



# A quantitative analysis of voiding cystourethrogram features confirms the association between high-grade vesicoureteral reflux with male sex, younger age, and hydronephrosis

Adree Khondker<sup>1</sup>, Jethro C.C. Kwong<sup>2</sup>, Priyank Yadav<sup>1</sup>, Justin Y.H. Chan<sup>2</sup>, Anuradha Singh<sup>3</sup>, Lauren Erdman<sup>4</sup>, Daniel T. Keefe<sup>1,5</sup>, Mandy Rickard<sup>1</sup>, Armando J. Lorenzo<sup>1</sup>

<sup>1</sup>Division of Urology, The Hospital for Sick Children, Toronto, ON, Canada; <sup>2</sup>Division of Urology, Department of Surgery, University of Toronto, ON, Canada; <sup>3</sup>Division of Medical Imaging, The Hospital for Sick Children, Toronto, ON, Canada; <sup>4</sup>Department of Computer Science, University of Toronto, Toronto, ON, Canada; <sup>5</sup>Department of Surgery, IWK Hospital, Halifax, NS, Canada

Cite as: Khondker A, Kwong JCC, Yadav P, et al. A quantitative analysis of voiding cystourethrogram features confirms the association between high-grade vesicoureteral reflux with male sex, younger age, and hydronephrosis. *Can Urol Assoc J* 2023;17(8):243-6. <http://dx.doi.org/10.5489/cuaj.8460>

## ABSTRACT

**INTRODUCTION:** Vesicoureteral reflux (VUR) is commonly diagnosed in the workup of urinary tract infections or hydronephrosis in children. Traditionally, VUR severity is graded subjectively based on voiding cystourethrogram (VCUG) imaging. Herein, we characterized the association between age, sex, and indication for VCUG, by employing standardized quantitative features.

**METHODS:** We included renal units with a high certainty in VUR grade (>80% consensus) from the qVUR model validation study at our institution between 2013 and 2019. We abstracted the following variables: age, sex, laterality, indication for VCUG, and qVUR parameters (tortuosity, ureter widths on VCUG). High-grade VUR was defined as grade 4 or 5. The association between each variable and VUR grade was assessed.

**RESULTS:** A total of 443 patients (523 renal units) were included, consisting of a 48:52 male/female ratio. The median age at VCUG was 13 months. Younger age at VCUG (<6 months) was associated with greater odds of severe VUR (odds ratio [OR] 2.0), and there was a weak correlation between age and VUR grade ( $p=-0.17$ ). Male sex was associated with increased odds of high-grade VUR (OR 2.7). VCUGs indicated for hydronephrosis were associated with high-grade VUR (OR 4.1) compared to those indicated for UTI only. Ureter tortuosity and width were significantly associated with each clinical variable and VUR severity.

**CONCLUSIONS:** Male sex, younger age (<6 months), and history of hydronephrosis are associated with both high-grade VUR and standardized quantitative measures, including greater ureter tortuosity and increased ureteral width. This lends support to quantitative assessment to improve reliability in VUR grading.

## INTRODUCTION

Vesicoureteral reflux (VUR) is an important prognostic factor in children for recurrent urinary tract infection (UTI) and may predispose patients to adverse kidney outcomes.<sup>1,2</sup> The reported incidence of VUR in all children ranges from 0.4–1.8%, while it is present in 30–50% of children who experience a UTI.<sup>3,4</sup> Importantly, children with VUR can benefit from prophylactic antibiotics and those with severe VUR will often require surgical intervention.<sup>4,5</sup> VUR severity is determined from a grade between 1 and 5 from voiding cystourethrogram (VCUG).<sup>6</sup> VUR severity has been associated with the risk of breakthrough UTI, renal scarring, and likelihood of self-resolution of VUR.<sup>7-10</sup>

