Original Article

Associated ultrasound findings improve the accuracy of twinkling artifacts in kidney stone diagnosis

Rujiluck Rojthamrong, M.D.(1)(2)
Kewalee Sasiwimonphan, M.D.(1)(2)
From (1) Department of Radiology, Faculty of Medicine, Chulalongkorn University, (2) King Chulalongkorn Memorial Hospital the Red Cross Society, 1873 Rama IV Road, Pathumwan, Bangkok, Thailand.
Address correspondence to K.S. (e-mail: kewalees@gmail.com)

Abstract

**Background:** Twinkling artifact (TA) in color Doppler ultrasound is commonly used as a sign of kidney tract stone detection but the accuracy is limited as compared with unenhanced computed tomography (CT).

**Objective:** Define the associated ultrasound findings that may improve the accuracy of TA compared with CT for diagnosing kidney stones.

**Materials and Methods:** Prospective study was conducted on 128 TAs in patients sent for unenhanced CT KUB and performed color Doppler ultrasound on the same day. TA sizes and associated sonographic signs were recorded and analyzed with receiver operating characteristic curves (ROCs). The diagnostic reference was the CT scan.
Results: There was a total of 128 TAs with the size of 3.95 mm (2.7-6 mm). Only 30 TAs showed as kidney stones in CT. The sizes of kidney stones in CT were 5.4 mm (3.4-6.4 mm) which represented a significant difference in TA size \((P = 0.002)\). ROC curve analysis showed that 5 mm would be the optimal size of TA for kidney stone predictions. Other significant signs for improved diagnosis include echoic foci \((P = 0.039)\), posterior shadows \((P = 0.001)\), long TA tails \((P = 0.001)\) and 2nd approach TA \((P = 0.001)\). Then a predictive AT model was created to predict kidney stones, which moderately improved diagnosis accuracy for kidney stones with good agreement.

Conclusion: The combination of TA and other sonographic signs are moderately associated with kidney stone diagnosis including TA size (> 5 mm), posterior acoustic shadow, long TA tail, junctional line location and focal Caliectasis.

Keywords: Twinkling artifact, Color doppler ultrasound, Kidney stones, Renal calculi.

Introduction

Kidney stones are one of the most frequent concerns of people referred to emergency centers, with an occurrence likelihood of 12% for males and 6% for females [1]. Ultrasound (US) and computed tomography (CT) scans are widely used as imaging modalities to diagnose the disease [2]. CT scans without intravenous contrast administration are widely accepted as the reference standard imaging technique to confirm or deny the diagnosis of urinary calculi. Despite the high sensitivity and specificity values of CT, the cost-effectiveness ratio, high radiation doses and motion artifact disadvantages of this application warrant the development of alternative methods with similar reliability.
Recently, there has been growing awareness of the overuse of CT and the associated radiation effects arising from the evaluation of patients with acute flank pain [3]. The development of US probe technologies and US imaging techniques has improved the sonographic image quality and enhanced the utility of US in diagnosing urinary tract calculi.

In 1996, Rahmouni et al. [4] defined a “twinkling” artifact as a color Doppler (CD) artifact generated by a strongly reflecting medium: [4] the result was a rapidly changing mixture of red and blue color signals arising from a point behind a stationary object, which represents a substantial improvement for the sonographic detection of stones [5-6]. The TA on color Doppler US could be a good and safe alternative imaging modality with results comparable to those from noncontrast-enhanced CT for the sensitive detection of urolithiasis <5 mm. The initial study showed that the sensitivity, accuracy, and positive predictive values of TA for the detection of calculus were 94%, 94%, and 100%, respectively [7]. However, a study in 2016 found that isolated sonographic twinkling artifacts have a high false-positive rate (60%) for the diagnosis of renal calculi [8]. There are also conflicting results regarding the sensitivity and specificity of twinkling artifacts. Some studies have reported that the accuracy of the artifact depends on both the setting of the device and the shape of the stone. It has been argued that this artifact is also observed in many parts of the kidney where no stones exist [9].

In considering the size of the detected stone, Aytac and Ozcan [5] concluded in 1999 that TA helps differentiate a small stone from other small echogenic structures. Studies such as those of Yavuz et al. [10] and Hanafi et al. [7] have revealed that the TA has been a useful CDUS tool for detecting small urinary stones. However, Masch et al. [8] showed that an isolated TA has a high false-positive rate (60%) for the diagnosis of renal calculi in patients without known urolithiasis. The specificity and positive likelihood ratio for the diagnosis of renal calculi improved but declined in sensitivity when additional diagnostic features such as sonographic twinkling artifacts and echogenic foci were present. A limitation of this study was the retrospective design, which involved recruitment of the study population based on finalized abdominal ultrasound reports containing
the words “twinkle” or “twinkling” in reference to suspected urinary tract calculus; this delayed the collection of data between the index ultrasound and the reference standard CT in the study (mean delay of 8 days).

