

Stone-Mode Ultrasound for Determining Renal Stone Size

Philip C. May, MD,^{1,2} Yasser Haider,² Barbrina Dunmire, MS,² Bryan W. Cunitz, MS,²
Jeff Thiel,³ Ziyue Liu, PhD,⁴ Matthew Bruce, PhD,² Michael R. Bailey, PhD,^{1,2}
Mathew D. Sorensen, MD, MS,^{1,5} and Jonathan D. Harper, MD¹

Abstract

Purpose: The purpose of this study was to measure the accuracy of stone-specific algorithms (S-mode) and the posterior acoustic shadow for determining kidney stone size with ultrasound (US) *in vivo*.

Materials and Methods: Thirty-four subjects with 115 renal stones were prospectively recruited and scanned with S-mode on a research US system. S-mode is gray-scale US adjusted to enhanced stone contrast and resolution by minimizing compression and averaging, and increasing line density and frequency. Stone and shadow width were compared with a recent CT scan and, in 5 subjects with 18 stones, S-mode was compared with a clinical US system.

Results: Overall, 84% of stones identified on CT were detected on S-mode and 66% of these shadowed. Seventy-three percent of the stone measurements and 85% of the shadow measurements were within 2 mm of the size on CT. A posterior acoustic shadow was present in 89% of stones over 5 mm versus 53% of stones under 5 mm. S-mode visualized 78% of stones, versus 61% for the clinical system. S-mode stone and shadow measurements differed from CT by 1.6 ± 1.0 mm and 0.8 ± 0.6 mm, respectively, compared with 2.0 ± 1.5 mm and 1.6 ± 1.0 mm for the clinical system.

Conclusions: S-mode offers improved visualization and sizing of renal stones. With S-mode, sizing of the stone itself and the posterior acoustic shadow were similarly accurate. Stones that do not shadow are most likely <5 mm and small enough to pass spontaneously.

Introduction

ULTRASOUND (US) HAS PROVEN effective in the diagnosis of symptomatic renal and ureteral stones in the emergency department setting; however, US is limited for treatment planning by low sensitivity, user dependence, and inaccurate stone sizing.¹ The sensitivity of US for diagnosing renal stones compared with CT varies widely, from 24% to 70%.^{2–6} The accuracy of US for sizing stones compared with CT has also been shown to differ widely, particularly with size and depth.^{2–6} Anecdotally, many urologists feel US reports a significantly larger size than CT. This has resulted in a general lack of trust in US by the urology community as a primary modality for imaging renal stones, as accurate detection and sizing are important for surgical treatment planning and monitoring stone burden.

Our group has focused on improving the accuracy of stone sizing and detection by creating algorithms optimized for the imaging of kidney stones versus soft tissues.⁷ This includes reducing the use of smoothing algorithms and the compression of high signal-intensity regions to improve visualization

and sharpen contrast at the edge of stones and in the posterior shadow.^{8,9} These techniques have been incorporated into an US imaging modality, coined the stone-specific mode, or S-mode.

A complementary strategy for improving stone sizing accuracy with US is the use of the posterior acoustic shadow. Our *in vitro* study showed measuring the width of the posterior shadow was more accurate than measuring stone size in the traditional manner and did not differ with stone depth.¹⁰

The goal of this study was to assess the accuracy of S-mode US and the posterior acoustic shadow for the sizing of renal stones *in vivo*.

Materials and Methods

We performed a prospective study of kidney stone patients at the University of Washington and Puget Sound Veteran's Affairs hospitals. Institutional Review Board approval was obtained along with informed consent from all subjects.

¹Department of Urology, University of Washington, Seattle, Washington.

²Center for Industrial and Medical Ultrasound, University of Washington Applied Physics Lab, Seattle, Washington.

³Department of Radiology, University of Washington, Seattle, Washington.

⁴Department of Biostatistics, Indiana University, Indianapolis, Indiana.

⁵Division of Urology, Department of Veteran Affairs Medical Center, Seattle, Washington.

Study population

Screening took place between January 1, 2015, and March 31, 2016. Subjects were required to be over 18 years of age and to have at least one kidney stone and a CT scan performed within 100 days of their clinic appointment without intervening intervention. Subjects with staghorn calculi, ureteral stones, residual fragments following lithotripsy, or ureteral stents were excluded.

