

Clinical Study

Reliability of the Grading System for Voiding Cystourethrograms in the Management of Vesicoureteral Reflux: An Interrater Comparison

Süleyman Çelebi,¹ Seyithan Özaydın,² Cemile Beşik Baştaş,² Özgür Kuzdan,¹
Cankat Erdoğan,² Mehmet Yazıcı,¹ İsmail Caymaz,³ and Serdar Sander¹

¹Department of Pediatric Urology, Kanuni Sultan Suleyman Education and Research Hospital, 34300 Istanbul, Turkey

²Department of Pediatric Surgery, Kanuni Sultan Suleyman Education and Research Hospital, 34300 Istanbul, Turkey

³Department of Pediatric Radiology, Baskent University Faculty of Medicine, Istanbul, Turkey

Correspondence should be addressed to Süleyman Çelebi; celebisuleyman@hotmail.com

Received 23 August 2015; Revised 15 December 2015; Accepted 16 December 2015

Academic Editor: Darius J. Bagli

Copyright © 2016 Süleyman Çelebi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aim. Vesicoureteral reflux (VUR) is one of the most common conditions seen in pediatric urology. Fortunately, there are many treatment options for this disorder. The grading system for VUR varies among doctors, and the literature on its reliability is sparse. Here, we assessed the effectiveness of the current VUR grading system. *Methods.* A series of 40 voiding cystourethrogram (VCUG) studies were selected. Four pediatric urologists (PU) and four pediatric radiologists (PR) independently graded each VCUG and then agreed on a uniform interpretation. For statistical analysis the intraclass correlation coefficient (ICC) was applied to assess interrater agreement. *Results.* ICC values ranging from 0.82 to 0.88 reflected the strong reliability of VCUG for grading cases of VUR among pediatric urologists and radiologists as separate groups, and the reliability between the two groups was also good, as indicated by an ICC of 0.89. Despite the high ICC, disagreement existed between raters; the lowest agreement was associated with middle grades (III and IV). *Conclusions.* The interrater reliability of the international grading system for VUR was high but imperfect. Thus, grading differences at middle grades can profoundly influence the type of treatment pursued.

1. Introduction

Vesicoureteral reflux (VUR), defined as retrograde flow of urine from the bladder back up the ureter into the kidney, is diagnosed in 30% to 40% of children who present with urinary tract infections (UTIs). It is a congenital condition that may resolve or improve over time [1]. Because of the relatively high prevalence of renal scarring, it is prudent to understand how to identify VUR, potential problems associated with chronic VUR, and the most effective therapeutic strategies.

Currently, the standard test for diagnosing VUR is the voiding cystourethrogram (VCUG), which is also used to classify the severity of reflux. It is very important that physicians reach a consensus regarding the stages of VUR, because staging determines whether each child should simply be

closely observed, receive prophylactic antibiotics, or undergo endoscopic treatment or surgery. However, the reliability of interrater reflux grading has largely been ignored in the literature [2]. Thus, the goal of this study was to investigate the accuracy of the current staging system by assessing interrater reliability among radiologists and pediatric urologists.

2. Methods

After approval by the local institutional review board, we recruited children with primary VUR after their occurrence of UTI between 2012 and 2013. Patients who underwent VCUG and had bladder exstrophy or other abnormalities including ectopic ureterocele and ureteral duplication were excluded from the study. The VCUGs of these patients were shown to four pediatric urologists and four radiologists who

TABLE 1: Assessment of VUR grading among pediatric urologists.

Pediatric urologists (PU)	Intraclass correlation coefficient	95% confidence interval	
		Lower bound	Upper bound
Pediatric urologist 1	0.86	0.79	0.89
Pediatric urologist 2	0.85	0.80	0.88
Pediatric urologist 3	0.83	0.81	0.88
Pediatric urologist 4	0.87	0.83	0.88

are experienced in urology; these physicians were asked to grade VUR (from I to V) according to the international system of radiographic grading of vesicoureteric reflux [3]. The interrater reliability between the pediatric urologists and radiologists was given a grade of 0–5. Reviewers were blinded to the report and other readings. The responses from the pediatric urologists and radiologists were compared, and discrepancies were adjudicated. For statistical analysis, the intraclass correlation coefficient (ICC) was used to calculate interrater reliability. ICC > 0.72 indicated adequate reliability, whereas an ICC > 0.8 indicated almost perfect reliability.

