Acta Radiologica Open
9(4) 1–7

DOI: 10.1177/2058460120922147
journals.sagepub.com/home/arr
SAGE

Reproducibility of calcium scoring of the coronary arteries: comparison between different vendors and iterative reconstructions

Kyu Sung Choi[1], Whal Lee[2,3,4], Joon Hyung Jung[5] and Eun-Ah Park[2,3,4]

Abstract

Background: The coronary artery calcium scoring (CCS) has been widely used for cardiac risk stratification for asymptomatic patients.

Purpose: To assess the reproducibility of CCS performed on four different computed tomography (CT) scanners, and compare the variability between two reconstruction algorithms, filtered back projection (FBP), and iterative reconstruction (IR).

Material and Methods: A CCS phantom was made from agar and contained 23 pieces of chicken bones. The phantom was repeatedly scanned using four different CT scanners: Toshiba; GE; Philips; and Siemens. Images were reconstructed using FBP and IR. Agatston and volume scores of total bone fragments were calculated and the overall differences between the instruments were evaluated using the Friedman test. Comparison of the Agatston and volume scores between the two reconstruction algorithms, for each instrument, was evaluated using the Wilcoxon signed rank test.

Results: The difference in the Agatston scores was significantly different between the four machines ($P = 0.001$). The Toshiba scanner yielded the highest score followed by Philips, GE, and Siemens scanners. There was no difference in the CCS evaluated using the two reconstruction algorithms, except in case of the Siemens scanner ($P = 0.032$).

Conclusion: CCS performed on different scanners varied significantly. In the Toshiba, Philips, and GE scanners, there was no significant difference in the CCS determined using either an IR or the FBP algorithm. In the Siemens scanner, applying the IR algorithm resulted in a slightly different scores, which might not be clinically significant.

Keywords

Iterative reconstruction, coronary artery calcium score, reproducibility, interplatform, computed tomography, inter-vendor variability

Received 22 October 2019; accepted 3 April 2020

Introduction

Coronary calcium scoring (CCS) has emerged as one of the most important methods for risk stratification and a reliable follow-up tool for coronary heart disease (1–3). In 1990, Agatston first proposed an algorithm to measure the burden of coronary calcification using electron beam computed tomography (EBCT) (4). Since then, the Agatston score has been widely used to predict the possibility of coronary artery events, such as acute myocardial infarction (5,6).

For a method to be credible, it is crucial that the variability in measurement is as low as possible.

[1]Graduate School of Medical Science and Engineering, Korea Advanced Institute for Science and Technology (KAIST), Daejeon, Republic of Korea
[2]Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea
[3]Department of Radiology, Seoul National University College of Medicine, Seoul, Republic of Korea
[4]Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Republic of Korea
[5]Department of Neuropsychiatry, Seoul National University Hospital, Seoul, Republic of Korea

Corresponding author:
Whal Lee, Department of Radiology, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul, 110-744, Republic of Korea.
Email: whal.lee@gmail.com

There are several studies assessing the variability in CCS using identical machines, interscan variability, and ways to reduce this variability (7,8). McCollough et al. (9) reported the standardized quantification of coronary artery calcium results from equivalent calcium scores, acquired using different computed tomography (CT) systems. However, the hardware and software used in CT has improved dramatically since Agatston first proposed the Agatston score for CCS. Multi-detector channel CT (MDCT) has replaced EBCT and various reconstruction algorithms have been proposed to improve image quality. Iterative reconstruction (IR) will eventually replace filtered back projection (FBP) reconstruction as the algorithm of choice. IR enhances the CT image quality considerably and has the potential to reduce radiation dose in CT angiography for coronary artery by reducing image noise (10,11). However, the effect of IR on CCS is yet to be evaluated.

The aim of the present study was to assess the variability in CCS performed on the CT scanners from four different manufactures (Toshiba, GE, Philips, and Siemens) and to evaluate the effect of the IR algorithm on CCS.

