



Application of the postnatal urinary tract dilation classification system to predict the need for surgical intervention among neonates and young infants

ULTRA
SONO
GRAPHY

Jisun Hwang^{1*}, Pyeong Hwa Kim^{2*}, Hee Mang Yoon², Sang Hoon Song³, Ah Young Jung², Jin Seong Lee², Young Ah Cho²

¹Department of Radiology, Hallym University Dongtan Sacred Heart Hospital, Hwaseong;

²Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul; ³Department of Urology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Purpose: The aim of this study was to validate the postnatal urinary tract dilation (UTD) classification system by correlating it with the need for surgical intervention.

Methods: Young infants who underwent ultrasound (US) examinations for prenatal hydronephrosis were retrospectively identified. The kidney units (KUs; right, left, or bilateral) were graded from UTD P0 (very low risk) to P3 (high risk) based on seven US criteria from the UTD system. Surgery-free survival curves were constructed using the Kaplan-Meier method. Univariable and multivariable Cox proportional-hazards regression analysis clustered by patients was performed. Interobserver agreement was analyzed using the weighted kappa coefficient.

Results: In total, 504 KUs from 336 patients (mean age, 18.3±15.9 days; range, 1 to 94 days; males, n=276) were included, with a median follow-up of 24.2 months. Fifty-eight KUs underwent surgical intervention. Significant differences were observed among the Kaplan-Meier curves stratified into UTD groups ($P<0.001$). The presence of anterior-posterior renal pelvic diameter ≥ 15 mm (hazard ratio [HR], 8.602; 95% confidence interval [CI], 1.558 to 43.065), peripheral calyceal dilation (HR, 8.190; 95% CI, 1.558 to 43.065), ureteral dilation (HR, 2.619; 95% CI, 1.274 to 5.380), parenchymal thickness abnormality (HR, 3.371; 95% CI, 1.574 to 7.223), bladder abnormality (HR, 12.209; 95% CI, 3.616 to 41.225) were significantly associated with the occurrence of surgery. The interobserver agreement was moderate to almost perfect agreement for US features ($\kappa=0.564-0.898$) and substantial for final UTD grades ($\kappa=0.716$).

Conclusion: The UTD classification system is reliable and appropriately stratifies the risk of surgical intervention.

Keywords: Hydronephrosis; Ultrasonography; Infant; Prognosis; Reproducibility of results

Key points: The postnatal urinary tract dilation (UTD) classification system is reliable and appropriately stratifies the risk of surgical intervention. The surgery-free survival rates were 100% (UTD P0), 100% (UTD P1), 86.7% (UTD P2), and 30.9% (UTD P3) at 1 year. An anterior-posterior renal pelvic diameter ≥ 15 mm, peripheral calyceal dilation, ureteral dilation, parenchymal thickness abnormality, and bladder abnormality were significantly associated with the occurrence of surgery.

ORIGINAL ARTICLE

<https://doi.org/10.14366/usg.22035>

eISSN: 2288-5943

Ultrasonography 2023;42:136-146

Received: March 2, 2022

Revised: June 9, 2022

Accepted: July 23, 2022

Correspondence to:

Hee Mang Yoon, MD, PhD, Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea

Tel. +82-2-3010-0906

Fax. +82-2-476-4719

E-mail: espoirhm@gmail.com,

hmyoon@amc.seoul.kr

*These authors contributed equally to this work.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2023 Korean Society of Ultrasound in Medicine (KSUM)



How to cite this article:

Hwang J, Kim PH, Yoon HM, Song SH, Jung AY, Lee JS, et al. Application of the postnatal urinary tract dilation classification system to predict the need for surgical intervention among neonates and young infants. Ultrasonography. 2023 Jan;42(1):136-146.

Introduction

Dilation of the renal pelvis is a common abnormality in fetal kidneys, being present in approximately 1% of pregnancies [1]. Although transient and physiologic urinary tract dilation (UTD) accounts for more than half of the cases, various congenital anomalies of the kidney and urinary tract (CAKUT) can be associated with perinatal hydronephrosis [2], and a postnatal follow-up ultrasound (US) evaluation is recommended.

In the past, the reporting methods for postnatal hydronephrosis were considerably diverse. According to a survey by the Society for Pediatric Radiology in 2014, about two-thirds of responders used the mild-moderate-severe system, and the remaining one-third used either the Society for Fetal Urology (SFU) system or the anterior-posterior renal pelvic diameter (APRPD) for grading UTD [3]. Regarding the SFU system, several studies have reported interobserver variability for distinguishing between SFU grade 1 versus 2 or SFU grade 3 versus grade 4 [4,5]. Finally, a new multidisciplinary consensus on the classification of prenatal and posterior UTD (the UTD classification system) was proposed in 2014 to establish a unified system using standard terminology that can correlate clinical outcomes and suggest management strategies based on risk stratification [6]. The most notable characteristic of the UTD classification system is the introduction of combinations of US elements, which were not included in the SFU system and other grading systems; therefore, it better reflects the structural changes in various types of CAKUT. It also distinguishes between central and peripheral calyceal dilation.

Several recent studies have assessed the ability of the UTD system to predict surgical interventions; however, those studies focused on ureteropelvic junction obstruction–like hydronephrosis and did not evaluate all of the UTD system's US elements [7,8]. Other studies have shown that the UTD system yielded interobserver agreement data similar to or higher than those of the SFU system [9,10]. One study reported moderate interobserver agreement for the UTD system, with the highest disagreement for calyceal dilation and bladder status, emphasizing that further efforts to improve the UTD classification are warranted [10]. According to a recent survey, many pediatric hydronephrosis specialists and most imaging providers still prefer the SFU system, APRPD, or the mild-moderate-severe system due to their simplicity and familiarity [11]. This study aimed to validate the postnatal UTD classification system by correlating it with the need for surgical intervention. The prognostic value of each US element used in the system was analyzed, as well as the interobserver agreement of the overall UTD grades and each US element using definitions more detailed than those used for the original UTD system and previous studies.

Materials and Methods

Compliance with Ethical Standards

This study was approved by the relevant institutional review board (approval number 2020-0826), and the requirement for informed consent was waived.

Patient Selection and Study Design

This retrospective, observational study was conducted according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [12].

