



# The Neurogenic Bladder Symptom Score (NBSS): a secondary assessment of its validity, reliability among people with a spinal cord injury

Blayne Welk<sup>1</sup> · Sara Lenherr<sup>2</sup> · Sean Elliott<sup>3</sup> · John Stoffel<sup>4</sup> · Angela P. Presson<sup>5</sup> · Chong Zhang<sup>5</sup> · Jeremy B. Myers<sup>2</sup>

Received: 24 July 2017 / Revised: 27 September 2017 / Accepted: 30 September 2017 / Published online: 29 November 2017  
© International Spinal Cord Society 2018

## Abstract

**Study design** Prospective cross-sectional study.

**Objectives** Validate the Neurogenic Bladder Symptom Score (NBSS) for people with spinal cord injury (SCI).

**Setting** United States (recruitment from community/tertiary neurourology clinics).

**Methods** We used data from a prospective observational study of people with a SCI who enrolled during December 2015–September 2016. Participants completed the NBSS and other measurement tools (SF-12 and SCI-QOL Bladder Management Complications tool). Data were used to determine the internal consistency (Cronbach's alpha), validity (hypothesis testing), and test–re-test reliability (using an intraclass correlation coefficient).

**Results** 609 people with a SCI had complete data. The median NBSS total score was 22 (IQR 15–30), and median quality of life was “mixed”. The Cronbach's alpha of the total and the incontinence, storage/voiding, and consequences domains was 0.85, 0.93, 0.76, and 0.49 respectively. All item to domain correlations were  $\geq 0.3$ , aside from 3/7 of the items from the consequences domain. Appropriate correlations between the NBSS domains and external variables and other questionnaires were observed, such as a moderate correlation between the SCI-QOL Bladder Management complications tool and the NBSS total score. For the reliability assessment, 174 people had 3 month followup data and did not have a significant change to their urologic health. The intraclass correlation coefficients were  $>0.75$  for all subdomains and the overall score.

**Conclusions** The NBSS demonstrated good validity and reliability in a large cohort of people with a SCI, and is a suitable tool to assess neurogenic bladder symptoms.

**Sponsorship** Patient-Centered Outcomes Research Institute (PCORI) Award CER14092138.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41393-017-0028-0>) contains supplementary material, which is available to authorized users.

---

✉ Blayne Welk  
blayne.welk@sjhc.london.on.ca

<sup>1</sup> Department of Surgery and Epidemiology & Biostatistics, Western University, London, ON, Canada

<sup>2</sup> Department of Surgery, Division of Urology University of Utah, Salt Lake City, UT, USA

<sup>3</sup> Department of Surgery, University of Minnesota, Minneapolis, MN, USA

<sup>4</sup> Department of Surgery, University of Michigan, Ann Arbor, MI, USA

<sup>5</sup> Department of Epidemiology, Department of Internal Medicine, University of Utah Health Care, Salt Lake City, Utah, USA

## Introduction

Neurogenic bladder dysfunction affects a heterogeneous group of people with varying bladder symptoms [1]. Assessment and care of this population is challenging due to different functional abilities, comorbidities, and bladder management strategies. The impact of a neurogenic bladder on a person's quality of life (QOL) can be significant, and genitourinary complications are a source of morbidity for many people [2, 3]. While there are urinary specific QOL measures, such as the Qualiveen [4] and the SCI-QOL Bladder Management Complications tool [5], they focus primarily on the impact of bladder issues on QOL, rather than directly measuring symptom burden. The Neurogenic Bladder Symptom Score (NBSS) is a validated 24 item questionnaire that measures bladder symptoms across 3 different domains: incontinence (scored 0–29), storage and voiding (scored 0–22), and consequences (scored 0–23);

there is a single general urinary QOL question scored from 0 (pleased) to 4 (unhappy) [6]. For all domains, a higher score represents a worse symptom burden or QOL.

The NBSS has only been validated in the original study population which included a mix of people with a spinal cord injury (SCI), multiple sclerosis (MS) or congenital neurogenic bladder [6]. Reviews of the SCI measurement literature have suggested that a validation in a predominantly SCI population would add to its face validity [7]. The objective of this study was to complete a secondary assessment of the validity and reliability of the NBSS using a large cohort of people with a SCI.

## Methods

Data from an ongoing multicenter prospective observational study measuring bladder related complications and QOL of people with a SCI over time was used (clinicaltrials.gov NCT02616081). Institutional ethics board approval was obtained from all participating institutions, and all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during the course of this research.

