Original Article

Obstet Gynecol Sci 2019;62(2):87-92
https://doi.org/10.5468/ogs.2019.62.2.87
pISSN 2287-8572 · eISSN 2287-8580

Obstetrics & Gynecology Science

Reliability of society of fetal urology and Onen grading system in fetal hydronephrosis

Hee Young Cho[1,2], Inkyung Jung[3], Young Han Kim[1], Ja-Young Kwon[1]

[1]Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Institute of Women's Life Medical Science, Yonsei University Health System, Seoul; [2]Department of Obstetrics and Gynecology, CHA Bundang Medical Center, CHA University, Seongnam; [3]Division of Biostatistics, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea

Objective

To evaluate the reliability of the Society for Fetal Urology (SFU) and Onen grading systems for fetal hydronephrosis in prenatal ultrasound according to the level of experience of the examiner.

Methods

We reviewed the prenatal ultrasound images of 146 fetuses (292 kidneys) that were diagnosed as having hydronephrosis between January 2005 and December 2014. One expert and two trainees assessed the prenatal renal ultrasound images using the SFU and Onen grading systems. The three examiners independently assessed each ultrasound image with both grading systems and reassessed the same images after 7 to 14 days. Cohen's kappa statistic was used to estimate intra- and inter-observer reliability in prenatal ultrasound images according to training level.

Results

The intra-observer reliability of the SFU grading system (κ 0.873–0.945) showed almost perfect agreement and that of the Onen grading system (κ 0.749–0.913) showed substantial to almost perfect agreement. The overall inter-observer reliability of the SFU grading system (κ 0.620–0.825) showed substantial to almost perfect agreement and that of the Onen grading system (κ 0.618–0.724) showed substantial agreement. The weighted kappa value of inter-observer agreement was 0.223 to 0.400 for SFU grade 1 and 0.064 to 0.346 for SFU grade 3. For Onen grading, the inter-observer agreement was 0.012 to 0.214 for grade 2 and 0.193 to 0.334 for grade 3.

Conclusion

Both the SFU and Onen grading systems showed good intra-observer agreement in prenatal ultrasonography. The inter-observer agreement was decreased in SFU grades 1 and 3 and Onen grades 2 and 3. Therefore, more focus should be given to SFU grades 1 and 3 and Onen grades 2 and 3 for trainees.

Keywords: Fetus; Hydronephrosis; Fetal hydronephrosis; Prenatal ultrasonography; Grading system

Introduction

Antenatal hydronephrosis is defined as abnormal renal pelvis dilatation with or without changes in the renal parenchyma, and its prevalence is approximately 1–4% of all pregnancies [1]. As the severity of hydronephrosis measured using ultrasonography became known as an important prognostic factor [2], the Society for Fetal Urology (SFU) grading system was introduced by the SFU in 1993 [3]. Since then, this system has been most commonly used to determine the severity of fetal and pediatric hydronephrosis. Some studies have shown that the hydronephrosis severity according to the SFU

Received: 2018.04.05. Revised: 2018.07.19. Accepted: 2018.08.09.

Corresponding author: Ja-Young Kwon
Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Institute of Women's Life Medical Science, Yonsei University Health System, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
E-mail: jaykwon@yuhs.ac
https://orcid.org/0000-0003-3009-6325

grading system and perinatal outcomes are correlated [4,5]. Hydronephrosis with low SFU grades usually resolve spontaneously and show good prognosis; however, hydronephrosis with high SFU grades show various features, making prognosis difficult to predict [6]. Furthermore, the SFU grading system differentiates the degree of hydronephrosis according to renal pelvic dilatation, calyceal dilatation, and the presence of cortical thinning, which makes the grading system rather subjective, as follows: grade 0 = no hydronephrosis, grade 1 = only visualized renal pelvis, grade 2 = dilatation of a few but not all calyces, grade 3 = dilatation of virtually all calyces, and grade 4 = calyceal dilatation and parenchymal thinning.

