BMJ Open

Coronary artery disease risk prediction using machine learning with clinical variables and medical image-derived patient-specific insights: protocol for the retrospective GeoCAD cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-054881
Article Type:	Protocol
Date Submitted by the Author:	26-Jun-2021
Complete List of Authors:	Adikari, Dona; University of New South Wales; Prince of Wales Hospital and Community Health Services, Cardiology Gharleghi, Ramtin; University of New South Wales Moses, Daniel; University of New South Wales; Prince of Wales Hospital and Community Health Services, Cardiology Ooi, Sze-Yuan; University of New South Wales; Prince of Wales Hospital and Community Health Services, Cardiology Department Beier, Susann; University of New South Wales
Keywords:	Coronary heart disease < CARDIOLOGY, Computed tomography < RADIOLOGY & IMAGING, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

TITLE PAGE

Title: Coronary artery disease risk prediction using machine learning with clinical variables and medical image-derived patient-specific insights: protocol for the retrospective GeoCAD cohort study

First author: Dona Adikari MBBS[ab]

Second author: Ramtin Gharleghi BSc[a]

Third author: Daniel Moses BSc MBBS MEngSc PhD[ab]

Fourth author: Sze-Yuan Ooi MBBS MD[ab]

Fifth author: Susann Beier BE ME PhD[a]

[a] The University of New South Wales, Sydney, Australia

[b] Cardiology Department, The Prince of Wales Hospital, Sydney, Australia

Corresponding author:

Dr Dona Adikari

Eastern Heart Clinic, The Prince of Wales Hospital

Barker Street, Randwick NSW 2031

Email: z5283805@ad.unsw.edu.au

Word count: 4,088

Dates of the study: 09/06/2021 to 21/09/2022

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 recognizes that adverse anatomical features of coronary arteries enhance atherogenic haemodynamics which in turn mediate the localisation and progression of plaques. We propose a novel approach predicated on advanced computed tomography coronary angiography (CTCA) data and state-of-the-art machine learning methods to address the gap in our understanding of anatomical risk for CAD. The early implementation of personalised preventive therapies in susceptible individuals may be the key to addressing the growing burden of CAD.

Methods and analysis

GeoCAD is a retrospective cohort study in 5,000 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 profile CTCA images with respect to variations in anatomical shape and associated haemodynamic risk comprising an individual's anatomical risk, develop a machine-learning algorithm for the rapid assessment of anatomical risk directly from unprocessed CTCA images and build a novel CAD risk model combining traditional risk factors with novel anatomical biomarkers to improve the accuracy of CAD risk prediction.

Ethics and dissemination

The study protocol has been approved by the St Vincent's Hospital Human Research Ethics Committee, Sydney. The project outcomes will be published in peer-reviewed and biomedical journals, scientific conferences and as a higher degree research thesis.

ARTICLE SUMMARY

Strengths and limitations of this study

- GeoCAD is a retrospective cohort study to assess anatomical risk in 5,000 adult patients who have undergone computed tomography coronary angiography (CTCA) 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 cohort study which limits the external validity of the findings.

KEYWORDS

Cardiovascular events

Coronary artery disease

Computed tomography coronary angiography

Machine learning

Risk factors

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 later implementation of statin therapy and the 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 from which it has been inferred that there are other risk factors for atherosclerosis that have not been identified.[5-6] Even more concerningly, ST-segment elevation myocardial infarct (STEMI) patients without standard modifiable risk factors (SMuRFs) have been shown to have significantly worse in-hospital outcomes compared to those with one or more risk factor.[7] Contemporary scoring algorithms such as PREDICT in New Zealand and QRISK3 in the United Kingdom proved promising in improving the accuracy of cardiovascular risk estimation in vulnerable high-risk subpopulations by incorporating additional demographic predictors such as socioeconomic deprivation and ethnicity.[8, 9] Thus, there is a tremendous opportunity to shift the paradigm from intervention to a greater focus on a comprehensive approach to primary prevention of cardiovascular disease with demonstrated potential for improved risk prediction using additional risk factors for atherosclerosis.

Anatomical biomarkers and the haemodynamic risk which they encompass explain, at least partially, some of the variance in susceptibility to cardiovascular disease among individuals and may 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 biological response of the arterial wall.[11] Systemic risk factors have been compounded to create current probabilistic risk scores.[1] 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.[11] These distinct regions are characterised by low wall shear stress (WSS), which enhances atherogenic molecular, cellular and vascular responses.[12] A low shear-dependent mass transfer mechanism for atherogenesis was first proposed by Caro et al. in 1971.[13, 14] It was observed that early atherosclerotic lesions developed preferentially in regions which experienced low WSS along the outer wall of arterial bifurcations in a series of cadaver human arteries. This led to the conclusion that cholesterol accumulated in low WSS regions in arteries because its diffusional efflux from the arterial wall to intraluminal blood was inhibited by the reduced concentration gradient.

This formed the understanding that WSS directly modulates the haemodynamic environment of the arterial wall and can enhance the predilection for atherosclerosis in localized regions.[15] Subsequent studies validated this proposal whereby low WSS (<0.5 Pa) was found to stimulate an atherogenic endothelial phenotype, which is 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 chemotactic peptide 1 and adhesion molecules such as vascular cell

adhesion molecule 1.[16, 17] In addition to low WSS, time-averaged WSS (TAWSS) (<0.5 Pa) was also identified as a key regulator in the vascular pathophysiology of atherosclerosis.[18]

Recognising that WSS and the endothelial response is in turn mediated by the coronary anatomy measured through its geometric variables led to the so-called 'anatomy of risk' hypothesis.[11, 12] In essence, anatomy has direct effects on vascular fluid mechanics and the resulting local haemodynamic factors influence endothelial structure and function.[15] As such, it is increasingly accepted that haemodynamic factors may enable more accurate cardiovascular risk prediction beyond clinical risk scores. This concept of geometric risk factors was first proposed by Friedman et al. in a study of pulsatile flow through casts of human aortic bifurcations in 1983.[6] They identified four geometric features of arterial bifurcations with sufficient variability among individuals to cause significant variability in WSS distribution. These were a flow divider which was offset from the aortic axis, an inward curvature in the aorta as the flow divider was approached, a markedly angulating daughter branch and an asymmetrical T-shaped bifurcation. The data suggested that the localisation and progression of plaques in susceptible arterial segments with atherogenic haemodynamics is mediated by corresponding adverse geometric features. Furthermore, they proposed that these geometrical risk factors may contribute to the variance in disease susceptibility to atherosclerosis among individuals which is unexplained by systemic risk factors.

Recent computational studies have built on Friedman's early work, leading to the discovery of several geometric features which can significantly influence WSS (Table 1).[19-26] Despite the progress in recent years, there are areas in which further work is needed to provide important new information. In

particular, investigating the link between haemodynamics and clinical outcomes is critical to our understanding of anatomical risk and will be relevant to identifying individuals without SMuRFs at risk of developing CAD. Such work has previously been limited by the lack of advanced imaging data, computational resources and large-scale population studies. 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. One of these is the integration of coronary artery calcium score (CACS) with clinical risk equations, which has been shown to have incremental predictive value for CAD.[27, 28] In addition, the use of machine learning approaches has now made it feasible to investigate the relationship between shape features, haemodynamic parameters and clinical outcomes, enabling a fast and practical system for risk assessment.[29] This provides a powerful framework to translate advanced imaging analyses to clinical practice, using novel anatomical biomarkers to develop improved risk models for CAD.

