Inter-observer Agreement between Urologists and Radiologists in Interpreting the Computed Tomography Images of Emergency Patients with Renal Colic

Jun Young Hong[1], Dong Hoon Lee[2]*, In Ho Chang[3], Sung Bin Park[4], Chan Woong Kim[5], Byung Hoon Chi[6]

Purpose: Low-dose non-enhanced computed tomography (LDCT) has been shown to provide low radiation exposure with proper diagnostic accuracy compared to standard dose non-enhanced computed tomography (SDCT) in patients with renal colic. The goal of our study is to estimate the accuracy of LDCT and SDCT interpretation by emergency medicine residents who primarily treated patients with renal colic.

Materials and Methods: Thirty sample images of both LDCT and SDCT from renal colic patients were extracted from January 2013 to December 2015 in a tertiary teaching hospital. Five emergency medicine residents interpreted 60 image samples over a time span of 3 weeks. The presence of a ureteric stone, the stone's size and location, and signs of obstruction were recorded in the reports. A total of 300 reports were compared with formal readings by a radiologist. The inter-observer agreement and kappa value were calculated for comparative analysis.

Results: Identification of ureteric stones showed almost perfect inter-observer agreement on SDCT (kappa value: 0.93), and the percentage of agreement was 96.7%. However, on LDCT, the inter-observer agreement was substantial (kappa value: 0.73), and the percentage of agreement was 88.0%.

Conclusion: Using SDCT, emergency medicine residents had almost perfect inter-observer agreement in interpreting the CT images of patients with renal colic compared to a radiologist. However, when using LDCT, they had a lower inter-observer agreement.

Keywords: emergency department; non-enhanced computed tomography; radiation dose; renal colic; urolithiasis.

INTRODUCTION

Approximately 12 percent of males and 6 percent of females will experience urolithiasis during their lifetime, and up to 50 percent of these individuals will experience a recurrence of urolithiasis within 10 years [1-3]. Renal colic is a common symptom seen in the emergency department (ED). In the United States, more than a million patients are treated for urolithiasis in an emergency department over the span of a year[4].

In the past, intravenous urography (IVU) was the imaging method of choice for diagnosing urolithiasis. However, unenhanced helical computed tomography (CT) has become the standard for diagnosing acute flank pain, and has replaced IVU as the best initial diagnostic imaging modality in patients with renal colic [5]. Furthermore, CT examination is often repeated to assess the progress of the condition. In 2007, Broder et al. reported that approximately half of the patients who had been diagnosed with urolithiasis in the ED received two more CT scans over the course of their condition and that approximately 30 percent of these patients underwent more than three scans[6]. The risk of cancer is increased at a rate greater than 1/1000 per abdominal CT scan, and the risk is higher in young patients [7,8]. Therefore, a means to reduce radiation exposure is needed, and low-dose CT (LDCT) was studied as a diagnostic modality.

The correct interpretation of urolithiasis by an emergency physician via CT images could be advantageous for the early diagnosis and treatment of renal colic patients. SDCT interpreted by emergency physicians has an appropriate percentage of inter-observer agreement compared with formal reporting by a radiologist[9]. However, there has not been a study that evaluated the accuracy of the LDCT interpretation of urolithiasis by emergency physicians. In this study, we compared the accuracy of LDCT interpretation by emergency medicine residents with radiologists.

METHODS

This study was approved by the institutional review board of the Chung-Ang University Hospital (IRB No. C2016023). Written informed consent was obtained from each participant. We have residency program in major of emergency medicine for 4 years. Five emergency medicine (EM) residents (two junior and three

[1]Department of Emergency Medicine, College of Medicine, Chung-Ang University, Seoul, Republic of Korea.
[2]Department of Emergency Medicine, College of Medicine, Chung-Ang University, Seoul, Republic of Korea
[3]Department of Urology, College of Medicine, Chung-Ang University, Seoul, Republic of Korea.
[4]Department of Radiology, College of Medicine, Chung-Ang University, Seoul, Republic of Korea.
[5]Department of Emergency Medicine, College of Medicine, Chung-Ang University, Seoul, Republic of Korea.
[6]Department of Urology, College of Medicine, Chung-Ang University, Seoul, Republic of Korea.
*Correspondence: Department of Emergency Medicine, College of Medicine, Chung-Ang University, Seoul, Republic of Korea.
Tel: 82-2-6299-3109. E-mail: emdhlee@cau.ac.kr.
Received March 2017 & Accepted November 2017

Table 1. Diagnostic performance of identifying urolithiasis

	agreement (95% CI)	Kappa	Sensitivity(%)	Specificity(%)	PPV† (%)	NPV‡ (%)
total CT	92.3(89.3-95.4)	0.85	90.9(86.5-95.3)	94.1(90.0-98.1)	94.9(91.5-98.4)	89.4(84.3-94.6)
SDCT	96.7(93.8-99.6)	0.93	96.7(92.0-100)	96.7(92.9-100)	95.1(89.5-100)	97.8(94.6-100)
LDCT	88.0(82.7-93.3)	0.73	87.6(81.2-94.0)	88.9(79.3-98.4)	94.8(90.4-99.3)	75.5(63.5-87.4)

†PPV, positive predictive value; ‡NPV, negative predictive value

senior residents) of Chung-Ang University Hospital were included to compare the accuracy of interpretation of LDCT.

