The Value of Shadowing and the Twinkling Artifact in the Diagnosis of Ureteral Stones: A Single-center Study

Meng Wang, Qi Ma, Yang Chen, Jie Li, Cheng Wang, Yuma Jin, Yingchun Zhang, Hanbing Chen, Songtao Liu, Caishan Wang, Yujing Sheng, Xue Bai, Ying Zhang, Wenwen Lin, and Xinyu Feng

OBJECTIVE To determine and evaluate the value of shadowing and the twinkling artifact (TA) for the diagnosis of ureteral stones.

MATERIALS AND METHODS Related ultrasound images from 117 patients with suspected ureteral stones were consecutively collected with optimized machine settings, confirmed by computed tomography and then retrospectively reviewed by 12 physicians who were classified into 3 groups according to their experience levels: elementary, intermediate, and advanced. The shadowing/TA grades were separately evaluated by all the participating physicians in a blinded manner, and the consistency was verified using Kendall’s coefficient of concordance (Kendall’s W). Furthermore, the diagnostic performance was compared among the groups stratified by physicians’ clinical experience levels and ureteral stone sizes.

RESULTS Using shadowing/TA as indicators for ureteral stones, Kendall’s W for the TA evaluation was higher than that for shadowing among all the participating physicians and subgroups (P < .05). Furthermore, with no difference in specificity at 100%, the sensitivity of the isolated TA was superior to that of shadowing in groups stratified by the physicians’ clinical experience levels and stone sizes, respectively (P < .05). However, for the respective comparisons of shadowing and the TA among groups stratified by stone sizes, as ureteral stones became larger, the detection sensitivities all significantly increased (P ≤ .001).

CONCLUSION Among physicians, subjective evaluation of the TA is more consistent and has better diagnostic sensitivity than that of shadowing for the diagnosis of ureteral stones, and the stone size may play an important role in the detection sensitivity of these 2 indicators. UROLOGY 126: 39–44, 2019. © 2019 Elsevier Inc.

Ureteral stones are the most common cause of acute flank pain in the emergency department. Regarding the corresponding diagnostic methods for ureteral stones, both nonenhanced computed tomography (CT) and ultrasonography have been accepted as first-line modalities according to clinical experience. CT has been always considered the gold standard due to its high sensitivity and specificity. However, the cost and the irradiation risk of CT have been considered 2 obvious drawbacks. Thus, real-time ultrasonography may be a more suitable choice than CT to some extent.²

Notably, the low sensitivity of ultrasound machines results in a problematic false-negative rate for ureteral stone detection, and decreases its applicable value. As we all know, a hyperechogenic region with shadowing along the ureter is considered a common indicator for the presence of a ureteral stone, although the sensitivity of this feature is low, especially for smaller stones and stones in the middle of the ureteral tract.³-⁷ Interestingly, in 1996, Rahmouni et al⁸ proposed another artifact that occurs behind urinary stones in color Doppler mode called the “twinkling artifact” (TA), which was defined as a rapidly changing mixture of colors behind a highly reflective structure on color Doppler ultrasound, and the TA was found to be another indicator of urinary stones on color Doppler imaging in their study. Later, Shababa et al assessed the capacity of the TA on color Doppler imaging to detect renal stones relative to acoustic shadowing and showed that the presence of the TA was a robust indicator of renal calculi, and the color signature enabled easier stone detection than shadowing.⁹

Conflict of interest: There are no conflicts of interest associated with the writing of this manuscript.

Ethical standards: This study was approved by Institutional Review Board (IRB) of the Second Affiliated Hospital of Soochow University. A waiver of informed consent was obtained from the IRB due to no risk posed to patients.

From the Department of Ultrasound, The Second Affiliated Hospital of Soochow University, Suzhou, China; the School of Mathematics and Physics, Soochow University of Science and Technology, Suzhou, China; and the Department of Ultrasound, Qilu Hospital, Shandong University, Jinan, China.