US system

All subjects were scanned with a research US instrument (VDAS; Verasonics, Inc., Redmond, WA) and curvilinear imaging probe (C5-2; Philips Ultrasound, Bothell, WA). No spatial compounding, speckle reduction, or no other backend processing implemented. The transducer operated at 4.5 MHz for both transmit and receive, which provided greater resolution than the conventional 3.2 MHz. A high scanning line density (256 lines/frame) was used to improve resolution. The operator was limited to adjusting the gain and the focal position in a step toward reducing operator dependence. The scanning session was recorded with screen capture software.

A subset of 5 subjects was also scanned with a clinical US system (Aixplorer and XC6-1; probe, Supersonic Imagine, France), performed by the same sonographer and during the same session as the S-mode scan for direct comparison. Scans were conducted using the renal preset with spatial compounding turned off and the resolution map (SR) set to 4. The operator, a trained sonographer, reduced the dynamic gain for improved stone contrast and adjusted the focus and gain as preferred.

Protocol

Subjects underwent imaging of one or both kidneys based upon stone location by a single sonographer. Since the focus of this study was on stone sizing, and not sensitivity or specificity, the sonographer was informed of the location and number of stones expected based on the CT images; as such, we do not report sensitivity or specificity, but refer to the number of stones seen with S-mode US as the number detected. The total number of stones, stone location, and presence or absence of a posterior acoustic shadow were recorded. Each stone was imaged from multiple angles. After completion of the study, select images were saved from the study recordings that demonstrated the maximal dimension of the stone and the posterior acoustic shadow.

The images were presented to three reviewers blinded to the CT stone size, including a sonographer, urologist, and US engineer. Each reviewer independently measured the size of the stone and the size of the posterior acoustic shadow. Each displayed image included a drawing line, and each reviewer was instructed to place each end of the drawing tool to mark the edges of the stone or shadow. No indication of size was reported to the reviewers. The shadow was measured ~ 1 cm behind the stone.

The results were averaged and compared with the maximum size on CT. Due to the retrospective nature of the CT scans, the specific protocol varied; the majority of studies were non-contrast enhanced scans of the abdomen and pelvis (CT KUB). Slice thickness was between 2.5 and 3.0 mm. Stones were measured using a chest–abdomen window, in the axial and

coronal orientations, at a zoom factor of 4. The maximum width served as the reference measurement for assessing S-mode US accuracy. The primary outcome was the absolute value of the difference in measured size between CT and US.

Statistical analysis

The interclass correlation coefficient between the three reviewers was 0.80 for the S-mode stone measurement and 0.85 for the S-mode shadow data. A coefficient above 0.75 indicates excellent reproducibility among measurements and appropriateness of averaging the results.¹¹ Linear regression models were used to compare S-mode measurements to CT and the clinical system, and to individually evaluate the effects of stone size, laterality, and pole location. The results are reported as bias (measured size–reference size) and standard deviation for the 95% confidence interval. Pearson's and Spearman's correlation coefficients were calculated to evaluate the effect of body mass index (BMI) and depth on the difference between S-mode and CT size. Weighted Kappa coefficients and their 95% confidence intervals were calculated to evaluate the concordance between S-mode and CT with respect to three clinically relevant size categories (≤ 5 , 5.1–10, and >10 mm based on CT size). Two-sided $p < 0.05$ were considered significant. Statistical analysis was performed using SAS 9.4 (SAS Institute, Cary, NC).

Results

Forty-six subjects were recruited, with 35 subjects and 115 renal stones included in the analysis. Of the 35 subjects, 29 were male and 6 were female with a mean age of 53 ± 16 , a mean BMI of 29.4 ± 6.2 kg/m², and mean interval between CT and US of 34 ± 32 days. Reasons for exclusion included no identifiable stones on CT (2), postlithotripsy fragments (1), ureteral stones only (2), over 100 days between CT and US (1), inability to record study images (2), and inability to correlate stones between US and CT (3).

