



Published in final edited form as:

World J Urol. 2018 May ; 36(5): 727–732. doi:10.1007/s00345-017-2156-8.

Retrospective comparison of measured stone size and posterior acoustic shadow width in clinical ultrasound images

Jessica C. Dai^{1,*}, Barbrina Dunmire², Kevan M. Sternberg³, Ziyue Liu⁴, Troy Larson⁵, Jeff Thiel⁶, Helena C. Chang¹, Jonathan D. Harper¹, Michael R. Bailey^{1,2}, and Mathew D. Sorensen^{1,7}

¹Department of Urology, University of Washington School of Medicine, Seattle, WA

²Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, Seattle, WA

³Division of Urology, University of Vermont, Burlington, VT

⁴Department of Biostatistics, Indiana University, Indianapolis, IN

⁵Department of Urology, University of Florida, Gainesville, FL

⁶Department of Radiology, University of Washington School of Medicine, Seattle, WA

⁷Division of Urology, Department of Veteran Affairs Medical Center, Seattle, WA

Abstract

Purpose—Posterior acoustic shadow width has been proposed as a more accurate measure of kidney stone size compared to direct measurement of stone width on ultrasound (US). Published data in humans to date have been based on a research US system. Herein, we compare these two measurements in clinical US images.

Methods—Thirty patient image sets where computed tomography (CT) and US images were captured less than one day apart were retrospectively reviewed. Five blinded reviewers

*Corresponding author: Jessica C. Dai, Department of Urology, University of Washington, 1959 NE Pacific Street, Box 356510, Seattle, WA 98195, jcdai@uw.edu.

Author contributions:

JC Dai: Project development, data collection and management, data analysis, manuscript writing/editing

B Dunmire: Project development, data collection and management, data analysis, manuscript writing/editing

KM Sternberg: Project development, manuscript editing

Z Liu: Data analysis, manuscript editing

T Larson: Data collection and management, manuscript editing

J Thiel: Data collection, manuscript editing

HC Chang: Data collection, manuscript editing

JD Harper: Project development, manuscript editing

MR Bailey: Project development, manuscript writing/editing

MD Sorensen: Project development, data collection, manuscript editing

Conflict of Interest: Michael R. Bailey, Barbrina Dunmire, and Mathew D. Sorensen have equity in and consulting agreements with SonoMotion Inc. which has licensed intellectual property from the University of Washington related to this technology. For the remaining authors, no competing conflicts of interest exist.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study formal consent is not required.

independently assessed the largest stone in each image set for shadow presence and size. Shadow size was compared to US and CT stone sizes.

Results—Eighty percent of included stones demonstrated anacoustic shadow; 83% of stones without a shadow were ≤ 5 mm on CT. Average stone size was $6.5 \text{ mm} \pm 4.0$ on CT, $10.3 \text{ mm} \pm 4.1$ on US, and $7.5 \text{ mm} \pm 4.2$ by shadow width. On average, US overestimated stone size by $3.8 \text{ mm} \pm 2.4$ based on stone width ($p < 0.001$) and $1.0 \text{ mm} \pm 1.4$ based on shadow width ($p < 0.0098$). Shadow measurements decreased misclassification of stones by 25% among three clinically relevant size categories ($\leq 5 \text{ mm}$, $5.1 - 10 \text{ mm}$, $> 10 \text{ mm}$), and by 50% for stones $\leq 5 \text{ mm}$.

Conclusions—US overestimates stone size compared to CT. Retrospective measurement of the acoustic shadow from the same clinical US images is a more accurate reflection of true stone size than direct stone measurement. Most stones without a posterior shadow are $\leq 5 \text{ mm}$.

Keywords

Ultrasonography; calculi; nephrolithiasis; urolithiasis; computed tomography; size

Introduction

Stone size is a critical factor in determining management options for urolithiasis [1, 2]. Previous studies have demonstrated that ultrasound (US) tends to overestimate average stone size by 1.5 to 2.2 mm [3–5]. However, interest in US for the diagnosis and management of nephrolithiasis has grown amongst concerns regarding cost and radiation exposure associated with computed tomography (CT), which remains the imaging modality of choice [6–8].

