


BMJ Open A new and automated risk prediction of coronary artery disease using clinical endpoints and medical imaging-derived patient-specific insights: protocol for the retrospective GeoCAD cohort study

Dona Adikari ,^{1,2} Ramtin Gharleghi,³ Shisheng Zhang,³ Louisa Jorm,⁴ Arcot Sowmya,⁵ Daniel Moses,^{5,6} Sze-Yuan Ooi,^{1,2} Susann Beier³

To cite: Adikari D, Gharleghi R, Zhang S, *et al.* A new and automated risk prediction of coronary artery disease using clinical endpoints and medical imaging-derived patient-specific insights: protocol for the retrospective GeoCAD cohort study. *BMJ Open* 2022;**12**:e054881. doi:10.1136/bmjopen-2021-054881

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-054881>).

Received 26 June 2021
Accepted 05 June 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Dona Adikari;
dona.adikari@unsw.edu.au

ABSTRACT

Introduction Coronary artery disease (CAD) is the leading cause of death worldwide. More than a quarter of cardiovascular events are unexplained by current absolute cardiovascular disease risk calculators, and individuals without clinical risk factors have been shown to have worse outcomes. The ‘anatomy of risk’ hypothesis recognises that adverse anatomical features of coronary arteries enhance atherogenic haemodynamics, which in turn mediate the localisation and progression of plaques. We propose a new risk prediction method predicated on CT coronary angiography (CTCA) data and state-of-the-art machine learning methods based on a better understanding of anatomical risk for CAD. This may open new pathways in the early implementation of personalised preventive therapies in susceptible individuals as a potential key in addressing the growing burden of CAD.

Methods and analysis GeoCAD is a retrospective cohort study in 1000 adult patients who have undergone CTCA for investigation of suspected CAD. It is a proof-of-concept study to test the hypothesis that advanced image-derived patient-specific data can accurately predict long-term cardiovascular events. The objectives are to (1) profile CTCA images with respect to variations in anatomical shape and associated haemodynamic risk expressing, at least in part, an individual’s CAD risk, (2) develop a machine-learning algorithm for the rapid assessment of anatomical risk directly from unprocessed CTCA images and (3) to build a novel CAD risk model combining traditional risk factors with these novel anatomical biomarkers to provide a higher accuracy CAD risk prediction tool.

Ethics and dissemination The study protocol has been approved by the St Vincent’s Hospital Human Research Ethics Committee, Sydney—2020/ETH02127 and the NSW Population and Health Service Research Ethics Committee—2021/ETH00990. The project outcomes will be published in peer-reviewed and biomedical journals, scientific conferences and as a higher degree research thesis.

Strengths and limitations of this study

- ⇒ GeoCAD is a retrospective cohort study to assess anatomical risk in 1000 adult patients who have undergone CT coronary angiography for suspected coronary artery disease (CAD).
- ⇒ We propose a novel approach predicated on our current understanding of clinical and additional demographic risk factors, coronary artery calcium scoring and machine learning methods to non-invasively determine the relationship between shape features, wall shear stress and the risk of clinical endpoints in a large population.
- ⇒ This provides an unprecedented opportunity to translate advanced imaging analyses to clinical practice, using novel anatomical biomarkers to develop improved risk models for CAD.
- ⇒ This is a single-centre study which potentially limits the patient cohort considered and the findings may thus be limited to such cohort.

INTRODUCTION

The landmark Framingham Heart Study, which was commenced in 1948, established the principle of coronary risk profiling using a simple equation with clinical risk factors independently predictive of coronary artery disease (CAD) and remains commonly used today.¹ However, CAD is still the leading cause of death worldwide despite the implementation of statin therapy and a movement towards aggressive low-density lipoprotein (LDL) cholesterol lowering.^{2–4} In fact, more than a quarter of cardiovascular events are unexplained by clinical risk equations, surmising that there are other risk factors for atherosclerosis that have not been identified.^{5,6} Even more concerning, ST-segment elevation myocardial infarction (STEMI) patients without standard modifiable risk



factors (SMuRFs) have significantly worse in-hospital outcomes compared with those with one or more risk factors.⁷ Contemporary scoring algorithm studies such as PREDICT in New Zealand and QRISK3 in the United Kingdom showed promising improvements in the accuracy of cardiovascular risk estimation in vulnerable high-risk subpopulations by incorporating additional demographic predictors such as socioeconomic indicators and ethnicity.^{8,9} Inevitably, there is a tremendous opportunity for improved CAD risk prediction by identifying the remaining risk indicators which may yield a paradigm shift from intervention to a greater focus on primary prevention.

Anatomical biomarkers encompass haemodynamic risk which explain, at least in part, some of the variance in susceptibility to cardiovascular disease among individuals and thus can help to improve cardiovascular risk identification and stratification.^{6,10,11} Specifically, atherosclerosis is the manifestation of the complex interplay between the triad of systemic risk factors, haemodynamic factors and the physiological response of the arterial wall.¹⁰ Systemic risk factors have been compounded to create current probabilistic risk scores,¹ yet the latter two, haemodynamic factors and the physiological response, remain ignored in clinical risk assessments. However, it has been observed that atherosclerotic plaques form and progress preferentially at geometrically predisposed locations such as arterial bifurcations, despite the fact that the entire arterial tree is exposed to systemic risk factors.¹⁰ These distinct regions are characterised by low wall shear stress (WSS), which is known to enhance atherogenic molecular, cellular, and vascular responses.¹² A low shear-dependent mass transfer mechanism for atherogenesis was first proposed by Caro *et al*,^{13,14} and it was later demonstrated that cholesterol accumulates in low WSS arterial regions because of the inhabitation of diffusional efflux from the arterial wall to the intraluminal blood due to the reduced concentration gradient.¹³ This formed the understanding that WSS directly modulates the haemodynamic environment of the arterial wall and can enhance the predilection for atherosclerosis in localised regions.¹⁵ Subsequent studies validated this hypothesis, whereby low WSS (<0.5 Pa) was found to stimulate an atherogenic endothelial phenotype, characterised by greater endothelial proliferation under the influence of vasoconstrictors and mitogenic substances such as endothelin I, angiotensin II and platelet-derived growth factor B, apoptogenic stimuli such as oxidised LDL and tumour necrosis factor α , inflammatory mediators such as monocyte chemoattractant peptide 1 and adhesion molecules such as vascular cell adhesion molecule 1.^{16,17} Later, in addition to instantaneous low WSS, cardiac cycle time-averaged low WSS was also identified as a key regulator in the vascular pathophysiology of atherosclerosis.¹⁸ As such, it is increasingly recognised that haemodynamic factors can form a valuable indicator for higher accuracy cardiovascular risk prediction beyond commonly used clinical risk scores.