Previous studies have shown that VUR affects more males than females, with higher grades at younger ages;<sup>8,11-13</sup> however, the subjectivity in VUR grading and institutional differences have limited the generalizability of these associations by reducing our certainty in classifying VUR.<sup>14-16</sup> This may explain the contradictory findings that have hindered the development of more individualized approaches to VUR imaging.<sup>17</sup> With the current drive to reduce invasive imaging in children<sup>18,19</sup> and the need to reliably assess VUR severity, there is a role for standardized assessment of VUR grading. Our group has recently developed qVUR, a multi-center,

validated model, which determines quantitative features from VCUGs to predict VUR grade.<sup>17</sup>

The purpose of this study was to investigate the association of objective parameters of VUR severity (grade and quantitative features on VCUG) with clinical information, such as age, sex, and indication for imaging.

## METHODS

### Setting, study population, variables

qVUR is a multi-institutional, machine-learning model that was trained from the imaging repository at The Hospital for Sick Children (Toronto, ON, 2013–2019).<sup>17</sup> In qVUR, VCUGs are split into two renal units (i.e., left

and right) and assessed for VUR grade by >5 clinical raters. Due to the highly subjective nature of VUR grading, only renal units with very high agreement (agreement between >80% of raters) were included in this study.<sup>14,15</sup> Renal units, or patients, were excluded for unavailable clinical information, VUR grade 1, or major congenital abnormalities.

Patient age, sex, and indication for VCUG were abstracted. The indication for VCUG was then abstracted and categorized into: UTI, hydronephrosis (HN, including pelviectasis, caliectasis), both, and other (followup outside VUR, query duplex, congenital abnormality, lower urinary tract symptoms, sepsis, decreasing renal function, unknown).

The primary outcome for VUR severity was severe or high-grade VUR, defined as VUR grade 4 or 5 (dichotomous). We also assessed VUR severity by quantitative features (continuous): tortuosity and ureteral width, as described below.

### Image annotation

To determine quantitative features from each VCUG, we used standardized feature labelling and annotation, provided in qVUR.<sup>16,17</sup> Briefly, two features of each renal unit were annotated and approximated within the model: ureter tortuosity and maximum ureter width. Each feature has been validated to correlate with VUR severity, and qVUR is freely available for use at: <https://sickkidsurology.shinyapps.io/qVUR/>.

### Statistical analysis

Descriptive statistics were tabulated between renal units. Comparative statistics between renal units were performed using the Kruskal-Wallis test for continuous variables. Fisher's exact test and Chi-squared test were used for categorical variables. Spearman's rank correlation was calculated for continuous associations. All statistical analyses were performed using Matlab 2021a (Mathworks), with a p-value of <0.05 used for statistical significance.

## RESULTS

A total of 6288 renal units (from 3144 VCUGs) were identified in the study period and screened for VUR. Of these, 1935 renal units had documented VUR, with 1220 meeting inclusion criteria, and 523 (43%) reached at least 80% agreement on the individual VUR grade between the raters.

The 523 renal units came from 443 patients, including 363 patients with unilateral reflux and 80 patients with bilateral reflux. The population characteristics are provided in Table 1. The median age at VCUG was 13

	Unilateral reflux	Bilateral reflux	Unilateral vs. bilateral	Overall
	n (%)	n (%)	p	Median/n (IQR/%)
Number of patients	363 (82%)	80 (18%)	–	443
Number of renal units*	363 (69%)	160 (31%)	–	523
Age (months)				
Median age (IQR)	13 (IQR 4, 51)	13 (IQR 3, 37)	0.11	13 (IQR 3, 44)
<6 months	127 (35%)	68 (43%)	0.12	195 (37%)
>6 months	236 (65%)	92 (57%)		328 (63%)
Sex				
Male	165 (45%)	86 (54%)	0.09	251 (48%)
Female	198 (54%)	74 (46%)		272 (52%)
VUR grade				
2	87 (24%)	31 (19%)	0.16	118 (23%)
3	151 (42%)	78 (49%)		229 (44%)
4	66 (18%)	20 (13%)		86 (16%)
5	59 (16%)	31 (19%)		90 (17%)
Laterality				
Left	212 (58%)	–	–	292 (56%)
Right	151 (42%)	–		231 (44%)
Indication				
UTI + HN	31 (9%)	12 (8%)	0.62	43 (8%)
HN	72 (19%)	34 (21%)		106 (20%)
UTI	175 (48%)	84 (53%)		259 (50%)
Other	85 (23%)	30 (19%)		105 (20%)

