# First impressions count: How imaging order impacts ultrasound accuracy in renal stones assessment

Kai-Ho Fok, Lara Gabrielle L. Lim, Deelan Patel, Hui Ming Lin, Michael Ordon, Errol Colak, Guan Huang
Unity Health Toronto - St. Michael's Hospital, Toronto, ON, Canada

**Cite as:** Fok K-H, Lim LGL, Patel D, et al. First impressions count: How imaging order impacts ultrasound accuracy in renal stones assessment. *Can Urol Assoc J* 2026 November 20; Epub ahead of print http://dx.doi.org/10.5489/cuaj.9721

Published online November 20, 2026

**Corresponding author:** Dr. Kai-Ho Fok, Unity Health Toronto - St. Michael's Hospital, Toronto, ON, Canada; Kai.fok@utoronto.ca

***

## ABSTRACT

**Introduction:** Accurate detection and measurement of renal stones guides clinical decision-making, as stone size predicts spontaneous passage and appropriate surgical management. Non-contrast computed tomography (NCCT) is the reference standard, but ultrasonography (US) is increasingly used for its accessibility, lack of ionizing radiation, and cost. We evaluated US performance at our tertiary referral center, assessing the impact of imaging sequence (NCCT first or US first) on diagnostic accuracy.

**KEY MESSAGES**

- US overestimates the size and presence of stones compared to NCCT.
- Specificity of stone detection on US may be improved by having a prior NCCT.
- Mimickers of stones on US include renal vascular calcifications, milk of calcium cysts, and sloughed papillae.
- Overreliance of twinkle artifact may increase false positive US stone detection rate.

**Methods:** A single-center, retrospective study was performed using data from January 1, 2017, to September 30, 2022. Patients who underwent both an US and NCCT within 15 days of each other were included. Each stone was matched between modalities. Descriptive statistics, sensitivity, and specificity of US stone detection were calculated. A qualitative review of missed and overcalled stones was also performed.

**Results:** A total of 206 patients with a total of 368 stones were identified on NCCT, with 121 patients having US first. Overall, sensitivity and specificity of US were 64.4% and 40.3%, respectively. When US was performed first, sensitivity was 66.7% and specificity dropped to


© 2026 Canadian Urological Association

18.0%. In contrast, when NCCT was performed first, sensitivity was 58.7% and specificity increased to 76.4%.
**Conclusions:** In our cohort, US sensitivity and specificity in detecting renal stones were consistent with prior reports. We demonstrate that the specificity of US can be improved if a NCCT is done first. Stone size tends to be overestimated on US, especially for smaller stones.

## INTRODUCTION

Accurate detection and measurement of renal stones are critical for guiding clinical management, as stone size is a major determinant in predicting spontaneous passage and selecting appropriate treatment strategies. Imaging plays a central role in the diagnosis and monitoring of disease burden. Non-contrast computed tomography (NCCT) remains the reference standard for renal stone evaluation due to its near-perfect sensitivity (97–100%) and specificity (>97%).[1,2] However, concerns regarding radiation exposure, cost, and availability have led to increased interest in ultrasonography (US) as an alternative, particularly in specific populations where ionizing radiation or access to NCCT is a concern.

Several studies have shown that US tends to overestimate stone size, especially for stones under 5 mm.[3-6] This average overestimation of about 1.8 mm has important clinical implications, as treatment recommendations between conservative and surgical management are closely tied to stone size and location. Rates of spontaneous stone passage drop significantly for every millimetre of increased stone size beyond 5 mm. As a result, the American Urological Association guidelines caution the use of US as a single modality to recommend treatment to patients.[7] Under the simplified assumption that stones ≤4 mm are managed conservatively and those >4 mm warrant intervention, studies have found that 14 to 75 percent of patients might be offered treatment based on US alone, while NCCT would suggest observation.[5,6]

The sensitivity of US for detecting renal stones varies widely, with reported rates ranging from 24% to 78.9%, depending on factors such as stone size, location, and operator experience.[3-6,8-10] Specificity tends to be high, greater than 90%, but with some reporting quite low which might be a result of methodological differences.[10] Detection accuracy improves with increasing stone size and varies by anatomical location. For example, Kanno et al. (2014) reported significantly lower detection rates for stones in the upper pole calyces.[9] Operator experience also plays a role in detection performance, with more experienced sonographers achieving higher accuracy.[6] Despite these challenges, US remains an essential diagnostic tool in many clinical contexts.[11]

