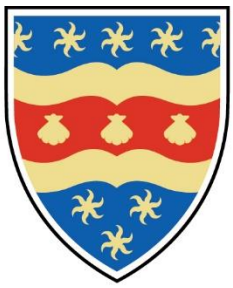

UNIVERSITY OF PLYMOUTH

AN EVALUATION OF THE INCREMENTAL VALUE OF COMPUTED TOMOGRAPHIC BIOMARKERS IN CARDIOVASCULAR RISK PREDICTION

by

CHUN LAP PANG

A thesis submitted to University of Plymouth in fulfilment for the degree of

DOCTOR OF PHILOSOPHY

Peninsula Medical School

July 2020

ACKNOWLEDGEMENT

Over the last 3 years it has been a pleasure to collaborate and learn from the Derriford CT physicist (Mr Robert Loader), the academics (Dr Yinghui Wei, Dr Jaime Peters, Professor Chris Hyde, Dr William Lee and Mr Simon Briscoe) and the clinicians (Dr Gareth Morgan-Hughes, Dr Vikram Raju and Professor Carl Roobottom). The administrative staff (Jo Smith, Margaret Hart and support staff at Primary Care, University of Plymouth, UK) and colleagues (Dr Varut Vardhanabhuti, Dr Tinu Purayil, Dr Sean Tenant, Dr Prageeth Dissanyake, Ms Joanne Hooper, Ms Charlotte Land, the CT booking team and CT radiographers at Derriford Hospital, UK) have all contributed to making this project a positive experience. The Head of School (Dr Tom Sulkin) at the School of Radiology has been supportive from the outset and throughout this whole process. The financial generosity of Plymouth NHS Hospitals Trust and Peninsula CLAHRC for sponsoring this degree has been hugely appreciated. I also received support from the Radiology Academy and the Imaging Department of Derriford Hospital during the two years of out-of-programme leave. Without the support of the Directorate of the Imaging Department of Derriford Hospital, I would not have been able to continue with my clinical commitments whilst simultaneously pursuing this research. Without the individuals and institutions above, I would not have been able to finish this thesis. Finally, I would like to thank my wife (Nicola), dog (Lisa) and now daughter (Mia) for patiently supporting me throughout.

Details of contribution: Professor Carl Roobottom and Dr Tinu Purayil conceptualised the feasibility study. Mr Robert Loader and Dr Vikram Raju devised the scanning protocol for the feasibility study. Professor Carl Roobottom and Dr Gareth Morgan-Hughes subjectively analysed the acquired coronary data from the feasibility study. Professor Chris Hyde conceptualised the systematic review. Dr Jaime Peters and Dr Yinghui Wei provided statistical support. Dr Jaime Peters provided methodological support. Mr Simon Briscoe

provided support for searching the literature. Professor Carl Roobottom, Professor Chris Hyde and Dr Jaime Peters all commented on the thesis and the publications derived from it.

Chun Lap Pang

AUTHOR'S DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee. Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a studentship by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula (NIHR CLAHRC South West Peninsula) and Plymouth Hospitals NHS Trust.

Publication:

Pang CL, Pilkington N, Wei Y, Peters J, Roobottom C, Hyde C. A methodology review on the incremental prognostic value of computed tomography biomarkers in addition to Framingham risk score in predicting cardiovascular disease: the use of association, discrimination and reclassification. BMC Cardiovasc Disord. 2018;18(1):39.

Word count of main body of thesis:

Signed:

Student number: 10480125

Date:

LIST OF CONTENTS

LIST OF CONTENTS

LIST OF CONTENTS

LIST OF CONTENTS

ABBREVIATIONS

AHA – American Heart Association

ACS – Acute coronary syndrome

Δ AUC - The difference in AUC

AUC – Area under the receiver operating characteristics curve

ATP III – Adult Treatment Panel III

Bpm – Beat per minute

BSCCT – British Society of Cardiovascular Computed Tomography

BMI – Body mass index

CABG – Coronary artery bypass graft

CACS – Coronary artery calcium score

CHARMS – CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies

CHD – Coronary artery/ heart disease

CIs – Confidence intervals

CIMT - Carotid intimal-medial wall thickness

CONFIRM – Coronary CT Angiography Evaluation for Clinical Outcomes: An International Multicenter Registry

CT – Computed tomography

CTCA – computed tomography coronary angiogram

ABBREVIATIONS

CTCA –computed tomography coronary angiography

CVD – Cardiovascular disease

DHS – Diabetes Heart Study

DLP – Dose-length product

ECG – Electrocardiogram

EBCT – Electron beam CT

ECG - Electrocardiogram

EISNER – Early Identification of Subclinical Atherosclerosis by Noninvasive Imaging Research

