



Accuracy of subjective vesicoureteral reflux timing assessment: supporting new voiding cystourethrogram guidelines

Dabin Ji¹ · Derrick E. Ridley² · J. Damien Grattan-Smith³ · Joseph P. Williams³ · Angela M. Arlen⁴ · Andrew J. Kirsch⁵ · Courtney M. McCracken⁶ · Christopher S. Cooper⁷ · Wesley W. Durrence¹ · Michelle A. Lightfoot⁵ · Michael L. Garcia-Roig⁵

Received: 1 September 2019 / Revised: 24 December 2019 / Accepted: 11 February 2020 / Published online: 17 March 2020

© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Background Bladder volume at the onset of vesicoureteral reflux (VUR) is an important prognostic indicator of spontaneous resolution and the risk of pyelonephritis.

Objective We aim to determine whether pediatric urologists and pediatric radiologists can accurately estimate the timing of reflux by examining voiding cystourethrogram (VCUG) images without prior knowledge of the instilled contrast volume.

Materials and methods Total bladder volume and the volume at the time of reflux were collected from VCUG reports to determine the volume at the onset of VUR. Thirty-nine patients were sorted into three groups: early-/mid-filling reflux, late-filling and voiding only. Thirty-nine images were shown to three pediatric urologists and two pediatric radiologists in a blinded fashion and they were then asked to estimate VUR timing based on the above categories. A weighted kappa statistic was calculated to assess rater agreement with the gold standard volume-based report of VUR timing.

Results The mean patient age at VCUG was 3.1±2.9 months, the median VUR was grade 3, and 20 patients were female. Overall agreement among all five raters was moderate ($k=0.43$, 95% confidence interval [CI] 0.36–0.50). Individual agreement between rater and gold standard was slight to moderate with kappa values ranging from 0.13 to 0.43.

Conclusion Pediatric radiologists and urologists are unable to accurately and reliably characterize VUR timing on fluoroscopic VCUG. These findings support the recently published American Academy of Pediatrics protocol recommending the routine recording of bladder volume at the onset of VUR as a standard component of all VCUGs to assist in a more accurate assessment of the likelihood of resolution and risk of recurrent urinary tract infections.

Keywords Bladder · Children · Urogenital tract · Vesicoureteral reflux · Voiding cystourethrogram · Volume

✉ Michael L. Garcia-Roig
mgarciaroig@emory.edu

¹ Mercer University School of Medicine, Savannah, GA, USA

² Emory University School of Medicine, Atlanta, GA, USA

³ Division of Pediatric Radiology, Department of Radiology, Emory University School of Medicine, Children's Healthcare of Atlanta, Atlanta, GA, USA

⁴ Division of Pediatric Urology, Department of Urology, Yale School of Medicine, New Haven, CT, USA

⁵ Division of Pediatric Urology, Department of Urology, Emory University School of Medicine, Children's Healthcare of Atlanta, 5730 Glenridge Dr., Suite 200, Atlanta, GA 30328, USA

⁶ Department of Pediatrics, Emory University School of Medicine, Children's Healthcare of Atlanta, Atlanta, GA, USA

⁷ Division of Pediatric Urology, Department of Urology, University of Iowa Hospitals & Clinics, Iowa City, IA, USA

Introduction

With an estimated prevalence of 0.4–1.8% in the general pediatric population and 30% among those with a history of febrile urinary tract infection (UTI), vesicoureteral reflux (VUR) is a common urological diagnosis affecting children [1, 2]. VUR management is individualized based on a given patient's likelihood of spontaneous resolution and UTI, underlying renal pathology and parental preferences. In many cases, VUR spontaneously improves or resolves over time; however, observation may not be an optimal option for those at risk for breakthrough febrile UTIs or renal scarring.

The initial management for the majority of children with VUR is nonsurgical with continuous antibiotic prophylaxis or observation until spontaneous resolution. The decision to undergo surgical intervention in a child with VUR is impacted by assessing multiple factors that influence spontaneous VUR

resolution, including the initial grade of reflux, bladder volume at the onset of VUR, age, gender, anatomical abnormalities, and bladder and bowel dysfunction [3, 4]. Of the aforementioned variables, bladder volume at the time of VUR is one of the most important prognostic indicators of spontaneous resolution as VUR occurring at earlier filling volumes is associated with a lower likelihood of spontaneous resolution as well as an increased likelihood of breakthrough febrile UTI [5, 6].