It is important to notice that coronary anatomy governs the localised development of WSS within the arterial tree

Box 1 Candidate anatomical biomarkers and haemodynamic variables for coronary artery disease

Geometric biomarkers

- ⇒ Flow divider which is offset from the aortic axis.
- ⇒ Inward curvature.
- ⇒ Marked angulating daughter branches.
- ⇒ Asymmetrical T-shaped bifurcation.
- ⇒ Bifurcation angle.
- ⇒ Cardiac curvature.
- ⇒ Vessel diameter.
- ⇒ Inflow angle.
- ⇒ Tortuosity.

Haemodynamic parameters

- ⇒ Wall Shear Stress (WSS).
- ⇒ Time-averaged WSS.

and thus mediates the endothelial response,¹⁵ formulating the 'Anatomy of Risk' hypothesis.^{10,12} While haemodynamic factors are difficult to assess in vivo, coronary anatomical characteristics are apparent in standard medical images and may offer a pathway into future integration into standard clinical CAD risk assessments.

The concept of arterial geometric risk was first proposed by Friedman *et al* in a study of pulsatile flow through casts of human aortic bifurcations in 1983,⁶ which identified geometric bifurcations features causing significant variability in WSS distribution. Recent computational studies have built on Friedman's early work, leading to the discovery of several anatomical features which can significantly influence WSS (Box 1).^{19–26} Despite the progress in recent years, investigating the link between coronary haemodynamics and clinical outcomes remains critical to our understanding of anatomical risk and is likely directly relevant to identifying individuals without SMuRFs at risk of developing CAD.

Meaningful progress towards such understanding has been hindered by the lack of advanced imaging technology and computational resources, prohibiting large-scale population studies until recently. The evolution of computed tomography coronary angiography (CTCA) technology with improved spatial and temporal resolution has enabled a wide range of new applications in the field of preventive cardiology, such as the integration of coronary artery calcium scoring with clinical risk equations, with incremental predictive value for CAD risk.^{27,28} Combined with the increase in processing power and storage facilitating high-fidelity (mainly medical images-based) big data efforts coupled with the rise of machine learning approaches, fast and practical automated systems for better CAD risk assessment are now not a distant vision but a near future opportunity.²⁹ Traditional machine learning methods (logistic regression, k-nearest neighbours, support vector machines, tree-based algorithms) have previously been used for risk stratification.^{30–32} More recent methods, including deep neural networks, now outperform these earlier attempts.^{33–36} These latest

developments in the field are thus a powerful framework for the translation of advanced imaging analyses into clinical CAD risk assessment practice.

Still, cardiac CT requires unfavourable radiation exposure and some studies attempted to leverage non-cardiac imaging to investigate CAD risk factors.^{37–39} Deep learning models have shown promising results in using low-dose CT imaging for lung cancer screening,³⁷ and risk factors such as blood pressure, smoking history and diabetes, have been successfully identified in retinal vasculature from retinal images only,³⁸ showing correlation with CAD risk and all-cause mortality.³⁹ This showcases the potential for general investigation of the anatomy of risk and patient-specific image-derived biomarkers, as these may not just be linked to cardiac CT but can also be deployed to a range of available imaging modalities.

Other noteworthy approaches in better CAD risk prediction includes machine learning systems including systemic lifestyle factors combined with data from wearable devices together with traditional risk factors,⁴⁰ and a similar deep learning system, aimed at including localised markers by automatically predicting coronary artery calcium scores.⁴¹ These works showcase the potential of such efforts, which may be especially relevant when considering better risk assessments for specific subgroups including more vulnerable populations.^{8,9}

Here, we propose a novel approach to build on this previous knowledge and to non-invasively determine the relationship between shape features, WSS and the risk of clinical endpoints in a large population, with the aim to generate a superior CAD risk prediction model. To the best of our knowledge, vessel geometry and its haemodynamic impact has not been accounted for in CAD risk models to date, and our approach thus offers an unprecedented opportunity to study detailed anatomical biomarkers driving haemodynamic processes linked to CAD in addition to calcium scoring and standard risk assessment. State-of-the-art machine learning methods will be applied to develop a practical system to generate new insights into previously unexplained susceptibility in many individuals without SMuRFs. Our expert team is well positioned to build such a sophisticated CAD risk model using machine learning algorithms. Specifically, SB and team previously developed the Coronary Atlas, the world's first and largest three-dimensional CT computational atlas describing the detailed statistical anatomy of the coronary tree.^{11 42–44} This led to the introduction of a new coronary shape parameter—the inflow angle, defined as the angle with which the proximal vessel enters the bifurcation plane, as well as the first classification of coronary shape features.^{11 43} The Coronary Atlas provides a systematic and comprehensive framework to integrate large-scale datasets from multiple individuals and to generate new insights into the relationship between coronary anatomy and WSS patterns, which we then successfully predict directly using machine learning.^{22 45} This has elucidated the understanding of WSS in individuals with direct implications for individual CAD susceptibility and