To improve stone size accuracy with US, the posterior acoustic shadow width has been proposed as an adjunctive measure. Using a research US platform, this approach is more accurate than direct measurement of stone size in phantom models and human subjects [9, 10]. On US, renal stones appear as a hyperechoic signal with a hypoechoic shadow extending behind the stone. Whereas system settings and US imaging modality can affect the appearance of the hyperechoic boundaries, therefore impacting measured stone size, the acoustic shadow is generally unaffected [11].

In 2016, Sternberg and colleagues published a retrospective multi-institutional study of patients undergoing formal renal US and CT [5]. They reported an average overestimation of US stone size of 2.2 mm, with even greater overestimation for small stones $\leq 5 \text{ mm}$ (mean 3.3 mm). In the present study, a subset of these US images was retrospectively reviewed for the presence of a posterior acoustic shadow. Shadow width was sized and compared to reported US and CT measurements. This is the first validation of posterior acoustic shadow measurements in clinical images obtained using a commercial US system.

Methods

As described in the multi-institutional study by Sternberg et al [5], data were originally collected from patients undergoing formal renal US and low-dose non-contrast CT within 1

day of each other, for three cohorts at the University of Vermont Medical Center, Massachusetts General Hospital, and Dartmouth-Hitchcock Medical Center. Length, width, and height were determined on axial, coronal, and sagittal sections for the largest stone on CT, and reference stone size was determined by the largest dimension. Clinical US examinations were performed by licensed sonographers and formally read by radiologists as part of clinical care. US stone size was based on the largest reported stone measurement in any dimension. Imaging was optimized per the sonographer's preference without software or hardware system modifications.

In the present study, de-identified CT data and US images from the University of Vermont cohort were obtained for retrospective review. Stones with incomplete imaging and images with multiple overlapping stones were excluded. The largest stone in each remaining image set was then independently reviewed by 5 individuals (a sonographer, US engineer, endourologist, endourology fellow, and PGY4 urology resident) at the University of Washington for the presence and size of the posterior acoustic shadow. All reviewers were blinded to original US and CT stone measurements. Stone depth on US was recorded as a surrogate for body mass index (BMI).

Shadow Sizing Protocol

The method for sizing stone shadows has been reported previously [10]. US images were displayed using a MATLAB™ (MathWorks, Natick, MA), which includes two moveable guide lines and a caliper. Reviewers were instructed to use the caliper to measure the shadow width, and the guide lines to help delineate the shadow borders as needed (Fig 1). There was also a check box if the reviewer felt no shadow was present. All measurements were made approximately 1 cm posterior to the stone. No numerical information was included to indicate stone size or caliper length.

Statistical Analysis

Stone shadow was deemed present if at least one reviewer made a size measurement. This was reported as a dichotomous variable. Absolute bias between stone and shadow measurements on US was determined and compared to the reference CT stone size using a linear mixed effects model to account for within-stone correlations. This was reported as mean bias. Average CT stone size, US stone size, and shadow width were also calculated.

Based on CT size measurements, all shadowing stones were placed into clinically relevant size categories previously used within the literature (≤ 5 mm, 5.1–10 mm, and >10 mm) [4, 5, 12]. Within these groups, concordance of each stone's CT size with its reported US size and average shadow measurement was determined. For each size category, average US stone size, shadow size, and magnitude of deviation from CT stone size were, also calculated.

Inter-rater reliability of shadow measurements was assessed by intra-class correlation (ICC) with a 95% confidence interval. Two-sided p-values < 0.05 were considered statistically significant. All analyses were performed in SAS 9.4 (SAS Institute, Cary, NC, USA). This retrospective study received institutional review board approval.

Results

Forty-four image sets were evaluated for inclusion. Fourteen were excluded because of incomplete imaging, lack of imaging reports, or multiple overlapping stones in the images precluding identification of the primary stone or stone shadow. Of the 30 stones included for review, one was a ureteral stone and the remainder were renal stones. The mean CT stone size was 6.5 mm \pm 4.0. Mean reported US stone size was 10.3 mm \pm 4.1, at a mean depth of 6.8 cm \pm 2.1.

Overall, 80% (24/30) of included stones demonstrated a posterior acoustic shadow. Average shadow width was 7.5 mm \pm 4.2, with ICC of 0.86 (95% CI 0.77, 0.94) among all reviewers. Of those stones without a shadow, 83% (5/6) were \leq 5 mm. A post-hoc power analysis showed a sample size of 14 pairs as sufficient to obtain 80% power at type I error level 0.05 in comparing US stone measurements.

