

Original Article

Int Neurourol J 2026;30(2):155-162
https://doi.org/10.5213/inj.2652052.026
pISSN 2093-4777 · eISSN 2093-6931

INTERNATIONAL NEUROUROLOGY JOURNAL
INJ
OPEN ACCESS

Interobserver Agreement of Filling Cystometry Parameters in Children With Neurogenic Lower Urinary Tract Dysfunction

Yunus Emre Genc[1], Faruk Arslan[1], Onur Can Ozkan[2], Osman Akyuz[3], Raziye Ergun[4], Ahmet Ozgur Guctas[1], Selcuk Yucel[2], Kamil Cam[1], Tufan Tarcan[1,5], Cagri Akin Sekerci[2]

[1]Department of Urology, School of Medicine, Marmara University, Istanbul, Turkey
[2]Division of Pediatric Urology, Department of Urology, School of Medicine, Marmara University, Istanbul, Turkey
[3]Department of Urology, Trabzon Kanuni Training and Research Hospital, Trabzon, Turkey
[4]Division of Pediatric Urology, Department of Pediatric Surgery, Kocaeli State Hospital, Kocaeli, Turkey
[5]Department of Urology, School of Medicine, Koç University, Istanbul, Turkey

Purpose: Urodynamic studies (UDS) are the gold standard for evaluating pediatric lower urinary tract dysfunction (LUTD). However, previous studies have shown that the interpretation of cystometric findings may vary according to the observer's level of experience and the criteria used for evaluation. This study aimed to investigate interobserver differences in the reporting of cystometric parameters in children with spina bifida.

Methods: Patients who underwent UDS between 2020 and 2025 for neurogenic LUTD secondary to spina bifida were retrospectively reviewed. Studies that did not reach 80% agreement for compliance with good urodynamic practice standards were excluded. Detrusor overactivity (DOA) and bladder compliance were independently evaluated by 4 pediatric urologists and 3 urologists experienced in neuro-urology. Agreement was assessed using Fleiss' kappa test.

Results: A total of 100 children, including 53 females and 47 males, were included; the median age was 6.5 (range, 1–17) years. Agreement of at least 80% was reached in 86% of evaluations for DOA and 82% of evaluations for compliance. Based on the ratings of 7 observers, the Fleiss' kappa value was 0.60 for DOA, indicating moderate agreement, and 0.46 for bladder compliance, indicating weak agreement.

Conclusions: This study demonstrated substantial interobserver variation in the interpretation of UDS, even among experienced centers and observers. Because UDS are essential for the diagnosis, treatment, and follow-up of patients with spina bifida, standardized criteria are needed to improve interobserver consistency in this patient group. Multicenter prospective studies involving observers with varying levels of experience are warranted to improve concordance in UDS reporting.

Keywords: Urodynamics; Neurogenic bladder dysfunction; Spinal dysraphism; Child; Interobserver variability

- **Grant/Fund Support:** This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
- **Research Ethics:** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Marmara University, School of Medicine, Ethics Committee (No: Ref: 09.2025.25-0787). Written informed consents from legal guardians were received and archived during the study.
- **Conflict of Interest:** No potential conflict of interest relevant to this article was reported.

Corresponding author: Cagri Akin Sekerci https://orcid.org/0000-0002-0334-2466
Division of Pediatric Urology, Department of Urology, School of Medicine, Marmara University, Fevzi Çakmak Mah., Muhsin Yazicioglu Cad. No:10 Ust Kaynarca, Pendik, Istanbul, Turkey
Email: cagri_sekerci@hotmail.com
Submitted: February 14, 2026 / **Accepted after revision:** March 17, 2026

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Neurogenic lower urinary tract dysfunction (LUTD) in children is predominantly attributable to congenital spinal dysraphism, with spina bifida representing the leading cause and the largest subgroup requiring long-term neuro-urological surveillance and intervention. Improvements in healthcare systems, access to evidence-based treatment, and proactive management protocols have substantially reduced mortality rates. As survival has improved, long-term disease-related morbidities and quality-of-life impairment have become better characterized through extended follow-up data [1].