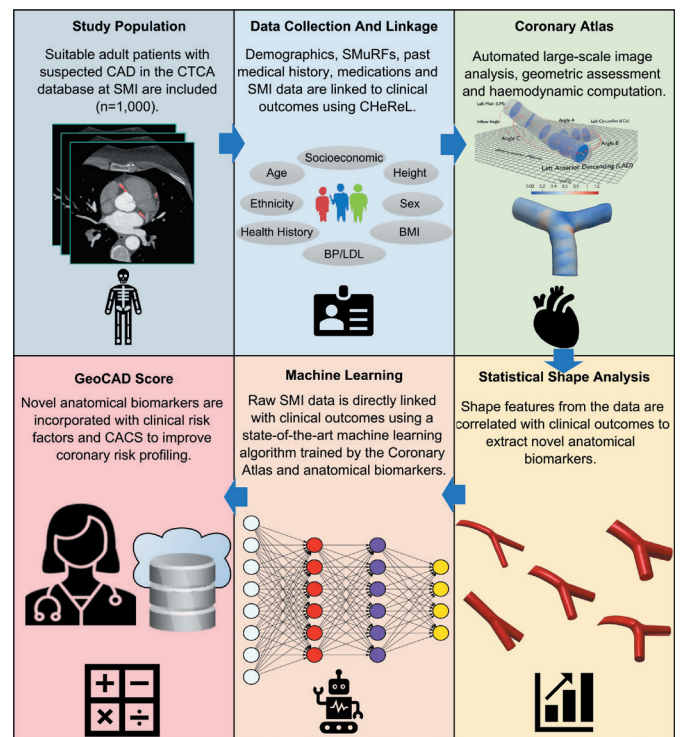


Figure 1 GeoCAD study flow chart. BMI, body mass index; BP, blood pressure; CACS, coronary artery calcium score; CAD, coronary artery disease; CHeReL, Centre for health record linkage; CTCA, CT coronary angiography; LDL, low-density lipoprotein; SMI, spectrum medical imaging; SMuRF, standard modifiable risk factor.

underpins the current proposal to address the gap in our understanding of anatomical risk for CAD. The identification of susceptible individuals and the early implementation of targeted therapies based on patient-specific data may take us one step closer to the Holy Grail of preventive cardiology.

METHODS AND ANALYSIS

Patient and public involvement

Patients/the public were not directly involved in the research. However, the concept of the study was designed to address the gap in our understanding of susceptibility to CAD in the one quarter of individuals without standard clinical risk factors who suffer from unexplained cardiovascular events. The study outcomes will be disseminated in peer-reviewed journals, scientific conferences and as a higher degree research thesis, which will provide a powerful framework to translate the findings into clinical practice to improve coronary risk profiling in the general population.

Objectives

The primary objective of the GeoCAD study is:

1. To identify novel anatomical biomarkers to improve the accuracy of CAD risk prediction.

The secondary objectives of the GeoCAD study are (figure 1):

1. To profile CTCA images of a large population with respect to variations in anatomical shape and associated haemodynamic risk, comprising an individual's anatomical risk.
2. To develop a machine-learning algorithm for the rapid assessment of anatomical risk directly from unprocessed CTCA images.
3. To develop a novel CAD risk model combining traditional risk factors with anatomical risk.

Study type

GeoCAD is a retrospective cohort study (figure 1). It is a proof-of-concept study to test the hypothesis that advanced image-derived patient-specific information can accurately predict long-term cardiovascular events.

Study population

Retrospectively, 1000 adult patients referred for CTCA due to suspected CAD will be identified from the CTCA database at Spectrum Medical Imaging, Sydney, Australia. We will identify patients who have undergone at least two CTCA scans from 2010 onwards (due to available CTCA image resolution) to allow comparison of geometry and plaque features over time. We will use the oldest records available to allow for a longer follow-up period. The patients will be selected and screened and patients who meet all of the inclusion criteria and none of the exclusion criteria will be selected for the study.

Inclusion criteria

- ▶ Patients who were referred for at least two CTCA scans for investigation of suspected CAD from 2010 onwards at Spectrum Medical Imaging.
- ▶ Age: 18 years or older.

Exclusion criteria

- ▶ Patients who have had a prior myocardial infarction (MI), percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG).

Data collection

Imaging and associated data will be collected from Spectrum Medical Imaging and will include the following:

- ▶ CTCA digital imaging and communication in medicine (DICOM) files.
- ▶ Coronary dominance.
- ▶ Presence or absence of the ramus intermediate artery.
- ▶ Coronary artery calcium score.
- ▶ Location, severity and plaque composition of all lesions according to the 16-segment American Heart Association classification.⁴⁶

Clinical data will be collected from Spectrum Medical Imaging and from administrative datasets linked by the NSW Centre for Health Record Linkage (CHeReL) (Admitted Patient Data Collection (APDC), the Registry of Births, Deaths and Marriages, and the Australian Coordinating Registry Cause of Death Unit Record File). APDC records include contain diagnoses coded according to the International Classification of Diseases, 10th Revision,

Australian Modification and procedures coded according to the Australian Classification of Health Interventions. Clinical data will include the following:

Demographic data (age, sex).

SMuRFs (hypertension, diabetes mellitus, dyslipidaemia, smoking).

A medical history (eg, prior MI, PCI or CABG).

Medication history.

5. Clinical outcomes (all-cause death, cardiovascular death, coronary angiography, hospitalisation for heart failure, non-fatal MI, non-fatal stroke, revascularisation and unstable angina requiring hospitalisation).

7. Major adverse cardiovascular events will be defined as cardiovascular death, non-fatal MI and non-fatal stroke.

Data governance

Data management practices will follow the principles of the Australian Code for the Responsible Conduct of Research. A research data management plan for the project has been established and managed using the University of New South Wales (UNSW) ResToolkit platform. All research data will be classified according to UNSW Classification Standards and handled in accordance to UNSW data handling guidelines.