Exc. – Excluding

FOV – Field of view

FRS – Framingham Risk Score

FSRS – Framingham Stroke Risk Score

GFR – Glomerular filtration rate

GE – General Electric

HDL – High density lipoprotein

HF – Heart failure

HNR – Heinz Nixdorf Recall

hsCRP – High-sensitivity C-reactive protein

ABBREVIATIONS

kVp – Peak tube voltage

ICA – Invasive coronary angiogram

IV – Intravenous

IDI - Integrate discrimination index

I^2 – I-square

Inc. – Including

LAD – Left anterior descending artery

LDL – Low density lipoprotein

LIMA – Left internal mammary artery

LVEF – Left ventricular ejection fraction

MACE – Major adverse cardiac events

MESA – Multi-Ethnic Study of Atherosclerosis

MDCT – Multi-detector CT

MI – Myocardial infarction

NRI - Net reclassification index

n/a – Not specified or available

mSv – Millisieverts

NHS – National Health Service

ABBREVIATIONS

PROSPERO – International database of prospectively registered systematic reviews in health and social care (Centre for Reviews and Dissemination, University of York)

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses

QALY – Quality-adjusted life years

QUIPS - Quality In Prognosis Studies

rIDI - Relative integrate discrimination index

r - Correlation coefficient

RCA- Right coronary artery

RCRI - Revised Cardiac Risk Index

SD – Standard deviation

SCORE – Systematic Coronary Risk Evaluation

SPECT – Single-photon emission computed tomography

TACS – Thoracic aorta calcium score

TIA – Transient ischaemic attack

UK – United Kingdom

Abstract

Predicting cardiovascular events is an important subject in the developed world as it is a major cause of morbidity and mortality. Identifying those at risk of developing cardiovascular disease is key as there are treatments available to reduce the risk of future events. The most well-known prediction tool is the Framingham Risk Score (FRS), a multivariate cardiovascular risk prediction model. The Framingham cohort identified some of the most fundamental risk factors that shape modern cardiovascular prevention, however, it is not a perfect model.

The imperfect nature of cardiovascular risk prediction based on FRS forms the starting point of this research journey. In the search for a better prediction tool, a logical approach would be to improve on an existing model, rather than 'reinventing the wheel'. This philosophy underpins this piece of work, which focuses on finding a tool that improves identification of subclinical disease. From my clinical practice in radiology, the value of cardiovascular CT biomarkers became an obvious area to investigate. Over the course of my research, I realised both cardiovascular (CVD) risk prediction models and CVD CT biomarkers have evolved over a similar period. The scope of my research demanded my attention to focus on FRS as a base model, though there are many other CVD risk prediction models. Similarly, there are multiple cardiovascular CT biomarkers that have been proposed. The best studied CT biomarker in terms of predicting CVD events is undoubtedly coronary calcium score (CACS). Considering the evolving nature of CT technology and the deeper understanding of CVD pathophysiology, there are two other up-and-coming biomarkers, namely thoracic calcium score (TACS) and coronary artery stenosis, which broaden the scope of investigating potentially useful biomarkers.

Embedding CT biomarkers within Framingham Risk Score formed the framework investigation. Derived from this was a journey of discovery that led me to learn the rapidly expanding knowledge of prognosis research. My initial investigation was conducting a systematic review and meta-analysis of the incremental value of discussed CT biomarkers. This was followed by investigating the reporting standard of the Framingham Model within the realm of incremental value added by CT biomarkers. Finally, performing a feasibility study to look at whether the coronary arteries can be assessed during routine oncological whole-body CT imaging. I would like to illustrate and share my learning in the subsequent chapters.

Chapter 1 – Introduction

1.1 Framingham risk score & prediction models in cardiovascular disease

CVD is common amongst adults and causes significant morbidity and mortality worldwide. The absence of risk factors at 50 years of age is associated with a very low lifetime risk for CVD and a longer median survival (1) and is the rationale behind implementing intensive preventative strategies and modification of daily living. Coronary heart disease (CHD) is the most common manifestation of CVD but many individuals with CVD are initially free of CHD (2), thus adding to the attraction of having prediction tools. To implement preventative measures and targeted therapy, such as lipid lower drugs and lifestyle adjustments, identifying those at risk has been keenly advocated. However, it is vitally important that any prediction model used in CVD risk prediction is accurate to avoid inappropriate risk categorisation, which can lead to either over- or under-treatment.

The original FRS was derived from a mostly Caucasian population and was one of the first models to provide an estimation of CVD risk. The FRS used CHD death, non-fatal myocardial infarction, stable angina and unstable angina as endpoints. The Wilson 1998 version incorporated age, gender, systolic blood pressure, total or low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, diabetes mellitus and smoking in their multivariate model (3, 4). In general, prediction models are never perfect (5) and frequently have methodological weaknesses meaning very few stand the test of time (6). The FRS has extensive validation and is one of the exceptions (5), however, it remains imperfect and attempts have been made to improve its risk prediction, particularly within the “intermediate” category (7). Like other prediction tools, different iterations were developed to improve upon earlier versions. The subsequent version (ATP III 2002) excluded diabetes and included blood pressure treatment, whilst excluding stable and unstable angina as endpoints (8). Apart from CHD, cerebrovascular disease, peripheral artery disease (PVD)

and aortic aneurysms account for the remaining cases of CVD. The D'Agostino 2008 version was similar to the ATP III 2002 version but considered more outcomes including coronary insufficiency or angina, fatal or non-fatal ischaemic or haemorrhagic stroke, transient ischaemic attack and heart failure (9). All of the described iteration models have been externally validated numerous times in different settings and populations, with most studies showing poor calibration. Therefore, there still remains a drive to improve even the most validated CVD risk prediction tool.

