Original Article | Cardiovascular Imaging

eISSN 2005-8330
https://doi.org/10.3348/kjr.2022.0826
Korean J Radiol 2023;24(4):284-293

Korean Journal of Radiology
KJR

A Modified Length-Based Grading Method for Assessing Coronary Artery Calcium Severity on Non-Electrocardiogram-Gated Chest Computed Tomography: A Multiple-Observer Study

Suh Young Kim[1,2], Young Joo Suh[3], Na Young Kim[3], Suji Lee[3], Kyungsun Nam[3], Jeongyun Kim[3], Hwan Kim[3], Hyunji Lee[3], Kyunghwa Han[3], Hwan Seok Yong[4]

[1]Department of Radiology, Gangneung Asan Hospital, University of Ulsan College of Medicine, Gangneung, Korea
[2]Department of Medicine, Yonsei University Graduate School, College of Medicine, Seoul, Korea
[3]Department of Radiology, Research Institute of Radiological Science, Center for Clinical Imaging Data Science, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea
[4]Department of Radiology, Korea University Guro Hospital, Korea University College of Medicine, Seoul, Korea

Objective: To validate a simplified ordinal scoring method, referred to as modified length-based grading, for assessing coronary artery calcium (CAC) severity on non-electrocardiogram (ECG)-gated chest computed tomography (CT).
Materials and Methods: This retrospective study enrolled 120 patients (mean age ± standard deviation [SD], 63.1 ± 14.5 years; male, 64) who underwent both non-ECG-gated chest CT and ECG-gated cardiac CT between January 2011 and December 2021. Six radiologists independently assessed CAC severity on chest CT using two scoring methods (visual assessment and modified length-based grading) and categorized the results as none, mild, moderate, or severe. The CAC category on cardiac CT assessed using the Agatston score was used as the reference standard. Agreement among the six observers for CAC category classification was assessed using Fleiss kappa statistics. Agreement between CAC categories on chest CT obtained using either method and the Agatston score categories on cardiac CT was assessed using Cohen's kappa. The time taken to evaluate CAC grading was compared between the observers and two grading methods.
Results: For differentiation of the four CAC categories, interobserver agreement was moderate for visual assessment (Fleiss kappa, 0.553 [95% confidence interval {CI}: 0.496–0.610]) and good for modified length-based grading (Fleiss kappa, 0.695 [95% CI: 0.636–0.754]). The modified length-based grading demonstrated better agreement with the reference standard categorization with cardiac CT than visual assessment (Cohen's kappa, 0.565 [95% CI: 0.511–0.619 for visual assessment vs. 0.695 [95% CI: 0.638–0.752] for modified length-based grading). The overall time for evaluating CAC grading was slightly shorter in visual assessment (mean ± SD, 41.8 ± 38.9 s) than in modified length-based grading (43.5 ± 33.2 s) ($P < 0.001$).
Conclusion: The modified length-based grading worked well for evaluating CAC on non-ECG-gated chest CT with better interobserver agreement and agreement with cardiac CT than visual assessment.
Keywords: Calcium; Coronary vessels; Thorax; Tomography, X ray computed

INTRODUCTION

Coronary artery calcium (CAC) scores obtained from electrocardiogram (ECG)-gated cardiac computed tomography (CT) scans have been established as strong predictors of future cardiovascular events in the general

Received: September 1, 2022 **Revised:** December 19, 2022 **Accepted:** February 4, 2023
Corresponding author: Young Joo Suh, MD, PhD, Department of Radiology, Research Institute of Radiological Science, Center for Clinical Imaging Data Science, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.
• E-mail: rongzusuh@gmail.com

population [1-3]. CAC quantification and severity classification are generally performed using ECG-gated cardiac CT scans. However, the importance of evaluating CAC on non-ECG-gated chest CT scans has been increasing, especially after the initiation of CT lung cancer screening in high-risk populations [4]. The 2016 guidelines of the Society of Cardiovascular Computed Tomography (SCCT)/Society of Thoracic Radiology (STR) recommend that CAC should be reported on all non-contrast chest CT scans of patients aged ≥ 40 years with an estimation of severity as none, mild, moderate, or severe [5].

Previously, it has been verified that CAC could be assessed with high reliability on chest CT with a high correlation between the CAC scores obtained from chest CT and those obtained from cardiac CT scans [6-10]. Consequently, several methods for grading CAC severity on chest CT have been proposed, such as Agatston scoring, ordinal scoring, and visual estimation [11-13]. Due to the lack of an optimal method for grading CAC on chest CT, the SCCT/STR guidelines left the final decision of the scoring method to the individual centers [5]. Although visual estimation is a simple and quick method, it has the disadvantages of high subjectivity and consequent high interobserver variability owing to the lack of clear grading criteria. Ordinal scoring methods can reduce interobserver variability but are somewhat complex to apply in clinical practice [14]. If a simplified method with clear and objective criteria for CAC severity classification is developed, it can contribute to CAC evaluation and cardiovascular risk stratification on chest CT. Therefore, we developed a simple ordinal scoring method modified from the length-based grading method suggested in a previous study [15]. This study aimed to validate the modified length-based grading method on chest CT by comparing it with a visual estimation of interobserver agreement, agreement with the CAC category on ECG-gated cardiac CT, and evaluation time.

