



Reliability of the international spinal cord injury urodynamic basic data set version 1.0

Kittamet Dejkiengkraikul^{1,2} · Sintip Pattanakuhar¹ · Tuankasfee Hama^{1,3} · Siam Tongprasert¹

Received: 20 April 2020 / Revised: 23 May 2020 / Accepted: 27 May 2020 / Published online: 8 June 2020
© The Author(s), under exclusive licence to International Spinal Cord Society 2020

Abstract

Study design A cross-sectional study.

Objective To assess interrater and intrarater reliability of the International Spinal Cord Injury (SCI) Urodynamic Basic Data Set (UBS) version 1.0.

Setting Urodynamic clinic at Maharaj Nakorn Chiang Mai Hospital.

Methods Two raters independently analyzed urodynamic tracings from 50 patients and completed the UBS twice, each test 1 month apart. The interrater and intrarater reliability of this data set were analyzed using Kappa, Weighted Kappa, and the Intraclass correlation coefficient (ICC).

Results Of the 50 patients, 72% were male. The mean (SD) age was 48.2 (16.6) years. The median time (IQR) since the injury was 27 months (0–101 months). The interrater reliability of the items of UBS were substantial to almost perfect (0.78–0.99). The intrarater reliability of the first rater was fair to almost perfect (0.37–1.00). The intrarater reliability of the second rater was moderate to almost perfect (0.51–1.00). Relatively low interrater and intrarater reliability were observed in bladder compliance and urethral function items.

Conclusion The first version of UBS has acceptable interrater and intrarater reliability on most items. Although bladder compliance and urethral function have problematic interrater and intrarater reliability, they have been adjusted in the second version. Due to its simplicity and reliability, UBS is clinically useful for urodynamic assessment in people with SCI.

Introduction

Neurogenic lower urinary tract dysfunction (NLUTD) is a common problem for people with spinal cord injury (SCI). It leads to many complications and can cause premature death. A urodynamic study is crucial in the evaluation and planning for the management of NLUTD [1–3]. Unfortunately, its interpretation is subjective, which could result in

different viewpoints among raters and within the same rater. Therefore, the standard recording tool is needed to reduce these errors.

The International SCI Urodynamic Basic Data Set (UBS) (Supplementary data 1) was developed by the International Spinal Cord Society (ISCoS), the American Spinal Injury Association (ASIA), the International Continence Society (ICS), and the European Association of Urology (EAU), aiming to introduce a standardized format for collecting and reporting information based on an urodynamic study [4]. This data set is also a part of the International SCI Data Sets [5]. The UBS contains nine different sections, which include the following: performing date, bladder sensation during filling cystometry, detrusor function, compliance during filling cystometry, urethral function during voiding, detrusor leak point pressure, maximum detrusor pressure, cystometric bladder capacity, and post-void residual volume.

Before using a data set in clinical settings, one should be confident that it has acceptable validity and reliability. Therefore, it is necessary to conduct a psychometric study

Supplementary information The online version of this article (<https://doi.org/10.1038/s41393-020-0496-5>) contains supplementary material, which is available to authorized users.

✉ Siam Tongprasert
siam.tongpr@cmu.ac.th

¹ Department of Rehabilitation Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

² Chiangmai Ram Hospital, Chiang Mai, Thailand

³ Department of Rehabilitation Medicine, Yala Hospital, Yala, Thailand

of the UBS before using it with patients. Focusing on reliability, it consists of interrater and intrarater components. Interrater reliability is the consistency of the result rated by different persons. Intrarater or test-retest reliability is the consistency of the result rated by the same person at different time points. The validity of the urodynamic study is difficult to evaluate since there has been no gold standard measurement for reporting urodynamic results in people with SCI yet [4]. Thus, the objective of this study is to evaluate reliability, including the interrater and intrarater reliability of the UBS.

Methods

Fifty urodynamic tracings from 50 patients with SCI were included in our study. All of them were recorded during the urodynamic study in compliance with the ICS standard [6]. Demographic data were collected by referring to the patient's history documented in medical records and were collected using the International SCI Core Data Set [7] by the first author (KD). Then, two raters with different experiences in the urodynamic study were brought in to interpret the data. The first rater was a rehabilitation medicine resident, who has had 1 year of experience in urodynamics (TH). The other rater was a rehabilitation medicine consultant, who has had 5 years of experience in urodynamics (SP). Both raters had practiced undergoing urodynamic interpretation using the ISCoS training cases [8] before the study began. After that, they independently interpreted all tracings and completed the UBS. Both raters interpreted the same urodynamic tracings again 1 month after the first evaluation to assess the intrarater reliability. The reliability were evaluated by the first author (KD).