Consecutive patient data were collected at a single tertiary hospital between 2010 and 2017. The inclusion criteria were as follows: (1) infants younger than 3 months and (2) initial postnatal kidney US conducted for prenatal hydronephrosis. The exclusion criteria were as follows: (1) follow-up US data not available for at least 30 days (n=116); (2) kidney US within 48 hours after birth, which might lead to underestimation of hydronephrosis, except if emergency US was performed because of a high suspicion of CAKUT from prenatal US (n=13); (3) normal US without measurable pelvic dilation (n=5); and (4) infants that previously underwent an *in utero* intervention (n=3). In duplex kidneys, either the upper or lower moiety that showed a higher grade of hydronephrosis was included. In patients with posterior urethral valves, one kidney of each pair was included for analysis. Finally, 504 kidney units (KUs) from 336 patients were included in this study.

US Imaging and Analysis

Over 8 years, US examinations were performed by multiple radiologists, including radiology trainees, pediatric radiology fellows, and faculty pediatric radiologists, using iU-22 equipment (Philips Healthcare, Bothell, WA, USA). The kidney US examinations conducted by radiology trainees were supervised by faculty pediatric radiologists. A 5–8 MHz sector transducer was primarily used for evaluation, and a 5–12 MHz linear transducer was additionally used for detailed assessment.

Two board-certified pediatric radiologists (P.H.K. and J.H.; 5 and 8 years of experience in radiology, respectively) independently reviewed the US images and were blinded to any clinical data and the original US report. The reviewers recorded the following imaging features based on previous research on the postnatal UTD system and pediatric urological terms (Table 1) [6,13]: the laterality of the involved KU (right, left, bilateral), APRPD, central calyceal dilation, peripheral calyceal dilation, parenchymal appearance abnormalities (altered echogenicity, cortical cysts, decreased corticomedullary differentiation), parenchymal thinning, ureteral dilation, and bladder abnormalities (wall thickening, ureterocele,

Table 1. Definitions of ultrasound features of the urinary tract dilation system

| Urinary tract dilation system Item | Definition |
|--|--|
| Anterior-posterior renal pelvic diameter | Measured on transverse images at the maximal diameter of the intrarenal pelvis |
| Central calyceal dilation | Dilation of the central calyces of the kidney |
| Peripheral calyceal dilation | Dilation of the peripheral calyces of the kidney |
| Parenchymal appearance abnormalities | |
| Altered echogenicity | Increased echogenicity subjectively assessed by comparison with the adjacent liver or spleen |
| Cortical cysts | Presence of cortical cysts as a result of renal dysplasia |
| Decreased corticomedullary differentiation | Loss of normal corticomedullary differentiation |
| Parenchymal thickness abnormality | The medulla is not seen, or parenchymal thickness is less than one-fourth of those in contralateral kidneys. In case of bilateral involvement, less than one-fourth of the reference medullary thickness. |
| Ureteral dilation | Entire or almost entire dilation. Transient visualization of the ureter is considered normal postnatally and excluded. |
| Bladder abnormalities | |
| Wall thickening | Thickening of bladder wall and/or trabeculated bladder with irregular mucosa. Wall thickness >3 mm in a well-distended bladder and >5 mm in an under-distended bladder. Thickness is measured in the wall away from the trigone. |
| Ureterocele | Cystic structure within the posterior wall of the bladder |
| Dilated posterior urethra | Abnormally dilated posterior urethra |

dilated posterior urethra). A baseline consensus on the items used in the UTD system was reached through discussion of 25 cases before the imaging analysis. Any discrepancies between the reviewers were resolved by the consensus reading of another board-certified radiologist (blinded; 12 years of experience in radiology).

In previous studies [6,13], parenchymal thinning was subjectively assessed, whereas the present study defined parenchymal thinning as being present when the medulla was not seen or when parenchymal thickness was less than one-quarter of that in the contralateral kidney. In cases of bilateral renal involvement, thinning was considered to be present when the medulla was less than one-quarter of the reported reference medullary thickness in 1-month-old infants (6.06±1.24 mm in the right kidney and 6.07±1.33 mm in the left kidney) [14]. Thickening of the bladder wall was defined as a thickness >3 mm in a well-distended bladder and >5 mm in an under-distended bladder [15]. Thickness measurements were performed in the wall distal to the trigone [15]. The UTD system classifies the kidneys into three groups: UTD P1 (low risk), UTD P2 (intermediate risk), and UTD P3 (high risk for postnatal uropathies) based on the US features [6]. In this system, APRPDs from 10 to 15 mm and ≥15 mm are considered to be associated with low and intermediate risk, respectively. Then, if the additional US features for intermediate risk (peripheral calyceal dilation, ureteral dilation) or high risk (parenchymal appearance abnormalities, parenchymal thinning, bladder abnormalities) are present, the UTD groups are upgraded regardless of the APRPD. The presence of an APRPD <10 mm and no other abnormality (no calyceal or ureteral dilation,

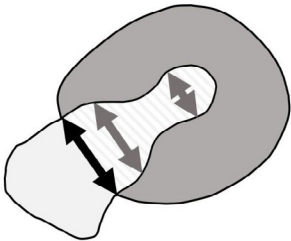
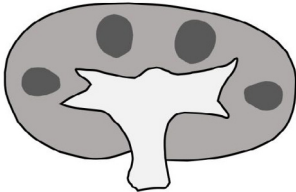
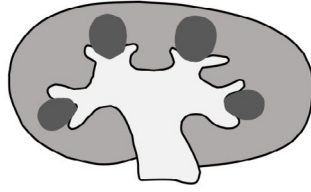
normal renal parenchyma, normal bladder) is defined as normal in the UTD system. This "normal category" was referred to as "P0," following previous studies (Fig. 1) [10,16].

Clinical Data

Clinical data, including age, sex, and surgical intervention, were collected. Voiding cystourethrography (VCUG) and technetium-99m mercaptoacetyl triglycine (MAG3) renal scanning were carried out at the clinicians' discretion. The presence of vesicoureteral reflux (VUR), a significant obstruction ($T_{1/2} >20$ min and a rising uptake curve with no response to diuretics on a renogram), and a differential renal function (DRF) value <40% in the hydronephrotic kidney were recorded for the children with available studies [17]. The DRF is estimated based on the number of counts generated from each kidney during the uptake phase after radiotracer injection, and a DRF <40% is commonly utilized as a cut-off value to guide the management of children with hydronephrosis [18,19]. To evaluate VUR, VCUG studies that were performed within 3 months of kidney US were retrieved. The MAG3 scan results with the highest impairment were recorded if the patients underwent multiple studies during follow-up.