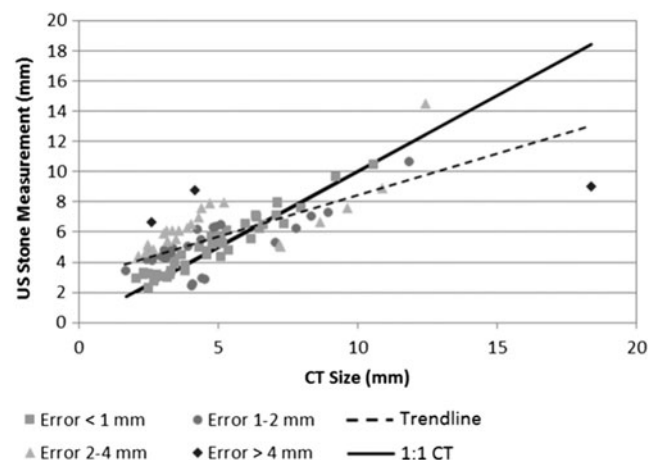


FIG. 1. Correlation of stone size measurement between US S-mode and CT. Each measurement is color- and shape-coded based on the difference from the CT measurement. The solid *black line* represents a 1:1 correspondence between the S-mode and CT measurements. The *dashed line* represents the correlation between the S-mode and CT measurements. Most stones are within 2 mm of the size determined on CT. US = ultrasound.

TABLE 1. AVERAGE STONE SIZE DETERMINED BY CT AND S-MODE US

Size group	No. of stones	CT size (mm)	S-mode US stone size (mm)	Difference ^a (mm)	Bias ±95% CI	p ^b
All stones	93	5.0±2.7	5.6±2.0	1.4±1.3	0.58±0.38	0.0028
≤5 mm	59	3.4±0.8	4.7±1.5	1.5±1.0	1.24±0.35	<0.0001
5–10 mm	29	6.7±1.4	6.5±1.2	1.0±0.7	0.29±0.46	0.19
>10 mm	5	12.9±3.7	10.7±2.3	3.0±3.7	-2.17±3.88	0.27

^aDifference represents average absolute difference between CT and S-mode size measurements.

^bp>0.05 is significant for no bias; nonsignificance is also dependent on sample size. US=ultrasound.

Overall, 84% (97 of 115) of stones identified on CT were detected on S-mode. Detectability was not affected by laterality or pole location; size did impact detectability ($p<0.0001$) as nonvisualized stones had an average CT size of 3.1 ± 1.0 mm compared with 4.9 ± 2.7 mm for detected stones. Ninety-five percent of stones missed with S-mode were smaller than 5 mm; only one stone missed was >5 mm. On average, S-mode overestimated stone size for stones identified on CT as less than 6 mm and underestimated stone size for stones identified on CT as greater than 6 mm (Fig. 1). Forty-four percent (41 of 93) of S-mode stone measurements were within 1 mm of the CT size and 73% (68 of 93) were within 2 mm of the CT size. Only five (5%) of the S-mode stone measurements were more than 3 mm discrepant from the CT size. There was no correlation between stone size accuracy and BMI or depth; less bias was observed for right-sided stones versus left-sided stones. The average difference in size measurement for all stones and within select size groupings is presented in Table 1. S-mode differed from CT by 1.3 ± 1.0 mm for stones ≤ 10 mm and 3.0 ± 3.7 mm for stones >10 mm. The concordance of S-mode to CT for clinically relevant size categories is presented in Table 2. Overall, 73% of stones (68 of 93) were within the same size category on both S-mode and CT, including 64% for stones ≤ 5 mm, 93% for stones between 5 and 10 mm, and 60% for stones ≥ 10 mm.

With S-mode a discrete shadow was present on 89% of stones over 5 mm, but only 53% of stones under 5 mm ($p<0.0001$) (Fig. 2). In total, 55% (30 of 56) of the shadow measurements were within 1 mm and 85% (47 of 55) were within 2 mm of the CT size. Table 3 compares the stone and shadow measurements for stones that shadowed. The difference in size measurement using shadow was 1.1 ± 0.9 mm for stones ≤ 10 mm and 4.5 ± 5.1 mm for stones >10 mm. The shadow was more accurate than direct stone measurement for stones ≤ 5 mm, with a concordance of 80%, but under-

estimated stones >5 mm, resulting in a concordance for stones 5–10 mm and >10 mm of 55% and 25%, respectively. Of the 10 stones in the 5–10 mm category underestimated with the shadow, the average underestimation was 1.0 mm, with the majority of stones under 6.5 mm on CT.