Study design

This Study retrospectively reviewed images of renal colic patient performed in emergency department. Five emergency medicine residents interpreted 60 patient CT scans over a time span of 3 weeks and reported total 300 cases. A simple reporting method was provided to the EM residents. Each interpretation was recorded on the reporting form, which included brief clinical information. The case report form included the presence of ureteric stones, their size and location, and signs of obstruction. Other clinical findings that were unrelated to ureteric stones were recorded to create a descriptive clinical picture. The participants' reports were compared for inter-observer agreement with reports by a professional radiologist.

Sampling Images

974 patient image samples were composed of unenhanced abdominal pelvic CT conducted in the emergency department from January 2013 to December 2015. During this period, another study was conducted to compare the diagnostic efficacy of LDCT with SDCT (Title: Diagnostic Trial of Low-Dose CT for the Detection of Urolithiasis IRB No. C2013234(1194)). All 30 LDCT and SDCT image samples were randomly extracted from 974 patient image samples and those were anonymized and randomized. Total 60 patient CT images were used for the interpretation.

CT protocol

All of the unenhanced CT studies were performed using a 256-MDCT scanner (Brilliance iCT, Philips Healthcare, Cleveland, OH, USA). All patients underwent a scan using the standard- or low-dose protocol from the proximal aspect of the T12 vertebra to the distal aspect of the symphysis pubis in the supine position. The standard-dose protocol and low-dose protocol was achieved at a manually set peak tube voltage of 120 kVp and 100kVp, with automated Z-axis dose modulation by the scout image (DoseRight, Philips Healthcare, Cleveland, OH, USA),and the tube current was limited to 150 mAs and 100mAs, respectively. The remaining scanning parameters were as follows: detector configuration, 128x0.625; pitch, 0.915; beam collimation, 80 mm; rotation time, 0.4 sec; and helical acquisition. Image noise was reduced by iterative reconstruction in the acquired scan images and could reduce the radiation dose from 5.77 mSV to 1.34 mSV.

Sample size and statistical analysis

To compare the accuracy of diagnostic performance utilizing LDCT by EM residents with a radiologist, the inter-observer agreement was used. The kappa coefficient was calculated using the R statistical computing program (R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/). We considered a kappa value of ≤ 0.19 as poor, a kappa value of 0.20-0.39 as fair, a kappa value of 0.40-0.59 as moderate, a kappa value of 0.60-0.79 as substantial, and a kappa value of ≥ 0.80 as almost perfect[10]. If the expected lower boundary for a kappa one-sided 95% confidence interval (CI) was 0.5 and the expected preliminary kappa value and prevalence were 0.73 and 0.5, respectively, based on a previous study, a minimum of 146 subjects were required for this study of inter-observer agreement by 2 raters. We estimated sample size using the kappaSize library statistical program in R-project (R Core Team [2012]. R: A language and environment for statistical computing; R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/)

RESULTS

This study included 44 men and 16 women. The mean age was 47.5 years (inter-quartile range: 34.25 to 59.75). Overall, 55% (n = 33) of the CT images were positive for urolithiasis, and 45% (n = 27) were negative for urolithiasis. All five EM residents who participated in this study had experience with more than 1000 scans for SDCT and fewer than 100 scans for LDCT. When identifying ureteric stones on SDCT, the percentage of agreement between residents and radiologists was 96.7%, and the inter-observer agreement was near perfect (kappa value; 0.93). However, ureteric stones were identified at a percentage of agreement of 88.0%, and the inter-observer agreement was substantial (kappa value; 0.73) on LDCT scans. The LDCT interpretation by an EM resident had a 75% negative predictive value compared with the interpretation conducted by a radiologist. This was significantly low compared with the 98% of agreement on SDCT scans (**Table 1**).