Surgical intervention was carried out for patients with symptomatic hydronephrosis (such as abdominal pain, recurrent urinary tract infection, or gross hematuria), continuous or increasing high-grade hydronephrosis on repeated US, persistent high-grade VUR, DRF <40% or decline of DRF >5% between repeated MAG3 scans, and delayed renal tissue transit time on a MAG3 scan [20].

| | APRPD | Calyceal dilation | Ureter dilation | Parenchyma and/or bladder abnormality |
|----|----------|-------------------|-----------------|---------------------------------------|
| P0 | <10 mm | None | None | None |
| P1 | 10–15 mm | Central | None | None |
| P2 | ≥15 mm | Peripheral | Yes | None |
| P3 | ≥10 mm* | Any dilation* | Any | Yes |

| APRPD | Central calyceal dilation | Peripheral calyceal dilation |
|---|---|---|
|  <p>Measured on transverse image at the maximal diameter (black line) of intrarenal pelvis (shaded region)</p> |  <p>Small protrusions at the margins of the renal pelvis</p> |  <p>Cupping the medullary pyramids</p> |

Definitions newly added to the original UTD classification system

- Ureter dilation: entire or almost entire dilation. transient visualization of the distal ureter (<4 mm) is excluded
- Parenchymal thinning: absent medulla, medulla thickness < ¼ of contralateral kidney
- Bladder wall thickening: >3 mm in distended, >5 mm in under-distended bladder, measured away from the trigone

Fig. 1. Urinary tract dilation (UTD) risk stratification. The kidneys were classified into four groups: UTD P0 (normal), P1 (low risk), P2 (intermediate risk), and P3 (high risk for postnatal uropathies). First, the hydronephrotic kidneys were separated by anterior-posterior renal pelvic diameter (APRPD), and then the stratification could be upgraded depending on the most concerning ultrasound features of calyceal dilation, ureteral dilation, renal parenchyma, and bladder abnormality. The criteria in the shaded box must be met in each category. Renal collecting system should be abnormally dilated (i.e., either abnormal APRPD or calyceal dilation should be present) to grade P3 (asterisk).

Statistical Analysis

Continuous variables are expressed as means with standard deviations or medians with ranges. Categorical variables are presented as proportions. Intergroup comparisons of categorical variables were assessed using the Pearson chi-square test as appropriate.

Surgery-free survival was defined as the time from enrollment (the date of an initial postnatal US) until either the operation date or the date that the patient was last known to have not yet undergone an operation (the date of the last follow-up US). Survival curves were constructed using the Kaplan-Meier method. Univariable Cox proportional hazards regression analysis clustered by patients was performed to compare the occurrence of surgical intervention among the UTD groups (UTD P0–P3). C-statistics with 95% confidence intervals (CIs) were also calculated to estimate the discriminative performance of the UTD system using the Z test. To minimize overestimation, optimism-corrected C-statistics were calculated by subtracting optimism from the original values using the bootstrap method. C-statistics were regarded as follows: >0.8 indicated excellent discrimination, 0.7–0.8 indicated good discrimination, 0.6–0.7 indicated some clinical value, and <0.6 indicated no clinical

value [21].

Additionally, univariable and multivariable Cox proportional hazards regression analysis clustered by patients was performed to determine the predictors of surgical intervention among the US features of the UTD system. Variables with P<0.1 in the univariate analysis were incorporated into the multivariable analysis. In the multivariable analysis, the backward elimination method was used only for factors with P<0.05. Interobserver agreement for US features and UTD classification were analyzed using weighted kappa coefficients. The kappa values were interpreted as follows: moderate agreement, 0.41–0.60; substantial agreement, 0.61–0.80; and almost perfect agreement, ≥0.81 [22].

Statistical analyses were conducted with SPSS for Windows version 20.0 (IBM Corp., Armonk, NY, USA), MedCalc version 20.009 (MedCalc Software, Ostend, Belgium), and R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). A P-value <0.05 was considered to indicate statistical significance.

Table 2. Patient characteristics

| Variable | Value |
|-----------------------------------|------------|
| Sex | |
| Male | 276 (82.1) |
| Female | 60 (17.9) |
| Age (day) | 18.3±15.9 |
| Side | |
| Right | 29 (8.6) |
| Left | 135 (40.2) |
| Both | 172 (51.2) |
| UTD grade | |
| P0 | 101 (20.0) |
| P1 | 201 (39.9) |
| P2 | 151 (30.0) |
| P3 | 51 (10.1) |
| CAKUT | |
| VUR | 8 |
| Dysplastic kidney | 4 |
| Duplex kidney | 11 |
| Duplex kidney and UPJ obstruction | 1 |
| UPJ obstruction | 48 |
| UVJ obstruction | 6 |
| Primary megaureter | 2 |
| PUV | 4 |
| Surgical intervention | |
| Ureterocelectomy | 3 |
| Ureterocele puncture | 1 |
| Partial nephrectomy | 2 |
| Simple nephrectomy | 1 |
| Pyeloplasty | 37 |
| Reimplantation | 7 |
| Ureteroneocystostomy | 3 |
| Valve ablation | 4 |

Values are presented as number (%) or mean±SD.

Data for sex, patient age, and sidedness are per person. Data for UTD grade, CAKUT, and surgical intervention are per kidney unit.

UTD, urinary tract dilatation; CAKUT, congenital anomalies of the kidney and urinary tract; VUR, vesicoureteral reflux; UPJ, ureteropelvic junction; UVJ, ureterovesical junction; PUV, posterior urethral valves; SD, standard deviation.