The sensitivity and specificity values from prior studies are still inconclusive. Most of the prior studies are retrospective studies that used small sample sizes and multiple confirmation methods, such as surgery, computed tomography or the collection of a stone extracted by the patient during micturition.

The associated ultrasound findings of the TA in CDUS compared to CT results for the evaluation of kidney stones have not been well established.

Our hypothesis was that associated findings with twinkling artifacts might have been significantly related to accuracy in kidney stone diagnosis. The purpose of our study is to assess the correlation between sonographic TA findings and CT results for the diagnosis of kidney stones.

Imaging findings of TAs associated with renal stones have been documented in the literature. In some cases, grayscale US simultaneously shows multiple findings, enabling confident diagnosis of stones. However, in some cases, grayscale US and CDUS findings are equivocal or inconclusive, but there is high clinical suspicion of renal disease. Because the diagnosis of stones from US is not always certain, a clear way of expressing the likelihood of renal stones is needed. This can be achieved by linking imaging findings of TA with scores categorizing imaging findings as positive, negative, or indeterminate for diagnosing renal stones.

In this study, we introduce a more systematic approach to use TA for the diagnosis of stones. This article represents a feasibility study of our standardized reporting format system using a cohort of patients who underwent CT of the abdomen within 6 hours before or after an US examination. We undertook this study to assess the diagnostic accuracy of the system and the usefulness of specific imaging findings, and to evaluate its reproducibility.
Materials and methods

Patients
Ethics approval was provided by the institutional review board, the Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand (No. 617-62). A written consent was obtained from all study participants. From August 2019 to September 2020, we prospectively studied patients in our hospital who were suspected of having kidney stones. Thereafter, 40 patients underwent a CT scan when the subsequent ultrasound was performed on the same day, which did not result in a known CT result at the time of the ultrasound.

Sonographic technique
On the same day and after the CT scan, the patients directly underwent a limited sonographic scan of the kidneys (LOGIQ™ E9, GE Healthcare, Milwaukee, WI, USA). The examination was performed by two radiologists using a curved low-frequency probe (1-6 MHz), color frequency 2.5 MHz, Gain 60-70 and the focus zone is located below the suspicious lesion and a high pulse repetition frequency with which the machine's scale was greater than 60 cm/s [11]. The pulse repetition frequency is defined as the number of pulses sent per second. We used a high pulse repetition frequency (PRF) to exclude aliasing artifacts arising from small vessels. This is a key technical factor for identifying and documenting twinkling artifacts. The twinkling artifacts documented rapidly changing color mix immediately behind suspect focal lesions on Color Doppler ultrasound [4]. The radiologists in this study had 3 and 13 years of experience and were not aware of the CT findings at that time.

The kidneys were evaluated with both grayscale and color Doppler imaging. TAs were detected, and associated findings were observed. The area at the interface between the renal sinus and renal cortex was defined as the junctional line, and this is shown in Figure 1. The length of the TA was defined as the tail of the TA if it was longer than its width, as shown in Figure 2. Once a TA was found, it was again detected in different US approaches and was categorized as a 2nd approach TA.
The radiologists filled out a standardized form indicating factors such as size, location, presence or absence of the long tail of the TA, the presence of a posterior acoustic shadow, evidence of hydronephrosis, and areas of high echogenicity with associated shadowing on grayscale images.

Figure 1. Junctional line defined as the area of the interface between the renal sinus and renal cortex (dashed line).

Figure 2. TA image with a length longer than its width was defined as the tail of the TA (arrow).
Computed tomography technique
CT examinations were performed on MDCT scanners (Discovery CT 750 HDCT, GE Healthcare, Milwaukee, WI, USA) with a slice thickness/increment of 2.5/1.0 mm and a 120-kV tube potential. Tube current modulation was applied. Patients were scanned from the diaphragm to the pubic bone.

Data analysis
After independent evaluation of the CT scan and Doppler US of stones, the result of each imaging modalities were compared. The data were analyzed using STATA version 15.1. Continuous variables are presented as the mean ± SD or median (IQR) and n (%). Categorical data are reported as frequencies and proportions. P-values correspond to independent t tests or Mann-Whitney tests and chi-square tests. There was the AUC analysis defining the optimal cutoff size of the TA. Univariate and multivariate renal stones were evaluated using logistic regression analysis as well as model validation. The significance level was set at 0.05 for all statistical tests.