Nearly a decade has passed since most large-scale studies evaluating the performance of US were published, prompting us to investigate whether advances in US technology or technique have mitigated its known limitations in stone detection. Specifically, this study aims to assess the accuracy of US in detecting and measuring renal stones, using NCCT as the reference standard,


© 2026 Canadian Urological Association

in a large, real-world cohort. Furthermore, we investigate whether the order in which imaging is performed (NCCT first or US first) influences stone detection and measurement. Although this factor has been mentioned in prior literature, it remains poorly characterized.

## METHODS

A single centre retrospective cohort study was performed using data between January 1, 2017, and September 30, 2022. Patients who underwent both an US and NCCT for renal stones within 15 days of each other were identified from the radiology information system and included. The 15-day window between studies was chosen to capture patients with images where their stones are unlikely to have changed in the interval. This window is consistent with prior literature with capture windows varying between same day and 90 days.[6,12] The earliest pair of US and NCCT was included for patients who had multiple pairs (n = 25 were excluded on this basis). US and NCCT studies with and without renal stones were included. Renal and ureteral stones were present in the studies but only renal stones were analyzed. To mitigate potential biases introduced by reader or operator fatigue, patients who had more than 8 stones were excluded (n = 34). In such cases, the clinical focus would typically be on overall stone burden rather than individual stone detection, and precise mapping would be less feasible or clinically relevant. The study was approved by our institutional research ethics board.

NCCT were performed on a 64-slice multidetector computed tomography scanner (Revolution, LightSpeed 64, or Optima 64, General Electric Medical Systems, Milwaukee, WI) with soft tissue window images reconstructed at 3 × 3 mm for axial series and 5 × 5 mm for the coronal and sagittal series. The scan coverage is from above the superior margin of the kidneys to the inferior margin of the symphysis pubis.

Renal ultrasound examinations were performed by a team of 16 sonographers with experience ranging from 5 to 10 years. Examinations were performed on either a Canon Aplio i500 (Canon Medical Systems, Tochigi, Japan) or Philips IU22 system (Philips Healthcare, Andover, MA, USA), using a 6MHz curvilinear transducer supplemented by a 10 MHz curvilinear transducer where body habitus permitted. Examinations followed a standardized imaging protocol as follows: Patients were scanned in supine and left lateral decubitus positions, with intercostal and oblique windows as required. The departmental imaging protocol required real-time scanning of both kidneys, with acquisition of sagittal and transverse images covering the upper, mid, and lower poles. When a stone was identified, it was documented in 2D grayscale and with colour Doppler and measured with calipers. Hydronephrosis, when present, was documented in 2D and colour Doppler with calipers obtained in the transverse plane at the renal hilum.

A fellowship trained abdominal radiologist reviewed patient charts and imaging, documenting the laterality and anatomical location (renal pole) of each stone. Measurements were made in all three planes and stone size was determined using the longest axis on both NCCT and US for each stone. Corresponding stones were matched between modalities based on


© 2026 Canadian Urological Association

location and size for comparison. Difficult or ambiguous cases were adjudicated by a second senior fellowship-trained abdominal radiologist.

Descriptive statistics are reported as mean with standard deviation or median with interquartile range. These were compared using Wilcoxon signed rank test for paired samples and two tailed t-tests for unpaired samples. Sensitivity and specificity were calculated at the per-stone level. A true positive required a matched stone in the same location on both modalities and a false negative was a stone on NCCT without a corresponding stone on US in that location. A false positive was a US-reported stone without a corresponding NCCT finding. A true negative was a study pair that was stone-free on both modalities. Two-by-two contingency tables are shown for sensitivity and specificity calculations. These were stratified by stone location and side.

To assess clinically relevant stone detection, a simplified scenario was applied consistent with prior studies. In this scenario, stones ≤4 mm are recommended observation and stones >4 mm are recommended treatment. The sensitivity and specificity of US for detecting clinically relevant stones were calculated and incorrect recommendations were quantified (i.e. recommending treatment when observation was indicated, and recommending observation when treatment was warranted).