Address correspondence to: Meng Wang M.D., Department of Ultrasound, The Second Affiliated Hospital of Soochow University, Suzhou 215004, China; Jie Li M.D., Department of Ultrasound, Qilu Hospital, Shandong University, Jinan 250012, China. E-mails: wangmeng601@163.com; jieli301@163.com

Submitted: November 20, 2018, accepted (with revisions): January 25, 2019

https://doi.org/10.1016/j.urology.2019.01.021
0090-4295

© 2019 Elsevier Inc.
All rights reserved.
Recently, many researchers have conducted clinical studies on the value of shadowing/TA for diagnosing urinary stones. However, the true value of these indicators remains controversial. A retrospective review of these corresponding studies revealed that although shadowing/TA was sometimes graded according to the length or the area behind a urinary stone, differences in grading shadowing/TA among observers have not been investigated. Additionally, whether grading shadowing/TA can aid all physicians in making precise decisions remains a confusing problem. Therefore, in this study, using shadowing and the TripdelA as isolated indicators for the diagnosis of ureteral stones, the aim herein was to evaluate the consistency of observers' evaluations and to compare the diagnostic performance of shadowing/TA among groups stratified by the clinical experience levels of the observers and ureteral stone size.

**MATERIALS AND METHODS**

Between April 2015 and January 2018, patients with acute flank pain with or without hematuria were referred to the Second Affiliated Hospital of Soochow University Ultrasonic Imaging Centre for further evaluation. This retrospective study involving evaluation of ureteral stone images was conducted after receiving approval from the local institutional review board, and the requirement for patient informed consent was waived. The medical records of a total of 312 patients were reviewed clinically for reevaluation of ureteral stones (n = 117). Conversely, the patients who were confirmed by CT not to suffer from ureteral stones were designated as the control group (n = 31), which included those with spontaneous abdominal pain (n = 10), appendicitis (n = 6), ureterostenosis (n = 4), ovarian torsion (n = 1), epididymitis (n = 3), colonic disease (n = 2), and nonpositive findings (n = 5).

**Ultrasound Examination and CT Protocols**

The ultrasound examinations were performed using a LOGIQ P5 ultrasound machine (GE Ultrasound Korea, Seongnam-si, Gyeonggi-do, Korea) with a convex 3-7 MHz probe. The patients who were referred to the ultrasound department underwent B-mode ultrasonography and color Doppler ultrasonography of the kidneys and ureters, respectively. When the patients were scanned, both flanks were carefully checked in both the supine and prone positions. First, the condition of the kidneys was examined, and the ureters were imaged and recorded from the renal pelvis to the ureterovesical junction in a sequential scanning manner. When a suspicious hyperechogenic region was found, the size and location as well as the shadowing pattern were recorded. Next, color Doppler ultrasonography was performed at the same site under optimized conditions for identifying the TA, with reference to our previous study (Table 1).

**Image Evaluation**

All the included ultrasound images were archived on a commercial digital workstation (Neosoft Co. Ltd., Shenyang, China). Then, the findings from the B-mode and color Doppler mode ultrasound images were assigned 1 of 4 grades using the classification standards from previous studies (Table 2 and supplementary Fig. 2). The final impressions of whether shadowing/TA was displayed or not were determined by the observer and categorized as “present” or “absent”.

Twelve physicians with different levels of clinical experience evaluated the shadowing and TA grades. The participants were classified into 3 groups including an elementary group (<5 years of experience), an intermediate group (5-10 years of experience) and an advanced group (>10 years of experience), with 4 participants in each group. The examiners, who were trained on the principles of this study before the evaluation, were first shown 8 different standard images reflecting the degrees of shadowing and TA. Then, the 12 examiners separately evaluated the images in a blinded manner, and the shadowing and TA grades were recorded.

**Statistical Analysis**

The imaging data were evaluated by 12 physicians divided into 3 groups according to their clinical experience levels, and the final results were saved as raw data for further evaluation. To perform consistency comparisons, Kendall’s coefficient of concordance (Kendall’s W) was selected to evaluate the observer variability regarding the use of the shadowing/TA grading, which was defined as:

\[
W = \frac{12 \sum_{i=0}^{n} (R_i - \bar{R})}{m^2(n^3 - n)}
\]

in this study, where m is the number of observers (m = 12), n is the number of objects (n = 86), R is the grade evaluation made

<table>
<thead>
<tr>
<th>Imaging Mode</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>B mode</td>
<td>Gain = 70%; Frequency = 3.3 MHz; Depth: below the ureteral stone; MI = 0.9; TGC: median</td>
</tr>
<tr>
<td>Color Doppler mode</td>
<td>Color Gain = 21; Color-write priority = 100%; Depth: below the ureteral stone; WF = 739 Hz; PRF = 4.6 KHz; MI = 0.9; TGC: median</td>
</tr>
<tr>
<td>Computed tomography</td>
<td>120 kVp; 200 mA; Slice thickness = 3 mm; Field of view = 50 cm; Window width = 744; Window level = 301</td>
</tr>
</tbody>
</table>