We propose a novel approach predicated on our current understanding of clinical and additional demographic risk factors, CACS and machine learning methods to non-invasively determine the relationship between shape features, WSS and the risk of clinical endpoints in a large population. To the best of our knowledge, it is the first time that a machine learning approach has been applied to establish the link between cardiovascular outcomes and haemodynamics, predicted by detailed image-derived analysis. The use of advanced CTCA technology will overcome a key weakness of previous scoring algorithms which have been limited by the lack of additional patient-specific data and now offers an unprecedented opportunity to study detailed anatomical biomarkers for CAD other than CACS in

normal populations without manifest atherosclerosis. State-of-the-art machine learning methods can then be applied to develop a practical system to generate new insights into previously unexplained susceptibility in a large number of individuals without SMuRFs.

Our expert team is well positioned to build a sophisticated risk model to predict CAD using machine learning algorithms. We previously constructed the Coronary Atlas, the World's first and largest three-dimensional CT computational atlas describing the detailed statistical anatomy of the coronary tree.[10, 30, 31, 32] 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.[10, 31] 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 predicted directly using machine learning.[22, 33] This has directly contributed to our understanding of CAD and 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 to clinical practice in order 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 2). 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

5000 adult patients referred for CTCA for investigation of suspected CAD from 2010 onwards will be identified from the CTCA database at Spectrum Medical Imaging (SMI), Sydney. We will use the oldest records available to allow for a longer follow-up period. The first 6,000 consecutive patients from 2010 onwards will be recruited and screened to account for patients with exclusion criteria. The first 5,000 patients to meet all of the inclusion criteria and none of the exclusion criteria will be enrolled in the study.

Inclusion criteria:

- Patients who were referred for CTCA for investigation of suspected CAD from 2010 onwards at SMI
- 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 data will be collected from SMI and will include the following:

- CTCA digital imaging and communication in medicine (DICOM) files
- Dominance
- Presence or absence of the ramus intermediate artery
- CACS
- Location, severity and plaque composition of all lesions according to the 16-segment AHA classification.[34]

Clinical data will be collected from SMI and the Centre for Health Record Linkage (CHeReL) datasets (Admitted Patient Data Collection, the Registry of Births, Deaths and Marriages and the Australian Coordinating Registry Cause of Death Unit Record File). Data obtained from the CHeReL datasets will be coded based on the International Classification of Diseases, 10th Revision (ICD-10). Clinical data will include the following:

- Demographic data (age, sex).
- Standard modifiable risk factors (hypertension, diabetes mellitus, dyslipidaemia, smoking).
- Past medical history (e.g. prior MI, PCI or CABG).
- Medication history
- 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). Major adverse cardiovascular events (MACE) 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 will be established and managed using the University of New South Wales (UNSW) ResData platform. All research data will be classified according to UNSW Classification Standards and handled in accordance to UNSW data handling guidelines.

There is a central repository of CTCA images at SMI. We have clear guidelines on the cases that we will require as per the inclusion and exclusion criteria. Once we have a list of accession numbers we will download the DICOM files and reports to a local server inside the SMI firewall. We can then anonymise the cases within the SMI firewall and then copy the relevant parts of the anonymised cases to a password protected drive on a secure UNSW server for storage and analysis. We have written a MATLAB script to do this. UNSW Data Archive will be used for back-up.

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 is made available to the research team in a non-identifiable format.

Data analysis plan

Shape Features

Virtual models of the coronary anatomy will be reconstructed for each patient based on CTCA imaging. The left main (LM) bifurcation will be extracted and the relevent geometric features quantified using in-house python scripts. The extracted models will be smoothed using Taubin's algorithm to better represent the smoothness of arteries and remove artefacts. Vessel centrelines will be extracted using the Vascular Modelling Toolkit (VMTK).[35] Angles between vessels will be calculated based on the average centreline direction. For each vessel, the median diameter will be used for analysis. Tortuosity is measured for each vessel, defined as the length of the vessel divided by straight distance between the vessel end points. The curvature of vessel centerline will be measured according to the Frenet-Serret formulas with the average curvature used for analysis.[36]

Haemodynamic Indicators

Computational Fluid Dynamic (CFD) simulations will be carried out to establish blood flow patterns in the LM bifurcation for each patient. Transient simulations will be used to investigate flow conditions throughout the cardiac cycle with data from important time steps used for analysis. Non-newtonian behaviour of blood will be accounted for using the Carreu-Yasuda visocity model.[37] An automated workflow has been developed to handle setting up, solving and post-processing of CFD simulations, taking approximately 25 central processing unit hours for each patient. VMTK will be used to generate tetrahedral meshes with prismatic boundary layers and ANSYS CFX used for solving the

simulations. Settings used for CFD simulations are based on the expert recommendations as described in.[38]

Machine Learning

We have previously developed machine learning models to generate haemodynamic risk indicators based on the vessel geometry, avoiding the need for high computation cost associated with CFD.[33] Additional features such as demographic information and medical history will be incorporated into the model to improve the prediction accuracy. The performance of the machine learning model in predicting disease risk will be evaluated and compared with other risk models using 10-fold cross validation.

Statistical analysis plan

Continuous variables will be presented as mean (± standard deviation) and categorical variables as proportions (%). Comparisons between groups will be performed using independent student *t*-tests with Bonferroni correction for continuous variables and χ^2 or Fisher's exact tests for continuous 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 value <0.05 with Bonferroni correction will be considered significant.

We estimate that will need a sample size of 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), taking into account 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. 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. Statistical shape analyses and

haemodynamic simulations will be shared with other researchers on the Coronary Atlas website, GitHub and/or the Amazon Web Services (AWS) Public Dataset Program. We will not share any raw imaging data or unit record data with other researchers.

DISCUSSION

Several studies have suggested that bifurcation angle (Figure 3), defined as the angle between the daughter vessels after branching, is a geometric risk factor for atherosclerosis.[19-21] Two computational studies have found consistent observations between wider LM bifurcation angles and atherogenic haemodynamics.[19, 20] The first, showed that wider bifurcation angles (75° to 120°) correlated with lower WSS and the second, showed that wider-angled models (70° to 110°) strongly altered WSS distribution.[19, 20] Interestingly, a study investigating the high incidence of left anterior descending (LAD) artery disease in young MI patients, found that the LAD-left circumflex (LCx) angle was not significantly different in patients with stenotic LAD arteries compared to patients with normal arteries, whilst the LM-LAD angle was significantly wider in the stenotic group.[22]

The current state of evidence suggests that in stented populations with LM disease, there is a complex interaction between wider bifurcation angles as well as mechanical factors such as stent underexpansion and multiple layers of stent that confers an increased risk of adverse cardiovascular events.[39, 40] To the best of our knowledge, there are no studies evaluating the relationship between bifurcation angles and clinical outcomes in non-stented populations, which is critical to understanding the true biologic effect of the bifurcation angle and addressing the gap in our understanding of anatomical risk for CAD.

While much attention has been paid to the bifurcation angle and its relationship with WSS, several studies have shown that bifurcation angle alone has minimal haemodynamic impact.[22-24] One such study performed CFD in 101 models derived from CTCAs of asymptomatic subjects.[22] Other shape

characteristics (inflow angle, diameter and tortuosity) had stronger adverse effects on WSS distribution compared to bifurcation angle. A similar study found a strong correlation between tortuosity of the LM-LAD segment and low WSS in the proximal LAD as well as a weak correlation between tortuosity of the LM-LCx segment and low WSS in the proximal LCx.[23] There was no significant correlation between bifurcation angles and low WSS although this may have been due to the fact that all patients had similar bifurcation angles. Yet another study showed that cardiac curvature led to higher exposure to low WSS while bifurcation angle had a minor effect.[24]

Pinho et al. assessed several geometric parameters of the coronary arteries and their influence on WSS-based haemodynamic descriptors in the first statistical study of its kind using fluid-structure interaction simulations based on CTCA images.[25, 26] Higher cross-sectional areas of LM, LAD and LCx and higher tortuosity between LM-LCx were strongly correlated with low TAWSS in the LAD, as were higher angles between LM-LAD, LAD-LCx and LAD-septum to a lesser extent. In contrast, higher angles between LM-LCx negatively correlated with low TAWSS in the LAD.[25] In the right coronary artery (RCA), lower tortuosity and smaller cross-sectional areas of the right ventricular (RV) branch and a higher angle between the RCA and RV branch had the strongest correlation with low WSS.[26] Smaller cross-sectional areas of the RCA ostium also promoted lower WSS more propitious to atherosclerosis formation.