To evaluate our hypothesis that study order impacts US detection, sensitivity and specificity were measured separately for study pairs when US was performed first and when NCCT was performed first.

To better understand the causes of false positive findings, we conducted a qualitative review of 20 cases in which echogenic foci >4 mm were misinterpreted as renal stones on US, but not confirmed on NCCT.

## RESULTS

### Descriptive analysis

A total of 206 patients (90 female, 116 male) met our inclusion and exclusion criteria. Of these patients, 148 patients had at least one stone (58 patients did not have stones). A total of 368 stones were identified on NCCT (178 right kidney, 190 left kidney). Mean age was 59.8 ± 1 with a range of 21 to 89. Mean and median days between studies were 5.7 ± 5.3 days and 3.9 days (IQR 0.96 – 9.55) respectively.

In keeping with prior studies, US tended to overestimate the size of stones. The mean stone size on US was 5.6 ± 0.01 mm, while the mean size on NCCT was 3.83 ± 0.02 mm ($p<0.001$; CI [1.42, 2.03]). The mean discrepancy was 1.72 ± 2.8 mm. A comparison of US and NCCT stone size is seen in Figure 1 and Figure 2. Small stones (≤4 mm) were overestimated more than larger stones. Mean size discrepancy for small stones was 2.2 ± 0.12 mm, but was only 0.43 ± 0.46 mm for stones >4 mm ($p<0.001$; CI [1.08, 2.43]). The sensitivity and specificity of US stone detection were 64.4% and 40.3% respectively when compared against NCCT (Table 1). The positive predictive value was 73.4%.


© 2026 Canadian Urological Association

**Missed and overcalled stones**

There were 131 stones missed on US, but visualized on NCCT. These missed stones tended to be small as 111 were ≤4 mm (Figure 1). US had less sensitivity to detect upper pole stones in both kidneys and had similar sensitivity to detect stones between right and left kidneys. (Table 2) Eighty-six stones were detected on US and not identified on NCCT. We calculated specificity by determining the "true negative" count of pairs without stones in each location. All 86 "overcalled" stones were ≤4 mm (Figure 1).

**Clinical decision-making analysis – detection of stones requiring treatment**

Stone detection analysis for the simplified situation where clinical decision-making is determined by size alone is summarized in Table 3. If only US was used for decision-making, 23 patients (16.6%) would be recommended treatment when in fact treatment is not indicated based on size criteria. Similarly, if only US size criteria were used in decision-making, 14 patients (20.9%) needing treatment would be recommended observation.

**Impact of study order on stone detection and size estimation**

Out of the 206 patients, 121 had US performed first and 85 had NCCT performed first corresponding to 264 stones and 104 stones respectively. The sensitivity and specificity was 66.7% and 18.0% with a positive predictive value of 70.7% when US was performed first. In contrast, it was 58.7% and 76.4% with a positive predictive value of 82.4% when NCCT was performed first (Table 4 and 5). Mean discrepancy in size of stones was 1.8 ± 0.2 mm where US was performed first and 1.4 ± 0.4 mm where NCCT was performed first (p=0.3038; CI [-0.3484, 1.1141]). These findings suggest that performing NCCT first may reduce false positives on subsequent US.

**Qualitative review of cases**

Several contributing factors were identified in our qualitative analysis of 20 cases.

**Ultrasound mimickers of renal stones**

Various anatomical structures and pathological entities can create echogenic foci that resemble stones. Known mimickers include renal vascular calcifications, milk of calcium cysts, cortical scarring or calcification from prior infection, and sloughed papillae.[13] Prominent medullary pyramids, especially in the setting of dilated calyces, can appear echogenic. Moreover, segmental renal arteries that terminate near the apices of the pyramids may contribute to this appearance. In our analysis, 60% (12/20) of large (>4 mm) overcalled stones were localized to the medullary tip. Additionally, we identified three cases involving misinterpretation of a milk of calcium cyst, a cyst wall calcification, and a sloughed papilla, respectively. One patient with polyarteritis nodosa and a history of renal artery aneurysm embolization was found to have partially calcified aneurysms and echogenic embolization coils misdiagnosed as stones. These findings underscore the importance of understanding renal anatomy, ultrasound physics, and relevant clinical history to avoid misinterpretation.[13,14]


© 2026 Canadian Urological Association

**Sonographer bias**
Ultrasound remains highly operator-dependent and subject to cognitive bias. In one instance, a benign echogenic focus was falsely identified as a stone, likely due to the presence of multiple confirmed stones in the same kidney. This highlights the potential for confirmation bias to influence interpretation.