Appropriate cases matching the inclusion and exclusion criteria will be selected and their accession numbers noted. DICOM files and reports for cases will be downloaded from a central repository at Spectrum Medical Imaging to a local server inside the firewall. DM will semiautomatically anonymise and copy the data to secure password protected storage on UNSW servers through an encrypted channel. DM will not be involved in the analysis of linked data. The researchers analysing the data will have only access to the anonymised data. The provided data will be transferred to the Data Archive provisioned for this project (RDMP ID: D0240165), rated as appropriate for sensitive data, using the Data Archive web application. Data on UNSW Data Archive are encrypted and access to UNSW Data Archive is password protected and require connection to UNSW's virtual private network (VPN) with a valid university account.

The imaging data will be securely linked with the CHeReL datasets as follows:

1. Splitting, data integration and disclosure: Identifying information such as name, address and date of birth is separated from content information such as imaging data. All participants will be assigned an arbitrary Person Number which replaces identifying information. A research project-specific person number (PPN) will be made for each participant using an encrypted version of the arbitrary person number. All records for a participant will have the same PPN.
2. Creating a research dataset: Using the PPN, the research team can combine records for a participant without accessing identifying information. The data are made available to the analysing research team in a non-identifiable format.

Data analysis plan

Shape features

It is important to note that the analysis of the vessel geometry and its haemodynamics in the same patient years apart will provide critical and unprecedented insights into the development of stable CAD, allowing for the comparison of arterial geometry and plaque changes over time to elucidate the role of haemodynamics. Deep learning methods have gained significant popularity in image segmentation and analysis, particularly due to the success of U-Net in segmenting medical images.⁴⁷ Virtual models of the coronary anatomy will be reconstructed from the CTCA image using deep convolutional neural networks based on nnU-Net architecture,⁴⁸ as this method has been shown to work well in automated coronary artery segmentation.⁴⁹ After Taubin's algorithm smoothing and vessel centrelines extraction with Vascular Modelling Toolkit,⁵⁰ relevant geometric arterial tree features will be quantified using in-house python scripts. This includes the median branch diameters, tortuosities, curvature (Frenet-Serret formulas with the average curvature used for analysis).^{51,52} The processing time for each case is approximately 2 min on a single core 2.9GHz Xeon ES-2670.

Haemodynamic indicators

Haemodynamics will be computed using validated machine learning models,⁴⁵ taking less than one minute per case on a single core 2.9GHz Xeon ES-2670. This allows the generation of haemodynamic risk indicators based on vessel geometry, avoiding the need for high computation cost associated with standard computational modelling. Transient simulations will be used to investigate pulsatile flow conditions throughout the cardiac cycle. Non-Newtonian behaviour of blood will be accounted for using the Carreau-Yasuda viscosity model.⁵³ The haemodynamic modelling follows experts' recommendations for coronary modelling.⁵⁴

Machine learning

Building on our previous machine learning haemodynamics predictions from reconstructed models,⁴⁵ additional features such as demographic information and medical history will be incorporated into the model to improve the prediction accuracy. Locally connected layers,⁵⁵ will be used to build 2D feature maps from the global shape, clinical and demographics information, generating feature maps that can appropriately model the effect of this information in different regions of the bifurcation. Convolutional neural network layers are used to predict haemodynamic metrics, vessel response and expected disease development over the surface of the coronary vessels. The deep learning model will be used to generate pixelwise predictions, which can be correlated against the follow-up imaging to investigate localised plaque growth and progression based on haemodynamic descriptors, as well as overall risk metrics which will be evaluated versus the all-cause mortality. Additionally, random forest models⁵⁶ will be trained on the same data to

investigate performance of traditional machine learning methods versus deep learning, and potentially provide a more interpretable risk model. The performance of the trained models will be evaluated and compared using 10-fold cross validation. The Area Under Receiver Operating Characteristics Curve⁵⁷ metric will be used to compare predictions of the machine learning models to existing literature on machine learning risk models⁴¹ as well as traditional models. This allows for easy comparisons against other models as it is commonly reported and simple to interpret.

Statistical analysis plan

Additional statistical analysis will explore the relationships between our developed non-traditional potential risk factors and clinical endpoint data. Continuous variables will be presented as mean (\pm SD) and categorical variables as proportions (%). Comparisons between groups will be performed using independent Student's t-tests with Bonferroni correction for continuous variables and χ^2 or Fisher's exact tests for categorical variables. Univariate and multivariate analyses will be performed using Mantel-Haenszel logistic regression. Univariate variables with $p < 0.10$ will be included in the multivariate analysis. The discriminative performance of the multivariable model will be assessed using Harrell's c-statistic. Comparisons between the multivariable models will be assessed using net reclassification index. A two-tailed $p < 0.05$ with Bonferroni correction will be considered significant. Our sample size of 1000 will be sufficient because we estimated that we will need a sample size of at least 445 patients to show that a c-statistic of 0.80 is significantly different from the null hypothesis (assuming a c-statistic of 0.71 for the Framingham risk score), considering a p value of 0.05, power of 80% and event rate of 20%.

Ethics and dissemination

The study protocol has been approved by the St Vincent's Hospital Human Research Ethics Committee, Sydney—2020/ETH02127 and the NSW Population and Health Service Research Ethics Committee—2021/ETH00990. The committee granted a waiver of the usual requirement of consent. The project outcomes will be published in peer-reviewed and biomedical journals, scientific conferences and as a higher degree research thesis. Patient confidentiality will be maintained by not including any individually identifying information in publications. Non-identifiable data (statistical shape analyses and haemodynamic simulations) will be shared with other researchers on the Coronary Atlas website. We will not share any raw imaging data or unit record data with other researchers.