Comorbidities such as urinary tract infection and renal damage occur in almost 60% of affected patients if they are not managed appropriately, resulting in a 15% risk of renal deterioration and a 1% risk of end-stage renal failure. To identify patients at risk for these outcomes, international guidelines recommend baseline and periodic urodynamic studies (UDS) as a cornerstone of diagnosis and follow-up in patients with spina bifida, because UDS are considered the gold standard for monitoring LUTD [2].

UDS also guide therapeutic decision-making across the broad clinical spectrum of neurogenic LUTD, including the prescription of anticholinergic medication, initiation of clean intermittent catheterization, bladder outlet procedures, augmentation, and urinary diversion. Because UDS play a central role in treatment decisions, standardized and accurate interpretation is essential, particularly because many resulting interventions can have direct and irreversible effects on a patient's lifelong clinical course [3].

Accordingly, good urodynamic practice standards and recommended reporting terminology were described and updated by the International Children's Continence Society (ICCS) in 2015 [4] and by the International Continence Society (ICS) in 2016 [5].

Although some previous reports indicated good consistency among experienced observers for pressure measurement and artifact detection during cystometry [6, 7], most reports have found that UDS interpretation varies substantially among centers and physicians. Such variation may lead to misdiagnosis or differences in management strategies despite standardization efforts, consensus meetings, available UDS training modules, and emerging applications of artificial intelligence in interpretation [8-13].

The role of UDS is well established, particularly for pressure-flow studies that guide surgical decision-making in adult patients undergoing correction of bladder outlet obstruction [14]. Filling cystometry parameters, including detrusor or Valsalva leak point pressure and the presence of detrusor overactivity (DOA), are widely used for risk stratification in neurogenic LUTD [15] and, in selected cases, for evaluating stress and urgency urinary incontinence in nonneurogenic LUTD [16, 17]. However, few studies have systematically examined interobserver agreement in the evaluation of pediatric neurogenic LUTD, particularly for DOA and bladder compliance [7, 18]. This study aimed to measure interobserver agreement among experienced observers in a standardized setting for the evaluation of children with neurogenic LUTD at a tertiary referral center.

MATERIALS AND METHODS

After obtaining institutional ethics committee approval, we retrospectively reviewed archived UDS performed at our tertiary pediatric urology center between January 2020 and December 2025. The study population consisted of children with neurogenic LUTD secondary to spina bifida who underwent multichannel filling cystometry using an air-charged catheter-based urodynamic unit (Medical Measurement Systems, Dover, NH03820, USA). All data were anonymized before analysis.

Urodynamic acquisitions were performed according to the contemporary terminology and good urodynamic practice recommendations of the ICS [5]. Families were instructed to ensure that the rectum was emptied the night before the study to reduce artifactual bowel contractions. All cystometric studies were conducted in the presence of sterile urine culture findings, with patients in the supine or sitting position. Double-lumen air-charged 7F catheters were used to measure vesical pressure, and rectal balloon catheters were used to measure abdominal pressure, allowing continuous calculation of detrusor pressure. Before each study, catheters were calibrated at baseline atmospheric pressure, and zero referencing was confirmed at the level of the symphysis pubis. Bladder filling was performed with body-temperature sterile saline at an age-appropriate physiological filling rate, defined as 10% of expected bladder capacity (EBC) for age per minute or based on bladder diaries. EBC for age was calculated using the "age×30+30 mL" formula; for children younger than 2 years, it was calculated using the "body weight (kg)×7 mL" formula, as recommended by the ICCS [4]. The total infused volume during cystometry was also deter-

mined according to age-EBC and bladder capacity on voiding diaries.