**Twinkle artifact and overdiagnosis**
The twinkle artifact on color Doppler ultrasound is often used to aid in stone detection. Studies have shown that it improves sensitivity, particularly for small stones. Korkmaz et al. demonstrated enhanced detection for stones under 5 mm, while Mitterberger et al. reported twinkle artifact in 97% of stones between 3 mm and 10 mm.[15,16] However, as Sun et al. emphasized, reliance on twinkle artifact alone may increase the false positive rate.[6] Consistent with this, 95% (16 of 17) of the echogenic foci ≥4 mm that were falsely identified as stones in our study demonstrated twinkle artifact. Three cases lacked color Doppler imaging for review.

**Intermodality differences between US and NCCT**
Discrepancies between US and NCCT can also arise due to inherent limitations of each modality. In one case, two closely adjacent stones were resolved as separate stones on US, but appeared as a single stone on NCCT. Another case involved an echogenic focus interpreted as a stone on US that was subsequently identified as a parenchymal calcification on NCCT.

## DISCUSSION
This study presents one of the larger retrospective cohorts directly comparing US and NCCT for the detection and measurement of renal stones, with careful stone-by-stone matching between modalities. Our findings reinforce prior literature showing that US tends to overestimate stone size for smaller stones, and that the sensitivity of US is lower for stones located in the upper renal poles. Notably, in our cohort, stone size overestimation occurred almost exclusively in stones ≤4 mm. This finding is clinically significant, as the commonly used treatment threshold in the absence of symptoms or obstruction is 5 mm or greater.

When evaluating the clinical impact of relying on US alone for decision-making, we found that 16.6% of patients would have been recommended treatment unnecessarily, closely aligning with Ganesan et al.'s reported overtreatment rate of 14%.[5] This rate is considerably lower than the 75% reported in a recent study by Sun et al.[6] Conversely, 20.9% of patients who should have been considered for treatment would have instead been observed, a more favorable figure than the 39% under-treatment rate observed by Ganesan et al.[5] These results reinforce the recommendation that treatment decisions based on US alone should be discouraged.

Our observed US sensitivity of 64.4% is in line with more recent studies.[5,6] However, our study found a notably lower US specificity (40.3%) compared to most of the literature, where specificity is typically reported above 90%. This discrepancy likely stems from methodological differences. Many previous studies appear to assess only the largest stone per patient, whereas


© 2026 Canadian Urological Association

our analysis included all detected stones, regardless of size or number.[4,5] When the sensitivity and specificity based solely on the presence of any stone (rather than stone-for-stone matching), our values rose to 84.7% and 84.0%, respectively which more closely aligns with published data. Nevertheless, we feel that a key strength of our study is the rigorous mapping of stones between imaging modalities, considering both anatomical location and size to ensure accurate comparison.

Our study contributes novel insight into the impact of imaging order on US performance. We observed a significant improvement in specificity and positive predictive value when NCCT was performed first. This result suggests that prior knowledge of stone location or size may influence stone detection by sonographers and/or interpretation by radiologists. As many patients choose surveillance over treatment for their non-obstructive renal stones, perhaps a protocol that starts with a NCCT and is followed by US surveillance would allow for mitigation of the accuracy shortfalls of US as an imaging modality for the detection of renal stones.

**Limitations**

This study has several limitations that should be acknowledged. First, it was conducted at a single academic tertiary care center, which may limit the generalizability of the findings to other clinical settings with different patient populations or US expertise.

Second, we did not include patient body mass index (BMI), which has been shown in prior studies to significantly affect US sensitivity and image quality. Future studies incorporating BMI may provide additional insights into patient-specific factors influencing diagnostic performance.

Third, while we limited the interval between imaging studies, the mean time between US and NCCT was approximately 5.7 days. Although we reviewed clinical records to assess for potential stone migration or interval passage, this delay introduces the possibility of discordance due to changes in stone status over time. However given the short time frame on average between studies, we suspect the impact of this on the results is likely to be minimal.

