Published in final edited form as:

J Pediatr Urol. 2015 August ; 11(4): 176.e1–176.e7. doi:10.1016/j.jpuro.2015.03.006.

Predictive value of specific ultrasound findings when used as a screening test for abnormalities on VCUG

Tanya Logvinenko^{a,b}, Jeanne S. Chow^{a,c,d}, and Caleb P. Nelson^{a,c,e,*}

^aDepartment of Urology, Boston Children's Hospital, Boston, MA, USA

^bClinical Research Center, Boston Children's Hospital, Boston, MA, USA

^cDepartment of Radiology, Boston Children's Hospital, Boston, MA, USA

^dDepartment of Radiology, Harvard Medical School, Boston, MA, USA

^eDepartment of Surgery, Harvard Medical School, Boston, MA, USA

Summary

Background—Renal and bladder ultrasound (RBUS) is often used as an initial screening test for children after urinary tract infection (UTI). The 2011 AAP guidelines specifically recommend that RBUS be performed first, with voiding cystourethrogram (VCUG) to be performed only if the ultrasound is abnormal. While prior research has suggested that RBUS is neither sensitive nor specific for VCUG findings, such as vesicoureteral reflux (VUR), it is uncertain as to whether specific RBUS findings, alone or in combination, might make RBUS more useful as a predictor of VCUG abnormalities.

Aims—To evaluate the association of specific RBUS with VCUG findings, and determine whether predictive models that accurately predict patients at high risk of VCUG abnormalities, based on RBUS findings, can be constructed.

Methods and study sample—A total of 3995 patients were identified with VCUG and RBUS performed on the same day. The RBUS and VCUG reports were reviewed and the findings were classified. Analysis was limited to patients aged 0–60 months with no prior postnatal genitourinary imaging and no history of prenatal hydronephrosis.

Analysis—The associations between large numbers of specific RBUS findings with abnormalities seen on VCUG were investigated. Both multivariate logistic models and a neural network machine learning algorithms were constructed to evaluate the predictive power of RBUS

*Corresponding Author: Caleb P. Nelson, MD, MPH, Department of Urology, Boston Children's Hospital, 300 Longwood Avenue, HU-355, Boston, MA 02115, USA, Tel: (617) 355-3776, Fax: (617) 730-0474, caleb.nelson@childrens.harvard.edu.

Conflict of interest statement

None of the authors have any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work.

Ethical consent

This research was approved by the Institutional Review Board and complied with all ethical standards for human subjects research.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

for VCUG abnormalities (including VUR or bladder/urethral findings). Sensitivity, specificity, predictive values and area under receiving operating curves (AUROC) of RBUS for VCUG abnormalities were determined.

Results—A total of 2259 patients with UTI as the indication for imaging were identified. The RBUS was reported as “normal” in 75.0%. On VCUG, any VUR was identified in 41.7%, VUR grade >II in 20.9%, and VUR grade >III in 2.8%. Many individual RBUS findings were significantly associated with VUR on VCUG. Despite these strong univariate associations, multivariate modeling didn’t result in a predictive model that was highly accurate. Multivariate logistic regression built via stepwise selection had: AUROC=0.57, sensitivity=86% and specificity=25% for any VUR; AUROC=0.60, sensitivity=5% and specificity=99% for VUR grade >II; and AUROC=0.67, sensitivity=6% and specificity=99% for VUR grade >III. The best predictive model constructed via neural networks had: AUROC=0.69, sensitivity=64% and specificity=60% for any VUR; AUROC=0.67, sensitivity=18% and specificity=98% for VUR grade >II; and AUROC=0.79, sensitivity=32% and specificity=100% for VUR grade >III.

Conclusions—Even with the state-of-the-art predictive models, abnormal findings on RBUS provide a poor screening test for genitourinary abnormalities. Renal bladder ultrasound and VCUG should be considered complementary, as they provide important, but different, information.

Keywords

Urinary tract infection; Imaging; Vesicoureteral reflux; Pediatrics

Introduction

The 2011 recommendations from the American Academy of Pediatrics (AAP) for evaluation of children after an initial febrile UTI revised the previous edition’s recommendation [1] that all such children undergo renal and bladder ultrasound (RBUS) and VCUG. The new guidelines specify that: after a febrile UTI, children aged 2–24 months should undergo only RBUS, with VCUG reserved for those whose RBUS shows “hydronephrosis, scarring, or other findings that would suggest either high-grade VUR or obstructive uropathy” [2].