Clinical guidelines from the National Cholesterol Education Program and the American College of Cardiology and American Heart Association (AHA) advise using the ATP III 2002 model but it is worth noting that most clinical research has utilised the Wilson 1998 model. Although Framingham Wilson is not mentioned in the clinical guidelines, it is relevant to review this prediction model as many studies in the field of CVD risk prediction have externally validated it and used it to either assess the incremental value of new predictors, or for comparison with newly developed prediction models (5). FRS was developed on an American population who were asymptomatic at time of joining the original Wilson 1998 model inception. It is important to note that there are subsequent studies that implement FRS on a mixture of symptomatic and asymptomatic populations, which is not the original intended population of the model. There are many other versions and variations of FRS. This research explicitly did not set out to review all existing CVD risk prediction models. Owing to the nature of the literature and included studies, the research focus was on the Wilson 1998, ATP 2002 and D'Agostino 2008 versions. Other competing prediction models such as the SCORE (10) QRISK (11) models, which have been developed using European populations, have not been considered because sensible conclusions can only be drawn when a single, rather than multiple, model is evaluated at the outset.

1.2 Evaluation of multivariable risk prediction model & standard of reporting prediction model in cardiovascular risk prediction

Development and reporting of prediction models is generally poor in the medical and specialty literature (6, 12). The TRIPOD statement aims to improve the reporting of multivariable prediction, which can be applied to CVD risk prediction models (13). The initial development of new prediction studies should always include quantification of predictive performance of the developed model, in particular discrimination and calibration. Discrimination is seen as how good the model is at identifying higher risk individuals who go on to develop the outcome of interest and can be quantified with summary estimates, such as sensitivity, specificity and the area under the receiver operating characteristics curve (AUC). Predictive performance can be further quantified in terms of calibration, which look at whether the proportion classified as high risk indeed went on to develop the outcome of interest. Calibration can be quantified using Hosmer-Lemeshow 'goodness-of-fit" test. When the predictive performances are available, methods such as bootstrapping are applied. This is known as internal validation which is a necessary part of the development. Other issues arising at this stage, such as overfitting, optimism and miscalibration can be addressed. After a model is developed, it is important to evaluate the predictive performance in another population other than the one used for model development. This is known as external validation.

Damen et al summarised the current landscape of the CVD prediction literature comprehensively (5), examining 363 prediction models in the CVD literature. Overall, there were plenty of CVD risk prediction models for the general population but also a lot of methodological concerns. Apart from the well-known models (FRS, SCORE and QRISK), the competing models lack external validation leaving healthcare professionals uncertain of their value. The lack of power in some of the included studies is concerning. Furthermore, the

number of events in those studies is frequently less than 10 per variable (14, 15). Heterogeneous and selective reporting of the basic information regarding the prediction model was another issue, with disparities between studies making direct comparison difficult. The CVD outcomes were heterogeneous with more than 40 different definitions for fatal or non-fatal CHD and international classification codes being specified in less than a quarter of the studies. Most studies had a prediction horizon of 5 or 10 years, but there was marked variation (between 2 and 45 years), making comparison difficult. When the follow-up period does not match the prediction horizon, there is a risk of extrapolation, which has not been explored (5). Lastly, there is also a lack of indicators of the model's predictive performance. A mere 39% of included studies reported a measure of discrimination, 32% reported calibration and 27% reported both. Even when there were indicators, there was often insufficient information to allow calculation of individual model performance.

1.3 Imaging biomarkers: Computed tomographic biomarkers

Numerous imaging and biochemical markers have been investigated aiming to improve upon the FRS model (16). This is in keeping with increasing interest in novel biomarkers in the field of CVD risk prediction (17). In search of a surrogate biomarker that detects subclinical disease, CACS has been investigated for more than three decades (18) with some proposing screening with CT in the general population (19). TACS is considered a relative of CACS, whereby the Agatston method (20) is applied to the thoracic aorta (21). Computed tomographic coronary angiography (CTCA) has established itself in the acute chest pain setting and is now being investigated as a tool of reclassifying cardiac risk based on luminal stenosis (and other characteristics) in the CONFIRM cohort (22). These imaging biomarkers generate substantial interest and may add incremental value to traditional Framingham risk factors.