DISCUSSION

The role of adverse anatomical features in CAD risk remains unclear. Several studies have suggested that bifurcation angle (figure 2), defined as the angle between the daughter vessels after branching, is a geometric risk factor

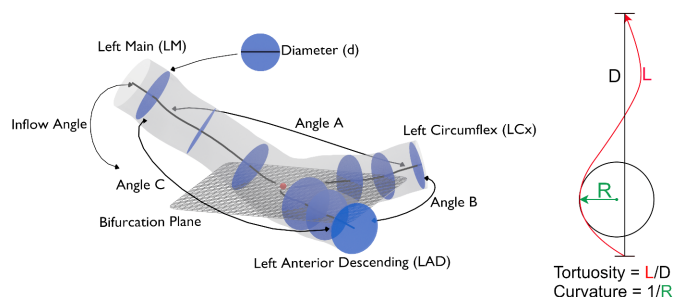


Figure 2 Three-dimensional representation of candidate anatomical biomarkers: (1) bifurcation angle (angle B), defined as the angle between the daughter vessels after branching, (2) inflow angle, defined as the angle with which the proximal vessel enters the bifurcation plane, (3) diameter, (4) curvature (1/radius) and (5) tortuosity (length/diameter).

for atherosclerosis.^{19–21} However several later studies have shown that bifurcation angle alone has minimal haemodynamic impact,^{22–24} and that in fact the combination with other shape characteristics (inflow angle, diameter and tortuosity) determine either a stronger or mitigating effects on WSS. Others showed that vessel tortuosity,^{23 58} curvature²⁴ and cross-sectional area,^{25 26} may also play a role in local WSS development.⁵⁹ Overall, inconsistent observations of geometric parameters in the literature suggest that anatomical risk factors remain little understood, possibly due to their complex three-dimensional structure with interdependent haemodynamic impact of several shape characteristics.²²

Current absolute cardiovascular disease risk calculators in Australia are based on the Framingham risk equation.¹ The model was developed to estimate an individual's five-year and 10-year risk of cardiovascular disease using a point-score algorithm including clinical risk factors (age, female sex, systolic blood pressure, total cholesterol, high-density lipoprotein cholesterol, smoking, diabetes and electrocardiographic left ventricular hypertrophy). A recent meta-analysis of validation studies evaluating the discriminative performance of the 10-year Framingham risk model found a pooled c-statistic of 0.68 (95% CI 0.66 to 0.69) to 0.71 (95% CI 0.66 to 0.76).⁵ From this modest discriminative power, it becomes clear that the adverse cardiovascular events in one-in-four patients remains unexplainable by the Framingham risk model, and that there is an urgent need to identify the remaining risk factors for atherosclerosis. Indeed, a recent study using two large multi-centre Australian registries showed that a substantial and increasing proportion of STEMI patients were individuals without SMuRFs.⁷ Moreover, 19% of patients were SMuRF-less, and this proportion increased from 14% to 23% during the study period. Concerningly, SMuRF-less patients had a higher in-hospital mortality rate than patients with one or more SMuRF (6% vs 4%, $p=0.032$). It is likely that advanced image-derived patient-specific information can account for some of these unexplained susceptibilities to atherosclerosis in SMuRF-less individuals, and even be detected through imaging analysis.

CTCA technology already has a well-established role in the field of preventive cardiology. The Scottish CT of the Heart and Prospective Multicentre Imaging Study for Evaluation of Chest Pain trials were landmark studies, showing that a CTCA-guided strategy improves clinical outcomes in symptomatic patients with stable angina, increasing the diagnostic certainty and frequency of CAD and the subsequent implementation of appropriate secondary prevention and revascularisation.^{60–62}

Still, the role of CTCA in asymptomatic patients with CAD remains somewhat uncertain. The Factor-64 trial has been the only randomised clinical trial to date to assess the prognostic value of routine CTCA screening for CAD in this population.⁶³ More than 900 high-risk diabetic patients were randomised for either CTCA or standard national guidelines-based optimal medical care, whereby, at four years follow-up, there was no difference in the primary outcome of death, non-fatal MI or unstable angina requiring hospitalisation. However, the trial was not adequately powered due to a lower than anticipated event rate. Similarly, a meta-analysis evaluating the prognostic value of CTCA in more than 6000 diabetic patients, whereby two-thirds were asymptomatic, observed a higher hazard ratio for obstructive CAD if revascularisation was included as an endpoint—meaning that CTCA in some of this population could have important prognostic implications.⁶⁴ Still, registry studies in broader asymptomatic populations have also suggested that CTCA findings (location, severity and plaque composition) have incremental prognostic utility beyond traditional risk factors alone.⁶⁵

Several studies have demonstrated the predictive value of the coronary artery calcium score in addition to traditional risk factors for CAD.^{27 28} The South Bay Heart Watch Study found that a calcium score higher than 300 combined with the Framingham risk score significantly improved the discriminative ability (c-statistic 0.68 vs 0.63, $p<0.001$).²⁷ Similarly, the St. Francis Heart Study showed that coronary artery calcium score was superior to the Framingham risk index for the prediction of atherosclerotic cardiovascular disease events (c-statistic 0.79 vs 0.69, $p=0.0006$).²⁸ It should also be noted that the distribution of calcium was found to be more significant in predicting cardiovascular events than the calcium score alone.^{66 67} Specifically, in more than 1200 participants from the Offspring and Third Generation cohorts of the Framingham Heart Study, it was shown that the number of coronary arteries with calcium, and especially the presence of calcium in the proximal dominant coronary artery, independently predicted coronary heart disease after adjustment for the Framingham risk score and coronary artery calcium score.⁶⁷ The addition of calcium distribution improved the discriminatory capacity of the multivariable model with the Framingham risk score and calcium score for coronary heart disease events (c-statistic 0.79 to 0.80 vs 0.77, relative integrated discriminatory index 0.14). This study confirmed the observations of an earlier analysis of 3262 participants in

the Multi-Ethnic Study of Atherosclerosis cohort, which showed that diffusely distributed calcium, as assessed by the number of coronary arteries with calcified plaque, significantly improved the capacity to predict cardiovascular events beyond the calcium score (c-statistic 0.67 vs 0.64, $p=0.0001$).⁶⁶