This use of RBUS as a screening test for abnormalities (such as VUR) that are diagnosed by VCUG has raised the question of whether RBUS is an accurate predictor of such findings. Previous analyses have suggested that RBUS is a poor screening test with variable sensitivity and specificity [3–5]. However, most of these studies have been very limited in the degree to which they have investigated the predictive power of specific RBUS findings (e.g. renal collecting system duplication, renal cysts, ureteral dilation, etc.) to accurately identify children most likely to have findings on VCUG. It is possible that certain RBUS findings, along or in combination through multivariate modeling, would make RBUS a more useful predictor than has been previously reported. If such models could be developed, they might offer a way of identifying children for whom early VCUG (after the initial UTI) would be more likely to yield significant findings.

The present study sought to evaluate the association of specific RBUS findings with VCUG abnormalities (and VUR in particular), and to determine whether multivariate predictive models can be constructed that accurately identify patients at high risk of VCUG abnormalities, based on RBUS findings.

Methods

Data source and patient selection

The present study was conducted on a database of patients age <60 months who underwent both RBUS and VCUG on the same day for a history of UTI, and who had no prior history of postnatal genitourinary imaging (based on either records of prior imaging in the system or documentation of prior outside imaging in the report) or prenatal hydronephrosis. The details of the cohort have previously been described [3]. In summary: billing between 1/1/2006 and 12/31/2010 was retrospectively reviewed, the patients were classified, and the findings from the radiology reports (and clinical records, when available) were abstracted into the database (images themselves were not reviewed). Institutional review board approval was obtained for the study.

Imaging data abstraction and classification

As previously described [3], the RBUS and VCUG findings were categorized based on the radiology report. When a specific finding was not commented upon (e.g. bladder diverticulum), it was assumed that this finding was not present. The RBUS findings included: renal dilation (hydronephrosis, pelviectasis, pelvocaliectasis, caliectasis, pelvic/calyceal dilation, extra-renal pelvis, fullness, prominence); ureteral dilation (hydroureter, ureterectasis, megaureter); parenchymal findings (abnormal echogenicity or cortico-medullary differentiation, cortical thinning/scarring, cysts, ectopia, duplication, hypotrophy or size discrepancy, agenesis, calcifications); and bladder findings (bladder wall thickening, trabeculation, diverticulum, ureterocele, dilated posterior urethra, and debris). Dilation was graded based on the clinical reading provided in the report, using the descriptive mild-moderate-severe scale, which remains the most widely used grading scale among North American pediatric urologists [6]. At the present institution, these grades approximate to the Society for Fetal Urology (SFU) scale [7] as follows: mild = SFU grade 1–2, moderate = SFU grade 3, severe = SFU grade 4. Renal parenchymal findings included: abnormal echogenicity, abnormal cortico-medullary differentiation, cortical thinning/scarring, cysts, ectopia, duplication, hypotrophy/atrophy or renal size discrepancy, agenesis, or calcifications. Bladder findings included: bladder wall thickening, trabeculation, diverticulum, ureterocele, dilated posterior urethra, and debris.

The VCUG findings included: VUR (graded on the five-point international grading system [8]); diverticulum (peri-ureteral or otherwise); trabeculation, ureterocele, and urethral findings (e.g. PUV). Because there is disagreement among some clinicians regarding which findings on a VCUG are “significant”, three different VCUG thresholds for a “positive” study were defined, varying by VUR grade. At the most relaxed threshold (Threshold A), any degree of VUR (or bladder diverticulum, trabeculation, ureterocele, or urethral abnormalities) qualified as a “positive” VCUG. For the middle threshold (Threshold B),

VUR > grade II (or bladder diverticulum, trabeculation, ureterocele, or urethral abnormalities) qualified as a “positive” VCUG. For the most stringent threshold (Threshold C), only VUR > grade III (or bladder diverticulum, trabeculation, ureterocele, or urethral abnormalities) qualified as a “positive” VCUG.

Data analysis and construction of predictive model

To investigate how individual RBUS findings were associated with each threshold of VCUG findings, the following were used: a logistic regression to derive odds ratios (OR); *P*-values corresponding to each association; receiver operating characteristic (ROC) curves were constructed; the area under the ROC curve (AUROC) was calculated for each predictor.

