A prospective study comparing direct radionuclide cystography with voiding cystourethrography in the detection and grading of vesicoureteric reflux in children

Dr. MD Khizer Razak and Dr. Sudhakar R

Abstract
Background: Vesico ureteric reflux (VUR) is an important preventable cause of renal failure in pediatric urology. It may be present antenatally and an early detection is crucial to contain the progressive renal damage it can inflict in the early years of life. Voiding Cystourethrography (VCUG), and more recently Direct Radionuclide Cystography (DRCG) are standard tests used to detect and grade VUR. These tests have distinct advantages and disadvantages.

Aims & Objectives: The aims of this study were 1. To compare the efficacy of DRCG in detecting VUR as compared to VCUG (VCUG being taken as the reference standard). 2. To propose a viable grading of VUR on the DRCG.

Methods: A prospective, cross-sectional study that enrolled 54 children was performed. Inclusion criteria were: 1. Children less than 12 years being evaluated for VUR, 2. a sterile urine culture at the time of the tests and 3. Both tests performed within 24 hours of each other (same day study). Demographic data, presenting complaint, provisional diagnosis and presence & grade of reflux on both the tests were documented. Specific statistical parameters in detecting and grading VUR with a DRCG were calculated.

Results: DRCG detected 22 of the 32 renal units that had VUR on the VCUG. It had a sensitivity of 69%, specificity of 100%, positive predictive value of 100%, negative predictive value of 88%, accuracy 91% and Coefficient of Agreement -Kappa of 0.76 (highly significant) in detecting VUR as compared to VCUG. Low grade (IRSC I, II) VUR was likely to be missed on DRCG. Whenever DRCG demonstrated a VUR, the VCUG also showed it. Despite minor differences, we note considerable agreement between the two systems (accuracy - 77.78%, coefficient of agreement with VCUG - Kappa of 0.48, significant) in the grading of VUR.

Conclusions: DRCG compared modestly with VCUG in detecting VUR. A voiding study is suggested to improve the sensitivity. A practical three tier grading of VUR on DRCG with a superior statistical correlation (accuracy -90.76%, coefficient of agreement with VCUG - Kappa of 0.773, highly significant) with the IRSC grading on the VCUG is proposed: A – radioactivity in the ureter (corresponding to IRSC grade I), B – radioactivity in the undilated /minimally dilated pelvicalyceal system, (corresponding to IRSC grades II / III), C – radioactivity in the dilated pelvicalyceal system (corresponding to IRSC grades IV / V). Of these, A and B are Low Grade while C is High Grade VUR.

Keywords: Scintigraphy, pediatric, imaging, grading, reflux

Introduction
Vesico ureteric reflux (VUR) is defined as the abnormal retrograde flow of urine from the urinary bladder into the ureter across the vesicoureteric junction (VUJ). It is a common urologic problem in the pediatric population. The normal VUJ is akin to a valve that allows unidirectional flow from the ureter into the urinary bladder. This is primarily attributed to a submucous oblique tunnel, the ureter traverses as it enters the urinary bladder. VUR is the second most common functional cause of obstruction in the urinary tract (after pelviureteric junction obstruction) and is common in infancy and early childhood [1]. It has a detrimental effect on the growth and development of the kidney in particular, and the somatic growth in general. Most children with VUR are known to be in the lower weight percentile groups [2]. Also, studies show that the renal parenchyma is most vulnerable to infection in the first three years of life [3]. Untreated VUR predisposes to urinary tract infections that may progress into the spectrum of reflux nephropathy and renal failure. Up to 50% of infants and children with UTI are known to have VUR [4] and hence any young child presenting with UTI is screened for VUR. When detected early, VUR is eminently manageable to prevent the potentially deleterious effects on the kidney.
Therefore, considerable emphasis is laid on an early diagnosis of VUR.