All tracings were reviewed for inclusion by 4 certified pediatric urologists, and studies were retained if consensus reached 80%. Studies with excessive artifacts, including catheter dislocation, incomplete baseline stabilization, or uncontrolled abdominal contractions or crying, were excluded. Studies that did not meet the 2020–2021 ICS technical requirements for good urodynamic practice, including calibration, presence of vital waves, artifact recognition, and accurate labeling, were also excluded from the final dataset, as were patients with missing clinical data.

For each eligible UDS, the following parameters were extracted: baseline vesical and abdominal pressures, compliance measurements, presence and type of DOA, detrusor leak point pressure, end-filling detrusor pressure, and maximum cystometric capacity. Nonphasic DOA referred to terminal detrusor contractions, defined as a single involuntary detrusor contraction occurring at cystometric capacity that cannot be voluntarily suppressed and typically results in voiding or urinary leakage, and sustained detrusor contractions, defined as involuntary detrusor contractions during the filling phase that are maintained without spontaneous resolution and persist until voiding or termination of cystometry [5, 19]. Bladder compliance was calculated as change in volume divided by change in detrusor pressure during cystometric filling, using pressure values free from DOA and artifacts [5]. The selected UDS tracings were sent to observers through automated online anonymized software for blinded rating. In filling cystometry, the presence and type of DOA (absent, phasic, or nonphasic) and bladder compliance status (hypocompliant or normocompliant) were independently evaluated by 7 observers, including 4 board-certified pediatric urologists and 3 board-certified urologists experienced in neuro-urology. The observers had 10 to 30 years of practice experience and an approximate annual reading volume of 150 to 520 UDS per year. All observers were routinely involved in the clinical interpretation of UDS in children with neurogenic LUTD at our institution. Before the interpretation phase, the observers reviewed the interpretation criteria based on ICS recommendations [5]. The required sample size was calculated as 66 to 88 UDS, assuming a 2-category outcome and 7 independent raters, with mild to moderate agreement, a null kappa of 0.30, a target kappa of 0.50, 80% statistical power, and an alpha error of 0.05 [20].

All statistical analyses were conducted using IBM SPSS Statistics ver. 27.0 (IBM Co., USA). Normality was evaluated using the Kolmogorov-Smirnov test and visual assessment of histogram curves. Results are presented as medians with minimum and maximum values. Qualitative data were assessed using the chi-square test, and quantitative data were analyzed using the Mann-Whitney U-test. Consensus was defined as agreement of more than 80% among observers' ratings. A P-value of <0.05 was considered statistically significant. Interobserver agreement was calculated using Fleiss' kappa test, and the level of agreement was interpreted according to McHugh clustered report for Fleiss', Cohen, and weighted kappa statistics [21].

RESULTS

Of 125 UDS, 25 were excluded because they did not comply with good urodynamic practice standards: 4 had catheter dislocation, 13 had incorrect artifact labeling, 2 had inappropriate baseline calibration, and 6 included uncontrolled abdominal contractions. A total of 100 children, including 53 females and 47 males, were independently analyzed by 7 observers. The median age of the participants was 6.5 (range, 1–17) years. The median cystometric capacity of the included studies was 122 (range, 22–450) mL, the median maximum detrusor pressure during filling was 36 (range, 0–57) cm H_2O, the median detrusor pressure at the end of filling was 23 (range, 0–42) cm H_2O, and the median compliance was 10.2 (range, 1.4–580) mL/cm H_2O.

The Fleiss' kappa value for DOA was 0.60 among all patients, indicating moderate agreement. For compliance, the kappa value was 0.46, indicating weak agreement among the 7 observers (Table 1).

When the 4 pediatric urologists and the 3 urologists experienced in neuro-urology were analyzed separately, the kappa score among pediatric urologists was 0.546 for DOA and 0.484 for compliance. Among the urologists experienced in neuro-urology, the kappa score was 0.706 for DOA and 0.526 for compliance.