A comprehensive predictive model for each VCUG threshold was then constructed, using multiple logistic regression with stepwise covariate selection and individual abnormal findings on RBUS as predictors. Such models are clinically useful in that they allow for easy interpretation of each covariate in the multivariate model via the OR. However, logistic models only allow for linear relationships between predictors and logit of the probability of detecting an abnormality on a VCUG and, thus, are limited only to cases of a linear decision boundary. To address the situation of a possibly non-linear decision boundary, another set of predictive models was constructed using a more-complex neural networks machine learning algorithm [9]. A single hidden layer neural network with 10 nodes in the hidden layer was used. Neural networks with different numbers of nodes in the hidden layer lead to comparable results. The neural network fit classification models do not permit interpretation of the effect of each individual predictor covariate in the final model. However, they allow for complex non-linear decision boundaries, making the prediction more flexible compared to logistic regression [10]. For each of the predictive models, ROC curves were built and evaluated for sensitivity, specificity, PPV and NPV.

As a measure of performance for each of the predictive models, concordance statistics and AUROC, which assesses the ability of the model to discriminate between the subjects meeting and not meeting criteria for each VCUG threshold, were computed and evaluated [11]. An AUROC of 0.51–0.65, 0.66–0.80, and >0.8 was considered to represent a model with poor, fair, and good fit, respectively. All statistical analyses and model construction were performed using packages base, and nnet of R statistical software [12–14].

Results

A total of 2259 patients, aged 0–60 months and with UTI as the indication for imaging, were identified. Characteristics of the sample are shown in Table 1. On VCUG, any VUR was identified in 41.7%, VUR grade >II in 20.9%, and VUR grade >III in 2.8%. The number of participants with a positive VCUG under Threshold A was 975/2259 (43.2%), under Threshold B was 528/2259 (23.4%), and under Threshold C was 137/2259 (6.1%).

On univariate analysis, many RBUS findings were significantly associated with findings on VCUG (Table 2). Any abnormality on RBUS was strongly associated with Threshold A (OR=1.33, *P*=0.0032), Threshold B (OR=2.01, *P*<0.0001), and Threshold C (OR=4.03, *P*<0.0001). Both renal dilation and ureteral dilation were strongly associated with all three

thresholds, with the strongest associations with Threshold C (Table 2). Parenchymal abnormalities as a class were also associated with all three thresholds. However, certain findings were not individually associated with any threshold, including renal cysts, renal agenesis (solitary kidney) and renal ectopia. Only urothelial thickening was significantly associated with all three thresholds. Bladder findings such as bladder wall thickening, ureterocele, trabeculation, and debris were all strongly associated with the higher threshold (Threshold C), but not consistently with the lower thresholds (A and B). Bladder findings as a group were associated with all three thresholds.

Despite these strong univariate associations, multivariate modeling did not result in a predictive model that was highly accurate (Table 3), and none of the models met the definition of a “good fit”. Multivariate logistic regression built via stepwise selection had AUROC=0.57, sensitivity=86% and specificity=25% for Threshold A; AUROC=0.60, sensitivity=5% and specificity=99% for Threshold B; and AUROC=0.67, sensitivity=6% and specificity=99% for Threshold C. The best predictive model constructed via neural networks with single hidden layer and 10 nodes had AUROC=0.69, sensitivity=64% and specificity=60% for Threshold A; AUROC=0.67, sensitivity=18% and specificity=98% for Threshold B; and AUROC=0.79, sensitivity=32% and specificity=100% for Threshold C. For the multivariate logistic predictive models, the fit based on the observed AUROC statistics ranged from poor (0.57 for Threshold A and 0.60 for Threshold B) to fair (0.67 for Threshold C). For the neural networks machine learning algorithm, the predictive models performed slightly better, having a fair fit with AUROC concordance statistic of 0.69 for Threshold A, 0.67 for Threshold B and 0.79 for Threshold C.

Discussion

In Urology and Radiology literature there has been much discussion on the issue of the accuracy of ultrasound as a tool with which to identify abnormalities, such as VUR, that are seen on VCUG. The perceived morbidity of the VCUG, with its requirement for urethral catheterization and ionizing radiation, has naturally led to efforts to determine if ultrasound is sufficient to serve as a proxy in place of the VCUG for a subset of children. The recent publication of the AAP guidelines for the management of children aged 2–24 months after a first febrile UTI, with their recommendation for *de facto* use of RBUS as a screening test to determine which children should undergo VCUG, has resulted in increased interest in the question of just how useful RBUS can be for this purpose.

