PEER REVIEW HISTORY

ARTICLE DETAILS

Title (Provisional)

Cohort profile: AI driven national Platform for CCTA for clinicaL and industriaL applicatiOns (APOLLO)

Authors

Baskaran, Lohendran; Leng, Shuang; Dutta, Utkarsh; Teo, Lynette; Yew, Min Sen; Sia, Ching-Hui; Chew, Nicholas; Huang, Weimin; Lee, Hwee Kuan; Vaughan, Roger; Ngiam, Kee Yuan; Lu, Zhongkang; Wang, Xiaohong; Tan, Eddy Wei Ping; Cheng, Nicholas Zi Yi; Tan, Swee Yaw; Chan, MY; ZHONG, LIANG

VERSION 1 - REVIEW

Reviewer **1**

Name **Song, Lei**

Affiliation **Chinese Academy of Medical Sciences & Peking Union Medical College Fuwai Hospital**

Date **22-Jul-2024**

COI **None**

In the manuscript entitled “Cohort profile: AI driven national Platform for CCTA for clinicaL and industriaL applicatiOns (APOLLO)”, Dr. Mark Chan and Dr. Liang Zhong et al. conduct a hybrid, retrospective-prospective, open-label, observational, multi-centre study, to build an AI-driven platform for CAD assessment using CCTA in the context of the multi-ethnic population of Singapore.

As there is an unmet need for integrated information from patient demographics, lab tests, imaging data as CCTA and clinical outcomes, AI technique will play a key role to address this issue. This cohort data from Asian population is highly valuable and necessary. However, there are some important issues still need to address and consider:

1. Patients included from 1 January 2007 to 31 December 2025. So not like the description in Page 7, Line 54, patients were not only retrospectively but also prospectively recruited. If it's correct, what's the target number of this study, the authors mentioned a "5000" CAD registry in the Graphical abstract, but not in the main text.

2. The enrolled subjects cross 18 years from 2007 to 2025, considering the devices and technique progression, there will be a huge heterogeneity among patients demographics, imaging modalities, as well as treatment strategies. How to eliminate the bias from these differences?

3. The definition of clinical endpoints need to be more accurate, as the definition of MACE, should not described as "not limited to", and also, such events of "arrhythmias", is hard to identify and usually not included in the MACEs, or should be more detailed and feasible in practice.

4. Patients will be prospectively followed up till 31 December 2025, this time is exactly same as the recruit time, how to perform clinical follow-up for the last enrolled patient?

5. AI-driven national platform for CCTA is core purpose of this study, more detailed information should be described, including the process of automation of anonymization, reporting, scoring, quantification and finally, how to integrate these data and establish the AI model.

6. The author should clearly report the process of size calculation to provide enough power building the four AI model, as well as how to built a clinical risk model, the sample size calculation, the predict event rate over the 5-year follow-up.

7. More limitation should be discussed in the discussion, as the difficulties in data collecting and sharing, the extrapolation of conclusion.

Reviewer	**2**
Name	**Szczeklik, W**
Affiliation	**Jagiellonian University Medical College, Internal Medicine**
Date	**01-Sep-2024**
COI	**NO**

Thank you for the possibility to review: „Cohort profile: AI driven national Platform for CCTA for clinicaL and industriaL applicatiOns (APOLLO)”.

This interesting paper reports on the project APOLLO that aim to create Singapore national platform for collection, diagnosis, interpretation (automatized) and predicting CAD using CCTA.

This is one of the largest of such studies and as such wormly welcomed, as there are not many studies with such large population cohort analysed in prospective manner - especially in Asia.

In the manuscript authors report on the 1/project - explaining the methodology and 2/on the characteristics of the cohort.

I have no major comments. It is comprehensible, and nicely written.