Three imaging modalities have been used to detect VUR - Voiding Cystourethrogramy, Radionuclide Cystography (Direct and Indirect) and Voiding Cystosonography. Voiding Cystourethrogramy is the conventional and most accepted method and is generally considered the reference standard for diagnosis [3]. It provides good anatomic delineation that is particularly important in imaging associated structural anomalies (e.g. posterior urethral valves) and provides for an accurate grading. Its main drawbacks include bladder catheterization and radiation exposure.

Radionuclide Cystography (Direct and Indirect) are relatively newer tests that have minimal radiation exposure but inferior anatomic delineation. Voiding Cystosonography, a recent extended application of conventional ultrasonography does not involve radiation exposure and is useful both antenatally and postnatally to screen for indirect signs of VUR. It is neither very sensitive nor specific in detecting VUR.

An accurate grading of VUR is crucial, as the higher grades are associated with progressive renal scarring - the principal factor governing the management of VUR [6]. Currently the International Reflux Study Committee criteria are universally adopted for the grading of VUR on the VCUG. Renal scarring is best detected on a dimercaptosuccinate (DMSA) scan [7-9].

The management of VUR depends on the etiology and grade of VUR, whether unilateral or bilateral, age of child at the time of detection and most importantly whether renal scarring is present or not [10]. A conservative approach is adopted for the lower grades (I – III) of primary VUR and selected cases of secondary VUR and involves long term chemoprophylaxis and surveillance. A surgical approach implies ureteric reimplantation and is adopted in those with persistent VUR, repeated bouts of breakthrough infections and VUR secondary to selected structural & functional anomalies of the urinary tract [11-13]. The VUR in this category is usually of high grade (IV – V). Newer minimally invasive methods like cystoscopic sub ureteral injection of bulking agents (e.g. collagen) beneath the reflushing orifice are also in vogue with encouraging results. Both the conservative and surgical management options, have had good success rates in arresting scarring [14, 15].

The present study compares DRCG with VCUG, two standard tests, in the detection and grading of VUR. During the early recruitment into the study, there was no internationally accepted practical grading for VUR on the DRCG and this study aimed to propose a viable grading for the same [16].

Materials and Methods
The study was conducted at the Department of Radiology and Pediatric Surgery at the Azeezia medical college and Research Centre, Kollam between July 2018 and December 2019. The study was approved by the Institutional Ethical Review Board.

A total of 54 children (108 renal units) were studied. The inclusion criteria were:

-  N<12 years being evaluated for VUR
-  A sterile urine culture at the time of VCUG
-  Both tests, the VCUG and DRCG performed within 24 hours of each other (same day study)

Voiding Cystourethrogramy
This was usually the first of the two tests and conducted at the at the Department of Radiology, Azeezia Medical College and Research Centre, Kollam in the following manner -

-  A parent or familiar adult was allowed to accompany the patient throughout the procedure to reassure the child
-  The child was administered a prophylactic antibiotic (Inj. Gentamycin @ 5mg/kg body weight) half an hour prior to the test.
-  Under aseptic precautions and local anesthetic gel, the bladder was catheterized with an Infant feeding tube of size 5 – 7.
-  The child was placed on an X-ray table having under couch tube and fluoroscopic facility. Intermittent fluoroscopy was employed to limit the radiation involved.
-  Diluted contrast medium (Urograffin 76% and i.v. physiological saline in the ratio 1: 4) at room temperature was instilled under fluoroscopy. Bladder capacity was assumed to be approximately 10 ml/kg body weight unless previously estimated to be otherwise and instillation stopped with pericatheter leak of contrast.
-  The older child was asked to void when he / she felt a strong urge while in children below 2 years and in those with known neurogenic bladder states, the bladder was filled till the child voided around the catheter
-  Voiding films were taken in all. Full bladder / post void images were taken where necessary
-  Oblique films were taken whenever possible and in males the posterior urethra was carefully imaged.

-  Where VUR was present, it was graded according to the International Reflux Study Committee (IRSC) classification. The VCUG was reported on the same day of the test and the patient reported for the DRCG.

