Correlation between Measured and Visual Scoring of Coronary Artery Calcification

Jubal R. Watts Jr. MD1,*, Daniel J. Dauer MD1, Naomi S. Fineberg PhD2, Matthew Budoff MD3, HrudayaNath MD1

1Department of Cardiopulmonary Radiology, University of Alabama Hospitals. JT N370, 619 19th Street South Birmingham, AL 35249-6830, United States
2Department of Biostatistics, UAB School of Public Health.RPHB 343B, 1530 3rd Avenue South Birmingham, AL 35294-0022, United States
3Department of Cardiology, Harbor-UCLA LA BioMed CT Reading Center. 1000 West Carson Street, Box 405, Torrance, CA 90509, United States

*Corresponding Author: jrwatts@uabmc.edu

Abstract  Coronary artery calcification (CAC) is a well-known marker of subclinical coronary atherosclerosis, which is always detectable in non ECG-gated routine chest CT examinations, and its visual estimation is correlated to clinical outcomes. Agatston scoring is not routinely performed on these examinations. We sought to validate a visual scoring scheme we derived against ECG-gated CT’s and compare our system with another previously published visual scoring scheme in a different cohort of lung cancer screening participants. 50 COPDGene participants received, regular dose full inspiration (non-gated high mA), and low dose expiration CT (non-gated low mA) and ECG-gated CT’s at the same time. CAC was visually scored by 3 readers using our total visual scoring (TVS) method and compared to the Agatston score. The second portion of the study involved visual and Agatston scoring of a larger sample of 198 lung cancer screening patients, comparing visual scoring described by Shemesh et. al. and our TVS method. For the COPDGene participants, scores were highly correlated among readers (all ICC≥0.92), between the ECG-gated CT, non-gated low mA CT, and the non-gated low dose CT (all p<0.001), and with the Agatston score (all ICC≥0.90). For the cancer screening cohort there was very good agreement of our system and Shemesh scores. Correlations between reader scores, our system, Shemesh scores, and Agatston scores were also very good, ranging from 0.81-0.96. We derived cutoff values corresponding to Agatston risk quartiles for our system and Shemesh. There was excellent correlation of visual scoring with Agatston scoring on ECG-gated and non-gated CT. In lung cancer screening CT’s both ours and Shemesh visual scoring correlated well with Agatston scores and with each other. Visual scoring may predict clinically significant CAC in major Agatston categories.

Keywords  Visual Calcium Scoring, Agatston Scores, Coronary Calcium, NLST, COPDGene

1. Introduction

Atherosclerosis is the leading cause of death in the developed nations[1-3]. Coronary artery calcification (CAC) quantified using electrocardiographically (ECG) gated computed tomography (CT) is well studied as a marker of subclinical coronary artery disease (CAD). A recent publication demonstrated that Agatston score > 300 is associated with significantly higher incidence of adverse cardiac events [4]. CAC is always detectable in noncontrast chest CT examinations, performed for any reason. Visual scoring of CAC on non ECG gated low dose CT scans (LDCT) of the chest performed for lung cancer screening has been demonstrated to be predictive of death from cardiovascular disease in a recent publication[5]. A more recent publication has also demonstrated Agatston scoring of CAC on LDCT for lung cancer screening is predictive for both all cause mortality and cardiovascular events[6]. This work, by Jacobs et. al., also demonstrated a hazard ratio for all-cause mortality of greater than 6 for CACAgatston scores greater than 100 when compared with a CAC score of zero[6]. With the results of the National Lung Cancer Screening Trial (NLST) demonstrating a lung cancer specific and all cause mortality benefit to low dose lung cancer screening in individuals at high risk for lung cancer[7], and more recent recommendation by the US Disease Prevention Task Force’s recommendation to use LDCT screening for lung cancer [8], the use of lung cancer screening LDCT will likely increase. Despite promising advances in lung cancer diagnosis and therapy, we know that more smokers will die from cardiovascular disease than will survive or succumb to lung cancer[3]. Agatston scoring performed on low dose non-gated CT for lung cancer screening correlates well with
the assessment of the presence or absence of CAC and appropriately categorizes people into major Agatston score ranks[9, 10]. This information may be valuable to referring clinicians allowing more appropriate cardiovascular risk stratification as Agatston scores have been shown to add independent and incremental benefits to cardiovascular event prediction above that of Framingham risk scores[11]. Unfortunately, Agatston scoring is not routinely performed on low dose lung cancer screening CT’s, or other CT examinations of the thorax as it usually requires postprocessing on an independent workstation with calcium scoring software, not routinely available on most diagnostic picture archiving and communication system (PACS). Though prior research has shown that visual scoring of CAC can be predictive of mortality, only one study we are aware of has correlated visual scoring of CAC with Agatston scores, and that was on conventional regular dose CT exams[12]. We sought to develop an easy semi-quantitative visual scoring system, total visual scoring(TVS), to correlate with measured Agatston scores from ECG gated CT, non ECG gated regular dose and non gated LDCT. In this study we wanted to initially validate our TVS scheme against ECG gated calcium scoring CT’s available through the COPDGene trial as a ‘proof of concept’, and then to test our system in a larger population of non gated LDCT exams in NLST, comparing it to the method described by Shemesh et al in a similar population[5] and correlated both of these methods with the measured Agatston score. The Shemesh method evaluated only the linear extent of calcium in a vessel and we assumed that a visual scoring scheme that evaluated thickness as well as linear extent may correlate better with the Agatston scores. We also believe that correlation of visual score to the more widely accepted Agatston score can be more readily incorporated in to the familiar coronary artery disease risk stratification and prevention[13].