Shadow width more closely approached CT stone size than US stone size (Fig 2), particularly for smaller stones. Reported US stone size consistently overestimated CT stone size with a mean bias of 3.8 mm \pm 2.4 ($p < 0.001$), while shadow width demonstrated a mean bias of 1.0 mm \pm 1.4 ($p < 0.0098$). The difference in mean absolute bias between US stone and shadow measurements was statistically significant ($p < 0.0001$). Using the shadow measurement, there was a greater than 3-fold improvement in the number of stones measuring within 1 mm of the CT measurement (10/24), compared to reported US stone size (3/24).

Shadowing stones were categorized by CT size: \leq 5 mm ($n = 12$), 5.1–10 mm ($n = 7$), and >10 mm ($n = 5$) (Table 1). Mean CT stone size, mean reported US stone size, and mean shadow size for each group is listed in Table 2a. Using reported US stone size, 58% of all stones were misclassified; using shadow size, misclassification was reduced to 33%. The greatest degree of misclassification was among stones \leq 5 mm, where misclassification was reduced from 91.7% using US stone measurements to 41.7% using shadow measurements. Absolute deviation from CT stone size for this subgroup was reduced by 63%, from 4.3 mm \pm 2.7 to 1.6 mm \pm 1.6 on average, respectively, with notably less variability (Table 2b).

Discussion

Prior phantom and human studies using a research ultrasound platform have indicated that posterior acoustic shadow width measurement improves the accuracy of stone size measurement on US [9, 10]. This is the first study to validate this technique in human subjects using a clinical ultrasound system at a second institution. We demonstrate that retrospective measurement of shadow width improves size accuracy on US compared to reported stone size, and approaches the 1 mm accuracy of CT reported by Kishore and colleagues [13]. Shadow measurements also reduced misclassification of stones in clinically relevant size categories by 25% overall and 50% for stones \leq 5 mm. Though the overall degree of size misclassification in this study is higher than reported by others using the same size categories (27–28% overall and 17–60% for stones \leq 5 mm, respectively) [4, 12], this may be explained by the exclusion of non-shadowing stones from size classification, over

80% of which were ≤ 5 mm. As smaller stones have a high likelihood of spontaneous passage, improved size concordance using shadow measurements is clinically relevant [1, 2] and may impact clinical management decisions [14].

Posterior acoustic shadow measurements may be a particularly useful technique in the acute setting, where there is growing interest in ultrasound as first-line imaging for patients with suspected nephrolithiasis [6, 15]. This approach addresses the inherent tradeoffs with CT between optimizing size accuracy and minimizing radiation exposure. By improving stone size accuracy and identification of small, passable stones on US, triage and management decisions may be facilitated without CT, which can be reserved only for ambiguous cases. This may be most useful for children, young women, and recurrent stone formers, where minimizing ionizing radiation is a priority.

The software interface in this study simulates sizing calipers on standard ultrasound platforms, making this technique easy to apply within existing clinical ultrasound systems without any additional hardware or software modifications. Moreover, caliper use should already be familiar to practicing sonographers, radiologists, and clinicians. This allows the technique to be easily integrated into clinical use.

This is a small retrospective study subject to the potential biases of such a study design, and it has several additional limitations. The US images obtained were performed for clinical purposes, and were not specifically optimized for the detection or sizing of the posterior acoustic shadow. Moreover, reviewed images were limited to only those saved within the formal study, rather than the full complement of ultrasound images seen in real time by the sonographer. However, this is reflective of US imaging in the “real world,” where the stone or posterior acoustic shadow may be sub-optimally captured due to time or operator dependence, and providers are dependent on the technical expertise of their sonographers.

As images were de-identified, no clinical data on body mass index or body habitus were available. Such factors potentially influence image quality and the appearance of the posterior acoustic shadow. In vitro, the accuracy of US stone size worsened with increasing depth, but this does not hold true for shadow measurements [9]. In this study, average stone depth for all included stones was $6.9 \text{ cm} \pm 2.1$, with a maximum depth of 11.8 cm. This suggests that the shadow measurement remains feasible for obese patients.