The results of the interpretation of size and location of ureteric stones were perfect in terms of the inter-observer agreement (kappa value; 0.85, 0.95) on SDCT and

Table 2. Diagnostic performance of sign of obstruction, stone size and location

	Sign of urinary obstruction		Stone size(5 mm)		Stone location	
	Agreement (95% CI)	Kappa	Agreement (95% CI)	Kappa	Agreement (95% CI)	Kappa
total CT	76.3 (71.5 - 81.2)	0.52	85.0 (80.9 - 89.1)	0.76	89.3 (85.8 - 92.8)	0.84
SDCT	86.0 (80.4 - 91.6)	0.71	91.3 (86.8 - 95.9)	0.85	91.3 (86.8 - 95.9)	0.93
LDCT	66.7 (59.0 - 74.3)	0.34	78.7 (72.0 - 85.3)	0.66	78.7 (72.0 - 85.3)	0.76

were substantial for inter-observer agreement (kappa value; 0.66, 0.76) on LDCT (**Table 2**). Sign of obstruction results had a kappa value of 0.71 on SDCT and 0.34 on LDCT (**Table 2**).

DISCUSSION

Rafi et al. compared the accuracy of interpretation of conventional CT scans by emergency physicians for patients with renal colic, and the results had a sensitivity of 92%, a specificity of 99%, and a kappa value of 0.89[9]. These results indicate that emergency physicians could interpret the images of SDCT almost perfectly for patients with renal colic in the ED. Therefore, emergency physicians used non-enhanced CT to evaluate patients in many EDs who presented with renal colic. Recently low-dose, non-enhanced helical CT was studied in patients with renal colic to reduce the radiation threat of SDCT. Therefore, there have been several reports that LDCT had high sensitivity and specificity for the diagnosis of urolithiasis when interpreted by radiologists and urologists[11,12].

In our study, the kappa value was 0.93 (**Table 1**), which was similar to that found in Rafi's previous study. Kwon et al. reported that a recent survey, LDCT in patients with renal colic demonstrated similar sensitivity and specificity compared with the conventional Standard-dose CT (SDCT)[13-15]. However, there have not been studies on the accuracy of LDCT interpretation performed by emergency physicians. In this study, we compared the agreement of interpretation on LDCT the kappa value was 0.73, which is a lower value than that of SDCT. Thus, when LDCT was used in the ED and the result was read by an emergency medicine resident, some patients could have been misdiagnosed, although the final confirmation of interpretation was made by a radiologist.

Yang et al. reported that the diagnostic performance of low-dose appendiceal CT was influenced by the amount of a physician's experience with both low- and standard-dose CT interpretation[12]. Urologists with an appropriate amount of experience seem to frequently be in agreement with radiologists on LDCT scans. However, our participants (emergency physician residents) had worked with more than 1000 scans of SDCT for a year; therefore, they were familiar with images of SDCT. According to this study, emergency medicine residents could find urolithiasis in the images of SDCT as well as a radiologist could and could interpret the exact location and size. Therefore, there was minimal difficulty in making a clinical decision with SDCT. In contrast, the images from LDCT were coarser than those of SDCT because of the low radiation amount. Emergency medicine residents had no experience with interpreting images from LDCT prior to this study. Each resident had worked with fewer than 100 scans of LDCT, and they had not trained in the interpretation of LDCT during the study period. Therefore, they were not familiar with the coarse and low-quality LDCT images. In this study, emergency medicine residents simply read the images of LDCT based on previous knowledge and competence with SDCT. To improve the accuracy of interpreting LDCT images, emergency medicine residents may be required to have sufficient experience and training.

In this study, sign of obstruction, size, and location of ureteric stones had substantial to almost excellent inter-observer agreement (kappa value; 0.71, 0.85, 0.93) compared with formal readings on SDCT (Table 2). In contrast, a fair to substantial inter-observer agreement (kappa value: 0.34, 0.66, 0.76) was observed on LDCT scans. The signs of obstruction and the size and location of ureteric stones are important for determining the prognosis and first-line treatment for renal colic patients (16). Therefore, emergency medicine residents should be trained in the interpretation of LDCT.

When emergency medicine residents could find stones in LDCT or SDCT, they had little difficulty with interpreting the characteristics of urolithiasis. When they had not been trained to interpret the low-quality images of LDCT, it was difficult to deduce the presence of a stone. Therefore, if they have more experience with LDCT images and receive training on the interpretation of these images, LDCT might be as useful in the assessment of urolithiasis as SDCT.

LDCT has been reported to have adequate diagnostic performance while reducing the risk of cancer from radiation as compared with SDCT in a variety of diseases[17-19]. Accordingly, LDCT was used for the examination of several diseases in some EDs. If emergency physicians can properly interpret LDCT without waiting for a formal reading, they can potentially determine the appropriate treatment course and prognosis in the ED more expediently. As our study showed, the interpretation of LDCT by emergency medicine residents had low inter-observer agreement compared with formal reading. For proper interpretation with LDCT scans in renal colic patients, additional experience and education may be required.