Overall, the mean agreement rate for the interpretation of DOA was 86%. Complete agreement (100%) was reached in 50% of patients; 35% were classified as having phasic DOA, and the remaining 15% were classified as having no DOA. When the agreement threshold was set at 80% (>5 of 7 raters in agreement), 75% of all patients met this criterion. Among UDS with more than 80% agreement, 28 patients were rated as having no DOA, 46 as having phasic DOA, and 1 as having nonphasic

Table 1. Interrater agreement of filling phase parameters in cystometric studies (n = 100)

Variable	Detrusor overactivity (DOA)			Compliance	
	No DOA	Phasic DOA	Nonphasic DOA	Normocompliant	Hypocompliant
Observer 1 (%)	43	49	5	65	33
Observer 2 (%)	31	52	17	55	45
Observer 3 (%)	34	58	8	35	65
Observer 4 (%)	41	57	2	42	58
Observer 5 (%)	44	49	5	58	40
Observer 6 (%)	37	59	3	33	66
Observer 7 (%)	35	60	3	40	58
Kappa score	0.60			0.46	
Agreement rate	86%			82%	
Level of agreement	Moderate			Weak	

In interpretation of DOA, "inconclusive study/abstention" rates were 3 for observer 1, 2 for observer 5, 1 for observer 6, 2 for observer 7. In interpretation of compliance, "inconclusive study/abstention" rates were 2 for observer 1, 2 for observer 5, 1 for observer 6, 2 for observer 7.

Table 2. Descriptive analyses of patients with an achieved agreement of > 80% for detrusor overactivity and compliance (n = 100)

Variable	> 80% agreement in DOA (n = 75)			> 80% agreement in compliance (n = 57)		
	Agreement on no DOA (n = 28)	Agreement on phasic DOA (n = 46)	P-value	Agreement on normocompliance (n = 28)	Agreement on hypocompliance (n = 29)	P-value
Age (yr)	6 (2–17)	5.5 (1–15)	0.180	6 (2–17)	8 (1–17)	0.962
Sex			0.620			0.463
Male	17	21		14	15	
Female	11	25		14	14	
Cystometric capacity/EBC ratio (%)	104 (23–208)	56 (8–166)	0.001	106 (23–108)	71 (8–166)	< 0.001
Maximum detrusor pressure (cm H_2O)	13 (2–39)	36 (15–57)	< 0.001	13 (2–41)	35 (15–57)	< 0.001
End-filling detrusor pressure (cm H_2O)	9 (1–28)	16 (4–42)	< 0.001	9.5 (1–22)	18 (8–42)	< 0.001
Compliance (mL/cm H_2O)	25 (4.3–580)	6.7 (1.4–32.5)	< 0.001	25.1 (4–580)	7.6 (1.4–17.4)	< 0.001

DOA, detrusor overactivity; EBC, expected bladder capacity.

DOA.

In descriptive analyses, significant differences were observed between patients without and with phasic DOA in the cystometric capacity/EBC ratio (104% vs. 56%, P = 0.001), maximum detrusor pressure (13 cm H_2O vs. 36 cm H_2O, P < 0.001), end-filling detrusor pressure (9 cm H_2O vs. 16 cm H_2O, P < 0.001), and compliance (25 mL/cm H_2O vs. 6.7 mL/cm H_2O, P < 0.001) (Table 2). In the remaining cohort, 71% agreement was observed in 8 patients, 57% agreement in 12, and 43% agreement in 5 patients.

For the interpretation of compliance among all patients, the mean agreement rate was 82%. Complete agreement was present in 37% of patients, of whom 21% were rated as hypocompliant and 16% as normocompliant. Agreement of more than 80% was reached in 57% of patients; 29% were rated as hypocompliant, and 28% were considered normocompliant.