\*Used as denominator for statistical analysis. HN: hydronephrosis; IQR: interquartile range; UTI: urinary tract infection.

months (interquartile range [IQR] 3, 44). Of the 523 renal units, 272 (52%) were from female patients and 231 (44%) were right-sided. No significant differences in age, sex, VUR grade, or indication for VCUG were observed between patients with unilateral and bilateral reflux.

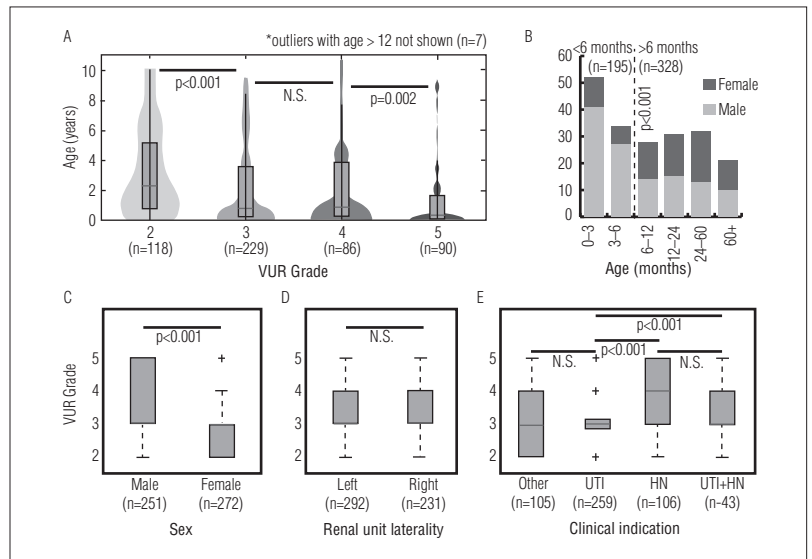
The median age at presentation for patients with grade 5 VUR was four months (IQR 1, 19), while patients with grade 2 VUR were 27 months (IQR 9, 62) at imaging ( $p < 0.001$ ) (Figure 1A). There is a significant weak correlation between age and VUR grade ( $\rho = -0.17$ ,  $p < 0.001$ ), as shown in Figure 1A. Renal units from patients with VCUG before six months of age had increased odds of having high-grade VUR than imaging occurring at  $> 6$  months (odds ratio [OR] 2.0, 95% confidence interval [CI] 1.4, 2.9,  $p < 0.001$ ) (Figure 1B). The male/female ratio for patients  $< 6$  months was 4:1 compared to 1:1 patients  $> 6$  months at imaging ( $p < 0.001$ ). Male sex was associated with increased odds of high-grade VUR (OR 2.7, 95% CI 1.8, 3.9,  $p < 0.001$ ) and greater median VUR grade ( $p < 0.001$ ) (Figure 1C) compared to female sex.

There was no significant difference in VUR grade between left- and right-sided VUR ( $p = 0.40$ ) (Figure 1D). VCUGs indicated in the workup for HN were associated with increased odds of having high-grade VUR compared to those with only UTI concerns (OR 4.1, 95% CI 2.6, 6.3,  $p < 0.001$ ) (Figure 1E). VUR grade was strongly associated with tortuosity and maximum ureteral dilatation (Figures 2A, 2D). Male sex, age  $< 6$  months, and HN were associated with a significantly greater tortuosity and increased maximum ureteral dilatation (Figures 2B–D, 2F–H).