Results

Patient Characteristics

A total of 336 patients (mean age, 18.3±15.9 days; range, 1 to 94 days; males, n=276) who showed UTD on postnatal US were included, with a median follow-up of 24.2 months. Among the patients, 164 patients had unilateral UTD, and 172 patients had

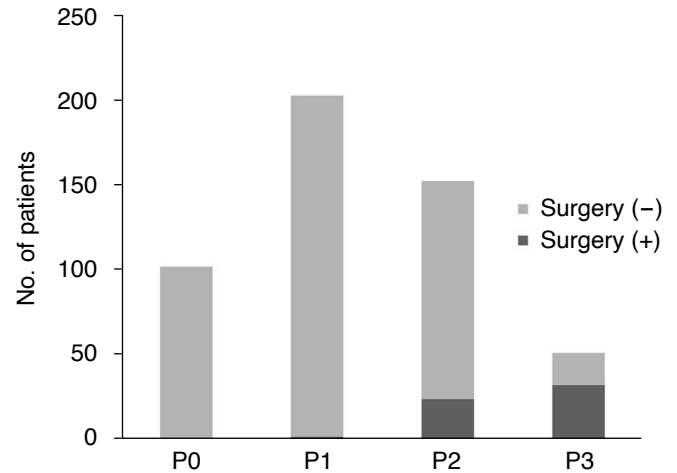


Fig. 2. Component bar chart showing the proportions of kidney units (KUs) that underwent surgical intervention in each urinary tract dilation (UTD) group. Of the 504 KUs, 0% (0 of 101 KUs) with UTD P0, 1.0% (2 of 201 KUs) with UTD P1, 15.9% (24 of 151 KUs) with UTD P2, and 62.8% (32 of 51 KUs) with UTD P3 underwent surgical intervention.

bilateral UTD. After selecting one KU for each patient with posterior urethral valves (n=4), 504 KUs were included in the study. Of the 504 KUs, 20.0% (101/504), 39.9% (201/504), 30.0% (151/504), and 10.1% (51/504) were classified as UTD P0, P1, P2, and P3, respectively. Eighty-four KUs were diagnosed as CAKUT. Patient characteristics are shown in Table 2.

Clinical Outcomes of the UTD Groups

In total, 58 KUs (0% [0 of 101 KUs] in UTD P0, 1.0% [2 of 201 KUs] in UTD P1, 15.9% [24 of 151 KUs] in UTD P2, and 62.7% [32 of 51 KUs] in UTD P3) underwent surgical intervention (Figs. 2, 3). The median age at surgery was significantly younger in patients with UTD P3 than in those with UTD P1 (1.6 months vs. 14.6 months, P=0.004) and UTD P2 (1.6 months vs. 6.7 months, P<0.001). The median time interval between MAG3 scan and kidney US was 16 days (interquartile range, 41 days). The proportions of KUs with significant obstruction and impaired DRF on MAG3 scan were significantly different among the UTD groups (P<0.001) in the KUs with available studies (Fig. 3). The KUs with impaired DRF were more common in association with UTD P3, compared with UTD P1 (P=0.001) and UTD P2 (P=0.001). KUs with significant obstructions on MAG3 scans were more common in the UTD P3 group than in all other UTD groups (P<0.01), and more common in the UTD P2 group than in the UTD P1 group (P<0.001). There were no significant intergroup differences in terms of the presence of VUR (P=0.056).

Fig. 4 summarizes the analysis of the likelihood of patients remaining surgery-free during follow-up. Significant differences were

Table 3. Results of Cox proportional hazards regression analysis clustered by patients for ultrasound factors predicting the requirement for surgery

| Variable | Univariable | | | Multivariable | | |
|------------------------------------|--------------------|---------------|---------|--------------------|--------------|---------|
| | Unadjusted HR | 95% CI | P-value | Adjusted HR | 95% CI | P-value |
| Bilaterality | 0.937 | 0.545–1.611 | 0.814 | | | |
| APRPD (mm) | | | | | | |
| <10 | Reference category | | | Reference category | | |
| 10–15 | 3.549 | 1.436–8.772 | 0.006 | 0.944 | 0.290–3.070 | 0.924 |
| ≥15 | 37.686 | 17.741–80.053 | <0.001 | 8.602 | 3.125–23.677 | <0.001 |
| Central calyceal dilation | 14.014 | 1.940–101.237 | 0.007 | Eliminated | | |
| Peripheral calyceal dilation | 31.184 | 9.754–99.705 | <0.001 | 8.190 | 1.558–43.065 | 0.013 |
| Ureteral dilation | 10.225 | 5.900–17.720 | <0.001 | 2.619 | 1.274–5.380 | 0.009 |
| Parenchymal appearance abnormality | 8.021 | 4.052–15.875 | <0.001 | Eliminated | | |
| Parenchymal thickness abnormality | 18.295 | 10.691–31.308 | <0.001 | 3.371 | 1.574–7.223 | 0.002 |
| Bladder abnormality | 21.734 | 10.077–46.872 | <0.001 | 12.209 | 3.616–41.225 | <0.001 |

HR, hazard ratio; CI, confidence interval; APRPD, anterior-posterior renal pelvic diameter.

