```
• begin
•     using Pkg; Pkg.activate(".")
•     using Images, FileIO, PlutoUI, CSV, DataFrames, PrettyTables, Latexify
• end
```

Table of Contents

Abstract

Agatston scoring is limited by factors such as arbitrary thresholding and does not include all the calcium information in computed tomography scans. Agatston scoring's limitations lead to problems with reproducibility and accuracy and, therefore, may not correlate well with known calcium mass. A scoring technique that removes the need for thresholding and quantifies calcium mass more accurately and reproducibly is needed.

This study adapted two techniques to create two novel calcium scoring approaches that include all the calcium information within an image. The integrated Hounsfield technique was adjusted for calcium mass quantification and the volume fraction technique. Integrated calcium mass and volume fraction calcium mass were calculated and compared to known calcium mass, along with Agatston scoring and spatially weighted calcium scoring on simulated CT scans across different patient sizes, energies, and motion levels.

The simulation was created, based on previously validated software, to match the scanning parameters of a 320-slice CT scanner. To simulate different patient sizes, additional fat rings were added, which resulted in a small-sized phantom of 30x20cm, a medium-sized phantom of 35x25cm, and a large-sized phantom of 40x30cm. Three calcification inserts of different diameters (1, 3, and 5 mm) and different hydroxyapatite (HA) densities were placed within the phantom. All calcium scoring measurements were repeated across different kVs (80-135 kV), patient sizes (small, medium, large), calcium insert sizes (1-5 mm), and calcium insert densities (25-800 mg/cm3). Physical phantom scans, acquired by a previous group, were then used to validate the results of the simulation study.

Both integrated and volume fraction calcium mass yielded lower root mean squared error (RMSE) and deviation (RMSD) values than Agatston scoring in all accuracy comparisons. Integrated and volume fraction calcium mass also yielded lower RMSE and RMSD values than Agatston or spatially weighted calcium scoring in every reproducibility comparison. Specifically, the RMSE and RMSD of the low-density calcifications for the integrated calcium mass technique vs. known calcium mass were 0.495 mg and 0.494 mg, respectively. The RMSE and RMSD values for the volume fraction calcium mass technique were 0.585 mg and 0.575 mg, respectively. Compared to Agatston scoring, which produced RMSE and RMSD values of 3.509 mg and 2.240 mg, respectively. Reproducibility across all patient sizes, calcium densities, and tube voltages, integrated calcium mass produced RMSE and RMSD values of 0.708 mg and 0.692 mg, respectively. Volume fraction calcium mass produced RMSE and RMSD values of 0.783 mg and 0.776 mg, respectively. Spatially weighted calcium scoring produced RMSE and RMSD values of 4.722 mg and 4.714 mg, respectively. Agatston scoring produced RMSE and RMSD values of 0.982 mg and 0.980 mg, respectively. The percentage of false-negative (CAC=0) scores produced by integrated calcium mass, volume fraction calcium mass, spatially weighted calcium scoring, and Agatston scoring were 11.111%, 11.111%, 42.593%, and 38.889%, respectively. The percentage of false-positive (CAC>0) scores produced by integrated calcium mass, volume fraction calcium mass, spatially weighted calcium scoring, and Agatston scoring were 11.111%, 11.111%, 16.667, and 0.0%, respectively.

1. Introduction

The coronary artery calcification (CAC) score is a standard atherosclerotic marker [CITE]. CAC scoring is a test performed using computed tomography (CT) that measures the amount of calcium buildup within the walls of the heart's arteries and is an essential predictor of coronary heart disease [CITE]. The leading cause of death in the United States is coronary heart disease, killing 659,000 people annually, and it is the third leading cause of mortality worldwide [CITE].

Agatston scoring is the most common CAC scoring technique [CITE] and is a good predictor of major adverse cardiac events (MACE) [CITE]. Although, studies have shown that a significant number of patients have been characterized as having no calcium (CAC=0) while still developing MACE [CITE]. This is possibly due in part to the intensity thresholding requirements associated with the Agatston scoring technique, so other approaches like spatially weighted calcium scoring have been put forth as an alternative to Agatston scoring. Spatially weighted calcium scoring improves upon traditional calcium scoring by avoiding thresholding and has been shown to predict MACE more accurately [CITE], [CITE]. Spatially weighted calcium scoring is still limited in distinguishing non-zero CAC from noise. It also lacks quantitative insight as it is an arbitrary score without any direct physical association.

The Agatston technique can be used to estimate calcium volume and calcium mass. The calcium mass score acquired via the Agaston technique is fundamentally similar to traditional Agatston calcium scoring and suffers from many of the same limitations inherent to the Agatston scoring approach [CITE], [CITE]. A previous study also showed that calcium mass quantification via the Agatston approach produces up to 50% underestimation of calcium mass for large patients [CITE].

A CAC measurement technique that improves the accuracy, reproducibility, sensitivity, and specificity of the measured calcium mass would help to improve coronary heart disease diagnosis and stratify risk more accurately, especially for undetectable and nearly undetectable calcium levels. A more quantitative calcium mass calculation would likewise help improve risk stratification and diagnosis of coronary heart disease. The integrated intensity or Hounsfield unit (HU) technique recently demonstrated the ability to accurately assess coronary artery cross-sectional area, even past the visible threshold [CITE]. Similarly, volume fraction mass quantification has demonstrated the ability to accurately measure the mass of lung tissue in pulmonary computed tomography (CT) scans. These techniques have not been adapted for calcium mass quantification and CAC scoring.

This study adapted the integrated HU and volume fraction mass quantification techniques to quantify coronary artery calcium in CT scans by accounting for the partial volume effect. These new calcium mass quantification techniques, integrated calcium mass, and volume fraction calcium mass were evaluated on simulated CT scans where the total calcium mass was known. These techniques were compared to the known calcium mass to determine which technique was the most robust and accurate compared to the gold standard Agatston scoring technique and the recently introduced spatially weighted calcium scoring technique.

Throughout this study, Agatston scoring refers to the calcium mass score derived from the Agatston approach. Spatially weighted calcium scoring refers only to the calcium score without associated physical units. Integrated calcium mass and volume fraction calcium mass refer only to the calcium mass calculated via the integrated HU approach and volume fraction mass quantification technique, respectively. This study's calcium scoring algorithms are publicly available at https://github.com/Dale-Black/CalciumScoring.jl [13].

2. Method

2.1 - Simulation

The simulation study was set to match the scanning parameters of the 320-slice CT scanner (Canon Aquilion One, Canon America Medical Systems, Tustin, CA) as previously reported [CITE]. The X-ray spectrum was created with an interpolating polynomial model [CITE]. The linear attenuation coefficients were created from their chemical composition [CITE]. Poisson noise was added to simulate quantum noise. The simulation did not include Compton scatter, but beam hardening was included. A 3200x2200 pixel digital phantom was designed based on an anthropomorphic thorax phantom with a size of 30x20cm (QRM-Thorax, QRM, Möhrendorf, Germany). To simulate different patient sizes, additional fat rings emulated by a mixture of 20% water and 80% lipid were added, which resulted in a medium-sized phantom of 35x25cm and a large-sized phantom of 40x30cm. There were nine calcification inserts within the thorax with different densities and sizes. Three calcification inserts of different diameters (1, 3, and 5 mm) and different hydroxyapatite (HA) densities were placed within each phantom. For the normal density study, the HA densities in the inserts were 200, 400, and 800 mg/cm3. For the low-density study, the densities were changed to 25, 50, and 100 mg/cm3. Each phantom also contained a 10 mm diameter calibration rod. All phantom sizes and density levels were scanned using 80, 100, 120, and 135 kV tube voltages. For small, medium, and large patient sizes, the exposure value was adjusted to 0.9, 2.0, and 5.4 mR, respectively, resulting in similar noise levels for different-sized phantoms.

Simulation materials and geometries are shown in Figure 1. Acquisition and reconstruction parameters for the simulated and physical phantoms are shown in Table 1. The calibration rods were all 10 mm in diameter. All calcium scoring measurements were repeated across each kV, patient size, calcium insert size, and calcium insert density.