The result of the VCUG was withheld from the consultant performing the DRCG and vice versa till documentation & interpretation.

Direct Radio-nuclide Cystography
Bladder filling was natural and the older children were instructed to indicate when they have a strong urge to void. About 100 – 500 micro Curie of 99mTcDTPA was injected percutaneously into a full bladder by a suprapubic puncture under ultrasound guidance.

-  The child was positioned with his / her back facing the Gamma camera and a full bladder image was obtained.
-  In the co-operative children, continuous voiding images were obtained with a maximum exposure time of 100 seconds. All the images were accessible in real time through the computer to acquire adequate exposures for documentation & interpretation.
-  Multiple voiding images were obtained in patients where the initial images were negative prior to terminating the study.
-  In infants and younger children where, voiding exposures were not possible, immediate post void images were obtained.

The grading of VUR was as per an Institutional Grading (IG) protocol based on the distribution and configuration of the radionuclide activity in the urinary tract.
The description in Table 1 below and the illustrative figures (Figure 1 a, b, c) thereafter exemplify the criteria adopted.

<table>
<thead>
<tr>
<th>INSTITUTIONAL GRADE (IG)</th>
<th>CRANIAL DISTRIBUTION OF RADIOACTIVITY AND DELINEATION OF URINARY TRACT ANATOMY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lower ureter</td>
</tr>
<tr>
<td>2</td>
<td>Upper ureter</td>
</tr>
<tr>
<td>3</td>
<td>Undilated pelvicalyceal system</td>
</tr>
<tr>
<td>4</td>
<td>Mild dilatation of the pelvicalyceal system</td>
</tr>
<tr>
<td>5</td>
<td>Gross dilatation of the pelvicalyceal system</td>
</tr>
</tbody>
</table>

Fig 1: a) DRCG with right grade 4 VUR, b) DRCG with right grade 3 VUR, c) DRCG with bilateral grade 5 reflux according to the IG protocol (Note - the side is as visualized from behind the patient)

The demographic details, presenting complaint, provisional diagnosis and the test results (presence/absence and grading of VUR in individual renal units on VCUG and DRCG) were entered into a proforma and tabulated into a master chart.

The data thus obtained were statistically analyzed to yield the following:
- Number of renal units with VUR in each test
- Sensitivity (True positives / True + False positives X 100), and Specificity (True negatives / True + False negatives X 100) of DRCG as compared to VCUG in detecting VUR
- Positive predictive value (True positives / True positives + False negatives X 100) and Negative predictive value (True negatives / True negatives + False negatives X 100) of DRCG as compared to VCUG in detecting VUR
- Observed agreement or Accuracy (True positives + True negatives / Total number of units X 100) of DRCG versus VCUG in detecting VUR
- Coefficient of agreement - Kappa between DRCG and VCUG in detecting VUR
- Distribution and differences in grading of VUR between the tests
- Observed agreement or Accuracy of DRCG versus VCUG in grading VUR
- Coefficient of agreement - Kappa between DRCG and VCUG in grading of VUR

Results
54 children (n = 108 renal units) were enrolled for the tests.

Demographic Data
The mean age was 4.5 years (range: 1 month - 10 years). There were 43 males and 11 females (M: F: 4:1). The distribution of presenting complaints / provisional diagnosis is tabulated below (Table 2)
Table 2: The distribution of presenting complaints / provisional diagnosis

<table>
<thead>
<tr>
<th>PRESENTING COMPLAINTS / PROVISIONAL DIAGNOSES</th>
<th>NO. OF PATIENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptomatic UTI</td>
<td>25</td>
</tr>
<tr>
<td>Antenatally detected hydronephrosis / asymptomatic</td>
<td>15</td>
</tr>
<tr>
<td>Secondary VUR eg. Neurogenic bladder</td>
<td>5</td>
</tr>
<tr>
<td>Known cases of VUR on follow up</td>
<td>5</td>
</tr>
<tr>
<td>Intravesical obstructive uropathy</td>
<td>4</td>
</tr>
</tbody>
</table>