S-mode visualized 78% of stones (14 of 18) versus 61% (11 of 18) for the clinical system. Size with S-mode was ~ 0.4 and 0.8 mm more accurate for the stone and shadow results, respectively, compared with the clinical system (Table 4). The variance was also smaller with S-mode compared with the clinical system. The sample size was too small to consider statistical significance.

Discussion

Clinical US has limited utility as a standalone modality for imaging stones due to poor stone sizing accuracy and sensitivity.^{2–6} We detail a number of findings that suggest S-mode US is more accurate at sizing renal stones than both previously published series and in direct comparison to a clinical US system. Estimated stone size from S-mode using the stone or acoustic shadow was within 1 mm of the CT size (95% confidence interval) for stones <10 mm. Accurate sizing for stones under 10 mm is critical, as decisions regarding intervention versus observation, and selecting a type of intervention, are commonly made based on the assumption of accurate stone sizing.¹² S-mode also performed well in classifying stones into clinically relevant size categories

TABLE 2. STONE SIZE CONCORDANCE BETWEEN CT AND US S-MODE STONE MEASUREMENTS

CT (mm)	S-mode		
	≤5 mm	5–10 mm	>10 mm
≤5 mm	38	21	0
5–10 mm	2	27	0
>10 mm	0	0	3

K (agreement)=0.540.

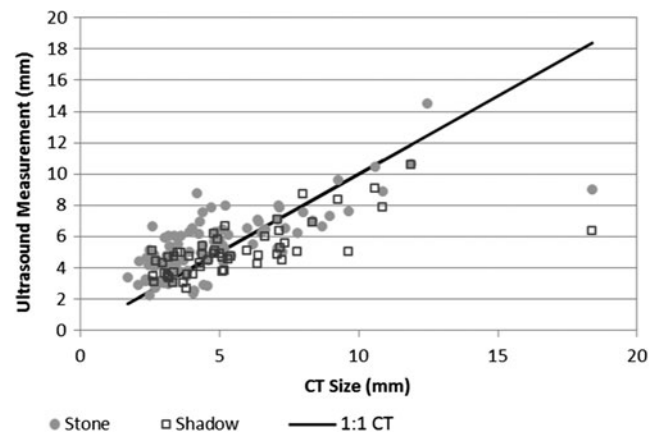


FIG. 2. Comparison of S-mode stone versus the S-mode shadow measurement. Both modalities track the CT measurement well, with most stones within 2 mm of the CT-determined size.

TABLE 3. COMPARISON BETWEEN AVERAGE ABSOLUTE DIFFERENCE FOR S-MODE STONE AND SHADOW

	N	Size CT (mm)	Difference		Bias	
			S-mode US shadow (mm)	S-mode US stone ^a (mm)	Stone shadow	p ^b
All stones	56	5.5±2.9	1.3±1.7	1.4±1.5	0.92±0.25	<0.0001
≤5 mm	30	3.7±0.7	0.8±0.6	1.5±1.0	0.78±0.32	<0.0001
5–10 mm	21	6.8±1.4	1.5±1.0	1.0±0.8	1.04±0.43	<0.0001
>10 mm	4	12.9±3.7	4.5±5.1	3.2±4.2	1.28±1.06	0.019

^aOnly stones that shadowed are included.

^b*p*>0.05 is significant for no bias.

(Table 2). Only 7% of stones >5 mm were misclassified as smaller than 5 mm, while 36% of stones ≤5 mm were misclassified greater than 5 mm by an average of 2.3 mm.

This study is the first to describe the clinical features and utility of the posterior acoustic shadow in human subjects. A shadow was present in 66% of all stones, including 89% of stones over 5 mm and 53% of stones under 5 mm, demonstrating a statistically significant difference in shadow presence based on stone size. In addition, shadow width demonstrated improved sizing performance for stones <5 mm in size (Table 3). These features of the posterior acoustic shadow have not been previously reported and are potentially useful adjuncts in the management of renal stones with US. If an object in the kidney that does not shadow is a stone, it is likely <5 mm, and thus likely to pass spontaneously.¹²

S-mode also demonstrated improved detection of stones than previously published series.^{2–6} S-mode detected 83% of stones overall, including detection of 66% of stones ≤3 mm and 74% of stones ≤4 mm. Although this is higher than previous reports of US sensitivity for stone diagnosis, this rate is not a direct representation of the true clinical sensitivity of S-mode because our sonographer was not blinded to the CT results. In a direct comparison to a clinical US system, S-mode visualized 78% of stones versus 61%, a difference that was not statistically significant based on a small sample size.