Segmenting regions of interest (ROIs) is an important step in calcium measurement. For this study, segmentations were done automatically based on previous work by van Praagh et al. [CITE] and adapted for simulated phantoms. The automatic segmentation approach effectively segments calcium inserts based on the known geometry of the simulated phantom without requiring manual intervention. This is important because much of this study is focused on the accurate measurement of calcium mass for inserts below the threshold of visual detectability. In patients, coronary artery centerline estimates can be extracted in non-contrast CT scans by automatic methods, such as atlas-based approaches [CITE]. Recently, deep learning has shown promise in automatically segmenting cardiac anatomy and has the potential to accurately segment coronary artery centerlines in non-contrast CT scans using a supervised learning approach in patient images like the OrCaScore dataset [CITE]. The coronary artery centerlines can then automatically generate ROIs for calcium mass measurement of coronary artery calcification.
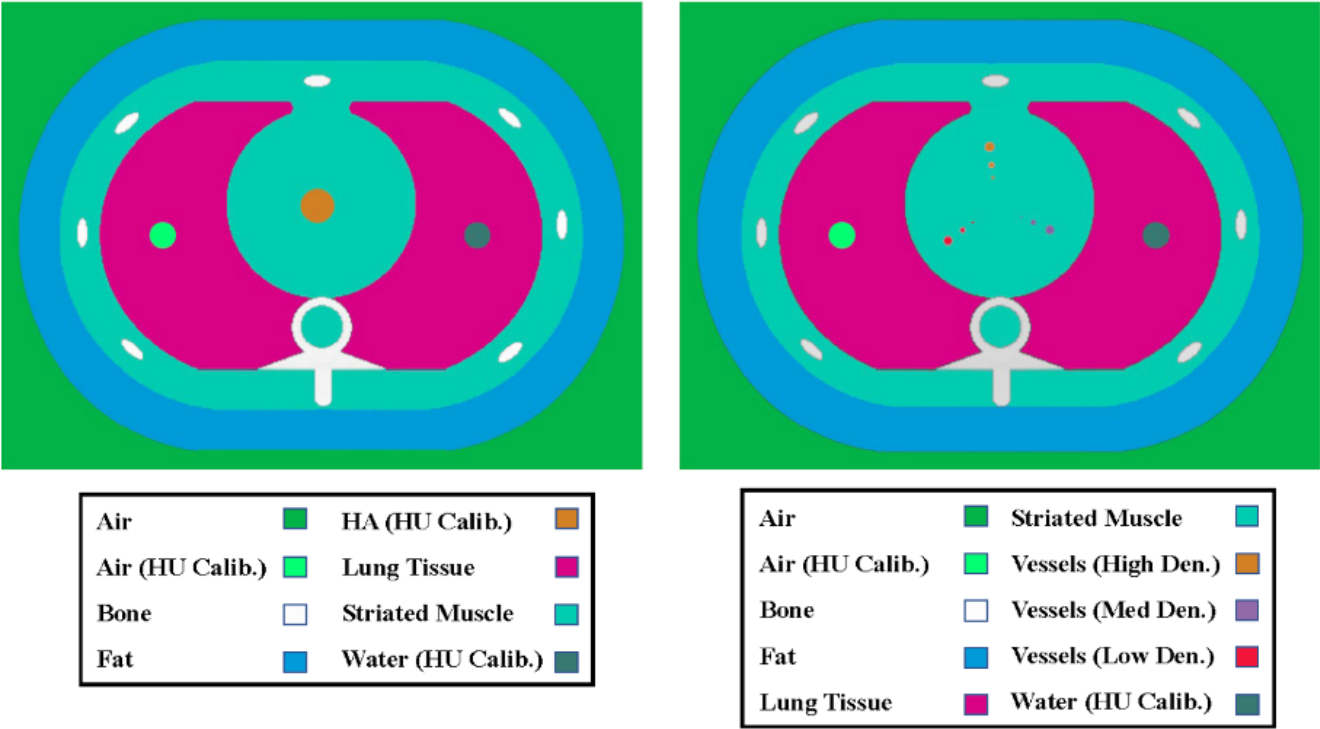
Fig. 1 Shows a sketch of the simulated phantom with the colors highlighting the different materials in the simulated phantoms.

`df_sim =`

	Parameter	Simulation	Physical
1	"Manufacturer"	"Canon"	"Canon"
2	"CT System"	"Aquilion One Vision"	"Aquilion One Vision"
3	"Reconstruction"	"FBP"	"FBP"
4	"Tube Voltage (kV)"	"80, 100, 120, 135"	"120"
5	"Exposure Small (mR)"	"0.9"	"15"
6	"Exposure Medium (mR)"	"2"	"N/A"
7	"Exposure Large (mR)"	"5.4"	"84"
8	"Slice Thickness (mm)"	"0.5"	"3.0"
9	"Matrix Size Small (pixels)"	"640 x 440"	"512 x 512"
10	"Matrix Size Medium (pixels)"	"740 x 540"	"N/A"
11	"Matrix Size Large (pixels)"	"840 x 640"	"512 x 512"
12	"Detector Element Width (mm)"	"0.5"	"UNSURE"
13	"Detector Thickness (mm)"	"0.5"	"UNSURE"

Table 1. Acquisition and reconstruction parameters for the simulated and physical phantoms.

To study the effect that motion has on Agatston scoring, integrated calcium mass, volume fraction calcium mass, and spatially weighted calcium scoring, we simulated motion using a previously reported random motion filter [CITE: http://proceedings.mlr.press/v102/shaw19a.html] and [CITE: https://www.sciencedirect.com/science/article/pii/S0169260721003102]. We then performed repeated analysis on this simulated motion data to better understand the effect that motion has on all four calcium scoring techniques.

2.2 - Physical Phantom

This study also utilized an anthropomorphic thorax phantom (QRM-Thorax, QRM, Möhrendorf, Germany) with an insert containing calcium (Cardiac Calcification Insert (CCI), QRM, Möhrendorf, Germany). All images were acquired by G van Praagh et al. [CITE], and a limited subset of scans was included in this analysis. The cardiac calcification insert phantom consisted of nine calcification inserts made up of hydroxyapatite (HA). Within the cardiac calcification insert phantom, two calibration rods consisting of water-equivalent material and 200 mgHAcm-3 were also present. The calcifications had diameters and lengths of 1.0, 3.0, and 5.0 mm. Three different densities were present in the phantom for each calcification size: 200, 400, and 800 mgHAcm-3. Two different patient sizes, small and large, were included in this analysis by the addition of a fat ring. This fat ring increased the phantom size from 300 × 200 mm to 400 × 300 mm. Four scans, two small and two large patient scans were included in this analysis. Figure 2 Shows axial slice views of various phantoms included in this study. Fig. 2A shows a simulated small phantom with normal-density inserts (200, 400, 800 mg/cm^3) and a tube voltage of 120 kV. Fig. 2B shows a simulated large phantom with low-density inserts (25, 50, 100 mg/cm^3) and a tube voltage of 120 kV with simulated motion. Fig. 2C shows an axial slice of the calibration rod (200 mg/cm^3) for a simulated medium phantom with a tube voltage of 100 kV. Fig. 2D shows an axial slice view of a physical QRM Thorax Phantom with a Cardio Calcification Insert. The red arrow shows simulated beam hardening artifact, and the blue arrows show simulated streaking artifact caused by motion.

Segmenting regions of interest is an important step in calcium measurement. For this study, segmentations were done automatically based on previous work by van Praagh et al. [CITE] and adapted for the Julia programming language [CITE].
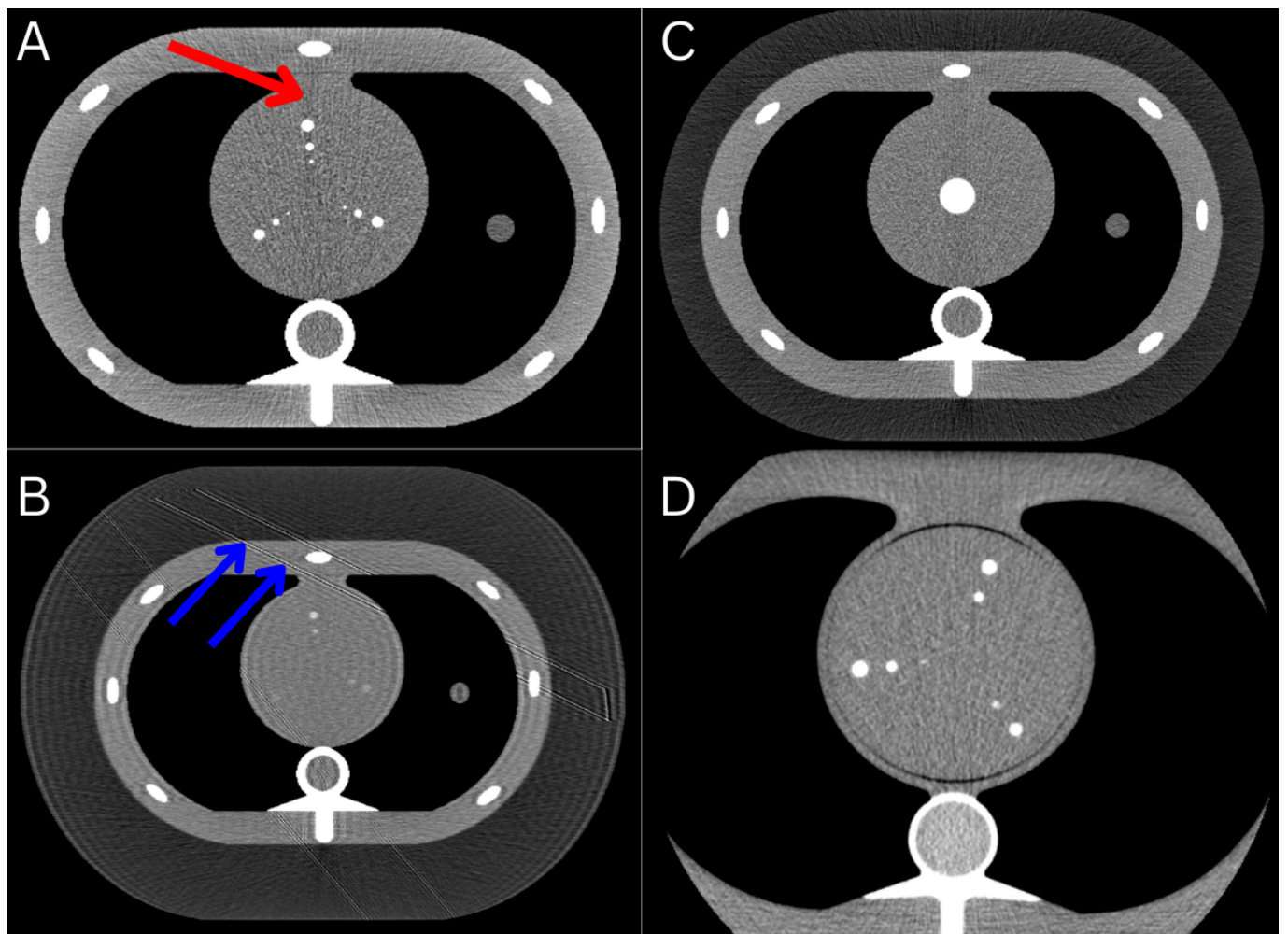
Fig. 2 Shows axial slice views of various phantoms included in this study. (A) shows a simulated small phantom with normal-density inserts (200, 400, 800 mg/cm^3) and a tube voltage of 120 kV. (B) shows a simulated large phantom with low-density inserts (25, 50, 100 mg/cm^3) and a tube voltage of 120 kV with simulated motion. (C) shows an axial slice of the calibration rod (200 mg/cm^3) for a simulated medium phantom with a tube voltage of 100 kV. (D) shows an axial slice view of a physical QRM Thorax Phantom with a Cardio Calcification Insert. The red arrow shows simulated beam hardening artifact, and the blue arrows show simulated streaking artifact caused by motion.