3. Results

Unilateral VUR was diagnosed in 40 children (28 girls, 12 boys). The median age was 12 months, and most children (91%) were enrolled after their recurrent UTI. The UTI prior to enrollment was both febrile and symptomatic in 24 children, only febrile in 7, and only symptomatic in 9 children. All cases of VUR were diagnosed using VCUG. In total, 40 VCUGs were reviewed by 4 pediatric urologists and 4 radiologists, yielding a total of 320 observations. Among the pediatric urologists, ICC scores were consistently > 0.8, between 0.83 and 0.87, indicating good reliability (Table 1). Among the radiologists' evaluations, the ICC scores were consistent, ranging from 0.82 to 0.85 (Table 2), indicating the reliability of VCUG grading in this group of physicians. The ICC values demonstrated the strong reliability of VCUG for grading VUR cases among pediatric urologists and radiologists, and the interrater agreement was 0.88 (Table 3).

A closer inspection of scoring among pediatric urologists and radiologists revealed significant discrepancies between grades III and IV. For pediatric urologists, one rater graded 7 cases (17.5%) as III, while the other raters graded 16 cases (40%) as III. Similar ratings were given by the radiologists; 1 radiologist graded 6 cases (15%) as III, while the others graded 15 cases (37.5%) as III (Table 3). The same type of discrepancy also occurred for cases given a grade of IV. For pediatric urologists, 10 cases (25%) were graded as IV by one rater, while another rater graded 20 cases (50%) as IV. The radiologists used similar grades; 1 radiologist graded 9 cases (22.5%) as IV, while the other radiologist graded 17 cases (42.5%) as IV (Table 4).

Although ICC scores were consistently > 0.7 (0.73), there was also a discrepancy between cases which were graded III among the pediatric urologists and radiologists. The pediatric urologists rated 47 renal cases as grade III, while the radiologists scored 39 renal cases as grade III (Table 4).

TABLE 2: Assessment of VUR grading among pediatric radiologists.

Radiologists	Correlation coefficient	95% confidence interval	
		Lower bound	Upper bound
Radiologist 1	0.84	0.82	0.88
Radiologist 2	0.83	0.81	0.89
Radiologist 3	0.85	0.83	0.89
Radiologist 4	0.82	0.79	0.88

4. Discussion

VUR is present in 30% to 40% of children with a UTI and is associated with a higher risk of renal scarring. Reports from the 1960s and 1970s, when VUR was less frequently recognized, revealed that renal scarring due to VUR was the etiology of 50% of hypertension cases and 30% of end-stage renal disease (ESRD) cases in children [4]; therefore, the current standard of care includes imaging to assess the presence and extent of VUR [5]. Due to the risk of renal scarring, common clinical practice guidelines now state that VCUG is recommended after the first episode of febrile UTI for all children, depending on sex, age, and clinical presentation [6, 7]. The advantage of this method is the ability to grade reflux severity using the widely accepted 5-level International Scale [5]. The majority of children affected by this condition have low-grade VUR (grades I-II); the strategy depends on the hypothesis that reflux, especially VUR of grade III or greater, increases the risk of recurrent UTIs and renal scarring [8, 9]. Thus, grading must be accurate so that reflux of grade III or higher can reliably be distinguished from lower grade reflux to guide the decision [10, 11].