People were recruited from neurourology and rehabilitation clinics in the United States, and through an open online portal between December 2015 to September 2016 [8]. Briefly, the inclusion criteria were age of at least 18 years with an acquired SCI. People completed an extensive standardized telephone interview and patient reported outcome measures (including the NBSS) at initial enrollment. They then completed assessments (including the patient reported outcome measures such as the NBSS) at 3-month intervals over a 1 year period. These assessments were carried out independently by the participants and triggered using a reminder system and a standardized online gateway (that was unchanged between time points). Additional variables used for the hypothesis testing validity assessment included, first, the Short Form-12 (SF-12), an extensively studied general QOL tool; scores are standardized to a mean of 50 and a standard deviation of 10, and a higher score is interpreted as better QOL [9]. The physical domain questions were modified for people with SCI [10]. Second the computer adaptive version of the SCI-QOL Bladder Management Complications tool [5] which generates a standardized score (mean score of 50 and a standard deviation of 10), and a higher score equates to more complications. Third, responses to questions about the presence of a renal stone procedure in the 3 months prior to enrollment (binary variable), the number of urinary infections in the past year (continuous variable), and a hospitalization for urinary infections in the prior year (binary variable). Our

hypotheses regarding these relationships are outlined in Table 2 of the results. As this study was ongoing, an interim data set after 9 months of study recruitment was used for this study.

## Statistical analysis

Medians and interquartile ranges were used, except for the calculation of measurement error which required the mean and standard deviation of the NBSS total score and individual domains. For our reliability assessment where the NBSS was measured at two time points, we used an intra-class correlation coefficient ( $ICC_{2,1}$ ) = 0.85 with a lower confidence bound of 0.75,  $\alpha = 0.05$ , and  $\beta = 0.20$  to estimate a minimal sample size for the reliability analysis of 79 people [11].

Internal consistency was assessed using Cronbach's alpha (with values  $\geq 0.70$  considered good internal consistency) [12], and with item to domain correlations (a Pearson correlation coefficient  $\geq 0.30$  was considered a moderate to strong association). While the full cohort was used to assess validity, a subset of this cohort (based on data availability and treatment stability) was used to assess reliability. Reliability was assessed using a test–re-test methodology comparing the original NBSS score with the repeat NBSS score at the 3-month reassessment using an  $ICC_{2,1}$  [13]; this analysis was restricted to the subset of people who had experienced no change in their urinary health based on the available data. They were identified by selecting people who self-reported no hospitalizations, surgeries, changes to medications, or changes to bladder management method between enrollment and the first 3-month follow-up. The validity of the domains was assessed by testing hypothesized correlations between NBSS domains and other variables collected in the study. Correlations  $>0.70$  were considered strong,  $0.70-0.30$  considered moderate and those  $<0.30$  considered weak [14]. The Spearman's rank (for ordinal variables), point-biserial (for binary variables), or Pearson's correlation coefficient (for continuous variables) were used as appropriate.

Finally, the measurement error for each of the domains, and for the total score was determined. This was estimated both using the standardized mean difference (SMD, half of the between person standard deviation at baseline), and standard error of the mean (SEM, calculated as between person standard deviation at baseline multiplied by the square root of 1 minus the  $ICC_{2,1}$ ) [15]. The smallest real difference (SRD), using a 90% confidence interval was calculated from the SEM for both group level comparisons (SEM multiplied by 1.64) and individual comparisons (SEM multiplied by 2.31) [16]. Group level SRD are the most relevant, and identify differences between groups,

**Table 1** Description of the study cohort demographics

	SCI patient cohort ( <i>N</i> = 609)
Age (years)	48 (IQR 36–57)
Male	410/609 (67%)
Body-mass index (kg/m <sup>2</sup> )	25.1 (IQR 21.8–29.7)
Highest education level	
High school or less	122 (20%)
Some college	225 (37%)
Bachelor's degree	164 (27%)
Some graduate degree	7 (1%)
Graduate degree	89 (15%)
Income	
<\$20 000	123 (20%)
\$20–40 000	115 (19%)
\$40 000–75 000	117 (19%)
>75 000	150 (25%)
Do not know	74 (12%)
Years since SCI	11.5 (IQR 5.8, 22.7)
SCI Level	609/609 (100%)
Cervical	285/609 (47%)
Thoracic	265 (44%)
Lumbar/sacral	42 (7%)
ASIA class	
A	167 (28%)
B	75 (12%)
C	61 (10%)
D	42 (7%)
E	3 (<1%)
Unknown	258 (43%)
Voiding method	
Indwelling catheter/urostomy	143 (23%)
Condom catheter	28 (5%)
Intermittent catheters	381 (63%)
Spontaneous/triggered voiding	57 (9%)

whereas individual SRD identifies the real difference when considering change in a single individual's score.