Multiple strategies for combining the strengths of S-mode stone and shadow sizing were investigated to improve stone size category concordance. These included averaging stone width and shadow width and assigning stones without an identifiable shadow a size of 5.0 mm. This size was based on a binary classification tree analysis that showed stones >4.2 mm were likely to shadow. In addition, incorporating the resolution limits of CT at ±1 mm, all stones within 1 mm of the CT size were considered concordant. These results show an improved concordance of 82% and 90% for the <5 mm and 5–10 mm categories, respectively.

TABLE 4. COMPARISON BETWEEN S-MODE AND CLINICAL ULTRASOUND SYSTEM IN DETERMINING STONE SIZE

	No. of stones	Size CT (mm)	Absolute difference from CT	
			S-mode (mm)	Clinical system (mm)
Stone	14 (S-mode) 11 (Clinical)	4.5±2.2	1.6±1.0	2.0±1.5
Shadow	7	4.5±2.2	0.8±0.6	1.6±1.0

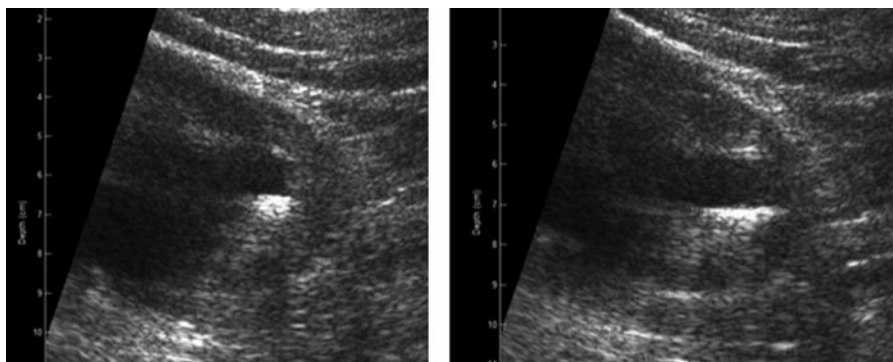
While overestimation of stone size with US is common, underestimation of stones larger than 10 mm was seen in both our study and the study by Kanno et al.³ The degree of size underestimation was generally mild (Fig. 1). In these instances, although the measured size was underestimated, the US operator still perceived that the stone was qualitatively large. The longest dimension of larger stones can be missed in a single two-dimensional US image. The real-time US imaging display, however, reveals the stone in many planes that can indicate a large stone. In our protocol, the result of the size comparison between CT and S-mode was highly dependent upon the choice of images from the S-mode study; the sonographer picked a single image that best represented the largest dimension, and all reviewers sized stones/shadows based on the same selected image. Figure 3 demonstrates how one of these larger stones was underestimated. The stone was initially sized at 9.0 mm on S-mode, in comparison to 18.4 mm on CT. However, selecting an alternate image would have resulted in a much more accurate stone size (17.5 mm). In general, the accuracy of US stone sizing is lost when the stone cannot be clearly resolved from tissue or other stones on selected images or if the image selected does not properly reflect the largest dimension.

Measurement of stone size and shadow size under S-mode showed similar results compared with CT. This is likely because the improved accuracy of stone sizing with S-mode blunted the potential benefit of measuring the width of the shadow. Shadow width did demonstrate improved performance for stones ≤5 mm and a statistically significant association with stone size, which is useful diagnostic information as smaller stones were less likely to shadow than larger stones. Fowler et al. disregarded any hyperechoic objects that did not shadow on US, which may have contributed to the low sensitivity reported in their article.²

Clinical US systems have high variability in the sizing of stones in comparison to CT. Table 4 shows a standard deviation higher for the clinical imaging system than S-mode. Measuring shadow size appeared to improve sizing accuracy with the clinical system, although this study was not powered to detect a difference in the two modalities for the clinical system. S-mode stone and S-mode shadow were both more accurate than the clinical system.