Craig et al. [12] reported near perfect agreement (kappa 90% to 91%) when three radiologists separately graded contrast VCUGs. However, Kronemer et al. [13] reported that there was divergent grade interpretation in 20 of 39 patients with VUR when 2 radiologists separately read the studies. Metcalfe et al. [2] also analyzed reflux grades and concluded that although the overall VUR grading of VCUGs was shown to be reliable, agreement was highest at the extremes of the scale (grades I and V); scoring discrepancies were more common at the middle grades (II–IV). One of the last issues from Greenfield et al. reported that there was divergent grade interpretation in 9 of 61 ureters initially assessed as middle grade (15%). Of these 9 discrepancies 7 (78%) were adjudicated to the higher grade. Greenfield et al. concluded that discrepancies in the assessment of

TABLE 3: Agreement between PR and PU on VCUG assessments.

	Intraclass correlation coefficient	95% confidence interval	
		Lower bound	Upper bound
Pediatric urologists	0.887	0.879	0.937
Radiologists	0.854	0.839	0.864

TABLE 4: Adjudication of discrepancies in VUR grade among pediatric urologists (PU) and pediatric radiologists (PR).

Grade	PU1	PU2	PU3	PU4	PR1	PR2	PR3	PR4
1	4 (10%)	3 (7.5%)	3 (7.5%)	4 (10%)	5 (12%)	3 (7.5%)	4 (10%)	3 (7.5%)
2	4 (10%)	5 (12.5%)	4 (10%)	3 (7.5%)	6 (15%)	5 (12.5%)	4 (10%)	3 (7.5%)
3	9 (22%)	15 (37.5%)	7 (17.5%)	16 (40%)	6 (15%)	15 (37.5%)	8 (20%)	10 (25%)
4	14 (35%)	10 (25%)	20 (50%)	10 (25%)	16 (40%)	9 (22%)	15 (37%)	17 (42%)
5	9 (22%)	7 (17%)	6 (15%)	7 (17%)	7 (17%)	8 (20%)	9 (22.5%)	7 (17.5%)

intermediate grade vesicoureteral reflux were noteworthy and added in particular that there was considerable disagreement in the evaluation of intermediate grades of reflux [14].

Our study found that, among groups of pediatric urologists and radiologists, and also in comparisons between the two groups, the ICC value was close to 0.9, indicating reliable grading. There was disagreement regarding grade in up to 20% of individual ureteral readings. We found that most discrepancies concerned VUR of grades III and IV, because both grades subjectively depend on the appearance of the renal calyx without numerical values being taken into account. This finding suggests that in clinical practice there may be reasonable doubt when categorizing children with reflux to grade III. This issue becomes especially important if one is not going to treat a child with grade III reflux with prophylaxis or perform follow-up imaging.

The grading of reflux using the International Scale remains the major means of categorizing patients and determining treatment [15]. Although currently there are different objective parameters such as bladder volume [16, 17] and ureteral diameter [18] to determine the resolution rates, still nearly common principle is using VUR grade for deciding the treatment. In some cases, a discrepancy exists between the degree of dilatation of the pyelocalyceal system and the distal ureter which in turn makes grading difficult. Resolution rate of grade 4 is less than grade 3 and so will require more surgery [19, 20]. Therefore, grade 3 and 4 distinction is important but as our results demonstrated, VUR grade often varies on the middle grades (3 and 4) depending on the observer.

Confusion in this staging system arises from the fact that the current 5-level grading system cannot easily be applied to VCUGs, which only include characteristics of four stages: stage 1, affecting the ureter but reflux does not reach the renal pelvis; stage 2, affecting the ureter with reflux reaching the renal pelvis; stage 3, affecting the renal calyceal system; and stage 4, gross dilatation and kinking of the ureter with papillary impressions no longer visible. More objective and quantitative data are required to divide VCUG findings into five stages.

While the current VUR grading system focuses primarily on the radiographic appearance of the upper tract, VUR

grading may be difficult when discrepancies exist between the degree of dilation of the pyelocalyceal system and the ureter. McMillan et al. [16] and Knudson et al. [17] demonstrated bladder volume at the onset of reflux as another independent factor affecting resolution rates, concluding that bladder volume is an objective value that could be utilized to provide prognostic information regarding the spontaneous resolution of vesicoureteral reflux (VUR) in patients. This knowledge could benefit unresolved cases where there were noteworthy discrepancies in the assessment of the grade of VUR. The present study by Cooper et al. [18] revealed the ureteral diameter and concluded that resolution may be more accurately predicted by the appearance of the distal ureter and concluded that ureteral diameter ratio (UDR) correlates with grade of reflux and proves more predictive of the ultimate clinical outcome in children with primary reflux than grade alone [18].