Please consider addidng a short description of possible limitations of the study and planned substu

---

Reviewer	**3**
Name	**Deckers, Jaap**
Affiliation	**University Medical Center Utrecht, Department of Cardiology**
Date	**13-Sep-2024**
COI	**none**

The growing burden of CAD in Asia and the emergence of novel CT-based risk markers have prompted the investigators to build an AI-driven platform for detailed CAD assessment of - amongst others I assume - CCTA findings in the context of the multi-ethnic population of Singapore. The current paper (only) describes the baseline characteristics of the first 1884 (of 4196) patients. Follow-up is being conducted and the final results of the APOLLO study are eagerly awaited. Until that time, however, publication of the current status of the study in the BMJ Open Journal seems somewhat premature.

VERSION 1 - AUTHOR RESPONSE

Reviewer #1

Prof. Lei Song, Chinese Academy of Medical Sciences & Peking Union Medical College Fuwai Hospital

Comments to the Author:
In the manuscript entitled “Cohort profile: AI driven national Platform for CCTA for clinicaL and industriaL applicatiOns (APOLLO)”, Dr. Mark Chan and Dr. Liang Zhong et al. conduct a hybrid, retrospective-prospective, open-label, observational, multi-centre study, to build an AI-driven platform for CAD assessment using CCTA in the context of the multi-ethnic population of Singapore.

As there is an unmet need for integrated information from patient demographics, lab tests, imaging data as CCTA and clinical outcomes, AI technique will play a key role to address this issue. This cohort data from Asian population is highly valuable and necessary. However, there are some important issues still need to address and consider:

1. Patients included from 1 January 2007 to 31 December 2025. So not like the description in Page 7, Line 54, patients were not only retrospectively but also prospectively recruited. If it's correct, what's the target number of this study, the authors mentioned a "5000" CAD registry in the Graphical abstract, but not in the main text.

Reply: Thank you for pointing that out, and we apologize for the confusion. We have revised the description for clarity (Page 8 of revised manuscript). The study utilizes a hybrid recruitment approach, both retrospective and prospective, targeting a total of 5,000 CAD patients.

For the retrospective arm, patients who underwent CT scans between 2007 and 2017 will be screened and included if they meet the inclusion criteria. Outcomes will be obtained through a review of medical records and national registries, with follow-up continuing until December 31, 2025.

For the prospective arm, patient recruitment occurred from October 2021 to July 2024. Clinical events and outcomes will be tracked for a period of five years following enrollment, with data collected from hospital medical records and national databases.

2. The enrolled subjects cross 18 years from 2007 to 2025, considering the devices and technique progression, there will be a huge heterogeneity among patients demographics, imaging modalities, as well as treatment strategies. How to eliminate the bias from these differences?

Reply: While the original plan was to screen retrospective patients from 2007 to 2017, the actual retrospective enrollment occurred from 2015 to 2017. The prospective patients were recruited between 2021 and 2024. After thorough review, we found minimal heterogeneity in patient demographics and imaging modalities, as there were no significant changes in the CT scanners used during this period. Additionally, patients who had undergone any form of coronary revascularization (percutaneous coronary intervention or coronary artery bypass grafting) were excluded from the study, which helps mitigate heterogeneity in treatment strategies.

3. The definition of clinical endpoints need to be more accurate, as the definition of MACE, should not described as "not limited to", and also, such events of "arrhythmias", is hard to identify and usually not included in the MACEs, or should be more detailed and feasible in practice.

Reply: We have revised the definition of the MACE endpoint for greater clarity. The phrase "not limited to" and the inclusion of "arrhythmias" have been removed. The updated definition for Major Adverse Cardiovascular Events (MACE) now includes acute myocardial infarction, stroke, heart failure, and percutaneous or surgical revascularization (Page 11 of revised manuscript).

4. Patients will be prospectively followed up till 31 December 2025, this time is exactly same as the recruit time, how to perform clinical follow-up for the last enrolled patient?