Material and Methods

Coronary calcium phantom

The coronary calcium phantom used in this study was made of agar and chicken bones (Fig. 1). Agar (cell-culture and electrophoresis grade) was dissolved in water (5 g in 500 mL), by heating in a regular microwave oven, and gently poured into a plastic container. On cooling, agar solidified to form a gel foam. Cooked and dried chicken bones were broken into small fragments, a few millimeters in size, with a hammer. Twenty-three bone pieces of varying size (range = 1.4–6.0 mm; mean size = 2.88 ± 1.06 mm) were collected and inserted into the agar gel foam using needles.

Scanning protocols

The agar phantom was scanned five times each using four different CT scanners (Toshiba, GE, Philips, and Siemens). The phantom was moved randomly between consecutive scans to mimic the positioning variability observed during actual patient scanning in the clinical setting. All the CT scans were performed using a sequential and prospective acquisition. CCS was performed using manufacturer recommended protocols for each of the scanners (Table 1). The acquired CT images were reconstructed into 2.5–3-mm-thick slices using FBP and IR algorithms, which was used for comparison. A 0.5-mm scan was acquired using the GE scanner as a reference, and the reference image was not used for comparison.

Scoring methods

CT images were analyzed using Rapidia (Infinitt, Seoul, Republic of Korea). Using images from a 0.5-mm reference scan, acquired using the GE scanner, all the bone pieces were located and serially numbered from 1 to 23. Manufacturer-recommended calcium scoring protocols for each of the CT machines were

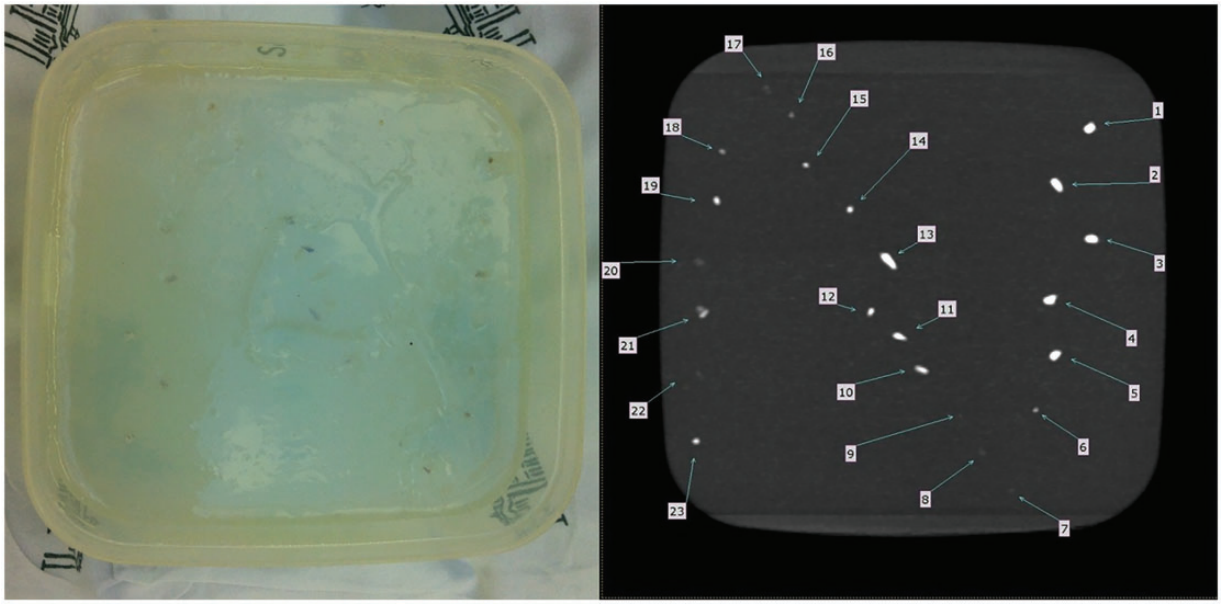

Fig. 1. An in vitro agar phantom for coronary calcium scoring.