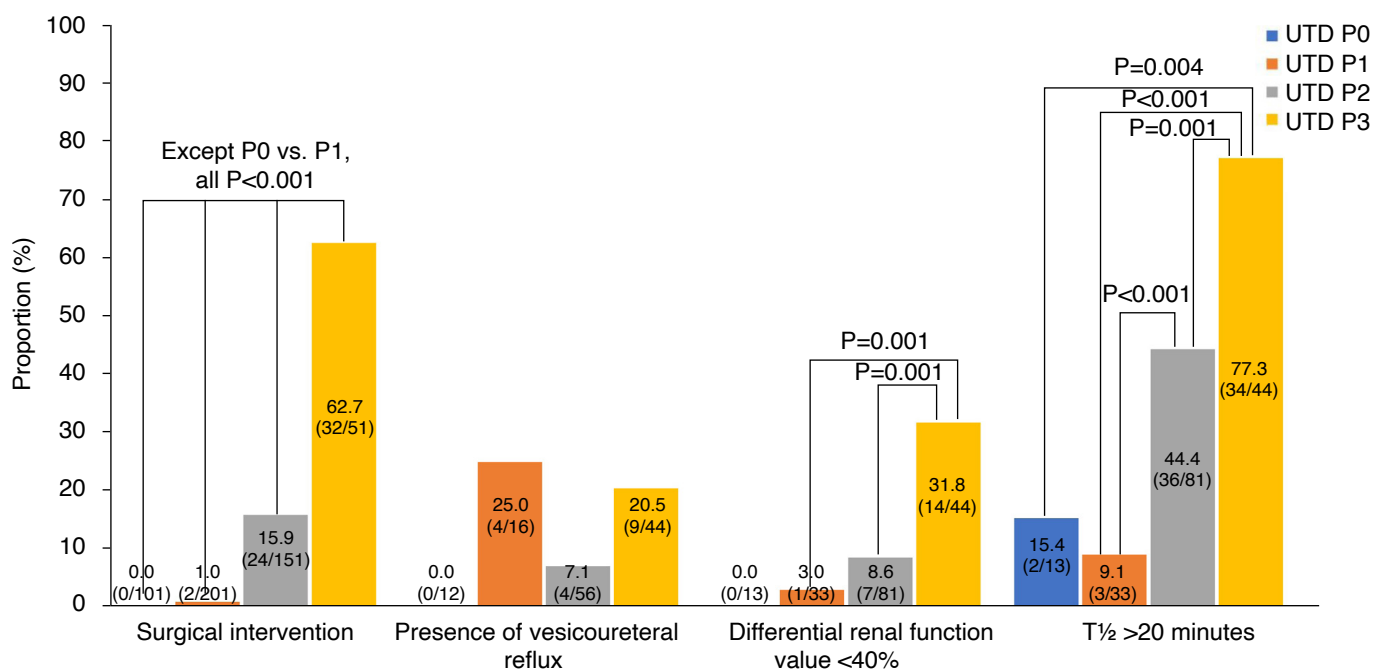


Fig. 3. Bar graphs showing the proportions of renal units regarding surgical intervention, presence of vesicoureteral reflux, decreased differential function (value <40%), and a significant obstruction (T_{1/2} >20 min) on mercaptoacetyltriglycine renal scans among the urinary tract dilation (UTD) groups. P-values for significant difference, as indicated in the bar graphs, were calculated by the Fisher exact test or Pearson chi-square test with the Bonferroni correction.

observed among the UTD groups, as shown in Kaplan-Meier curves (P<0.001). Furthermore, the optimism-corrected C-index for risk stratification of the UTD system was 0.893, indicating an excellent discriminative performance. The surgery-free survival rates were 100% (P0), 100% (P1), 86.7% (P2), and 30.9% (P3) at 1 year; 100% (P0), 98.3% (P1), 81.8% (P2), and 28.3% (P3) at 5 years;

and 100% (P0), 98.3% (P1), 60.2% (P2), and 28.3% (P3) at 9.4 years of follow-up.

US Features Associated with Surgical Intervention

The results of the multivariable Cox regression analysis for US features affecting the requirement for surgery are summarized in

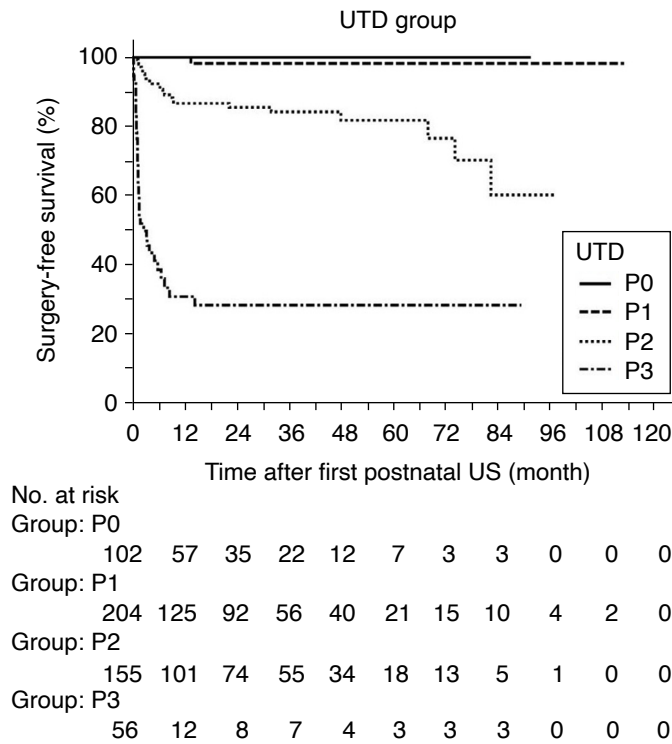


Fig. 4. Kaplan-Meier curves for surgery-free survival stratified into urinary tract dilation (UTD) groups determined by the first postnatal ultrasound (US) examination.

Table 4. Interobserver agreement for ultrasound features in the urinary tract dilation classification system

| Ultrasound feature | Kappa index | 95% CI |
|---|-------------|-------------|
| Central calyceal dilation | 0.585 | 0.494–0.676 |
| Peripheral calyceal dilation | 0.708 | 0.645–0.772 |
| Overall parenchyma appearance | 0.817 | 0.674–0.960 |
| Echogenicity | 0.725 | 0.543–0.908 |
| Cortical cysts | 0.898 | 0.758–1.000 |
| Decreased corticomedullary differentiation | 0.564 | 0.345–0.783 |
| Parenchymal thinning | 0.732 | 0.627–0.838 |
| Ureteral dilation | 0.678 | 0.535–0.821 |
| Overall bladder abnormality | 0.873 | 0.699–1.000 |
| Ureterocele | 0.748 | 0.410–1.000 |
| Wall thickening | 0.856 | 0.577–1.000 |
| Dilated posterior urethra | 0.665 | 0.227–1.000 |
| Final urinary tract dilation classification | 0.716 | 0.672–0.760 |
| P0 vs. P1–P3 | 0.577 | 0.485–0.669 |
| P0–P1 vs. P2–P3 | 0.760 | 0.702–0.819 |
| P0–P2 vs. P3 | 0.822 | 0.743–0.902 |

CI, confidence interval.

Table 3. Bilateral dilation was not a significant factor ($P=0.814$). Multivariable analysis revealed that APRPD ≥ 15 mm compared to APRPD < 10 mm (adjusted hazard ratio [HR], 8.602; 95% CI, 3.125 to 23.677; $P<0.001$), peripheral calyceal dilation (adjusted HR, 8.190; 95% CI, 1.558 to 43.065; $P=0.013$), ureteral dilation (adjusted HR, 2.619; 95% CI, 1.274 to 5.380; $P=0.009$), parenchymal thickness abnormality (adjusted HR, 3.371; 95% CI, 1.574 to 7.223; $P=0.002$), and bladder abnormality (adjusted HR, 12.209; 95% CI, 3.616 to 41.225; $P<0.001$) were significant predictors of surgical intervention (Fig. 5).