A lack of consensus on VUR grade staging may lead to a discrepancy in clinical decisions. The poor agreement on moderate grades may stem from differences in judging dilation of the calyceal system. The combined review of images by multiple reviewers, or improving the current grading system with including numerical value of calyceal dilatation, may help reduce grading discrepancies.

A discrepancy exists between the degree of dilatation of the pyelocalyceal system and renal parenchymal damage which in turn makes grading difficult [21, 22]. This may reflect situations where the degree of upper tract dilation and parenchymal damage upon which VUR grade is assigned is not well matched [23]. At VUR diagnosis using a combination of grade, age, volume at reflux onset, and presentation with a history of prenatal hydronephrosis and renal parenchymal damage improves predictive ability regarding early reflux resolution. Combining these individual factors may improve the decision-making process regarding reflux management.

In conclusion, VCUG has long been a mainstay of the diagnosis and grading of VUR, and our study confirmed the reliability of this method. However, discrepancies arise in grading abnormalities of the calyceal system seen on VCUGs, which could greatly impact the treatment method used. The combined review of images by multiple reviewers,

or improving the current grading system, may help reduce grading discrepancies.

Competing Interests

The authors declare that there is no conflict of interests.

References

- [1] S. P. Greenfield, M. Ng, and J. Wan, "Resolution rates of low grade vesicoureteral reflux stratified by patient age at presentation," *Journal of Urology*, vol. 157, no. 4, pp. 1410–1413, 1997.
- [2] C. B. Metcalfe, A. E. MacNeily, and K. Afshar, "Reliability assessment of international grading system for vesicoureteral reflux," *The Journal of Urology*, vol. 188, no. 4, pp. 1490–1492, 2012.
- [3] R. L. Lebowitz, H. Olbing, K. V. Parkkulainen, J. M. Smellie, and T. E. Tamminen-Möbius, "International system of radiographic grading of vesicoureteric reflux," *Pediatric Radiology*, vol. 15, no. 2, pp. 105–109, 1985.
- [4] G. Ardissino, L. Avolio, V. Dacco et al., "Long-term outcome of vesicoureteral reflux associated chronic renal failure in children. Data from the Italkid Project," *The Journal of Urology*, vol. 172, no. 1, pp. 305–310, 2004.
- [5] J. Wan, S. J. Skoog, W. C. Hulbert et al., "Section on urology response to new guidelines for the diagnosis and management of UTI," *Pediatrics*, vol. 129, no. 4, pp. 1051–1053, 2012.
- [6] R. Stein, H. S. Dogan, P. Hoebcke et al., "Urinary tract infections in children: EAU/ESPU guidelines," *European Urology*, vol. 67, no. 3, pp. 546–558, 2015.
- [7] W. H. Foresman, W. C. Hulbert Jr., and R. Rabinowitz, "Does urinary tract ultrasonography at hospitalization for acute pyelonephritis predict vesicoureteral reflux?" *The Journal of Urology*, vol. 165, no. 6, pp. 2232–2234, 2001.
- [8] A. Hoberman, M. Charron, R. W. Hickey, M. Baskin, D. H. Kearney, and E. R. Wald, "Imaging studies after a first febrile urinary tract infection in young children," *The New England Journal of Medicine*, vol. 348, no. 3, pp. 195–202, 2003.
- [9] G. Zamir, W. Sakran, Y. Horowitz, A. Koren, and D. Miron, "Urinary tract infection: is there a need for routine renal ultrasonography?" *Archives of Disease in Childhood*, vol. 89, no. 5, pp. 466–468, 2004.