MATERIALS AND METHODS

Patients

The institutional review board of Severance Hospital approved this retrospective study and waived the requirement for informed consent (IRB No. 4-2022-1141). Among patients who underwent cardiac CT between January 2011 and December 2021 and non-contrast-enhanced chest CT within 90 days, 120 patients (mean age ± standard deviation [SD], 63.1 ± 14.5 years; range, 19–91 years; male, 64) were randomly selected for this study. To determine the minimum sample size for interobserver agreement studies with multiple raters, we used a confidence interval (CI) construction for κ-statistics, as recommended by Rotondi et al. [16]. We calculated the required sample size to be 101 cases for each observer for an expected kappa statistic of 0.7 with a width of 0.2, in which six observers were noted for agreement between CAC severity categories between chest and cardiac CT for classification of a multinomial scale involving four categories (no, mild, moderate, and severe) with expected frequencies of 40%, 20%, 20%, and 20%, respectively. The prevalence of each category used in this study was determined empirically based on a previous study [17], and CAC categories were determined based on the CAC score on cardiac CT: none (score = 0), mild (score 1–100), moderate (score 101–400), and severe (score > 400).

CT Image Acquisition

Chest CT images were acquired using one of the nine different systems (Sensation 64, Somatom Definition Flash, Somatom Force, Somatom Definition AS+; Siemens Healthineers; Discovery CT750 HD, Revolution EVO, Revolution CT, LightSpeed VCT: GE Healthcare, and iCT 256, Philips Healthcare). Images were obtained by modulating the tube voltage (usually at 120 kVp) and current (standard mAs and low mAs), leading to two different categories of radiation doses: standard and low dose. In addition, the slice thickness (thin sections [1 or 1.25 mm], medium sections [2 or 3 mm], and thick sections [5 mm]) and image reconstruction algorithms (filtered back projection and iterative reconstruction) varied. Detailed imaging parameters and their modulations are presented in Supplementary Table 1.

Non-contrast cardiac CT for calcium scoring was acquired using one of eight different systems (Somatom Definition Flash, Somatom Force, Somatom Definition AS+; Siemens Healthineers; Revolution EVO, Revolution CT, LightSpeed VCT; GE Healthcare; iCT 256: Philips Healthcare; and Aquilion ONE: Toshiba). Detailed imaging parameters and their modulations are presented in Supplementary Table 2.

CAC Scoring and Severity Categorization on Cardiac CT

On cardiac CT scans, CAC scores were measured using the Agatston score, which defines a calcific lesion on CT with a threshold of 130 Hounsfield units and an area ≥ 1 mm^2 [11]. We classified CAC severity into four categories based on the Agatston score: none (score = 0), mild (score 1–100), moderate (score 101–400), and severe (score > 400). Notably, we used the category assessed on cardiac CT as the reference standard for CAC severity.

Reader Study for CAC Severity Grading on Chest CT

Three board-certified cardiothoracic radiologists with 5–10 years of experience in cardiac and chest CT examinations (observers 1–3) and three radiology residents with 3–4 years of experience (observers 4–6) independently evaluated CAC severity on chest CT scans. In this study, patient information was anonymized, and CT images of the patients were reviewed using a dedicated server (AVIEW Research™, Coreline Soft Inc.), which could only be accessed by the six observers in this study to review and grade the CT images (Fig. 1). The observers were blinded to the CAC severity category on cardiac CT and other observer interpretations. Radiologists evaluated CAC severity using two different scoring methods (method 1, visual assessment; method 2, modified length-based grading) in a sequential order from scoring methods 1 to 2 over two rounds per reader and using a separate round for each method. A minimum of 2 weeks was scheduled as a washout period between each scoring round to reduce recall bias. Representative cases from each category were provided to the readers before scoring round 1. Definitions of each scoring method are provided in Table 1. For scoring method 1, observers performed an overall visual assessment of none, mild, moderate, or severe CAC for a patient's entire coronary artery [13]. For scoring method 2, observers performed a modified length-based grading, developed based on previous publications [15]. The coronary artery was divided into the left main, left anterior descending,

Table 1. Definitions of CAC Grading Methods on Non-Electrocardiogram-Gated Chest CT

CAC Grading Methods	CT Findings
Visual assessment [13]	
None, mild, moderate, or severe according to published method [13]	
Conventional length-based method [15]	
Score for each main coronary artery	
Score = 0	If CAC is not detectable
Score = 1	If CACs are tiny and had a total length < 3 mm
Score = 2	If CACs are small and had a total length of 3–5 mm
Score = 3	If CACs are moderate and had a total length of 6–11 mm
Score = 5	If CACs are large and had a summed length of 12–25 mm
Score = 9	If CACs are extremely large and had a summed length > 25 mm
CAC severity category	
None	Total sum of the scores of all four main coronary artery = 0
Mild	Total sum of the scores of all four main coronary artery = 1–3
Moderate	Total sum of the scores of all four main coronary artery = 4–8
Severe	Total sum of the scores of all four main coronary artery ≥ 9
Modified length-based method*	
CAC severity category	
None	If CAC was not detectable
Mild	If CAC had a total length < 12 mm in the most severe artery
Moderate	If CAC had a total length of 12–25 mm in the most severe artery
Severe	If CAC had a total length > 25 mm in the most severe artery

*If the number of vessels with CAC is three or four, the category was upgraded one by one from that determined. CAC = coronary artery calcium, CT = computed tomography