Predictive model
Our predictive model was based on a significant relationship existing between objective factors of the TA and a stone diagnosis in CT. The model consists of a scoring of imaging findings previously described to have an association with the presence or absence of kidney stones [5, 11, 13]. The objective imaging findings convey the radiologist's decision-making process to the clinician, and the final score reflects the corresponding implications for patient management.
Results

A total of 128 TAs was identified and 46 TAs (35.9%) were found in male patients. The median size of the TAs was 3.95 mm (2.7 to 6 mm) in color Doppler US. There were only 30 TAs found as kidney stones in the CT scan. The median size of the renal stones was 5.4 mm (3.4 to 6.4 mm) in CT and 2.8 mm (2.3 to 6.6 mm) in color Doppler US. Echoic foci were detected in 84 TAs (65.6%). A posterior acoustic shadow was detected in 11 TAs (8.6%). The long tails and 2nd approach TA were detected in 65 (50.8%) and 108 (84.4%) TAs, respectively. A significant difference was found between the size of the TA and the presence of stones in CT ($P = 0.002$). The table 1 described patients’ characteristics in findings of TAs.

Table 1. Characteristics of identified TAs (n=128).

<table>
<thead>
<tr>
<th>Variable</th>
<th>n (%), mean ± SD, or medians with IQRs*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>46 (35.9%)</td>
</tr>
<tr>
<td>BMI (ml/min/1.73 m²)</td>
<td>23.22 ± 3.4</td>
</tr>
<tr>
<td>TA size (mm)</td>
<td>3.95 (2.7-6)</td>
</tr>
<tr>
<td>Location: Cortex</td>
<td>21(16.4)</td>
</tr>
<tr>
<td>Location: Junctional line</td>
<td>72(56.3)</td>
</tr>
<tr>
<td>Location: Renal sinus</td>
<td>35(27.3)</td>
</tr>
<tr>
<td>Echoic foci</td>
<td>84(65.6)</td>
</tr>
<tr>
<td>Posterior acoustic shadow</td>
<td>11(8.6)</td>
</tr>
<tr>
<td>Caliectasis</td>
<td>7(5.5)</td>
</tr>
<tr>
<td>Stone size in CT</td>
<td>4.5(3.4-6.4)</td>
</tr>
</tbody>
</table>

*interquartile ranges
In Table 2 we describe the factors that significantly affected the accuracies of TA diagnoses, including echoic foci ($P = 0.039$), posterior acoustic shadow ($P = 0.001$), long TA tail ($P = 0.001$), location at the junctional line ($p=0.028$), caliectasis ($p=0.013$) and 2nd approach TA ($P = 0.001$). ROC curve analysis of the sizes of TAs (Figure 3) and the presence of stones in CT showed that 5 mm would be the optimal size for predicting renal stones. Additionally, the results showed that the location in the renal cortex was significantly related to no stone on CT scan and 35 TAs at the renal sinus indicated 10 stones that were confirmed with CT.

**Table 2.** Associated findings that significantly supported the diagnosis of renal stones from TA findings (total n=128).

<table>
<thead>
<tr>
<th>Stone-predicting factors</th>
<th>CT result positive (n=38)</th>
<th>CT result negative (n=90)</th>
<th>Odds ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA Size (mm)</td>
<td>5.45(3.3-6.6)</td>
<td>3.5(2.4-4.8)</td>
<td>1.19(1.04-1.37)</td>
<td>0.01</td>
</tr>
<tr>
<td>Longer tail of TA</td>
<td>28(73.7%)</td>
<td>37(41.1%)</td>
<td>4.01(1.63-10.32)</td>
<td>0.001</td>
</tr>
<tr>
<td>Posterior acoustic shadow</td>
<td>8(21.1%)</td>
<td>3(3.3%)</td>
<td>7.73(1.68-47.26)</td>
<td>0.001</td>
</tr>
<tr>
<td>Caliectasis</td>
<td>5(13.2%)</td>
<td>2(2.2%)</td>
<td>6.67(1.01-71.97)</td>
<td>0.013</td>
</tr>
<tr>
<td>Echoic foci</td>
<td>30(78.9%)</td>
<td>54(60%)</td>
<td>2.5(0.97-7)</td>
<td>0.039</td>
</tr>
<tr>
<td>Location at junctional line</td>
<td>27(71.1%)</td>
<td>45(50%)</td>
<td>2.45(1.02-6.14)</td>
<td>0.028</td>
</tr>
<tr>
<td>Location at renal cortex</td>
<td>1(2.6%)</td>
<td>20(22%)</td>
<td>0.09(0-0.65)</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Figure 3. Receiver operating characteristic (ROC) curves for the diagnosis of renal calculus by color Doppler TA and echoic foci. TA, AUC=0.679 (95% CI 0.579-0.778); echoic foci, AUC=0.671 (95% CI 0.568-0.776).