Samples

Urodynamic tracings were sampled from the database at the urodynamic clinic, Department of Rehabilitation Medicine, Faculty of Medicine, Chiang Mai University. All tracings were recorded during the urodynamic studies that were performed between July 2015 and June 2017. Before being interpreted, all tracings were de-identified by urodynamic nurses. This was done by covering the identification sections.

Inclusion criteria

The Urodynamic tracings of patients with SCI from both traumatic and nontraumatic causes were included in the study.

Exclusion criteria

The Urodynamic tracings that had poor quality due to artifacts [artifacts from the urodynamic machine (poor transducer placement) and/or artifacts from the patient (bowel movements)], were excluded from this study.

Sample size

According to the recommendation given by the International SCI Data Set Committee regarding reliability testing, 50 urodynamic tracings were evaluated in this study [9].

Materials

International SCI Core Data Set [7].

International SCI Urodynamic Basic Data Set version 1.0 [4].

Statistical analysis

The program SPSS version 23 for Windows was used for statistical analyses. Interrater and intrarater reliability were analyzed using the kappa statistic and intraclass correlation coefficient (ICC) [10, 11]. Unweighted Kappa statistics were used for categorical parameters. Weighted Kappa was used for the ordinal parameters. The ICC was used for continuous parameters. The interpretation of Kappa and ICC values were determined using the following criteria: ≤ 0 , poor agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, almost perfect agreement [11]. Therefore, the reliability is acceptable when the Kappa and ICC are more than or equal to 0.7 [12].

Result

Demographic data of the sampled tracings

Urodynamic tracings from 70 patients were reviewed. Twenty tracings were excluded. For instance, 12 tracings were excluded due to artifacts from the patients, whereas eight tracing were excluded due to artifacts from the urodynamic machine. Of 50 patients whose tracings included in this study, 72% were male. The mean (SD) age of the patients was 48.2 (16.6) years. The median (IQR) time since the injury occurred was 27 (0–101) months. Eighty-six percent of tracings were from people with suprasacral lesions and the remaining were from those with sacral and subsacral lesions. The most common neurological level of

injury and ASIA Impairment Scale (AIS) were of the thoracic level (62%) and AIS A (44%), respectively. (Table 1)

Table 2 shows the interrater and intrarater reliability of the UBS. Weighted Kappa statistics were used in the items “bladder sensation” and “detrusor function” since no “unknown” choices were selected in these two items,

Table 1 Demographic data of the patients with SCI whose tracings included in the study.

Parameters	N (50)
Age (years) ^a	48.2 (16.6)
Time since injury (months) ^b	27 (0–101)
Gender ^c	
Male	72
Female	28
Etiology ^c	
Traumatic	70
Non-traumatic	30
Neurological level ^c	
C2-C7	24
T1-T12	62
L1-L5	12
S1-S4	2
ASIA impairment scale (AIS) ^c	
A	44
B	14
C	8
D	34

SCI spinal cord injury, ASIA American Spinal Injury Association.

^aMean (standard deviation; S.D).

^bMedian (interquartile range; IQR).

^cPercent.

Table 2 Interrater and intrarater reliability of the international spinal cord injury urodynamic basic data set.

Item	Interrater reliability	1st intrarater reliability	2nd intrarater reliability
Bladder sensation during filling cystometry ^a	0.81 (0.70–0.93)	0.76 (0.63–0.90)	0.76 (0.63–0.90)
Detrusor function ^a	0.78 (0.63–0.93)	0.72 (0.48–0.96)	0.85 (0.69–1.00)
Compliance during filling cystometry ^b	0.56 (0.25–0.87)	0.83 (0.61–1.00)	0.51 (0.19–0.82)
Urethral function during voiding ^b	0.32 (0.17–0.46)	0.37 (0.20–0.55)	0.52 (0.34–0.70)
Detrusor leak point pressure ^c	NA	NA	NA
Maximum detrusor pressure ^c	0.91 (0.83–0.95)	0.95 (0.91–0.97)	0.99 (0.98–1.00)
Cystometric bladder capacity ^c	0.99 (0.99–1.00)	1.00	1.00
Post-void residual volume ^c	0.93 (0.87–0.96)	0.99 (0.98–1.00)	0.95 (0.91–0.97)

NA not applicable.

^aWeighted kappa.

^bUnweighted kappa.