2. Materials and Methods

2.1. COPDGene Cohort

The Genetic Epidemiology of COPD (COPDGene) study is a large, NIH case-control study of genetic markers and the predilection of developing COPD. Participants were non-Hispanic white and African American smokers aged 45 to 80 years with a minimum 10 pack-year smoking history. Exclusion criteria are outlined in the study design by Regan et.al.[14]. All participants received a full dose (120 kVp; 430 mA; 0.625mm collimation) supine CT during full inspiration (non-gated high mA) and another low dose (120 kVp; 100 mA; 0.625mm collimation) supine scan in forced expiration (non-gated low mA), as per the COPDGene protocol[14]. The inspiratory scan technique is comparable to the technique of a routine noncontrast chest CT and the low dose expiratory scan technique is similar to that used in lung cancer screening. The COPDGene screening center at UCLA then recruited 50 patients to also undergo an additional ECG gated limited field of view CT for CAC evaluation (gated high mA - 120 kVp, 430 mA; 2.5mm collimation) at the time of the usual COPDGene trial CT scans. Of these 50 patients, 33 were female and 17 male with a mean age of 62 years and mean pack year smoking history of 41 years. The study was approved by the COPDGene Executive Committee and further details of the COPDGene CT acquisitions are outlined in a prior publication[10].

All three CT series were reconstructed with contiguous 2.5 mm collimation and our TVS scheme was utilized to score the CAC. The right coronary artery (RCA), left anterior descending coronary artery (LAD - including the left main coronary artery and diagonal branches) and left circumflex artery (LCx – including the obtuse marginal branches) were scored both in terms of linear extent and percent of the calcified vessel wall. Linear extent was scored as follows: 0 = No visible calcification; 1 = 1-3 images with calcification; 2 = 4-5 images with calcification; 3 = 6-10 images with calcification; 4 = >10 images with calcification. The thickness of the calcification was scored as follows: 0 = No visible calcification; 1 = any calcification < 25% of vessel diameter; 2 = 1-3 images with 25% or more of the vessel calcified; 3 = >3 images with 25% or more of vessel calcified. Examples of thickness of coronary arterial calcification in the LAD of <25% are shown in Figure 1, and CAC > 25% in the LAD are shown in Figure 2. The individual vessel score was calculated as the product of the linear extent and thickness (ranging from 0-12). The total score was calculated as the sum of the three vessel scores (ranging from 0-36).

TVS method scoring was performed independently by two dedicated cardiopulmonary radiologists (readers 1 and 2) and one post-graduate year-2 resident (reader 3). Agatston scoring was performed on both gated and non-gated exams as previously described. A CT threshold of 130 Hounsfield units (HU) was used and 2 contiguous pixels were required for identification of a calcified lesion. The total coronary calcium score was determined by summing individual lesion scores from each of the main coronary arteries (Left Main, LAD, LCx, RCA).
Figure 1. Part A and Part B. Examples of calcification in the left anterior descending coronary artery less than 25% of the vessel wall diameter (circles).

Figure 2. Part A and Part B. Examples of calcification involving the left anterior descending coronary 25% or greater of the vessel diameter (circles).

The Agatston Scores were log transformed (log_{10} (value +1)) before calculating repeated measures analysis of variance (RM ANOVA) and the Intraclass Correlation Coefficient. TVS scores were analyzed using Friedman’s non-parametric RM ANOVA and the Wilcoxon Signed Rank Test. All correlations involving total visual scores are Spearman Rank Correlations. Agatston scores were compared between the three techniques (ECG-gated high mA, non-gated high mA, non-gated low mA). TVS scores were compared among readers as well as with the Agatston score.