Reply: Thank you for your comment, and we apologize for the confusion. The date of December 31, 2025, refers to the end of follow-up for the retrospective study. Retrospective patient recruitment took place between 2015 and 2017, and clinical outcomes will be collected through reviews of medical records and national registries, with follow-up continuing until that date. For prospectively recruited patients, clinical events and outcomes will be tracked for a period of five years following their enrollment (Page 8 of revised manuscript).

5. AI-driven national platform for CCTA is core purpose of this study, more detailed information should be described, including the process of automation of anonymization, reporting, scoring, quantification and finally, how to integrate these data and establish the AI model.

Reply: We have provided more detailed information on the AI-driven platform in the revised manuscript (Pages 12 and 13 of revised manuscript):

The images, data, and analyses will be uploaded to an AI platform for de-identification, analysis, integration, and automated reporting. A container will encapsulate all AI solutions developed during the study, allowing for seamless deployment across third-party environments, including laptops, cloud platforms, and both Windows and Linux operating systems. The toolkits can also be integrated with commercial third-party software platforms.

(1) **AI anonymization:** The anonymization of pixel data follows a pipeline consisting of the following steps:

- Extracting personal data from the DICOM metadata that may be present in the pixel image. This information defines a set of words the pipeline searches for within the image.
- Preprocessing the image to enhance contrast and reduce noise.
- Deploying a convolutional neural network (CNN) for alphabet recognition, which identifies characters within the image.
- Matching and removing personal data by cross-referencing identified words with those found in the DICOM metadata.

(2) **AI stenosis grading:**

- Coronary artery tree detection: Error-tolerant graph neural network technology [Gao et al., 2019] is integrated into the platform. Building on prior work by Huang et al. [Huang et al., 2018], we use an enhanced 3D U-Net model to identify coronary arteries. A graph U-Net model further filters these candidates based on topological, positional, and image features (**Supplementary Figure 1**). Non-coronary segments and discontinuous arteries are either removed or reconnected as necessary. The result is a coronary artery tree that is easily mapped due to its graph structure.
- Joint stenosis grading and plaque quantification on 3D images: Stenosis grading and plaque quantification are performed simultaneously by an algorithm combining a 3D U-Net model and a 3D image classifier (**Supplementary Figure 2**). The U-Net generates segmentation masks for the lumen, calcified plaque, and non-calcified plaque. These are then used as inputs for the image classifier, which outputs stenosis grades and plaque types.

(3) **AI Agatston score analysis:** AI-based Agatston scoring begins with the segmentation of calcified plaque on non-contrast CT scans (**Supplementary Figure 3**), leveraging CNNs [Lessmann et al., 2018; Wolterink et al., 2016]. A novel approach in our platform involves combining non-contrast and contrast CT scans, aligned through multimodal image

registration. A deep learning multitask network analyzes both plaque and calcification. This interpretable multitask learning algorithm provides more accurate analysis.

(4) **AI epicardial adipose tissue (EAT):** AI-based EAT quantification uses 2D axial slices (**Supplementary Figure 4**), with segmentation achieved through fully convolutional networks (e.g., U-Net) or fully annotated CTs.

(5) **AI reporting:** The AI-generated reports include Agatston scoring and stenosis grading. Automated tasks include:

- CCTA image quality evaluation
- Heart segmentation
- EAT segmentation and analysis
- Aorta segmentation
- Detection and registration of the coronary artery tree
- Agatston scoring and stenosis grading

- Gao HY, Ji SW. Graph U-Nets [online]. 2019. https://arxiv.org/abs/1905.05178 (accessed 20 Apr 2024).
- Huang W, Huang L, Lin Z, *et al*. Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images. Annu Int Conf IEEE Eng Med Biol Soc. 2018;2018:608-11. doi: 10.1109/EMBC.2018.8512328.
- Lessmann N, van Ginneken B, Zreik M, *et al*. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. IEEE Trans Med Imaging. 2018;37:615-25. doi: 10.1109/TMI.2017.2769839.
- Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. Med Image Anal. 2016;34:123-36. doi: 10.1016/j.media.2016.04.004.

