Interrater Reliability in Pediatric Urodynamic Tracings: A Pilot Study

Anne G. Dudley,* Daniel P. Casella, Chelsea J. Lauderdale, Shilin Zhao, Heidi Chen, Stacy T. Tanaka, John C. Thomas, John C. Pope, IV, Mark C. Adams, John W. Brock, III, Chevis N. Shannon and Douglass B. Clayton

From the Division of Pediatric Urology, Monroe Carell Jr. Children's Hospital at Vanderbilt (AGD, DPC, CJL, STT, JCT, JCP, MCA, JWB, DBC), Department of Neurological Surgery, Vanderbilt University Medical Center (CNS) and Center for Quantitative Science, Vanderbilt University School of Medicine (SZ, HC), Nashville, Tennessee

Purpose: Urodynamic studies are crucial to neuropathic bladder management and they often determine surgical intervention. However, current evidence indicates that interpretations show poor agreement across physicians. We sought to determine the interrater reliability of urodynamic interpretation in our practice. We hypothesized that there would be strong correlation among pediatric urologists of similar training in a single academic practice.

Materials and Methods: We retrospectively identified patients with neuropathic bladder who underwent urodynamics at our institution between 2014 and 2015. An anonymous electronic survey (phase I) was developed with 20 clinical scenarios, each containing a brief history, a single urodynamic tracing and an accompanying fluoroscopic image. Faculty members assessed each tracing by an online instrument developed using urodynamic reports and published literature. The primary outcome was statistical correlation across raters as measured by the Spearman correlation coefficient. In a followup study (phase II) we investigated the sources of variability in urodynamic interpretations.

Results: Six faculty members completed the study with a response rate of 100%. In comparing urodynamic interpretation across raters, the faculty demonstrated a weak to strong correlation (r_s 0.39–0.61, $p < 0.001$). A strong correlation was found for fluoroscopic and clinical decision making variables, while electromyography synergy and detrusor overactivity demonstrated weaker correlation across physicians.

Conclusions: Faculty interpretations of urodynamic tracings showed only moderate agreement despite a close working relationship and similar training at a single institution. Variability in interpretation can strongly impact patient treatment. Therefore, further work is needed to standardize the reporting and interpretation of urodynamic studies to optimize patient care.

Key Words: urinary bladder, neurogenic; urodynamics; reproducibility of results; urologists; observer variation

Abbreviations and Acronyms

EMG = electromyography
ICCS = International Children's Continence Society
MOMS = Management of Myelomeningocele Study
NSBPR = National Spina Bifida Patient Registry
VUDS = videourodynamics

Accepted for publication October 4, 2016.

No direct or indirect commercial incentive associated with publishing this article.

The corresponding author certifies that, when applicable, a statement(s) has been included in the manuscript documenting institutional review board, ethics committee or ethical review board study approval; principles of Helsinki Declaration were followed in lieu of formal ethics committee approval; institutional animal care and use committee approval; all human subjects provided written informed consent with guarantees of confidentiality; IRB approved protocol number; animal approved project number.

Supported by National Center for Advancing Translational Sciences/National Institutes of Health Grant UL1 TR000445 (REDcap).

* Correspondence: Department of Urologic Surgery, Vanderbilt University Medical Center, AA-1302 Medical Center North, 1161 21st Ave. South, Nashville, Tennessee 37232 (telephone: 623-6150; e-mail: annedudleymd@gmail.com).

THE management of neuropathic bladder relies heavily on the results of urodynamic testing. UDS are used to report outcomes in multi-institutional and multidisciplinary trials in the published literature. Recent guidelines published by ICCS guide the interpretation of pediatric urodynamic tracings, although the final report remains subjective.[1] Previous studies

0022-5347/17/1974-0001/0
THE JOURNAL OF UROLOGY®

http://dx.doi.org/10.1016/j.juro.2016.10.053
Vol. 197, 1-6, April 2017
Printed in U.S.A.

show that physicians who interpret pediatric urodynamic tracings demonstrate poor interrater reliability, which is a statistical measure of rater agreement.[2] Additionally, our experience as a participating center for the NIH (National Institutes of Health) MOMS highlighted the challenges of interpreting pediatric urodynamic studies across physicians.

We sought to understand the variability across our group for interpreting VUDS in patients with neuropathic bladder. We believed that our faculty composition was homogenous in nature, given our close working relationship and similar practice patterns. We hypothesized that there would be a strong correlation among pediatric urologists of similar training in interpreting urodynamic tracings in a single academic practice.