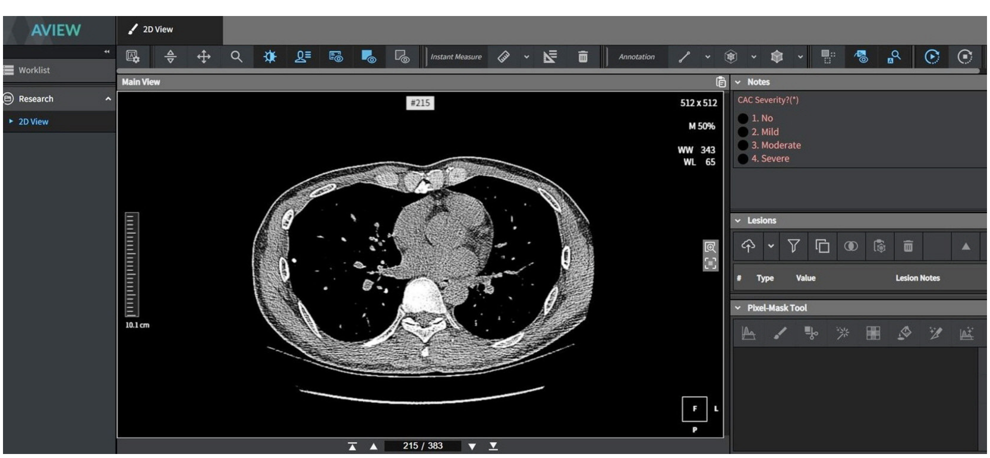

Fig. 1. Platform for coronary artery calcium grading. The observers independently reviewed and graded the computed tomography images using a dedicated server (AVIEW Research™, Coreline Soft Inc.).

left circumflex, and right coronary arteries, and CAC severity grading was evaluated based on calcium deposition in the most severe artery. CAC severity was classified as none if CAC was not detectable, mild if CAC had a total length < 12 mm, moderate if CAC had a total length of 12–25 mm, and severe if CAC had a total length > 25 mm. If the number of vessels with calcium was three or four, the category was upgraded one by one from that determined above. A detailed description of the development and validation of the modified length-based grading method is provided in Supplementary Materials and Methods, Supplementary Table 3, and Supplementary Figure 1. Additionally, CAC grading was performed using the length-based grading method by two observers (observers 2 and 5) for comparison between the previous length-based grading and our modified length-based grading method [15].

Study Endpoint

The primary endpoint was the agreement of the CAC severity categories between each method assessed by chest CT and the standard reference categories on cardiac CT. The secondary endpoints were the agreement between the six observers for each method and the evaluation time of the two different CAC grading methods.

Subgroup Analysis

Subgroups were determined according to the scan parameters (radiation dose, slice thickness, tube voltage, and reconstruction kernel) of chest CT. The "low-dose" and "standard dose" subgroups were created based on the radiation dose. Three subgroups based on slice thickness were designated as follows. One group used a thin slice thickness different from that used in cardiac CT (1 or 1.25 mm). Another group used a slice thickness similar to that of cardiac CT (2.5 or 3 mm). Finally, the other study used a thick slice thickness (5 mm). In addition, subgroups were created according to tube voltage (same or different kVp between cardiac and chest CT). Furthermore, the "soft" or "sharp" subgroups were based on the kernel.

Statistical Analysis

Categorical variables are expressed as frequencies and percentages. Continuous variables were expressed as either mean ± SD or median with 25th to 75th percentile values. Agreement among the six observers for CAC category classification (discrimination of the four CAC categories) was assessed using Fleiss' kappa statistics. The agreement between chest and cardiac CT for each observer was analyzed using Cohen's and weighted kappa values. Kappa values were interpreted as follows: < 0.4, poor; 0.41–0.6, moderate; 0.61–0.8, good; and 0.81–1.0, excellent agreement. The evaluation time for CAC grading was compared between the grading methods using a general linear model with a covariance pattern. Statistical analyses were performed using commercially available software (R package version 4.0.0, SAS Institute Inc.), and *P*-values < 0.05 were considered statistically significant.

RESULTS

Patient Characteristics

Cardiac and chest CT examinations were performed on the same day in 59 (49.2%) cases, resulting in a median time gap between cardiac and chest CT of 0 days (25–75th percentile, 0–8.25 days). Cardiac and chest CT scans were acquired using the same CT scanner in 54 (45%) cases and from different CT scanners in 66 (55%) cases. Forty-five (37.5%) patients underwent cardiac CT for health checkups and 75 (62.5%) for diagnostic purposes. Forty-three (35.8%) patients underwent chest CT for lung cancer screening and 77 (64.2%) underwent diagnostic and clinical management.

The CAC categories on cardiac CT were classified as none, mild, moderate, and severe in 46 (38.3 %), 25 (20.8%), 22 (18.3%), and 27 (22.5%) cases, respectively. The average Agatston scores of the four CAC severity categories on cardiac CT were 0 for none, a median of 42 (25–75th percentile 17.2–55.1) for the mild category, a median of 240.6 (25–75th percentile 185.8–262.6) for the moderate category, and a median of 633.6 (25–75th percentile 527.2–1404.1) for the severe category. The distribution of CAC severity, categorized by the six observers for each patient, is shown in Supplementary Table 4.

