Endourology and Stones

Accuracy of Ex Vivo Semiautomatic Segmentation of Urinary Stone Size in Computed Tomography Compared With Manual Size Estimation in Radiographic Correlation


OBJECTIVE
To evaluate the accuracy of semiautomated segmentation of urinary stone size in computed tomography (CT) compared with manual measurement.

MATERIALS AND METHODS
A total of 103 patients (32f, 71m; mean age 52 years ± 18) that were diagnosed with urolithiasis and collected stones received standardized ex vivo CT-scans and radiography of the stones. Stone size was segmented semiautomatically using commercial software (syngo.via, Siemens, Germany) and compared with manual caliper measurement on digital radiography.

RESULTS
Mean size was 4.4 mm in CT and 4.6 mm in radiography. Depending on number of stones analyzed per patient, estimation of stone size showed moderate to excellent correlation for both methods. There was no significant difference in overall size measurement.

CONCLUSION
Semiautomatic segmentation of urinary stone size in CT is possible and reduces measurement errors, allowing more precise estimation especially for smaller concrements. Neighboring stones may hamper segmentation of stone size.

Urinary stone formation is a widespread disease with an increasing incidence and prevalence worldwide that appears even more pronounced in industrialized countries.

If conservative and/or expectative management (observation) is not an option, there is a variety of modern treatment modalities, such as shock wave lithotripsy, percutaneous nephrolithotomy, and ureterorenoscopy with laser-lithotripsy.1-3 Commonly, stones are stratified into 4 size groups, those measuring up to 5 mm, 5-10 mm, 10-20 mm, and those measuring over 20 mm in the largest diameter.2 Furthermore, evaluation of procedure selection depends on localization of calculi and is generally described using 4 different levels of stone localization such as kidney, lower pole, proximal, and distal ureter.1-3

The diagnostic and therapeutic pathway for treatment of urinary calculi is mainly based on the size of the stone. Therefore, accurate estimation of stone size in computed tomography (CT) for suspected urolithiasis is crucial for adequate treatment decision and patient triage.1-5 Stone size is the most important parameter to estimate the likelihood of spontaneous stone passage and the decision of the type of intervention.1,2 It has been reported that ureteral stones up to 4 mm pass in 95% within 40 days.6 Noncontrast CT is the gold-standard for diagnosing and measuring urinary stones.2

With regard to stone size, most studies consider only the maximum transverse diameter as measured on noncontrast CT imaging. Several studies have shown that measuring stone size on transverse CT images alone often underestimates the burden of stone disease.7-9 When examining small asymptomatic stones or large symptomatic stones, accuracy and reproducibility are important during the imaging process. Although CT is the most accurate imaging modality to linearly measure stone size, interobserver variability can occur based on how a radiologist manually measures stone size.10 Automated CT measurements could dramatically reduce interobserver variability, and it would help standardize stone measurement variability among radiologists and institutions.10

Especially in smaller stones that are mostly found in ureterolithiasis, the inter-reader variability is known to be very high for manual caliper measurements and measurement errors must be taken into account.11 Lidén et al have...
shown that segmentation algorithm-based CT methods can significantly improve the inter-reader agreement for stone sizing. Furthermore, they have shown that 3-dimensional (3D) reconstruction CT stone sizing improves the accuracy of sizing maximal stone diameter compared with 2D CT imaging only. Recently, it was demonstrated that automated renal stone volume measurement by CT is more reproducible than linear measurement.

Hence, inaccurate stone sizing can lead to incorrect treatment decisions. This can lead to deferred treatment or even over-treatment. Furthermore, the success (eg, stone-free rate; re-intervention rate) of the selected procedure depends predominantly on stone size.

However, guidelines panels, such as the high impact EAU guidelines panel have difficulties of suggesting an exact stone size threshold for the recommendation of conservative treatment and therefore include the word “small” instead of a maximal diameter length. This is predominantly because of individual differences in stone shape and ureter width, but might also be due to interobserver variability of stone sizing, leading to heterogeneous results.

Contrary to percutaneous nephrolithotomy, the success (eg, regarding stone-free rate, re-treatment rate) of ureterorenoscopy is strongly dependent on stone size. This principle is true for shock wave lithotripsy alike, showing a high risk of ureteral obstruction with colic and steinstrasse leading to high rates of retreatment, especially with stone size over 20 mm.