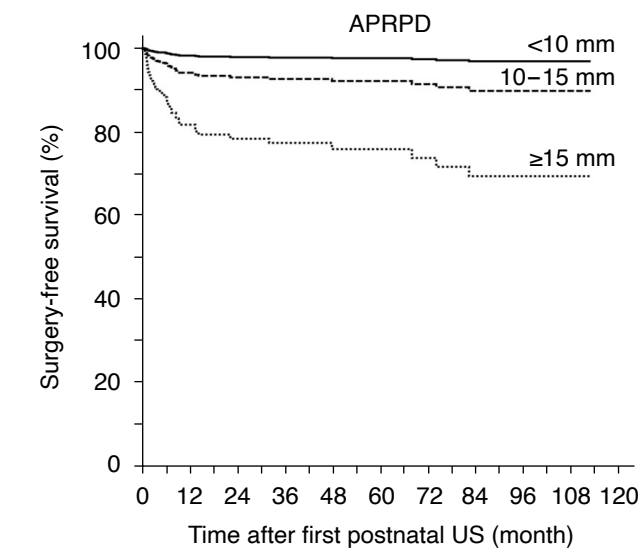
Interobserver Agreement

The reviewers showed moderate agreement for central calyceal dilation ($\kappa=0.585$) and decreased corticomedullary differentiation ($\kappa=0.564$) (Table 4). The kappa values indicated substantial to almost perfect agreement for the other US features included in the UTD classification system (range, 0.665 to 0.898). The reviewers showed substantial agreement regarding assessment of the final UTD grades (UTD P0–P3) ($\kappa=0.716$). For classifying UTD P0–P1 versus UTD P2–P3 and UTD P0–P2 versus UTD P3, the interobserver agreement was substantial ($\kappa=0.760$) and almost perfect ($\kappa=0.822$), respectively.

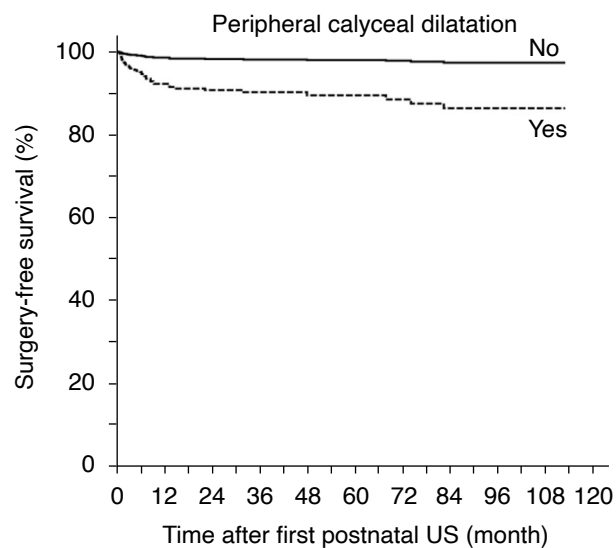
Discussion

The present study showed that the UTD system appropriately stratified the risk of surgical intervention with excellent discriminative performance. Most of the intermediate and high-risk US features in the UTD system were significant predictors of surgical intervention, and bladder abnormality was the most powerful predictor. The interobserver agreement was moderate to almost perfect for US features and substantial for the overall UTD grade. Therefore, the UTD system is a useful US-based risk stratification tool for infants with UTD.

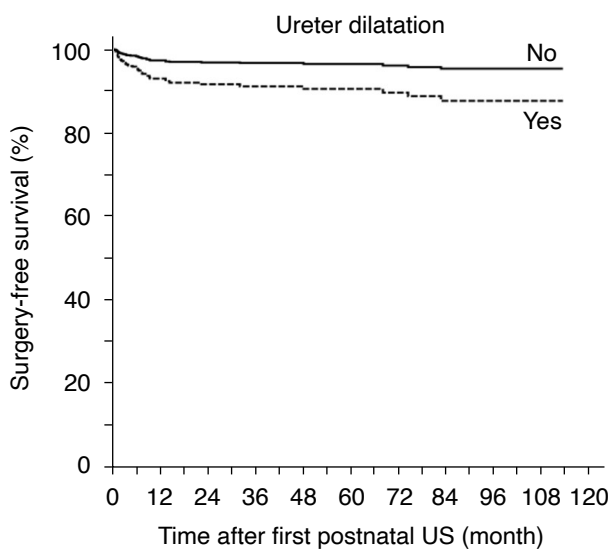
Several studies have been conducted to validate the UTD system for predicting various clinical outcomes regarding surgical intervention [7,8,23], urinary tract infections [7,8], the resolution of UTD [24], and long-term renal injury [8]. In the present study, in concordance with previous studies [7,8,23], the UTD system appropriately stratified the risk of surgical intervention. Unlike this study, previous studies only included ureteropelvic junction obstruction–like hydronephrosis, excluding various conditions associated with ureteral or bladder abnormalities, such as VUR and posterior urethral valves; therefore, the wide spectrum of CAKUT was not evaluated using the UTD system [7,8]. Furthermore, the correlation of each US feature of the UTD system with the risk of surgical intervention was first analyzed. Most of the intermediate- and high-risk US features defined in the UTD system were significant