Voiding cystourethrogram (VCUG) is the gold standard for diagnosing and evaluating the grade and severity of VUR. However, a recent survey revealed significant variability in VCUG technique and protocol among 65 children's hospitals in the United States and Canada [7]. This inconsistency raised concerns about patient safety as VCUG is an invasive test and its interpretation influences the management of VUR. In 2016, the American Academy of Pediatrics (AAP) developed a recommended protocol for VCUG with the purpose of standardizing VCUG technique, including reporting the volume of contrast infused and the bladder volume at which the onset of reflux occurred [8]. Concurrently, the 2019 American College of Radiology (ACR) VCUG practice parameter states that bladder volume at the onset of reflux is only recommended, not required [9]. Despite the AAP recommended protocol, there is still variation in practice. If bladder volumes are not recorded during the study, the urologist often estimates VUR timing by reviewing the fluoroscopy images. Our study aimed to determine whether pediatric radiologists and pediatric urologists can accurately estimate the timing of reflux by examining VCUG images without prior knowledge of the volume of contrast instilled.

Materials and methods

Institutional review board approval was obtained for the study. Patients diagnosed with vesicoureteral reflux were retrospectively identified at an institution where the volume of contrast instilled is routinely recorded. Studies were selected from those performed between February 2006 and May 2013. Patients were included if they were younger than 18 years old at the time of the study, underwent a VCUG with recorded bladder volumes (both total and at VUR onset), and were diagnosed with primary VUR. Patients were excluded if they had a diagnosis of secondary VUR (neurogenic bladder, posterior or anterior urethral valves, ureterocele, urethral stricture, bladder bowel dysfunction). A total of 39 patients were selected to satisfy calculated power requirements.

Typical protocol for a VCUG involves a scout image, followed by images during estimated early, mid and late bladder filling. Once reflux is identified, the ureter and renal collecting system are imaged. A voiding image is captured, as well as post void image. Fluoroscopic screen

capture is used for imaging. Bladder volumes were measured using the amount of contrast material instilled, as determined by volume markers on the contrast bottle label. Maximum bladder capacity was considered as the instilled contrast volume at which spontaneous voiding occurred. Cyclical VCUG studies were performed for all selected patients. Total bladder volume and volume at the time of reflux, age at VUR diagnosis, gender and VUR index were collected from the VCUG reports. The patients were sorted into three groups of volume-based VUR timing: early-/mid-filling reflux (VUR onset <75% bladder filling), late-filling reflux (VUR onset at 75–100% bladder filling), and voiding-only reflux (VUR onset only during micturition).

Three fellowship-trained pediatric urologists and two pediatric radiologists were shown all available fluoroscopic images from each of the 39 included patients. The pediatric urologist reviewers included a fellow (M.L.G.-R.) and 2 practicing pediatric urologists with 1 (A.J.K.) and 18 (M.A.L.) years of post-fellowship experience at the time of study. The pediatric radiologist reviewers had 17 (J.D.G.-S.) and 25 (J.P.W.) years of post-fellowship experience at the time of study. All available images from each study were compiled in slide show format and presented to the reviewers in sequential order. An example of VCUG images shown to each reviewer is provided in Fig. 1. Reviewers had not participated in the care of these patients and were blinded to the associated radiology report, recorded bladder volumes and each other's answers. To account for bilateral VUR and cyclic studies, the onset of initial VUR in the image series presented was tested whether unilateral or bilateral, or on first or second cycle of VCUG. Reviewer answers were compared against the timing of onset of VUR as reported by the radiologist (% filling=volume at VUR/total bladder filling at VCUG).

A weighted kappa statistic with associated 95% confidence interval (CI) was calculated and used to assess rater agreement with the gold standard volume-based interpretation of VUR timing among the three pediatric urologists and two pediatric radiologists. The following ranges proposed by Landis and Koch [10] were used to interpret the degree of agreement between raters based on the value of the weighted kappa: (<0, poor), (0–<0.2, slight), (0.2–<0.4, fair), (0.4–<0.6, moderate), (0.6–<0.8, substantial), (0.8–1.0, almost perfect). Exact agreement was also tabulated. Analysis was conducted using SAS v. 9.4 (SAS Institute Inc., Cary, NC), and statistical significance was assessed at the 0.05 level.

Results

Patient demographics

A total of 39 patients were randomly selected from a historical cohort of children with VUR, with 13 patients in each VUR

Fig. 1 Example of a voiding cystourethrogram (VCUG) report shown to reviewers. Reviewers were asked to estimate volume at the onset of vesicoureteral reflux (VUR) and sort patients into three groups: early-/mid-filling reflux (VUR onset <75% bladder filling), late filling (75–100%) and voiding only (during micturition). Sections circled in red represent VUR during VCUG. Red circles indicate onset of VUR during the study



timing category (early-/mid-filling reflux, late-filling reflux, and voiding-only reflux) meeting inclusion criteria. Detailed demographics for the patients meeting inclusion criteria are outlined in Table 1.