SAS 9.4 and R 3.4.0 were used for at statistical analysis, and a two-sided  $p < 0.05$  was considered significant and 95% confidence intervals were reported where applicable. The Consensus-based standards for the selection of health measurement Instruments checklist was used to ensure complete study reporting [17].

## Results

There were 644 people with a confirmed SCI who had initiated study enrollment during the interim study period.

Of these, data from 14 people were not used as they failed to complete the enrollment process, and data from 21 people could not be used due to missing NBSS data. Our final group of 609 people had a median age of 48 (IQR 36–57) years, were 67% were male, and most used CIC (63%, Table 1). Our cohort was representative of various socioeconomic groups and educational backgrounds. The median NBSS total score at study enrollment was 22 (IQR 15–30), and the median NBSS QOL was “Mixed” (IQR “Mostly unsatisfied”—“Mostly satisfied”). The median NBSS domain scores were incontinence (9, IQR 2–14), storage and voiding (7, IQR 4–10), and consequences (6, IQR 5–8).

## Validity assessment

The item-to-domain correlations were all moderate to strong ( $\geq 0.30$ ) for the incontinence and storage and voiding domains; 3 questions from the consequences domains had only weak item-to-domain correlations (Supplementary Appendix 1). Cronbach's alpha was calculated for the incontinence (0.93), storage & voiding (0.76), consequences (0.49), and total score (0.85). Correlations were assessed between the NBSS or its components and other variables collected as part of the existing study protocol (Table 2). People with missing data for the additional measurement tools required for the hypothesis testing validity assessment were excluded from that specific correlation (<1%). There was a correlation of  $r = 0.50$  between the SCI-QOL Bladder Management Complications tool and the NBSS consequences domain and weak correlations between the SCI-QOL Bladder Management complications tool and the NBSS total score ( $r = 0.28$ ) and QOL question ( $r = 0.29$ ) were observed. The SF-12 physical and mental domains had a weak negative correlation with the NBSS QOL question. The NBSS consequences domain had weak to moderate correlations with prior UTIs and renal/bladder stone procedures.

## Reliability assessment

Of the 609 people, 349 had 3-month followup data. Of these, 163 were excluded due to a potential change in their bladder function or general health (they had a reported change to their medications, bladder management strategy, or a new surgery or hospitalization), and an additional 12 were excluded due to missing NBSS data in the follow-up assessment. In our final cohort of 174 presumably urologically stable people, the test–retest reliability based on an ICC<sub>2,1</sub> was  $>0.75$  for all domains, the QOL question, and the total score (Table 3).

**Table 2** Assessment of hypothesized relationships between NBSS components and external measures

NBSS component	Comparison measure	Hypothesized correlation	Actual correlation <sup>a</sup> ( <i>r</i> , 95% CI)
NBSS QOL question	Bladder management complications score	Moderate +	+0.29 (0.22–0.36)
NBSS QOL question	SF12-physical domain	Weak –	–0.16 (–0.24––0.08)
NBSS QOL question	SF12-mental domain	Weak –	–0.28 (–0.35––0.20)
NBSS consequences domain	History of renal/bladder stone procedures	Weak +	+0.16 (0.08–0.23)
NBSS consequences domain	Number of UTIs in the prior year	Weak +	+0.51 (0.45–0.57)
NBSS consequences domain	Hospitalizations for UTIs in the prior year	Weak +	+0.21 (0.13–0.28)
NBSS consequences domain	Bladder management complications score	Moderate +	+0.50 (0.44–0.56)
NBSS total score	Bladder management complications score	Moderate +	+0.28 (0.20–0.35)

Correlation strength was defined as low <0.3, moderate 0.3–0.7, or high as >0.7. A positive correlation is expected when the scale orientations are the same (ie, lower score means worse QOL), or negative when the scale orientations are different

<sup>a</sup>All correlations are significant ( $p < 0.01$ )