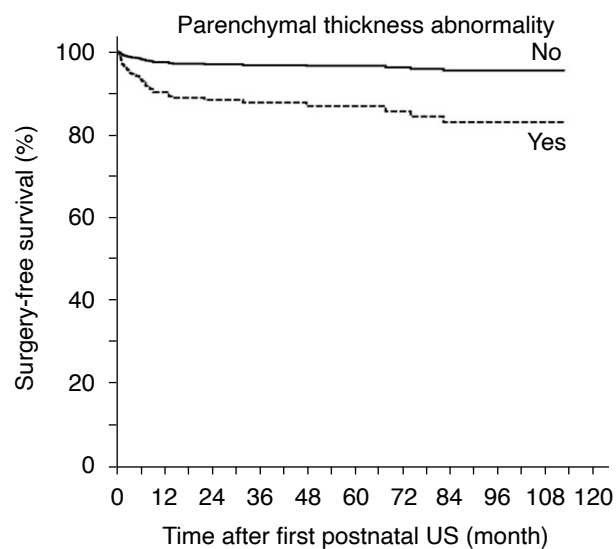
A



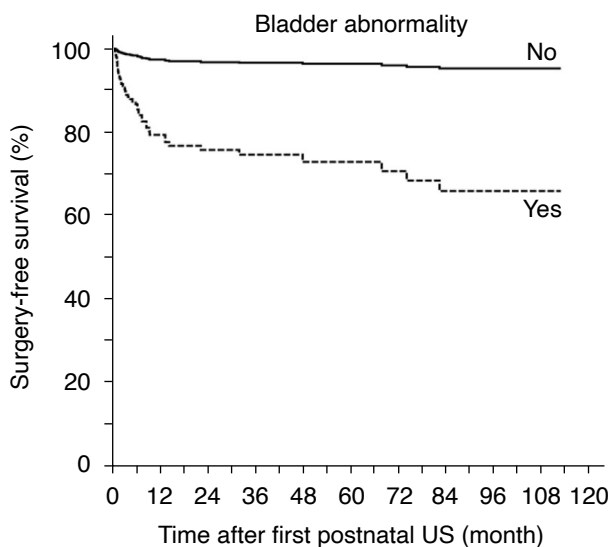
B



C



D



E

Fig. 5. Multivariable-adjusted Kaplan-Meier curves for surgery-free survival according to anterior-posterior renal pelvic diameter (APRPD) (A), as well as the presence of peripheral calyceal dilatation (B), ureteral dilatation (C), parenchymal thickness abnormality (D), and bladder abnormality (E). US, ultrasonography.

predictors of surgical intervention, supporting the prognostic impact of the UTD system. Remarkably, most of the children with UTD P3 required surgical intervention before 1 year of follow-up, with a surgery-free survival rate of 30.9% at 1 year. Children with high-risk US features (UTD P3) should be followed up closely by monitoring them for deterioration of symptoms or renal function, which would necessitate surgery.

However, an abnormal parenchymal appearance was not associated with the occurrence of surgery in this study. In the patients with parenchymal appearance abnormalities, seven of 17 KUs were not treated by surgery; nearly non-functioning kidneys on MAG3 scans were noted in two KUs, and medullary nephrocalcinosis was confirmed as the cause of parenchymal appearance abnormalities in two KUs. Therefore, it can be inferred that parenchymal appearance abnormalities may reflect severely impaired kidney function that cannot benefit from surgical intervention or might be findings associated with other diseases rather than dysplastic parenchymal changes in hydronephrotic kidneys in some cases.

It seemed that UTD P1 was clinically equivalent to normal hydronephrosis variants. The KUs with UTD P1 had similar survival curves to those of KUs with UTD P0, and only two KUs with UTD P1 underwent surgical intervention (ureteroneocystostomy) for high-grade VUR. This finding is in accordance with a previous study that reported no cases of surgical intervention or long-term renal injury among children with UTD P1 [8].

In this study, interobserver agreement in terms of overall UTD grade was better than or similar to the results of previous studies [9,10]. Han et al. [9] reported that the interobserver agreement associated with the UTD system was significantly higher than that associated with the SFU system in the first assessment ($\kappa=0.53-0.89$ vs. $\kappa=0.36-1.00$); however, their study was limited because bladder and ureteral abnormalities were not included in the assessment of overall UTD grades due to a lack of the required images. Nelson et al. [10] studied the interobserver agreement regarding each US feature ($\kappa=0.222-0.895$) and overall UTD grades ($\kappa=0.421$), and both results were poorer than those of the present study. The authors suggest that a lack of clear definitions for parenchymal thinning and bladder wall thickening in the UTD system contributes to substantial variability among raters. In contrast, the present study added definitions of abnormal parenchymal thickness and bladder wall thickening to minimize inconsistency, which might have led to the improved interobserver agreement herein.

In this study, moderate interobserver agreement was found regarding central calyceal dilation measurements; this level of agreement was lower than the levels for the other US features included in the UTD system. This might have resulted in moderate

agreement for classifying UTD P0 versus P1–P3. Similarly, previous publications have pointed out that it could be difficult to evaluate central calyceal dilation when only a few calyces are dilated or to clearly distinguish central from peripheral calyceal dilation [9,10]. The distinction between central and peripheral calyceal dilation is one of the important additions to the UTD system relative to the SFU system. Because calyceal dilation is a continuous variable, it is sometimes not easy to exactly differentiate between central or peripheral calyceal dilation. The two raters in this study reviewed all the available US images for each patient. When just a few peripheral calyces are dilated, it is possible to misinterpret this as central calyceal dilation. However, the differentiation between UTD P0 and P1–P3 is of minor clinical significance, and this issue would not degrade the clinical utility of the UTD classification system.

The present study had several limitations of note. First, the UTD system was not validated for other clinical outcomes, such as urinary tract infection or conditions associated with long-term renal injury, such as end-stage renal disease, hypertension, or proteinuria. Events of febrile urinary tract infection were not sufficiently documented in the electronic medical records, and some of the children were revealed to have been treated for urinary tract infections at local clinics, limiting the accuracy of the evaluation of these variables due to the study's retrospective design. Additionally, various non-clinical factors, such as the parents' or surgeons' preferences or institutional surgical indications shared by multidisciplinary teams, contributed to the timing of surgical interventions. Future studies are warranted to prospectively assess the validity of the UTD system for predicting long-term renal injury in the broad spectrum of CAKUT. Second, it was not possible to analyze the correlation between antenatal UTD classification and surgical intervention, because most of the patients were referred from other institutions and the antenatal kidney US results were not available.

In conclusion, the postnatal UTD classification system appropriately stratified clinical outcome risk in terms of the need for surgical intervention. Most intermediate and high-risk US features defined in the UTD system significantly predicted the occurrence of surgical intervention, and the presence of a bladder abnormality was the most powerful predictor. Each US feature and the overall UTD grades were reliable, with moderate to almost perfect interobserver agreement, and setting definitions for some previously vaguely defined US features appeared to be helpful. The authors expect that the UTD classification system will become a useful common language for describing UTD in infants. It can provide useful information for follow-up evaluation, management, and counseling parents regarding patients' prognoses.