Detection of VUR on the VCUG
Considering VCUG as the Gold Standard, VUR was detected in a total of 32 / 108 renal units (35%) in 21 patients. Of these 15 were males and 6 females; the VUR was bilateral in 11(22 renal units) and unilateral in 10 (10 renal units) patients. Males predominated in unilateral (7 / 10 patients) and bilateral (8 / 11 patients) reflux reflecting the overall sex distribution of cases.

Detection of VUR on the DRCG
VUR was detected in a total of 22 /108 renal units (20.1%) in 15 patients, 10 males and 5 females. Males predominated in unilateral (5 / 8 patients) and bilateral (5 / 7 patients) reflux. Figure 2 illustrates the number of renal units with and without VUR in both the tests.

DRCG did not detect VUR in 10 of 32 renal units that showed VUR on the VCUG. Of these, 3 were of Grade I, 1 of Grade II and 6 of Grade III IRSC grades. There were none of IRSC grade IV / V that were missed by the DRCG. When DRCG was compared to VCUG in the detection of VUR, the following were the specific statistical correlates which can be calculated from the following Table 3.

Table 3: Showing the distribution of renal units with respect to detection of VUR

<table>
<thead>
<tr>
<th></th>
<th>VCUG</th>
<th>DRCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE POSITIVES</td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>FALSE NEGATIVES</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>FALSE POSITIVES</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>TRUE NEGATIVES</td>
<td>76</td>
<td>-</td>
</tr>
</tbody>
</table>

Sensitivity (22 / 32 X 100) - 69%
Specificity (76 / 76 X 100) - 100%
Positive Predictive Value (22 / 22 X 100) - 100%
Negative Predictive Value (76 / 86 X 100) - 88%
Accuracy (96/108 X 100) - 91%
Coefficient of Agreement (Kappa) - 0.76
(with a one tailed p < 0.001, significant)

Grading of VUR on the DRCG
Figure 3. summarizes the distribution of VUR grades on the VCUG (IRSC grade) and the DRCG (Institutional grade) for the renal units detected to have VUR on the respective tests.

Fig 2: Detection of VUR in renal units.

Fig 3: shows the distribution of the grades of VUR, in the 32 renal units, which were detected on VCUG and DRCG.
The distribution of reflux grades in the 22 units with VUR documented on both tests is indicated in figure 4.

Fig 4: Shows the distribution of the grades of VUR, in the 22 renal units which were detected on both VCUG and DRCG.

Overall, the grading of VUR on the DRCG was comparable to that on the VCUG. If Grade 0 was taken into account then identical grades were seen in 76 out of the 108 renal units, giving a total of 84 renal units, which had exact grading in both tests. DRCG failed to detect VUR in 10 out of 32 renal units that showed VUR in VCUG. Of these 3 were Grade I, 1 of Grade II and 6 of Grade III. Thus the higher grades of VUR (IRSC grades IV and V on VCUG) were always picked up on DRCG while the lower grades (IRSC grades I – III) were occasionally missed.

Of the 22 renal units that manifest VUR on both tests, identical grades of VUR were reported in 8. It differed in 14 of 22 renal units, in all the difference was a single grade. The higher grade was usually on the VCUG except for 2 renal units a higher grade was seen on the DRCG.

Table 4: The distribution of renal units, the presence/absence of VUR, and the grades assigned to them in the two tests in the form of a grid
The accuracy of grading on DRCG as compared to VCUG was 78% and, the coefficient of agreement Kappa was 0.48 (with a one tailed p value of 0.00001, highly significant).

Discussion
VUR is a common problem in the pediatric age group. It predisposes children to recurrent UTI and its sequel reflux nephropathy is a leading cause of renal failure. Maximal renal injury occurs during the first three years of life. Early detection and optimal management of VUR, medically or surgically, arrests this process.