Table 1. CT protocols used for scanning the agar coronary calcium phantom.

Scanner	GE MEDICAL SYSTEMS, Discovery CT750HD	Philips, Ingenuity CT	Siemens SOMATOM Definition	Toshiba, Aquilion ONE
Acquisition mode	Sequential	Sequential	Sequential	Sequential
ECG synchronization	Prospective (70%)	Prospective (70%)	Prospective (70%)	Prospective (70%)
Peak voltage (kV)	120	120	120	120
Spatial resolution (mm)	0.35	0.32	0.36	0.35
Tube current (mA)	75	57	52	40
CTDIvol (mGy)	7.2	3.8	6.2	6.6
Rotation time (ms)	228	420	330	350
Reconstruction algorithm	FBP/ASIR 50%	FBP/I5	B35f/I36f	FC12n/FD12 AIDR STD

CT, computed tomography.

used to automatically locate bone inserts, which were >130 Hounsfield units (HU). The Agatston score and volume score of each bone piece was measured. Calcium score of the whole phantom was defined as the sum of all the scores from individual pieces.

Comparison of calcium scores

Total calcium scores from different CT scanners and different reconstruction algorithms were compared. Friedman test was used for the overall comparison of the Agatston scores and volume scores between various CT scanners. Wilcoxon signed rank test with Bonferroni correction was performed as a post-hoc analysis. Only FBP reconstruction data was used for the comparison. Wilcoxon signed rank test was used for the comparison of the Agatston scores and volume scores between FBP and IR algorithms. All statistical analyses were performed using MedCalc software (version 16.2.1, MedCalc Software); $P < 0.05$ was considered statistically significant.

Results

Comparison of number of detected calcifications

Of the 23 bone pieces, only 8–14 pieces were detected for each scanner, since other pieces were too small and their HU values were too low to be detected. There were significant differences in detected number of bone pieces between different scanners, both with FBP and IR algorithms ($P = 0.007$ and $P = 0.013$). The Agatston score obtained from the Siemens scanner detected the least number of calcifications (Table 2) among the four vendors, with both FBP and IR algorithms. However, there were no significant differences in number of detected bone pieces for each vendor, comparing FBP and IR algorithms (all $P > 0.05$).

Table 2. Agatston scores evaluated from CT images acquired using four different CT scanners.

	Detected calcifications (n)		Agatston score	
	FBP	IR	FBP	IR
GE	12.0 ± 1.0	12.6 ± 1.1	153.4 ± 7.7	157.8 ± 6.6
Philips	11.6 ± 0.9	11.6 ± 0.9	166.7 ± 4.2	166.9 ± 3.9
Siemens	8.8 ± 1.3	10.2 ± 0.8	115.0 ± 5.1	124.1 ± 5.4
Toshiba	11.0 ± 0.7	11.0 ± 0.7	224.5 ± 14.4	225.6 ± 12.2

Values are given as mean ± SD.
CT, computed tomography; FBP, filtered back projection; IR, iterative reconstruction.

Table 3. Volume scores evaluated from CT images acquired using four different CT scanners.

	Detected calcifications (n)		Volume score	
	FBP	IR	FBP	IR
GE	11.8 ± 1.3	12.8 ± 1.3	143.3 ± 8.4	146.6 ± 8.2
Philips	11.6 ± 0.9	11.6 ± 0.9	157.5 ± 6.3	156.3 ± 5.2
Siemens	9.0 ± 1.2	9.0 ± 1.2	126.2 ± 5.1	130.1 ± 4.2
Toshiba	11.0 ± 0.7	11.0 ± 0.7	169.2 ± 9.3	169.3 ± 8.9

Values are given as mean ± SD.
CT, computed tomography; FBP, filtered back projection; IR, iterative reconstruction.