2.2. NLST Cohort

The second component of the study involved visual scoring of a larger cohort of participants in the NLST and was approved by the NLST Research Committee. NLST recruited participants between the ages of 55-75, with 30 pack years’ or more of smoking history. All the participants gave consent to review of the deidentified images for research purposes at the time of the enrollment. Two hundred year 0 LDCT’s, 20 from each of the 10 Lung Screening Study (LSS) sites of the NLST were randomly selected. Details of the LDCT technique used in NLST was previously described[15]. All the examinations were performed on multidetector scanners (4-16 detectors) at 120 kVp, during single breathhold and mAs ranging from 40-60 depending on the size of the patient. The images were reconstructed with 2.0 or 2.5 mm slice thickness. Participant identifiers were electronically removed from all the images. Two examinations were excluded from analysis as the data were corrupted and the exams were not reviewable. Agatston scoring was performed on the remaining 198 exams using TeraRecon application (Aquarius Work Station, TeraRecon, San Mateo, CA). The two experienced cardiopulmonary radiologists and the post-graduate year-2 radiology resident independently performed visual scoring, as described by Shemeshet. al.[5](Shemesh method) on these 198 exams. These three readers also independently scored the 198 exams using our TVS scoring method described above.

For reference, the Shemesh method involves the visual scoring of the left main coronary artery, LAD, RCA, and LCx separately on a 4 point scale (Ranging from 0 to 3) with the linear extent of calcification scored as absent (0), mild (1), moderate (2), or severe (3). Calcification is mild when it involves less than 1/3rd of a vessel length, moderate when it involves greater than 1/3rd but less than 2/3rd of the vessel length, and severe when the calcification involves greater 2/3rd of the vessel length. The individual vessel scores
Correlation between Measured and Visual Scoring of Coronary Artery Calcification

(ranging from 0 to 3) are then added to give the total score (ranging from 0 to 12) [5].

All three scores were not normally distributed and non-parametric statistical methods were used. TVS scores and the Shemesh method scores were correlated with Agatston scores, with each other, as well as with Agatston risk quartiles: 1st quartile - absence of any calcium (Agatston score of 0); 2nd quartile – Agatston score of 1 to 99, 3rd quartile – Agatston score of 100 to 299, and 4th quartile – Agatston score ≥ 300. Friedman’s non parametric RM ANOVA was used to compare scores between the 3 readers. Spearman’s rank correlation and Wilcoxon signed rank test were used for comparisons between 2 readers. Intraclass correlation coefficient (ICC) was used to correlate the 3 measurements.

3. Results

3.1. COPDGene Cohort

The Agatston scores for the different COPDGene groups were as follows (mean ± standard deviation): ECG-gated CT - 588±1335, Non-gated high mA - 922±1838, Non-gated low mA - 1020±1950. The p value of the RM ANOVA was <0.001. The Intraclass Correlation Coefficient was 0.88 (0.77, 0.94). Although there is overlap, the gated high mA score is significantly lower than the other two measurements. The non-gated high mA score is also lower than the non-gated low mA score. Table 1 shows differences in TVS scores between readers and among techniques. The TVS scores of the 3 readers are highly correlated with each other (all ICC ≥ 0.92, all p<0.001). Differences among readers were small and consistent for all three methods. Reader Two scored significantly lower than Reader Three with Reader One having intermediate values (all p<0.002). The TVS scores were significantly correlated with the gated Agatston’s scores for all readers. The non-gated Agatston’s scores also correlated with the gated score (all p<0.001). There were no significant differences among these correlations.

3.2. NLST Cohort

For the second component of the study involving NLST participants100 females (mean and median age of 60.93 and 60, respectively) and 100 males (mean and median age of 61.57 and 60.5, respectively) were included. The correlation between the scores obtained with TVS and the Shemesh method were excellent, with r values greater or equal to 0.95 for all three readers. The r values for TVS and Agatston score correlation ranged from 0.91 to 0.94 for the three readers, and the r values for Shemesh and Agatston score correlation ranged from 0.86 to 0.92 (all p<0.001).