The false positive ratio for TAs in our study was approximately 0.5, and the false positive ratios for locations at the junctional line, renal sinus and renal cortex were 0.5, 0.28, and 0.22, respectively. However, the high false positive rate for the location at the junctional line was accompanied by the highest sensitivity, approximately 0.71, as compared to approximately 0.26 and 0.03 for the renal sinus and renal cortex, respectively.

From this ROC curve, the optimal cutoff TA size was 5 mm with a sensitivity of 52.6%, specificity of 78.9%, PPV of 51.3%, NPV of 79.8%, and accuracy of 71.1% (p=0.001). There were 5 variables that significantly predicted the diagnosis of the kidney stones when evaluating the TA as shown in Table 3. We proposed a predictive model based on the use of these factors to rate each TA in order to improve the precision of the diagnosis. We found that the use of 5 variables or the use of 4 variables (excluding caliectases because of the extremely low prevalence) did not exhibit significantly different diagnostic values; therefore, we suggested the use of 4 predictive variables to create a predictive model and show how to use the predictive model in some cases in Figure 4.
**Table 3. Significant predicting factors.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>Adjusted OR</th>
<th>95.0% C.I. for OR</th>
<th>P-value</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA Size (&gt;5 mm)</td>
<td>1.016</td>
<td>2.761</td>
<td>1.10-6.92</td>
<td>0.030</td>
<td>1</td>
</tr>
<tr>
<td>Posterior acoustic shadow</td>
<td>1.586</td>
<td>4.884</td>
<td>1.06-22.44</td>
<td>0.042</td>
<td>2</td>
</tr>
<tr>
<td>Long tail of TA</td>
<td>0.935</td>
<td>2.548</td>
<td>1.01-6.41</td>
<td>0.047</td>
<td>1</td>
</tr>
<tr>
<td>Location at junctional line</td>
<td>1.032</td>
<td>2.806</td>
<td>1.12-7.01</td>
<td>0.027</td>
<td>1</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.576</td>
<td></td>
<td></td>
<td></td>
<td>Total 5</td>
</tr>
</tbody>
</table>

**Figure 4.** A 51-year-old man with an example of a positive result from the predictive model of nephrolithiasis (A and B). CDUS image (A) shows a TA in the mid pole of the left kidney at the junctional line, with a size of 6.6 mm and a long tail of the TA. The total score of this TA is 3, positive. The unenhanced axial CT image (B) obtained on the same day shows renal stones at the mid pole of the left kidney. In contrast, a 47-year-old woman with an example of a negative result from the predictive model (C and D). The CDUS image (C) shows a TA in the lower pole of the right kidney at the junctional line, with a 4.0 mm size. No demonstrable posterior shadow or long tail of the TA was noted. The total score of this TA is 1, representing a negative result. The unenhanced axial CT image (D) obtained on the same day shows no calculus.
Then, we additionally collected TAs (53 TAs in total) for model testing. The predicted probabilities for scores=3, 4, and 5 were 60%, 72.7% and 88%, respectively as shown in Table 4. The cutoff scores of 3 or above were positive, with a sensitivity of 75%, specificity of 88%, PPV of 87.5%, NPV of 75.9% (p=0.05) and AUC of 0.82 (95% CI 0.71-0.92). The agreement with the model is 81.13%.

**Table 4. Results of testing the predictive model.**

<table>
<thead>
<tr>
<th>Score model</th>
<th>Predicted probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7.1%</td>
</tr>
<tr>
<td>1</td>
<td>17.2%</td>
</tr>
<tr>
<td>2</td>
<td>35.1%</td>
</tr>
<tr>
<td>3</td>
<td>60.0%</td>
</tr>
<tr>
<td>4</td>
<td>72.7%</td>
</tr>
<tr>
<td>5</td>
<td>88.0%</td>
</tr>
</tbody>
</table>