Descriptive analyses revealed that the cystometric capacity/EBC ratio (106% vs. 71%, P < 0.001), maximum detrusor pressure (13 cm H_2O vs. 35 cm H_2O, P < 0.001), end-filling pressure (9.5 cm H_2O vs. 18 cm H_2O, P < 0.001), and compliance (25.1 mL/cm H_2O vs. 7.6 mL/cm H_2O, P < 0.001) differed between patients with normocompliance and those with hypocompliance (Table 2). Among the remaining patients, 27% had 71% agreement, 15% had 57% agreement, and 1 patient had 43%

agreement in the assessment of compliance among observers.

When a threshold of 10 mL/cm H_2O was used to categorize patients as hypocompliant, 49 patients met this criterion. Among these 49 "true hypocompliant patients," agreement of more than 80% for a hypocompliance rating was detected in only 25 patients (51%); in the remaining 24 patients, the agreement rate was less than 80%. In comparisons between these groups (>80% agreement vs. <80% agreement for hypocompliance), no statistically significant differences were observed in age (P=0.366), sex (P=0.990), cystometric capacity/EBC ratio (P=0.478), maximum detrusor pressure (P=0.711), end-filling pressure (P=0.211), or compliance (P=0.496).

DISCUSSION

Although UDS interpretation is often acknowledged to vary substantially across centers and observers because of differences in available resources and measurement criteria, this issue has not been systematically examined in the current literature [4, 5, 16]. This study therefore provides real-world evidence supporting a principle that has been repeatedly cited but incompletely studied.

The first study to focus on observer-related differences in UDS interpretation was published by Venhola et al. [9] in 2003. That analysis included 15 children without neurogenic LUTD, and 4 pediatric urologists independently rated bladder function, including detrusor function, compliance, and bladder capacity, and urethral or sphincteric function, including urethral pressure profile, sphincteric function, and uroflow analysis. The authors reported kappa values for each pairwise combination among the 4 observers, ranging from 0.37 to 1.00 for detrusor function and from 0.06 to 0.60 for bladder compliance. Despite limitations related to the relatively small number of patients and raters, that study was the first to suggest that interpretation of UDS may differ among observers and therefore represented an important contribution to the field. In our cohort, the Fleiss' kappa value for bladder compliance among all raters was 0.46, which may better reflect agreement in a multirater setting.

In 2006, Zimmern et al. [6] conducted a multicenter study of filling cystometrogram pressure measurements in an adult series, using 4 raters and 36 UDS randomly selected from 9 continence treatment centers. They reported excellent intraclass correlation coefficients of 0.90 to 0.996 for vesical, abdominal, and detrusor pressures at baseline and at maximum cystometric capacity. They concluded that, despite previously reported variation in qualitative measurements by Venhola et al. [9], quantitative analyses showed high concordance among well-trained reviewers. In our study, quantitative data were not measured or compared because only prelabeled standard UDS tracings that complied with good urodynamic practice standards were used.

In another study of DOA and pelvic floor activity in UDS, Bael et al. [18] used 247 investigations from 202 children extracted from the European Bladder Dysfunction Study [22]. Their panel, composed of 2 investigators from participating centers and one independent urologist, re-evaluated UDS obtained from the participating centers. The reported concordance rate for DOA was 37%. DOA was scored as "throughout filling," "before voiding only," or "none or slight" and was defined as an increase in detrusor pressure of at least 15 cm H_2O. This cutoff value may explain the higher concordance rate (74%) for "none or slight DOA." That study was limited by the absence of an appropriate methodological analysis and by the relatively small number of observers. Reliability was assessed only by the concordance rate, without presentation of a kappa value or correlation coefficient. Another limitation compared with our study is that the authors did not compare concordance or agreement rates for bladder compliance, which is considered a key cystometric component in the evaluation of neurogenic LUTD.