Beyond calcium scoring, machine learning-based approaches have been the latest focus of the field and enable the effective processing of even very large datasets with promising potential for cloud-based clinical integration. However, key challenges in such an undertaking are the comparability, and reproducibility across different clinical cohorts, imaging specifications and scan protocols, and of course most importantly, the assurances of patient confidentiality and data security.⁶⁸

Machine learning methods have been predominantly used in conjunction with medical images and other medical data^{69 70} to train multiple non-linear classifiers (support vector machine, logistic regression, tree-based models, deep neural networks) to predict mortality rates.^{71 72} CTCA applied deep learning applications allowed detection and quantification of calcified plaques,^{73–75} as well as correlating calcium score to mortality.⁴¹ Standard blood test results are also often included in machine learning models for risk stratification.⁷⁶

While promising, these machine learning methods are not matured enough to replace the traditional Framingham score,⁷⁷ and further research and exploration of the field is required. Existing machine learning methods usually rely on generalised adverse features for CAD risk prediction which may lead to low reproducibility.⁶⁸ Additionally, current machine learning approaches,^{37–41 71 72} focus primarily on systemic risk factors. This does not consider the observed trends that particular locations within the coronary tree, for example, bifurcations,¹⁰ are at significantly higher risk of disease. More advanced comprehensive machine learning risk prediction and intervention recommendation systems are at an early stage of algorithm development, and to our knowledge, there is no prior work on a comprehensive machine learning incorporating haemodynamic information within CAD risk models.

In summary, there is a tremendous opportunity to improve the accuracy of CAD risk prediction by integrating additional patient-specific anatomical risk with traditional risk models. We hope that incorporating haemodynamic metrics, which can provide significantly more granular information beyond the traditionally used models can better predict the expected vessel response and future outcomes. The use of anatomical surrogate markers for CAD will enable us to extend the application of CTCA-guided risk prediction from diseased individuals to normal populations without atherosclerosis, generate new understandings of disease mechanisms and its development in individuals, and open future pathways for application to imaging modalities without or with reduced radiation. This unprecedented opportunity has been underpinned by advanced imaging analysis,

sophisticated computational technology and state-of-the-art machine learning algorithms, which offer a fast and practical approach for CAD risk assessment in large-scale population studies. Understanding the mechanism of personal susceptibility to atherosclerosis opens up the opportunity for early implementation of targeted therapies and may be a key in addressing the growing burden of CAD, especially in individuals without SMuRFs.

Author affiliations

¹Faculty of Medicine, The University of New South Wales, Sydney, New South Wales, Australia

²Cardiology Department, The Prince of Wales Hospital, Sydney, New South Wales, Australia

³School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, New South Wales, Australia

⁴Centre for Big Data Research in Health, The University of New South Wales, Sydney, New South Wales, Australia

⁵School of Computer Science and Engineering, The University of New South Wales, Sydney, New South Wales, Australia

⁶Department of Medical Imaging, The Prince of Wales Hospital, Sydney, New South Wales, Australia

Contributors DA contributed to the study design, drafting the manuscript and revising it critically for important intellectual content. RG, SZ and DM contributed to revising the manuscript. LJ, AS and SO contributed to final approval of the version to be published. SB has supervised the process and assisted in the manuscript draft and revisions. All authors contributed to the study design and conception, revising the manuscript critically for important intellectual content and final approval of the version to be published.

Funding DA is supported by an Australian Government research training programme scholarship. Award/grant number not applicable. This research is supported by NHMRC Ideas grant (2012474) and the NSW Cardiovascular Research Capacity Programme Early-Mid Career (EMC) Researcher Grant (EMC78).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Dona Adikari <http://orcid.org/0000-0002-8166-3222>

REFERENCES

- Anderson KM, Wilson PW, Odell PM, *et al*. An updated coronary risk profile. A statement for health professionals. *Circulation* 1991;83:356–62.
- World Health Organization. *World health statistics 2019: monitoring health for the SDGs, sustainable development goals*. Geneva: World Health Organization, 2019.
- Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian simvastatin survival study (4S). *Lancet* 1994;344:1383–9.
- Cannon CP, Blazing MA, Giugliano RP, *et al*. Ezetimibe added to statin therapy after acute coronary syndromes. *N Engl J Med* 2015;372:2387–97.
- Damen JA, Pajouheshnia R, Heus P, *et al*. Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. *BMC Med* 2019;17:109.