Rater agreement

Answers provided by the three pediatric urologists and two radiologists were compared to the recorded volume-based onset of VUR. Details of rater agreement among pediatric urologists and pediatric radiologists are outlined in Table 2. Overall agreement among all five raters was moderate ($k=0.43$, 95% CI 0.36–0.50). Individual agreement between rater and the volume-based result was slight to moderate with kappa values ranging from 0.13 to 0.43. Interpretation among

pediatric urologists (kappa scores ranging from 0.13 to 0.36) was less accurate than interpretation among pediatric radiologists (kappa scores ranging from 0.32 to 0.43) compared to the gold standard. However, both were fair at best. Interobserver variability was similar among pediatric radiologists and urologists ($k=0.53$ vs. 0.50). Pediatric radiologists and urologists did not consistently identify any of the VUR timing groups. The percentage of exact agreement among pediatric urologists ranged from 52.6% to 72.5% while the percentage of exact agreement among pediatric radiologists was 70%; the percentage of exact agreement with pediatric urologists and volume-based VUR timing ranged from 37.5% to 55% while the percentage of exact agreement with pediatric radiologists and volume-based VUR timing ranged from 47.5% to 57.5% (Table 2). As demonstrated in Table 3, the pediatric urologists’ accuracy, represented as percent agreement, was generally consistent across the volume-based VUR timing groups (between 41% and 53% agreement on average). However, the pediatric radiologists were more accurate when interpreting studies for the late-filling and voiding-only reflux groups compared to the early-/mid-filling reflux groups (56% and 69% agreement vs. 31%, respectively).

Table 1 Demographic information for patients whose VCUG images were included in the survey (A), and number of studies included with early/mid, late or voiding VUR timing as defined by bladder volume during the study (B)

A. Patient demographics	
Female – <i>n</i> (%)	20 (51%)
Mean±SD age at diagnosis	3.1±2.9 months
Median VUR grade	3
B. Number VUR timing – <i>n</i> (%)	
Early/mid filling	13 (33.3%)
Late filling	13 (33.3%)
Voiding	13 (33.3%)

SD standard deviation, VCUG voiding cystourethrogram, VUR vesicoureteral reflux

Discussion

Our study reveals that pediatric radiologists and urologists are unable to retrospectively estimate VUR timing solely based on observing static fluoroscopic VCUG images. Accurate interpretation of VUR timing requires recording the volume of

Table 2 Agreement among pediatric urologists and pediatric radiologists

Individual agreement	Overall agreement among reviewers		
	Weighted kappa	95% CI	% exact agreement
Uro 1 vs. Uro 2	0.65	(0.47–0.84)	72.5%
Uro 1 vs. Uro 3	0.51	(0.30–0.72)	60.5%
Uro 2 vs. Uro 3	0.51	(0.29–0.73)	65.8%
Overall	0.50	(0.37–0.63)	52.6%
Agreement with pediatric urologist and volume-based VUR timing			
Uro 1	0.36	(0.12–0.60)	55%
Uro 2	0.13	(–0.10–0.36)	37.5%
Uro 3	0.27	(0.02–0.52)	50%
Agreement between pediatric radiologists			
Rad 1 vs. Rad 2	0.53	(0.30–0.75)	70%
Agreement with pediatric radiologist and volume-based VUR timing			
Rad 1	0.43	(0.14–0.57)	57.5%
Rad 2	0.32	(0.09–0.54)	47.5%
Overall agreement with all five raters			
	0.42 (0.36–0.50)		

Interpretation of the agreement between the independent examiners according to the weighted kappa: <0 as poor, 0–<0.2 as slight, 0.2–<0.4 as fair, 0.4–<0.6 as moderate, 0.6–<0.8 as substantial, 0.8–1.0 as almost perfect
CI confidence interval, *Rad* radiologist, *Uro* urologist, *VUR* vesicoureteral reflux

contrast instilled at the time of reflux and maximum bladder capacity by the radiologist or radiology technician during VCUG. Our findings support the recently published American Academy of Pediatrics (AAP) protocol recommending the routine recording of bladder volume at the onset of VUR as a standard component of all VCUGs to more accurately assess the likelihood of resolution and risk of recurrent UTI [8].

When discussing treatment options for VUR with families, it is important to have an accurate assessment of the patient's condition as options range from continuous antibiotic prophylaxis to surgical intervention depending on the outcome and interpretation of the VCUG study. VCUGs are used to guide management and treatment of an individual child with VUR; therefore, our study supports the recommendation that the broader community of pediatric health care providers adopt into practice the current VCUG

protocol established by the AAP. Standardizing an evidence-based protocol in medicine provides a key strategy for improving health care by reducing the variance in practice, minimizing patient risk, improving validity of the imaging results, and allowing outcomes to be accurately compared between individuals and institutions [11, 12].