In 2017, Dudley et al. [8] developed an anonymous electronic survey with 20 clinical scenarios from children with neuropathic bladder. Twenty videourodynamic studies from children were rated by 14 pediatric urologists from 7 institutions. The authors used a 10-item form for each patient to assess bladder capacity, presence of DOA, bladder compliance, end-filling pressure, electromyographic activity or dyssynergia, treatment choice, bladder morphology, and bladder neck morphology. Using Fleiss' kappa to assess multirater agreement, they found a kappa value of 0.349 for the presence or absence of DOA and 0.4681 for bladder compliance. Compared with our findings, their results for bladder compliance were similar (0.4681 vs. 0.46), whereas agreement for DOA was substantially lower than in our series (0.349 vs. 0.60). This discrepancy may reflect broader implementation of modern good urodynamic practice standards and the greater expertise that clinicians have progressively gained in UDS methodology. The larger number of centers (7 vs. 1) and smaller number of patients (20 vs. 100) may also have affected overall agreement compared with our study. In addition, preselection of studies in our cohort based on 80% expert consensus may have introduced selection bias by enrich-

ing the dataset with more interpretable tracings, thereby increasing kappa scores.

For intraobserver agreement, Dejkriengkraikul et al. [23] examined UDS from 50 adults with spinal cord injury in 2020, using the International Spinal Cord Injury Urodynamic Basic Data Set Version 1.0. Intrarater reliability for 2 rehabilitation medicine physicians was 0.72 and 0.85 for detrusor function and 0.83 and 0.51 for bladder compliance, respectively. Interrater agreement between the 2 observers was 0.78 for detrusor function and 0.56 for bladder compliance. Detrusor function was rated as "normal," "neurogenic DOA," "underactive detrusor," "acontractile detrusor," or "unknown." For bladder compliance, a cutoff value of <10 mL/cm H_2O was used to define hypocompliance. The authors emphasized the importance of using a standardized reporting checklist to improve both intra- and interrater agreement. That study was limited by the inclusion of only 2 raters and by the predominance of patients with suprasacral spinal cord injury (86%). In addition, because the raters were specialists in rehabilitation medicine rather than urologists, the findings may not be directly comparable with urological interpretation in the present study.

Few studies have examined adult female cohorts undergoing UDS for lower urinary tract symptoms, such as urinary incontinence or voiding difficulty. The first such study was conducted by Whiteside et al. [10] in 2006, who examined UDS from 100 females, with a particular focus on detrusor function during the filling and voiding phases. The kappa scores for 3 female urologists and 3 urogynecologists were 0.59 and 0.36, respectively, for filling and voiding detrusor function. In another female cohort of 28 patients with voiding difficulty, detection of bladder outlet obstruction was assessed by 5 observers using 6 well-known nomograms [24]. Kappa scores among observers ranged from 0.548 with the Solomon–Greenwell nomogram to 0.788 with the Chassagne nomogram.

Variation in interpretation may also arise from artifacts within UDS tracings. Canmemiş and Ozel examined 90 pediatric UDS grouped according to the number of artifacts per study [7] and found that both inter- and intraobserver agreement between 2 raters decreased as the number of artifacts increased. They reported a kappa score of 0.321 for compliance and no agreement for detrusor function or type of overactivity in the no-artifact group. Compared with our findings, their interpretations of detrusor function and type of overactivity appear more variable, even in artifact-free investigations. In our study, tracings with artifacts were excluded by consensus, which may have reduced variability in the interpretation of inappropriate cystometry findings, overestimated the overall agreement rates, and limited the generalizability of the results. In contrast, intraobserver agreement appeared to remain within acceptable limits, with a kappa score >0.7, even for studies with a high number of artifacts. The authors stated that follow-up UDS interpreted by the same physician may reduce bias in interpretation.

As another potential source of variability in UDS interpretation and urodynamically guided management, Smith et al. [25] compared *post hoc* interpretation with real-time interpretation in 2009. In their analysis, comparison of *post hoc* and live interpretation by 2 independent observers demonstrated no superiority of real-time interpretation. Nevertheless, the authors concluded that unrecorded and heuristic observations made at the time of the study may influence the diagnostic interpretation of urodynamic evaluation and subsequent management strategy.