Any infant or young child presenting with a UTI should be investigated to rule out VUR. The initial screening test is usually an ultrasound that relies heavily on the indirect signs of VUR like hydronephrosis, hydrourtereonephrosis or a lower hydroureter to suggest the diagnosis. Since many years, antenatal ultrasonography has facilitated an early detection of VUR. The VCUG is the investigation of choice to confirm and characterize VUR. Of late, a DRCG is being adopted as an alternative or complementary investigative tool in the armamentarium for the diagnosis of VUR.

VCUG is considered the reference gold standard against which newer tests are compared because most available data over decades have conceded that it is highly sensitive in the detection of VUR. Some studies claim comparable or better sensitivity with a DRCG [5]. The comparison between VCUG and DRCG has been marred by the intermittent nature of low grade VUR, performance of the tests at different time periods, lack of an equivalent grading system on the DRCG and the intrinsic differences between the tests. Nevertheless, few studies have compared both the tests in the detection of VUR. Dikshit et al. [22] found a sensitivity of 95% and specificity of 95.8% for DRCG as compared to VCUG. They note that the lower grades of VUR may be missed on DRCG and the VCUG may occasionally fail to detect higher grades. In a study of 105 refluxing units, McLaren et al. [30] showed DRCG to be more sensitive (91% vs. 45%) in detecting VUR in younger infants, irrespective of the grade. Of course, the few VUR missed on the DRCG were mostly low grade. Hence higher grades of VUR were better detected on DRCG as compared to VCUG. The better sensitivity of DRCG is often ascribed to continuous real time observation over a relatively longer period than the VCUG.

Polito et al. [37] also found DRCG to be more accurate than fluoroscopic cystourethrography, when all the grades were considered together (p=0.00001) or when severe VUR was considered alone (p=0.004). Other authors like Fretzayas et al. [53] and Chapman et al. [63] report that VUR is detected equally well by DRCG and VCUG.

This study reveals a modest sensitivity (69%) and negative predictive value (88%) of the DRCG in detecting VUR; the specificity (100%) and Positive Predictive value (100%), accuracy (90.7%) and Coefficient of Agreement (0.76, one tailed p = 0.000000) were excellent. A recent study by Jose et al. [64] showed similar figures (sensitivity - 75.86%, specificity and positive predictive value of 100% and negative predictive value of 89%).

Like most other authors [30, 37, 54, 60], we note that DRCG may occasionally fail to detect the lower grades (IRSC I to III) reflux rather than the higher grades. It has been suggested that the intermittent nature of low grade reflux and the overlapping bladder radioactivity may contribute to this loss of sensitivity. However similar factors including minimal reflux of unopacified fluid and variation of VUR on consecutive VCUGs [65] may also result in the lower grades being missed on the VCUG.

On the other hand, a failure to detect IRSC grade III VUR on the DRCG in 6/32 renal units with reflux in this study is unacceptable. Reliance on immediate post void studies obtained where voiding studies were not possible could account for the lack of detection in a few instances but other reasons remain elusive. Such deficiencies suggest the need for further standardization of test methodology.

Differences in grading of VUR in both the tests are described. The observed accuracy in grading of VUR on the DRCG (77.8%) and the coefficient of agreement in grading (Kappa) 0.48 between the DRCG and VCUG were satisfactory. However we reiterate that the superior anatomic detail with a VCUG allows for the IRSC grading while a similar anatomical Institutional Grading adopted in this study for the DRCG is relatively less exacting. It is possible to differentiate radionuclide activity in the urinary tract in two places – the ureter and pelvicalyceal system, and segregate mild dilatation from gross dilatation of the urinary tract.

Based on our observations, we would like to propose the following grading for VUR on the DRCG (Table 5)
When the 22 common refluxing units were reassigned grades as per the suggested three tier system, there was near total agreement on the grading between the two tests.