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 (Table 1).[22]

Table 1: Candidate anatomical biomarkers and haemodynamic variables for coronary artery disease.

Geometric biomarkers
• Flow divider which is offset from the aortic axis • Inward curvature in the aorta as the flow divider is approached • Markedly angulating daughter branch • Asymmetrical T-shaped bifurcation • Bifurcation angle • Cardiac curvature • Diameter • Inflow angle • Tortuosity
Haemodynamic parameters
WSS Time-averaged WSS

WSS = wall shear stress

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- 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, we can infer that more than a quarter of cardiovascular events are unexplained by the Framingham risk model and that there are other risk factors for atherosclerosis which have not yet been identified. Indeed, a recent study using two large multicentre Australian registries, showed that a substantial and increasing proportion of STEMI patients were individuals without SMuRFs.[7] 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% versus 4%, p=0.032). Advanced image-derived patient-specific information may account for some of the unexplained susceptibility to atherosclerosis in SMuRF-less individuals.

CTCA technology already has a well-established role in the field of preventive cardiology. The Scottish Computed Tomography of the Heart (SCOT-HEART) and Prospective Multicentre Imaging Study for Evaluation of Chest Pain (PROMISE) trials were landmark studies which showed that a CTCA-guided strategy improves clinical outcomes in symptomatic patients with stable angina by increasing the diagnostic certainty and frequency of CAD and the subsequent implementation of appropriate secondary prevention and revascularisation.[41-43]

The role of CTCA in asymptomatic patients with CAD remains uncertain. The Factor-64 trial has been the only randomised clinical trial to date to assess the prognostic value of routine screening for CAD

using CTCA in this population.[44] Nine-hundred high-risk diabetic patients were randomized to CTCA or standard national guidelines-based optimal medical care. At four years, 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. A meta-analysis evaluating the prognostic value of CTCA in 6,225 diabetic patients, 4,145 of whom were asymptomatic, observed a higher hazard ratio for obstructive CAD in the studies that included revascularisation in the endpoints compared with those that did not, suggesting that CTCA in this population could have prognostic implications by identifying patients who may be appropriate for revascularisation.[45] 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.[46]

Several studies have demonstrated the incremental predictive value of the CACS, in addition to traditional risk factors for CAD.[27, 28] The South Bay Heart Watch Study found that a CACS >300 combined with the Framingham risk score significantly improved the discriminative ability of the Framingham risk score (c-statistic 0.68 vs 0.63, p<0.001).[27] Similarly, the St. Francis Heart Study showed that CACS 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] Furthermore, the distribution of calcium has been shown to be incremental to its presence and extent in predicting cardiovascular events.[47, 48] An analysis of 1,268 participants from the Offspring and Third Generation cohorts of the Framingham Heart Study showed that the number of coronary arteries with calcium, and especially the presence of calcium in the proximal dominant coronary artery, as detected by CTCA, independently

predicted coronary heart disease after adjustment for the Framingham risk score and CACS.[48] The addition of calcium distribution improved the discriminatory capacity of the multivariable model with Framingham risk score and CACS 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 3,262 participants in the MESA (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 CACS (c-statistic 0.67 vs 0.64, p=0.0001).[47]

There is a tremendous opportunity to improve the accuracy of CAD risk prediction by integrating additional patient-specific anatomical risk with traditional risk models. Geometry shapes flow; adverse geometric features of coronary artery bifurcations enhance atherogenic WSS patterns which govern the localisation and progression of plaques.[11] The distribution of atherosclerosis, in turn, has been demonstrated to predict cardiovascular events independently of systemic risk factors.[47, 48] The use of anatomical surrogate markers for plaque distribution, such as bifurcation angle, rather than CACS or the number of coronary arteries with calcified plaques will enable us to extend the application of CTCA-guided risk prediction from diseased individuals to normal populations without atherosclerosis. 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 to risk assessment in large-scale population studies. More than a quarter of cardiovascular events remain unexplained by systemic risk factors, and individuals without SMuRFs

have been shown to have poor outcomes.[5, 7] Understanding the mechanism of personal susceptibility to atherosclerosis and the early implementation of targeted therapies in susceptible individuals may be the key to addressing the growing burden of CAD.

REFERENCES

1. Anderson KM, Wilson PW, Odell PM et al. An updated coronary risk profile. A statement for health professionals. Circulation. 1991;83(1):356-62.

2. World health statistics 2019: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2019.

3. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet. 1994;344(8934):1383-9.

4. Cannon CP, Blazing MA, Giugliano RP et al. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. N Engl J Med. 2015;372(25):2387-97.

5. 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(1):109.

6. Friedman MH, Deters OJ, Mark FF et al. Arterial geometry affects hemodynamics. A potential risk factor for athersoclerosis. Atherosclerosis. 1983;46(2):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(21):e013296.

8. Pylypchuk 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(10133):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. Medrano-Gracia P, Ormiston J, Webster M et al. A computational atlas of normal coronary artery anatomy. EuroIntervention. 2016;12(7):845-54.

11. Morbiducci U, Kok AM, Kwak BR et al. Atherosclerosis at arterial bifurcations: evidence for the role of haemodynamics and geometry. Thromb Haemost. 2016;115(3):484-92.

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(10):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(1046):109-59.

14. Caro CG. Discovery of the role of wall shear in atherosclerosis. Arterioscler Thromb Vasc Biol. 2009;29(2):158-61.

15. Nerem RM. Vascular fluid mechanics, the arterial wall, and atherosclerosis. J Biomech Eng. 1992;114(3):274-82.

16. Malek AM, Alper SL, Izumo S. Hemodynamic shear stress and its role in atherosclerosis. JAMA. 1999;282(21):2035-42.

17. Friedman MH, Hutchins GM, Bargeron CB et al. Correlation between intimal thickness and fluid shear in human arteries. Atherosclerosis. 1981;39(3):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(7):1411-27.

19. Chaichana T, Sun Z, Jewkes J. Computation of hemodynamics in the left coronary artery with variable angulations. J Biomech. 2011;44(10):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(14):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(5):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(9):1570-82.

23. Malvè 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(1):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(3):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(2):257-70.time

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(2):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(5):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(5):1519-33.

30. Beier S. The Coronary Atlas 2020 [Available from: https://www.coronaryatlas.org/.

31. 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(1):82-90.

32. Medrano-Gracia P, Ormiston J, Webster M et al. Construction of a coronary artery atlas from CT angiography. Med Image Comput Comput Assist Interv. 2014;17(Pt 2):513-20.

33. Gharleghi R, Samarasinghe G, Sowmya A et al. Deep Learning for Time Averaged Wall Shear Stress Prediction in Left Main Coronary Bifurcations. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020: IEEE.

34. 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(4 Suppl):5-40.

35. Antiga L, Steinman DAJItomi. Robust and objective decomposition and mapping of bifurcating vessels. 2004;23(6):704-13.

36. Crenshaw HC, Edelstein-Keshet LJBomb. Orientation by helical motion—II. Changing the direction of the axis of motion. 1993;55(1):213-30.

37. Razavi A, Shirani E, Sadeghi MJJob. Numerical simulation of blood pulsatile flow in a stenosed carotid artery using different rheological models. 2011;44(11):2021-30.