There is no single powerful ultrasound parameter for predicting postnatal renal function [7]; however, follow-up serial ultrasound examinations are appropriate to easily detect the progression of hydronephrosis.

In neonates, Onen proposed a new grading system in 2006 by modifying the SFU grading system to show the severity better and to make follow-up more practical [8]. The Onen grading system combined SFU grades 1 and 2 into Onen grade 1, and divided SFU grade 4 according to the degree of renal parenchymal loss (less than 50%, Onen grade 3; more than 50%, Onen grade 4). There are important differences between the two grading systems; however, there is only little information about the reliability of these grading systems for prenatal ultrasound. Moreover, both grading systems are subjective and can be influenced by the expertise of the examiner, and they have been little investigated.

Therefore, we conducted this study to evaluate the intraobserver and inter-observer agreement of the SFU and Onen grading systems, and to investigate the agreement of prenatal hydronephrosis diagnosis between the two grading systems according to the level of experience of the examiner.

Materials and methods

The study population consisted of 292 kidneys in 146 fetuses with hydronephrosis diagnosed between March 2005 and December 2013. The institutional review board in Yonsei University Health System approved the study protocol (IRB No. 4-2015-1167). This study included pregnant women who underwent prenatal ultrasonography and their neonates who underwent postnatal ultrasonography for evaluating the kidneys. Cases complicated with cystic renal disease, single kidney, duplicated collecting system, or horseshoe kidney were excluded from the study. All pregnant women underwent ultrasonography for the assessment of fetal hydronephrosis in the prenatal period, and all neonates underwent abdominal ultrasonography to evaluate the grade of hydronephrosis after birth. The prenatal ultrasound images were randomly selected including one axial view and one sagittal view, and all images were collected into a file for review. All personal information was eliminated to protect patient confidentiality, and the examiners were blinded to the ultrasound examination reports. One expert (H.Y.C.) with 10 years of experience in prenatal ultrasound and two trainees (M.H.C. and J.H.C.) with less than 1 year of experience participated in the assessments. Before reviewing the images, all examiners were provided with written instruction material about the SFU and Onen grading systems (Fig. 1). The three examiners independently assessed each ultrasound image with both grading systems and performed reassessment of the same images after 7 to 14 days.

The intra- and inter-observer agreements of the three examiners about the prenatal ultrasound images were evaluated using non-weighted Cohen's kappa and weighted kappa statistics. A kappa value of 0.81 to 0.99 was considered an almost perfect agreement; 0.61 to 0.80, substantial agreement; 0.41 to 0.60, moderate agreement; 0.21 to 0.40, fair agreement; 0.01 to 0.20, slight agreement; and below 0, poor or less than chance agreement [9].

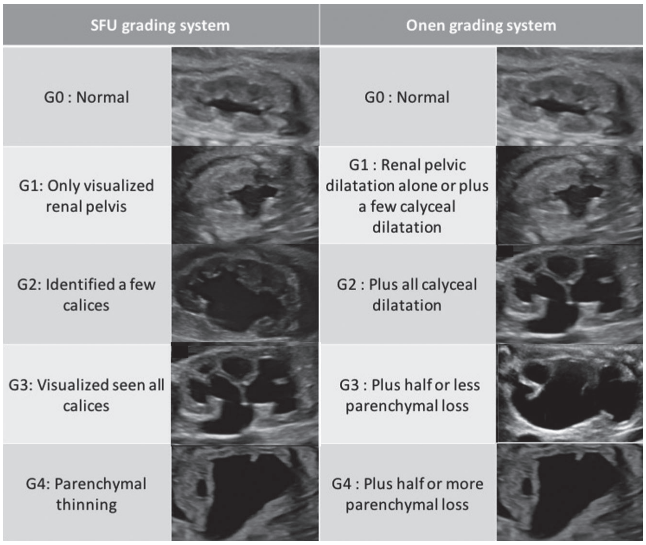
SFU grading system
Onen grading system
G0 : Normal
G0 : Normal
G1: Only visualized renal pelvis
G1 : Renal pelvic dilatation alone or plus a few calyceal dilatation
G2: Identified a few calices
G2 : Plus all calyceal dilatation
G3: Visualized seen all calices
G3 : Plus half or less parenchymal loss
G4: Parenchymal thinning
G4 : Plus half or more parenchymal loss

Fig. 1. Ultrasound images evaluated using the SFU grading system and the Onen grading system. SFU, Society for Fetal Urology.