Agreement of CAC Grading among Different Observers

The interobserver agreement for discriminating between the four CAC categories was moderate for method 1 (Fleiss kappa, 0.553 [95% CI: 0.496–0.610]) and good for method 2 (Fleiss kappa, 0.695 [95% CI: 0.636–0.754]) (Table 2).

Table 2. Agreement of CAC Grading Among Six Observers

	Fleiss Kappa	95% CI
Visual assessment	0.553	0.496–0.610
Modified length-based method	0.695	0.636–0.754

CAC = coronary artery calcium, CI = confidence interval

https://doi.org/10.3348/kjr.2022.0826

Agreement of CAC Grading between Chest and Cardiac CT

Method 2 demonstrated better agreement with the standard reference cardiac CT for all six observers than method 1 (Cohen's kappa of 0.565 [95% CI: 0.511–0.619] for method 1 and 0.695 [95% CI: 0.638–0.752] for method 2) (Table 3). The agreement matrices analyzed by the six observers are presented in Supplementary Table 5. Discrepancies in grading between the categories based on the Agatston scores on cardiac CT and by the observers on chest CT were observed to be 29.7% (a total of 214 cases in six observers [35.7 cases per observer]) with method 1 and 21.9% with method 2 (a total of 158 cases in six observers [26.3 cases per observer]). Of 214 incorrect cases with method 1, 80 (37.4%) were assessed as having a higher category, and 134 (62.6%) were classified as having a lower category compared with cardiac CT (Fig. 2). In method 2, the observers overestimated the categories in 87 cases (55.1%) and underestimated them in 71 cases (44.9%) (Supplementary Table 5).

In the subgroup analysis according to experience in cardiothoracic imaging, the agreement of CAC grading between chest and cardiac CT was similar between the cardiothoracic radiologist and radiology resident groups in both methods 1 and 2 (Table 3). Notably, method 2 showed a higher agreement with cardiac CT than method 1 in both subgroups (Table 3). Table 4 shows the kappa values of the two methods according to the radiation dose, slice thickness, tube potential, and reconstruction kernel subgroups. The modified length-based method consistently demonstrated better agreement results than conventional visual assessment in all subgroup analyses of radiation dose, slice thickness, and tube potential of chest CT. In the analysis of the reconstruction kernel subgroups, the sharp reconstruction kernel subgroup had a higher agreement (Cohen's kappa 0.578–0.743) than those in the soft reconstruction kernel subgroup (Cohen's kappa 0.557–0.670).

Evaluation Time of Two Different CAC Grading Methods

The overall evaluation time for CAC grading of all observers was shorter in method 1 (mean 41.8 ± 38.9 s) than in method 2 (mean 43.5 ± 33.2 s) ($P < 0.001$) (Fig. 3). However, the difference between the two methods was not sufficiently large to achieve clinical significance, and the evaluation time was even shorter in method 2 for one observer (observer 3). If the data from observer 1 was considered an outlier, the mean overall evaluation time for the remaining five observers was 34.5 ± 47.9 s in method 1 and 35.5 ± 23.9 s in method 2 ($P < 0.001$ for comparison).

Table 3. Agreement of CAC Grading Between Non-ECG-Gated Chest CT and ECG-Gated Cardiac CT

	Cohen's Kappa	95% CI	Weighted Kappa	95% CI
Visual assessment				
Overall	0.565	0.511–0.619	-	-
Cardiothoracic radiologists	0.575	0.504–0.646	-	-
Observer 1	0.574	0.470–0.678	0.704	0.620–0.787
Observer 2	0.604	0.497–0.711	0.762	0.690–0.834
Observer 3	0.693	0.595–0.791	0.805	0.738–0.872
Radiology residents	0.567	0.499–0.635	-	-
Observer 4	0.552	0.443–0.660	0.702	0.619–0.785
Observer 5	0.624	0.521–0.727	0.753	0.679–0.826
Observer 6	0.529	0.419–0.639	0.705	0.630–0.780
Modified length-based method				
Overall	0.695	0.638–0.752	-	-
Cardiothoracic radiologists	0.688	0.623–0.753	-	-
Observer 1	0.703	0.604–0.802	0.825	0.761–0.888
Observer 2	0.635	0.531–0.738	0.788	0.722–0.854
Observer 3	0.723	0.627–0.819	0.834	0.771–0.897
Radiology residents	0.703	0.637–0.769	-	-
Observer 4	0.712	0.614–0.810	0.841	0.784–0.898
Observer 5	0.688	0.589–0.787	0.831	0.774–0.888
Observer 6	0.722	0.627–0.817	0.835	0.773–0.898

CAC = coronary artery calcium, ECG = electrocardiogram, CI = confidence interval, CT = computed tomography

https://doi.org/10.3348/kjr.2022.0826

Fig. 2. Discrepancies in grading between electrocardiogram (ECG)-gated computed tomography (CT) and non-ECG-gated chest CT. **A:** A case of moderate coronary artery calcium (CAC) in an 81-year-old female. On ECG-gated calcium scoring CT, the Agatston score was 244.7. **B:** On non-ECG-gated, low-dose chest CT performed at Sn 150 kVp (150 kVp with tin filter), three of six observers downgraded this case as mild in visual assessment. All observers correctly graded this case as moderate in the modified length-based grading method. **C:** A case of severe CAC in a 52-year-old male. The Agatston score was 881.3 on ECG-gated calcium scoring CT. **D:** Four of six observers downgraded this case on non-ECG-gated, standard dose chest CT scanning at 120 kVp as mild or moderate in visual assessment. All observers correctly graded this case as severe in the modified length-based grading method.