Comparison between different scanners

The difference in the Agatston scores and volume scores between the four CT machines was significant ($P = 0.0018$). The Toshiba scanner yielded the highest Agatston score followed by the Philips, GE, and Siemens scanners. The scores were significantly different in a pairwise comparison of the subgroups (Tables 2 and 3, Fig. 2a). There were significant inter-vendor differences ($P = 0.003$) in the volume scores. In the pairwise comparison, there was no significant difference between GE and Philips ($P = 0.068$), and

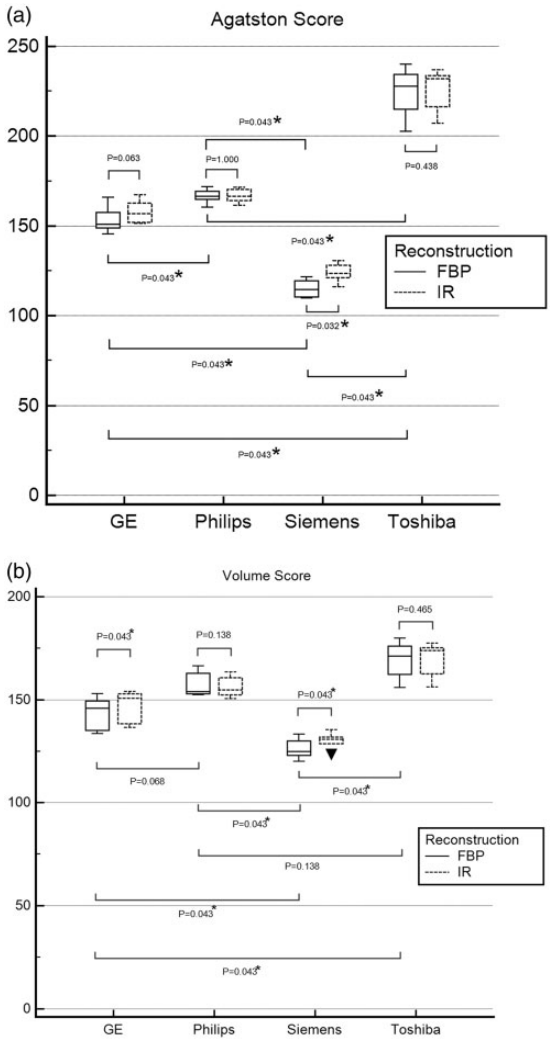

Fig. 2. Comparison of (a) Agatston score and (b) volume score determined from CT images acquired using four different CT scanners (Toshiba, Philips, GE, and Siemens) and following two different reconstruction algorithms (iterative reconstruction and filtered back projection).
CT, computed tomography.

Philips and Toshiba ($P = 0.138$), while all other combinations showed significant differences ($P = 0.043$ for all comparisons) (Tables 2 and 3, Fig. 2b). Agatston score differences were in the range of 13.3–109.48 (–31.0% to +34.7% when comparing with the calcium score from a Philips scanner, which was the median value of four machines). The differences in the volume scores were relatively smaller, in the range of 11.68–42.92 (–19.8% to +7.4% compared to Philips data).

Comparison between FBP and IR algorithms for different scanners

The Agatston and volume scores, obtained using IR, were different for different scanners. In the case of the Siemens scanner, there was an increase in the Agatston score between FBP and IR ($P = 0.032$). The mean difference was 9.1. In the Toshiba, Philips, and GE scanners, the Agatston score from FBP reconstruction was comparable to that of IR. Agatston scores and volume scores for FBP and IR are shown in Tables 2 and 3. There was an increase in the volume score, obtained from IR, in the case of the Siemens and GE scanners ($P = 0.043$ for both) (Fig. 2b). Mean differences were 3.3 and 3.9, respectively. There was no significant difference in the case of the Toshiba and Philips scanners (Fig. 2b).