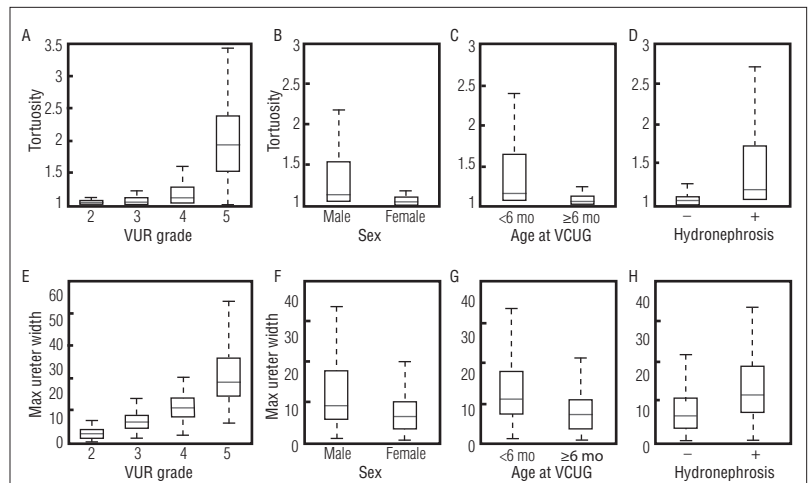
## DISCUSSION

Current studies lack reliability in VUR grading to determine associations between VUR severity and clinical characteristics. By associating quantitative features on VCUG, with qVUR, with clinical characteristics, we can be more certain in the strength of their associations. By demonstrating known clinical characteristics are associated with these quantitative features, we can begin implementing qVUR measures as a more reliable metric for VUR severity. This may be clinically useful in determining tailored workup for children with VUR and inform future reporting standards for VCUG.

There is a high degree of variability in VUR grading between clinicians when interpreting VCUGs, which limits the validity of previous studies assessing the association of VUR severity and clinical outcomes.<sup>4,14,15,20</sup> For example, Schaeffer et al showed that there was consensus on VUR grade by three radiologists in only 43% of renal units.<sup>14</sup> The strength of this study stems



**Figure 1.** (A) Distribution of age for patients with each VUR grade. (B) percent of patients with severe (grade 4+) VUR based on age at VCUG with comparison between  $< 6$  and  $> 6$  months with % male/female per group. Comparison of VUR grade between (C) male vs. female, (D) left vs. right renal unit, and (E) indication for VCUG. VCUG: voiding cystourethrogram; VUR: vesicoureteral reflux.



**Figure 2.** (A) Tortuosity by VUR grade with (B) sex, (C) age, or (D) indication for VCUG. (E) Maximum ureter width by VUR grade, with (F) sex, (G) age, or (H) indication for VCUG. All comparisons are significant ( $p < 0.001$  on Kruskal Wallis test). VCUG: voiding cystourethrogram; VUR: vesicoureteral reflux.

from the use of a consensus grade from at least five experts to ensure reliable outcome assessment.

In our study, we corroborate findings from previous studies with a validated approach using both VUR grade and quantitative VUR features. For example, multiple studies have shown VUR prevalence to decline with increasing age and to have an older onset in females.<sup>11,12,21</sup> We confirm this correlation between age, sex, and both VUR prevalence and grade. Capozza et al suggest that the natural history of VUR is different between males and females;<sup>13</sup> younger males are likely

to have higher-grade VUR, which will likely self-resolve, potentially due to higher detrusor pressures,<sup>22</sup> while non-infant males and females are likely to have a similar prevalence of VUR grade. Our results agree with this finding, as younger males are found to be the most at-risk for severe VUR (compared to older females).