Fourth, while NCCT is widely accepted as the reference standard for stone detection, it has limited sensitivity for detecting stones under 3 mm, particularly with 3 mm slice thickness. To address this, we reviewed NCCT images in axial, sagittal, and coronal planes and excluded suspected false negatives by chart review; however, it remains possible that some small stones seen on US were not detected on NCCT and may have been misclassified.

Finally, this study involved detailed, independent image review by investigators, which may not reflect real-world clinical US practice. Our sonographers and radiologists were not blinded when performing and reading between the two modalities of studies which was a decision made to reflect real-world clinical practice. However, there is potential expectation/confirmation bias due to the nature of this methodology. Prior available studies are routinely used to assist sonographers and radiologists in their interpretation. However, there are some scenarios where sonographers and interpreting radiologists may not have access to paired


© 2026 Canadian Urological Association

NCCT imaging or the time for such granular evaluation, potentially leading to different diagnostic accuracy.

## CONCLUSIONS

In this large, retrospective cohort study directly comparing US and NCCT for renal stone detection, we found that US continues to overestimate stone size in stones ≤4 mm. The sensitivity varies based on stone location and is decreased for the upper poles. While the overall sensitivity of US was consistent with recent literature, specificity was notably lower due to our rigorous stone-by-stone comparison methodology, which included smaller and multiple stones typically omitted in prior studies. Clinical decision-making based on US alone would have resulted in both overtreatment and undertreatment in a substantial minority of patients, reinforcing concerns about its limitations as a standalone diagnostic tool.

Importantly, our study highlights that the order of imaging significantly influences diagnostic accuracy, with US demonstrating improved specificity and positive predictive value when preceded by NCCT. Qualitative review of false positive cases revealed that stone mimickers, reliance on twinkle artifact, and operator bias contribute meaningfully to diagnostic inaccuracies. These findings raise the possibility that US performance may be optimized through targeted quality improvement efforts and should be further investigated in prospective studies. Ultrasound remains a valuable modality, especially in settings where NCCT is less accessible or contraindicated. However, clinicians must be aware of its limitations, and careful consideration should be given to imaging sequence, interpretation context, and the clinical decision thresholds that rely on accurate stone characterization.

DRAFT


© 2026 Canadian Urological Association

**REFERENCES**

1. Worster A, Preyra I, Weaver B, et al. The accuracy of noncontrast helical computed tomography versus intravenous pyelography in the diagnosis of suspected acute urolithiasis: A meta-analysis. *Ann Emerg Med* 2002;40:280-6. https://doi.org/10.1067/mem.2002.126170
2. Fulgham PF, Assimos DG, Pearle MS, et al. Clinical effectiveness protocols for imaging in the management of ureteral calculous disease: AUA technology assessment. *J Urol* 2013;189:1203-13. https://doi.org/10.1016/j.juro.2012.10.031
3. Ray AA, Ghiculete D, Pace KT, et al. Limitations to ultrasound in the detection and measurement of urinary tract calculi. *Urology* 2010;76:295-300. https://doi.org/10.1016/j.urology.2009.12.015
4. Sternberg KM, Eisner B, Larson T, et al. Ultrasonography significantly overestimates stone size when compared to low-dose, noncontrast computed tomography. *Urology* 2016;95:67-71. https://doi.org/10.1016/j.urology.2016.06.002
5. Ganesan V, De S, Greene D, et al. Accuracy of ultrasonography for renal stone detection and size determination: Is it good enough for management decisions? *BJU Int* 2017;119:464-9. https://doi.org/10.1111/bju.13605
6. Sun R, Sommer E, Ganesan C, et al. Limitations of ultrasound compared with computed tomography for kidney stone surveillance. *Can Urol Assoc J* 2025;19:E229-37. https://doi.org/10.5489/cuaj.9043
7. Pearle MS, Matlaga BR, Antonelli JA, et al. Surgical management of kidney and ureteral stones: AUA guideline (2026) part I: Evaluation and treatment of patients with kidney and/or ureteral stones. *J Urol* 2026;215:113-23. https://doi.org/10.1097/JU.0000000000004842
8. Viprakasit DP, Sawyer MD, Herrell SD, et al. Limitations of ultrasonography in the evaluation of urolithiasis: A correlation with computed tomography. *J Endourol* 2012;26:209-13. https://doi.org/10.1089/end.2011.0177
9. Kanno T, Kubota M, Sakamoto H, et al. The efficacy of ultrasonography for the detection of renal stone. *Urology* 2014;84:285-8. https://doi.org/10.1016/j.urology.2014.04.010
10. Ahmed F, Askarpour MR, Eslahi A, et al. The role of ultrasonography in detecting urinary tract calculi compared to CT scan. *Res Rep Urol* 2018;10:199-203. https://doi.org/10.2147/RRU.S178902
11. Smith-Bindman R. Ultrasonography vs. CT for suspected nephrolithiasis. *N Engl J Med* 2014;371:2531. https://doi.org/10.1056/NEJMc1412853
12. Kanno T, Kubota M, Sakamoto H, et al. Determining the efficacy of ultrasonography for the detection of ureteral stone. *Urology* 2014;84:533-7. https://doi.org/10.1016/j.urology.2014.04.047
13. Durr ES, Khan AN, Craig M, et al. Sonographic mimics of renal calculi. *J Ultrasound Med* 2004;23:1361-7. https://doi.org/10.7863/jum.2004.23.10.1361
14. Wieczorek AP, Wozniak MM, Tyloch JF. Errors in the ultrasound diagnosis of the kidneys, ureters and urinary bladder. *J Ultrason* 2013;13:308-18. https://doi.org/10.15557/JoU.2013.0031
15. Korkmaz M, Aras B, Sanal B, et al. Investigating the clinical significance of twinkling artifacts in patients with urolithiasis smaller than 5 mm. *Jpn J Radiol* 2014;32:482-6. https://doi.org/10.1007/s11604-014-0337-6