Limitations

The participants who enrolled in this study were in one tertiary medical center. Therefore, the sample could not represent the accuracy of interpretation of LDCT by an emergency physician. However, our participants had a similar accuracy of interpretation on SDCT compared with a previous study that included emergency physicians. We used a lower greyscale monitor compared to radiologists, who use a higher greyscale monitor for formal reading. This could have affected the diagnostic accuracy of our participants due to the lower imaging quality. However, emergency physicians do not use a high-resolution monitor for readings in the ED setting. Further, in a real ED setting, emergency physicians take detailed histories and conduct physical examinations of patients before reading the CT results. Our study included only brief patient information prior to reading.

CONCLUSIONS

When SDCT was performed in the ED for patients with renal colic, emergency medicine residents had a high level of agreement of interpretation compared with radiologists. However, on low-dose unenhanced CT, emergency medicine residents had relatively lower levels of agreement of interpretation with the use of SDCT compared with a radiologist.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Bartoletti R, Cai T, Mondaini N, et al. Epidemiology and risk factors in urolithiasis.

Urol Int. 2007;79 Suppl 1:3-7.

2. Curhan GC. Epidemiology of stone disease. Urol Clin North Am. 2007;34:287-93.
3. Sierakowski R, Finlayson B, Landes RR, Finlayson CD, Sierakowski N. The frequency of urolithiasis in hospital discharge diagnoses in the United States. Invest Urol. 1978;15:438-41.
4. Brown J. Diagnostic and treatment patterns for renal colic in US emergency departments. Int Urol Nephrol. 2006;38:87-92.
5. Türk C, Petřík A, Sarica K, et al. EAU guidelines on diagnosis and conservative management of urolithiasis. European urology. 2016;69:468-74.
6. Broder J, Bowen J, Lohr J, Babcock A, Yoon J. Cumulative CT exposures in emergency department patients evaluated for suspected renal colic. The Journal of emergency medicine. 2007;33:161-8.
7. [No authorlisted]. Radiation and your patient: a guide for medical practitioners. Ann ICRP. 2001;31:5-31.
8. Brenner D, Elliston C, Hall E, Berdon W. Estimated risks of radiation-induced fatal cancer from pediatric CT. AJR Am J Roentgenol. 2001;176:289-96.
9. Rafi M, Shetty A, Gunja N. Accuracy of computed tomography of the kidneys, ureters and bladder interpretation by emergency physicians. Emergency Medicine Australasia. 2013;25:422-6.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. biometrics. 1977159-74.
11. Kwon JK, Chang IH, Moon YT, Lee JB, Park HJ, Park SB. Usefulness of Low-dose Nonenhanced Computed Tomography With Iterative Reconstruction for Evaluation of Urolithiasis: Diagnostic Performance and Agreement between the Urologist and the Radiologist. Urology. 2015;85:531-8.
12. Yang HK, Ko Y, Lee MH, et al. Initial Performance of Radiologists and Radiology Residents in Interpreting Low-Dose (2-mSv) Appendiceal CT. AJR Am J Roentgenol. 2015;205:W594-611.
13. Poletti P-A, Platon A, Rutschmann OT, Schmidlin FR, Iselin CE, Becker CD. Low-dose versus standard-dose CT protocol in patients with clinically suspected renal colic. American Journal of Roentgenology. 2007;188:927-33.
14. Niemann T, Kollmann T, Bongartz G. Diagnostic performance of low-dose CT for the detection of urolithiasis: a meta-analysis. AJR Am J Roentgenol. 2008;191:396-401.
15. Kulkarni NM, Uppot RN, Eisner BH, Sahani DV. Radiation dose reduction at multidetector CT with adaptive statistical iterative reconstruction for evaluation of urolithiasis: how low can we go? Radiology. 2012;265:158-66.
16. Glenn M. Preminger M, Co-Chair; Hans-Goran Tiselius, MD, PhD, Co-Chair; Dean G. Assimos, MD, Vice-Chair; Peter Alken, MD, PhD; A. Colin Buck, MD, PhD; Michele Gallucci, MD, PhD; Thoma Knoll, MD, PhD; James E. Lingeman, MD; Stephen Y. Nakada, MD; Margaret Sue Pearle, MD, PhD; Kemal Sarica, MD, PhD; Christian Turk, MD, PhD; J. Stuart Wolf, Jr., MD. 2007 Guideline for the Management of Ureteral Calculi. american urological association. 2007.
17. Berrington de González A, Mahesh M, Kim K-P, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Archives Of Internal Medicine. 2009;169:2071-7.
18. Keyzer C, Tack D, de Maertelaer V, Bohy P, Gevenois PA, Van Gansbeke D. Acute appendicitis: comparison of low-dose and standard-dose unenhanced multi-detector row CT. Radiology. 2004;232:164-72.
19. Team NLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. The New England journal of medicine. 2011;365:395.