Over the last decade, evidence and guidelines have recommended more judicious use of VCUGs.<sup>18</sup> Our study suggests that certain clinical factors may predispose patients to greater risk of severe VUR, which adds nuance to VCUG ordering patterns. For example, in our study, a history of HN is associated with higher VUR grade, as also shown in previous studies.<sup>12</sup> With current guidelines recommending surgery in VUR grade 4 or 5, this raises a question of whether a VCUG would affect management in patients without HN, as there would be a lower likelihood of showing high-grade VUR.

### Limitations

Limitations of the current study include its retrospective methodology and largely case-control analysis. Additionally, the patients studied in this work may correspond to patients who have more significant VUR than the general population because patients included in this analysis were indicated for VCUG and patients with VUR grade I were excluded. Extending upon this, as it is a strength to have a high certainty in VUR grade label, it also introduces a selection bias against patients with uncertain imaging or more complex patients. Patients who have higher-grade VUR are more likely to be symptomatic, prompting investigation at a younger age, which may introduce bias to these results. Lastly, this study lacks prospective data associating VUR severity with definite clinical outcomes, such as recurrent or breakthrough UTIs, reduced kidney function, or need for surgical intervention — primary considerations in current guidelines.<sup>4,23</sup>

### CONCLUSIONS

Male sex, younger age, and HN were all associated with more severe VUR grade, ureter tortuosity, and dilatation. While further validation is needed, we demonstrate a strong association between validated measures for VUR severity and clinical characteristics, which informs the epidemiology of VUR, while highlighting the need for more standard VUR grading.

COMPETING INTERESTS: The authors do not report any competing personal or financial interests related to this work.

This paper was reviewed by the CUA 2023 Scientific Program Co-Chairs as part of the CUA prize essay contest, and was presented at the CUA annual meeting, June 23–25, 2023 in Montreal, QC.