TVS and Shemesh method scores were correlated with Agatston scores and classified according to major Agatston risk quartiles (Table 2).

| Table 1. Total Visual Score (TVS) differences between techniques and readers. |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| Reader | ECG-Gated Inspiration | Non-gated Inspiration | Non-gated Expiration |
| 1 | 9.1±11.4 (0.95) | 10.4±12.3 (0.94) | 8.0±12.3 (0.90) |
| 2 | 7.7±10.3 (0.96) | 8.6±11.7 (0.96) | 7.3±11.7 (0.96) |
| 3 | 9.6±11.6 (0.95) | 11.4±12.4 (0.94) | 9.3±12.0 (0.93) |

Intraclass Correlation (r) 0.95 0.94 0.92

Data given as mean±s.d. (correlation coefficient with Agatston score)

| Table 2. Range of values and median values of visual scoring methods for Agatston score risk groups. Median values in bold. |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| Agatston Score Risk Groups | TVS | Shemesh Method |
| Reader | 0 (n=72) | 1 to 99 (n=52) | 100 to 299 (n=30) | ≥300 (n=44) |
| Reader | 0 (n=72) | 1 to 99 (n=52) | 100 to 299 (n=30) | ≥300 (n=44) |
| TVS | 0.1±0.3 (0, 1) | 3.4±3.7 (0, 19) | 9.0±7.1 (0, 33) | 19.5±8.7 (3, 36) |
| 1 | 0.1±0.3 (0, 1) | 3.8±3.7 (0, 18) | 7.9±4.9 (2, 19) | 20.8±8.8 (4, 36) |
| 2 | 0.4±0.8 (0, 3) | 4.9±4.0 (0, 16) | 10.5±6.5 (2, 36) | 21.4±8.2 (6, 36) |
| 3 | 0.1±0.3 (0, 1) | 1.8±1.2 (1, 5) | 3.5±2.1 (0, 9) | 6.1±2.6 (1, 15) |
| 2 | 0.1±0.3 (0, 1) | 1.9±1.4 (0, 6) | 3.1±1.6 (1, 7) | 5.7±2.2 (2, 10) |
| 3 | 0.5±0.9 (0, 4) | 2.8±1.8 (0, 9) | 4.3±2.1 (1, 10) | 7.2±2.8 (2, 12) |

Data are given as mean±s.d. (minimum, maximum)
### Table 3. Cutoff values for TVS and Shemesh scores for major Agatston risk quartiles

<table>
<thead>
<tr>
<th>Agatston Score</th>
<th>TVS</th>
<th>Shemesh Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0-99</td>
<td>1 to 5</td>
<td>1 to 2</td>
</tr>
<tr>
<td>100-299</td>
<td>6 to 12</td>
<td>3 to 5</td>
</tr>
<tr>
<td>≥300</td>
<td>≥13</td>
<td>≥6</td>
</tr>
</tbody>
</table>

Percent classified correctly

| Reader 1      | 76%     | 72%           |
| Reader 2      | 80%     | 71%           |
| Reader 3      | 70%     | 62%           |

1.5 to 2.5% were misclassified by 2 categories and none by 3
2.5 to 5.0% were misclassified by 2 categories and none by 3

Using either TVS or the Shemesh method, both the mean values and the median values divide nicely along the lines of the major Agatston score quartiles, demonstrating these scores may substitute for measured Agatston scores.

Cutoff values corresponding to major Agatston risk quartiles were derived for both TVS and the Shemesh method (Table 3). Cutoff values for TVS scores that would correspond to Agatston score groups were calculated as 0, 1 to 5, 6 to 12 and ≥13. Cutoff values for Shemesh scores that would correspond to Agatston score groups were calculated as 0, 1 to 2, 3 to 5 and ≥6. The Shemesh scores did not divide as well as TVS scores as the smaller range of values made it harder to discriminate the 1 to 99 and 100 to 299 categories.

TVS and Shemesh scores that best predicted Agatston scores >100 were also identified. A TVS >5 had an average sensitivity of 89% for the three readers for predicting Agatston scores. The average specificity was 90%. The positive predictive value averaged 85% while the negative predictive value averaged 93%. Using a Shemesh score of ≥2 to predict Agatston scores >100 resulted in an average sensitivity of 83% for the three readers. The average specificity was 89%. Positive predictive value averaged 79% while negative predictive value averaged 90%. For each parameter, TVS performed better than Shemesh scores.

### 4. Discussion

Cardiovascular disease is the most common cause of death in smokers and assessment of CAC in this population may be beneficial, particularly in patients without known CVD. Given recent data suggesting that chest CT can characterize COPD both qualitatively as well as quantitatively and results from the NLST showing that LDCT screening confers a significant mortality benefit, chest CT examinations among smokers are sure to increase[7, 10]. The Agatston CAC scoring method has been well validated as a predictor of cardiovascular outcomes in numerous prior studies. However, Agatston scoring is not routinely performed on non-cardiac chest CT exams. If one could demonstrate good correlation of a visual scoring system that also correlates with the Agatston scoring system, this would potentially be useful for a busy clinical practice. To our knowledge, no other group has compared visual and Agatston scoring on ECG-gated, high mA (typical calcium scoring protocols), non-gated low mA (lung cancer screening protocols), and non-gated high mA (typical clinical chest CTs) performed at the same time. We have shown in this pilot investigation through the COPDGene cohort that our semi-quantitative visual scoring system of CAC from non-gated chest CT (conventional or low dose) can predict clinically significant CAC as previously validated via the Agatston method.

