Published in final edited form as:

Acad Radiol. 2019 July ; 26(7): 885–892. doi:10.1016/j.acra.2018.08.010.

Robustness of textural features to predict stone fragility across CT acquisition and reconstruction parameters

Taylor Moen, Andrea Ferrero, PhD, and Cynthia McCollough, PhD*

Department of Radiology, Mayo Clinic, Rochester, MN

Abstract

Rationale and Objectives: Previous studies have demonstrated that quantitative relationships exist between stone fragility at lithotripsy and morphological features extracted from computed tomography (CT) scans. The goal of this study was to determine if variations in scanner model, patient size, radiation dose, or reconstruction parameters impact the accuracy of the prediction of renal stone fragility in an *in vitro* model.

Materials and Methods: Sixty-seven kidney stones were scanned using routine single and dual energy stone protocols, mimicking average and large patient habitus. Low dose scans were also performed. Each scan was reconstructed with routine protocol parameters, and with thinner (0.6mm) or thicker (3mm) images, two different reconstruction kernels, and iterative reconstruction at two strengths. Fragility of each stone was measured in a controlled *ex vivo* experiment. A single predictive model was developed from a reference CT protocol configuration and applied to data from each CT acquisition and reconstruction parameter tested to obtain estimated stone comminution times.

Results: None of the investigated protocols showed a significant variation in the accuracy of stone fragility classification, except for the ones with the most aggressive iterative reconstruction and/or with thicker images. In these protocols, a number of stone fragility assessments changed from fragile to hard (or vice versa), compared to their ground truth measurement.

Conclusion: Prediction accuracy of stone fragility models developed from CT data is robust to expected variations in CT stone protocols used for quantification tasks. This finding facilitates their future adoption to different clinical practices.

Keywords

Computed tomography; kidney calculi; percutaneous nephrolithotripsy; stone fragility prediction

*Corresponding Author 200 First Street SW, Rochester, MN 55905, USA, Phone: (507) 284-2511 Fax: (507) 266-3661
mccollough.cynthia@mayo.edu.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the AAPM or the National Institutes of Health.

Selected sections presented during poster presentation at AAPM 2017

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Annual estimates of economic burden associated with kidney stones now exceed \$5 billion (1), and kidney stone prevalence is on the rise. Affected population in the US has reached over 8% in 2010 (2) compared to 5% in 1994 (3). Unenhanced computed tomography (CT) is the recommended method for diagnosing nephrolithiasis in most cases, as it generates images that provide detailed information about the presence, size, and location of kidney stones (4,5). Additional stone characterization, including mineral composition, is possible if the CT exam is performed with a dual-energy acquisition (6–9). Because treatment options for kidney stones can differ significantly based on stone size and composition, the information provided by CT is essential to identify and evaluate an individual's stone burden and determine the optimal treatment plan (4).

Surgical intervention is used for the 10–20% of kidney stones that fail to pass spontaneously (10). One common procedure is percutaneous nephrolithotomy (PCNL), a minimally invasive stone fragmentation and removal technique using an ultrasonic lithotripter (11). Procedures that fragment stones for removal can vary significantly in length due to the fragility of the stone. If known, stone fragility - defined as the time to comminution by a given surgical procedure (12) - can be used to better evaluate treatment options and prevent unnecessary complications.

Variability in stone fragility has been widely demonstrated. A previous study examining stone fragility based on stone composition concluded that knowing the composition alone may not accurately predict the fragility (13). However, other studies examining homogeneity (14), surface roughness (15), and internal stone structures visible with CT(16,17) have demonstrated that internal and external stone morphological features show correlation with kidney stone fragility. A recent study was demonstrated that quantitative relationships exist between stone fragility and stone morphological features extracted from routine single-energy CT (SECT) and dual-energy CT (DECT) scans (12). These relationships may be used to develop a predictive model of stone fragility for a specific stone removal procedure – e.g., PCNL or extracorporeal shock wave lithotripsy (ESWL) – that could facilitate stratification of symptomatic patients requiring active treatment for stone removal (12).

To be widely applicable to clinical practice, predictive models of stone fragility need to be robust to differences in clinical CT protocols used for suspected urinary stones. In this investigation, we derived a predictive stone fragility model and used it to determine how patient size, radiation dose, scanner model, and image reconstruction impact the accuracy of the prediction of renal stone fragility using SECT and DECT in an *in vitro* model.