2.3 - Agatston Scoring

Agatston scoring is defined at a tube voltage of 120 kV [CITE], but recent papers have shown how Agatston scoring can be adjusted for use in low-dose scans (70, 80, and 100 kV) [CITE]. For this study, we assumed an exponentially decreasing trendline and extrapolated beyond to account for a higher tube voltage of 135 kV. This results in an extrapolation formula shown in Equation 1, where y_{thresh} corresponds to the extrapolated threshold (HU) and TV corresponds to the tube voltage (kV). All kV-specific thresholds used in this study are shown in Table 2.

$$y_{thresh} = (378)e^{-0.009(TV)}$$

$$(1)$$

	Tube Voltage (kV)	Threshold (HU)
df_agat_thresh =		
1	80	177
2	100	145
3	120	130
4	135	112

Table 2. Tube voltage adapted thresholds for Agatston scoring.

2.4 - Integrated Calcium Mass

As previously reported [CITE], the integrated Hounsfield technique assumes that although the partial volume effect influences the HU of a particular voxel, the total integrated HU within an ROI is conserved. This study addresses the issue of non-detectable CAC by adjusting the integrated HU technique for use in CAC scoring, calling it integrated calcium mass. The cross-sectional area equation (Eq. 2) can be adjusted for use in three-dimensional regions (Eq. 3) and applied to calcium mass quantification. Fig. 3A shows the cross-section of a simulated coronary plaque with the measurements that need to be computed for each image. S_{Obj} is the intensity (HU) of a region within the plaque containing pure calcium with no partial volume affected voxels. S_{Bkg} is the measured intensity (HU) of a ring-like section of the coronary lesion with no calcium. I is the sum of every voxel intensity (HU), including those voxels affected by the partial volume effect. V is the total number of voxels in the cross-section multiplied by the voxel size. Eq. 4 shows how to convert the volume of the object (V_{Obj}) to mass (M_{Obj}), where $\rho_{S_{Obj}}$ is the density of the calcification, specific to the object intensity (S_{Obj}).

It is impractical to accurately measure the intensity of a calcification with no partial volume effect (S_{Obj}) for small calcifications. Therefore, this study utilized a calibration rod and computed a volume based on the measured intensity of this calibration rod S_{Obj}, then converted the volume to mass by multiplying by the known density of that rod. Integrating over the entire ROI should remove any effect from noise, assuming noise only affects individual voxels.

$$I = [(A - CSA) \times S_{Bkg}] + [CSA \times S_{Obj}]$$
$$CSA = \frac{I - (A \times S_{Bkg})}{S_{Obj} - S_{Bkg}} \tag{2}$$

$$I = [(V - V_{Obj}) \times S_{Bkg}] + [V_{Obj} \times S_{Obj}]$$
$$V_{Obj} = \frac{I - (V \times S_{Bkg})}{S_{Obj} - S_{Bkg}} \tag{3}$$

$$M_{Obj} = V_{Obj} \times \rho_{S_{Obj}} \tag{4}$$

2.5 - Volume Fraction Mass

The volume fraction calcium mass technique is similar to integrated calcium mass, but instead of calculating the calcium within an entire ROI, the percent of calcium contained within one voxel is calculated. The percent calcium contained within each voxel is then summed up within a ROI to obtain the total percentage of calcium for that ROI (Eq. 5). Given the known size of the ROI and density of the calibration rod, the volume and mass of calcium can be calculated (Eq. 6).

Equation 5 shows how to calculate the percentage of calcium contained within each voxel. i is the intensity of one voxel of interest (HU), S_{Obj} is the intensity of pure background, which can be obtained from a ring-like object as seen in Fig. 3B, and S_{Obj} is the intensity of pure calcium which can be obtained from a calibration rod of any density. The result, k_i, is then the percentage of calcium within one voxel i. The entire region of voxels is then summed to give K which is the total percentage of calcium contained within the ROI.

Equation 6 shows how to calculate the volume of calcium V_{Obj}, given the total percentage of calcium within an ROI K and the known volume of that ROI V_{ROI}. This can then be converted into a calcium mass M_{Obj} given the density of the calibration rod. Fig. 3 shows the parameters needed to calculate mass via the integrated calcium mass technique (Fig. 3A) and the volume fraction calcium mass technique (Fig. 3B).

$$k_i = \frac{i - S_{Bkg}}{S_{Obj} - S_{Bkg}}$$
$$K = \sum_i k_i$$

(5)

$$V_{Obj} = K \times V_{ROI}$$
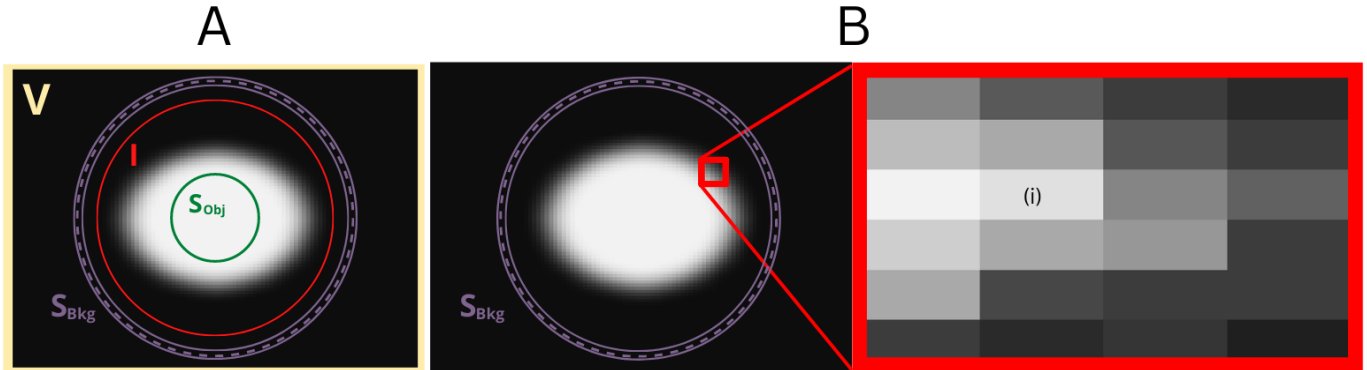$$M_{Obj} = V_{Obj} \times \rho_{S_{Obj}}$$

(6)

Fig 3. Shows two identical simulated vessel lumens with ROIs for calcium measurement.(A) Shows the ROIs needed for the integrated calcium mass technique. The central ROI that yields S_{Obj} and the background ROI that yields S_{Bkg} are unaffected by the partial volume effect, while the object ROI used to calculate I is affected by the partial volume effect. (B) Shows the ROIs needed for the volume fraction technique and a zoomed-in portion of the simulated vessel lumen, where i is the individual voxel intensity. S_{Bkg} is a ring-like background ROI unaffected by the partial volume effect. S_{Obj} is the mean intensity of known calcium density obtained from a calibration rod unaffected by the partial volume effect.