6. The author should clearly report the process of size calculation to provide enough power building the four AI model, as well as how to build a clinical risk model, the sample size calculation, the predict event rate over the 5-year follow-up.

Reply: Sample Size Calculation: Conventional sample size calculations rely on a predefined margin of error; however, in AI, estimating the required sample size prior to experimentation is not always feasible because this margin of error cannot be established until the deep learning (DL) model development process has begun. Instead, during DL analysis, we will use cross-validation techniques, such as k-fold cross-validation and hold-out test sets, to ensure robust statistical confidence in our model. These methods allow us to derive confidence intervals and assess model performance iteratively.

To further ensure statistical rigor, we will also perform a post-hoc power analysis to evaluate the actual power achieved by the analysis and report these findings. Additionally, previous studies provide support for the adequacy of our dataset size. For example, a coronary artery calcium DL project using 377 subjects achieved over 90% accuracy [Singh et al., 2021]. Similarly, Commandeur et al. [Commandeur et al. 2020] demonstrated improved predictive performance using AI on cardiac CT in 1912 asymptomatic subjects, achieving a higher AUC (0.82 vs. 0.77, $P < 0.05$) compared to conventional methods like coronary artery calcium

scoring. These results suggest that the dataset collected in our current study should be sufficient to train the DL models effectively and achieve high performance for each specific aim.

In previous studies, the 5-year rate of the primary endpoint in the CTA group was reported as 2.3% in one study [Newby et al., 2018], while another study observed a primary endpoint event in 164 out of 4,996 patients (3.3%) over a median follow-up of 25 months [Douglas et al., 2015]. Based on these findings, we predict the event rate in our current study to be approximately 3% over the 5-year follow-up period. With a predicted event rate of 3%, we aim to capture at least 100 events to ensure robust model estimation. Following the general rule of 10 events per predictor variable [Peduzzi et al., 1996], we anticipate including 10 predictors in our models, resulting in a required sample size of approximately 3333 patients.

Clinical Risk Model Development: This study aims to build a clinical risk model that incorporates parameters derived from AI calcium score, AI epicardial adipose tissue (EAT), AI stenosis, and AI plaque characteristics to predict patient outcomes or progression of atherosclerosis. In addition to the AI-derived features, traditional clinical and demographic data will also be included. The risk model will be developed using logistic regression, Cox proportional hazards models, or machine learning algorithms. By combining the AI-derived features with traditional risk factors, patients will be categorized into risk groups (e.g., low, intermediate, high risk) based on output probabilities or risk scores. After model construction, a post-hoc power analysis will be conducted to ensure that the sample size used is sufficient to detect meaningful associations and that the model possesses adequate statistical power.

These descriptions have been incorporated into the revised manuscript (Pages 14-16 of revised manuscript).

- Singh G, Al'Aref SJ, Lee BC, *et al*. End-to-end, pixel-wise vessel-specific coronary and aortic calcium detection and scoring using deep learning. Diagnostics (Basel). 2021;11:215. doi: 10.3390/diagnostics11020215.
- Commandeur F, Slomka PJ, Goeller M, *et al*. Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study. Cardiovasc Res. 2020;116:2216-25. doi: 10.1093/cvr/cvz321.
- SCOT-HEART Investigators; Newby DE, Adamson PD, Berry C, *et al*. Coronary CT angiography and 5-year risk of myocardial infarction. N Engl J Med. 2018;379:924-33. doi: 10.1056/NEJMoa1805971.
- Douglas PS, Hoffmann U, Patel MR, *et al*. Outcomes of anatomical versus functional testing for coronary artery disease. N Engl J Med. 2015;372:1291-300. doi: 10.1056/NEJMoa1415516.
- Peduzzi P, Concato J, Kemper E, *et al*. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373-1379.