- 6 Friedman MH, Deters OJ, Mark FF, *et al.* Arterial geometry affects hemodynamics. A potential risk factor for atherosclerosis. *Atherosclerosis* 1983;46:225–31.
- 7 Vernon ST, Coffey S, D'Souza M, *et al.* St-Segment-Elevation myocardial infarction (STEMI) patients without standard modifiable cardiovascular risk Factors-How common are they, and what are their outcomes? *J Am Heart Assoc* 2019;8:e013296.
- 8 Pylpchuk R, Wells S, Kerr A, *et al.* Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study. *Lancet* 2018;391:1897–907.
- 9 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357:j2099.
- 10 Morbiducci U, Kok AM, Kwak BR, *et al.* Atherosclerosis at arterial bifurcations: evidence for the role of haemodynamics and geometry. *Thromb Haemost* 2016;115:484–92.
- 11 Medrano-Gracia P, Ormiston J, Webster M, *et al.* A computational atlas of normal coronary artery anatomy. *EuroIntervention* 2016;12:845–54.
- 12 Antoniadis AP, Mortier P, Kassab G, *et al.* Biomechanical modeling to improve coronary artery bifurcation stenting: expert review document on techniques and clinical implementation. *JACC Cardiovasc Interv* 2015;8:1281–96.
- 13 Caro CG, Fitz-Gerald JM, Schroter RC. Atheroma and arterial wall shear. observation, correlation and proposal of a shear dependent mass transfer mechanism for atherogenesis. *Proc R Soc Lond B Biol Sci* 1971;177:109–59.
- 14 Caro CG. Discovery of the role of wall shear in atherosclerosis. *Arterioscler Thromb Vasc Biol* 2009;29:158–61.
- 15 Nerem RM. Vascular fluid mechanics, the arterial wall, and atherosclerosis. *J Biomech Eng* 1992;114:274–82.
- 16 Malek AM, Alper SL, Izumo S. Hemodynamic shear stress and its role in atherosclerosis. *JAMA* 1999;282:2035–42.
- 17 Friedman MH, Hutchins GM, Bargeron CB, *et al.* Correlation between intimal thickness and fluid shear in human arteries. *Atherosclerosis* 1981;39:425–36.
- 18 Dolan JM, Kolega J, Meng H. High wall shear stress and spatial gradients in vascular pathology: a review. *Ann Biomed Eng* 2013;41:1411–27.
- 19 Chaichana T, Sun Z, Jewkes J. Computation of hemodynamics in the left coronary artery with variable angulations. *J Biomech* 2011;44:1869–78.
- 20 Dong J, Sun Z, Inthavong K, *et al.* Fluid-structure interaction analysis of the left coronary artery with variable angulation. *Comput Methods Biomech Biomed Engin* 2015;18:1500–8.
- 21 Ikeda U, Kuroki M, Ejiri T, *et al.* Stenotic lesions and the bifurcation angle of coronary arteries in the young. *Jpn Heart J* 1991;32:627–33.
- 22 Beier S, Ormiston J, Webster M, *et al.* Impact of bifurcation angle and other anatomical characteristics on blood flow - A computational study of non-stented and stented coronary arteries. *J Biomech* 2016;49:1570–82.
- 23 Malvé M, Gharib AM, Yazdani SK, *et al.* Tortuosity of coronary bifurcation as a potential local risk factor for atherosclerosis: CFD steady state study based on in vivo dynamic CT measurements. *Ann Biomed Eng* 2015;43:82–93.
- 24 Chiastra C, Gallo D, Tasso P, *et al.* Healthy and diseased coronary bifurcation geometries influence near-wall and intravascular flow: a computational exploration of the hemodynamic risk. *J Biomech* 2017;58:79–88.
- 25 Pinho N, Castro CF, António CC, *et al.* Correlation between geometric parameters of the left coronary artery and hemodynamic descriptors of atherosclerosis: FSI and statistical study. *Med Biol Eng Comput* 2019;57:715–29.
- 26 Pinho N, Sousa LC, Castro CF, *et al.* The impact of the right coronary artery geometric parameters on hemodynamic performance. *Cardiovasc Eng Technol* 2019;10:257–70.
- 27 Greenland P, LaBree L, Azen SP, *et al.* Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. *JAMA* 2004;291:210–5.
- 28 Arad Y, Newstein D, Roth M, *et al.* Rationale and design of the St. Francis heart study: a randomized clinical trial of atorvastatin plus antioxidants in asymptomatic persons with elevated coronary calcification. *Control Clin Trials* 2001;22:553–72.
- 29 Liang L, Liu M, Martin C, *et al.* A machine learning approach to investigate the relationship between shape features and numerically predicted risk of ascending aortic aneurysm. *Biomech Model Mechanobiol* 2017;16:1519–33.
- 30 Singh A, Gutttag JV. A comparison of non-symmetric entropy-based classification trees and support vector machine for cardiovascular risk stratification. In: *2011 annual International Conference of the IEEE engineering in medicine and biology Society*. IEEE, 2011.
- 31 Khateeb N, Usman M. Efficient heart disease prediction system using k-nearest neighbor classification technique. *Proceedings of the international conference on big data and internet of thing*, 2017.
- 32 Colombet I. Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression. *Proceedings of the AMIA Symposium*, 2000.
- 33 Krittanawong C, Virk HUH, Kumar A, *et al.* Machine learning and deep learning to predict mortality in patients with spontaneous coronary artery dissection. *Sci Rep* 2021;11:1–10.
- 34 Sajeev Set *al.* Deep Learning to improve heart disease risk prediction. In: *Machine learning and medical engineering for cardiovascular health and intravascular imaging and computer assisted stenting*. Springer, 2019: 96–103.
- 35 Bharti R, Khamparia A, Shabaz M, *et al.* Prediction of heart disease using a combination of machine learning and deep learning. *Comput Intell Neurosci* 2021;2021:8387680.
- 36 Tison GH, Singh AC, Ohashi DA, *et al.* Abstract 21042: cardiovascular risk stratification using off-the-shelf Wearables and a Multi-Task deep learning algorithm. *Circulation* 2017;136:A21042.
- 37 Chao H, Shan H, Homayounieh F, *et al.* Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. *Nat Commun* 2021;12:1–10.
- 38 Poplin R, Varadarajan AV, Blumer K, *et al.* Prediction of cardiovascular risk factors from retinal fundus Photographs via deep learning. *Nat Biomed Eng* 2018;2:158–64.
- 39 Cheung CY, Xu D, Cheng C-Y, *et al.* A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng* 2021;5:498–508.
- 40 Huang W, Ying TW, Chin WLC, *et al.* Application of ensemble machine learning algorithms on lifestyle factors and wearables for cardiovascular risk prediction. *Sci Rep* 2022;12:1–12.
- 41 Zeleznik R, Foldyna B, Eslami P, *et al.* Deep convolutional neural networks to predict cardiovascular risk from computed tomography. *Nat Commun* 2021;12:1–9.
- 42 Beier S. The coronary atlas, 2020. Available: <https://www.coronaryatlas.org/>
- 43 Medrano-Gracia P, Ormiston J, Webster M, *et al.* A study of coronary bifurcation shape in a normal population. *J Cardiovasc Transl Res* 2017;10:82–90.
- 44 Medrano-Gracia P, Ormiston J, Webster M, *et al.* Construction of a coronary artery atlas from CT angiography. *Med Image Comput Assist Interv* 2014;17:513–20.
- 45 Gharleghi R. Deep learning for time averaged wall shear stress prediction in left main coronary bifurcations. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020.
- 46 Austen WG, Edwards JE, Frye RL, *et al.* A reporting system on patients evaluated for coronary artery disease. Report of the AD hoc Committee for grading of coronary artery disease, Council on cardiovascular surgery, American heart association. *Circulation* 1975;51:5–40.
- 47 Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on medical image computing and computer-assisted intervention*. Springer, 2015.
- 48 Isensee F, Jaeger PF, Kohl SAA, *et al.* nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021;18:203–11.
- 49 Gharleghi R, Adikari D, Ellenberger K, *et al.* Automated segmentation of normal and diseased coronary arteries - The ASOCA challenge. *Comput Med Imaging Graph* 2022;97:102049.
- 50 Antiga L, Steinman DA. Robust and objective decomposition and mapping of bifurcating vessels. *IEEE Trans Med Imaging* 2004;23:704–13.
- 51 Kashyap V, Gharleghi R, Li DD, *et al.* Accuracy of vascular tortuosity measures using computational modelling. *Sci Rep* 2022;12:1–10.
- 52 Crenshaw HC, Edelstein-Keshet L. Orientation by helical motion—II. changing the direction of the axis of motion. *Bull Math Biol* 1993;55:213–30.
- 53 Razavi A, Shirani E, Sadeghi MR. Numerical simulation of blood pulsatile flow in a stenosed carotid artery using different rheological models. *J Biomech* 2011;44:2021–30.
- 54 Gijssen F, Katagiri Y, Barlis P, *et al.* Expert recommendations on the assessment of wall shear stress in human coronary arteries: existing methodologies, technical considerations, and clinical applications. *Eur Heart J* 2019;40:3421–33.
- 55 Chen Y-h, Moreno IL, Sainath T, *et al.* Locally-connected and convolutional neural networks for small footprint speaker recognition. *Interspeech* 2015.