2.6 - Spatially Weighted Calcium Scoring

This study reimplemented the previously described spatially weighted calcium scoring technique [CITE] to compare integrated calcium mass and volume fraction calcium mass against a more recently proposed calcium scoring approach. The same set of voxels included in the integrated calcium mass and volume fraction calcium mass regions of interest were included in the spatially weighted calcium scoring calculations. Each voxel was assigned a weight using a normal distribution function with means and standard deviations obtained from 100 mg/cm3 calcium rod measurements across multiple simulated scans. These measurements are shown in Table 3. Only small patient sizes were used for measurements, and the signal-to-noise ratio was kept constant between different energies.

```
df_swcs_cal =
```

	kV	Mean	Std
1	80	201.49	34.304
2	100	174.366	28.171
3	120	158.264	22.966
4	135	152.881	21.078

Table 3. Means and standard deviations of the 100 mg/cm^3 calcium calibration rods.

2.7 - Statistical Analysis

All calcium scoring calculations and analyses were performed using the programming language Julia [CITE]. Root-mean-square error (RMSE) and root-mean-square deviation (RMSD) were calculated for all linear regression measurements to test for accuracy (RMSE) and precision (RMSD). Equation 7 shows how to calculate RMSE and RMSD. N is the total number of data points, $\hat{y}$ is the calculated calcium masses, and y is either the ground truth calcium masses (RMSE) or the linear regression-based calcium masses (RMSD), which is computed based on the calculated calcium masses.

$$RMS = \sqrt{\frac{\sum |y - \hat{y}|^2}{N}} \tag{7}$$

3. Results

3.1 - Simulated Phantom

3.1.1 - Accuracy

Agatson scoring, integrated calcium mass, and volume fraction calcium mass can calculate mass directly, making the comparison between known mass and those techniques straightforward. The spatially weighted approach produces an arbitrary score correlated with the Agatston score [CITE]. A mass calculation derived from spatially weighted calcium scoring is not practical due to the lack of thresholding in the spatially weighted technique.

Linear regression was performed for integrated calcium mass, volume fraction calcium mass, and Agatston scoring against known calcium mass on all stationary and motion-affected simulated phantoms. The low-density and normal-density phantoms were examined separately.

The correlation coefficient (r), RMSE, and RMSD values were calculated for Agatston scoring, integrated calcium mass, and volume fraction calcium mass. Linear regression used the known calcium mass as the reference. Integrated calcium mass on normal-density phantoms produced an r-correlation coefficient, RMSE, and RMSD value of 1.000, 0.677 mg, and 0.602 mg, respectively. Volume fraction calcium mass on normal-density phantoms produced an r-correlation coefficient, RMSE, and RMSD value of 1.000, 0.609 mg, and 0.599 mg, respectively. Agatston scoring on normal-density phantoms produced an r-correlation coefficient, RMSE, and RMSD value of 1.000, 1.650 mg, and 0.906 mg, respectively. All normal-density accuracy measurements are shown in Table 4.

Integrated calcium mass on low-density phantoms (Fig. 4A) produced an r-correlation coefficient, RMSE, and RMSD value of 0.991, 0.495 mg, and 0.494 mg, respectively. Volume fraction calcium mass on low-density phantoms (Fig. 4B) produced an r-correlation coefficient, RMSE, and RMSD value of 0.989, 0.585 mg, and 0.575 mg, respectively. Agatston scoring on low-density phantoms (Fig. 4C) produced an r-correlation coefficient, RMSE, and RMSD value of 0.789, 3.509 mg, and 2.240 mg, respectively.

A similar analysis was performed for all motion-affected phantoms, and the results can be seen in Table 4 The trend continued to show that integrated calcium mass and volume fraction calcium mass outperformed Agatston scoring for both low-density and normal-density phantoms.
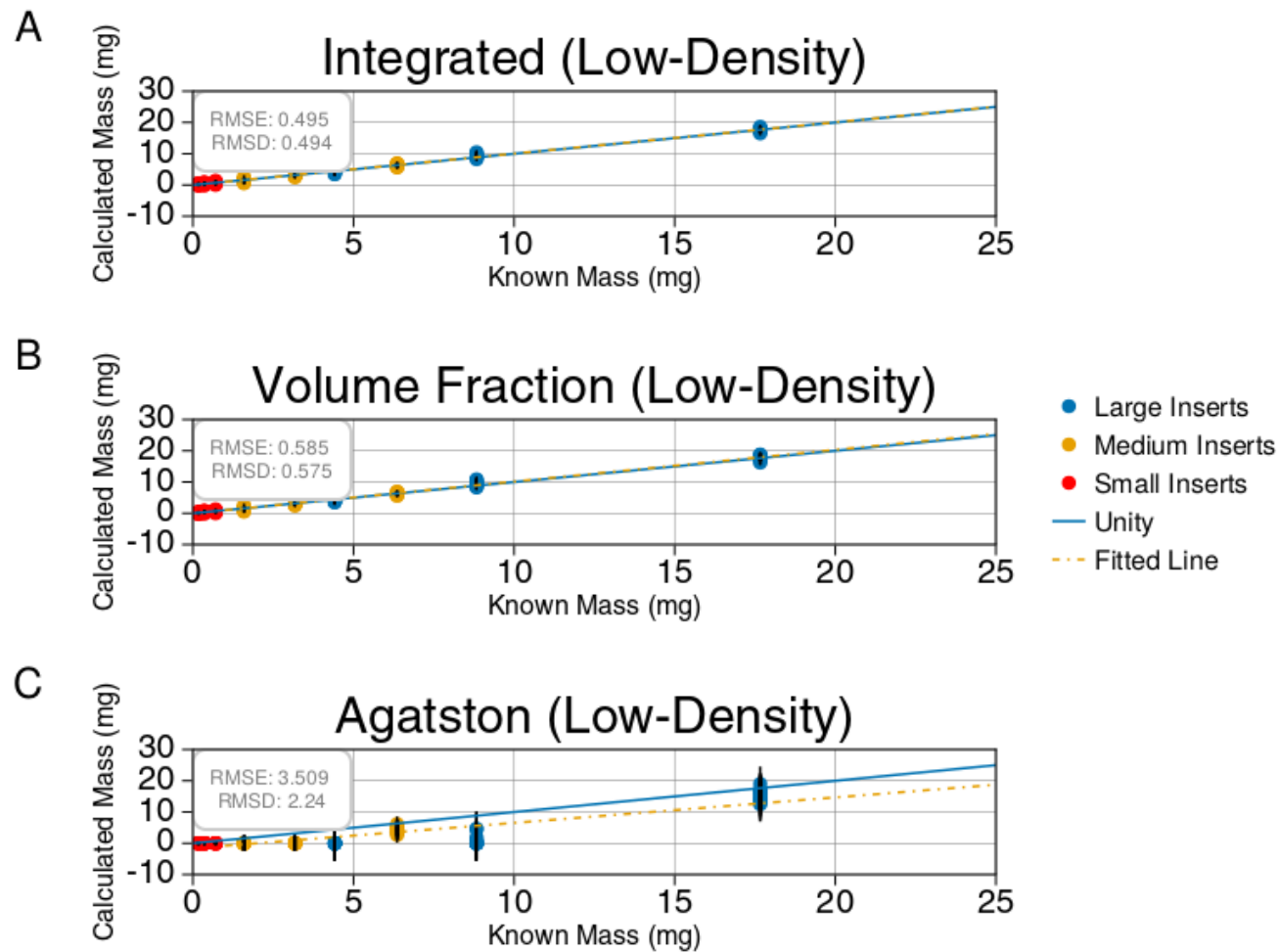
Fig. 4 Shows the linear regression analysis comparing measured calcium to the known calcium for the low-density (25, 50, 100 mg/ cm3) stationary phantoms. Every tube voltage (80, 100, 120, 135 kV) and size (small, medium, large) is included in the analysis. (A) shows the results of integrated calcium mass. (B) shows the results of the volume fraction method. (C) shows the results of Agatston mass scoring. The best fit line, along with the root mean squared error and root mean squared deviation values are shown in each plot.

3.1.2 – Reproducibility

This study included another set of simulated images with identical geometry to understand how repeatable the four different calcium scoring techniques are. The only variation between the two groups of images comes from the random quantum noise associated with the simulation itself. First, all false-negative values were removed from the analysis and then the reproducibility of the results was calculated. Then the results for the stationary phantoms were plotted, comparing measurements from the first set of images against the second set of images for all four scoring techniques (Fig. 5). The RMSE and RMSD values for integrated calcium mass were 0.708 mg and 0.692 mg, respectively (Fig. 5A). The RMSE and RMSD values for volume fraction calcium mass were 0.783 mg and 0.776 mg, respectively (Fig. 5B). The RMSE and RMSD values for Agatston scoring were 0.982 mg and 0.980 mg, respectively (Fig. 5C). The RMSE and RMSD values for spatially weighted calcium scoring were 4.722 mg and 4.714 mg, respectively (Fig. 5D).