### REFERENCES

- Leonardo CR, Filgueiras MFT, Vasconcelos MM, et al. Risk factors for renal scarring in children and adolescents with lower urinary tract dysfunction. *Pediatr Nephrol* 2007;22:1891-6. <https://doi.org/10.1007/s00467-007-0564-5>
- Cooper CS. Diagnosis and management of vesicoureteral reflux in children. *Nat Rev Urol* 2009;6:481-9. <https://doi.org/10.1038/nrurol.2009.150>
- Sargent MA. Opinion: what is the normal prevalence of vesicoureteral reflux? *Pediatr Radiol* 2000;30:587-93.
- Tekgül S, Riedmiller H, Hoebeke P, et al. EAU guidelines on vesicoureteral reflux in children. *Eur Urol* 2012;62:534-42. <https://doi.org/10.1016/j.eururo.2012.05.059>
- Investigators RT. Antimicrobial prophylaxis for children with vesicoureteral reflux. *N Engl J Med* 2014;370:2367-76. <https://doi.org/10.1056/NEJMoa1401811>
- Blickman JG, Taylor GA, Lebowitz RL. Voiding cystourethrography: The initial radiologic study in children with urinary tract infection. *Radiology* 1985;156:659-62. <https://doi.org/10.1148/radiology.156.3.4040642>
- Goldraich NP, Goldraich IH. Followup of conservatively treated children with high- and low-grade vesicoureteral reflux: A prospective study. *J Urol* 1992;148:1688-92. [https://doi.org/10.1016/S0022-5347\(17\)37003-9](https://doi.org/10.1016/S0022-5347(17)37003-9)
- Alsaywid BS, Saleh H, Deshpande A, et al. High-grade primary vesicoureteral reflux in boys: Long-term results of a prospective cohort study. *J Urol* 2010;184:1598-603. <https://doi.org/10.1016/j.juro.2010.04.021>
- Estrada CR, Passerotti CC, Graham DA, et al. Nomograms for predicting annual resolution rate of primary vesicoureteral reflux: results from 2462 children. *J Urol* 2009;182:1535-41. <https://doi.org/10.1016/j.juro.2009.06.053>
- Kirsch AJ, Arlen AM, Leong T, et al. Vesicoureteral reflux index (VURx): A novel tool to predict primary reflux improvement and resolution in children less than 2 years of age. *J Pediatr Urol* 2014;10:1249-54. <https://doi.org/10.1016/j.jpuro.2014.06.019>
- Hannula A, Venhola M, Renko M, et al. Vesicoureteral reflux in children with suspected and proven urinary tract infection. *Pediatr Nephrol* 2010;25:1463-9. <https://doi.org/10.1007/s00467-010-1542-x>
- Vachvanichsanong P, Dissaneewate P, McNeil E. Primary vesicoureteral reflux: A 26-year experience in a single center. *Nephrology* 2016;21:335-40. <https://doi.org/10.1111/nep.12615>
- Capozza N, Gulia C, Heidari Bateni Z, et al. Vesicoureteral reflux in infants: What do we know about the gender prevalence by age. *Eur Rev Med Pharmacol Sci* 2017;21:5321-9.
- Schaeffer AJ, Greenfield SP, Ivanova A, et al. Reliability of grading of vesicoureteral reflux and other findings on voiding cystourethrography. *J Pediatr Urol* 2017;13:192-8. <https://doi.org/10.1016/j.jpuro.2016.06.020>
- Metcalfe CB, MacNeilly AE, Afshar K. Reliability assessment of international grading system for vesicoureteral reflux. *J Urol* 2012;188:1490-2. <http://dx.doi.org/10.1016/j.juro.2012.02.015>
- Khondker A, Kwong JCC, Rickard M, et al. A machine learning-based approach for quantitative grading of vesicoureteral reflux from voiding cystourethrograms: Methods and proof of concept. *J Pediatr Urol* 2022;18:78.e1-7. <https://doi.org/10.1016/j.jpuro.2021.10.009>
- Khondker A, Kwong JCC, Yadav P, et al. Multi-institutional validation of improved vesicoureteral reflux assessment with simple and machine learning approaches. *J Urol* 2022;208:1314-22. <https://doi.org/10.1097/JU.0000000000002987>
- Ming JM, Lee LC, Chua ME, et al. Population-based trend analysis of voiding cystourethrogram ordering practices in a single-payer healthcare system before and after the release of evaluation guidelines. *J Pediatr Urol* 2019;15:152-e1. <https://doi.org/10.1016/j.jpuro.2018.12.009>
- Gaske MJ, Applegate KE, Boylan J, et al. The Image Gently campaign: Working together to change practice. *AJR Am J Roentgenol* 2008;190:273-4. <https://doi.org/10.2214/AJR.07.3526>
- Arlen AM, Kirsch AJ, Leong T, et al. Validation of the ureteral diameter ratio for predicting early spontaneous resolution of primary vesicoureteral reflux. *J Pediatr Urol* 2017;13:383.e1-6. <https://doi.org/10.1016/j.jpuro.2017.01.012>
- Caragata R, Bălănescu L, Bălănescu R, et al. The epidemiology of vesicoureteral reflux in children—a nine-year retrospective study. *Rom J Urol* 2019;18:17-20. <https://doi.org/10.1155/2019/9560247>
- Yeung C, Sreedhar B, Leung YV, et al. Correlation between ultrasonographic bladder measurements and urodynamic findings in children with recurrent urinary tract infection. *BJU Int* 2007;99:651-5. <https://doi.org/10.1111/j.1464-410X.2006.06580.x>
- Peters CA, Skoog SJ, Arant BS, et al. Summary of the AUA guideline on management of primary vesicoureteral reflux in children. *J Urol* 2010;184:1134-44. <https://doi.org/10.1016/j.juro.2010.05.065>

CORRESPONDENCE: Dr. Armando J. Lorenzo, Hospital for Sick Children, Toronto, ON, Canada; armando.lorenzo@sickkids.ca