Results

The baseline clinical characteristics of the 292 kidneys in the study population are shown in Table 1. The gestational age at ultrasound was 31.5 gestational weeks (28.0–35.5 weeks), and the amniotic fluid index at measurement was normal (15.5±4.8). At the time of measurement, the median renal pelvis anteroposterior diameter was 1.6 cm (1.1–8.4 cm). According to grading and examiners, we presented the distribution of the number of kidneys to the SFU and Onen grades for each assessment (Table 2). Examiners including A, B, and C reviewed the prenatal ultrasound images.

The intra-observer agreements of hydronephrosis for the SFU grading system among the three examiners were in the range of 0.734 to 0.880 and 0.873 to 0.945 for Cohen's kappa and weighted kappa, respectively (Table 3). The weighted kappa statistics showed almost perfect agreement for the three examiners. The Cohen's kappa values of the Onen grading system were 0.496 to 0.847, and showed moderate to almost perfect agreement. Weighted kappa statistics also showed substantial to almost perfect agreement (0.749–0.913) in the Onen grading system. Cohen's kappa values were calculated to investigate the difference in the SFU and Onen grading systems, and examiner A showed a statistically significant intra-observer agreement (–0.238, 95% confidence interval –0.368 to –0.110).

Table 1. Basic characteristics of the study population

Variables (n=292)	Mean±SD or median (min–max)
Real pelvis AP diameter (cm)	1.6 (1.1–8.4)
Gestational age at measurement (wk)	31.5 (28.0–35.5)
Amniotic fluid index at measurement	15.5±4.8
Gestational age at delivery (wk)	38.0 (36.0–40.2)

SD, standard deviation; AP, anteroposterior.

The inter-observer agreements among the three examiners in both grading systems were evaluated (Table 4). The comparison of expert B and trainee A for SFU grading showed a weighted kappa value of 0.825, an almost perfect agreement. For Onen grading, the weighted kappa value was 0.724, meaning a substantial agreement. For the comparison of expert B and trainee C, both grading systems showed substantial agreement for weighted kappa value (SFU, κ 0.620; Onen, κ 0.618).

However, the inter-observer agreement of trainees and expert was decreased in SFU grades 1 and 3 and Onen grades

Table 2. Distribution of the kidneys according to grading and examiners with the Society for Fetal Urology and Onen grading systems

Examiners	Grading system	Kidneys (n=292)							
		First assessment				Second assessment			
		1	2	3	4	1	2	3	4
A	SFU	115	47	27	103	124	46	32	90
	Onen	152	32	50	58	164	41	61	26
B	SFU	112	47	35	98	111	49	26	106
	Onen	157	36	49	50	156	30	46	60
C	SFU	123	67	47	55	118	73	35	66
	Onen	188	37	35	32	190	27	39	36

SFU, Society for Fetal Urology.

Table 3. Intra-observer agreement of hydronephrosis of prenatal ultrasound with the SFU and Onen grading system

Examiners	SFU		Onen		Difference between SFU and Onen
	Cohen's κ	Weighted κ	Cohen's κ	Weighted κ	Cohen's κ
A	0.734 (0.658:0.809)	0.873 (0.825:0.920)	0.496 (0.402:0.590)	0.749 (0.699:0.799)	–0.238 (–0.368:–0.110)
B	0.880 (0.819:0.941)	0.945 (0.915:0.975)	0.791 (0.715:0.867)	0.895 (0.850:0.938)	–0.090 (–0.169:0.016)
C	0.764 (0.719:0.869)	0.891 (0.847:0.935)	0.847 (0.779:0.915)	0.913 (0.872:0.954)	0.054 (–0.014:0.122)

SFU, Society for Fetal Urology.