The correlation coefficient and best-fit line were also calulated for each calcium scoring technique and are shown in Table 4. The results were then repeated for the motion-affected simulated phantom data. Again, similar trends were seen in reproducibility analysis for both the stationary scans and motion-affected scans. These results are shown in Table 4.
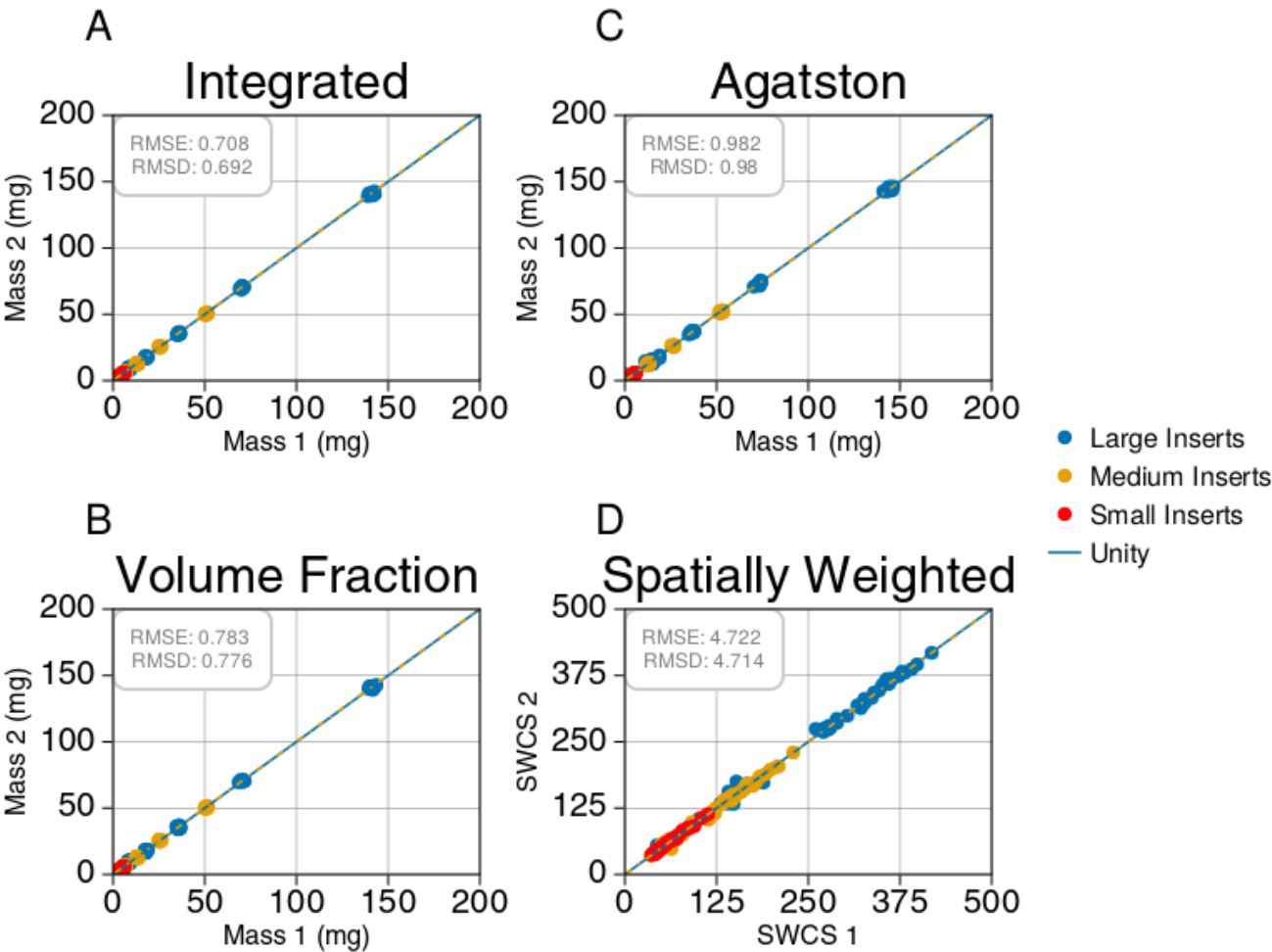
Fig. 5 Shows reproducibility measurements for all four scoring techniques on the stationary phantoms. Every tube voltage (80, 100, 120, 135 kV), size (small, medium, large), and density (low, normal) is included in the analysis. The measurements from the first set of images were plotted against the second set of images for each technique. Integrated mass (A), Volume fraction (B), Agatston mass scoring (C), and spatially weighted calcium scoring (D) are shown along with the root mean squared error and root mean squared deviation values.

3.1.3 - Sensitivity and Specificity

The percentage of false-negative (CAC=0) and false-positive (CAC>0) scores was also calculated to understand the sensitivity and specificity of all four calcium scoring techniques. Any region containing known calcium that resulted in a CAC score of zero was determined to be a false-negative (CAC=0) score, and any region of pure background that resulted in a positive calcium score was determined to be a false-positive (CAC>0) score. This was simple to obtain for Agatston scoring, as Agatston returns zero as a possible score.

For spatially weighted calcium scoring, the score is always greater than zero, making a score of zero impossible. This is unhelpful since noise will eventually become a dominating factor for low-density and small calcifications making it difficult to distinguish between a small amount of calcium and noise, so spatially weighted calcium scores for regions of pure background were calculated across all simulated phantoms, and the mean and standard deviation of these scores were calculated. Any spatially weighted calcium score containing regions of known calcium that resulted in a score less than the mean background score plus one standard deviation was then considered a false-negative score. Likewise, any spatially weighted calcium score greater than the mean score of pure background plus one standard deviation was considered a false-positive (CAC>0) score for regions containing only background.

A similar approach for determining false-negative (CAC=0) and false-positive (CAC>0) scores was applied to integrated calcium mass and volume fraction calcium mass. Both calcium mass techniques result in real masses from $(-\infty, \infty)$. For low-density and small calcifications, the noise will dominate the calcium mass results making it difficult to distinguish between a small amount of calcium and noise, so a mean and standard deviation mass was obtained in the same way as spatially weighted calcium scoring. Any integrated calcium mass or volume fraction calcium mass result containing regions of known calcium that resulted in a mass less than the mean background mass plus one standard deviation was then considered a false-negative score. Likewise, integrated calcium mass or volume fraction calcium mass result greater than the mean mass of pure background plus one standard deviation was considered a false-positive (CAC>0) mass for regions containing only background.

The percentage of false-negative (CAC=0) and false-positive (CAC>0) scores were then calculated for all four techniques. For the stationary simulated phantoms, integrated calcium mass and volume fraction calcium mass techniques produced the fewest false-negative (CAC=0) results (21 out of 216 scores), and spatially weighted calcium scoring produced the most (92 out of 216 scores). While Agatson scoring produced a total of 84 false-negative (CAC=0) out of 216 scores. The integrated calcium mass and volume fraction calcium mass techniques produced 8 false-positive (CAC>0) results out of 72 scores. Spatially weighted calcium scoring produced 12 false-positive (CAC>0) out of 72 total scores, and Agatston scoring produced zero false-positive scores (CAC>0). These results are shown in Fig. 6. Results for the sensitivity and specificity of each technique, evaluated on the motion-affected phantoms, are shown in Table 4.

