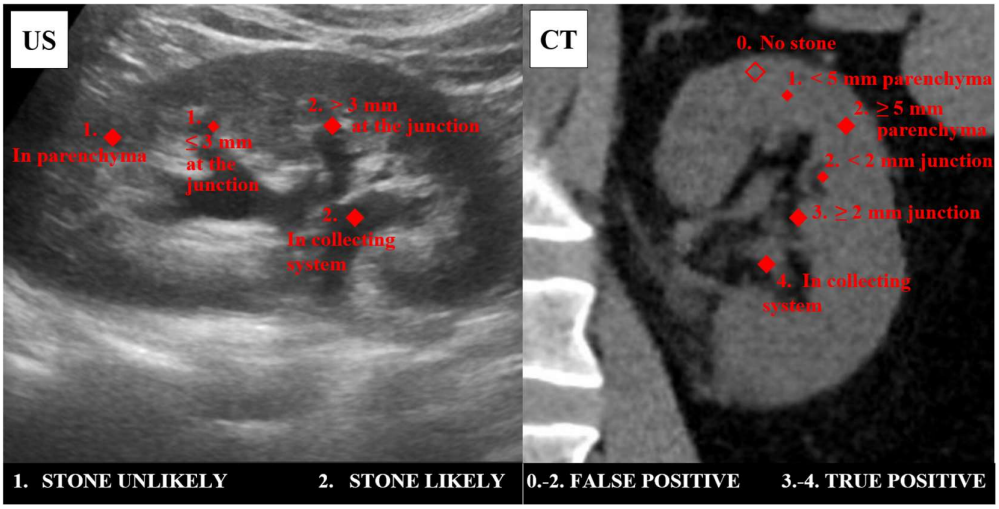
Dylan Taylor

TITLE:
Variables predicting false positive stones on renal sonograms

INTRODUCTION AND OBJECTIVES:
Kidney stone's affect one in eleven Americans, often prompting work up in the emergency department (ED) and diagnosis with non-contrast abdominal-pelvic computed-tomography (CT). Existing studies and trials comparing initial CT and ultrasonography (US) have failed to demonstrate significant difference in serious adverse events, diagnostic accuracy or return visits to the ED, which has resulted in disagreement among existing algorithms on whether CT or US is best for initial imaging. US for nephrolithiasis reduces radiation but overestimates stone size often leading to subsequent CT which increases radiation exposure and over-diagnosis. In the case of falsely positive US leading to subsequent CT a patient will have received unneeded radiation. We aimed to identify demographic and sonographic factors to help clinicians more confidently rule out "stones" noted on US reports. We performed retrospective review of patients who had CT scan of the kidneys following US and created US and CT Stone Likelihood Scores (US-SLS, CT-SLS) based on stone size and location.

METHODS:
Renal US demonstrating a single "stone" or "calculus" on sonographic report followed by CT within 24 hours during the years 2019 and 2020 were included. Demographic and patient characteristics were collected by chart review while imaging and stone related data were collected from the US or CT images and the radiographic reports. To facilitate logistic regression, each "stone" was assigned an ordinal Stone Likelihood Score based on stone size and location on both US and CT depicted and explained in *Figure 1*. Stone size differentiation at the parenchymal junction was set at 2 mm for CT because this is the minimum size able to be captured in a basket during ureteroscopy while 3 mm was utilized for US because this was the optimal sized obtained from receiver operating characteristic curve analysis with CT-SLS as the dependent variable. Statistical analysis was preformed using SPSS statistical software version 27 with true positive (TP) and False positive (FP) stones on CT scan being compared with independent sample t-tests for continuous variables and Chi-square tests for categorical variables. Univariate and multivariate regression analysis was performed to generate odds ratio for each variable

Figure 1: Sketch of the US-SLS and CT-SLS



Stone size was the maximum measurable diameter. There is no score of 0 for US because only positive radiologic reads of renal ultrasounds were incorporated in this study. Scores of 2 on US-SLS and scores of 3 or 5 on CT-SLS possess features more likely to be associated with renal stones rather than mimic condition.

RESULTS:

820 cases of CT scans following US within 24 hours, were identified. Of which 228 (27.8%) US reports documented renal stones, and 141 (17.2%) documented solitary stones with 88 TPs and 53 FPs yielding a Positive Predictive Value (PPV) of 62.41% for the radiologists' interpretation. Raw variables significantly associated with TP were higher BMI (mean (±SD), FP = 28.08 (±7.03) vs. TP = 31.76 (±8.99), p = 0.02), depth to stone (FP = 6.33 (±1.79) vs. TP = 7.23 (±1.92), p = 0.007), presence of hydronephrosis (FP = 13 (24.5%) vs. TP = 37 (42%), p = 0.04), presence of stone shadow (FP = 17 (32.1%) vs. TP = 46 (52.3%), p = 0.02), stone size (FP = 4.66 (±2.66) vs. TP = 6.94 (±5.60), p = 0.001), location in the collecting system (FP = 23 (43.4%) vs. TP = 53 (60.2%), p = 0.02). All variables are displayed in *Table 1* and *Table 2*. Univariate predictors of TP aligned with significance testing. On multivariate analysis of these significant variables, nonparenchymal location in the collecting system (Odds Ratio (OR) : 5.65, Confidence Interval (CI): 1.45 to 22.06, p = 0.01) and hydronephrosis (OR: 2.37, CI: 1.001 to 5.59, p = 0.05) were the only significant predictors of TP stones. Twinkle and shadow were combined into a new variable which was highly correlated with TP (OR: 4.36, CI: 1.56 to 12.19, p = 0.01) and carried the highest specificity and positive predictive value (PPV). US-SLS correlated with TP more than any other variable (OR: 8.54, CI: 2.93 to 24.81, p < 0.0001) and correctly classified 101 stones, with sensitivity and specificity of 94.3 and 34.0 and negative predictive value (NPV) and PPV of 78.3 and 70.3, respectively.

Table 1a. T-tests of TP and FP variables

	n	ALL	True positive	False positive	P Val
Demographic					
Age	141	47.81(±20.02)	47.72(±20.24)	47.96(±19.85)	0.94
BMI	132	30.36(±8.46)	31.76(±8.99)	28.08(±7.03)	0.02
Sonographic					
Stone size (mm)	141	6.08(±4.83)	6.94(±5.60)	4.66(±2.66)	0.001
Stone HU	141	161.40(±36.81)	159.33(±37.63)	164.82(±35.48)	0.39
Depth to Stone (cm)	141	6.89(±1.92)	7.23(±1.92)	6.33(±1.79)	0.007
Depth to Kidney (cm)	140	4.43(±1.64)	4.61(±1.66)	4.13(±1.57)	0.09
Cortical Thickness (cm)	141	1.39(±0.47)	1.44(±0.44)	1.31(±0.52)	0.12
Gain (dB)	141	49.54(±8.67)	48.58(±9.12)	51.13(±7.70)	0.09

HU = Houndsfield units

Table 1b. Chi-square tests of TP and FP variables

	n	ALL	True positive	False positive	P value
Demographic variables					
Sex	M	45(31.9%)	28(31.8%)	17(32.1%)	0.97
	F	96(68.1%)	60(68.2%)	36(67.9%)	
History of Stones	N	40(28.4%)	21(23.9%)	19(35.8%)	0.13
	Y	101(71.6%)	67(76.1%)	34(64.2%)	
Smoking History	N	85(60.3%)	58(65.9%)	27(50.9%)	0.08
	Y	56(39.7%)	30(34.1%)	26(49.1%)	
Diabetes Mellitus	N	105(74.5%)	66(75.0%)	39(73.6%)	0.85
	Y	36(25.5%)	22(25.0%)	14(26.4%)	
Hypertension	N	80(56.7%)	49(55.7%)	31(58.5%)	0.74
	Y	61(43.3%)	39(44.3%)	22(41.5%)	
History of CAD/PVD/PAD	N	128(90.8%)	81(92%)	47(88.7%)	0.50
	Y	13(9.2%)	7(8%)	6(11.3%)	
Previous UTI	N	92(65.2%)	55(62.5%)	37(69.8%)	0.38
	Y	49(34.8%)	33(37.5%)	16(30.2%)	
Presenting features					