- 56 Boulesteix A-L, Janitza S, Kruppa J, *et al.* Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2012;2:493–507.
- 57 Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 2007;115:654–7.
- 58 Buradi A, Mahalingam A. Impact of coronary tortuosity on the artery hemodynamics. *Biocybern Biomed Eng* 2020;40:126–47.
- 59 Doutel E, Pinto SIS, Campos JBLM, *et al.* Link between deviations from Murray's law and occurrence of low wall shear stress regions in the left coronary artery. *J Theor Biol* 2016;402:89–99.
- 60 SCOT-HEART investigators. Ct coronary angiography in patients with suspected angina due to coronary heart disease (SCOT-HEART): an open-label, parallel-group, multicentre trial. *Lancet* 2015;385:2383–91.
- 61 Newby DE, Adamson PD, *et al.*, SCOT-HEART Investigators. Coronary CT angiography and 5-year risk of myocardial infarction. *N Engl J Med* 2018;379:924–33.
- 62 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.
- 63 Muhlestein JB, Lappé DL, Lima JAC, *et al.* Effect of screening for coronary artery disease using CT angiography on mortality and cardiac events in high-risk patients with diabetes: the FACTOR-64 randomized clinical trial. *JAMA* 2014;312:2234–43.
- 64 Celeng C, Maurovich-Horvat P, Ghoshhajra BB, *et al.* Prognostic value of coronary computed tomography angiography in patients with diabetes: a meta-analysis. *Diabetes Care* 2016;39:1274–80.
- 65 Cho I, Al'Aref SJ, Berger A, *et al.* Prognostic value of coronary computed tomographic angiography findings in asymptomatic individuals: a 6-year follow-up from the prospective multicentre international confirm study. *Eur Heart J* 2018;39:934–41.
- 66 Blaha MJ, Budoff MJ, Tota-Maharaj R, *et al.* Improving the CAC score by addition of regional measures of calcium distribution: multi-ethnic study of atherosclerosis. *JACC Cardiovasc Imaging* 2016;9:1407–16.
- 67 Ferencik M, Pencina KM, Liu T, *et al.* Coronary artery calcium distribution is an independent predictor of incident major coronary heart disease events: results from the Framingham heart study. *Circ Cardiovasc Imaging* 2017;10:e006592.
- 68 Benjamin MM, Rabbat MG. Machine learning-based advances in coronary computed tomography angiography. *Quant Imaging Med Surg* 2021;11:2208–13.
- 69 Kadem M, Garber L, Abdelkhalik M. Hemodynamic modeling, medical imaging, and machine learning and their applications to cardiovascular interventions. *IEEE Rev Biomed Eng* 2022;PP:RBME.2022.3142058.
- 70 Corrigan FE, Gleason PT, Condado JF, *et al.* Imaging for Predicting, Detecting, and Managing Complications After Transcatheter Aortic Valve Replacement. *JACC Cardiovasc Imaging* 2019;12): :904–20.
- 71 Samad MD, Ulloa A, Wehner GJ. Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning. *JACC Cardiovasc Imaging* 2019;12:681–9.
- 72 Ulloa Cerna AE, Jing L, Good CW, *et al.* Deep-learning-assisted analysis of echocardiographic videos improves predictions of all-cause mortality. *Nat Biomed Eng* 2021;5:546–54.
- 73 Kurkure U, Chittajallu DR, Brunner G, *et al.* A supervised classification-based method for coronary calcium detection in non-contrast CT. *Int J Cardiovasc Imaging* 2010;26:817–28.
- 74 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* 2017;37:615–25.
- 75 Martin SS, van Assen M, Rapaka S, *et al.* Evaluation of a deep Learning-Based automated CT coronary artery calcium scoring algorithm. *JACC Cardiovasc Imaging* 2020;13:524–6.
- 76 Li D, Xiong G, Zeng H, *et al.* Machine learning-aided risk stratification system for the prediction of coronary artery disease. *Int J Cardiol* 2021;326:30–4.
- 77 Tesche C, Brandt V. *Calling for a new Framingham: machine learning in cardiovascular risk Assessment—The key for improved outcome prediction.* Washington DC: American College of Cardiology Foundation, 2021: 626–8.