- [10] A. Hoberman and R. Keren, "Antimicrobial prophylaxis for urinary tract infection in children," *The New England Journal of Medicine*, vol. 361, no. 18, pp. 1804–1806, 2009.
- [11] J. M. Smellie, U. Jodal, H. Lax, T. T. Möbius, H. Hirche, and H. Olbing, "Outcome at 10 years of severe vesicoureteric reflux managed medically: report of the international reflux study in children," *Journal of Pediatrics*, vol. 139, no. 5, pp. 656–663, 2001.
- [12] J. C. Craig, L. M. Irwig, J. Christie et al., "Variation in the diagnosis of vesicoureteric reflux using micturating cystourethrography," *Pediatric Nephrology*, vol. 11, no. 4, pp. 455–459, 1997.
- [13] K. A. Kronemer, S. Don, G. D. Luker, and C. Hildebolt, "Soft-copy versus hard-copy interpretation of voiding cystourethrography in neonates, infants, and children," *American Journal of Roentgenology*, vol. 172, no. 3, pp. 791–793, 1999.
- [14] S. P. Greenfield, M. A. Carpenter, R. W. Chesney, J. M. Zerlin, and J. Chow, "The RIVUR voiding cystourethrogram pilot study: experience with radiologic reading concordance," *Journal of Urology*, vol. 188, no. 4, pp. 1608–1612, 2012.
- [15] G. Roussey-Kesler, V. Gadjos, N. Idres et al., "Antibiotic prophylaxis for the prevention of recurrent urinary tract infection in children with low grade vesicoureteral reflux: results from a prospective randomized study," *Journal of Urology*, vol. 179, no. 2, pp. 674–679, 2008.
- [16] Z. M. McMillan, J. C. Austin, M. J. Knudson, C. E. Hawtrey, and C. S. Cooper, "Bladder volume at onset of reflux on initial cystogram predicts spontaneous resolution," *The Journal of Urology*, vol. 176, no. 4, pp. 1838–1841, 2006.
- [17] M. J. Knudson, J. C. Austin, Z. M. McMillan, C. E. Hawtrey, and C. S. Cooper, "Predictive factors of early spontaneous resolution in children with primary vesicoureteral reflux," *Journal of Urology*, vol. 178, no. 4, pp. 1684–1688, 2007.
- [18] C. S. Cooper, K. K. Birusingh, J. C. Austin, M. J. Knudson, and P. D. Brophy, "Distal ureteral diameter measurement objectively predicts vesicoureteral reflux outcome," *Journal of Pediatric Urology*, vol. 9, no. 1, pp. 99–103, 2013.
- [19] J. P. Vallee, M. P. D. Vallee, S. P. Greenfield, J. Wan, and J. Springate, "Contemporary incidence of morbidity related to vesicoureteral reflux," *Urology*, vol. 53, no. 4, pp. 812–815, 1999.
- [20] R. Weiss, T. Tamminen-Möbius, O. Koskimies et al., "Characteristics at entry of children with severe primary vesicoureteral reflux recruited for a multicenter, international therapeutic trial comparing medical and surgical management. The International Reflux Study in Children," *The Journal of Urology*, vol. 148, no. 5, part 2, pp. 1644–1649, 1992.
- [21] P. J. Karanicolas, M. Bhandari, H. Kreder et al., "Evaluating agreement: conducting a reliability study," *The Journal of Bone & Joint Surgery—American Volume*, vol. 91, no. 3, pp. 99–106, 2009.
- [22] G. Montini, L. Rigon, P. Zucchetta et al., "Prophylaxis after first febrile urinary tract infection in children? A multicenter, randomized, controlled, noninferiority trial," *Pediatrics*, vol. 122, no. 5, pp. 1064–1071, 2008.
- [23] C. W. Schwab Jr., H.-Y. Wu, H. Selman Grahame, H. H. Smith, H. M. Snyder III, and D. A. Canning, "Spontaneous resolution of vesicoureteral reflux: a 15-year perspective," *Journal of Urology*, vol. 168, no. 6, pp. 2594–2599, 2002.