© 2026 Canadian Urological Association

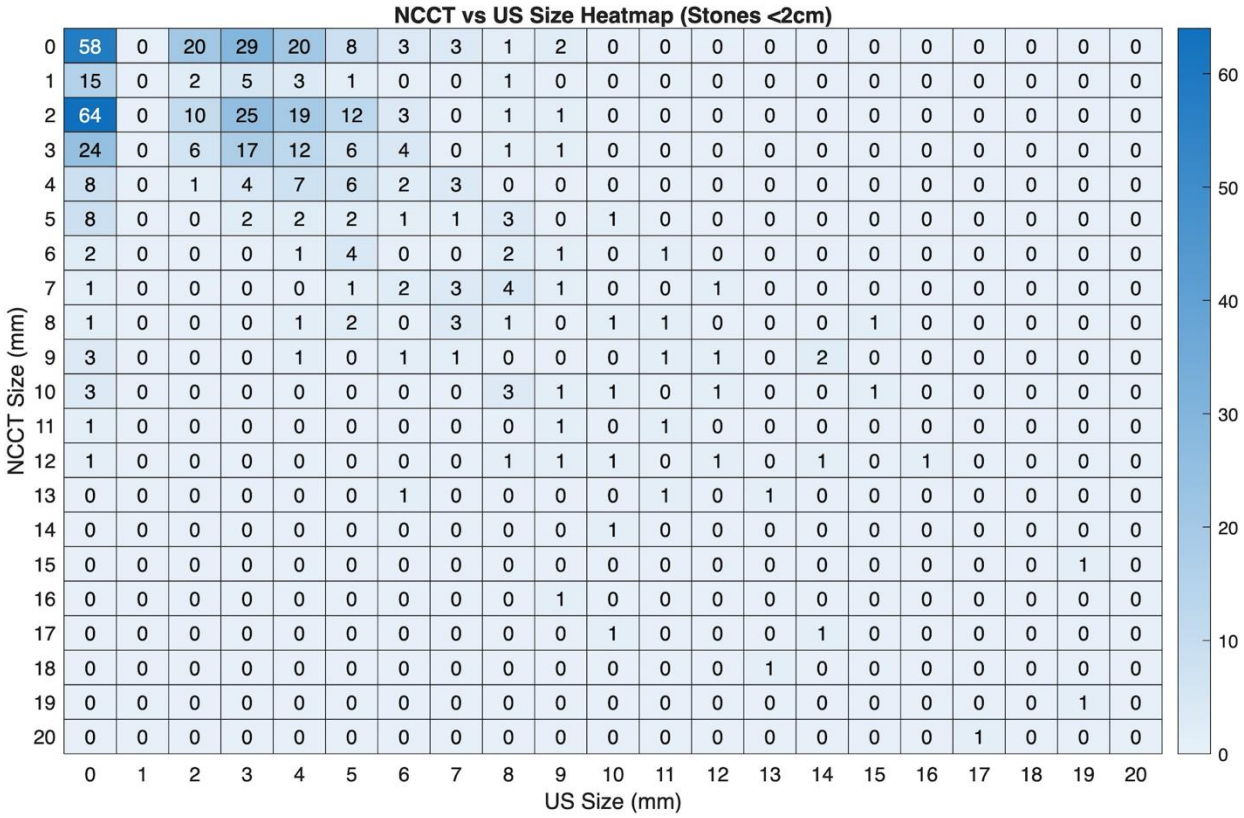
16. Mitterberger M, Aigner F, Pallwein L, et al. Sonographic detection of renal and ureteral stones: Value of the twinkling sign. *Int Braz J Urol* 2009;35:532-9. https://doi.org/10.1590/S1677-55382009000500004

