The RIVUR Voiding Cystourethrogram Pilot Study: Experience with Radiologic Reading Concordance

Saul P. Greenfield,* Myra A. Carpenter, Russell W. Chesney, J. Michael Zerin and Jean Chow

From the Department of Urology, State University of New York at Buffalo School of Medicine & Biomedical Sciences, Department of Pediatric Urology, Women & Children's Hospital of Buffalo, Buffalo, New York (SPG), Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina (MAC), Department of Pediatrics, Le Bonheur Children's Medical Center, Memphis, Tennessee (RWC), Department of Radiology, Children's Hospital of Michigan, Detroit, Michigan (JMZ), and Department of Radiology, Children's Hospital Boston, Boston, Massachusetts (JC)

Abbreviations and Acronyms

UTI = urinary tract infection

VCUG = voiding cystourethrogram

VUR = vesicoureteral reflux

Study received institutional review board approval.

* Correspondence: Women & Children's Hospital of Buffalo, 219 Bryant St., Buffalo, New York 14222 (telephone: 716-878-7393; FAX: 716-878-7096; e-mail: spg@buffalo.edu).

Purpose: Two reference radiologists independently review voiding cystourethrograms for the National Institutes of Health sponsored RIVUR (Randomized Intervention for Children with Vesicoureteral Reflux) trial for children with vesicoureteral reflux. A pilot study was required from all clinical centers before enrolling patients.

Materials and Methods: Digital images were reviewed. Responses were compared and discrepancies adjudicated by teleconference to a final assessment.

Results: A total of 75 studies from 19 sites were reviewed. Discrepancies in vesicoureteral reflux grade level were noted on the left and right side in 11 (15%, kappa 0.85) and 12 (16%, kappa 0.83) ureters, respectively. Other areas of disagreement were the presence of paraureteral diverticulum (left 11%, kappa 0.31; right 9%, kappa 0.34), urethral anatomy (15%, kappa 0.33), whether the child voided (8%, kappa 0.21), the presence of ureteral duplication (left 7%, kappa 0.64; right 3%, kappa 0.78) and the presence of bladder trabeculation (5%, kappa 0.32). Of 83 ureters in which reflux was seen there was grade disagreement about 23 (28%). Of 61 ureters initially assessed as grade II or III reflux by both readers, there was disagreement on 9 (15%). Of these 9 discrepancies 7 (78%) were adjudicated to the higher grade (grade III).

Conclusions: Discrepancies in the assessment of intermediate grade vesicoureteral reflux were noteworthy. Recommendations for patients with grade II or III reflux advanced by studies which rely on a single reading, which categorize only grade III or higher reflux as significant, may not be valid.

Key Words: vesico-ureteral reflux, radiology, classification

Contrast voiding cystourethrograms have long been a mainstay of the diagnosis of vesicoureteral reflux.[1] Advantages are the ability to grade the severity of reflux using the widely accepted 5-level International Scale, visualization of bladder wall and bladder neck anatomy, and assessment of the male urethra during the voiding phase.[1,2] Despite certain inherent and unavoidable variability in technique, such as the amount of contrast material infused, the timing of spot fluoroscopic imaging and the reluctance of some children to void during fluoroscopy, the grading of reflux using the International Scale remains the major means of categorizing patients and determining treatment. However, the literature is sparse on the reliability

0022-5347/12/1884-1608/0
THE JOURNAL OF UROLOGY®

http://dx.doi.org/10.1016/j.juro.2012.06.032
Vol. 188, 1608-1612, October 2012
Printed in U.S.A.

of interobserver determinations of reflux grade.[3,4] Interobserver concordance has been shown to be high but imperfect.

Most recently several studies have suggested that children with grade III or greater reflux be treated more aggressively and that those with grade I or II reflux be observed without any intervention, medical or surgical.[5–7] These authors argue that the presence of low grade reflux is of little clinical consequence. Therefore, it would seem critical to be able to reliably distinguish between grade II and higher grade reflux.