Methods and Materials

IRB protocol approval was not required for this non-patient study. However, biospecimen approval was obtained from the institutional biospecimen committee.

Data Acquisition

Sixty-seven human kidney stones from 51 patients were placed in distilled water and hydrated for 24 hours. No more than 2 stones were taken from the same patient to reduce intra-patient correlation. The average stone volume was 702 mm³ (range: 127–5762 mm³), whereas the stone composition was pure uric acid for 15 stones, pure nonuric acid for 24 stones, and mixed for 28 stones. The stones were then embedded into pork gelatin that was placed in several ice cube trays. The trays were placed centrally into a water phantom of 35 cm lateral dimension, modeling an average patient size. The phantom was then scanned using a clinical DECT system (SOMATOM Force, Siemens Healthcare, Germany, hereinafter referred to as system A), with the routine stone composition protocol at our institution (Table 1a). Variations to the routine protocol were individually investigated, including use of a 45cm phantom to model a large patient, use of half the radiation dose, use of a SECT renal stone protocol, and use of a different scanner model from the same manufacturer (SOMATOM Definition Flash, Siemens Healthcare, hereinafter referred to as system B) (Table 1a). Data from each scan were reconstructed with the same parameters used by our clinical protocol (Table 1b) [routine: filtered back projection with 300 mm field-of-view, 1mm image thickness, and quantitative medium-sharp kernel (Qr40 for system A and D30 for system B)], as well as with the following protocol variations: a) minimum image thickness (0.6mm); b) “thick” image thickness (3mm); c) smoother kernel (Qr32 for system A, D24 for system B); d) sharper kernel (Qr54 for system A, D50 for system B); e) average strength iterative reconstruction (ADMIRE level 3 for system A and SAFIRE level 3 for system B); f) strong iterative reconstruction (ADMIRE level 5 for system A and SAFIRE level 5 for system B). A workflow describing the investigated variations in CT acquisition and reconstruction parameters is shown in Figure 1.

Extraction of CT features and measurement of stone fragility

For each scan and reconstruction combination, stones were identified and segmented with our in-house stone analysis software. Texture analysis was performed to quantify stone morphology, as described previously (12). 18 CT features were automatically extracted for each stone, including volume, porosity, CT Number ratio, 12 features describing internal morphology and 3 describing surface morphology. The CT features extracted from the routine protocol with the average 35-cm phantom size were used to derive our predictive model of stone fragility. Of all the possible combinations of CT features, the linear regression model that resulted in the highest adjusted r-squared was used as the predictor of comminution time for this study. Subsequently, this single linear regression model was applied to the set of texture features extracted using the different acquisition and reconstruction permutations. This approach resulted in a predicted time to comminution for each stone in each CT protocol investigated.

Fragility of the stones was measured by disintegrating each stone in a controlled *ex vivo* environment (Figure 2), measuring the time required to obtain full clearance of all stone material in the vial. To the best of our abilities we reproduced the stone disintegration and removal techniques used during PCNL procedures *in vivo* (12).

Evaluation of robustness of stone fragility prediction

Root mean squared error (RMSE (s)) between the measured and predicted comminution time was calculated and compared for each protocol variation. In addition, a classification scheme was applied to the measured comminution times for each kidney stone. The lower quartile was comprised of the stones with the lowest measured comminution times. They were therefore classified as fragile stones. Hard stones were in the upper quartile, and contained the stones with the 25% largest measured comminution times. The remaining stones were in the middle two quartiles and were classified as average stones. The values that define the upper and lower quartiles of the measured comminution times were then used to classify the predicted stone comminution times as fragile, average, or hard. The predicted stone classification for each acquisition and reconstruction were then compared to the classification of measured comminution times. The number of stones where the predicted classification matched the measured classification was recorded as being correctly classified. The remaining stones were recorded as being misclassified. If the misclassified stones were predicted as hard when they were measured as fragile, or if they were predicted as fragile when they were in fact hard stones they were recorded as being severely misclassified. The number of misclassified stones from the routine protocol – which was used to derive the predictive model as described above - was taken as reference for comparison with all other protocol variations investigated. The number of severely misclassified stones for each acquisition and reconstruction were similarly compared. A z-test was used to determine the statistical significance of the difference in classification performance obtained with each protocol variation, using the routine protocol as reference. A diagram describing the evaluation of robustness of stone fragility prediction is shown in Figure 3.