To investigate the effect of IR in detecting tiny calcifications with low calcium scores, calcium scores of calcification observed in each of the five scans processed using FBP and IR algorithms were averaged to obtain a Bland–Altman plot (Fig. 3). In the range of average scores <10, Agatston scores from IR were significantly higher than the scores from FBP for all scanners except the Toshiba scanner ($P = 0.0078$ for GE, $P = 0.0156$ for Philips, $P = 0.0313$ for Siemens, and $P = 0.916$ for Toshiba). In the range of scores >10, Agatston scores were higher for FBP than IR for all vendors except Toshiba. However, the differences were not statistically significant ($P > 0.05$ for all comparisons). The Agatston score from a Toshiba scanner showed a relatively higher agreement between FBP and IR when compared to the other vendors, except for an outlier with a score < 10 (Fig. 3d). Moreover, larger calcifications tended to show less variability and have almost identical values. Small calcifications showed larger variability (Fig. 3).

Discussion

CCS is a widely accepted non-invasive tool to assess risk stratification of coronary artery events (1–3). Many studies suggest that there are no significant differences in CCS between different CT scans, vendors, and scoring software (7,9). However, Willemink et al. (12) recently reported that there could be significant inter-vendor variability in the Agatston scores from state-of-the-art CT machines, which can lead to inappropriate risk stratification, and re-stratification may lead to subsequent loss of early treatment.

In the present study, we investigated the variability in calcium scoring using different CT scanners. We observed significant differences, and our result differs from the findings of McCollough et al. (9), who applied standardization at a set noise level of 20 HU for all the scanners. We did not apply any standardization in the present study, which might be the reason for the observed differences. However, we followed the regular clinical protocol, which can lead to significantly different CCS. This is in line with the recent study by Willemink et al. (12).