CT and US stone measurements were obtained by a single reviewer, whereas shadow measurements were evaluated across five reviewers. However, this was accounted for with a linear mixed effects model. As an additional sensitivity analysis, a second provider reviewed the US images and re-measured stone size in a blinded fashion. Though average US stone size measurements were significantly different between reviewers ($10.3 \text{ mm} \pm 4.1$ vs. 8.4 ± 3.2 , $p < 0.0001$; ICC 0.72), this did not alter stone size misclassification, or the superior accuracy of shadow measurement compared to US stone size ($p < 0.0001$ for reviewer 1 and $p < 0.0099$ for reviewer 2). Such variation is not unique to US, as CT stone size measurements have also been reported to vary among radiologists [15, 16]. Moreover, our reviewers included 2 individuals naïve to the shadow measuring technique, suggesting that the high ICC for shadow measurements is not dependent on experienced reviewers.

Despite these limitations, this is the first study to assess posterior acoustic shadow in human subjects with a commercial US unit in the clinical setting. Moreover, it validates the improved accuracy of this technique utilizing images obtained by blinded sonographers at a second institution. This technique requires no software or hardware modification and is available for immediate clinical use. Further multi-institutional studies with greater sample sizes may better clarify the accuracy, reproducibility, transferability, and clinical utility of this approach to measuring stone size on US.

Conclusions

Direct measurement of stones on US consistently overestimates stone size compared to CT imaging, while the posterior acoustic shadow width appears more accurate. This has been shown using clinical US images obtained in human subjects. On average, the shadow overestimates CT stone size by about 1 mm and decreases misclassification of stones within clinically relevant size categories, particularly for stones ≥ 5 mm. Most stones without a posterior shadow are ≥ 5 mm. These are clinically significant findings, as these stones have the highest probability of spontaneous passage.

Acknowledgments

This work is part of a large collaborative effort, and we appreciate the help of our many collaborators at the University of Vermont, University of Washington (UW) Center for Industrial and Medical Ultrasound, the UW Department of Urology, and within National Institute of Diabetes and Digestive Kidney Diseases (NIDDK) Program Project DK043881.

Funding: Funding was provided by the National Space Biomedical Research Institute through National Aeronautics and Space Association (NASA) Grant NCC 9-58 and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Grant DK043881.

Abbreviations

CT	computed tomography
US	ultrasound
BMI	body mass index
ICC	intra-class correlation

References

1. Hubner WA, Irby P, Stoller ML. Natural history and current concepts for the treatment of small ureteral calculi. *Eur Urol.* 1993; 24:172–176. [PubMed: 8375436]
2. Assimos D, et al. Surgical Management of Stones: American Urological Association/Endourological Society Guideline, PART II. *J Urol.* 2016; 196:1161–1169. [PubMed: 27238615]
3. Fowler, Ka B., Locken, JA., Duchesne, JH., Williamson, MR. US for detecting renal calculi with nonenhanced CT as a reference standard. *Radiology.* 2002; 222:109–13. [PubMed: 11756713]
4. Ray AA, Ghiculete D, Pace KT, Honey RJD. Limitations to ultrasound in the detection and measurement of urinary tract calculi. *Urology.* 2010; 76:295–300. [PubMed: 20206970]
5. Sternberg KM, et al. Ultrasonography Significantly Overestimates Stone Size When Compared to Low-dose, Noncontrast Computed Tomography. *Urology.* 2016; 95:67–71. [PubMed: 27289025]