False-Negative Scores (CAC=0)
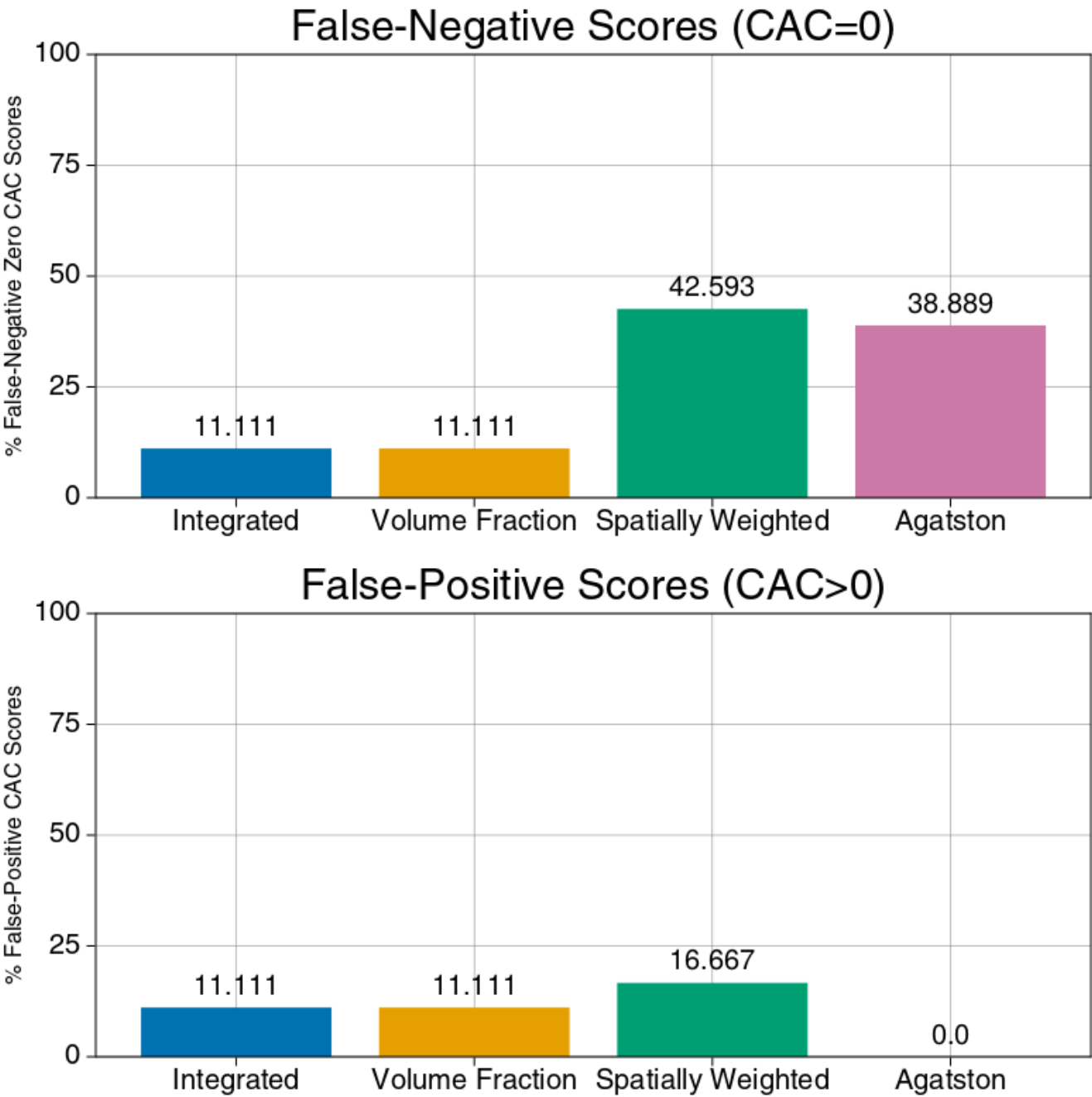

False-Positive Scores (CAC>0)

Fig 6. Shows the percentage of false-negative (CAC=0) and false-positive (CAC>0) scores on the stationary phantoms. Every tube voltage (80, 100, 120, 135 kV), size (small, medium, large), and density (low, normal) is included in the analysis.

3.2 - Physical Phantom

The physical phantom scans acquired by Praagh et al. [CITE] were also used for analysis. Two small and two large phantom scans acquired by the CANON scanner (Table 1) were included in this analysis to validate the simulation results. Spatially weighted calcium scoring requires a collection of 100 mg/cm^3 calibration rod measurements for the weighting function. The physical QRM phantom scans acquired by Praagh et al. [CITE] only contained a 200 mg/cm^3 calibration rod. Therefore, spatially weighted calcium scoring was excluded from the analysis of the physical phantom.

3.2.1 - Accuracy

Linear regression was performed for integrated calcium mass, volume fraction calcium mass, and Agatston scoring against known calcium mass on all four physical phantom scans. Only normal-density (200, 400, 800 mg/cm^3) is included in the physical QRM phantom.

The correlation coefficient (r), RMSE, and RMSD values were calculated for Agatston scoring, integrated calcium mass, and volume fraction calcium mass. Linear regression used the known calcium mass as the reference. Integrated calcium mass on the normal-density physical phantom (Fig. 7A) produced an r-correlation coefficient, RMSE, and RMSD value of 0.981, 8.811 mg, and 5.752 mg, respectively. Volume fraction calcium mass on the normal-density physical phantom (Fig. 7B) produced an r-correlation coefficient, RMSE, and RMSD value of 0.975, 8.559 mg, and 6.647 mg, respectively. Agatston scoring on the normal-density physical phantom (Fig. 7C) produced an r-correlation coefficient, RMSE, and RMSD value of 0.973, 22.485 mg, and 4.653 mg, respectively.

All normal-density physical phantom accuracy measurements, including the best-fit line and the r-correlation coefficient, are shown in Table 4.
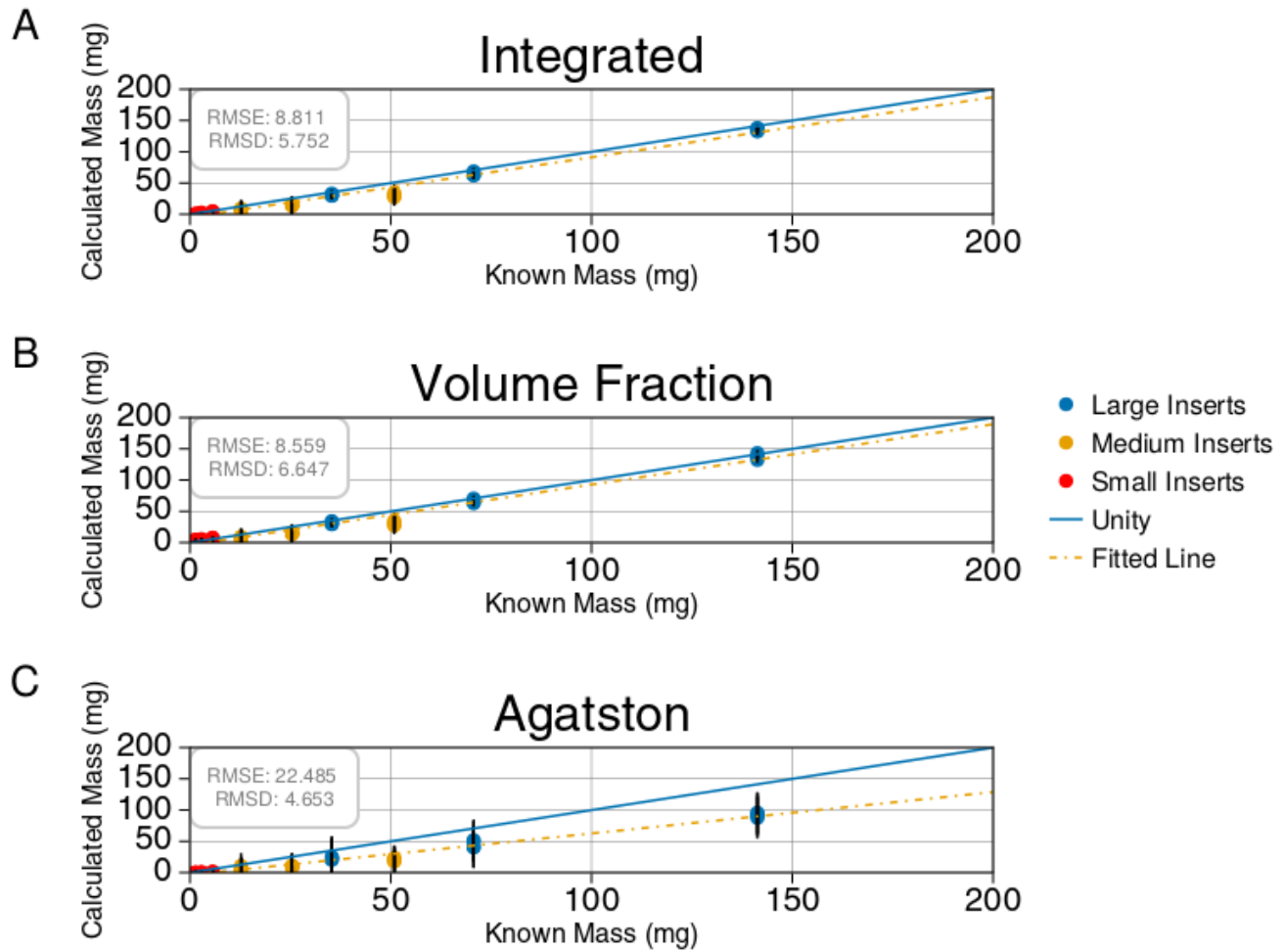
Fig. 7 Shows the linear regression analysis comparing measured calcium to the known calcium for the normal density (200, 400, 800 mg/ cm3) physical phantom scans at a tube voltage of 120 kV. (A) shows the results of integrated calcium mass. (B)shows the results of the volume fraction method. (C) shows the results of Agatston mass scoring. The best fit line, along with the root mean squared error and root mean squared deviation values, are shown in each plot.

3.2.2 - Reproducibility

To analyze the reproducibility of integrated calcium mass, volume fraction calcium mass, and Agatston scoring on the physical phantom, the first small phantom, and large phantom scans were compared to the second small phantom and large phantom scans. These scans were acquired under identical settings, with the only variation coming from a rotation of the phantom. All false-negative values were removed from the analysis, and then the reproducibility of each technique was calculated.

The RMSE and RMSD values for integrated calcium mass were 2.838 mg and 1.906 mg, respectively (Fig. 8A). The RMSE and RMSD values for volume fraction calcium mass were 3.203 mg and 1.777 mg, respectively (Fig. 8B). The RMSE and RMSD values for Agatston scoring were 2.144 mg and 1.792 mg, respectively (Fig. 8C).