Semiautomatic estimation of stone diameter by computer-assisted segmentation of the stone and derived 3D calipers can be a powerful tool in reducing measurement errors. Especially the largest diameter is crucial for ureter passage. Linear measurements in the transversal plane only in CT may be hampered by multiple sources of errors.

The purpose of our study, therefore, was to evaluate ex vivo the accuracy and reproducibility of a commercial semiautomated segmentation software tool (syngo.via, Siemens, Germany) for urinary stone size estimation compared with manual measurement of real stone size on digital radiography.

METHODS
The study was approved by the local ethical review board with general consent of the patients for scientific use of biological material.

We retrospectively evaluated a total of 103 patients that were diagnosed with urolithiasis. All patients were asked to collect passed stones after diagnosis and all patients retrieved stone samples.

All stone samples underwent ex vivo digital radiography and standardized ex vivo CT (Figs. 1-3).

Imaging
Digital radiography of all stone samples was performed using Polydoros R80/Ysio (Siemens, Erlangen, Germany) with standardized imaging parameters (45 kV, 2.2 mAs).

Ex vivo CT was performed of all stone samples on Siemens Definition Flash scanner (Siemens, Erlangen, Germany), using a standardized low dose none contrast dual energy protocol (Sn140/100 kV, 70/90 ref. qual. mAs).

Direct contact or overlapping position of stones was avoided during both imaging acquisitions.

Size Measurements
Stone size was segmented semiautomatically using commercial software (syngo.via, VB20A, Siemens, Germany) and compared with manual caliper measurement on digital radiography.
Noninferiority of the segmentation tool was assessed. Digital radiography size estimation was assumed as representing nearest approximation to real stone size.

For manual caliper measurement, digital radiography images of the stones were analyzed using a certified PACS workstation (Centricity PACS 4.0, GE Barrington, IL). For exact measurement, zoom factor and windowing was free to be chosen. The maximum in plane diameter of the largest stone visible was estimated on calibrated images.

Semiautomatic stone size segmentation was performed using syngo.via software. Maximum diameter of the largest stone visible was measured using the generic segmentation tool. Soft tissue image reconstructions (131f, 1.0 mm slice thickness, 1 mm increment) were used for semiautomatic analysis.

Statistics

The sample size was computed using the stone size as the primary outcome. It was computed on the basis of a noninferiority hypothesis, estimating that a small difference in measured stone size would be acceptable. On manual measurements, the mean stone size was previously estimated as 4 mm (±3 mm).27 We hypothesized that a measurement error between 5% and 10% might be acceptable with the automatic segmentation technique. Consequently, the noninferiority margin was fixed at 7.5% of the mean stone size. An estimated 87 patients were necessary for 80% power with significance at 5%, rounded up to 90 subjects, a total of 103 consecutive patients were evaluated.

Results were expressed as means and standard deviations (SD) for continuous variables, and frequencies and percentages for categorical variables.

The noninferiority hypothesis was assessed using a 2-sided 95% confidence interval for the true difference between the two paired means of stone size measured for the largest diameter. For continuous variables, comparisons between groups were performed using a paired Student’s t test. Statistical significance was defined as P < .05.

RESULTS

Our patient collective consisted of 103 patients (32 female [31%], 71 male [69%]; mean age 52 ± 18 years). All patients collected and retrieved stone samples after diagnosis of urolithiasis. All stone samples underwent ex vivo digital radiography and ex vivo CT.

The overall mean stone size measured was 4.6 mm (±1.46 mm SD, median 4.4 mm) in the manual caliper size estimations. In semiautomatic size estimation, we measured a mean maximum diameter of 4.4 mm (±1.45 mm SD, median 4.3 mm). There was an overall mean difference of −0.18 mm between the two measurement approaches.

In the subgroup analysis of the stones ≤4.3 mm (n = 52) a mean size of 3.39 mm (±0.58 mm SD, median 3.4 mm) was measured for manual caliper size estimation. In semiautomatic size estimation, we measured a mean maximum diameter of 4.2 mm (±1.44 mm SD, median 3.8 mm) for the same subgroup respectively.