6. Smith-Bindman R, et al. Ultrasonography vs Computed Tomography for Suspected Nephrolithiasis. *N Engl J Med*. 2014; 371:1100–1110. [PubMed: 25229916]
7. Ferrandino MN, et al. Radiation Exposure in the Acute and Short-Term Management of Urolithiasis at 2 Academic Centers. *J Urol*. 2009; 181:668–673. [PubMed: 19100573]
8. Coursey, Ca, et al. ACR Appropriateness Criteria® Acute Onset Flank Pain–Suspicion of Stone Disease. *Ultrasound Q*. 2012; 28:227–233. [PubMed: 22902840]
9. Dunmire B, et al. Use of the Acoustic Shadow Width to Determine Kidney Stone Size with Ultrasound. *J Urol*. 2016; 195:171–176. [PubMed: 26301788]
10. May P, et al. Stone-Mode Ultrasound for Determining Renal Stone Size. *J Endo*. 2016; 30:958–962.
11. Dunmire B, et al. Tools to Improve the Accuracy of Kidney Stone Sizing with Ultrasound. *J Endourol*. 2015; 29:147–152. [PubMed: 25105243]
12. Kanno T, et al. The efficacy of ultrasonography for the detection of renal stone. *Urology*. 2014; 84:285–288. [PubMed: 24908592]
13. Kishore TAA, Pedro RN, Hinck B, Monga M. Estimation of Size of Distal Ureteral Stones: Noncontrast CT Scan Versus Actual Size. *Urology*. 2008; 72:761–764. [PubMed: 18701151]
14. Ganesan V, De S, Greene D, Torricelli FCM, Monga M. Accuracy of ultrasonography for renal stone detection and size determination: is it good enough for management decisions? *BJU Int*. 2017; 119:464–469. [PubMed: 27459091]
15. Sternberg KM, Littenberg B. Trends in imaging use for the evaluation and follow-up of kidney stone disease: A single center experience. *J Urol*. 2017; 198:383–388. [PubMed: 28161351]
16. Patel SR, et al. Automated renal stone volume measurement by noncontrast computerized tomography is more reproducible than manual linear size measurement. *J Urol*. 2011; 186:2275–2279. [PubMed: 22014818]
17. Lidén M, Andersson T, Geijer H. Making renal stones change size-impact of CT image post processing and reader variability. *Eur Radiol*. 2011; 21:2218–2225. [PubMed: 21698465]

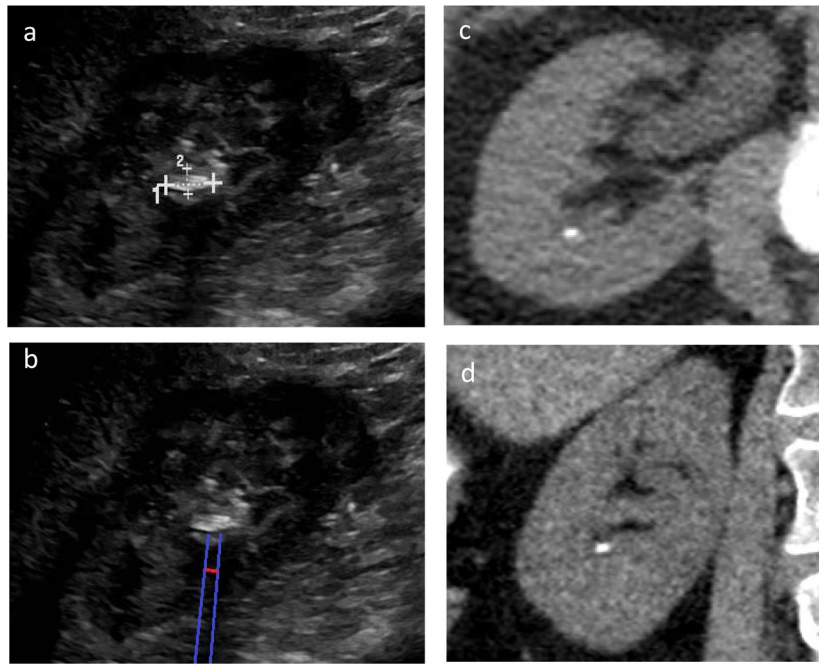


Fig. 1. Example of ultrasound (US) stone and shadow measurement. (a) original US image, (b) reported (US) stone size, and (c) US shadow size. The blue lines correspond to moveable guide lines used to facilitate delineation of the shadow. The red line corresponds to the measurement caliper, positioned about 1 cm posterior to the stone. This stone measured 1.6 mm on CT, 9.8 mm on US, and 4.9 mm using the shadow measurement averaged across 5 unique reviewers (standard deviation 1.23 mm, range 3.1–6.2 mm).