While the Agatston scores can vary between 0 and thousands, the scores that predict a significant probability of cardiac events are relatively narrow and recent data from MESA found that an Agatston score higher than 300 is associated with a greater than 9 fold increase in major adverse cardiac events[4] and the same value is also used in the recent AHA guidelines for CAD risk stratification and prevention[16]. Additionally, the recent work by Jacobs et. al. demonstrated a hazard ratio for all cause mortality greater than 6 for Agatston scores > 100 compared with zero as measured on low dose lung cancer screening CT exams [6].Our results from the NLST population show excellent correlation between two visual coronary calcification scoring schemes and calculated Agatston scores and appear to satisfactorily discriminate between those that are associated with higher adverse events. Our TVS system did allow better prediction of Agatston scores >100 when compared with the Shemesh method (the positive predictive value of TVS >5 averaged 85% while the positive predictive value of Shemesh >2 averaged 79%). To our surprise, we did not observe better overall correlation with our TVS system, which weights thickness of calcification in addition to linear
extent, than with the Shemesh’s method which only weights the linear extent of calcification. Agatston scoring also weights the thickness of the calcium in its area and density measurements. The median or mean values for both visual scoring schemes divided nicely along major Agatston risk quartiles and either would be valuable in providing added cardiovascular risk stratification, with no additional testing, radiation, or cost to the patient. Using cutoff values we derived, the radiologist could quickly place the patient in an Agatston risk quartile correctly in the majority of cases (Table 3). From the data from Shemesh et al., a visual score of 4 or greater was predictive of cardiovascular death (odds ratio of 2.1 compared with a score of 0 and adjusted for sex, age, and pack-years of smoking)\[5\]. A Shemesh score of 4 would place an individual in the top 299 Agatston risk quartile using the cut-off points we derived, and a Shemesh score of 6 or greater would place the patient in the highest risk quartile (Agatston score 300 or greater). An advantage of a visual scoring method that can discriminate clinically significant and insignificant values of Agatston scores may also provide more compelling evidence for patients’ physicians to act upon such description compared to non-standardized descriptions of mild, moderate or significant CAC in the CT reports. Excellent agreement between 3 readers of varying levels of experience also suggests that the methods can be used by any radiologist familiar with CT coronary artery anatomy. A recent publication also describes a novel visual CAC scoring scheme that was also effective in stratifying patients into various Agatston risk quartiles on non-gated chest CT exams[12]. We were unaware of this visual scheme during our experiment but would be interested in comparing this method with others in future projects.

Limitations of the study include the lack of ECG gated ‘normal dose’ CT exams in the NLST participants, which is the standard currently for Agatston calcium scoring; however, a prior published report[9] has shown good correlation with Agatston scores in low dose non-gated exams and typical gated coronary calcium scoring exams. Additionally, data we obtained using COPDGene ECG-gated CT’s and non-gated low dose CT’s demonstrated good agreement between visual scoring of the ECG-gated and non-gated exams. In this larger NLST data set we found that a visual score of \(\geq 13\) allowed appropriate classification of most exams into the highest quartile, and the majority of patients were placed appropriately in each of the 4 Agatston quartiles using the cutoff values described above (appropriate quartile grouping ranged between 70 and 80% for the three readers). Additional limitations include a lack of mortality data and relatively small sample size which could be remedied with future studies as more CAC visual scoring data becomes available from the NLST.

**5. Conclusion**

The TVS visual scoring scheme we developed correlated well with Agatston scores in ECG-gated and non ECG-gated exams, as demonstrated in the COPDGene cohort of this study. Testing our TVS visual scoring scheme with the previously published scheme developed by Dr. Shemeshin the larger NLST cohort demonstrated high correlation with each other and with major Agatston risk quartiles for readers of varying experience levels. An easy semi-quantitative visual scoring system, such as our TVS system or Shemesh’s system could be of significant benefit to patients undergoing lung cancer screening or receiving non-gated CT examinations of the chest for other indications, allowing radiologists to quickly stratify patients into major Agatston risk quartiles.

REFERENCES