The up to 50% underestimation for the Agatston scoring masses (Fig. 7C) means the calculated masses are, on average, smaller than the calculated masses from the integrated and volume fraction calcium mass techniques. Since RMS values are sensitive to the magnitude of the inputs, the smaller RMSE and RMSD values for Agatston scoring, relative to integrated calcium mass and volume fraction calcium mass techniques, are explained by the underestimated mass calculations of the Agatston technique.

The correlation coefficient and best-fit line were also calculated for each calcium scoring technique and are shown in Table 4.
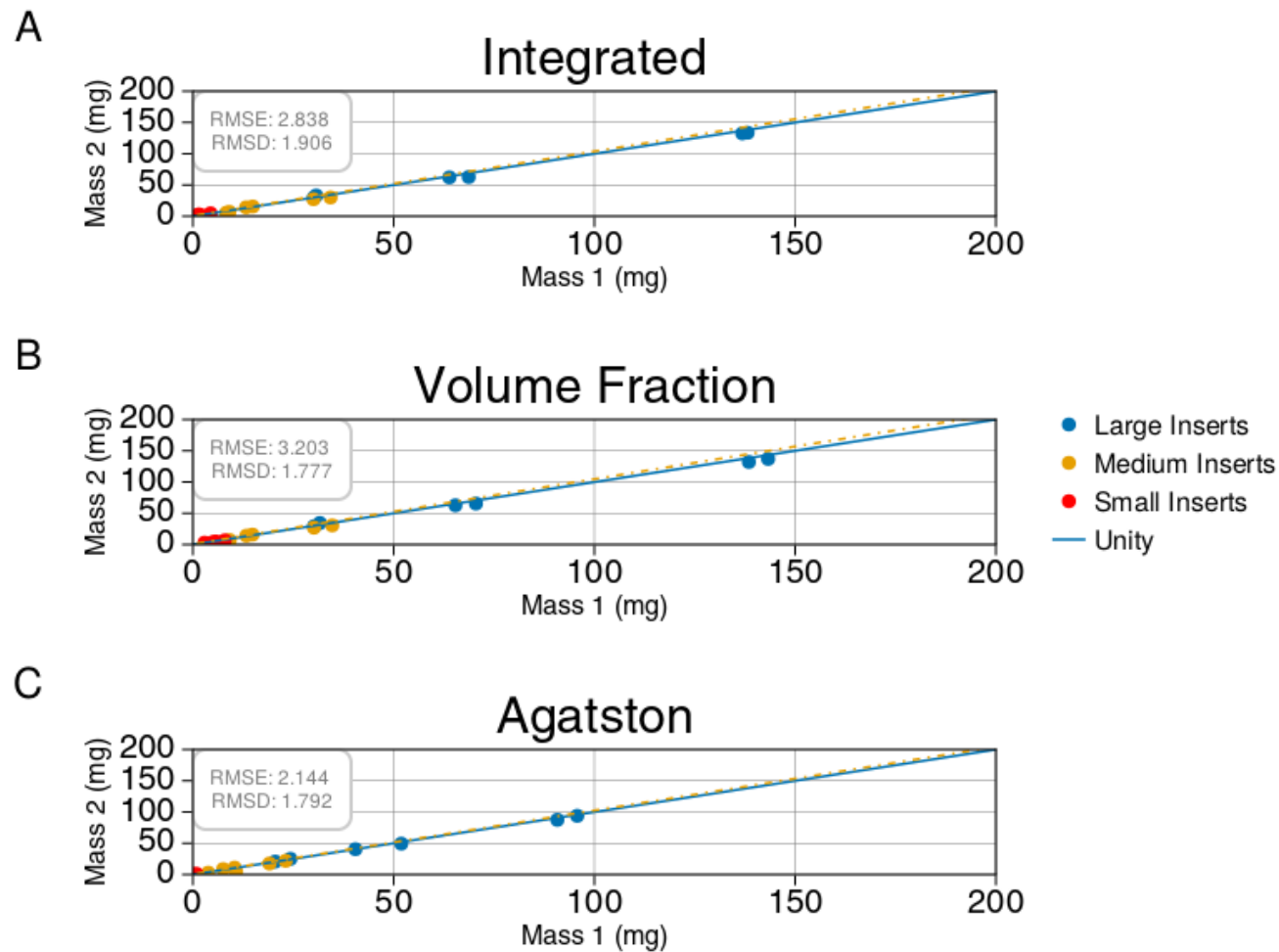
Fig. 8 Shows reproducibility measurements on the physical phantom scans at a tube voltage of 120 kV. Measurements from two scans were plotted against a second set of measurements from two different scans at different angles. Integrated mass (A), Volume fraction (B), Agatston mass scoring (C) are shown along with the root mean squared error and root mean squared deviation values.

3.2.3 - Sensitivity and Specificity

The approach used on the simulated phantoms that determined the cutoff value for false-negative (CAC=0) and false-positive (CAC>0) scores, based on the mean and standard deviation of pure background, was also utilized for physical phantom sensitivity and specificity analysis.

For the physical phantom scans, integrated calcium mass and volume fraction calcium mass techniques produced six false-negative (CAC=0) scores out of 36 total scores. While Agatson scoring produced a total of 7 false-negative (CAC=0) out of 216 scores.

The integrated calcium mass and volume fraction calcium mass techniques produced one false-positive (CAC>0) result out of 12 scores. Agatston scoring produced zero false-positive scores (CAC>0). These results are shown in Fig. 9. and are summarized in Table...

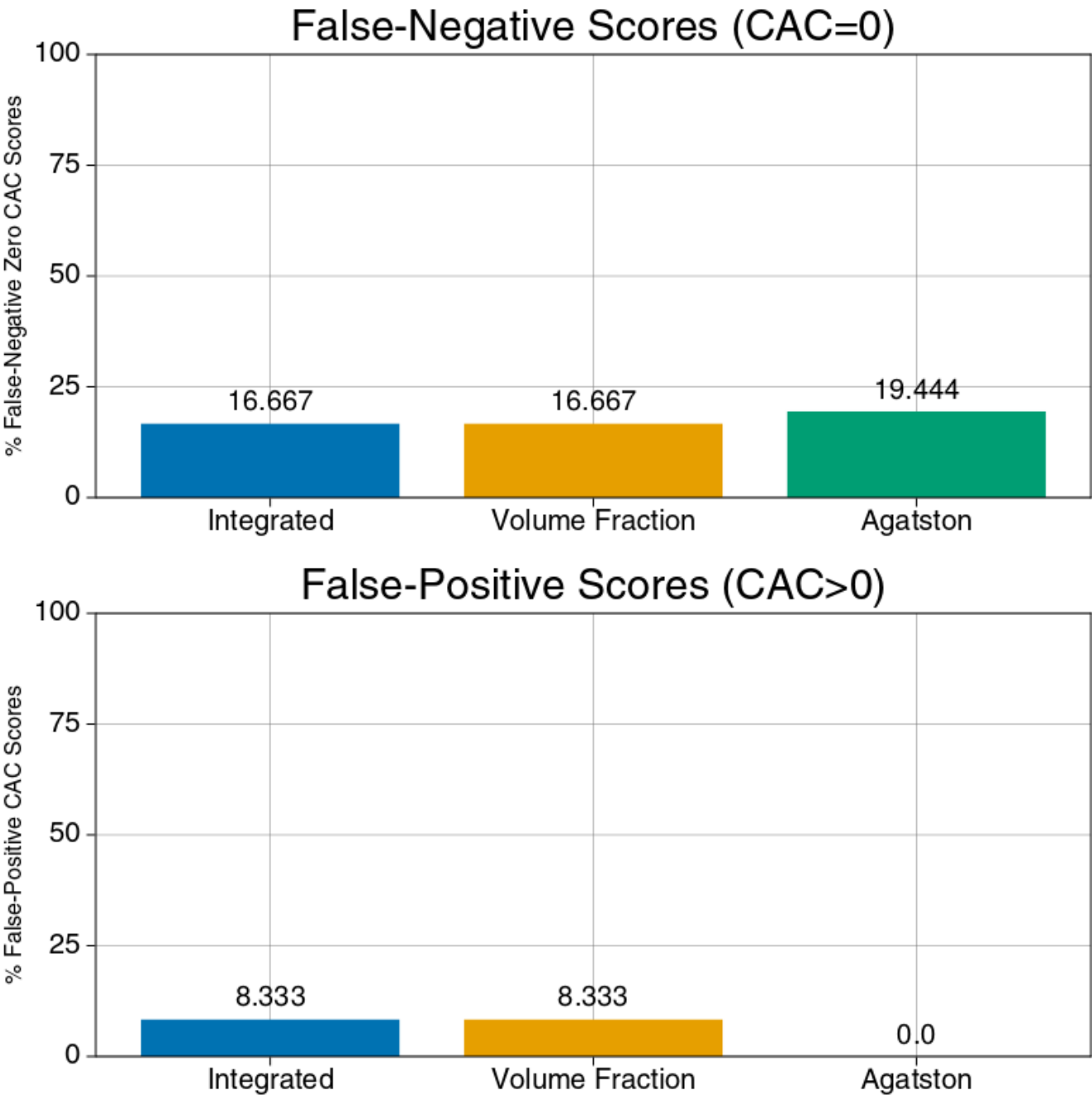## False-Negative Scores (CAC=0)

False-Positive Scores (CAC>0)

Fig 9. Shows the percentage of false-negative (CAC=0) and false-positive (CAC>0) scores for the physical phantom scans at a tube voltage of 120 kV.