38. Gijsen 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. 2019;40(41):3421-33.

39. Girasis C, Serruys PW, Onuma Y et al. 3-Dimensional bifurcation angle analysis in patients with left main disease: a substudy of the SYNTAX trial (SYNergy Between Percutaneous Coronary Intervention with TAXus and Cardiac Surgery). JACC Cardiovasc Interv. 2010;3(1):41-8.

40. Chen S, Zhang J, Ye F et al. Final kissing balloon inflation by classic crush stenting did not improve the clinical outcomes for the treatment of unprotected left main bifurcation lesions: the importance of double-kissing crush technique. Catheter Cardiovasc Interv. 2008;71(2):166-72.

41. investigators S-H. 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(9985):2383-91.

42. Newby DE, Adamson PD, Berry C et al. Coronary CT Angiography and 5-Year Risk of Myocardial Infarction. N Engl J Med. 2018;379(10):924-33.

43. Douglas PS, Hoffmann U, Patel MR et al. Outcomes of anatomical versus functional testing for coronary artery disease. N Engl J Med. 2015;372(14):1291-300.

44. Muhlestein JB, Lappé DL, Lima JA 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(21):2234-43.

45. 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(7):1274-80.

46. 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(11):934-41.

47. 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(12):1407-16.

48. 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(10).

AUTHORS' CONTRIBUTIONS

DA contributed to the study design, drafting the manuscript and revising it critically for important intellectual content. RG and DM contributed to revising the manuscript. SB and SO are joint first authors. They contributed equally to the study design and conception, revising the manuscript critically for important intellectual content and final approval of the version to be published.

ACKNOWLEDGEMENTS

Nil

FUNDING

DA is supported by an Australian Government research training program scholarship. Award/grant number not applicable.