This study is not without limitations. Our retrospective study of children with VUR identified by querying our electronic medical records carries inherent limitations in the ability to accurately identify all patients meeting the inclusion criteria. Additionally, our study is limited by a sample size of 39 patients. In addition, the number of raters was small (three urologists and two radiologists), which may not be an accurate representation of the variability among all subspecialists. However, the current study design is common when assessing the accuracy of a diagnostic test [13–15]. An additional limitation is that all raters were from the same institution. Further, it is unclear how these results apply to contrast-enhanced voiding urosonography as VUR timing was defined with standard fluoroscopic VCUG. This is a consideration for future study.

Conclusion

Our study supports the 2016 VCUG protocol established by the AAP. Pediatric radiologists and technicians should routinely record bladder volume at the onset of VUR and at maximum bladder capacity as a standard component of performing VCUG in order to provide a more accurate

Table 3 Percent agreement by volume-based VUR timing groups

Reviewer	Early-/mid-filling reflux	Late-filling reflux	Voiding-only reflux
Uro 1	54%	57%	54%
Uro 2	46%	43%	23%
Uro 3	46%	58%	46%
Urologist average	49%	53%	41%
Rad 1	23%	76%	69%
Rad 2	38%	36%	69%
Radiologist average	31%	56%	69%

Rad radiologist, *Uro* urologist, *VUR* vesicoureteral reflux

interpretation of the test outcome. Implementing the AAP's standard protocol will improve patient care by assisting clinicians, patients, and families in making accurate and informed decisions regarding VUR management.

Compliance with ethical standards

Conflicts of interest None

References

1. Sargent MA (2000) What is the normal prevalence of vesicoureteral reflux? *Pediatr Radiol* 30:587–593
2. Skoog SJ, Peters CA, Arant BS Jr et al (2010) Pediatric vesicoureteral reflux guidelines panel summary report: clinical practice guidelines for screening siblings of children with vesicoureteral reflux and neonates/infants with prenatal hydronephrosis. *J Urol* 184:1145–1151
3. Estrada CR Jr, Passerotti CC, Graham DA et al (2009) Nomograms for predicting annual resolution rate of primary vesicoureteral reflux: results from 2,462 children. *J Urol* 182:1535–1541
4. Knudson MJ, Austin JC, McMillan ZM et al (2007) Predictive factors of early spontaneous resolution in children with primary vesicoureteral reflux. *J Urol* 178:1684–1688
5. Alexander SE, Arlen AM, Storm DW et al (2015) Bladder volume at onset of vesicoureteral reflux is an independent risk factor for breakthrough febrile urinary tract infection. *J Urol* 193:1342–1346
6. McMillan ZM, Austin JC, Knudson MJ et al (2006) Bladder volume at onset of reflux on initial cystogram predicts spontaneous resolution. *J Urol* 176:1838–1841
7. Palmer BW, Ramji FG, Snyder CT et al (2011) Voiding cystourethrogram — are our protocols the same? *J Urol* 186:1668–1671
8. Frimberger D, Bauer SB, Cain MP et al (2016) Establishing a standard protocol for the voiding cystourethrography (VCUG). *J Pediatr Urol* 12:361
9. American College of Radiology-Society for Pediatric Radiology (2019) Practice Parameter for the Performance of Fluoroscopic and Sonographic Voiding Cystourethrography in Children. <https://www.acr.org/-/media/ACR/Files/Practice-Parameters/VoidingCysto.pdf>. Accessed 1 Sep 2019
10. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
11. Pronovost PJ, Berenholtz SM, Goeschel CA et al (2006) Creating high reliability in health care organizations. *Health Serv Res* 41:1599–1617
12. Rozich JD, Howard RJ, Justeson JM et al (2004) Standardization as a mechanism to improve safety in health care. *Jt Comm J Qual Saf* 30:5–14
13. Carpenter MA, Hoberman A, Mattoo TK et al (2013) The RIVUR trial: profile and baseline clinical associations of children with vesicoureteral reflux. *Pediatrics* 132:e34–e45
14. Daswani DD, Shah VP, Avner JR et al (2016) Accuracy of point-of-care lung ultrasonography for diagnosis of acute chest syndrome in pediatric patients with sickle cell disease and fever. *Acad Emerg Med* 23:932–940
15. Varshney T, Mok E, Shapiro AJ et al (2016) Point-of-care lung ultrasound in young children with respiratory tract infections and wheeze. *Emerg Med J* 33:603–610

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