ORCID: Jisun Hwang: <https://orcid.org/0000-0002-7593-2246>; Pyeong Hwa Kim: <https://orcid.org/0000-0003-4276-8803>; Hee Mang Yoon: <https://orcid.org/0000-0001-6491-5734>; Sang Hoon Song: <https://orcid.org/0000-0003-4888-483X>; Ah Young Jung: <https://orcid.org/0000-0002-7427-6240>; Jin Seong Lee: <https://orcid.org/0000-0002-8470-4595>; Young Ah Cho: <https://orcid.org/0000-0001-6722-121X>

Author Contributions

Conceptualization: Hwang J, Kim PH, Yoon HM, Song SH, Jung AY, Lee JS, Cho YA. Data acquisition: Hwang J, Kim PH. Data analysis or interpretation: Hwang J, Kim PH, Yoon HM. Drafting of the manuscript: Hwang J. Critical revision of the manuscript: Kim PH, Yoon HM, Song SH, Jung AY, Lee JS, Cho YA. Approval of the final version of the manuscript: all authors.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

- Ek S, Lidefeldt KJ, Varricio L. Fetal hydronephrosis; prevalence, natural history and postnatal consequences in an unselected population. *Acta Obstet Gynecol Scand* 2007;86:1463-1466.
- Nguyen HT, Herndon CD, Cooper C, Gatti J, Kirsch A, Kokorowski P, et al. The Society for Fetal Urology consensus statement on the evaluation and management of antenatal hydronephrosis. *J Pediatr Urol* 2010;6:212-231.
- Swenson DW, Darge K, Ziniel SI, Chow JS. Characterizing upper urinary tract dilation on ultrasound: a survey of North American pediatric radiologists' practices. *Pediatr Radiol* 2015;45:686-694.
- Keays MA, Guerra LA, Mihill J, Raju G, Al-Asheeri N, Geier P, et al. Reliability assessment of Society for Fetal Urology ultrasound grading system for hydronephrosis. *J Urol* 2008;180:1680-1682.
- Kim SY, Kim MJ, Yoon CS, Lee MS, Han KH, Lee MJ. Comparison of the reliability of two hydronephrosis grading systems: the Society for Foetal Urology grading system vs. the Onen grading system. *Clin Radiol* 2013;68:e484-e490.
- Nguyen HT, Benson CB, Bromley B, Campbell JB, Chow J, Coleman B, et al. Multidisciplinary consensus on the classification of prenatal and postnatal urinary tract dilation (UTD classification system). *J Pediatr Urol* 2014;10:982-998.
- Braga LH, McGrath M, Farrokhlyar F, Jegatheeswaran K, Lorenzo AJ. Associations of Initial Society for Fetal Urology grades and urinary tract dilatation risk groups with clinical outcomes in patients with isolated prenatal hydronephrosis. *J Urol* 2017;197:831-837.
- Melo FF, Mak RH, Simoes ES, Vasconcelos MA, Dias CS, Rosa LC, et al. Evaluation of urinary tract dilation classification system for prediction of long-term outcomes in isolated antenatal hydronephrosis: a cohort study. *J Urol* 2021;206:1022-1030.
- Han M, Kim HG, Lee JD, Park SY, Sur YK. Conversion and reliability of two urological grading systems in infants: the Society for Fetal Urology and the urinary tract dilatation classifications system. *Pediatr Radiol* 2017;47:65-73.
- Nelson CP, Lee RS, Trout AT, Servaes S, Kraft KH, Barnewolt CE, et al. Interobserver and intra-observer reliability of the urinary tract dilation classification system in neonates: a multicenter study. *J Urol* 2019;201:1186-1192.
- Vallasciani S, Bujons Tur A, Gatti J, Machado M, Cooper CS, Farrugia MK, et al. Hydronephrosis classifications: has UTD overtaken APD and SFU? A worldwide survey. *Front Pediatr* 2021;9:646517.
- Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007;4:e297.
- Vivier PH, Augdal TA, Avni FE, Bacchetta J, Beetz R, Bjerre AK, et al. Standardization of pediatric urological terms: a multidisciplinary European glossary. *Pediatr Radiol* 2018;48:291-303.
- Kadioglu A. Renal measurements, including length, parenchymal thickness, and medullary pyramid thickness, in healthy children: what are the normative ultrasound values? *AJR Am J Roentgenol* 2010;194:509-515.
- Paliwalla M, Park K. A practical guide to urinary tract ultrasound in a child: pearls and pitfalls. *Ultrasound* 2014;22:213-222.
- Nelson CP, Lee RS, Trout AT, Servaes S, Kraft KH, Barnewolt CE, et al. The association of postnatal urinary tract dilation risk score with clinical outcomes. *J Pediatr Urol* 2019;15:341.
- Hashim H, Woodhouse CR. Ureteropelvic junction obstruction. *Eur Urol Suppl* 2012;11:25-32.
- Viteri B, Calle-Toro JS, Furth S, Darge K, Hartung EA, Otero H. State-of-the-art renal imaging in children. *Pediatrics* 2020;145:e20190829.
- Agard H, Massanyi E, Albertson M, Anderson M, Alam M, Lyden E, et al. The different elements of the Urinary Tract Dilation (UTD) classification system and their capacity to predict findings on mercaptoacetyltriglycine (MAG3) diuretic renography. *J Pediatr Urol* 2020;16:686.
- Song SH, Park S, Chae SY, Moon DH, Park S, Kim KS. Predictors of renal functional improvement after pyeloplasty in ureteropelvic junction obstruction: clinical value of visually assessed renal tissue tracer transit in (99m)Tc-mercaptoacetyltriglycine Renography. *Urology* 2017;108:149-154.
- Wong MC, Ching JY, Ng S, Lam TY, Luk AK, Wong SH, et al. The discriminatory capability of existing scores to predict advanced colorectal neoplasia: a prospective colonoscopy study of 5,899 screening participants. *Sci Rep* 2016;6:20080.
- Crewson PE. Reader agreement studies. *AJR Am J Roentgenol* 2005;184:1391-1397.

23. Hodhod A, Capolicchio JP, Jednak R, El-Sherif E, El-Doray Ael A, El-Sherbiny M. Evaluation of urinary tract dilation classification system for grading postnatal hydronephrosis. *J Urol* 2016;195:725-730.
24. Braga LH, McGrath M, Farrokhyar F, Jegatheeswaran K, Lorenzo AJ. Society for Fetal Urology Classification vs urinary tract dilation grading system for prognostication in prenatal hydronephrosis: a time to resolution analysis. *J Urol* 2018;199:1615-1621.