**Table 1. Summary of scanning parameter settings**
by any observer, and \( \bar{R} \) is the mean grade of all observers for a specific object. The Kendall’s \( W \) is a nonparametric statistic used for assessing agreement among raters and is a normalization of the Friedman test statistic, ranging from 0 (no agreement) to 1 (complete agreement).\(^{21} \) Next, after stratification by the observer’s clinical experience level and stone size, the sensitivity and specificity of isolated shadowing/TA as an indicator for diagnosing ureteral stones were analyzed using paired \( t \) tests and one-way analysis of variance. All related comparisons were analyzed using SPSS software v. 25.0 (IBM, Armonk, NY), and a \( P \) value <.05 was considered statistically significant.

### RESULTS

The ultrasound images of 117 patients with suspicion of ureteral stones were included in the image evaluation process based on the inclusion criteria, and the demographics of the recruited patients and the stone characteristics are described in Table 3.

To compare the consistency in the observers’ evaluations between shadowing and the TA in patients with ureteral stones, all ureteral stones were graded by 12 physicians assigned to 3 groups according to their experience levels. The results of Kendall’s \( W \) were all confirmed to indicate significant consistency in the evaluation of both shadowing and the TA among all physicians and in different subgroups (\( P < .001 \)). More specifically, Kendall’s \( W \) for the shadowing vs TA evaluation procedures among the total physician group and the subgroups was 0.779 vs 0.859 (all physicians), 0.844 vs 0.878 (elementary), 0.789 vs 0.906 (intermediate), and 0.804 vs 0.871 (advanced). Furthermore, a paired \( t \) test of grade evaluation between shadowing and the TA showed that Kendall’s \( W \) for the TA grade was higher than that for the shadowing grade (\( P = .023 \)).

Among the total physician group’s image evaluations of all the included data, we found that all ureteral stones graded 2 and 3 for shadowing/TA were ultimately determined to have ureteral stones “present”, and those evaluated as grade 0 determined to be “absent”. However, there were some disagreements on the value of grade 1 shadowing/TA as an indicator of a ureteral stone. To further analyze the diagnostic performance, paired \( t \) tests of the sensitivity of using shadowing and the TA to identify ureteral stones showed that using the TA grade as an auxiliary tool was superior to using the shadowing grade among all the participating physicians and in each subgroup (\( P < .05 \)) (Table 4). However, regarding shadowing and the TA as indicators of ureteral stones, there was no significant difference in the sensitivity among the groups stratified by the observers’ clinical experience levels (\( P > .05 \)). Interestingly, based on the specificity evaluation, although there may be some small difference (0 or 1) in the grade classification in some conditions, no “present” results of the displayed shadowing/TA were found in the control group among all the observed physicians, and the specificity among all the groups reached 100%.

When the ureteral stones were stratified by the stone size into ≤5 mm, 5-10 mm, and ≥10 mm groups, one-way analysis of variance showed that with increasing stone size, the sensitivities of shadowing and the TA for diagnosing ureteral stones also increased (\( P < .001 \)). Furthermore, in groups with the same stone size, the paired \( t \)-test to evaluate the sensitivities of stratification using shadowing or the TA to identify ureteral stones showed that using the TA grade as an auxiliary tool was superior to using the shadowing grade (\( P < .001 \)) (Table 4).

### COMMENT

In this study, the diagnostic performance of the presence of shadowing/TA for detecting ureteral stones was evaluated based on a grading system. Notably, we first compared the observer variability in the use of acoustic shadowing and the TA for diagnosing ureteral stones and further confirmed the different roles of these 2 artifacts among physicians with different clinical experience. Images of these 2 artifacts under optimized scanning conditions for suspicious ureteral stones were acquired and retrospectively identified by 12 physicians. Statistically, the use of Kendall’s \( W \) to evaluate the consistency of the observers’ evaluations was reasonable and persuasive. Through variance analysis of the statistic in terms of the definition of Kendall’s \( W \), the corresponding evaluation of grade difference could be attributed to be due to the physicians (\( m \)), the included objects (\( n \)) and the evaluation procedures, which were referred to as \( 12 \sum_{i=0}^{n}(R - \bar{R}) \). Since all the ureteral stones were kept the