METHODS

Following approval by the Vanderbilt University institutional review board (IRB 151604), one of us (AGD) retrospectively identified pediatric patients with neuropathic bladder who were treated at our institution between 2014 and 2015. From this group 20 unique VUDS were randomly selected for inclusion. Study exclusion criteria included lack of a urodynamic tracing image, fluoroscopic image or neuropathic diagnosis. Of the patients 80% had a diagnosis of myelomeningocele, 10% had spinal cord injury and 10% had a tethered spinal cord. Average patient age was 12.6 years and 60% of the patients were female. Of the patients 60% were on anticholinergic therapy and 75% were on intermittent catheterization.

All studies were done for surveillance of neuropathic bladder or for new urological symptoms. The faculty members interpreting these studies were blinded to patient identity. All 6 clinical faculty members were enrolled and completed all portions of the study. Advanced practice providers and trainees were excluded. All faculty members regularly interpret VUDS and provide care in at least 1 setting, including the Multi-Disciplinary Spina Bifida Clinic, NSBPR, MOMs Trial and/or Complex Reconstruction Clinic.

Using the web based data collection system REDCap (Research Electronic Data Capture) we developed an anonymous online survey to assess VUDS interpretation.[3] The emailed survey contained 20 unique clinical scenarios, each with 1 urodynamic tracing using a device from Precision Urodynamics, Central Point, Oregon, or Laborie, Mississauga, Ontario, Canada and 1 fluoroscopic image (parts *A* and *B* of figure). At the time of VUDS 1 of 2 experienced urodynamic nurses performed the study and the treating physician was present for the majority of cases. The tracings included measurements of vesical pressure, rectal pressure via a rectal catheter, and calculated detrusor pressure and activity. Fluoroscopic images were taken during the filling portion of VUDS. [F1]

Bladder capacity was calculated using 2 formulas, including bladder capacity = (age + 2) × 30 cc and bladder capacity = weight in kg × 7 cc. For each scenario the urologist completed an instrument to assess VUDS (part *C* of figure). The instrument contained variables drawn from urodynamic reports, published literature, NSBPR variables and MOMS.[1,4,5] Because interpretation of urodynamic tracings across faculty members was our singular goal, patient specific clinical outcomes were not reported.

The instrument contained dichotomous questions designed to assess the interpretation of specific aspects of the urodynamic tracing in addition to the overall clinical assessment. We developed and internally tested the instrument in a group research setting but it was not externally validated. We intentionally did not provide strict definitions or training modules as the study aim was to assess the real practice variability of urodynamic interpretation.

We performed the statistical analysis using nonparametric statistics with SAS®, version 9.4 and R 3.1.1 (https://cran.r-project.org/bin/windows/base/old/3.1.1/). As our primary outcome we measured the correlation of urodynamic interpretations among faculty members as measured by the Spearman correlation coefficient (r_s), where a value of 1.0 corresponds to perfect correlation and a value of 0 represents no correlation. Statistical significance was set a priori at $p < 0.05$.

Our study consisted of 2 phases. Phase I assessed the correlation across raters for the clinical scenarios and individual instrument variables. We implemented phase II of the study to further explore the instrument variables with an unexpectedly weak correlation. We used construct validity, a technique used in survey research to analyze the content of the questions to ensure that all interpretation concepts were adequately tested.[6]

After identifying 3 clinical scenarios from phase I with weaker correlations, we developed a second anonymous online survey. The survey asked open-ended questions to elucidate the 3 most important urodynamic factors used for clinical decision making. To determine why detrusor pressure demonstrated only moderate correlation, we assessed working definitions of end fill detrusor pressure and detrusor leak point pressure, and compared these definitions to published definitions from the ICCS and the NSBPR terminology.[1,5]

RESULTS

All 6 invited faculty members completed both phrases of the study for a response rate of 100%. Faculty members had a median of 16.5 years (range 4 to 32) experience. Three faculty members (50%) had completed a pediatric fellowship at the institution and only 1 had practiced at another hospital previously. All others had spent the entire career at 1 institution.

When examining complete survey responses in phase I, we saw a moderate correlation across faculty responses (mean ± SD r_s 0.51 ± 0.07, range 0.38−0.8). When examining correlation by variable, the fluoroscopic variables of bladder shape and bladder neck status demonstrated strong correlation (r_s 0.72 and 0.69, respectively, see table). We [T1]