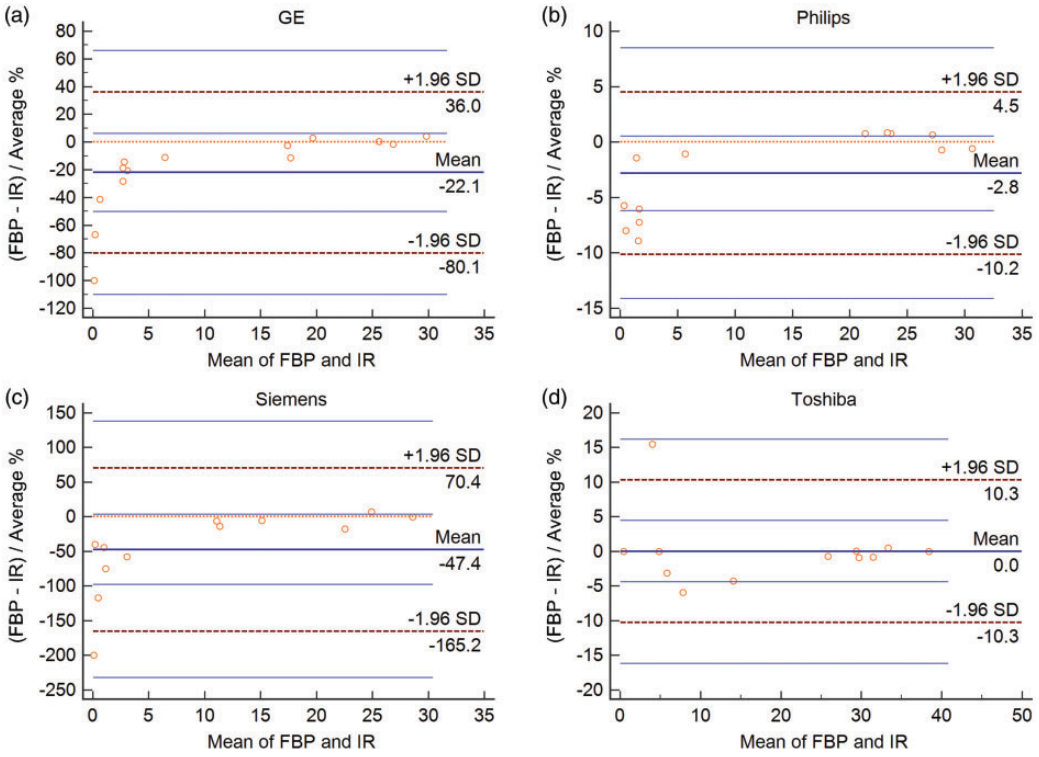

Fig. 3. Bland–Altman plot comparing the two algorithms (iterative reconstruction and filtered back projection) employed for reconstructing images acquired using (a) GE, (b) Philips, (c) Siemens, and (d) Toshiba CT scanners. Plotted scores are averages of calcium scores for five scans for each of the calcifications.
CT, computed tomography.

Several factors could affect this difference. First, the number of calcifications detected can be different. In the present study, the Agatston score obtained from the Siemens scanner detected the least number of calcifications (Table 2) among the four vendors, with both FBP and IR algorithms ($P = 0.007$ and $P = 0.013$). However, these undetected calcifications were very small and their total calcium score too low to explain all the differences. Furthermore, the volume difference and HU values can affect the significant difference of CCS. Volume scores are dependent solely on the number of voxels with a HU value >130, without considering the actual HU values of the detected calcifications (13), whereas Agatston scores are not only dependent on the lesion area occupied by calcification, but also on the HU values of calcifications (4). Though volume scores differ significantly, there is a lot more variation in the Agatston scores. This implies that the HU values may also differ with CT scanners, possibly frequently in a clinical setting, especially with state-of-the-art CT machines (12).

There also could be an issue of risk reclassification with different vendors. In standardized categories for the CCS, patients are categorized into risk groups by using Agatston scores as follows: 0 = absent calcification, very low risk; 1–10 = minimal calcifications, low risk; 11–100 = mild calcifications, intermediate risk; 101–400 = moderate calcifications, moderately high risk; and >400 = extensive calcifications, high risk (7,14). Our results suggest that differences between scanners are so high that patients may be classified in different risk groups depending on what scanner was used for the CCS. Moreover, the significance of zero calcium score has been highlighted because of a very high negative predictive value (up to 99%) for cardiovascular events in the next 2–5 years (14,15).

IR has been validated in many recent studies to reduce image noise significantly and resulted in a reduction of calcium scores by reducing "blooming artifacts" (3,16). In the present study, there was very little difference in the number of calcifications detected, using IR and FBP reconstruction algorithms. The Agatston score obtained from FBP was comparable to that of IR in three CT scanners (Toshiba, Philips, and GE), and was different for only one scanner (Siemens). In a Siemens scanner, though the Agatston score obtained from FBP was significantly higher than that obtained from IR ($P = 0.032$), the difference was relatively small (mean = 9.1) and might not be important in a clinical setting. There was no difference in the volume scores

obtained using FBP and IR algorithms in the case of the Toshiba and Philips scanners, and only small mean differences were observed between the Siemens and GE scanners (3.3 and 3.9, respectively). The data from the Siemens scanner show that when the calcification is dense, which implies higher CCS, the IR results in larger scores compared to FBP reconstruction. This results in lower total Agatston or volume scores for IR compared to FBP reconstruction. This result can be explained based on the ability of IR to detect larger number of small calcifications compared to FBP reconstruction, because of lower noise. Meanwhile, if the calcification is less dense, which means lower Agatston or volume scores, IR detects more calcifications, compared to FBP reconstruction, which makes total Agatston or volume scores obtained from IR higher than that of FBP reconstruction. This result can be explained based on the efficiency with which IR can measure smaller scores because of fewer blooming artifacts compared to FBP reconstruction. Overall, since most of the calcifications were small and less dense, total Agatston or volume scores obtained from IR were higher than the scores obtained from FBP reconstruction. Finally, the results from the Siemens scanner in the present study appear to be in line with the previous study of Schindler et al. (15).