The idea of using CACS as a biomarker was first documented in 1979 (23). CACS is a semi-automated method used to quantify the level of coronary calcium. The calcium indicates the result of inflammatory changes within the vessel wall of the coronary arteries as a result of atherosclerosis. Since then, the Agatston method has become the most cited method regarding the quantification of CACS (20). The Multi-Ethnic Study of Atherosclerosis (MESA) cohort recruited participants of Black, Hispanic and Chinese descent which specifically tackled the concern that FRS was primarily applicable to Caucasians (7). The MESA study also investigated whether other subclinical parameters predicted CVD, such as CACS and carotid intimal-medial wall thickness (cIMT) (7). Improvements in CHD risk prediction with other subclinical risk markers including cIMT, ankle-brachial index and pulse wave velocity were promising in other studies (8-12). Kavousi et al showed that CACS had the best independent risk prediction and best discriminatory ability when compared with 11 other measures of atherosclerosis, including both imaging characteristics and biomarkers (13). Studies have demonstrated the link between the risk of future CHD and mortality with CACS (14-19). To quantify atherosclerosis elsewhere, some have proposed investigating the quantity of thoracic aortic calcium (TAC), leading to the development of thoracic aorta calcium score (TACS). There is some evidence to suggest that there is correlation between TAC and coronary risk factors (20). One main criticism of CACS is that it can potentially miss non-calcified plaques, which can progress and cause myocardial infarction. Invasive coronary angiography (ICA) is the gold standard investigation for the evaluation of the coronary arteries and in many centres forms the basis for any subsequent intervention. Coronary revascularisation is performed based on semi-quantitative measures of luminal diameter narrowing of the artery visualised at the time of ICA. Multiple retrospective and prospective studies have demonstrated that CTCA has high sensitivity and specificity for the non-invasive detection of stenosis when compared with ICA (22-27). More recently, the PROMISE and SCOT-HEART studies focused on looking at the longer-term benefits of CTCA (28, 29). A systematic review supported the use of 64-slice CT to rule out significant

CAD but highlighted that more research is required, including identification of radiation burden on repetitive users (30).

1.4 Coronary artery disease screening using calcium score

There are some arguments for screening using CACS. CACS can detect calcification of the coronary arteries in people who are asymptomatic, many of whom would be classed as low risk when assessed by CVD risk prediction models. CACS have often been analysed categorically, most commonly as 0 (none), 1–99 (mild), 100–400 (moderate), and >400 (severe) (24) but can also be analysed as continuous data on a logarithmic scale with a higher CACS indicating higher risk. The idea is that the more extensive the calcification and the higher the CAC score, the greater the risk. Although coronary calcium is indicative of CHD in asymptomatic people, traditional risk factors class these patients as low risk. This forms the basis of “up classification” of low risk individuals using CACS. Treatment with statins can reduce that risk. Meta-analysis suggested that on average statins reduce 27% of fatal and non-fatal CHD outcomes and that there was no real harm, even when used among low risk individuals (25).

However, there are many unfavourable arguments against screening using CACS. Most importantly, CT will miss many of the most dangerous patches of arterial disease because they are not calcified. It would also depend on how much better CACS was than a CVD prediction model such as FRS, which remains an area of debate. Other uncertainties include the threshold of CACS that would trigger onward investigation and treatment, as well as whether it is cost effective. Another consideration is that it is unlikely that a low CACS score would prompt stopping of statins amongst those classed as high-risk individuals based on traditional risk factors. There is still a lack of definitive evidence that statins reduce events in

those with raised CACS (26). Whilst CACS appears attractive and the use of radiation is probably justified, there was insufficient evidence to support screening (27).

1.5 The notion of assessing incremental value

A new biomarker or diagnostic test very rarely replaces an existing test in its entirety. A notable exception applies to the screening test (28, 29). The diagnostic potential of a test or biomarker is conditional on the information obtained from patient history, physical examination and previous tests (30). It is important that a new test or biomarker adds information or value to the existing algorithm. To establish incremental value, head-to-head comparison between the model with new predictors and the reference baseline model (without the new predictors) is the recommended approach, but this may not happen in reality. There is evidence that claims better performance for markers when they are added to poorer performing reference models (31). A large number of novel biomarkers have been identified in the field of imaging and other areas of biomedical research, such as genomics and proteomics (32). These can potentially help improve prediction, diagnostic testing, prognosis, treatment choice and prevention (32, 33). The challenge for clinicians and medical research is to critically appraise the existing and new markers or tests (34). Unfortunately, both experts and lay people do not always understand the methodology of conducting prognostic research (33). A potential biomarker should be developed in a phased approach, as previously suggested (17, 35, 36).

In short, assessing the overall incremental value is the initial step and is commonly achieved by establishing the discriminatory ability of a new test in addition to existing prediction model. AUC is an often-used effect estimate to indicate the discriminative ability of a test of interest. The ROC curve is a plot of the sensitivity (true positive rate) against 1 minus specificity (false positive rate) for consecutive cut-offs for the probability of the outcome of interest, such as

the development of CVD. One of the limitations of AUC is that a threshold must be set, for example a threshold is required for continuous outcomes. AUC can be interpreted as the probability that one factor favours the development of a specific outcome. Useless predictions such as a flipping a coin result in an AUC of 0.5, whilst a perfect prediction has an AUC value of 1. The next step is to establish the potential improvement in predictive performance after adding a novel test to an existing predictive model. The change in area under the receiver operating curve (Δ AUC) quantifies a novel test's added value (37-39).