COMPETING INTERESTS

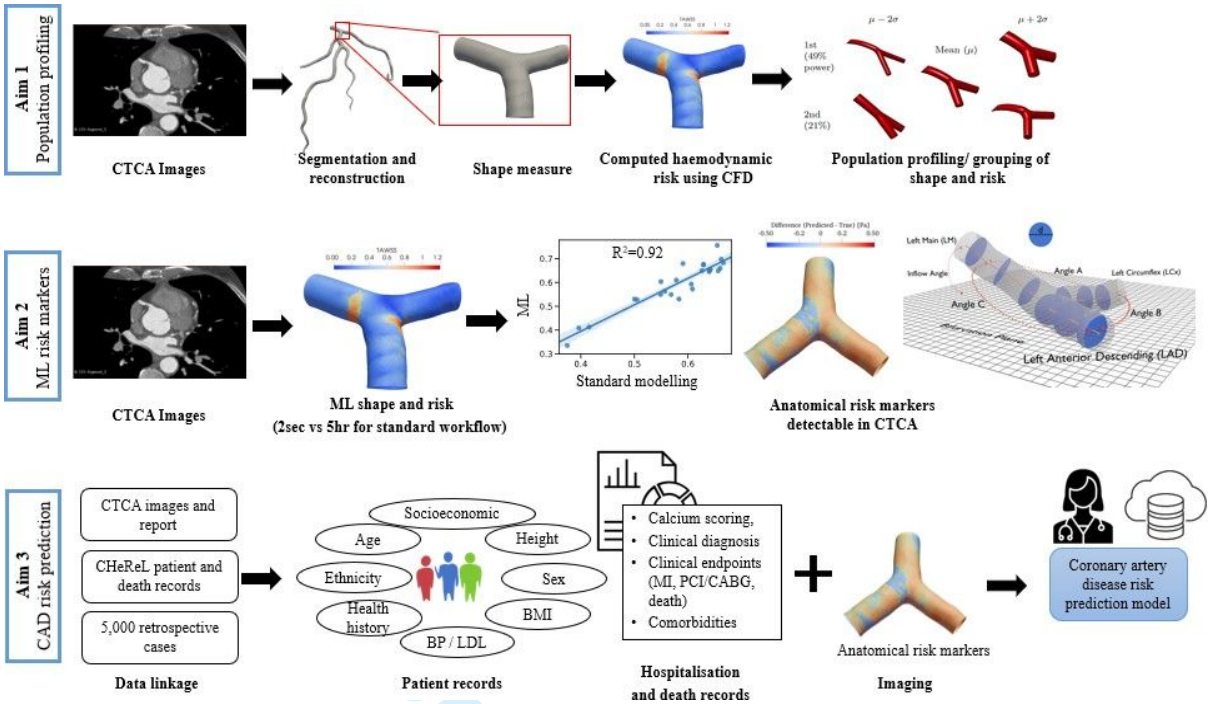
None declared

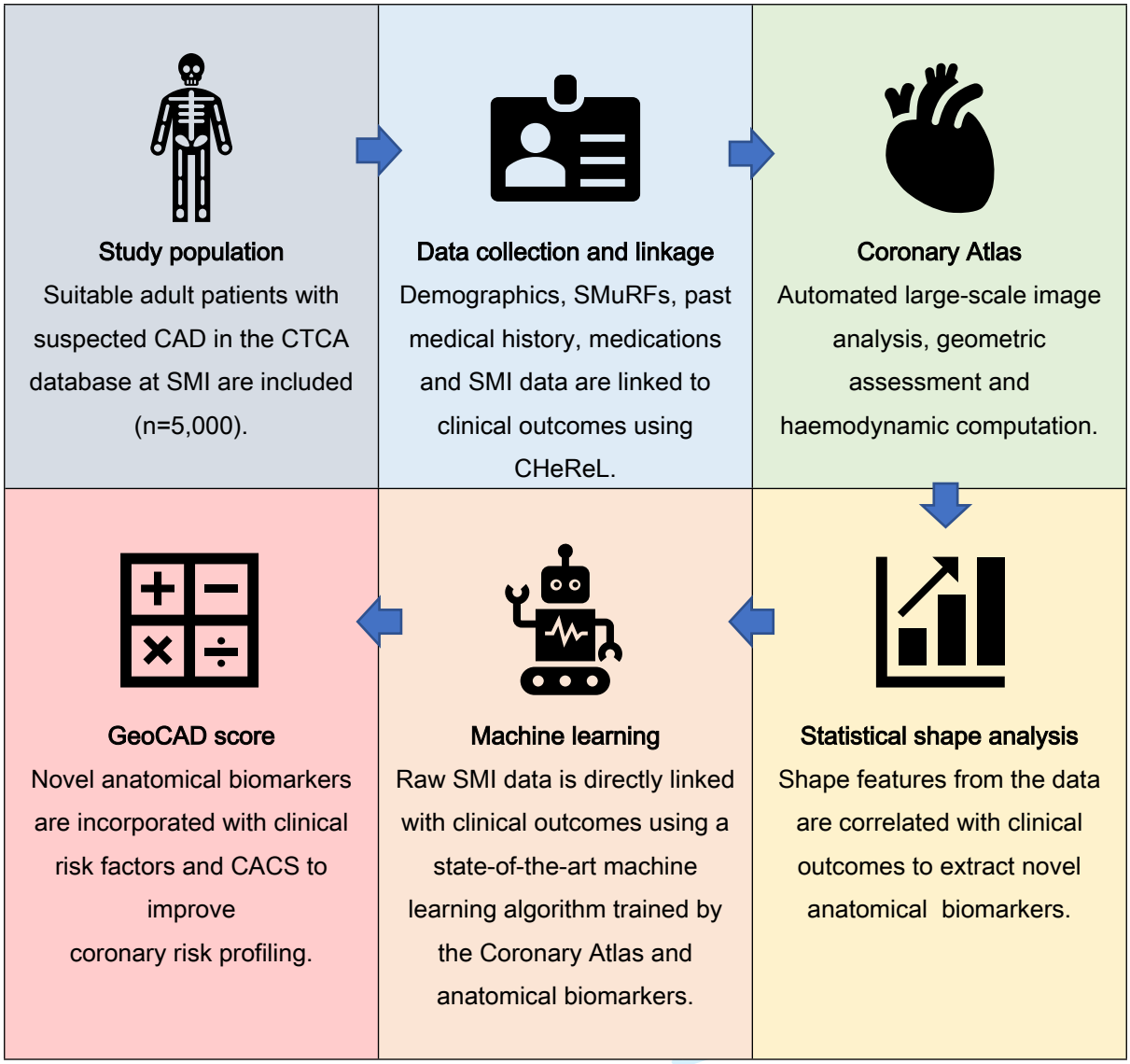
FIGURE LEGENDS

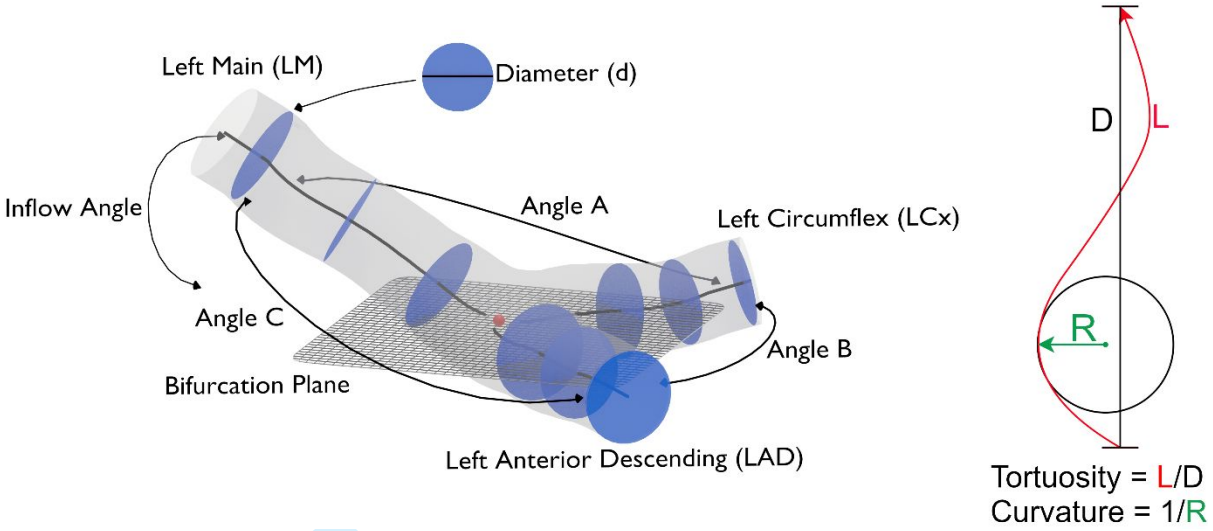
Figure 1: Secondary objectives of the GeoCAD study. CTCA = computed tomography coronary angiography, CFD = computational flow dynamics, ML = machine learning, CHeReL = Centre for Health Record Linkage, BP = blood pressure, LDL = low density lipoprotein, BMI = body mass index, MI = myocardial infarction, PCI = percutaneous coronary intervention, CABG = coronary artery bypass grafting

Figure 2: GeoCAD study flowchart – clockwise from top left to bottom left. CAD = coronary artery disease, CTCA = computed tomography coronary angiography, SMI = spectrum medical imaging, SMuRF = standard modifiable risk factor, CHeReL = Centre for Health Record Linkage, CACS = coronary artery calcium score

Figure 3: 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)







Reporting checklist for prediction model development/validation.

Based on the TRIPOD guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the TRIPODreporting guidelines, and cite them as:

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement.

		Reporting Item	Page Number
Title			
	#1	Identify the study as developing and / or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract			

	#2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
	#3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5
	#3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	9
Methods			
Source of data	#4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	9
Source of data	#4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	1
Participants	#5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	9
Participants	#5b	Describe eligibility criteria for participants.	10
Participants	#5c	Give details of treatments received, if relevant	n/a

Outcome	#6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	11
Outcome	#6b	Report any actions to blind assessment of the outcome to be predicted.	11
Predictors	#7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured	10
Predictors	#7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	11
Sample size	#8	Explain how the study size was arrived at.	9, 10, 13
Missing data	#9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	n/a
Statistical analysis methods	#10a	If you are developing a prediction model describe how predictors were handled in the analyses.	12, 13
Statistical analysis methods	#10b	If you are developing a prediction model, specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	12, 13
Statistical analysis methods	#10c	If you are validating a prediction model, describe how the predictions were calculated.	n/a
Statistical analysis methods	#10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	12, 13

Statistical analysis methods	#10e	If you are validating a prediction model, describe any model updating (e.g., recalibration) arising from the validation, if done	n/a
Risk groups	#11	Provide details on how risk groups were created, if done.	n/a
Development vs. validation	#12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	n/a
Results			
Participants	#13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 2
Participants	#13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	9, 10
Participants	#13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	n/a
Model development	#14a	If developing a model, specify the number of participants and outcome events in each analysis.	n/a
Model development	#14b	If developing a model, report the unadjusted association, if calculated between each candidate predictor and outcome.	n/a

Model specification	#15a	If developing a model, present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	n/a
Model specification	#15b	If developing a prediction model, explain how to the use it.	n/a
Model performance	#16	Report performance measures (with CIs) for the prediction model.	n/a
Model-updating	#17	If validating a model, report the results from any model updating, if done (i.e., model specification, model performance).	n/a
Discussion			
Limitations	#18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	3
Interpretation	#19a	For validation, discuss the results with reference to performance in the development data, and any other validation data	n/a
Interpretation	#19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	n/a
Implications	#20	Discuss the potential clinical use of the model and implications for future research	3

Other information

Supplementary information	#21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	n/a
Funding	#22	Give the source of funding and the role of the funders for the present study.	25

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-054881.R1
Article Type:	Protocol
Date Submitted by the Author:	26-Apr-2022
Complete List of Authors:	Adikari, Dona; University of New South Wales; Prince of Wales Hospital and Community Health Services, Cardiology Gharleghi, Ramtin; University of New South Wales Zhang, Shisheng; University of New South Wales Jorm, Louisa; University of New South Wales Centre for Big Data Research in Health Sowmya, Arcot; University of New South Wales Moses, Daniel; University of New South Wales; Prince of Wales Hospital and Community Health Services, Cardiology Ooi, Sze-Yuan; University of New South Wales; Prince of Wales Hospital and Community Health Services, Cardiology Department Beier, Susann; University of New South Wales
Primary Subject Heading:	Cardiovascular medicine
Secondary Subject Heading:	Cardiovascular medicine
Keywords:	Coronary heart disease < CARDIOLOGY, Computed tomography < RADIOLOGY & IMAGING, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

TITLE PAGE

Title: 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

First author: Dona Adikari MBBS[ab]

Second author: Ramtin Gharleghi BSc[a]

Third author: Shisheng Zhang BSc MPE[a]

Fourth author: Louisa Jorm BVSc MSc PhD[c]

Fifth author: Daniel Moses BSc MBBS MEngSc PhD[ab]

Sixth author: Arcot Sowmya BSc MSc MTech PhD[d]

Seventh author: Sze-Yuan Ooi MBBS MD[ab]

Eighth author: Susann Beier BE ME PhD[a]

[a] The University of New South Wales, Sydney, Australia

[b] Cardiology Department, The Prince of Wales Hospital, Sydney, Australia

[c] Centre for Big Data Research in Health, Faculty of Medicine, The University of New South Wales, Sydney, Australia

[d] School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia

Corresponding author:

Dr Dona Adikari

Eastern Heart Clinic, The Prince of Wales Hospital

Barker Street, Randwick NSW 2031

Email: dona.adikari@unsw.edu.au

Word count: 4354

Dates of the study: 17/03/2022 to 16/03/2027

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 computed tomography 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 1,000 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.