There are several limitations in this study. This is not an in vivo study. We did not use an anthropomorphic cardiac phantom with calcium insertion. Our agar phantom has a fixed amount of calcification, which makes it difficult to simulate variable calcification with variable HU values observed in clinical settings. Finally, the HU values of the phantom calcification were relatively low compared to real patients, which magnifies the observed variability and may lead to wrong risk stratification, because of the lower HU values and the more densely divided stratification.

In conclusion, CCS varied significantly between CT scanners from four different manufacturers, when evaluated using conventional FBP reconstruction. There was no difference in the CCS obtained using IR and FBP methods in the Toshiba, Philips, and GE scanners. However, in the Siemens scanner, applying the IR method resulted in a slightly higher CCS, which may not be significant in a clinical setting.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Whal Lee https://orcid.org/0000-0003-1285-5033

References

1. Wayhs R, Zelinger A, Raggi P. High coronary artery calcium scores pose an extremely elevated risk for hard events. J Am Coll Cardiol 2002;39:225–230.
2. Detrano RC, Wong ND, Doherty TM, et al. Prognostic significance of coronary calcific deposits in asymptomatic high-risk subjects. Am J Med 1997;102:344–349.
3. Gebhard C, Fiechter M, Fuchs TA, et al. Coronary artery calcium scoring: Influence of adaptive statistical iterative reconstruction using 64-MDCT. Int J Cardiol 2013;167:2932–2937.
4. Agatston AS, Janowitz WR, Hildner FJ, et al. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol 1990;15:827–832.
5. Arad Y, Spadaro LA, Goodman K, et al. Prediction of coronary events with electron beam computed tomography. J Am Coll Cardiol 2000;36:1253–1260.
6. Raggi P, Callister TQ, Cooil B, et al. Identification of patients at increased risk of first unheralded acute myocardial infarction by electron-beam computed tomography. Circulation 2000;101:850–855.
7. Weininger M, Ritz KS, Schoepf UJ, et al. Interplatform reproducibility of CT coronary calcium scoring software. Radiology 2012;265:70–77.
8. Gassenmaier T, Allmendinger T, Kunz AS, et al. In vitro evaluation of a new iterative reconstruction algorithm for dose reduction in coronary artery calcium scoring. Acta Radiol Open 2017;6:2058460117710682.
9. McCollough CH, Ulzheimer S, Halliburton SS, et al. Coronary artery calcium: a multi-institutional, multimanufacturer international standard for quantification at cardiac CT. Radiology 2007;243:527–538.
10. Leipsic J, Labounty TM, Heilbron B, et al. Estimated radiation dose reduction using adaptive statistical iterative reconstruction in coronary CT angiography: the ERASIR study. AJR Am J Roentgenol 2010;195:655–660.
11. Moscariello A, Takx RA, Schoepf UJ, et al. Coronary CT angiography: image quality, diagnostic accuracy, and potential for radiation dose reduction using a novel iterative image reconstruction technique-comparison with traditional filtered back projection. Eur Radiol 2011;21:2130–2138.
12. Willemink MJ, Vliegenthart R, Takx RA, et al. Coronary artery calcification scoring with state-of-the-art CT scanners from different vendors has substantial effect on risk classification. Radiology 2014;273:695–702.

13. Yoon HC, Greaser LE 3rd, Mather R, et al. Coronary artery calcium: alternate methods for accurate and reproducible quantitation. Acad Radiol 1997;4:666–673.
14. Youssef G, Kalia N, Darabian S, et al. Coronary calcium: new insights, recent data, and clinical role. Curr Cardiol Rep 2013;15:325.
15. Schindler A, Vliegenthart R, Schoepf UJ, et al. Iterative image reconstruction techniques for CT coronary artery calcium quantification: comparison with traditional filtered back projection in vitro and in vivo. Radiology 2014;270:387–393.
16. Renker M, Nance JW Jr., Schoepf UJ, et al. Evaluation of heavily calcified vessels with coronary CT angiography: comparison of iterative and filtered back projection image reconstruction. Radiology 2011;260:390–399.