# Reviewer #1 Comments

*Overview:*

This study is focused on an important limitation of Agatston score, which is a calcium thresholding defined as 130 HU. Black et al utilizes an integrated intensity and a volume fraction methods to overcome the above mentioned limitation. Topic is of significant interest, as due development of CT technology a new method of calcium scoring is needed. I have the following questions, concerns and/or suggestions for the authors to consider:

*Methods:*

1. Why were simulation phantoms used? And if Authors have an opportunity to use simulation phantom, why only one CT vendor was simulated? Wouldn’t it be more informative to present results from different CT systems? It is essential to increase the number of CT vendors to show clinical relevance. As already known, there is a discrepancy between CAC scored from different CT systems, therefore an integrated intensity and a volume fraction should be validated on different CT systems from different vendors.
   1. Simulation allows for a more controlled environment to test the limits of these new CAC scoring techniques. The simulation was chosen as a precursor to physical phantom studies and, eventually, patient studies to limit the number of variables involved in the comparison. The simulation package we have developed over the past few years does not include multiple different CT vendors. However, we are working on upgrading to a new simulation system that includes many different vendors for future studies.
2. I do believe that slice thickness should be the same. I do understand that Agatston methodology was created on 3.0 mm slice thickness, but just for the clarity of the results there should be added one column (Table 1.) presenting the same very same acquisition and reconstruction settings for all investigated methods. Otherwise, Authors enhance the results of simulated phantom.
   1. Thank you for pointing this out. Slice thickness in this simulation is limited to 0.5 mm, which makes us unable to include multiple slice thickness parameters without major adjustments to the simulation package. Although, we are currently continuing this work on physical phantoms with varied slice thicknesses, including 3.0 mm. The discussion section was adjusted further to highlight the limitation of slice thickness in this study.

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 29. 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, and this study is limited without a comparison to Agatston scoring at the gold standard slice thickness of 3 mm.

1. What is the size of simulated calcifications? This is not explained in the manuscript, therefore it is difficult to assess the performance of presented methods.
   1. Thank you for addressing this. The insert diameters are listed in the methods section as 1, 3, and 5 mm. The inserts were each 1.5 mm in length. This detail was not included in the original manuscript, and a sentence has been added to the methods section explaining this.

Three calcification inserts of different diameters (1, 3, and 5 mm), each with a length of 1.5 mm, and different hydroxyapatite (HA) densities were placed within each phantom.

1. In the point 2.3 Agatston scoring. Did authors also correct for the weighting factor when changing kV? Or only for the HU threshold for calcium detectability? The factor for 120 kV is defined as follows: 130‐199 HU, factor 1; 200‐299 HU, factor 2; 300‐399 HU, factor 3; and ≥ 400 HU, factor 4. If the threshold is changed, the factor also should be adjusted.
   1. Thank you for pointing this out. We added more details in the methods section explaining the correction for the weighting factor related to the kV-dependent threshold.

Additionally, the weighting factor was adjusted in a similarly kV-dependent manner, according to criteria outlined in Gräni, C. et al., and extrapolated for a tube voltage of 135 kV 19.

1. Is there any possibility to utilise the dynamic phantom for this study? As we know, motion creates motion artefacts which strongly affects Agatston score calculation. As motion was only simulated and was not reflected in physical phantom.
   1. The dynamic phantom images acquired by Praagh et al. are also limited. The motion is linear and therefore does not produce realistic motion artifacts like one would see in the heart. Also, the calcium rods included in the motion images were limited to large inserts (5 mm) and higher densities (> 200 mg/cc). This simulation study focuses more on the low-density regime since Agatston scoring is already reasonably effective for large, high-density calcifications. Because of the limitations of the previously acquired dynamic images, we are currently working on a dynamic phantom study using the CIRS Dynamic Cardiac Phantom. This future study will include much more realistic cardiac motion and inserts within the size and density regime that Agatston scoring is ineffective at measuring.

*Results:*

1. How did author compare 25 and 50, as by definition, these densities cannot be detected with 130 HU?
2. Fig 4 – please change the scale of y‐axis, it is difficult to see values. Is it possible to analyze densities separately? What the small, medium, and large insert mean? Is it about phantom size or about calcification size? Please, specify. I assume it is a patients’ size.
3. I assume there is an error in tables, as it shows 0/216 false negatives for Agatston score
4. Why the false‐negatives were excluded from reproducibility analysis? I do not see a rationale reason for this, it should be included.
5. May you display results in graphs (i.e. Fig 5) separately, by CAC density, radiation dose, and patient’s size?

# Reviewer #2 Comments

*Overview:*

Thank you for the opportunity to review this manuscript entitled 'Coronary artery calcium mass measurement based on integrated intensity and volume fraction techniques. The paper describes an interesting and relevant topic.   
Please consider the following general comments on this paper:

*General Comments:*

1. With the recommended Agatston threshold of 130 HU, together with the recommended noise level of about 24 HU, wouldn't simply lowering the Agatston scoring threshold to a lower level increase CAC sensitivity? This is also shown in Table 3. Please comment on this option, also in view of the proposed method in the current study and it's implementation in the clinic.
2. Please elaborate on the impact of missing Compton-scattering in the used simulation.
3. For the simulated phantom, which material was used in conjunction with the pure HA to obtain the different densities? Was is a combination with solid water, or with muscle?
4. Please provide an overview of CT number measurements between the simulated and real phantom, for (at least) the fat ring, material surrounding the calcification, all calcification densities, calibration disks.
5. Paragraph 2.3: heading title contains a typo (Agaston)
   1. Thank you for catching this. This has been fixed.
6. For both techniques, were scores calculated based 3 mm reconstructions? Or calculated based on other slice thicknesses, and subsequently recalculated to 3 mm equivalent scores?

# Editor's Comments

*Overview:*

I concur with two expert reviewers. The topic is of significant interest and there is a need for a new CAC scoring standard. However, there are some concerns raised by both reviewers. Please consider these comments/suggestions carefully and address them thoroughly.