Table 4. Subgroup Analysis for Agreement between Non-ECG-Gated Chest CT and ECG-Gated Cardiac CT According to Scan Parameters of Non-ECG-Gated Chest CT

	Cohen's Kappa	95% CI
Radiation dose		
Standard dose (n = 42)		
Visual assessment	0.587	0.488–0.686
Modified length-based method	0.689	0.589–0.789
Low dose (n = 78)		
Visual assessment	0.552	0.487–0.617
Modified length-based method	0.699	0.630–0.768
Slice thickness		
Slice thickness 1/1.25 mm (n = 83)		
Visual assessment	0.572	0.504–0.640
Modified length-based method	0.699	0.631–0.767
Slice thickness 2.5/3 mm (n = 16)		
Visual assessment	0.593	0.433–0.753
Modified length-based method	0.678	0.504–0.852
Slice thickness 5 mm (n = 21)		
Visual assessment	0.508	0.396–0.620
Modified length-based method	0.690	0.543–0.837
Tube voltage		
Same kVp between ECG-gated CT and non-ECG-gated CT (n = 107)		
Visual assessment	0.557	0.488–0.626
Modified length-based method	0.670	0.600–0.740
Different kVp between ECG-gated CT and non-ECG-gated CT (n = 13)		
Visual assessment	0.578	0.482–0.674
Modified length-based method	0.699	0.631–0.767
Kernel		
Soft kernel (n = 78)		
Visual assessment	0.557	0.488–0.626
Modified length-based method	0.670	0.600–0.740
Sharp kernel (n = 42)		
Visual assessment	0.578	0.482–0.674
Modified length-based method	0.699	0.631–0.767

95% CI: The 95% confidence interval (CI) are calculated by Wilson score method. ECG = electrocardiogram, CT = computed tomography

Comparison with the Conventional Length-Based Grading Method

The length-based grading method demonstrated lower agreement with the standard reference cardiac CT for the two observers than method 2 (weighted kappa of 0.757 [95% CI: 0.686–0.828] for observer 2 and 0.831 [95% CI: 0.774–0.888] for observer 5) (Table 5). The evaluation time for CAC grading of two observers was significantly shorter in method 2 (mean 38.8 ± 30.3 s) than in the length-based grading method (mean 60.8 ± 33.6 s) ($P < 0.001$).

DISCUSSION

We developed and validated a novel and simple ordinal scoring method to evaluate CAC severity on chest CT. Compared with visual estimation, the modified length-based grading method had a better agreement in CAC severity categorization with the standard reference cardiac CT. The modified length-based grading method for CAC on chest CT showed good agreement between the observers for the discrimination of CAC categories. However, the difference in evaluation time between the two methods was very short (1.7 s), which is not sufficient to achieve clinical significance.

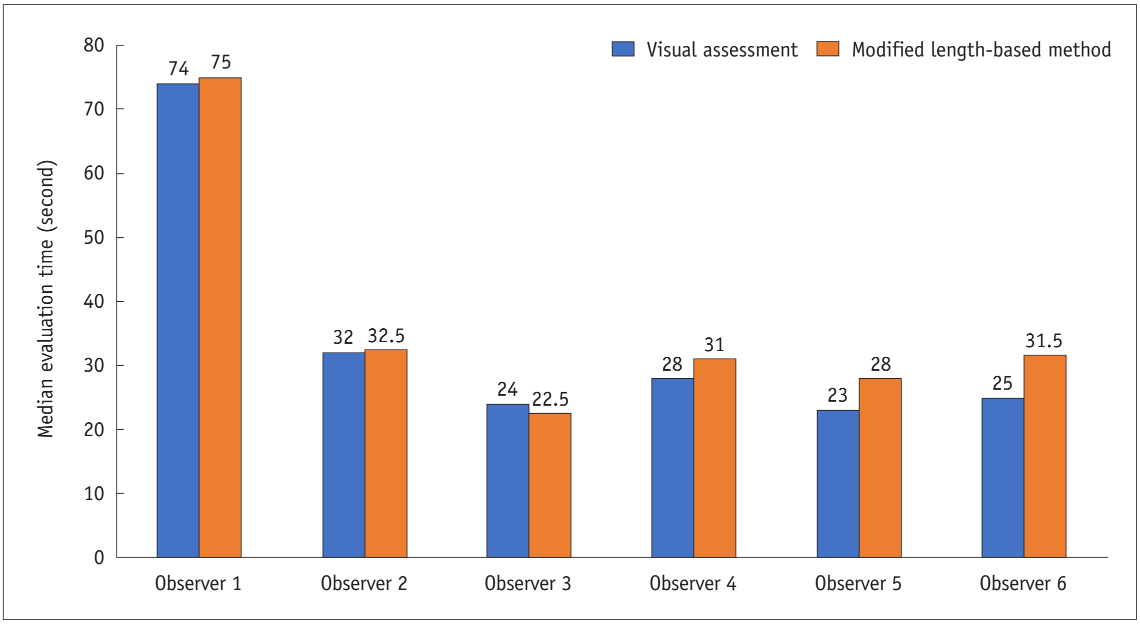

Fig. 3. Evaluation time (seconds) for coronary artery calcium grading. Bar graph showing the average evaluation times for each reader using the two methods. Observers 1–3, board-certified cardiothoracic radiologists; observers 4–6, radiology residents.