Results

Stone fragility model

The linear regression model that resulted in the highest adjusted r-squared for the reference dataset acquired using the routine protocol was a function of the percentage of voids within the stone volume (porosity), a metric describing stone uniformity (energy), and the mean curvature of its surface. All three variables significantly contributed to the final model ($p < 0.001$ for all) and are defined in Table 2. No multi-collinearity of the included variables was detected (variance inflation factors, VIF, < 2 for all), showing that each variable significantly and independently contributed to the estimation of stone fragility. The adjusted R-squared for this model was 0.62, meaning that 62% of the variability in comminution time for the population of stones was explained by the variations in the morphological features included in the parametric model.

$$\ln(\text{comminution time}) = -10.6725 \text{ porosity} + 18.6842 \text{ energy} + 35.5800 \text{ mean curvature} + 5.0221$$

Influence of CT Protocol on estimated stone fragility

The influence of CT protocol on the visual appearance of renal stones in CT images is shown in Figure 4 for three representative stones. The RMSE in comminution time, with respect to the measured comminution time, is reported in Figure 5 for each acquisition and

reconstruction variation. Across the acquisition settings investigated, the RMSE was < 15 s when compared to the routine dataset. Across the reconstruction parameters investigated, the smoother reconstruction kernel, the thick (3 mm) image thickness and strong iterative reconstruction (IR) yielded the highest RMSE values.

Evaluation of the routine dataset classification resulted in 20 stones that were misclassified relative to the measured comminution time classifications. The numbers of misclassified stones for each acquisition and reconstruction variation are shown in Figure 6. Only the thicker 3mm slice and the strong iterative reconstruction were statistically different than the reference dataset ($p < 0.01$).

The classification performance for every permutation of CT acquisition and reconstruction parameter investigated in comparison to the measured comminution data is shown in Figure 7. Only the thicker 3mm images at each acquisition, and the strong iterative reconstruction with the routine scan are statistically ($p < 0.01$) different than the routine data. Of the stones that were misclassified, the proportion of stones that were severely misclassified (e.g., fragile stones being classified as hard stones, or vice versa) is shown in Figure 8. Only the thicker 3mm slices at each acquisition and the strong iterative reconstruction on the system B produced any severely misclassified stones.

Discussion

In this study, we developed a model to predict stone comminution time from CT image data, using physical comminution time as truth. We then collected CT image data using different acquisition and reconstruction parameters relative to our routine stone protocol to determine if the developed model was robust to different acquisition and reconstruction parameters. That is, we sought to determine how well the prediction model worked when the evaluated image data differed in some way from the data on which the model was built.

The regression model developed in this study utilizes 3 different stone features, i.e., internal voids (porosity), the stone energy, and the mean curvature. The positive or negative coefficient related to the stone feature describes the impact of the feature on the estimated comminution time. A negative coefficient for the stone porosity variable implies that the more porous the stone is, the easier it is to break. A positive sign, as for the stone energy and mean curvature coefficients, implies the opposite: the more uniform the stone is, or the more round, the harder it is to break.

Our analysis showed that variations in patient size, dose, and scanner models do not significantly impact the predictability of the model. Similar results were obtained when testing common changes in the reconstruction parameters. However, when the image resolution degradation (which increases partial volume averaging) became significant (i.e., for very aggressive IR setting or 3 mm images), essential morphological features such as stone porosity and heterogeneity became more difficult to assess and the predictive model lost accuracy, as shown in Figure 4. Other reconstruction parameters such as use of thinner slices, use of average strength IR, or sharper or smoother kernels did not have a significant impact on the estimated stone comminution times.

Analysis of the RMSE for the investigated acquisition settings showed that multivariate models that combine morphological features of kidney stones can be applied across different patient sizes, acquisition protocols and scanner models, while retaining adequate prediction accuracy. However, we considered RMSE to be of limited value as a figure of merit to test the robustness of our prediction, owing to its inherent bias towards single stones with widely different estimated vs. measured comminution time. For that reason, we introduced the quartile-based classification.

The classification data demonstrate that the performance of the binned comminution time into fragile, average and hard stones confirmed the equivalent performance for each acquisition and reconstruction variation, except for the 3 mm images and the strong IR. This pattern of performance is also suggested by the number of severely misclassified stones, which are the ones that would have the greatest clinical impact, as a fragile stone that is predicted as hard, or vice versa, could potentially result in the choice of a suboptimal tool or procedure.