The RIVUR study, sponsored by the National Institutes of Health, is a large, multi-institutional, randomized and controlled trial comparing placebo to antimicrobial prophylaxis for reducing recurrent UTIs in children with grade I to IV reflux.[8,9] All VCUGs at randomization and study exit are read by 2 reference radiologists. These 2 radiologists read all VCUGs obtained to assess eligibility and during the study. After separate readings are submitted to a central Data Coordinating Center, any differences are adjudicated and a final reading is entered into the study database. Before beginning the trial and allowing each participating institution to start enrolling patients, a VCUG pilot study was required. The objectives of this pilot study were to evaluate the quality of VCUG imaging, to test the method of study transmission to the radiologists and the Data Coordinating Center, and to assess interobserver variability. Successful completion of the pilot study was required for a clinical site to begin recruiting patients into the RIVUR trial.

MATERIALS AND METHODS

Pilot study VCUGs were obtained from the 5 core RIVUR sites (10 studies each) and 12 satellite sites (11 sites submitted 2 studies each and 1 site submitted 3 studies). Studies submitted for the pilot were a convenience sample of VCUGs available from patients generally representative of the children expected to enroll in the RIVUR trial.

Digital images were placed on compact discs in DICOM (Digital Imaging and Communications in Medicine) format at the participating clinical sites and mailed to 2 pediatric radiologists with extensive experience in uroradiology who were designated to read VCUGs for the RIVUR study (JMZ, JC). The RIVUR pilot study was begun toward the end of the time of transition from standard x-ray films to digital radiography. In the few cases in which film was used, the files were digitized, or duplicate films were obtained and mailed to the 2 radiologists. All pilot images were de-identified. The pilot study was approved by the institutional review board at each participating center.

Reader software was not specified. A web based form captured the radiologist assessments, including whether the patient voided, the grade of reflux in each ureter, the presence of ureteral duplication, and the details of bladder wall anatomy (diverticuli, shape, trabeculation) and of male and female urethral anatomy. The responses were compared electronically and discrepancies were flagged for adjudication. The 2 radiologists convened by teleconference to adjudicate the discrepancies to a consensus assessment that was also entered into the web based data management system.

VCUG assessments by 2 radiologists were described by the number and percent of concordant and discrepant responses (duplication was only analyzed when both radiologists indicated the presence of VUR in the associated ureter). Agreement in VUR grade rating was assessed by weighted kappa.[10] Agreement of variables with nominal response categories was by unweighted (simple) kappa.[11] The kappa statistic is a conservative, chance corrected measure of agreement between the raters' evaluations of a categorical item.[12] Weighted kappa was used for the ordinal ratings of VUR grade so that greater weights were applied as the magnitude of discrepancies increased. Statistical analyses were performed with SAS® version 9.2 software.

RESULTS

A total of 19 sites participated in the pilot study and 79 VCUGs were submitted. Four studies (2 from each of 2 sites) were read by only 1 radiologist for logistical reasons and this report is based on 75 VCUGs which were read by both radiologists. Adjudication was required for 1 or more fields in 65 (87%) of the 75 pilot VCUGs.

The overall concordance of responses for the 75 studies was high, ranging from 84% to 100%, with more discrepancies occurring in the grading of VUR and classification of the urethra as normal, not normal or indeterminate. Major areas of disagreement are shown in table 1. Discrepancies in VUR grade were noted on the left and right side in 11 (15%) and 12 (16%) patients, respectively. In 15 ureters (10%) there was disagreement regarding the presence of a paraureteral diverticulum. In 11 patients (15%) there was disagreement regarding whether the urethra was normal and in 8 (11%) whether the urethra was classified as a spinning top. Less common were discrepancies regarding whether the child voided in 6 (8%), regarding ureteral duplication in 4 (5%) and the presence of bladder trabeculation in 4 (5%). Kappa coefficients were provided for all radiographic assessments but the 95% confidence intervals for most nominal responses were wide. The appropriate interpretation of individual kappa statistics requires consideration of the percent agreement, number of assessments (which is smaller for items embedded in skip patterns) and prevalence index (high for most variables).[13,14] For example, despite strong agreement on voiding during the study (92%, table 1), the corresponding kappa of 0.21 was low due to the high prevalence of voiding (reflecting the heavy imbal-