Table 4. Inter-observer agreement of each reviewer with the Society for Fetal Urology and Onen grading systems

Grading system	Examiner: A vs. B		Examiner: C vs. B	
	SFU	Onen	SFU	Onen
1	0.400 (0.234:0.556)	0.449 (0.287:0.612)	0.223 (0.060:0.385)	0.430 (0.268:0.592)
2	0.486 (0.324:0.648)	0.214 (0.051:0.376)	0.346 (0.184:0.509)	0.012 (–0.150:0.175)
3	0.346 (0.184:0.508)	0.334 (0.172:0.497)	0.064 (–0.099:0.226)	0.193 (0.031:0.355)
4	0.823 (0.661:0.986)	0.604 (0.442:0.767)	0.513 (0.351:0.675)	0.523 (0.361:0.686)
Cohen's κ	0.644 (0.556:0.732)	0.514 (0.419:0.610)	0.418 (0.323:0.513)	0.419 (0.290:0.483)
Weighted κ	0.825 (0.772:0.878)	0.724 (0.656:0.792)	0.620 (0.533:0.706)	0.618 (0.532:0.704)

SFU, Society for Fetal Urology.

2 and 3. In each SFU grade, the inter-observer agreement was fair and slight to fair for SFU grade 1 (κ 0.223–0.400) and grade 3 (κ 0.064–0.346), respectively. For Onen grading, the inter-observer agreement was slight to fair and fair for Onen grade 2 (κ 0.012–0.214) and grade 3 (κ 0.193–0.334), respectively.

Discussion

To our knowledge, this is the first study to investigate the inter- and intra-observer reliability of the SFU and Onen grading systems for fetal hydronephrosis in prenatal ultrasound images according to the level of experience of the examiner. Our results showed that the intra-observer agreement of the SFU grading system showed almost perfect agreement and that of the Onen grading system showed substantial to almost perfect agreement for the kidneys. The inter-observer agreement of the SFU grading system showed substantial to almost perfect agreement and the Onen grading system showed substantial agreement among the three examiners.

In 2008, the reliability of the SFU grading system was investigated. The inter-rater reliability showed fair to substantial agreement and the intra-rater reliability showed substantial to almost perfect agreement for staff and trainees. Moreover, Kim et al. [10] compared the reliability of the SFU and Onen grading systems for pediatric hydronephrosis diagnosed using ultrasonography. The inter-rater agreement was substantial and the intra-rater agreement was substantial to almost perfect in the SFU and Onen grading systems. Both the SFU and Onen grading systems showed good intra- and inter-observer agreements in the diagnosis and follow-up of pediatric hydronephrosis.

SFU grading is mainly used for the diagnosis of hydronephrosis in prenatal ultrasonography; however, there has been no study about its reliability. Our study demonstrated that the intra-observer agreement for the diagnosis of hydronephrosis in prenatal ultrasound showed almost perfect agreement in the SFU grading system and substantial to almost perfect agreement in the Onen grading system. The inter-observer agreement of expert B and trainee A showed almost perfect agreement for SFU grading and substantial agreement for Onen grading. For the comparison of expert B and trainee C, the SFU and Onen grading systems both showed substantial agreement. Our results demonstrated that the inter-observer agreement of the SFU grading system was significantly higher than that of the Onen grading system. Further, the inter-observer agreement for SFU grades 1 and 3 and Onen grades 2 and 3 were lower than for other grades. The reasons for these results are as follows. First, as SFU grading is mainly used in prenatal ultrasonography, the examiners may be more familiar with this system. Second, grades 2 and 3 in the Onen grading system may be difficult to distinguish between cases with dilatation of all calyces and half or less parenchymal loss. In 2013, Kim et al. [10] reported the reliability of the SFU and Onen grading systems for pediatric patients with hydronephrosis. This study showed good intra- and inter-observer agreement but relatively lower inter-observer agreement in SFU grades 1 and 2 and Onen grades 2 and 3. This result indicates that beginners may have difficulty in discriminating between renal pelvis dilatation only (SFU grade 1) and visualized renal pelvis with dilatation of a few calyces (SFU grade 2). However, our results showed lower inter-observer agreement in SFU grades 1 and 3. The reason for this discrepancy is unclear, although it could be because our study focused on prenatal ultrasound results and the previous study investigated postnatal ultrasound images.