DRAFT


© 2026 Canadian Urological Association

## FIGURES AND TABLES

**Figure 1.** NCCT vs. US size comparison for stones under 20 mm. The frequency of occurrence is displayed in each box. Only stones ≤20 mm are shown for legibility. The first column represents NCCT stone sizes of stones missed on US. The first row represents US stone sizes of stones overcalled on US but not visualized on NCCT.





© 2026 Canadian Urological Association

**Figure 2.** Bland-Altman plot of US vs. NCCT stone size estimations.

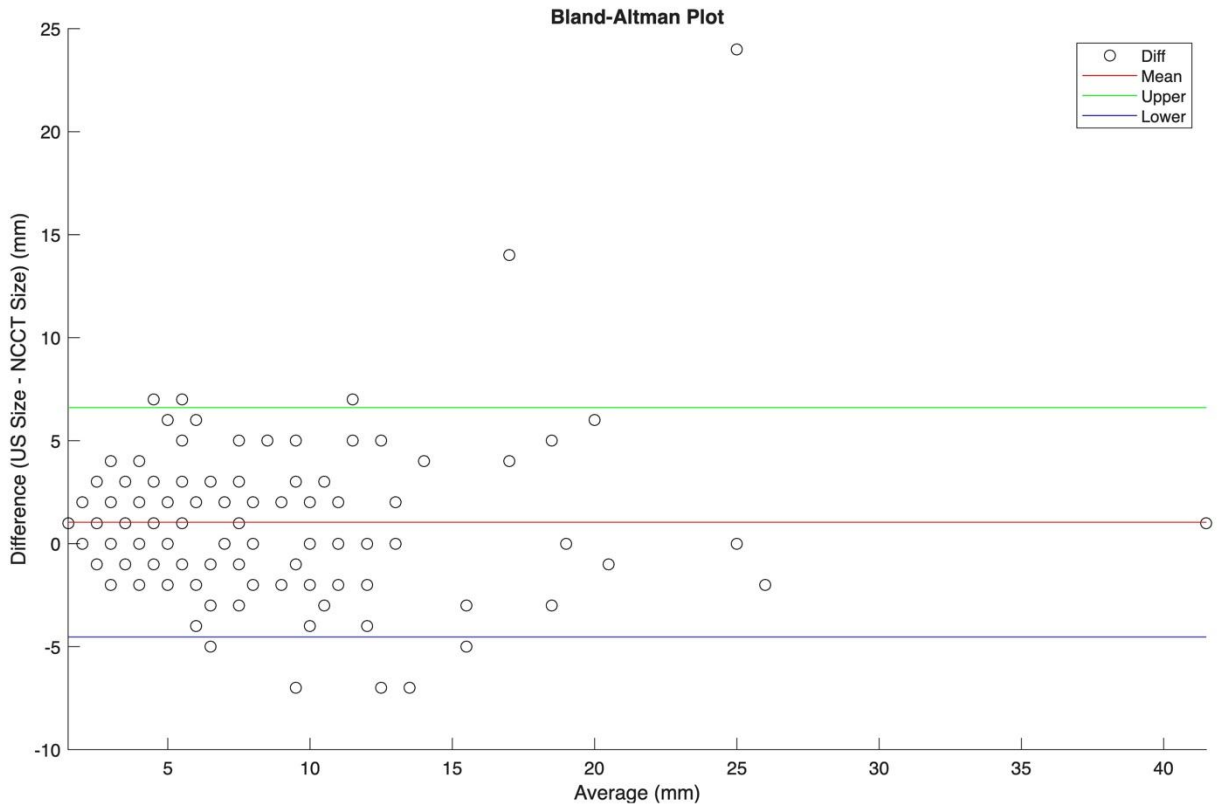



**Table 1. Contingency table showing US stone detection compared with NCCT stone detection**

| | NCCT stone present | NCCT stone not present | Total |
|---|---|---|---|
| US stone present | 237 | 86 | **323** |
| US stone not present | 131 | 58 | **189** |
| Total | **368** | **144** | **512** |


© 2026 Canadian Urological Association

**Table 2. Stone-level sensitivity, specificity, and positive predictive value detection, and overcall by anatomic location**

| Location | FN | TP | SEN (%) | FP | TN | SPEC (%) | PPV (%) |
|---|---|---|---|---|---|---|---|
| Both kidneys | 131 | 237 | **64.4 (368)** | 86 | 58 | **40.3 (144)** | **73.4 (323)** |
| | | | | | | | |
| Left kidney | 63 | 127 | **66.8 (190)** | 43 | 98 | **69.5 (141)** | **74.7 (170)** |
| Left UP | 15 | 14 | **48.3 (29)** | 3 | 179 | **98.4 (182)** | **82.4 (17)** |
| Left IP | 15 | 28 | **65.1 (43)** | 20 | 156 | **88.6 (176)** | **58.3 (48)** |
| Left LP | 24 | 72 | **75 (96)** | 19 | 137 | **87.8 (156)** | **79.1 (91)** |
| Left pelvis/UPJ | 9 | 13 | **59.1 (22)** | 1 | 187 | **99.5 (188)** | **92.9 (14)** |
| | | | | | | | |
| Right kidney | 68 | 110 | **61.8 (178)** | 43 | 97 | **69.3 (140)** | **71.9 (153)** |