ARTICLE SUMMARY

Strengths and limitations of this study

- GeoCAD is a retrospective cohort study to assess anatomical risk in 1,000 adult patients who have undergone computed tomography coronary angiography (CTCA) 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.

KEYWORDS

Cardiovascular events

Coronary artery disease

Computed tomography coronary angiography

Machine learning

Risk factors

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 infarct (STEMI) patients without standard modifiable risk factors (SMuRFs) have significantly worse in-hospital outcomes compared to 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 sub-populations 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. in 1971,[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 intra-luminal 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 chemotactic 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 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 (Table 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 sub-groups including more vulnerable populations.[8, 9]

Here, we propose a novel approach to build upon 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 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.

Table 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

1

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, 1,000 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 avaiable 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, and
- 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 (ICD-10-AM) and procedures coded according to the Australian Classification of Health Interventions. Clinical data will include the following:

Demographic data (age, sex),

- Standard modifiable risk factors (hypertension, diabetes mellitus, dyslipidaemia, smoking),
- Past medical history (e.g. prior MI, PCI or CABG),

- Medication history,
- 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), and
- Major adverse cardiovascular events (MACE) 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 semi-automatically 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 is encrypted and access to UNSW Data Archive is password protected and requires connection to UNSW's 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 is 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 (VMTK),[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 two minutes 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 models [56] will be trained on the same data to investigate performance of traditional machine learning methods versus deep learning, and potentially provide a more intrepretable risk model. The performance of the trained models will be evaluated and compared using 10-fold cross validation. The Area Under Receiver Operating Characteristics Curve (AUC)[57] metric will be used to compare predictions of the machine learning models to existing literature on machine learning risk models [41] as well as traditional models.. This allows for easy comparisons against other models as it is commonly reported and simple to intrepret.

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 (± standard deviation) and categorical variables as proportions (%). Comparisons between groups will be performed using independent student *t*-tests with Bonferroni correction for continuous variables

and χ^2 or Fisher's exact tests for continuous 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 value <0.05 with Bonferroni correction will be considered significant. Our sample size of 1,000 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 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] curvature,[24] 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- and ten-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 ten-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-of-four patients remain 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% versus 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 Computed Tomography of the Heart (SCOT-HEART) and Prospective Multicentre Imaging Study for Evaluation of Chest Pain (PROMISE) trials were landmark studies, showing that 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 6,000 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 1,200 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 3,262 participants in the MESA (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]

Whilst 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 overall 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.

REFERENCES

1. Anderson, K.M., et al., *An updated coronary risk profile. A statement for health professionals.* Circulation, 1991. **83**(1): p. 356-62.
2. *World health statistics 2019: monitoring health for the SDGs, sustainable development goals*. 2019, World Health Organization: Geneva.
3. *Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S).* Lancet, 1994. **344**(8934): p. 1383-9.
4. Cannon, C.P., et al., *Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes.* N Engl J Med, 2015. **372**(25): p. 2387-97.
5. Damen, J.A., 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**(1): p. 109.
6. Friedman, M.H., et al., *Arterial geometry affects hemodynamics. A potential risk factor for athersoclerosis.* Atherosclerosis, 1983. **46**(2): p. 225-31.
7. Vernon, S.T., 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**(21): p. e013296.
8. Pylypchuk, R., et al., *Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study.* Lancet, 2018. **391**(10133): p. 1897-1907.
9. Hippisley-Cox, J., C. Coupland, and P. Brindle, *Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study.* BMJ, 2017. **357**: p. j2099.
10. Morbiducci, U., et al., *Atherosclerosis at arterial bifurcations: evidence for the role of haemodynamics and geometry.* Thromb Haemost, 2016. **115**(3): p. 484-92.
11. Medrano-Gracia, P., et al., *A computational atlas of normal coronary artery anatomy.* EuroIntervention, 2016. **12**(7): p. 845-54.
12. Antoniadis, A.P., et al., *Biomechanical Modeling to Improve Coronary Artery Bifurcation Stenting: Expert Review Document on Techniques and Clinical Implementation.* JACC Cardiovasc Interv, 2015. **8**(10): p. 1281-96.
13. Caro, C.G., J.M. Fitz-Gerald, and R.C. Schroter, *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**(1046): p. 109-59.
14. Caro, C.G., *Discovery of the role of wall shear in atherosclerosis.* Arterioscler Thromb Vasc Biol, 2009. **29**(2): p. 158-61.
15. Nerem, R.M., *Vascular fluid mechanics, the arterial wall, and atherosclerosis.* J Biomech Eng, 1992. **114**(3): p. 274-82.
16. Malek, A.M., S.L. Alper, and S. Izumo, *Hemodynamic shear stress and its role in atherosclerosis.* JAMA, 1999. **282**(21): p. 2035-42.
17. Friedman, M.H., et al., *Correlation between intimal thickness and fluid shear in human arteries.* Atherosclerosis, 1981. **39**(3): p. 425-36.
18. Dolan, J.M., J. Kolega, and H. Meng, *High wall shear stress and spatial gradients in vascular pathology: a review.* Ann Biomed Eng, 2013. **41**(7): p. 1411-27.
19. Chaichana, T., Z. Sun, and J. Jewkes, *Computation of hemodynamics in the left coronary artery with variable angulations.* J Biomech, 2011. **44**(10): p. 1869-78.
20. Dong, J., et al., *Fluid-structure interaction analysis of the left coronary artery with variable angulation.* Comput Methods Biomech Biomed Engin, 2015. **18**(14): p. 1500-8.
21. Ikeda, U., et al., *Stenotic lesions and the bifurcation angle of coronary arteries in the young.* Jpn Heart J, 1991. **32**(5): p. 627-33.

22. Beier, S., 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**(9): p. 1570-1582.
23. Malvè, M., 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**(1): p. 82-93.
24. Chiastra, C., 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**: p. 79-88.
25. Pinho, N., 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**(3): p. 715-729.
26. Pinho, N., et al., *The Impact of the Right Coronary Artery Geometric Parameters on Hemodynamic Performance.* Cardiovasc Eng Technol, 2019. **10**(2): p. 257-270.
27. Greenland, P., et al., *Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals.* JAMA, 2004. **291**(2): p. 210-5.
28. Arad, Y., 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**(5): p. 553-72.
29. Liang, L., 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**(5): p. 1519-1533.
30. Singh, A. and J.V. Guttag. *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*. 2011. IEEE.
31. Khateeb, N. and M. Usman. *Efficient heart disease prediction system using K-nearest neighbor classification technique*. in *Proceedings of the international conference on big data and internet of thing*. 2017.
32. Colombet, I., et al. *Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression*. in *Proceedings of the AMIA Symposium*. 2000. American Medical Informatics Association.
33. Krittanawong, C., et al., *Machine learning and deep learning to predict mortality in patients with spontaneous coronary artery dissection.* Scientific reports, 2021. **11**(1): p. 1-10.
34. Sajeev, S., et 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*. 2019, Springer. p. 96-103.
35. Bharti, R., et al., *Prediction of heart disease using a combination of machine learning and deep learning.* Computational intelligence and neuroscience, 2021. **2021**.
36. Tison, G.H., et al., *Cardiovascular risk stratification using off-the-shelf wearables and a multi-task deep learning algorithm.* Circulation, 2017. **136**(suppl_1): p. A21042-A21042.
37. Chao, H., et al., *Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography.* Nature communications, 2021. **12**(1): p. 1-10.
38. Poplin, R., et al., *Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning.* Nature Biomedical Engineering, 2018. **2**(3): p. 158-164.
39. Cheung, C.Y., et al., *A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre.* Nature biomedical engineering, 2021. **5**(6): p. 498-508.
40. Huang, W., et al., *Application of ensemble machine learning algorithms on lifestyle factors and wearables for cardiovascular risk prediction.* Scientific Reports, 2022. **12**(1): p. 1-12.
41. Zeleznik, R., et al., *Deep convolutional neural networks to predict cardiovascular risk from computed tomography.* Nature communications, 2021. **12**(1): p. 1-9.