In practice, prenatal ultrasound grading of fetal hydronephrosis could affect the antenatal consultation about prognosis and could provide important information for the evaluation and management in the postnatal period. Therefore, accurate prenatal grading of fetal hydronephrosis is crucial, and this study might contribute to determining the grade of hydronephrosis. Further, as our study specifically addressed the difficulties that trainees experience in diagnosing the grade of hydronephrosis in prenatal ultrasound, effective training could be implemented based on the results of this study.

This study has some limitations. First, a randomized clinical trial is ideal but this was a retrospective study; therefore, there might have been selection bias in patients and examiner bias. Second, as we investigated reliability among only three examiners, the results of this study could be limited. Reliability needs to be evaluated with more examiners having diverse levels of experience. Third, we did not investigate which grading system is more correlated with the prognosis of neonatal hydronephrosis. At our institution, SFU grading is mainly used in prenatal ultrasound and Onen grading in postnatal ultrasound. Onen [11] reported that the Onen grading system has advantages for the follow-up of hydronephrosis and for determining the timing of surgical intervention in neonates; however, the Onen grading system has not been evaluated for hydronephrosis diagnosed using prenatal ultrasonography.

We conclude that both the SFU and Onen grading systems showed good intra-observer agreement in hydronephrosis evaluation. The inter-observer agreement was lower in SFU grades 1 and 3 and Onen grades 2 and 3, and more focus should be given to SFU grades 1 and 3 and Onen grades 2 and 3 for trainees. Further studies are necessary to determine which grading system is more relevant to the postnatal prognosis of fetal hydronephrosis.

Conflict of interest

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

Ethical approval

The study was approved by the Institutional Review Board of Yonsei University Health System (IRB No. 4-2015-1167) and performed in accordance with the principles of the Declaration of Helsinki. Written informed consents were obtained.

References

1. Gunn TR, Mora JD, Pease P. Antenatal diagnosis of urinary tract abnormalities by ultrasonography after 28 weeks' gestation: incidence and outcome. Am J Obstet Gynecol 1995;172:479-86.
2. Onen A, Jayanthi VR, Koff SA. Long-term followup of prenatally detected severe bilateral newborn hydronephrosis initially managed nonoperatively. J Urol 2002;168:1118-20.
3. Fernbach SK, Maizels M, Conway JJ. Ultrasound grading of hydronephrosis: introduction to the system used by the Society for Fetal Urology. Pediatr Radiol 1993;23:478-80.
4. Ulman I, Jayanthi VR, Koff SA. The long-term followup of newborns with severe unilateral hydronephrosis initially treated nonoperatively. J Urol 2000;164:1101-5.
5. Konda R, Sakai K, Ota S, Abe Y, Hatakeyama T, Orikasa S. Ultrasound grade of hydronephrosis and severity of renal cortical damage on 99m technetium dimercaptosuccinic acid renal scan in infants with unilateral hydronephrosis during followup and after pyeloplasty. J Urol 2002;167:2159-63.
6. Sidhu G, Beyene J, Rosenblum ND. Outcome of isolated antenatal hydronephrosis: a systematic review and meta-analysis. Pediatr Nephrol 2006;21:218-24.
7. Imaji R, Dewan PA. Calyx to parenchyma ratio in pelvi-ureteric junction obstruction. BJU Int 2002;89:73-7.
8. Onen A. An alternative grading system to refine the criteria for severity of hydronephrosis and optimal treatment guidelines in neonates with primary UPJ-type hydronephrosis. J Pediatr Urol 2007;3:200-5.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
10. 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:e484-90.

11. Onen A. Treatment and outcome of prenatally detected newborn hydronephrosis. J Pediatr Urol 2007;3:469-76.