| Right UP | 32 | 26 | **44.8 (58)** | 13 | 150 | **92.0 (163)** | **66.7 (39)** |
| Right IP | 14 | 31 | **68.9 (45)** | 12 | 162 | **93.1 (174)** | **72.1 (43)** |
| Right LP | 21 | 48 | **69.6 (69)** | 18 | 151 | **89.3 (169)** | **72.7 (66)** |
| Right pelvis/ UPJ | 1 | 5 | **83.3 (6)** | 0 | 200 | **100 (200)** | **100 (5)** |

Brackets for SEN, SPEC, and PPV denote denominator. FN: number of missed stones on US; FP: number of overcalled stones on US; IP: inter pole; LP: lower pole; TN: number of studies where both US and NCCT show no stones; SEN: sensitivity; SPEC: specificity; TP: number of correctly identified stones on US; UP: upper pole; UPJ: ureteropelvic junction.

DRAFT


© 2026 Canadian Urological Association

| **Table 3. Clinical decision-making analysis 2x2 contingency table for detection of stones >4 mm and ≤4 mm** | | | |
|---|---|---|---|
| | **NCCT stone >4 mm** | **NCCT stone ≤4 mm** | **Total** |
| US stone >4 mm | 77 | 59 | **136** |
| US stone ≤4 mm | 27 | 349 | **376** |
| Total | **104** | **408** | **512** |

| **Table 4. Contingency table showing stone detection in the subgroup of patients in whom US was performed first.** | | | |
|---|---|---|---|
| **US first** | **NCCT stone present** | **NCCT stone not present** | **Total** |
| US stone present | 176 | 73 | **249** |
| US stone not present | 88 | 16 | **104** |
| Total | **264** | **89** | **353** |

| **Table 5. Contingency table showing stone detection in the subgroup of patients in whom NCCT was performed first** | | | |
|---|---|---|---|
| **NCCT first** | **NCCT stone present** | **NCCT stone not present** | **Total** |
| US stone present | 61 | 13 | **74** |
| US stone not present | 43 | 42 | **85** |
| Total | **104** | **55** | **159** |

DRAFT


© 2026 Canadian Urological Association