**Table 3** Test–retest reliability of the NBSS over a 3 month period ( $n = 174$ )

	Score change (95% CI)	ICC <sub>2,1</sub> (95% CI)
Domains		
Incontinence	–1.03 (–1.7––0.37)	0.78 (0.71–0.83)
Storage and voiding	–0.28 (–0.61–+0.06)	0.83 (0.77–0.87)
Consequences	–0.28 (–0.57–+0.02)	0.76 (0.69–0.81)
NBSS QOL question	–0.20 (–0.31––0.09)	0.79 (0.71–0.84)
NBSS total	–1.59 (–2.5––0.67)	0.79 (0.72–0.85)

The intraclass correlation coefficient (ICC<sub>2,1</sub>) measures the agreement between the patients' initial score and the followup score

## Measurement error

The measurement error for the domains and the total score were very similar between the SMD and SEM (Table 4). The smallest real difference for group level comparisons (as would commonly be done in clinical research) ranged from 0.9 for the QOL question to 7.7 for the NBSS total score.

## Discussion

Choosing the right measurement tool for a research project can be challenging. In the field of neurourology, there are few tools that have been developed specifically for bladder-related QOL or symptoms, and most studies have used cross-validated instruments from other populations or questionnaires which have not been validated in a neurourology population [18]. Our multicenter prospective cohort study of the internal consistency, validity, and reliability of the NBSS in a large population of people with a SCI yielded results similar to the originally reported results in a group of people with different reasons for neurogenic bladder dysfunction [6]. For example, the

Cronbach's alpha for the overall NBSS was quite similar (0.85 vs. 0.89); however, for the consequences domain it was lower than previously measured (0.49 vs. 0.69). Cronbach's alpha is higher when the underlying construct is more consistent, and therefore a domain trying to capture urologic morbidity (which is quite variable among people) would be expected to be somewhat problematic. The initial validation study included a population of people with a SCI or MS with lower consequences scores, and this may explain why Cronbach's alpha for this domain was higher in the initial study population. The higher urologic morbidity generally seen in this larger cohort motivated by enrollment in a bladder-related QOL survey (and a significant number recruited from tertiary neurourology clinics) likely have more urinary consequences, and this intensifies the variability within this domain.

In general, the NBSS was consistent in its relationships with other measurement tools and clinical variables. Urinary specific QOL (the NBSS QOL question) only had a weak correlation with overall QOL (measured by the SF-12), which was expected given the multiple physical and social factors that determine overall QOL. The validated SCI-QOL bladder management complications tool (with questions about urinary tract infections, and the impact of bladder issues) was moderately correlated with the NBSS consequences domain, demonstrating the expected link between urinary morbidity and QOL. The magnitude of these correlations are in keeping with validation studies of other questionnaires in the neurogenic population (for example the I-QOL and Qualiveen-SF) [19, 20]. The test–retest reliability was appropriate, with ICC values >0.75 [21]. The values were however lower than the previously reported values of 0.91–0.86 (which were calculated in the original study using a median 3-week re-test period as compared to a 3 month test–re-test time period in this study). The longer the period between questionnaire administrations, the more likely a real (and potentially

**Table 4** Measurement error and smallest real difference

	Mean score (standard deviation) $N = 609$	Standardized mean difference	Standard error of the mean	Group: SRD <sub>90</sub>	Individual: SRD <sub>90</sub>
Incontinence	9.2 (6.9)	3.5	3.2	5.3	7.5
Storage and voiding	6.9 (4.2)	2.1	1.7	2.8	4.0
Consequences	6.6 (3.0)	1.5	1.5	2.4	3.4
NBSS QOL question	2.0 (1.2)	0.6	0.5	0.9	1.3
NBSS total	22.7 (10.3)	5.2	4.7	7.7	10.9

undetected) change will have occurred among the people, which results in a lower ICC. This likely explains the differences in the reliability measurement, and in addition it is also possible that some people did experience a real change in bladder symptoms that was not identified with the questions we had to use to detect change.

Statistically significant differences in questionnaire scores are dependent on sample size, which is why the SEM, and the SRD are useful characteristics to know. A difference between two groups of people that is greater than the SEM (which indicates the smallest detectable change) can be generalized as “a little better”, whereas a difference greater than the SRD (which indicates a more meaningful change) can be generalized as “a good deal better” [15]. For the total NBSS score, a change of 5–8 points, or 0.5–1.0 points on the QOL question is likely to represent a small but real change. This magnitude of change is consistent with other symptom scales, for example the American Urologic Association Symptom Score for benign prostatic hyperplasia [22].