Modern statistical methods include net reclassification index (NRI) and integrated discrimination index (IDI) which are alternatives to Δ AUC because the size effect of Δ AUC is often disappointingly small. Even a large Δ AUC is not sufficient evidence of clinical usefulness (40). Reclassification can be demonstrated using a table showing the 'upward' or 'downward' movement within a population. An 'upward' movement in categories means subjects are upgraded to a higher risk group, which means reclassifying upwards. A 'downward' movement indicates subjects are downgraded to a lower risk group, meaning reclassifying down. The overall improvement in reclassification can be quantified as a single effect estimate, known as NRI. One issue of NRI is that CVD risk categories often involve 3 groups, therefore the weighted version of NRI is not as straightforward (41). Another option is to calculate the integrated discrimination improvement (IDI), which considers improvements among multiple categories. It is important to note that there is ongoing debate regarding the nature of these measures (42, 43).

Calculation and reporting of AUC is a fundamental step in establishing the discriminatory ability of a test, followed by estimating the Δ AUC to quantify incremental value to existing model. The idea of reporting all of the reclassification measures alongside each other has been proposed (44). There are issues of multiple testing leading to a positive finding of

reclassification by chance. Calibration is often not reported and is essential for an informed decision regarding adoption of the new marker or test. Measures dealing with discrimination are regarded as early-stage research. Decision-analytic measure, such as net benefit, may be a better alternative statistical method (45). Currently, there is no consensus on what is the single best measure but transparent reporting is key (46).Ultimately, the proof of the advantages of adopting a potential biomarker should come from randomised comparison studies to quantify whether the use of the maker improves decision marking and patient relevant outcomes (34).

1.6 Incremental value of biomarkers in addition to Framingham Risk Score

Numerous cardiovascular risk prediction models have been developed between 1967 and 2013 (5). Of these models, only 132 models have been externally validated. To synthesise qualitative and quantitative data, a choice needs to be made to minimise the known heterogeneity in the field of cardiovascular risk prediction. The most effective way is to limit the investigation to a specific baseline model. FRS is the most validated model for all the cardiovascular risk prediction tools (5). The Wilson 1998 iteration is the most validated study (by 89 studies) (3), followed by the Anderson 1991 version (validated by 73 studies) (47), the D'Agostino 2008 version (validated by 44 studies) (9), ATP III 2002 version (validated by 31 studies) (8) and lastly the updated Anderson 1991 version (validated by 30 studies) (48). Given the evidence of validation, FRS has been chosen as the base model for the research process.

As outlined above, multiple studies have demonstrated the predictive ability of CACS in stratifying cardiovascular risk. Attempts have been made to incorporate CACS into routine clinical practice (49). The ability of CACS to reclassify individuals in the decision of initiating statin treatment is still not completely convincing (50). TACS is a less described biomarker

but probably shares similar pathophysiology to CACS. In the setting of chest pain, the diagnostic test accuracy of CTCA is comparable to ICA (the gold standard) (51-56). There is growing interest in utilising CTCA in general cardiovascular risk prediction. The CONFIRM registry is an ongoing multicentre cohort to investigate precisely that. Given the shortcomings of CACS outlined above (21), CTCA should theoretically complement CACS. A systematic review and meta-analysis of the incremental value of CACS, TACS and CTCA in addition to FRS is therefore a logical approach to gather the available evidence. In addition, methodological issues identified may help guide the design of any future trials.

1.7 Opportunistic assessment of the coronary arteries

Among a variety of subclinical indicators of cardiovascular disease, CACS is currently the strongest predictor of cardiovascular events independent of traditional risk factors (13). As discussed, the absence of coronary calcium does not preclude the rupture of non-calcified plaque. CTCA complements CACS because it identifies plaques of different composition and allows the assessment of luminal stenosis. Both CACS and CTCA can only be acquired if the set-up of the CT scan is configured at the outset. However, this information is not routinely obtained when a patient is having a CT scan.

Helical scanning of the thorax has been standard practice since the advent of spiral scanning technology allowing rapid acquisition of volumetric data. In contrast, ECG gating in the thorax has been exclusively performed for CTCA. Traditionally retrospective helical scanning has been performed with ECG gating to allow multiphase analysis of different parts of the cardiac cycle (57). However, ECG gated helical scans are of high dose compared to non-gated helical scans due to the requirement for low pitch scanning. More recently, there has been a move from helical to axial scanning due to the dose savings that can be achieved (58). Further radiation dose reductions can also be achieved by reducing the

amount of tube on time during the cardiac cycle. The majority of coronary artery segments can be evaluated on a single portion of the cardiac cycle (which is heart rate dependent) (59, 60). The combination of these techniques has resulted in radiation levels 80% less than helical scanning of the heart and less than non-ECG gated studies of the thorax (61). It is, therefore, possible that ECG gated studies of the thorax may result in diagnostic images of the coronary arteries at no additional cost regarding radiation dose or contrast administration. Such information could be particularly useful in patients where malignancy is suspected or patients undergoing CT of the thorax as part of the staging process before major surgery.