In the subgroup analysis of the stones >4.3 mm (n = 51), we measured a mean size of 5.72 mm (±1.08 mm SD, median 5.5 mm) for manual caliper estimation a diameter of 4.57 mm (±1.44 mm SD, mean 4.3 mm) for semiautomatic estimation respectively.

The noninferiority hypothesis was verified as the difference in the overall mean size between the two measurements was 0.18 mm with a confidence interval of 0.08-0.28 mm.

Mean size was 4.4 mm in CT and 4.6 mm in radiography. Depending on number of stones analyzed per patient, estimation of stone size showed moderate to excellent correlation for both methods. There was significant difference in overall size measurements (P < .005).

DISCUSSION

Accurate measurement of structures and objects in radiology remains a technical and methodical challenge. There are several factors that may influence measurement accuracy on the side of the measurement performing reader.28 Since reconstructed CT images have both limited spatial resolution and matrix size the accuracy of spatial measurements especially of small objects is technically limited. There is blurring of the margins due to these limits that may hamper exact definition of the size limits to measure. The larger the objects the smaller this influence. Since urinary stones are usually of smaller size, a substantial measurement error must be expected. Today there exist a variety of software tools of different vendors that allow for facilitated standardized size estimations and research has shown the superiority for reproducibility and accuracy compared with manual

FIGURE 3. Bland-Altman plot for variability of stone size between the different approaches $s_1$(radiography) − $s_2$(CT) (upper dotted line: mean + 1.96 SD, lower dotted line: mean − 1.96 SD, bold line: mean). (Color version available online.)
measurements. However, there exist various sources for measurement errors both in automatic segmentation and in manual caliper measurements. These are due to technical reasons as collimation size, reconstruction overlap, image matrix, image noise, segmentation errors but also due to observer-related errors such as choice of magnification, definition of object margins, choice of reconstruction kernel for measurement. Even if the semiautomatic size measurements meet the criteria for noninferiority, the stones analyzed were significantly smaller than manual caliper measurements due to these technical limits. This may be more important for smaller increments, since these effects will relatively increase for smaller sizes.

Since digital radiography images allow for calibrated measurements of stone size and have the lowest source for measurement errors compared with CT images we assumed the radiography size estimation as nearest approximation for true stone size. The remaining known sources for measurement error were assumed to be negligible in the clinical context of urolithiasis.

Even if 3D reconstructions for size estimations have been shown to be superior to considering only the maximum transverse diameter, measuring in the transverse plane still remains the standard in most studies. Deriving the maximum diameter from segmented stones considers all dimensions and may prevent underestimation of stone burden as described when considering transverse dimensions only.

To our knowledge, there is no publication that estimates the effects on accuracy of software-based segmentation tools with regard to the definition of real stone size in ex vivo scans (Figs. 1-3).

Our results showed good correlation between manual stone size definition and semiautomated segmentation of urinary stones. There was 1 outlier that was estimated with 6.3 mm maximum diameter on radiography, but the segmentation yielded a diameter of 4.2 mm (Fig. 4). Laboratory analysis revealed a mixed stone of calcium dihydrate and calcium phosphate. That stone had a very blurry, cloudy surface morphology. The segmentation software was not able to define the margins correctly.

There are limitations in our study. First, as we aimed at comparing size estimation by semiautomated segmentation with the real stone size measured in digital radiography, we hypothesized that the stone visualized on the projected radiography image was lying along its longest border. Oblique orientation of the stone during radiography might have underestimated stone size. Uncertainties with regard to real stone size have been described before. In contrast to Kishore et al, we performed ex vivo scans to prevent stone fragmentation during passage and therefore measurement bias. Second, CT images used for size estimation can be modified by the reconstruction parameters. We used our standard scan protocol with 0.6 mm collimation for image acquisition and 1 mm/1 mm slice thickness/increment for image reconstruction. This may yield in less noise for the prize of blurry edges. Acquisition parameters as described by other authors do not reflect recent technical standards. This balance between image noise and image sharpness is part of daily radiology routine, we consider this effect as not significant. Third, our study was not designed to support former results that showed the superiority of software-based automated segmentation tools for reproducibility and reduction of measurement errors. But our impression strongly supports these results. In daily radiology routine, our residents have a strong tendency to inaccurately measure urinary stones. Use of insufficient magnification and frantically size estimation in high-speed scheduled CT tend to overestimate stone size in our impression.