Many investigators have sought to report on the association between ultrasound and VCUG findings, and the results have been highly variable. Sensitivity has been reported from 18–79% and specificity from 41–99% [4, 5, 15–18]; these papers have tended to use different definitions of a positive RBUS (or no definition at all) and looked at varying VCUG outcomes. Much of the published literature is limited by verification bias (by which only some of the eligible participants underwent both RBUS and VCUG, often because the first study performed was normal) [4, 19, 20]. The current study is unique, in that it included a large sample of patients who underwent both RBUS and VCUG on the same day, with neither test being contingent on the findings of the first test performed. This practice was characteristic at the present institution during the study period, and tended to minimize the

verification bias in the sample. Much of the literature is otherwise limited by small numbers, narrow inclusion criteria, or tight (or broad) age ranges [5, 17, 21].

There is very little in the literature that assesses the relative contribution of specific individual RBUS findings to overall RBUS predictive values. The present group of authors have previously reported test characteristics for RBUS in this sample [3], and although a large number of specific RBUS findings were collected, the relative predictive contributions of these was not investigated. Instead, RBUS findings were characterized on a scale of relative severity based on clinical judgment, and findings above a certain severity were grouped to meet a certain overall threshold for a positive RBUS. Although several such thresholds were developed, based on the assumptions about relative severity of findings (i.e. it was assumed that severe hydronephrosis represented a higher severity level than mild hydronephrosis), these assumptions were not based on the model itself. Therefore, the ability in this previous study to determine the predictive value of such individual findings was limited. In this, the methodology mirrored that of Hoberman et al. [22], whose prospective evaluation included findings such as renal duplication and calculus, in addition to upper tract dilation, but did not break out of the individual predictors with regard to VCUG findings.

Individual predictors have been looked at in the past, albeit on a limited scale. For instance, Kenney et al. looked at the predictive value of ureteral dilation as an isolated finding on RBUS [23]. They found that ureteral dilation was a strong predictor of VUR, but did not determine the value of this finding within the context of other findings such as pelvocaliectasis or duplication. Similarly, a number of investigators have looked at the predictive value of urothelial thickening in the renal collecting system as a predictor of VUR [24–26]. Again, while this finding was associated with VUR in isolation, its utility as an independent predictor in the context of other findings was not reported. In contrast, in the present study, how individual RBUS findings perform as predictors for a range of VCUG finding thresholds have specifically been evaluated, while multiple other findings in a multivariate model have been controlled. The present study is the first to analyze all such findings comprehensively by using state-of-the-art modeling techniques.

What does this study add? In designing the study, it was hoped that the effort of modeling RBUS predictors would lead to an accurate, practical, decision-making tool that could be applied by clinicians to identify those children with a history of UTI in whom a VCUG would be “high-yield”. Children identified as low-risk could then be spared the VCUG. However, it was found that even with highly sophisticated modeling techniques, it was not possible to create a predictive model that was more than modestly successful in predicting VCUG findings. Renal bladder ultrasound is simply not a good predictor of VCUG findings, most glaringly for VUR of grade III or lower. In some cases, even high-grade VUR does not reveal itself through ultrasound.

The present results should be interpreted in light of their limitations. As a retrospective review, the ability to confirm all clinical events, such as UTI, was limited; many were diagnosed and treated in the community and could not be independently verified. This deficiency was mitigated by the fact that radiologists at the present institution routinely take detailed histories to confirm a clinical UTI prior to the VCUG; however, such histories are

subject to error and misclassification like any other record. Another limitation was that the radiology reports were used as the source of imaging findings; an independent reading was not performed. Both RBUS and VCUG findings were, therefore, subject to variability of interpretation (e.g. grading of VUR). Furthermore, as the radiologist reading one study was not systematically blinded to the findings of the other, interpretations may have been influenced by such knowledge. In most cases, however, different radiologists interpreted each study, and the RBUS was typically performed and read prior to the VCUG, which minimized the impact of this. To minimize verification bias, only patients who underwent both VCUG and RBUS on the same day were included. However, the data would not capture the small number of patients who underwent RBUS or VCUG separately, or who only had one test or the other. It is believed that local practice patterns during the study period in the community make this less likely due to widespread adherence to the 1999 AAP guidelines; however, there is no empirical evidence that this is the case. The routine practice of scheduling both RBUS and VCUG on the same day for patients with a history of UTI, as well as the routine completion of both tests regardless of findings on the first test, further minimizes these effects.