^cICC with (95% CI).

making them an ordinal parameter. Most of the items of UBS had substantial to almost perfect interrater reliability (0.78–0.99). Only two items had fair to moderate interrater reliability, namely the compliance during filling cystometry (0.56) and the urethral function during voiding cystometry (0.32). Focusing on the intrarater reliability, the intrarater reliability of the first rater were fair to almost perfect (0.37–1.00). The intrarater reliability of the second rater were moderate to almost perfect (0.51–1.00). It was noticed that the interrater and intrarater reliability of the detrusor leak point pressure could not be analyzed, since both raters answered the choices “not applicable” and “unknown” in most of the tracings.

Discussion

This study shows that most of the items included in the UBS had acceptable interrater reliability, as indicated by the Kappa and ICC values, which were more than 0.70 [12]. The interrater reliability of almost all items, even between raters with different urodynamic experience, was acceptable. It is noteworthy that, in this study, each rater has at least 1 year urodynamic experience. Therefore, it can be interpreted that the UBS is appropriate for any investigators, regardless of their urodynamic experiences.

However, low interrater and relatively low intrarater reliability on bladder compliance and urethral function during voiding were found. The compliance during filling cystometry parameter of the UBS had moderate interrater reliability (0.56) and almost perfect intrarater reliability for the first rater (0.83), but moderate intrarater reliability for the second rater (0.51). This might be because each rater picked up different points on tracing to calculate the compliance. Also, the low bladder compliance’s cut-off level in

the UBS was set from the expert consensus as lower than 10 mL/cm H₂O. This level is relatively low compared with the cut of value of our institute (20 mL/cm H₂O), which follows the standardization of terminology in NLUTD of the ICS Standardization Committee [13]. This could make the raters confused as well as lower the reliability of the compliance item. Urethral function during voiding cystometry had fair interrater reliability (0.32) and fair intrarater reliability for the first rater (0.37), but moderate intrarater reliability for the second rater (0.52). This might have occurred due to the effect of urethral function measurement. In this study, the urethral function was indirectly measured from an electromyography (EMG) of the pelvic muscles, of which the EMG signal is commonly interfered by nearby muscles.

In all of our knowledge, this is the first study investigating the reliability of the UBS. Focusing on the reliability of the other form of a urodynamic report, Venhola et al. investigated agreements in interpreting urodynamic measurements in children by four raters [14]. They found a good agreement on the detrusor function (Kappa = 0.37–1.00) and a slight agreement on the urethral function (Kappa = 0.09–0.27), which were comparable to the results of this study. The agreement of the bladder compliance report in the study of Venhola varied immensely (Kappa = 0.06–0.76) [14]. This supported our hypothesis, which stated that the inconsistency resulted from each rater picking up different points on the tracing to calculate compliance. In addition, another study from Whiteside et al. investigated urodynamic interpretation in people with female pelvic disorders by six raters [15]. They found a fair agreement on the urethral function (Kappa = 0.25), which is also comparable to the result of our study [15]. A different study from Dudley et al. evaluated interrater reliability in pediatric urodynamic tracings [16]. They also found low agreement on urethral function [16]. The consistent findings from each study emphasizes the difficulty in interpreting urethral function.

Our study had some limitations. Firstly, only two raters have participated in this study. Secondly, both raters were from the same department, so their level of agreement could be high due to them having participated in the same training program. Next, the exact time it took to complete the UBS was not recorded. However, both raters reported not spending more than 5 min on each tracing. Lastly, the interrater reliability of each item partly varied by the difficulty of questions. If a currently evaluated case was easy enough for both raters to correctly answer, the interrater reliability will be good, and *vice versa*. However, in clinical practice, urodynamic assessors would have guidance from the information regarding the history and physical examination of the patient whereas none of them was provided to the raters in this study. Henceforth, urodynamic interpretation in a real situation should be easier when compared with the present study and the effect of the

difficulty of the case on the interrater reliability should be attenuated.

From the authors' point of view, the advantages of the UBS are that it is concise and less time-consuming. Therefore, it is feasible for clinical use. However, the disadvantage of this data set is the absence of a clear separation between the filling and voiding phase assessment in some parameters. This makes it difficult for users to properly interpret detrusor function and urethral function. Fortunately, in the year 2018, the UBS working group has revised the UBS to version 2.0 [17] (Supplementary data 2). The detrusor function items are now divided into those in the voiding and filling phase. The low compliance cut-off level was also changed to <20 mL/cm H₂O [17], compared with initially being <10 mL/cm H₂O. These changes could clarify the terminology problems and help avoid misinterpretation. Therefore, it is hypothesized that the interrater and intrarater reliability of the second version of UBS would be greater than the first version investigated in this study. A further psychometric study to evaluate the reliability of the second version of UBS will need to be conducted in order to confirm this hypothesis.