Table 5. Agreement of CAC Grading between Non-ECG-Gated Chest CT and ECG-Gated Cardiac CT in Two Select Observers According to Assessment Methods

	Cohen's Kappa	95% CI	Weighted Kappa	95% CI	Average Evaluation Time (sec)
Observer 2					
Visual assessment	0.604	0.497–0.711	0.762	0.690–0.834	32
Modified length-based method	0.635	0.531–0.738	0.788	0.722–0.854	32.5
Length-based method	0.596	0.489–0.704	0.757	0.686–0.828	68.7
Observer 5					
Visual assessment	0.624	0.521–0.727	0.753	0.679–0.826	23
Modified length-based method	0.688	0.589–0.787	0.831	0.774–0.888	28
Length-based method	0.700	0.602–0.799	0.819	0.751–0.887	52.9

CAC = coronary artery calcium, ECG = electrocardiogram, CI = confidence interval, CT = computed tomography

Visual assessment seems attractive among the various methods for CAC severity grading on chest CT because it is simple and quick. Previous studies have reported that visual assessment of CAC on chest CT was feasible for risk stratification of coronary artery disease [18] and exhibited good agreement with the CAC category based on the Agatston score (weight kappa 0.67–0.78) [13,19]. Nevertheless, visual assessment has the disadvantages of high subjectivity and consequent high interobserver variability owing to the lack of clear criteria for grading [14]. Similar to other reported studies, our study demonstrated that interobserver agreement for discriminating between the four CAC categories was moderate for visual estimation (Fleiss kappa 0.553, 95% CI: 0.496–0.610) and the agreement with the modified length-based grading method by the same observer was moderate to good (Fleiss kappa 0.401–0.717).

Besides the visual assessment, a few ordinal scoring methods, such as artery-based grading, segment-based grading, or segment-involvement scores, have been suggested [12-15,20]. A previous study indicated that artery-based grading is the most reliable option for evaluating CAC severity on low-dose chest CT, considering the high agreement between observers and other grading methods [14]. Additionally, another study compared the reliability of the three artery-based ordinal grading methods (extent-based grading, Weston score, and length-based grading) and suggested that the length-based method was the most reliable option, exhibiting the best agreement between observers and cardiac CT [21]. However, the proposed artery-based grading methods are still somewhat complex and time-consuming in clinical practice, as opposed to the research setting. In our study, the agreement with the standard reference cardiac CT of the modified length-based grading method was better than that of the previous length-based method. Furthermore, the modified length-based grading method had a significantly shorter evaluation time than that of the length-based method.

In our study, the modified length-based grading method showed better agreement between observers and standard reference cardiac CT than visual estimation. Furthermore, the recorded time for evaluation via the dedicated server suggested that the modified length-based grading method had a slightly but insignificantly longer evaluation time than visual assessment.

We validated our modified length-based grading method using a chest CT dataset with various scanning protocols. The scan parameters, including slice thickness and reconstruction kernel, significantly affected the agreement of CAC severity assessed by cardiac and chest CT [22]. Previous studies have highlighted the importance of these technical factors in investigating the reliability of CAC assessments on chest CT [23-25]. For example, CT images obtained using a low-dose protocol with a reduced tube current or sharp kernel reconstruction have higher noise and induce a higher CAC score [26]. Additionally, a tiny CAC could be depicted with higher sensitivity using a thinner slice thickness, resulting in higher CAC scores [25,27]. In the subgroup analysis of the scan protocol, the modified length-based grading method was more effective in terms of higher agreement for cardiac CT than visual assessment, irrespective of radiation dose, slice thickness, or reconstruction.

The recent development of artificial intelligence (AI) makes automatic CAC scoring feasible [28,29]. Consequently, the clinical application of AI-based automatic CAC scoring has been extended to chest CT [30,31]. Automatic CAC scoring on low-dose chest CT showed excellent reliability with manual CAC scoring, but the reliability of CAC score-based severity categorization varies among datasets with different scan protocols [30,32] Therefore, the improvement of an AI-based automatic scoring algorithm specific to the scanning protocol is necessary to apply automatic CAC scoring to chest CT. In contrast, the modified length-based grading method suggested in our study has advantages in that it is less affected by the scan protocols. Notably, our modified length-based grading method can be helpful in the assessment of CAC severity on chest CT until optimization or popularization of automatic CAC scoring can be achieved.

Our study has some limitations. First, we did not analyze the prognostic value of the modified length-based grading method. Additional studies conducted across multiple centers could help to identify the prognostic value of this modified length-based grading method. Second, our modified length-based grading method considers only the calcium size in the grading, whereas the Agatston score is quantified from the product of calcium area and density. Nevertheless, our modified length-based grading method can be used to stratify patients' cardiovascular risk by a rapid and straightforward assessment of CAC severity in the clinical setting.

Therefore, the modified length-based grading method can be a good option for evaluating CAC severity on non-ECG-gated chest CT, showing better interobserver agreement and agreement with cardiac CT for CAC categorization than visual assessment. Importantly, this length-based grading method is less vulnerable to variations in scan parameters and provides good interobserver agreement and an acceptable evaluation time. The modified length-based grading method can be helpful for accurate and efficient cardiovascular risk stratification on non-ECG-gated chest CT scans and subsequent guidance for patient management.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2022.0826.