---

**Table 2.** Brief description of the grade evaluation standards for shadowing and the twinkling artifact

<table>
<thead>
<tr>
<th>Grade</th>
<th>Intensity</th>
<th>Margin</th>
<th>Visual Reference Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>1</td>
<td>Weak</td>
<td>Obsolete</td>
<td>Overwhelming the area behind the stone surface.</td>
</tr>
<tr>
<td>2</td>
<td>Stronger</td>
<td>Obvious</td>
<td>Overwhelming the area behind the stone surface.</td>
</tr>
<tr>
<td>3</td>
<td>Strongest</td>
<td>Obvious</td>
<td>Overwhelming the area behind the stone surface.</td>
</tr>
</tbody>
</table>

**Table 3.** Demographics of the included patients (\( n = 117 \)) and the stone characteristics of the ureteral stone patients (\( n = 86 \))

<table>
<thead>
<tr>
<th>Age (years, ( n = 117 ))</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤40</td>
<td>67</td>
</tr>
<tr>
<td>41-60</td>
<td>40</td>
</tr>
<tr>
<td>≥60</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stone size (mm, ( n = 86 ))</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤5</td>
<td>13</td>
</tr>
<tr>
<td>5-10</td>
<td>63</td>
</tr>
<tr>
<td>≥10</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stone location</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left/right</td>
<td>46/40</td>
</tr>
<tr>
<td>Upper ureter</td>
<td>32</td>
</tr>
<tr>
<td>Middle ureter</td>
<td>10</td>
</tr>
<tr>
<td>Lower ureter</td>
<td>44</td>
</tr>
</tbody>
</table>

---
same in this study, it was easy to conclude that Kendall’s W reflected the difference in the physicians’ subjective viewpoints and that using it to confirm the consistency of the evaluations was statistically logical. Our results strongly suggest that evaluating for the presence of the TA to diagnose ureteral stones is preferable to using shadowing.

In clinical practice, hyperechogenic regions associated with typical artifacts (such as shadowing and TA) may be useful for physicians to confirm the presence of ureteral stones as quickly as possible. Previously, shadowing below the hyperechogenic region had always been considered a useful indicator for the presence of a urinary stone; however, the presentation of shadowing is significantly affected by many factors, such as stone size, thickness, the nature of the tissue that is occupied, the space between the transducer and the stone, and the level of the preset focal zone.22 For these reasons, a relatively low sensitivity (overall: 45%) for ultrasonography had been found for the diagnosis of urinary stones in a meta-analysis.6 However, if the TA was included, the sensitivity of ultrasonography could be significantly improved. For example, Ripollès et al showed that the sensitivity using color Doppler ultrasound mode reached 90%.7 Likewise, Winkel et al attained a high sensitivity of 81% when B-mode and color Doppler ultrasonography were both used separately and in combination for detecting stones in the urinary tract.11 Recently, Sen et al found that almost 86.8% of ureteral stones could be detected using the TA identified on ultrasonography; thus, ultrasound was considered a good and safe imaging modality with comparable results to CT.18 In our study, according to the comparisons between the use of shadowing and the TA as indicators of ureteral stones, the sensitivity of ultrasonography increased from 57% (52%-62%) to 85% (82%-88%) overall. Furthermore, the same trend was present in each subgroup analysis (Table 4). Different from previous studies, we demonstrated that the sensitivity using the TA was higher than that of using shadowing and was not related to the clinical experience level. Additionally, the above meta-analysis regarding the value of ultrasonography also showed that the specificity of ultrasonography was very high (range: 87%-100%, overall: 94%),6 which was in accordance with our study. Interestingly, using shadowing or the TA as an indicator, we found that it was hard to identify false positive cases using sequential scanning mode if only ureteral stones were included. Similarly, Ripollès et al reported that using the TA to detect ureteral stones could achieve 100% specificity.7 We speculated that timely scanning with CT and rare calcification formation around the ureter may be 2 important reasons for the high specificity of the TA for the diagnosis of ureteral stones. Altogether, although we believe that shadowing and TA may both be important for recognizing ureteral stones with perfect specificity, the use of the TA was superior to that of shadowing in terms of the diagnostic sensitivity.

Stone size may be another important factor in the diagnosis of the ureteral stone. Furthermore, a positive correlation between the stone size and clarity of relevant acoustic shadowing was found previously.13 In addition, since the TA was reported, the role of stone size on the presence of the TA has also been investigated and evaluated. Mitterberger et al24 and Winkel et al11 could not find any correlation between the stone size and presence of the TA; however, Louvet16 first found that the TA was dependent on stone size, with larger stones producing higher artifact grades. Furthermore, Sorensen et al and Sen et al18 reported that when the stone size increased, the sensitivity of the TA on color Doppler US also increased. Our results showed that the sensitivity was related to the stone size in a positive trend for both shadowing and the TA.