Limitations of our study include the time interval between CT and US, where stone might have grown before the US. Because of variable slice thickness and protocol, there is the potential that some stones were inaccurately sized on CT, contributing to inaccuracy in our reference measurement. The sample size was small for comparing S-mode to the clinical US system. A more inclusive study would include

FIG. 3. *Left panel* shows US image used for the S-mode stone width measurement as reported in our results (9.0 mm, in comparison to 18.4 mm on CT). The *right panel* shows an alternate image of the same stone that was initially felt to be spuriously large. The width measurement on the *right panel* is actually considerably closer (17.5 mm) to the CT size of the stone.



multiple clinical US systems, which are outside the intent of this study. Although our detection rate is higher than previous reports of US sensitivity for stone diagnosis, this rate is not a direct representation of the true clinical sensitivity of S-mode because our sonographer was not blinded to CT results. This study also does not address multiple stones that cannot be individually resolved due to the resolution limits of US or CT.

Conclusions

S-mode offers improved visualization and sizing of renal stones. Measurement of the stone and shadow with S-mode was similarly accurate in comparison to CT, with a bias from CT size of ~ 1 mm. Measurement of stone size by measuring the shadow improved accuracy on the clinical machine. Kidney stones without a shadow are likely <5 mm, and thus, the majority will pass spontaneously.

Acknowledgments

This work is part of a large collaborative effort, and we appreciate the help of our many collaborators at the University of Washington Center for Industrial and Medical US, in the University of Washington Department of Urology, and within NIDDK Program Project DK043881. This material is the result of work supported by resources from the Veterans Affairs Puget Sound Healthcare System, Seattle, Washington. Funding was provided by National Space Biomedical Research Institute (NSBRI) through NASA NCC 9-58, and grants from the National Institute of Diabetes and Digestive and Kidney Diseases (DK043881 and DK092197).

Author Disclosure Statement

M.R.B., B.C., B.D., and M.S. have equity in SonoMotion, Inc., which has licensed this technology from the University of Washington. P.C.M., Y.H., J.T., Z.L., M.B., and J.D.H. have no competing financial interests.

References

1. Smith-Bindman R, Aubin C, Bailitz J, et al. Ultrasonography versus computed tomography for suspected nephrolithiasis. *N Engl J Med* 2014;371:1100–1110.
2. Fowler KAB, Locken JA, Duchesne JH, et al. US for detecting renal calculi with nonenhanced CT as a reference standard. *Radiology* 2002;222:109–113.
3. Kanno T, Kubota M, Sakamoto H, et al. The efficacy of ultrasonography for the detection of renal stone. *Urology* 2014;84:285–288.

4. 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.
5. Wong LM, Jenkins M. Accuracy of Ultrasonography for the Evaluation of Urolithiasis in Patients Undergoing Extracorporeal Shockwave Lithotripsy: Comparison with Non-contrast Computed Tomography. San Diego, CA: American Urological Association Annual Meeting Abstract, 2016, p. 1.
6. Larson T, Eisner BH. Can Ultrasonography be Used to Guide the Diagnosis and Management of Nephrolithiasis? San Diego, CA: American Urological Association Annual Meeting Abstract, 2016, p. 1.
7. Dunmire B, Lee FC, Hsi RS, et al. Tools to improve the accuracy of kidney stone sizing with ultrasound. *J Endourol* 2015;29:147–152.
8. Gabriel H, Shulman L, Marko J, et al. Compound versus fundamental imaging in the detection of subdermal contraceptive implants. *J Ultrasound Med* 2007;26:355–359.
9. Shabana W, Bude RO, Rubin JM. Comparison between color Doppler twinkling artifact and acoustic shadowing for renal calculus detection: An in vitro study. *Ultrasound Med Biol* 2009;35:339–350.
10. Dunmire B, Harper JD, Cunitz BW, et al. Use of the acoustic shadow width to determine kidney stone size with ultrasound. *J Urol* 2016;195:171–177.
11. Fleiss JL. Reliability of Measurement. The Design and Analysis of Clinical Experiments. Hoboken, NJ: John Wiley & Sons, Inc., 1999, pp. 1–32.
12. Preminger GM, Tiselius HG, Assimos DG, et al. 2007 Guideline for the management of ureteral calculi. *Eur Urol* 2007;52:1610–1631.

Address correspondence to:
Philip C. May, MD
Department of Urology
University of Washington
1959 NE Pacific Street
Box 356510
Seattle, WA 98195

E-mail: pmay727@uw.edu

Abbreviations Used

BMI = body mass index
CT = computed tomography
US = ultrasound