7. More limitation should be discussed in the discussion, as the difficulties in data collecting and sharing, the extrapolation of conclusion.

Reply: We have amended the manuscript to include the following points (Pages 20 and 21 of revised manuscript): This study has limitations. First, although APOLLO utilizes a national, multi-centre platform, there may still be inherent limitations regarding the completeness and consistency of data across different centres—an issue commonly encountered in multi-centre studies. Second, challenges may arise with the scalability of data sharing and the federated use of data across various healthcare entities. Third, the study focuses exclusively on a multi-ethnic Asian population, which may limit the generalizability of the findings to Western populations.

Differences in genetic, environmental, and lifestyle factors indicate that further validation studies are needed to ensure the broader applicability of the developed AI models.

Reviewer: 2
Dr. W Szczeklik, Jagiellonian University Medical College

Comments to the Author:
Thank you for the possibility to review: Cohort profile: AI driven national Platform for CCTA for clinicaL and industriaL applicatiOns (APOLLO)".

This interesting paper reports on the project APOLLO that aim to create Singapore national platform for collection, diagnosis, interpretation (automatized) and predicting CAD using CCTA. This is one of the largest of such studies and as such warmly welcomed, as there are not many studies with such large population cohort analysed in prospective manner - especially in Asia.

In the manuscript authors report on the 1/project - explaining the methodology and 2/on the characteristics of the cohort. I have no major comments. It is comprehensible, and nicely written. Please consider adding a short description of possible limitations of the study and planned sub-study.

Reply: This study has limitations. First, although APOLLO utilizes a national, multi-centre platform, there may still be inherent limitations regarding the completeness and consistency of data across different centres—an issue commonly encountered in multi-centre studies. Second, challenges may arise with the scalability of data sharing and the federated use of data across various healthcare entities. Third, the study focuses exclusively on a multi-ethnic Asian population, which may limit the generalizability of the findings to Western populations. Differences in genetic, environmental, and lifestyle factors indicate that further validation studies are needed to ensure the broader applicability of the developed AI models.

A planned sub-study could focus on comparing CAD characteristics across different Asian ethnic groups within the APOLLO study cohort. The aim would be to investigate potential ethnic-specific variations in CAD presentation, severity, and plaque characteristics. This sub-study would leverage CCTA combined with AI-driven analyses to: (1) assess the distribution and extent of coronary artery plaques across various ethnic groups (e.g., Chinese, Malay, Indian); (2) examine differences in plaque composition (e.g., calcified, non-calcified, or mixed plaques) and their association with cardiovascular risk factors (e.g., hypertension, diabetes); and (3) explore how genetic, environmental, and lifestyle factors contribute to CAD risk and progression among different ethnicities, potentially revealing unique risk profiles or protective factors in specific groups.

These descriptions have been incorporated into the revised manuscript (Pages 20 and 21 of revised manuscript).

Reviewer: 3
Dr. Jaap Deckers, University Medical Center Utrecht

Comments to the Author:

The growing burden of CAD in Asia and the emergence of novel CT-based risk markers have prompted the investigators to build an AI-driven platform for detailed CAD assessment of - amongst others I assume - CCTA findings in the context of the multi-ethnic population of Singapore. The current paper (only) describes the baseline characteristics of the first 1884 (of 4196) patients. Follow-up is being conducted and the final results of the APOLLO study are eagerly awaited. Until that time, however, publication of the current status of the study in the BMJ Open Journal seems somewhat premature.

Reply: We have revised Table 2 to include the baseline characteristics of all 4196 patients. The characteristics of these 4196 patients are largely consistent with those of the initial subset, ensuring the robustness of our findings. While the final results of the APOLLO study are still pending, we believe that presenting these comprehensive baseline characteristics at this stage provides valuable insight and contributes to the existing literature on CAD in the Asian population.