Limitations of our study are important to acknowledge. People were recruited through social media and neurourology/physiatry clinics, and therefore our data may be skewed towards those with a higher level of technological engagement or urologic complications; this potentially limits the external generalizability of our results. Our assessment of validity was limited to the variables and questionnaires that were included for the primary objective of this study. In this study, our reliability measurement likely underestimates the true reliability of the NBSS given the long time between administrations, and the lack of a question specifically asking about a change in bladder function or symptoms. As the SRD is derived from this reliability measurement, the actual SRD may actually be lower. As an example, when the original reliability estimates are used [6], the group level SRD for the total NBSS score is 5.1 as opposed to 7.7. Finally, while minimally important clinical change (which uses a relevant indicator of change which would influence management) is an attractive metric of meaningful change, we could not determine that with the current study data.

## Conclusions

The NBSS shows good internal consistency, validity, and reliability in a large population of people with SCI. However, the NBSS consequences domain had a low internal consistency, and this should be taken into account if it is to be used as a stand-alone domain.

**Acknowledgements** This work was (partially) supported through a Patient-Centered Outcomes Research Institute (PCORI) Award (CER14092138). Disclaimer: All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI.

**Compliance with ethical standards** All ethical standards for research using human subjects were met, in keeping with site-specific research ethic boards.

**Conflict of interest** The authors declare that they have no competing interests.

**Disclaimer** All statements in this report, including its findings and conclusions, are solely those 36 of the authors and do not necessarily represent the views of PCORI.

## References

- Powell CR. Not all neurogenic bladders are the same: a proposal for a new neurogenic bladder classification system. *Transl Androl Urol.* 2016;5:12–21.
- Ku JH. The management of neurogenic bladder and quality of life in spinal cord injury. *BJU Int.* 2006;98:739–45.
- 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.
- Costa P, Perrouin-Verbe B, Colvez A, Didier J, Marquis P, Marrel A, et al. Quality of life in spinal cord injury patients with urinary difficulties. Development and validation of qualiveen. *Eur Urol.* 2001;39:107–13.
- Tulsky DS, Kisala PA, Tate DG, Spungen AM, Kirshblum SC. Development and psychometric characteristics of the SCI-QOL bladder management difficulties and bowel management difficulties item banks and short forms and the SCI-QOL bladder complications scale. *J Spinal Cord Med.* 2015;38:288–302.
- Welk B, Morrow S, Madarasz W, Baverstock R, Macnab J, Sequeira K. The validity and reliability of the neurogenic bladder symptom score. *J Urol.* 2014;192:452–7.

7. Best KL, Ethans K, Craven BC, Noreau L, Hitzig SL. Identifying and classifying quality of life tools for neurogenic bladder function after spinal cord injury: a systematic review. *J Spinal Cord Med.* 2016;7:1–25.
8. Myers JB, Patel DP, Elliott SP, Stoffel JT, Welk B, Jha A, et al. Mending gaps in knowledge: collaborations in neurogenic bladder research. *Urol Clin North Am.* 2017;44:507–15.
9. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220–33.
10. Forchheimer M, McAweeney M, Tate DG. Use of the SF-36 among persons with spinal cord injury. *Am J Phys Med Rehabil.* 2004;83:390–5.
11. Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med.* 2001;20:3205–14.
12. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess.* 2003;80:99–103.
13. Bravo G, Potvin L. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. *J Clin Epidemiol.* 1991;44:381–90.
14. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* 2012;24:69–71.
15. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11:171–84.
16. Dvir Z. Difference, significant difference and clinically meaningful difference: the meaning of change in rehabilitation. *J Exercise Rehabil.* 2015;11:67–73.
17. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19:539–49.
18. 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. *Neurourol Urodyn.* 2016;35:8–14.
19. Schurch B, Denys P, Kozma CM, Reese PR, Slaton T, Barron R. Reliability and validity of the Incontinence Quality of Life questionnaire in patients with neurogenic urinary incontinence. *Arch Phys Med Rehabil.* 2007;88:646–52.
20. Bonniaud V, Bryant D, Parratte B, Guyatt G. Development and Validation of the Short Form of a Urinary Quality of Life Questionnaire: SF-Qualiveen. *J Urol.* 2008;180:2592–8.
21. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.
22. Rees J. Patients not P values. *BJU Int.* 2015;115:678–9.