42. Beier, S. *The Coronary Atlas*. 2020 13/09/2020]; Available from: https://www.coronaryatlas.org/.
43. Medrano-Gracia, P., et al., *A Study of Coronary Bifurcation Shape in a Normal Population.* J Cardiovasc Transl Res, 2017. **10**(1): p. 82-90.
44. Medrano-Gracia, P., et al., *Construction of a coronary artery atlas from CT angiography.* Med Image Comput Comput Assist Interv, 2014. **17**(Pt 2): p. 513-20.
45. Gharleghi, R., et al. *Deep Learning for Time Averaged Wall Shear Stress Prediction in Left Main Coronary Bifurcations*. in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. 2020. IEEE.
46. Austen, W.G., 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**(4 Suppl): p. 5-40.
47. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
48. Isensee, F., et al., *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.* Nature methods, 2021. **18**(2): p. 203-211.
49. Gharleghi, R., et al., *Automated Segmentation of Normal and Diseased Coronary Arteries-The ASOCA Challenge.* Computerized Medical Imaging and Graphics, 2022: p. 102049.
50. Antiga, L. and D.A.J.I.t.o.m.i. Steinman, *Robust and objective decomposition and mapping of bifurcating vessels.* 2004. **23**(6): p. 704-713.
51. Kashyap, V., et al., *Accuracy of vascular tortuosity measures using computational modelling.* Scientific reports, 2022. **12**(1): p. 1-10.
52. Crenshaw, H.C. and L.J.B.o.m.b. Edelstein-Keshet, *Orientation by helical motion—II. Changing the direction of the axis of motion.* 1993. **55**(1): p. 213-230.
53. Razavi, A., E. Shirani, and M.J.J.o.b. Sadeghi, *Numerical simulation of blood pulsatile flow in a stenosed carotid artery using different rheological models.* 2011. **44**(11): p. 2021-2030.
54. Gijsen, F., et al., *Expert recommendations on the assessment of wall shear stress in human coronary arteries: existing methodologies, technical considerations, and clinical applications.* 2019. **40**(41): p. 3421-3433.
55. Chen, Y.-h., et al., *Locally-connected and convolutional neural networks for small footprint speaker recognition.* 2015.
56. Boulesteix, A.L., 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**(6): p. 493-507.
57. Zou, K.H., A.J. O'Malley, and L. Mauri, *Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models.* Circulation, 2007. **115**(5): p. 654-657.
58. Buradi, A. and A. Mahalingam, *Impact of coronary tortuosity on the artery hemodynamics.* Biocybernetics and Biomedical Engineering, 2020. **40**(1): p. 126-147.
59. Doutel, E., et al., *Link between deviations from Murray's Law and occurrence of low wall shear stress regions in the left coronary artery.* Journal of theoretical biology, 2016. **402**: p. 89-99.
60. investigators, S.-H., *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**(9985): p. 2383-91.
61. Newby, D.E., et al., *Coronary CT Angiography and 5-Year Risk of Myocardial Infarction.* N Engl J Med, 2018. **379**(10): p. 924-933.
62. Douglas, P.S., et al., *Outcomes of anatomical versus functional testing for coronary artery disease.* N Engl J Med, 2015. **372**(14): p. 1291-300.

63. Muhlestein, J.B., 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**(21): p. 2234-43.
64. Celeng, C., et al., *Prognostic Value of Coronary Computed Tomography Angiography in Patients With Diabetes: A Meta-analysis.* Diabetes Care, 2016. **39**(7): p. 1274-80.
65. Cho, I., 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**(11): p. 934-941.
66. Blaha, M.J., et al., *Improving the CAC Score by Addition of Regional Measures of Calcium Distribution: Multi-Ethnic Study of Atherosclerosis.* JACC Cardiovasc Imaging, 2016. **9**(12): p. 1407-1416.
67. Ferencik, M., 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**(10).
68. Benjamin, M.M. and M.G. Rabbat, *Machine learning-based advances in coronary computed tomography angiography.* Quantitative Imaging in Medicine and Surgery, 2021. **11**(6): p. 2208.
69. Kadem, M., et al., *Hemodynamic modeling, medical imaging, and machine learning and their applications to cardiovascular interventions.* IEEE Reviews in Biomedical Engineering, 2022.
70. Corrigan, F.E., et al., *Imaging for predicting, detecting, and managing complications after transcatheter aortic valve replacement.* JACC: Cardiovascular Imaging, 2019. **12**(5): p. 904-920.
71. Samad, M.D., et al., *Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning.* JACC: Cardiovascular Imaging, 2019. **12**(4): p. 681-689.
72. Ulloa Cerna, A.E., et al., *Deep-learning-assisted analysis of echocardiographic videos improves predictions of all-cause mortality.* Nature Biomedical Engineering, 2021. **5**(6): p. 546-554.
73. Kurkure, U., et al., *A supervised classification-based method for coronary calcium detection in non-contrast CT.* The international journal of cardiovascular imaging, 2010. **26**(7): p. 817-828.
74. Lessmann, N., et al., *Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions.* IEEE transactions on medical imaging, 2017. **37**(2): p. 615-625.
75. Martin, S.S., et al., *Evaluation of a deep learning–based automated CT coronary artery calcium scoring algorithm.* Cardiovascular Imaging, 2020. **13**(2_Part_1): p. 524-526.
76. Li, D., et al., *Machine learning-aided risk stratification system for the prediction of coronary artery disease.* International Journal of Cardiology, 2021. **326**: p. 30-34.
77. Tesche, C. and V. Brandt, *Calling for a New Framingham: Machine Learning in Cardiovascular Risk Assessment—The Key for Improved Outcome Prediction?* 2021, American College of Cardiology Foundation Washington DC. p. 626-628.

AUTHORS' CONTRIBUTIONS

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.

ACKNOWLEDGEMENTS

Nil

FUNDING

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

COMPETING INTERESTS

None declared

1

2

3

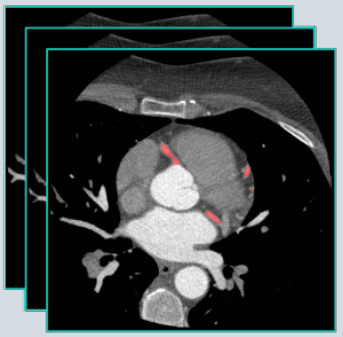
FIGURE LEGENDS

Figure 1: GeoCAD study flowchart. BMI = body mass index, BP = blood pressure, CACS = coronary artery calcium score, CAD = coronary artery disease, CHeReL = Centre for Health Record Linkage, CTCA = computed tomography coronary angiography, LDL = low-density lipoprotein, SMI = Spectrum Medical Imaging, SMuRF = standard modifiable 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)

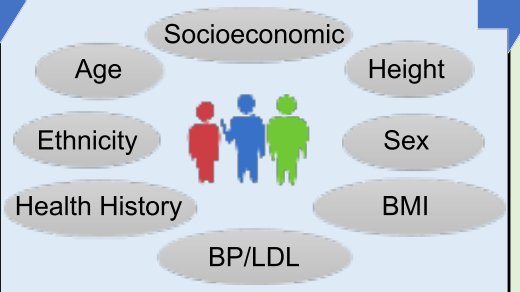
Study Population

Suitable adult patients with suspected CAD in the CTCA database at SMI are included (n=1,000).