This study has several limitations. First, the large parameter space prevented us from evaluating every possible combination of CT acquisition and reconstruction parameter. The acquisition and reconstruction parameters in this study were specifically chosen to span a variety of parameter combinations that are likely to be encountered in clinical practice. Another limitation exists in the *ex vivo* comminution time measurements. Measurements of this nature mimic clinical operation procedures; however, we are unable to account for differences in patient specific characteristics, including stone location and adhesion to the renal parenchyma. These differences may act as confounding variables in an *in vivo* evaluation of stone fragility and hence may limit the accuracy of a predictive model. A study using *in vivo* information and true patient stone comminution times is required for validation and is ongoing. Finally, the model developed and used in this study was generated from the same subset of stones; validation of the stone fragility model should be with independent stones in future work. This was acceptable, for this study was not designed to identify the accuracy of a CTbased predictive model of stone fragility, but rather to determine its robustness to variations in CT protocols that are to be expected in multicenter studies.

Conclusion

The results of this study show that when using a linear regression model to predict stone fragility from quantitative CT imaging biomarkers, the model is robust to changes in scan acquisition and most reconstruction parameters. Therefore, a stone fragility predictive model such as the one used in this study could be used at different institutions without the need to match imaging protocols or to add a dedicated image reconstruction series, as long as thin (1mm or less) images are produced and made available. A study to develop a predictive model of stone fragility from *in vivo* information, with comparison to clinical stone removal procedures, is ongoing.

Acknowledgments

This work was supported by the National Institutes of Health [DK100227], and the AAPM Research Seed Funding Initiative. The authors wish to thank Kristin Mara from the Department of Health Sciences Research for assistance with statistical analysis and Kristina Nunez for assistance with manuscript preparation.

Research reported in this article was supported by the AAPM Research Seed Funding Initiative and the O'Brien Urology Research Center Grant DK100227 from the National Institute of Diabetes and Digestive and Kidney Diseases.

Abbreviations:

CT	computed tomograph
DECT	dual-energy CT
ESWL	extracorporeal shock wave lithotripsy
IR	iterative reconstruction
PCNL	percutaneous nephrolithotomy
RMSE	root mean squared error
SECT	single-energy CT
VIF	variance inflation factors

References

1. Hyams ES, Matlaga BR. Economic impact of urinary stones. *Transl Androl Urol*. 2014;3(3):278. [PubMed: 26816777]
2. Scales CD, Smith AC, Hanley JM, Saigal CS. Prevalence of kidney stones in the United States. *Eur Urol*. 2012;62(1):160–5. [PubMed: 22498635]
3. Stamatelou KK, Francis ME, Jones CA, Nyberg LM, Curhan GC. Time trends in reported prevalence of kidney stones in the United States: 1976–1994. *Kidney Int*. 2003 5;63(5):1817–23. [PubMed: 12675858]
4. Fulgham PF, Assimios DG, Pearle MS, Preminger GM. Clinical effectiveness protocols for imaging in the management of ureteral calculous disease: AUA technology assessment. *J Urol*. 2013;189(4):1203–13. [PubMed: 23085059]
5. Coursey CA, Casalino DD, Remer EM, et al. ACR Appropriateness Criteria® acute onset flank pain—suspicion of stone disease. *Ultrasound quarterly*. 2012;28(3):227–33. [PubMed: 22902840]
6. Vrtiska TJ, Takahashi N, Fletcher JG, Hartman RP, Yu L, Kawashima A. Genitourinary applications of dual-energy CT. *AJR Am J Roentgenol*. 2010 6;194(6):1434–42. [PubMed: 20489081]
7. Primak AN, Fletcher JG, Vrtiska TJ, et al. Noninvasive differentiation of uric acid versus non-uric acid kidney stones using dual-energy CT. *Acad Radiol*. 2007 12;14(12):1441–7. [PubMed: 18035274]
8. Duan X, Li Z, Yu L, et al. Characterization of Urinary Stone Composition by Use of Third-Generation Dual-Source Dual-Energy CT With Increased Spectral Separation. *AJR American journal of roentgenology*. 2015;in press.
9. Qu M, Ramirez-Giraldo JC, Leng S, et al. Dual-energy dual-source CT with additional spectral filtration can improve the differentiation of non-uric acid renal stones: an ex vivo phantom study. *AJR Am J Roentgenol*. 2011 6;196(6):1279–87. [PubMed: 21606290]
10. Auge BK, Preminger GM. Surgical management of urolithiasis. *Endocrinol Metab Clin North Am*. 2002 12;31(4):1065–82.