	Phantom	Results	Metrics	Integrated
1	"Stationary"	"Accuracy (Normal Density)"	"r^2"	"1.0"
2	"Stationary"	"Accuracy (Normal Density)"	"RMSE"	"0.677"
3	"Stationary"	"Accuracy (Normal Density)"	"RMSD"	"0.602"
4	"Stationary"	"Accuracy (Normal Density)"	"Best Fit Line"	"y = 0.993x + 0.165"
5	"Stationary"	"Accuracy (Low Density)"	"r^2"	"0.991"
6	"Stationary"	"Accuracy (Low Density)"	"RMSE"	"0.495"
7	"Stationary"	"Accuracy (Low Density)"	"RMSD"	"0.494"
8	"Stationary"	"Accuracy (Low Density)"	"Best Fit Line"	"y = 1.002x + 0.016"
9	"Stationary"	"Reproducibility"	"r^2"	"1.0"
10	"Stationary"	"Reproducibility"	"RMSE"	"0.708"
	more			
38	"Physical"	"Sensitivity"	"False Positives"	"1 / 12"

Table 4. Summary of all the accuracy, reproducibily, sensitivity, and specificity measurements across every type of phantom (stationary, motion-affected, and physical phantom)

4. Discussion

Calcium mass was measured for different patient sizes, kVps, calcium sizes, and calcium densities, with and without motion on simulated phantoms, and validated in a subset of physical phantom scans acquired by Praagh et al. and shared with us for analysis. The calculated calcium mass was compared against the known mass of the calcium inserts.

The results indicate that integrated calcium mass and volume fraction calcium mass are more accurate, reproducible, sensitive, and specific than Agatston scoring (Table 4).

Agatston scoring has commonly been used in the past for predicting patient outcomes. However, a limitation of Agatston scoring is that it's only defined at 120 kVp, and a threshold of 130 HU is commonly used for calcium detection. However, the calcium attenuation coefficient is energy dependent, which makes scoring challenging when images are acquired at lower kVps, to reduce patient radiation dose. Recent reports have introduced correction factors for calcium measurements at lower kVps [CITE]. Another limitation of a thresholding approach for calcium measurement is that it is affected by partial volume effect and motion. We have introduced two new methods for calcium mass quantifications based on integrated intensity (HU) and volume fraction that address the above limitations.

Previous studies have shown that up to 5% of patients with suspected zero CAC (CAC=0) will experience MACE, despite no detectable calcium by Agatston scoring [CITE]. One question arises as to whether these patients had calcium that is not detectible by traditional Agatston scoring or simply no calcium. Integrated calcium mass, volume fraction calcium mass, and spatially weighted calcium scoring attempt to address this concern by removing the intensity thresholding requirements of standard Agatston scoring. This study shows that integrated calcium mass and volume fraction calcium mass are more sensitive to low-density calcifications than spatially weighted calcium scoring and Agatston scoring. The results showed that the percentage of false-negative (CAC=0) scores on the stationary simulated phantom were 11.111, 11.111, 42.593, and 38.889 for integrated calcium mass, volume fraction calcium mass, spatially weighted calcium scoring, and Agatston scoring, respectively (Figure 6). The substantial reduction in false-negative zero calcium scores for the integrated calcium mass and volume fraction calcium mass techniques compared to the existing techniques will help address the current limitation for patients with false-negative (CAC=0) scores.

The integrated calcium mass and volume fraction calcium mass techniques can detect calcifications that are currently indetectable by the Agatston scoring approach due to its thresholding requirement. Furthermore, a previous study has shown that calcium volume was positively and independently associated with major adverse cardiac event risk, and calcium density was inversely associated with major adverse cardiac event risk [CITE]. Another study has shown that calcium density score was the strongest positive independent predictor of major adverse cardiac events, compared to Agatston score, mass score, and volume score [CITE]. Disagreements between these studies are possibly related to the thresholding approach of Agatston scoring and poor reproducibility of Agatston scoring, which is also a limitation of all the traditional calcium (mass, volume, and density) scoring approaches based on the

Agatston technique. Integrated calcium mass and volume fraction calcium mass provide more accurate, reproducible, and quantitative approaches to calcium measurement. Future studies on patient data comparing Agatston scoring with integrated calcium mass and volume fraction calcium mass might help explain these seemingly contradictory results better

When acquiring the mass of calcium in the QRM phantom using Agatston scoring, it has been shown the mass scores can be significantly different from the known physical mass [CITE]. Our results (Figures 4 and 7) show that integrated calcium mass and volume fraction calcium mass measure calcium mass more accurately than Agatston scoring, on both simulated and physical phantoms. We also see an improvement in reproducibility when comparing integrated calcium mass and volume fraction calcium mass against Agatston scoring and spatially weighted calcium scoring (Fig. 5).

Although many parameters were considered in the simulation, only one type of reconstruction was utilized, which prevented us from addressing the effect that reconstruction type has on these calcium scoring techniques. Previous studies have shown good agreement between Agatston scores based on filtered back-projection, hybrid iterative reconstruction, and deep learning-based reconstruction [CITE].

Slice thickness plays an important role in calcium scoring, and traditional Agatston scoring is only defined at a slice thickness of 3 mm. Recent studies have shown that the accuracy and sensitivity of Agatston scoring are improved when slice thickness is decreased [CITE]. Our simulation was limited to 0.5 mm slice thickness which is expected to provide more accurate and sensitive comparisons for Agatston scoring. Nonetheless, future studies might provide insights by varying the slice thickness. This study was also limited by the lack of realistic cardiac hardware, such as stents, that are known to cause blooming or motion artifacts [CITE]. Future studies should account for this by including physical phantoms with realistic cardiac hardware or a robust patient data set.

Fewer total scans were included in the physical phantom analysis compared to the simulated analysis, resulting in fewer calcium scores. This is a limitation and becomes more pronounced in the sensitivity and specificity analysis, where false-negative (CAC=0) and false-positive (CAC>0) results are rare, to begin with. In addition, motion was not included in the physical phantom analysis. Likewise, the simulated motion is based on motion in magnetic resonance imaging. More physical phantom scans, including scans affected by motion, need to be included in future studies. This would provide more robust accuracy, reproducibility, sensitivity, and specificity measurements and would result in a more reliable expected percentage of false-negative (CAC=0) and false-positive (CAC>0) scores for each calcium quantification technique.

Previous studies show that Agatston scoring consistently underestimates calcium density and volume, with even further underestimation for low-density and motion-affected plaques [CITE]. Werf et el. indicates that low-density calcifications might fall below the 130 HU threshold because of blurring from motion which artificially reduces the Agatston score [CITE]. Our study is consistent with these

results (Fig. 4C and 7C); we showed that Agatston scoring produced the most false-negative (CAC=0) classifications for the simulated data and the physical phantom scans. Future studies are warranted in physical phantoms with lower-density calcification inserts (< 200 mg/cm3) to understand how integrated calcium mass and volume fraction calcium mass compares to Agatston mass within the low-density regime on physical data. Very high coronary artery calcium density (> 1000 mg/cm3) is quite rare, and Agatston scoring has already been shown to be a good predictor of cardiovascular disease within this subset of patients [CITE].

Tzolos et al. showed that Agatston scoring struggles to detect small calcifications in the coronary arteries of patients due to the threshold requirement of 130 HU and the minimum connected component requirement of 1 mm [33]. Our results are consistent with this study and indicate that the size of the calcium insert is a critical variable in accounting for false-negative (CAC=0) Agatston scores. The small inserts resulted in 40 false-negative (CAC=0) scores out of 216 total scores, whereas the medium and large inserts accounted for only 24 and 20 false-negative (CAC=0) scores, respectively. Density was also an important factor. Low-density calcifications resulted in 40 false-negative (CAC=0) scores out of 216 total scores, whereas the medium-density and high-density calcifications resulted in 32 and 12 false-negative (CAC=0) scores, respectively. Based on our results, integrated calcium mass and volume fraction calcium mass improved upon many of the issues associated with Agatston scoring and resulted in fewer false-negative (CAC=0) scores.

5. Conclusion

This study shows that both the integrated calcium mass and volume fraction calcium mass techniques improve sensitivity, accuracy, and reproducibility as compared to traditional Agatston scoring. This improvement in calcium scoring can potentially improve risk stratification for patients undergoing calcium scoring and further improve outcomes compared to Agatston scoring.

Acknowledgements

References