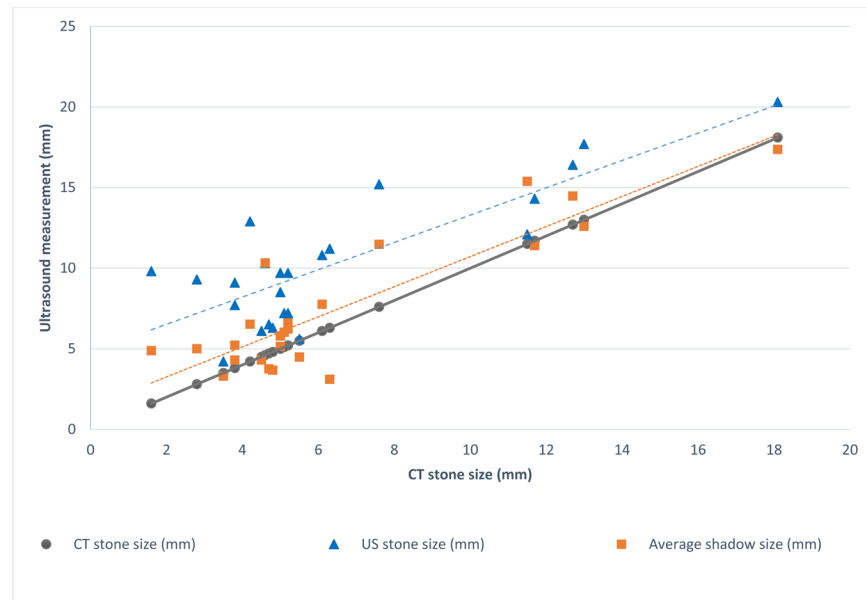


Fig. 2. Comparison among reported ultrasound (US) stone size, average posterior acoustic shadow size, and measured stone size on CT. The stone measurement is represented by the blue triangles and average shadow measurement by the orange squares. The blue (long) dashed line represents the trendline for the stone size on US and the orange (short) dashed line represents the trendline for the average shadow size on US. The solid gray line represents the 1:1 correspondence between the measurements. ICC for the averaged shadow measurements was 0.86. The maximal difference between CT and US stone size was 8.7 mm, whereas the maximal difference between CT and shadow size was 5.7 mm

Table 1

Concordance of stone size based on (a) reported US stone measurements and (b) mean posterior acoustic shadow measurement, for 3 clinically significant size categories. Percent concordance with CT size measurement within each size category is listed in (%)

	CT measurement			
		<i>0-5mm</i>	<i>5-10 mm</i>	<i>>10 mm</i>
a) US Stone measurement	<i>0-5 mm</i>	1 (8.3%)	0	0
	<i>5.1-10 mm</i>	9	4 (57.1%)	0
	<i>>10 mm</i>	2	3	5 (100%)
b) Mean posterior acoustic shadow measurement		CT measurement		
		<i>0-5mm</i>	<i>5-10 mm</i>	<i>>10 mm</i>
	<i>0-5 mm</i>	7 (58.3%)	2	0
	<i>5.1-10 mm</i>	4	4 (57.1%)	0
	<i>>10 mm</i>	1	1	5 (100%)

Table 2a

Mean reported ultrasound (US) stone measurement and mean posterior acoustic shadow measurement, categorized based on CT measurement of largest stone diameter

CT size	Mean CT Size (mean \pm SD)	Mean US Size (mean \pm SD)	Mean Posterior Acoustic Shadow Size (mean \pm SD)
0-5 mm (n=12)	4.0 mm \pm 1.0	8.4 mm \pm 2.3	5.2 mm \pm 1.9
5.1-10 mm (n=7)	5.9 mm \pm 0.9	9.6 mm \pm 3.2	6.5 mm \pm 2.7
>10 mm (n=5)	13.0 mm \pm 2.7	15.2 mm \pm 3.1	14.2 mm \pm 2.4

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2b

Deviation from CT stone size measurement of reported US stone measurement and average posterior acoustic shadow measurement, by size category. Magnitude of mean size difference from CT size is reported with standard deviation (SD)

CT size	% Overestimated By US Stone Size (mean \pm SD)	% Overestimated By Shadow Size (mean \pm SD)	% Underestimated By shadow Size (mean \pm SD)
0–5 mm (n=12)	100% (4.3 mm \pm 2.7)	66.7% (2.0 mm \pm 1.8)	33.3% (0.63 mm \pm 0.5)
5.1–10 mm (n=7)	100% (3.7 mm \pm 2.5)	71.4% (1.8 mm \pm 1.2)	28.6% (2.1 mm \pm 1.5)
>10 mm (n=5)	100% (2.8 mm \pm 1.6)	80% (1.7 mm \pm 1.6)	20% (0.55 mm \pm N/A)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript