradley SF, Cardenas DD, Colgan R, Geerlings SE, Rice JC, et al. Diagnosis, prevention, and treatment of catheter-associated urinary tract infection in adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases Society of America. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2010;50:625–63.
- Madden-Fuentes RJ, McNamara ER, Lloyd JC, Wiener JS, Routh JC, Seed PC, et al. Variation in definitions of urinary tract infections in spina bifida patients: a systematic review. *Pediatrics*. 2013;132:132–9.
- Everaert K, Lumen N, Kerckhaert W, Willaert P, van Driel M. Urinary tract infections in spinal cord injury: prevention and treatment guidelines. *Acta Clin Belg*. 2009; 64:335–40.
- Naber KG, Bergman B, Bishop MC, Bjerkklund-Johansen TE, Botto H, Lobel B, et al. EAU guidelines for the management of urinary and male genital tract infections. Urinary Tract Infection (UTI) Working Group of the Health Care Office (HCO) of the European Association of Urology (EAU). *Eur Urol*. 2001;40:576–88.
- Tsang B, Stothers L, Macnab A, Lazare D, Nigro M. A systematic review and comparison of questionnaires in the management of spinal cord injury, multiple sclerosis and the neurogenic bladder: comparison of questionnaires in the management of neurogenic bladder. *Neurourol Urodyn*. 2016; 35:354–64.
- Patel DP, Elliott SP, Stoffel JT, Brant WO, Hotaling JM, Myers JB. Patient reported outcomes measures in neurogenic bladder and bowel: A systematic review of the current literature: patient reported outcomes in neurogenic bladder and bowel. *Neurourol Urodyn*. 2016;35:8–14.
- Clark R, Welk B. Patient reported outcome measures in neurogenic bladder. *Transl Androl Urol*. 2016;5:22–30.
- Tractenberg RE, Groah SL, Rounds AK, Davis EF, Ljungberg IH, Frost JK, et al. Clinical profiles and symptom burden estimates to support decision-making using the Urinary Symptom Questionnaire for people with neurogenic bladder (USQNB) using intermittent catheters. *Phys Rehabil. Med*. 2021;13:229–40.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737–45.
- Tractenberg RE, Garver A, Ljungberg IH, Schladen MM, Groah SL. Maintaining primacy of the patient perspective in the development of patient-centered patient reported outcomes. *PLoS ONE*. 2017;12:e0171114.
- Tractenberg R, Frost J, Rounds AK, Ljungberg IH, Schladen MM, Groah SL. Validity of patient-centered, patient reported outcomes on the Urinary Symptom Questionnaires for people with neurogenic lower urinary tract dysfunction (NLUTD): the Urinary Symptom Questionnaires for people with NLUTD (USQNB) who void or use indwelling catheters. *Spinal Cord*. 2021. (In press).
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. American Educational Research Association; 2014. p. 230.



27. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011; 2:53–5.
28. Nunnally JC, Bernstein IH. *Psychometric theory.* 3rd ed. McGraw-Hill; 2008. p. 752.
29. Bollen KA. *Structural equations with latent variables.* Wiley; 1989. p. 514.
30. IBM How to cite IBM SPSS Statistics or earlier versions of SPSS—United States. 2016. <http://www.ibm.com/support>, <https://www-304.ibm.com/support/docview.wss?uid=swg21476197>. Accessed 15 Mar 2017.
31. Conrady S, Jouffe L. Bayesian networks and BayesiaLab: a practical introduction for researchers. Ch. 7. 2015 p. 165–98.
32. Biering-Sørensen F, Craggs M, Kennelly M, Schick E, Wyndaele J-J. International lower urinary tract function basic spinal cord injury data set. *Spinal Cord.* 2008;46:325–30.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of MMS (data curation, investigation, and acquisition of funding).

## AUTHOR CONTRIBUTIONS

RET: conceptualization, formal analysis, funding acquisition, investigation, methodology, supervision, validation, visualization, writing—original draft, and writing—review and editing; JKF: formal analysis, validation, visualization, writing—original draft, and writing—review and editing; FY: formal analysis, methodology, validation, visualization, writing—original draft, and writing—review and editing.; AKR: data curation, resources, and writing—review and editing; IHL: project administration, resources, acquisition of funding, and writing—review and editing; and SLG: funding

acquisition, investigation, project administration, resources, supervision, writing—review and editing.

## FUNDING

Two funders supported this work: USQNB-V: supported by NIDILRR grant #90IF0121 and USQNB-IDC: supported by Craig H Neilsen Foundation grant # 385077.

## COMPETING INTERESTS

The authors declare no competing interests.

## STATEMENT OF ETHICS

We certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during the course of this research.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to R.E.T.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.