Non-ECG gated helical scanning of the thorax is the established technique for the staging of cancer patients. However, a step-and-shoot ECG gated technique can be adopted with potential advantages for pulmonary imaging. Accurate measurement of lung nodules is influenced by cardiac related motion artefact and disproportionally affects various parts of the lungs (62). Similarly, peripheral pulmonary arteries move considerably with cardiac motion. Previous work using retrospective scanning found improvement in image quality when compared to non-gated studies in the context of pulmonary angiography studies (63, 64). Boehm et al argued that adopting gating did not improve the diagnostic test accuracy and that routine ECG gating was not justified (65). In contrast, others found improvement in image quality with the use of gating (66-68). To clarify the usefulness of gating for pulmonary imaging, we performed a prospective study of patients undergoing follow-up oncology scanning where the patients received both a gated and non-gated study of the thorax. The scans were then evaluated for cardiac related motion artefacts within the lungs and radiation dose. We also evaluated the coronary arteries to see how many coronary artery segments had diagnostic image quality.

1.8 Research questions

Previous cohort studies have investigated the value of CT biomarkers in addition to FRS. There has been no formal comparison of quantitative synthesis that compares the predictive performance of FRS and CT biomarkers between relevant cohort studies. Systematic reviews and potentially quantitative synthesis would be informative in the evaluation of the potential incremental value of CT biomarkers in addition to FRS across different populations. This would help in the implementation of a more accurate model that could positive affect clinical practice. In order to formally compare different cohort studies, prognostic information of the baseline FRS models and aggregate data from included cohort studies were collected or estimated. The aim was to conduct a systematic review in order to summarise and compare the overall predictive performance of different FRS models and the additional predictive performance provided by CT biomarkers in CVD risk prediction.

During the process of conducting the systematic review as proposed above, it became apparent that data synthesis was very difficult due to the variation in reporting practice. There was also a sense of optimism within the literature where most of the cohort studied claimed either improved discrimination or reclassification with additional CT biomarkers. To complicate matters further, the most popular effect estimates in reclassification, NRI, generated as much controversy as its popularity. These issues together presented an opportunity to investigate these matters further by comparing aggregate data between adequate and inadequate reporting practice.

Although the incremental value of CT biomarkers is yet to be established by the proposed systematic review, it is already technically possible to obtain information on the coronary arteries with slight modification of existing whole-body CT scanning. Oncology patients are often not asymptomatic but are at risk of cardiotoxicity related to chemoradiotherapy. They

are a vulnerable group where more subclinical information provided by CT biomarkers may be helpful in predicting treatment related CVD events. Given some of the described limitations of CACS and recent advances on CTCA, a slight modification of existing CT scanning techniques allows for opportunistic assessment of coronary arteries in oncology patients who are under surveillance after cancer treatments. Accordingly, a feasibility study to look at the technical aspects of assessing the coronary arteries by ECG-gating was conducted.

Chapter 2 Outcome measures

2.1 Outcome measures in prognostic studies in general

Prognostic studies look at subjects from a population at risk of developing a particular outcome within a specific period. The outcome of a study should ideally be objective, where the subjects either experienced or did not experience the outcome, and it should be well defined at the beginning. Reliable methods should be used to verify outcome measures, such as a death certificate and medical records. It is important to note the number of observed and expected events as these can then be used to calculate the observed:expected ratio, providing a rough indication of the overall model calibration. Calibration is an important aspect of predictive model performance in addition to discrimination, but is often overlooked.

Apart from the ideally stringent definition, it is important to realise that there is an additional dimension of time in prognostic outcomes. The period of interest where an outcome can potentially develop can be broadly seen as short term or long term. Of note, the Wilson 1998 model looked at the 10-year CVD risk whereas not all the subsequent validation cohorts used a matching prediction horizon. The variation in chosen prediction horizon is an important factor to consider when doing quantitative analysis. Prognostic prediction examines a longitudinal relationship whereas diagnostic prediction is mostly interested in a cross-sectional relationship. To determine whether an outcome of interest develops over time is a crucial in prognostic studies. To address changes in risk over time, survival or Kaplan Meier curves are used to incorporate the time element, rather than using odds ratio, relative risk or percentage. The follow-up period should be long enough to detect the outcome measures. Excluding those who are lost to follow up would underestimate the outcome of interest.