Focusing on the grade evaluation process in this study, all the observers in this study evaluated the grade scores and confirmed the final impressions regarding shadowing/T A in a subjective way, and disagreements regarding grade 1 findings among observers were considered the source of the difference in sensitivity. We also found that most of the grade 1 shadowing cases were considered to have “absent” ureteral stones, while most of the grade 1 TA cases could be easily differentiated visually and yielded “present” results for ureteral stones. Owing to the fact that the intensities of grade 1 shadowing and TA are so faint, recognition on images would be affected by the differentiation ability of the human eye and the complex physical conditions during scanning. More specifically, the reasons for this difference may be as follows: First, our eyes are more sensitive to color pixels than to grayscale.

<table>
<thead>
<tr>
<th>Group</th>
<th>Shadowing (95% CI)</th>
<th>TA (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical experience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>54% (49%-59%)</td>
<td>83% (79%-87%)</td>
<td>.002</td>
</tr>
<tr>
<td>Intermediate</td>
<td>57% (45%-68%)</td>
<td>83% (81%-84%)</td>
<td>.008</td>
</tr>
<tr>
<td>Advanced</td>
<td>61% (43%-80%)</td>
<td>89% (78%-99%)</td>
<td>.018</td>
</tr>
<tr>
<td>stone size (mm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5</td>
<td>30% (21%-39%)</td>
<td>66% (62%-70%)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>5-10</td>
<td>59% (54%-63%)</td>
<td>87% (83%-90%)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≥10</td>
<td>81% (73%-88%)</td>
<td>97% (99%-100%)</td>
<td>.001</td>
</tr>
<tr>
<td>Total (n = 86)</td>
<td>57% (52%-62%)</td>
<td>85% (82%-88%)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Cl, confidence interval; TA, twinkling artifact.
pixels in images. Regarding the findings for the region of interest, a grayscale difference must be sufficiently large to attract the attention of the sonographer due to the physiological characteristics of the human eye. While the intensity of grade 1 TA was not very strong, the obvious confirming value for the identification of a ureteral stone was still present for nonlinear color perception in humans. Second, although the related parameter settings have been improved, a spatial compounding effect creating shadows at different angles that are averaged can decrease the visibility of posterior acoustic shadowing. In contrast to shadowing, the presence of the TA had no relation to a partial volume effect and was independent of the body mass index of the patient.

There are some limitations to this study. First, the review of the images was performed retrospectively. Although all the patients recruited were consecutively selected, and the scanning machine settings were optimized, some influencing factors may have affected the display and evaluation of shadowing and the TA in an unclear manner. Second, although all the data were prospectively collected by a sonographer with optimized settings, selection bias could not be completely eradicated. In addition, the evaluation processes were performed using only still images under the optimized conditions for shadowing/TA rather than with videos. Third, there was a lack of false positive cases in this study. From our clinical experience, the only false positive cases may be due to ureteral lumen calcification, which is very rare, especially with our limited sample size. More data need to be collected in further prospective studies on acoustic shadowing and the diagnostic efficacy of TA. Finally, performing the sequential scanning methods was affected by some factors such as the patient’s body mass index, abdominal conditions, and operator’s scanning skills.

In summary, shadowing and the TA are both important artifacts for diagnosing ureteral stones. As an indicator of ureteral stones, subjective evaluation of the TA is more consistent and has better diagnostic sensitivity than shadowing, and this trend had no relation to the standards stratified by clinical experience level or stone size. In particular, different recognition of grade 1 shadowing/TA was found to be the source of the differences in diagnostic sensitivity. Additionally, the diagnostic sensitivity was also found to have a positive proportional relationship with the stone size stratification. All these results of our study may provide useful, complementary, and quantitative information on the use of shadowing and the TA in clinical practice.

Acknowledgment. The authors thank Jin Zhu, Longlong Tian and American Journal Experts (AJE) for providing assistance with manuscript preparation and editing. This study was supported by the Pre-Research Fund for Young Investigators of The Second Affiliated Hospital of Soochow University (SDFEYQN1702), the Superior Speciality of The Second Affiliated Hospital of Soochow University (XKQ2015009), the Natural Science Foundation of Jiangsu Province (BK20161222) and Science and Technology Development Planning of Shandong Province, China (2012GSF11847).

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.urol ogy.2019.01.021.

References