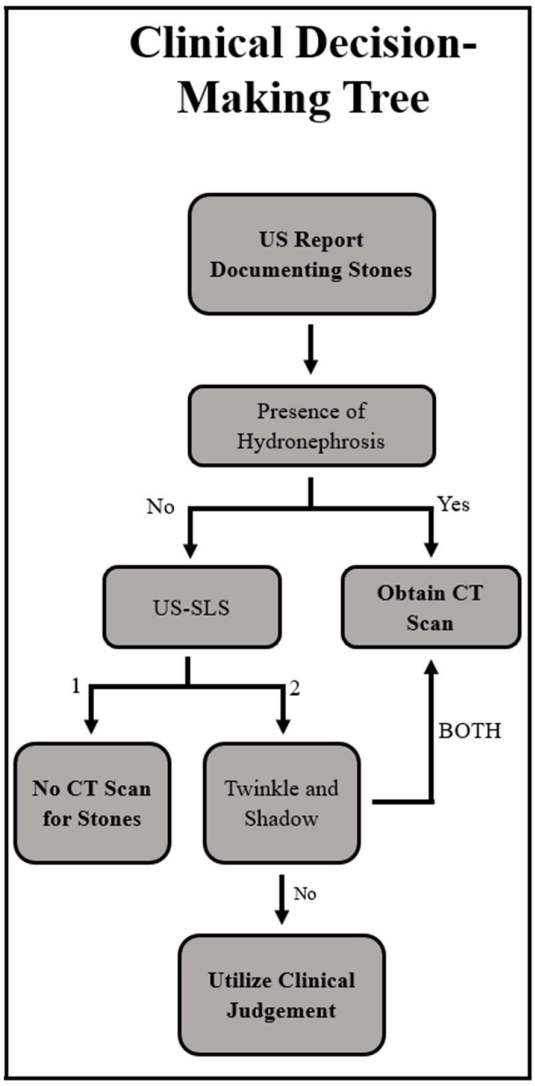
Gross Hematuria	N	112(79.4%)	69(78.4%)	43(81.1%)	0.70
	Y	29(20.6%)	19(21.6%)	10(18.9%)	
Flank/Abdominal Pain	N	33(23.4%)	21(23.9%)	12(22.6%)	0.87
	Y	108(76.6%)	67(76.1%)	41(77.4%)	
Pyelonephritis	N	98(69.5%)	59(67.0%)	39(73.6%)	0.41
	Y	43(30.5%)	29(33.0%)	14(26.4%)	
Sonographic Variables					
Frequency (Hz)	4	76(53.9%)	43(48.9%)	33(62.3%)	0.27
	4.5	40(28.4%)	28(31.8%)	12(22.6%)	
	> 4.5	11(7.8%)	6(6.8%)	5(9.4%)	
	< 4	14(9.9%)	11(12.5%)	3(5.7%)	
Twinkling	N	58(41.1%)	31(35.2%)	27(50.9%)	0.07
	Y	83(58.9%)	57(64.8%)	26(49.1%)	
Shadow	N	78(55.3%)	42(47.7%)	36(67.9)	0.02
	Y	63(44.7%)	46(52.3%)	17(32.1%)	
Twinkle and shadow	NEITHER	34(24.1%)	16(18.2%)	18(34.0%)	0.03
	TW	44(31.2%)	26(29.5%)	18(34.0%)	
	SH	24(17.0%)	15(17.0%)	9(16.9%)	
	BOTH	39(27.7%)	31(35.2%)	8(15.1%)	
Parenchymal Location	Y	16(11.3%)	5(5.7%)	11(20.8%)	0.02
	JUNCTION	49(34.8%)	30(34.1%)	19(35.8%)	
	N	76(53.9%)	53(60.2%)	23(43.4%)	
Hydronephrosis	N	91(64.5%)	51(58%)	40(75.5%)	0.04
	Y	50(35.5%)	37(42%)	13(24.5%)	
Kidney	R	68(48.2%)	42(47.7%)	26(49.1%)	0.88
	L	73(51.8%)	46(52.3%)	27(50.9%)	
Kidney Pole	UP	20(14.2%)	10(11.4%)	10(18.87%)	0.141
	MP	56(39.7%)	32(36.4%)	24(45.28%)	
	LP	65(46.1%)	46(52.3%)	19(35.85%)	
Echogenic artifact	N	116(82.3%)	72(81.8%)	44(83.0%)	0.857
	Y	25(17.7%)	16(18.2%)	9(17.0%)	
US-SLS	1	23(15.6%)	5(5.7%)	18(34.0%)	< 0.0001
	2	118(84.4%)	83(94.3%)	35(66.0%)	

CAD/PVD/PAD = Coronary artery disease, peripheral vascular disease, peripheral artery disease. TW = Twinkle only, SH = Shadow only, UP = Upper Pole, MP = Middle Pole, LP = Lower Pole.

CONCLUSION:

The US-SLS is an easy-to-use tool for ruling out clinically insignificant stones on subsequent CT scan with high sensitivity in comparison to other commonly used criteria. Given that the US-SLS has a higher sensitivity and equivalent NPV compared to a multivariate model of the significant variables in this study and is quick and straight forward to use it is likely more useful in the point of care setting. However, our US-SLS is not particularly specific (34.0%), which was minimally improved on multivariate models. In contrast the combination of twinkle and shadow was found to be a relatively specific (84.9%) finding. Among stones with a score of 2 on our US-SLS point of care providers can utilize the presence of both twinkle and shadow as well as hydronephrosis to further stratify stone likelihood in a quick and straight forward fashion depicted in *Figure 2*. Our study provides further support for US as the initial imaging modality for the work up of nephrolithiasis and provides Point of care providers with additional tools such as our US-SLS in combination with previously commonly used criteria like twinkle and shadow, and hydronephrosis to further stratify likelihood of clinically significant renal calculus on subsequent CT.

Figure 2. Decision-making tree for diagnosing renal stones



This decision-making tree demonstrates the proper interpretation of our US-SLS with additional variables to augment specificity.