Conclusions

Although many individual RBUS findings have a strong univariate that is associated with VUR and other findings on VCUG, these findings cannot be combined to provide a highly sensitive and specific predictive test, even with state-of-the-art predictive modeling techniques. The RBUS is a poor screening test for VCUG-identified abnormalities. The RBUS and VCUG should be considered as complementary, as they provide important, but different, information.

Acknowledgments

Funding acknowledgement

This research was supported by a K23 career development grant from NIDDK, an institute of the NIH (US). The study sponsor had no role in: the study design; the collection, analysis and interpretation of data; the writing of the manuscript; or the decision to submit the manuscript for publication.

References

1. American Academy of Pediatrics, Committee on Quality Improvement. Subcommittee on Urinary Tract Infection. Practice parameter: the diagnosis, treatment, and evaluation of the initial urinary tract infection in febrile infants and young children. *Pediatrics*. 1999; 103 (4 Pt 1):843–52. [PubMed: 10103321]
2. Urinary Tract Infection: Clinical Practice Guideline for the Diagnosis and Management of the Initial UTI in Febrile Infants and Children 2 to 24 Months. *Pediatrics*. 2011; 128(3):595–610. [PubMed: 21873693]
3. Nelson CP, et al. Ultrasound as a screening test for genitourinary anomalies in children with UTI. *Pediatrics*. 2014; 133 (3):e394–403. [PubMed: 24515519]
4. Lee JH, Kim MK, Park SE. Is a routine voiding cystourethrogram necessary in children after the first febrile urinary tract infection? *Acta Paediatr*. 2012; 101(3):e105–9. [PubMed: 22040289]
5. Mahant S, Friedman J, MacArthur C. Renal ultrasound findings and vesicoureteral reflux in children hospitalised with urinary tract infection. *Arch Dis Child*. 2002; 86 (6):419–20. [PubMed: 12023172]

6. Swenson DW, et al. Characterizing upper urinary tract dilation on ultrasound: a survey of North American pediatric radiologists' practices. *Pediatr Radiol*. 2014
7. Fernbach SK, Maizels M, Conway JJ. Ultrasound grading of hydronephrosis: introduction to the system used by the society for fetal urology. *Pediatr Radiol*. 1993; 23 (6):478–80. [PubMed: 8255658]
8. Medical versus surgical treatment of primary vesicoureteral reflux: report of the International Reflux Study Committee. *Pediatrics*. 1981; 67 (3):392–400. [PubMed: 7017578]
9. Ripley, BD. *Pattern Recognition and Neural Networks*. Cambridge University Press; 1996.
10. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform*. 2002; 35 (5–6):352–9. [PubMed: 12968784]
11. Steyerberg, EW. *Clinical Prediction Models: A Practical Approach to Development, Validation and Updating*. New York: Springer; 2009.
12. R Core Team. *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013.
13. Liaw AWM. Classification and Regression by random Forest. *R-News*. 2002; 2 (3):18–22.
14. Venables, WN.; Ripley, BD. *Modern Applied Statistics with S*. 4. New York: Springer; 2002.
15. Rickwood AM, et al. Current imaging of childhood urinary infections: prospective survey. *BMJ*. 1992; 304 (6828):663–5. [PubMed: 1571636]
16. Zamir G, et al. Urinary tract infection: is there a need for routine renal ultrasonography? *Arch Dis Child*. 2004; 89 (5):466–8. [PubMed: 15102643]
17. Tsai JD, et al. Screening high-grade vesicoureteral reflux in young infants with a febrile urinary tract infection. *Pediatr Nephrol*. 2012; 27 (6):955–63. [PubMed: 22374404]
18. Preda I, et al. Normal dimercaptosuccinic acid scintigraphy makes voiding cystourethrography unnecessary after urinary tract infection. *J Pediatr*. 2007; 151 (6):581–4. [PubMed: 18035134]
19. Foresman WH, Hulbert WC Jr, Rabinowitz R. Does urinary tract ultrasonography at hospitalization for acute pyelonephritis predict vesicoureteral reflux? *J Urol*. 2001; 165 (6 Pt 2): 2232–4. [PubMed: 11371951]
20. Montini G, et al. Value of imaging studies after a first febrile urinary tract infection in young children: data from Italian renal infection study 1. *Pediatrics*. 2009; 123 (2):e239–46. [PubMed: 19139086]
21. Smellie JM, Rigden SP, Prescod NP. Urinary tract infection: a comparison of four methods of investigation. *Arch Dis Child*. 1995; 72 (3):247–50. [PubMed: 7741578]
22. Hoberman A, et al. Imaging studies after a first febrile urinary tract infection in young children. *N Engl J Med*. 2003; 348 (3):195–202. [PubMed: 12529459]
23. Kenney IJ, Negus AS, Miller FN. Is sonographically demonstrated mild distal ureteric dilatation predictive of vesicoureteric reflux as seen on micturating cystourethrography? *Pediatr Radiol*. 2002; 32 (3):175–8. [PubMed: 12164349]
24. Fan NC, Tain YL. Sonographic finding of persistent renal pelvic wall thickening in children. *Pediatr Neonatol*. 2014; 55 (4):269–74. [PubMed: 24373671]
25. Robben SG, et al. Significance of thickening of the wall of the renal collecting system in children: an ultrasound study. *Pediatr Radiol*. 1999; 29 (10):736–40. [PubMed: 10525779]
26. Sorantin E, et al. The sonographically thickened wall of the upper urinary tract system: correlation with other imaging methods. *Pediatr Radiol*. 1997; 27(8):667–71. [PubMed: 9252432]