Availability of Data and Material

The datasets generated or analyzed during the study are not publicly available due medical confidentiality but are available from the corresponding author on reasonable request.

Conflicts of Interest

Kyunghwa Han, a contributing editor of the *Korean Journal of Radiology*, was not involved in the editorial evaluation or decision to publish this article. All remaining authors have declared no conflicts of interest.

Author Contributions

Conceptualization: Suh Young Kim, Young Joo Suh. Data curation: Suh Young Kim, Young Joo Suh. Formal analysis: Suh Young Kim, Young Joo Suh. Funding acquisition: Young Joo Suh. Investigation: Na Young Kim, Suji Lee, Kyungsun Nam, Jeongyun Kim, Hwan Kim, Hyunji Lee. Methodology:

Suh Young Kim, Young Joo Suh. Project administration: Young Joo Suh. Resources: Suh Young Kim, Young Joo Suh. Software: Young Joo Suh. Supervision: Young Joo Suh, Kyunghwa Han, Hwan Seok Yong. Validation: Young Joo Suh, Kyunghwa Han, Hwan Seok Yong. Visualization: Young Joo Suh, Kyunghwa Han, Hwan Seok Yong. Writing—original draft: Suh Young Kim, Young Joo Suh. Writing—review & editing: all authors.

ORCID iDs
Suh Young Kim
https://orcid.org/0000-0002-5101-0167
Young Joo Suh
https://orcid.org/0000-0002-2078-5832
Na Young Kim
https://orcid.org/0000-0003-1645-2434
Suji Lee
https://orcid.org/0000-0002-8770-622X
Kyungsun Nam
https://orcid.org/0000-0002-1673-0935
Jeongyun Kim
https://orcid.org/0000-0003-3023-3633
Hwan Kim
https://orcid.org/0000-0001-7353-9777
Hyunji Lee
https://orcid.org/0000-0001-7072-0001
Kyunghwa Han
https://orcid.org/0000-0002-5687-7237
Hwan Seok Yong
https://orcid.org/0000-0003-0247-8932

Funding Statement
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT).(No. 2021R1A2C4002195).

Acknowledgments
The authors would like to thank Sunggoo Kwon from Coreline Soft, Co.Ltd., for the technical support. Minju Kim (Department of Medical Statistics, Asan Medical Center, University of Ulsan College of Medicine) kindly provided statistical advice for this manuscript.

REFERENCES

1. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med* 2008;358:1336-1345
2. Arad Y, Goodman KJ, Roth M, Newstein D, Guerci AD. Coronary calcification, coronary disease risk factors, C-reactive protein, and atherosclerotic cardiovascular disease events: the St. Francis Heart Study. *J Am Coll Cardiol* 2005;46:158-165
3. Elias-Smale SE, Proença RV, Koller MT, Kavousi M, van Rooij FJ, Hunink MG, et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. *J Am Coll Cardiol* 2010;56:1407-1414
4. Reiter MJ, Nemesure A, Madu E, Reagan L, Plank A. Frequency and distribution of incidental findings deemed appropriate for S modifier designation on low-dose CT in a lung cancer screening program. *Lung Cancer* 2018;120:1-6
5. Hecht HS, Cronin P, Blaha MJ, Budoff MJ, Kazerooni EA, Narula J, et al. 2016 SCCT/STR guidelines for coronary artery calcium scoring of noncontrast noncardiac chest CT scans: a report of the Society of Cardiovascular Computed Tomography and Society of Thoracic Radiology. *J Cardiovasc Comput Tomogr* 2017;11:74-84
6. Kim SM, Chung MJ, Lee KS, Choe YH, Yi CA, Choe BK. Coronary calcium screening using low-dose lung cancer screening: effectiveness of MDCT with retrospective reconstruction. *AJR Am J Roentgenol* 2008;190:917-922
7. Jacobs PC, Isgum I, Gondrie MJ, Mali WP, van Ginneken B, Prokop M, et al. Coronary artery calcification scoring in low-dose ungated CT screening for lung cancer: interscan agreement. *AJR Am J Roentgenol* 2010;194:1244-1249
8. Budoff MJ, Nasir K, Kinney GL, Hokanson JE, Barr RG, Steiner R, et al. Coronary artery and thoracic calcium on noncontrast thoracic CT scans: comparison of ungated and gated examinations in patients from the COPD Gene cohort. *J Cardiovasc Comput Tomogr* 2011;5:113-118
9. Xie X, Zhao Y, de Bock GH, de Jong PA, Mali WP, Oudkerk M, et al. Validation and prognosis of coronary artery calcium scoring in nontriggered thoracic computed tomography: systematic review and meta-analysis. *Circ Cardiovasc Imaging* 2013;6:514-521
10. Wu MT, Yang P, Huang YL, Chen JS, Chuo CC, Yeh C, et al. Coronary arterial calcification on low-dose ungated MDCT for lung cancer screening: concordance study with dedicated cardiac CT. *AJR Am J Roentgenol* 2008;190:923-928
11. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M Jr, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol* 1990;15:827-832
12. Shemesh J, Henschke CI, Shaham D, Yip R, Farooqi AO, Cham MD, et al. Ordinal scoring of coronary artery calcifications on low-dose CT scans of the chest is predictive of death from cardiovascular disease. *Radiology* 2010;257:541-548
13. Chiles C, Duan F, Gladish GW, Ravenel JG, Baginski SG, Snyder BS, et al. Association of coronary artery calcification and mortality in the national lung screening trial: a comparison of three scoring methods. *Radiology* 2015;276:82-90
14. Suh YJ, Lee JW, Shin SY, Goo JM, Kim Y, Yong HS. Coronary