**Discussion**

The TA is a complex phenomenon. Described for the first time in 1996 by Rahmouni et al. [4], it consists of an intense color signal alternating between red and blue from behind some structures [4]. Two theories have been proposed to explain TAs. The first one was offered by Rahmouni et al. [4], who suggested that this artifact is generated by a strongly reflecting medium with a rough interface. When the incidental beam is reflected on the rough interface, the acoustic wave is split into a complex beam pattern caused by multiple reflections in the medium, resulting in a prolonged pulse duration of the transmitted sound signal; the Doppler units interpret this result as movement and thus assign it different colors. The second theory was offered by Kamaya et al. [13], who stated that the TA is caused by a narrow band of intrinsic sonographic machine noise, referred to as phase or clock jitter, which may be generated by slight random time fluctuations in the path lengths of transmitted and reflected acoustic waves [13].
The TA is a diagnostic sign for the reliable evaluation of stones in the urinary tract. Studies [2, 7, 10, 14] have revealed that twinkling artifacts are a useful color Doppler ultrasound tool for the detection of small urinary stones (smaller than 5 mm). However, a high false-positive rate, similar to that in our study, was noted in the study of Dillman et al. [15], with an overall false positive rate of approximately 50%.

A recent meta-analysis study [16] resulted in a pooled sensitivity and specificity in the diagnosis of urolithiasis, with ultrasonographic TA signs of 88.16% [95% confidence interval (CI): 87.07–89.19%] and 79.22% [95% CI: 73.41–84.26%], respectively. There was significant interstudy heterogeneity. The current body of evidence suggests that TAs may be useful as a complementary tool in the diagnostic workup of patients with suspected urolithiasis.

Louvet [17] first found that the TA was dependent on the stone size, with larger stones producing higher artifact grades. Furthermore, Sen et al. [18] reported that when the stone size increased, the sensitivity for the TA in color Doppler US also increased. Our study showed similar results: TA size was significantly related to detected stones in CT, with a moderate ROC area (AUC=0.679, 95% CI 0.579-0.778); the optimal cutoff TA size was 5 mm with a sensitivity of 52.6%, specificity of 78.9%, PPV of 51.3%, NPV of 79.8%, and accuracy of 71.1% (p=0.001).

Simon et al. [19] defined the tail of the TA as “twinkle power”, which supports the theory that microbubbles are present on kidney stones in humans [20, 21]. The stable crevice bubble theory is consistent with previous data that indicate an association between the surface roughness of a stone and the presence of twinkling, as a rougher surface would provide more crevices in which bubbles could form. This is probably a reason why stones usually produce more TA tails (more twinkle power) than other renal calcifications. We found that this associated finding is a positive predictor of kidney stones.
Regarding the location of the TA, the nature of the renal stone location is in the collecting system; thus, a TA located in the renal cortex does not signify a stone. Our study showed that the location in the renal cortex is significantly related to negative CT results. However, there is a TA in the renal cortex, which is a stone in CT due to diffusely increased parenchymal echogenicity with loss of corticomedullary differentiation, in this case, from a chronic kidney disease making it difficult to locate the actual location of this TA.

Our kidney stone predictive model was created to increase accuracy in the diagnosis of kidney stones especially small stones (subcentimeter) in sizes and to improve communication between radiologists and clinicians. The objective checklist includes imaging findings of TAs that are moderately to highly associated with kidney stones. In our study, TA size (≥5 mm), posterior acoustic shadow, long TA tail, junctional line location and Caliectasis each had a statistically significant association with renal stones in CT. The final score has a good confirmational value and expresses the reader’s degree of certainty regarding the presence or absence of kidney stones based on the TA. Communication of objective findings in a radiology report serves two purposes. First, radiologists are more accurate when systemically weighing the presence or absence of specific imaging findings before drawing conclusions. Second, a clear list of findings documents the information the radiologist used to determine the final likelihood of renal stones, allowing the referring clinician or surgeon to understand the radiologist’s decision-making process. There were limitations to our study. First, our research involved a small number of patients. Second, our study population was restricted to patients who was scheduled a CT scan from clinical indication of the stone disease which can cause selection bias. Third, as reported by Aytac and Ozcan [5], the twinkling sign depends on the color sensitivity and the acoustic output of the US unit. Therefore, with different US units, these results might not be reproducible.
Conclusion

In summary, many ultrasound findings with TAs are moderately to highly associated with the presence of kidney stones including the size of TA ≥ 5mm, posterior acoustic shadow, long TA tail, junctional line location and caliectasis. Our TA predictive model for kidney stones moderately improved the accuracy in identifying cases of renal stones with good agreement.

Future studies with larger patient populations will guide the development of the system and will likely provide a sufficiently large dataset for multivariate analysis of kidney stone-specific imaging findings. For now, our data suggest that the likelihood of kidney stones can be more confidently based on the presence of specific imaging findings.

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
The authors wish to thank the patients who participated in this study.
References