Table 1

Participant and imaging characteristics of the sample

	Children aged <60 months with a history of UTI as the indication for initial GU imaging (<i>n</i> =2259)
Sex	
Female	1787 (79.1%)
Male: uncircumcised	306 (13.6%)
Male: circumcised	50 (2.2%)
Male: circumcision status unknown	116 (5.1%)
Age	
0–1 months	78 (3.45%)
2–24 months	1701 (75.3%)
25–59 months	480 (21.25%)
Prior UTI history	
First UTI	1557 (68.9%)
Recurrent UTI	176 (7.8%)
Recurrence status unknown	526 (23.3%)
UTI fever status	
Febrile UTI	2045 (90.5%)
Non-febrile UTI	89 (3.9%)
Febrile history unknown	125 (5.5%)
VUR grade on VCUG	
No VUR	1317 (58.3%)
I or II	469 (20.8%)
II–III or III	410 (18.1%)
>III	63 (2.8%)

Table 2

Univariate associations of renal bladder ultrasound findings with abnormal VCUG, among 2259 children with both imaging studies.

Factor	N (%)	Threshold A			Threshold B			Threshold C		
		OR	P-value	AUROC	OR	P-value	AUROC	OR	P-value	AUROC
Any RBUS abnormality (vs none)	565 (25.0%)	1.33	0.0032	0.53	2.01	<0.0001	0.57	4.03	<0.0001	0.66
Renal dilation^d										
Any present (vs absent)	338 (15.0%)	1.36	0.0086	0.52	2.32	<0.0001	0.56	4.67	<0.0001	0.64
Categorized by severity										
None (reference)	1927 (85.3%)	–	–	0.52	–	–	0.56	–	–	0.63
Mild (vs reference)	300 (13.3%)	1.30	0.0325	0.52	2.12	<0.0001	0.56	3.57	<0.0001	0.63
Mild-moderate (vs reference)	13 (0.6%)	1.18	0.7674	0.52	2.36	0.1346	0.56	6.58	0.0048	0.63
Moderate (vs reference)	9 (0.4%)	2.75	0.1531	0.52	4.71	0.0213	0.56	17.55	<0.0001	0.63
Moderate-severe (vs reference)	9 (0.4%)	1.72	0.4199	0.52	4.71	0.0213	0.56	17.55	<0.0001	0.63
Ureter dilation										
Any Present (vs absent)	91 (4.0%)	1.43	0.0967	0.51	2.97	<0.0001	0.53	7.41	<0.0001	0.58
Categorized by severity										
None (reference)	2181 (96.5%)	–	–	0.51	–	–	0.52	–	–	0.57
Mild (vs reference)	44 (1.9%)	1.46	0.2149	0.51	2.86	0.0006	0.52	6.74	<0.0001	0.57
Mild-moderate (vs reference)	9 (0.4%)	1.07	0.9241	0.51	2.75	0.1332	0.52	8.98	0.0021	0.57
Moderate (vs reference)	11 (0.5%)	0.76	0.6645	0.51	1.29	0.7099	0.52	1.80	0.5780	0.57
Moderate-severe (vs reference)	5 (0.2%)	5.33	0.1347	0.51	13.73	0.0192	0.52	71.86	0.0001	0.57
Severe (vs reference)	1 (0.0%)	>100	0.9684	0.51	>100	0.9661	0.52	>100	0.9755	0.57
Parenchymal abnormality										
Present (vs absent)	244 (10.8%)	1.61	0.0005	0.52	2.10	<0.0001	0.54	2.75	<0.0001	0.57
Dysplasia (vs none)	29 (1.3%)	1.88	0.0959	0.50	2.71	0.0082	0.51	6.20	<0.0001	0.52
Cysts (vs none)	16 (0.7%)	1.02	0.9619	0.50	1.49	0.4579	0.50	2.23	0.2918	0.50
Atrophy or size discrepancy (vs none)	90 (4.0%)	1.16	0.4935	0.50	1.35	0.2085	0.51	2.26	0.0150	0.52
Solitary kidney (vs none)	1 (0.0%)	0.00	0.9698	0.50	0.00	0.9720	0.50	0.00	0.9839	0.50
Duplication (vs none)	80 (3.5%)	2.03	0.0024	0.51	2.39	0.0002	0.52	2.03	0.0525	0.52