### Table 6: the grading between the two tests

<table>
<thead>
<tr>
<th>Grade</th>
<th>DRCG</th>
<th>IRSC GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n-22</td>
<td>n-32</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>III</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>IV</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>V</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>

### Table 7: The distribution of all the renal units, the presence/absence of VUR, and the grades reassigned to them in the two tests according to the proposed three tier grading system.

Table 7 depicts a grid similar to Table 5 featuring the distribution of all the renal units, the presence/absence of VUR, and the grades reassigned to them in the two tests according to the proposed three tier grading system. The observed accuracy in grading of VUR on the DRCG (90.76%) and the coefficient of agreement in grading (Kappa = 0.773, one tailed p value of <0.001, significant) between the DRCG and VCUG were excellent.

### Detection of VUR

DRCG compared modestly with VCUG in the detection of VUR with a sensitivity of 69% and negative predictive value of 88%. The specificity (100%), positive predictive value (100%) and accuracy (90.74%) were excellent while coefficient of agreement with VCUG (Kappa – 0.76) was highly significant. Low grade VUR (IRSC I-III) was occasionally likely to be missed on DRCG.

### Grading of VUR

Despite minor differences in grading of VUR between the Institutional Grading of DRCG and IRSC grading of VCUG, we note considerable agreement between the two systems (accuracy = 77.78%, coefficient of agreement with VCUG - Kappa of 0.48, significant).

### Proposed Grading

A practical three tier system of grading on the DRCG with superior statistical correlation (accuracy -90.76%,
coefficient of agreement with VCUG - Kappa of 0.773, highly significant), with the IRSC grading on the VCUG is proposed.

Summary
Vesico ureteric reflux (VUR) or the abnormal retrograde flow of urine from the urinary bladder into the ureter across the vesico ureteric junction may be present antenatally and an early detection is crucial to contain the progressive renal parenchymal injury it can inflict in the vulnerable early years of life. An important preventable cause of renal failure in children, untreated VUR predisposes to urinary tract infections that may progress into the spectrum of reflux nephropathy. Besides other factors, the grade of VUR and extent of renal scarring dictate the plan of management, hence the importance of accurate imaging. Voiding Cystourethrogram (VCUG), and more recently Direct Radionuclide Cystography (DRCG) are standard tests used to detect and grade vesico ureteric reflux. Voiding Cystourethrogram is the conventional and most accepted method and is generally considered the reference standard for diagnosis. DRCG is a relatively newer test with minimal radiation exposure but inferior anatomic delineation.

The present study compared DRCG with VCUG in the detection and grading of VUR and adds to the scant literature on this subject. This was a prospective, cross-sectional, same day study that enrolled 54 children (108 renal units) being investigated for VUR. DRCG detected 22 of the 32 renal units that had VUR on the VCUG (sensitivity of 69%, specificity of 100%, positive predictive value of 100%, negative predictive value of 88%, accuracy 91% and Coefficient of Agreement - Kappa of 0.76, highly significant) in detecting VUR as compared to VCUG. Low Grade (Grade I, II) VUR was occasionally missed on the DRCG. Despite minor differences, there is considerable agreement between the two systems (accuracy - 77.78%, coefficient of agreement with VCUG - Kappa of 0.48, significant) in the grading of VUR.

DRCG compared modestly with VCUG in detecting VUR. A voiding study is suggested to improve the sensitivity.

The proposed practical three tier grading of VUR on the DRCG with a superior statistical correlation with the IRSC grading on the VCUG (accuracy -90.76%, coefficient of agreement with VCUG - Kappa of 0.773, one tailed p value of 0.0000000, highly significant), is as follows: A – IRSC grades I / II, B – radioactivity in the undilated /minimally dilated pelvicalyceal system, (corresponding to IRSC grades II / III), C – radioactivity in the dilated pelvicalyceal system (corresponding to IRSC grades IV / V).

References