To reduce variation in UDS interpretation, artificial intelligence-based models have previously been used to support standardized and accurate interpretation of both noninvasive [26] and invasive [13] UDS. Although the results are encouraging, these models still lack external validation, limiting their routine clinical use. At present, the most robust approach to consistent and reliable urodynamic evaluation appears to be strict adherence to rigorous standardization efforts and maintenance of high-quality urodynamic practice at reference centers.

This study has several limitations. First, all urodynamic investigations were performed in a single experienced center with a volume of 80 to 100 cases per month, and the investigated UDS were obtained using a homogeneous standardized protocol; therefore, the findings may not fully reflect real-world practice. Second, the retrospective design inherently limited control over confounding variables, such as repeat filling protocols and artifact management.

In conclusion, this study showed weak to moderate agreement in the evaluation of bladder compliance and DOA on cystometric studies, even among experienced, well-trained observers at a reference center for spina bifida. Notably, the assessment of compliance showed that no cystometric parameter reliably predicted whether observers would achieve low or high agreement. Given the high risk of upper urinary tract damage in this patient group, more detailed and cautious interpretation is warranted, as are protocols or registries designed to improve interobserver consistency and standardize interpretation. In this context, multicenter prospective studies involving observ-

ers with varying levels of experience and incorporating consensus methodologies, such as the Delphi process, are needed to improve the reliability and standardization of urodynamic evaluation.

AUTHOR CONTRIBUTION STATEMENT

· Conceptualization: *YEG, CAS, TT*
· Data curation: *YEG, FA, OCO, OA, RE, AOG, SY, KC, TT, CAS*
· Formal analysis: *YEG, CAS*
· Methodology: *YEG, AOG, CAS, TT*
· Project administration: *YEG, OCO, SY, KC, TT, CAS*
· Visualization: *YEG, FA, OCO, OA, RE*
· Writing - original draft: *YEG, CAS*
· Writing - review & editing: *SY, KC, TT, CAS*

ORCID

Yunus Emre Genc	0000-0001-8335-1292
Faruk Arslan	0000-0003-1921-0406
Onur Can Ozkan	0000-0002-4344-5195
Osman Akyuz	0000-0002-1402-7664
Raziye Ergun	0000-0001-5943-0505
Ahmet Ozgur Guctas	0000-0001-8429-0778
Selcuk Yucel	0000-0001-9837-5510
Kamil Cam	0000-0002-8275-5479
Tufan Tarcan	0000-0002-3387-3524
Cagri Akin Sekerci	0000-0002-0334-2466