Factor	N (%)	Threshold A			Threshold B			Threshold C		
		OR	P-value	AUROC	OR	P-value	AUROC	OR	P-value	AUROC
Renal ectopia (vs none)	9 (0.4%)	0.37	0.2218	0.50	0.94	0.9349	0.50	0.00	0.9789	0.50
Kidney stone (vs none)	4 (0.2%)	3.96	0.2336	0.50	9.89	0.0474	0.50	0.00	0.9786	0.50
Urothelial thickening (vs none)	42 (1.9%)	3.80	0.0002	0.51	5.01	< 0.0001	0.52	5.89	< 0.0001	0.53
Bladder abnormality										
Present vs absent	62 (2.7%)	1.99	0.0089	0.51	1.98	0.0110	0.51	5.43	< 0.0001	0.54
Bladder wall thickening (vs none)	21 (0.9%)	2.16	0.0889	0.50	1.31	0.5729	0.50	3.72	0.0195	0.51
Ureterocele (vs none)	4 (0.2%)	1.32	0.7829	0.50	3.29	0.2346	0.50	15.70	0.0061	0.51
Trabeculation (vs none)	5 (0.2%)	5.29	0.1367	0.50	4.94	0.0806	0.50	10.46	0.0105	0.51
Debris (vs none)	36 (1.6%)	1.86	0.0680	0.50	2.38	0.0111	0.51	5.46	< 0.0001	0.53

^aCases where the report identified “fullness” of the collecting system or “extra-renal pelvis” were classified as “mild” dilation

Thresholds for “positive” VCUG:

Threshold A: any degree of VUR, or bladder diverticulum, trabeculation, ureterocele, or urethral abnormalities; Threshold B: VUR > grade II, or bladder diverticulum, trabeculation, ureterocele, or urethral abnormalities; Threshold C: VUR > grade III, or bladder diverticulum, trabeculation, ureterocele, or urethral abnormalities.

P-values <0.05 are in bold.

AUROC, area under the receiver operating characteristic curve; RBUS, renal bladder ultrasound;

Table 3

Characteristics of predictive models.

Predictive model	Sensitivity	Specificity	PPV	NPV	AUROC
Logistic regression					
Any VUR ^a	86.3%	24.7%	53.7%	64.1%	0.5735
VUR grade >II ^b	5.1%	99.1%	70.6%	71.9%	0.6018
VUR grade >III ^c	6.0%	99.9%	77.8%	93.2%	0.6742
Neural networks^d					
Any VUR	64.2%	59.6%	61.6%	62.2%	0.6852
VUR grade >II	17.8%	97.8%	76.4%	74.5%	0.6726
VUR grade >III	31.6%	99.8%	92.5%	95.0%	0.7863

^a Logistic model with urothelial thickening, duplication, trabeculation and kidney stones (adjusted for sex, circumcision status in boys and febrile UTI)

^b Logistic model with any renal bladder ultrasound abnormality, urothelial thickening, duplication, kidney stones (adjusted for age, sex and febrile UTI)

^c Logistic model with any renal bladder ultrasound abnormality, ureter dilation, urothelial thickening, debris, first or recurrent UTI

^d All individual RBUS abnormalities, sex, age, circumcision status in boys, febrile UTI, first (vs recurrent) UTI