However, to our knowledge, there are no studies that evaluate not only reproducibility of measurements but also the accuracy related to the real stone size to be estimated in CT imaging. In this manuscript, we accurately compare size estimation of a software-based algorithm for stone size assessment in our standardized low-dose CT protocol with ex vivo digital radiography size measurement. In our results, the noninferiority hypothesis was verified, proving the applicability of the semiautomated segmentation approach. But due to technical limits as discussed, CT measurements especially of smaller stones may always be hampered by measurement errors.
CONCLUSION
Semiautomatic segmentation of urinary stone size in CT is possible and may allow for reduction of manual measurement errors, allowing more precise estimation especially for smaller concrements.

Even semiautomated segmentation of stone size not always results in estimation of real stone size. Measurement errors always have to be taken into account in the image interpretation when using computer-assisted support in routine radiology.

SUPPLEMENTARY DATA
Supplementary Data associated with this article can be found, in the online version, at doi:10.1016/j.jurology.2018.06.044.

References

EDITORIAL COMMENT
Computed tomography (CT) is now considered the gold standard imaging modality for the diagnosis of renal and ureteral stones. With this technology, there are additional capabilities, such as the semiautomatic measurement of a variety of stone parameters such as size and volume, which cannot similarly be obtained by traditional plain radiography.

In this study, the authors performed a retrospective trial to assess if ex vivo semiautomatic size measurement of a stone on CT was non-inferior to manual size measurement using calipers on digital plain radiography. The measurement on
Digital radiography was considered the gold standard. All measurements were reported as the maximal diameter in any plane. The threshold for non-inferiority was arbitrarily set at a difference of 7.5%. The mean size difference observed was 0.18 mm between the two techniques, which met the threshold, and thus semiautomatic measurement was non-inferior to manual measurement.

Although it is important to validate these semiautomatic measurements obtained via CT against a gold standard benchmark, a non-inferiority study comparing two different imaging modalities with two different measurement methods is likely not the ideal trial design to answer this question. Despite this significant limitation, the authors detected only <1 mm difference in size between the two imaging modalities, which is clinically insignificant. This suggests that semiautomatic size measurement via CT is accurate. More importantly, this study highlights that CT imaging has value beyond simply stone detection. Leveraging the captured image metadata to semi-automatically calculate clinically relevant stone parameters, such as size and volume will allow clinicians and researchers to better understand the natural history of stone disease and assess the efficacy of available treatment options.

Justin B. Ziemba, Division of Urology, Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

https://doi.org/10.1016/j.urology.2018.06.046

AUTHOR REPLY

We thank you for your interesting and helpful remarks.

Computed tomography (CT) is indeed nowadays widely accepted as gold standard for diagnosis of urinary stones. Beyond computer assisted software tools for more precise and replicable size estimation, there has been enormous evolution in scan technology over the last years resulting in further substantial reduction in radiation dose.1 Additionally new dual energy scan techniques allow for direct determination of stone composition.2 New photon counting detector based CT may further increase radiology performance in the future.3

The idea of our study originated due to the mistrust of urologists in the semiautomatic size estimations. Therefore we still consider our methodology of non-inferiority hypothesis as appropriate.

There will be further evolution of radiology imaging of urinary stones, especially ongoing research using dark field imaging seems to be promising for further imaging based stone characterisation.4 In contrast to the well-known absorption-based imaging, dark-field contrast is generated by diffuse angular deflections of the X-ray wavefront when being scattered at inherent substructures. Since the dark-field signal has been shown to be highly dependent, not only on the chemical composition of imaged samples, but also on the samples morphological structure well below the resolution limit of commonly used imaging detectors it may allow for further discrimination of urinary stone subtypes.5

T. Niemann, I. Jerjen, PhD, L Hefermehl, MD, Z Wang, PhD, M Stamparoni, Prof, K Lehmann, MD, RA Kubik-Huch, MD Prof; Department of Radiology, Kantonsspital Baden, Baden, Switzerland; Department of Urology, Kantonsspital Baden, Baden, Switzerland; University and ETH Zürich and Paul Scherrer Institut, Zürich, Switzerland

References

https://doi.org/10.1016/j.urology.2018.06.047