artery calcium severity grading on non-ECG-gated low-dose chest computed tomography: a multiple-observer study in a nationwide lung cancer screening registry. *Eur Radiol* 2020;30:3684-3691
15. Huang YL, Wu FZ, Wang YC, Ju YJ, Mar GY, Chuo CC, et al. Reliable categorisation of visual scoring of coronary artery calcification on low-dose CT for lung cancer screening: validation with the standard Agatston score. *Eur Radiol* 2013;23:1226-1233
16. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol* 2012;65:778-784
17. Einstein AJ, Johnson LL, Bokhari S, Son J, Thompson RC, Bateman TM, et al. Agreement of visual estimation of coronary artery calcium from low-dose CT attenuation correction scans in hybrid PET/CT and SPECT/CT with standard Agatston score. *J Am Coll Cardiol* 2010;56:1914-1921
18. Shemesh J, Henschke CI, Farooqi A, Yip R, Yankelevitz DF, Shaham D, et al. Frequency of coronary artery calcification on low-dose computed tomography screening for lung cancer. *Clin Imaging* 2006;30:181-185
19. Kim YK, Sung YM, Cho SH, Park YN, Choi HY. Reliability analysis of visual ranking of coronary artery calcification on low-dose CT of the thorax for lung cancer screening: comparison with ECG-gated calcium scoring CT. *Int J Cardiovasc Imaging* 2014;30 Suppl 2:81-87
20. Htwe Y, Cham MD, Henschke CI, Hecht H, Shemesh J, Liang M, et al. Coronary artery calcification on low-dose computed tomography: comparison of Agatston and ordinal scores. *Clin Imaging* 2015;39:799-802
21. Lee S, Suh YJ, Nam K, Lee K, Lee HJ, Choi BW. Comparison of artery-based methods for ordinal grading of coronary artery calcium on low-dose chest computed tomography. *Eur Radiol* 2021;31:8108-8115
22. Kim JY, Suh YJ, Han K, Choi BW. Reliability of coronary artery calcium severity assessment on non-electrocardiogram-gated CT: a meta-analysis. *Korean J Radiol* 2021;22:1034-1043
23. Wan YL, Tsay PK, Wu PW, Juan YH, Tsai HY, Lin CY, et al. Impact of filter convolution and displayed field of view on estimation of coronary Agatston scores in low-dose lung computed tomography. *Int J Cardiol* 2017;236:451-457
24. Christensen JL, Sharma E, Gorvitovskaia AY, Watts JP Jr, Assali M, Neverson J, et al. Impact of slice thickness on the predictive value of lung cancer screening computed tomography in the evaluation of coronary artery calcification. *J Am Heart Assoc* 2019;8:e010110
25. Mühlenbruch G, Thomas C, Wildberger JE, Koos R, Das M, Hohl C, et al. Effect of varying slice thickness on coronary calcium scoring with multislice computed tomography in vitro and in vivo. *Invest Radiol* 2005;40:695-699
26. Geyer LL, Schoepf UJ, Meinel FG, Nance JW Jr, Bastarrika G, Leipsic JA, et al. State of the art: iterative CT reconstruction techniques. *Radiology* 2015;276:339-357
27. van der Bijl N, de Bruin PW, Geleijns J, Bax JJ, Schuijf JD, de Roos A, et al. Assessment of coronary artery calcium by using volumetric 320-row multi-detector computed tomography: comparison of 0.5 mm with 3.0 mm slice reconstructions. *Int J Cardiovasc Imaging* 2010;26:473-482
28. Lee JG, Kim H, Kang H, Koo HJ, Kang JW, Kim YH, et al. Fully automatic coronary calcium score software empowered by artificial intelligence technology: validation study using three CT cohorts. *Korean J Radiol* 2021;22:1764-1776
29. Martin SS, van Assen M, Rapaka S, Hudson HT Jr, Fischer AM, Varga-Szemes A, et al. Evaluation of a deep learning–based automated CT coronary artery calcium scoring algorithm. *JACC Cardiovasc Imaging* 2020;13(2 Pt 1):524-526
30. Suh YJ, Kim C, Lee JG, Oh H, Kang H, Kim YH, et al. Fully automatic coronary calcium scoring in non-ECG-gated low-dose chest CT: comparison with ECG-gated cardiac CT. *Eur Radiol* 2023;33:1254-1265
31. van Velzen SGM, Lessmann N, Velthuis BK, Bank IEM, van den Bongard DHJG, Leiner T, et al. Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols. *Radiology* 2020;295:66-79
32. Kang HW, Ahn WJ, Jeong JH, Suh YJ, Yang DH, Choi H, et al. Evaluation of fully automated commercial software for Agatston calcium scoring on non-ECG-gated low-dose chest CT with different slice thickness. *Eur Radiol* 2023;33:1973-1981