REFERENCES

1. Sič A, Stojanović B, Đorđević M. Neurogenic bladder in children with myelomeningocele. Diseases 2025;13:117.
2. Stein R, Bogaert G, Dogan HS, Hoen L, Kocvara R, Nijman RJM, et al. EAU/ESPU guidelines on the management of neurogenic bladder in children and adolescent part I diagnostics and conservative treatment. Neurourol Urodyn 2020;39:45-57.
3. Abrahamson K, Bogaert G, Dellenmark-Blom M, Feitz W, Goyal A, Jouannic JM, et al. EAU – ESPU - ERN eUROGEN - ERN ITHACA – ERN ERKNet - IFSBH Guidelines on spinal dysraphism in children and adolescents. Arnhem (Netherlands): European Association of Urology; 2025.
4. Bauer SB, Nijman RJ, Drzewiecki BA, Sillen U, Hoebeke P; International Children's Continence Society Standardization Subcommittee. International Children's Continence Society standardization report on urodynamic studies of the lower urinary tract in children. Neurourol Urodyn 2015;34:640-7.
5. Rosier PFWM, 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.
6. Zimmern P, Nager CW, Albo M, Fitzgerald MP, McDermott S; Urinary Incontinence Treatment Network. Interrater reliability of filling cystometrogram interpretation in a multicenter study. J Urol 2006;175:2174-7.
7. Canmemis A, Ozel SK. The impact of artifact levels on intra- and interobserver agreement in pediatric urodynamic traces. J Pediatr Urol 2025;21:362-9.
8. Dudley AG, Adams MC, Brock JW, Clayton DB, Joseph DB, Koh CJ, et al. Interrater reliability in interpretation of neuropathic pediatric urodynamic tracings: an expanded multicenter study. J Urol 2018;199:1337-43.
9. 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.
10. 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.
11. Pourghazi F, Linder BJ, Alizad A, Fatemi M. Artifacts in urodynamic studies: a narrative review. Neurourol Urodyn 2025;44:1583-92.
12. Sekerci CA, Tanidir Y, Tarcan T, Yucel S. Consistency of uroflowmetry analysis in children among observers. Neurourol Urodyn 2023;42:662-8.
13. Hobbs KT, Choe N, Aksenov LI, Reyes L, Aquino W, Routh JC, et al. Machine learning for urodynamic detection of detrusor overactivity. Urology 2022;159:247-54.
14. Cohn JA. Urodynamics in the evaluation and management of bothersome lower urinary tract symptoms in men with a prostate. Neurourol Urodyn 2026;45:262-9.
15. Tanaka ST, Yerkes EB, Routh JC, Tu DD, Austin JC, Wiener JS, et al. Urodynamic characteristics of neurogenic bladder in newborns with myelomeningocele and refinement of the definition of bladder hostility: findings from the UMPIRE multi-center study. J Pediatr Urol 2021;17:726-32.
16. Nguyen LN, Nadeau G. Detrusor overactivity and urodynamics. Neurourol Urodyn 2026;45:279-85.
17. Tennstedt S; Urinary Incontinence Treatment Network. Design of

the Stress Incontinence Surgical Treatment Efficacy Trial (SISTEr). Urology 2005;66:1213-7.
18. Bael A, Verhulst J, Lax H, Hirche H, van Gool JD; European Bladder Dysfunction Study EU BMH1-CT94-1006. Investigator bias in urodynamic studies for functional urinary incontinence. J Urol 2009;182:1949-52.
19. Abrams P. Describing bladder storage function: overactive bladder syndrome and detrusor overactivity. Urology 2003;62(5 Suppl 2):28-37; discussion 40-2.
20. Donner A, Eliasziw M. Sample size requirements for reliability studies. Stat Med 1987;6:441-8.
21. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22:276-82.
22. Bael A, Lax H, de Jong TP, Hoebeke P, Nijman RJ, Sixt R, et al. The relevance of urodynamic studies for Urge syndrome and dysfunctional voiding: a multicenter controlled trial in children. J Urol 2008;180:1486-93; discussion 1494-5.
23. Dejkriengkraikul K, Pattanakuhar S, Hama T, Tongprasert S. Reliability of the international spinal cord injury urodynamic basic data set version 1.0. Spinal Cord 2021;59:613-7.
24. Vidal-Brandt AS, Castro-Núñez P, Noyola-Ávila I, Rodríguez-Muñoz U, Maldonado-Alcaraz E, Moreno-Palacios J. [Interobserver agreement in the diagnosis of bladder outlet obstruction in women]. Rev Med Inst Mex Seguro Soc 2023;61(Suppl 3):S422-8. Spanish.
25. Smith PP, Hurtado EA, Appell RA. Post hoc interpretation of urodynamic evaluation is qualitatively different than interpretation at the time of urodynamic study. Neurourol Urodyn 2009;28:998-1002.
26. Arslan F, Algorabi O, Ozkan OC, Turkan YS, Ersin Namli, Genc YE, et al. Artificial intelligence-based analysis of uroflowmetry patterns in children: a machine learning perspective. Neurourol Urodyn 2025;44:1575-82.