11. Lingeman J, Jr., Newmark J, Wong M. Classification and management of staghorn calculi. *Controversies in endourology* Philadelphia: Saunders. 1995:136–44.
12. Ferrero A, Montoya JC, Vaughan LE, et al. Quantitative prediction of stone fragility from routine dual energy CT: ex vivo proof of feasibility. *Acad Radiol.* 2016;23(12):1545–52. [PubMed: 27717761]
13. Williams JC, Saw KC, Paterson RF, Hatt EK, McAteer JA, Lingeman JE. Variability of renal stone fragility in shock wave lithotripsy. *Urology.* 2003 6;61(6):1092–6; discussion 7. [PubMed: 12809867]
14. Kim SC, Burns EK, Lingeman JE, Paterson RF, McAteer JA, Williams JC, Jr., Cystine calculi: correlation of CT-visible structure, CT number, and stone morphology with fragmentation by shock wave lithotripsy. *Urol Res.* 2007 12;35(6):319–24. [PubMed: 17965956]
15. Duan X, Qu M, Wang J, et al. Differentiation of calcium oxalate monohydrate and calcium oxalate dihydrate stones using quantitative morphological information from micro-computerized and clinical computerized tomography. *J Urol.* 2013 6;189(6):2350–6. [PubMed: 23142201]
16. Zarse CA, Hameed TA, Jackson ME, et al. CT visible internal stone structure, but not Hounsfield unit value, of calcium oxalate monohydrate (COM) calculi predicts lithotripsy fragility in vitro. *Urol Res.* 2007 8;35(4):201–6. [PubMed: 17565491]
17. Williams JC, Jr., Hameed T, Jackson ME, et al. Fragility of brushite stones in shock wave lithotripsy: absence of correlation with computerized tomography visible structure. *J Urol.* 2012 9;188(3):996–1001. [PubMed: 22819106]

Figure 1.
Workflow describing the study design.

Figure 2.
Ex vivo experiment to measure time to comminution for each stone.

Figure 3.
Evaluation of robustness of stone fragility prediction.

Figure 4.

Influence of CT protocol on the visual appearance of renal stones of similar volume (range 320–405 mm³) but very different comminution time at PCNL (21, 34, and 256 s, respectively). Same window width and window level used for all three stones.

Figure 5. RMSE (s) for (a) acquisition variations from routine reconstruction and (b) reconstruction variations from routine acquisition.

Figure 6.

Number of misclassified stones for (a) acquisition variations with routine reconstruction, and (b) reconstruction variations with routine acquisition. * indicates a statistically significant ($p < 0.01$) difference compared to the routine dataset, which was used to derive the fragility model.

Figure 7.

Number of stones misclassified with respect to the measured comminution times for each scan acquisition and reconstruction combination. A * indicates a statistically significant ($p < 0.01$) difference compared to the routine protocol data used to derive the fragility model.

Figure 8.

Number of misclassified stones and proportion that were severely misclassified for each acquisition and reconstruction combination.

Table 1.

(a) Data acquisition parameters (b) Data reconstruction parameters

	Routine	Large Patient	Half Dose	Single Energy	System B
Phantom	35cm water	45cm water	35cm water	35cm water	35cm water
Scanner	Force	Force	Force	Force	Flash
kVp	90/Sn150	90/Sn150	90/Sn150	120	100/Sn140
Rotation Time (s)	0.5	0.5	0.5	0.5	0.5
Collimation (mm)	128×0.6	128×0.6	128×0.6	192×0.6	32×0.6
CTDIvol (mGy)	17.15	26.56	8.33	12.61	13.35
Pitch	0.6	0.6	0.6	0.6	0.6
QRM	400/250	350/219	200/125	240	240/185

	System A	System B
Recon Style	FBP	FBP
FOV (mm)	300	300
Image Thickness (mm)	1	1
Kernel	Qr40	D30

Table 2.

Definition of CT features used in predictive model

Variable	Formula	Definition
Porosity	Void volume/real volume	Ratio of void volume to stone volume
Energy	$\sum_{i=1}^{num\ levels} \sum_{j=1}^{num\ levels} cooc(i, j)^2$	Uniformity of the Image
Mean Curvature	Average of vertex curvatures	Surface morphology