Conclusion

The first version of UBS has acceptable interrater and intrarater reliability on most items. Although bladder compliance and urethral function have problematic interrater and intrarater reliability, these issues have now been adjusted in the second version. Due to its simplicity and reliability, the UBS is clinically useful for urodynamic assessment in people with SCI.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions KD designed the research questions, collected, and analyzed the data, drafted the paper, and wrote the final version of the paper. SP collected and analyzed the data, drafted the paper, and wrote the final version of the paper. TH designed the research questions and commented on the final version of the paper. ST designed the research question, collected, and analyzed the data and commented on the final version of the paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The study protocol was exempted from ethical approval by the Institutional Ethics Committee of the Faculty of

Medicine, Chiang Mai University, Chiang Mai, Thailand since no identifiable data of the sampled tracings were exposed to the researchers. The authors certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during all courses of this study.

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

References

1. Todd AL, Donald RB, Graham HC, Bruce GG, Suzanne LG, Angela J. Bladder management for adults with spinal cord injury: a clinical practice guideline for health-care providers. *J Spinal Cord Med*. 2006;29:527–73.
2. Nosseir M, Hinkel A, Pannek J. Clinical usefulness of urodynamic assessment for maintenance of bladder function in patients with spinal cord injury. *Neurourol Urodyn*. 2007;26:228–33.
3. Danforth TL, Ginsberg DA. Neurogenic lower urinary tract dysfunction: how, when, and with which patients do we use urodynamics? *Urol Clin North Am*. 2014;41:445–52.
4. Biering-Sorensen F, Craggs M, Kennelly M, Schick E, Wyndaele JJ. International urodynamic basic spinal cord injury data set. *Spinal Cord* 2008;46:513–6.
5. Biering-Sorensen F, Charlifue S, DeVivo M, Noonan V, Post M, Stripling T, et al. International spinal cord injury data sets. *Spinal Cord* 2006;44:530–4.
6. Rosier P, Schaefer W, Lose G, Goldman HB, Guralnick M, Eustice S, et al. International continence society good urodynamic practices and terms 2016: Urodynamics, uroflowmetry, cystometry, and pressure-flow study. *Neurourol Urodyn* 2017;36:1243–60.
7. DeVivo M, Biering-Sorensen F, Charlifue S, Noonan V, Post M, Stripling T, et al. International spinal cord injury core data set. *Spinal Cord* 2006;44:535–40.
8. Biering-Sorensen F, Craggs M, Kennelly M, Schick E, Wyndaele JJ. The international spinal cord injury urodynamic basic data set version 1.0 training cases English [Online]. 2007 [Cited 2016 Nov 13]. <http://www.iscos.org.uk/international-sci-urodynamic-data-sets>.
9. Biering-Sorensen F, Alexander MS, Burns S, Charlifue S, DeVivo M, Dietz V, et al. Recommendations for translation and reliability testing of international spinal cord injury data sets. *Spinal Cord* 2011;49:357–60.
10. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85:257–68.
11. Sainani KL. Reliability statistics. *PM R* 2017;9:622–8.
12. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
13. Stohrer M, Goepel M, Kondo A, Kramer G, Madersbacher H, Millard R, et al. The standardization of terminology in neurogenic lower urinary tract dysfunction: with suggestions for diagnostic procedures. International Continence Society Standardization Committee. *Neurourol Urodyn* 1999;18:139–58.
14. Venhola M, Reunanen M, Taskinen S, Lahdes-Vasama T, Uhari M. Interobserver and intra-observer agreement in interpreting urodynamic measurements in children. *J Urol*. 2003;169:2344–6.
15. Whiteside JL, Hijaz A, Imrey PB, Barber MD, Paraiso MF, Rackley RR, et al. Reliability and agreement of urodynamics interpretations in a female pelvic medicine center. *Obstet Gynecol*. 2006;108:315–23.
16. Dudley AG, Casella DP, Lauderdale CJ, Zhao S, Chen H, Tanaka ST, et al. Interrater reliability in pediatric urodynamic tracings: a pilot study. *J Urol*. 2017;197:865–70.
17. Pannek J, Kennelly M, Kessler TM, Linsenmeyer T, Wyndaele JJ, Biering-Sørensen F. International spinal cord injury urodynamic basic data set (version 2.0). *Spinal Cord Ser Cases*. 2018;4:98.