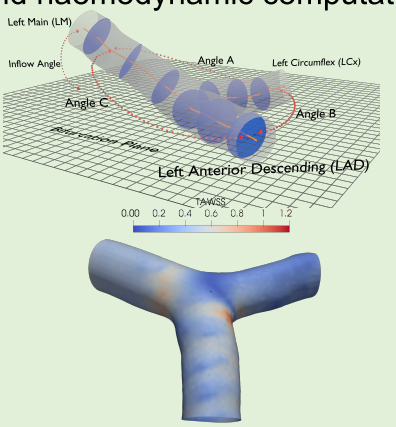
Data Collection And Linkage

Demographics, SMuRFs, past medical history, medications and SMI data are linked to clinical outcomes using CHeReL.



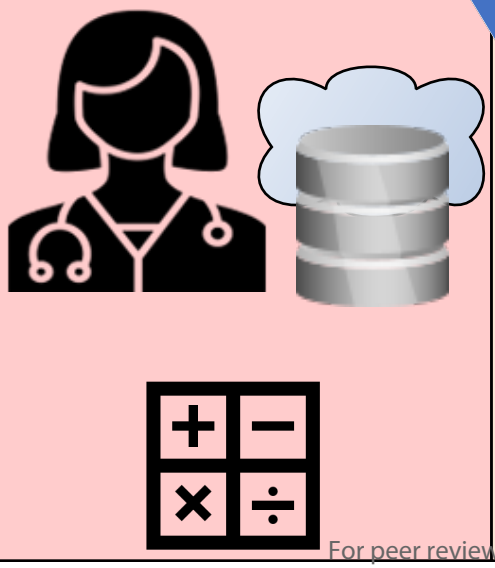
Coronary Atlas

Automated large-scale image analysis, geometric assessment and haemodynamic computation.



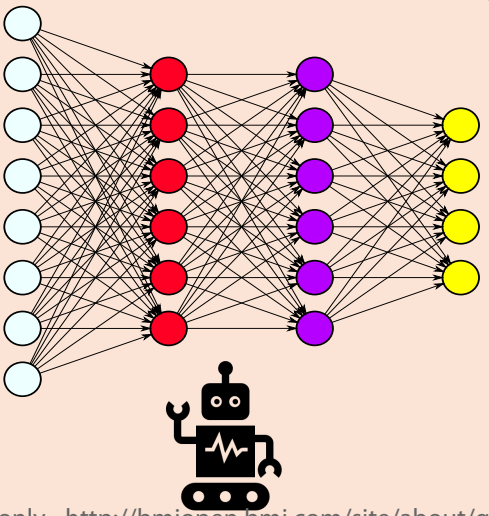
GeoCAD Score

Novel anatomical biomarkers are incorporated with clinical risk factors and CACS to improve coronary risk profiling.



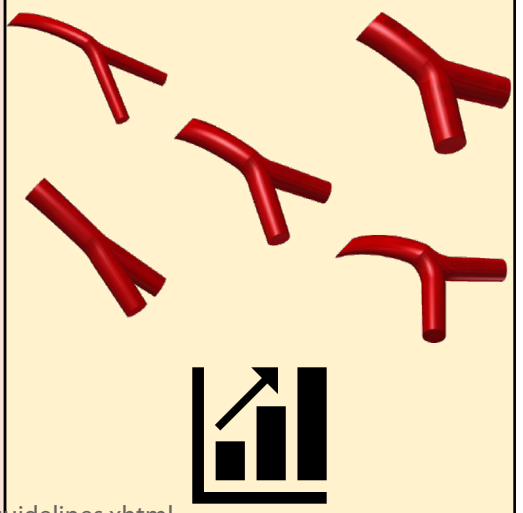
Machine Learning

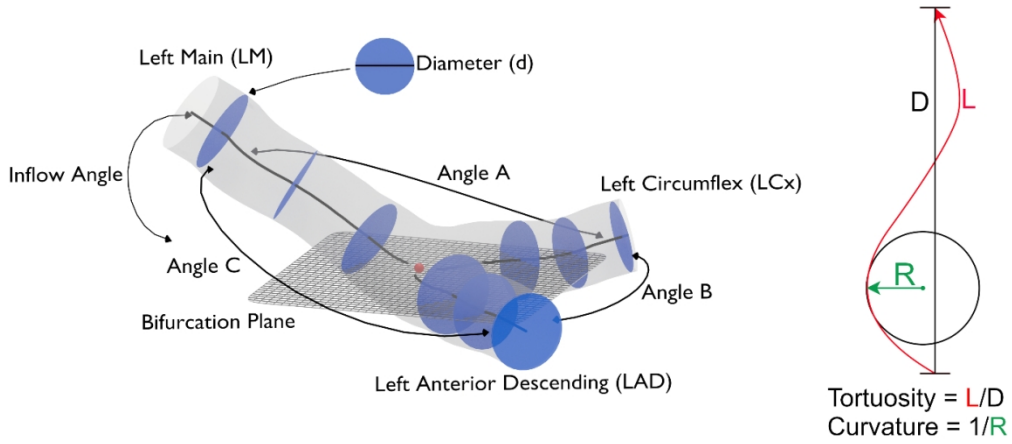
Raw SMI data is directly linked with clinical outcomes using a state-of-the-art machine learning algorithm trained by the Coronary Atlas and anatomical biomarkers.



Statistical Shape Analysis

Shape features from the data are correlated with clinical outcomes to extract novel anatomical biomarkers.





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)

402x176mm (87 x 87 DPI)

Reporting checklist for prediction model development/validation.

Based on the TRIPOD guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the TRIPODreporting guidelines, and cite them as:

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement.

		Reporting Item	Page Number
Title			
	#1	Identify the study as developing and / or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract			

	#2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
	#3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5
	#3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	9
Methods			
Source of data	#4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	9
Source of data	#4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	1
Participants	#5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	9
Participants	#5b	Describe eligibility criteria for participants.	10
Participants	#5c	Give details of treatments received, if relevant	n/a

Outcome	#6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	11
Outcome	#6b	Report any actions to blind assessment of the outcome to be predicted.	11
Predictors	#7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured	10
Predictors	#7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	11
Sample size	#8	Explain how the study size was arrived at.	9, 10, 13
Missing data	#9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	n/a
Statistical analysis methods	#10a	If you are developing a prediction model describe how predictors were handled in the analyses.	12, 13
Statistical analysis methods	#10b	If you are developing a prediction model, specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	12, 13
Statistical analysis methods	#10c	If you are validating a prediction model, describe how the predictions were calculated.	n/a
Statistical analysis methods	#10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	12, 13

Statistical analysis methods	#10e	If you are validating a prediction model, describe any model updating (e.g., recalibration) arising from the validation, if done	n/a
Risk groups	#11	Provide details on how risk groups were created, if done.	n/a
Development vs. validation	#12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	n/a
Results			
Participants	#13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 2
Participants	#13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	9, 10
Participants	#13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	n/a
Model development	#14a	If developing a model, specify the number of participants and outcome events in each analysis.	n/a
Model development	#14b	If developing a model, report the unadjusted association, if calculated between each candidate predictor and outcome.	n/a

Model specification	#15a	If developing a model, present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	n/a
Model specification	#15b	If developing a prediction model, explain how to the use it.	n/a
Model performance	#16	Report performance measures (with CIs) for the prediction model.	n/a
Model-updating	#17	If validating a model, report the results from any model updating, if done (i.e., model specification, model performance).	n/a
Discussion			
Limitations	#18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	3
Interpretation	#19a	For validation, discuss the results with reference to performance in the development data, and any other validation data	n/a
Interpretation	#19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	n/a
Implications	#20	Discuss the potential clinical use of the model and implications for future research	3

Other information

Supplementary information	#21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	n/a
Funding	#22	Give the source of funding and the role of the funders for the present study.	25