It is important that subjects included in a prognosis study all have similar risk profile so that meaningful conclusions can be drawn about the observed and expected outcomes. Within studies that adopt various FRS models, the included cohort studies in the systematic review are expected to incorporate a mixture of symptomatic and asymptomatic participants at recruitment. Even though all the studies used comparable outcomes, it is important to regard them differently as the stage of disease will clearly influence prognosis and analysis by disease stage may be necessary. There are different strata of risk with FRS models. For example, asymptomatic low risk individuals would clearly be different from asymptomatic high-risk subjects.

2.2 Cardiovascular outcomes in prediction model studies & systematic review

Most models in the cardiovascular prediction model literature define outcomes broadly based on two approaches. The first approach defines the outcome as either fatal or non-fatal CHD, for example cardiac death or ST segment elevation MI. The second approach involves the above definition in combination with other outcomes of CVD, as described in the first chapter, for example strokes or initiation of statin therapy. The evolvement of various cardiovascular outcome definitions and the different FRS iterations are linked. This is because some variables are better understood over time (for example during or after external validation), and there is further evidence to support or refute any adjustment of existing prediction models (69, 70). Methods of updating models include adjustment of the model's intercept for differences in outcome occurrences between the development and validation data sets, adjustment of the weights of specific predictors and addition, deletion or modification of predictors (34). For example, diabetes was removed from the base model of Wilson 1998 FRS during the development of ATP III 2002. The evidence considered at the time regarded Type 2 diabetes as CHD risk equivalent. People with Type 2 diabetes have a high 10-year risk for both fatal and non-fatal CHD and relatively poor long-term survival after

CHD (5). There is evidence that the definition of cardiovascular outcomes, in reality, is very heterogeneous with over 70 different definitions of CHD and a lack of use of internationally agreed outcome definition, for example the International Classification of Disease definitions (5). The heterogeneity of the outcome definition makes direct comparison difficult and leads to different estimated predictor size effects and model performances. The TRIPOD statement aims to improve the reporting of multivariable prediction models in development or being validated, with specific reference made to more stringent and clear reporting of outcomes and its definition (13). For this review, the primary outcome definition includes cardiac death and non-fatal MI. The diagnosis of an MI was based on a combination of symptoms, electrocardiographic signs, levels of creatinine kinase or troponin T or I, autopsy findings of nonfatal acute MI and coronary death (71). Secondary outcomes include all-cause mortality, major cerebrovascular events, surgical or non-surgical coronary revascularisation, angiographically defined new-onset peripheral vascular disease, hospitalisation for cardiac disease and the initiation of medical therapy for cardiac disease. Further adaptations of these may be required at the evidence synthesis stage.

2.3 Composite cardiovascular endpoints in systematic review

Composite CVD endpoints can be useful in the evaluation of infrequent outcomes and represent a broader range of the beneficial effects of an intervention (72, 73). The main rationale for considering a composite primary outcome is sample size (74). Composite endpoints may involve surrogate outcomes, clinical outcomes or a combination of both. However, clearer reporting of the construct of composite CVD outcomes, in which individual outcomes were selected for inclusion, should be mandatory (75). An effect estimate conditional on composite outcomes is different from a single outcome. There is evidence that the inclusion of each additional endpoint to the composite was associated with fewer participants (75). This is consistent with the argument that the smaller the sample size, the

more endpoints are included in the composite outcome to inflate the number of events. The choice of composite endpoints in cardiovascular trials can be made based on statistical testing (76). A qualitative approach to selecting single or composite CVD outcomes can be based on the prevalence of a target disease (74). The following case of heart failure illustrates that. The prevalence of heart failure positively correlates with left ventricular function. In a population that has just experienced an acute MI, for example in the Eplererone Post-Acute MI Heart Failure Efficacy and Survival Study, most of the participants were expected to experience CVD outcomes (77). The use of all-cause mortality instead of composite CVD endpoints would not be expected to result in a significant loss of sensitivity. In contrast, in a study of participants with baseline stable disease (preserved left ventricular function), a smaller proportion is expected to die from CVD causes. In this situation, the use of an all-cause mortality outcome may result in a loss of power for an intervention aimed at preventing CVD, for example in the Candesartan in Heart Failure Assessment of Reduction Mortality-Preserved Trial (78). Another alternative to a single outcome and single combined endpoint include co-primary outcomes, global index and hierarchical scoring or ranking where each will have respective advantages and disadvantages (74). The widespread use of composite cardiovascular outcome is expected to be a major source of heterogeneity.

2.4 Surrogate endpoints in the feasibility study

Surrogate outcomes can be used as substitutes for ‘hard’ or so-called patient relevant outcomes. These outcomes are pragmatic as they occur more quickly and therefore shorten the duration and resource expenses. Furthermore, surrogate markers can be used as markers to indicate the benefit of the intervention in the absence of ‘hard’ outcomes (79-81). It is essential that any marker used as a surrogate outcome must have an unambiguous association with subsequent patient-relevant outcomes and that there is sound pathophysiology underpinning them (34). As a proof of concept study, objective and

subjective image quality was chosen as an outcome to assess the technical adequacy of the proposed technique. Ideally, a diagnostic test accuracy design is better with ICA being the reference standard. However, this is very unlikely to be approved by the ethics committee, especially as the participants are asymptomatic. In the setting of chest pain, the diagnostic test accuracy of CTCA and ICA are comparable so hopefully this design will suffice.

Chapter 3 - Methods of the systematic review & meta-analysis

3.1 Searches, the screening process & the inclusion/ exclusion criteria

The systematic review is reported in line with Preferred Reporting Items for Systematic (PRISMA) (82). The protocol was registered on the international database of prospectively registered systematic reviews in health and social care (Centre for Reviews and Dissemination, University of York) (PROSPERO) (2015:CRD42015023795) (83). MEDLINE (OVID interface, indexed and non-indexed, 1946 onwards), EMBASE (OVID interface, 1974 onwards), Web of Science (Thomson Reuters interface) and the Cochrane Central Register of Controlled Trials (Wiley interface) were searched in July 2015 with assistance from an information specialist. The syntaxes were adapted appropriately for each database. Appendix 1 contains the search strategy used for MEDLINE. In addition, the bibliographies of all included studies were searched, and a snowballing technique was used. Snowballing involves looking at every reference within an included full text publication, in order to look for potential further studies eligible for inclusion. The searching process stopped when marked duplication within the pool of studies emerged. Authors of abstracts included in the title and abstract screening stage were contacted for full text publications. Grey literature and PhD theses were searched using the key words described above on search engines including EThOs and Open Grey (84, 85). Only full text publications were included. No language restrictions were applied. An update search was carried out in September 2017.

During the publication process, there was additional development of the screening process, data extraction and critical appraisal to include an additional independent assessor and a judicator in the event of disagreement.

Inclusion and exclusion criteria are displayed according to the characteristics based on the population, investigation, comparator, outcomes and setting (PICOS) format as follows:

1. Population (P)
    a. Participants of a cohort study who agreed to have their CVD risk evaluated using ‘novel’ CT biomarkers.
    b. Potential cohort studies can either be an independent cohort or part of an ongoing named cohort. If there are multiple publications from the same named cohort, only the publication with the longest average follow-up is incorporated into meta-analysis.
    c. The baseline risk of these participants was recorded in the format of FRS risk factors or risk score.
2. Investigation (I)
    a. Cross-sectional or diagnostic test accuracy studies are excluded.
    b. Both prospective and retrospective cohort studies that specifically examined Agatston Score/ CACS, TACS or CTCA, alone or in combination, as new predictors in addition to FRS, are included.
    c. Included cohort studies must report at least one summary statistic (indicating incremental value) or provide enough information which allows calculation of at least one summary estimate of the predictors of interest.
    d. Summary statistics (indicating incremental value) includes hazard ratio (HR), the difference in area under the curve (Δ AUC) (86, 87), c-statistic, category-based NRI (88, 89) or another recognised reclassification measure in the literature.
    e. Cohort studies that do not display any summary statistics are excluded.
3. Comparator (C)
    a. Cohort studies that compare the predictors of interest with the FRS model are included. In addition, the incremental value of the predictors of interest

beyond the FRS model needs to be demonstrated. Study design that focuses solely on head-to-head comparison between predictors of interest and the FRS model is excluded.

   b. Cohort studies that compare the predictors of interest with other CVD risk prediction model are excluded.

4. Outcome (O)
   a. All-cause mortality, composite cardiac event, incident hypertension and new treatment relating to onset of cardiac symptoms during the study are included.
   b. These outcomes should be obtained from a reliable source, such as death certificate or hospital records.
5. Setting (S)
   a. At recruitment of cohort studies, participants can either be asymptomatic, symptomatic or a combination of both. The definition of symptoms is broad, which includes both baseline CHD and/ or CVD.
   b. The aim of an included cohort study should be CVD risk evaluation. Cohort studies looking primarily at the evaluation of chest pain or CHD are excluded.

Title and abstract, followed by full-text screening were performed against the above criteria. The original Framingham cohort was constructed as an asymptomatic cohort at baseline. For the purpose of this review, it was initially felt that studies were included regardless of whether patients had CVD symptoms at baseline in order to capture the variability within the literature. The rationale behind this is because the definition of being “symptomatic” among cohort studies is heterogeneous and difficult to assess at the screening stage. Owing to a wide range of symptoms, such as chest pain or having CVD at baseline, the exact nature of baseline ‘symptoms’ will be separated into different strata subsequently. Cohort studies which set out to evaluate chest pain, rather than CVD risk prediction were excluded. The limitations to this approach will be discussed in the summary chapter.

3.2 Data extraction & bias assessment

The CHARMS checklist was used (90). Publications were classified according to the Framingham models whenever possible as defined by Wilson 1998, ATP III 2002 and D'Agnostino 2008 (91-93). Information on imaging parameters, patient demographics and relevant prognostic information was extracted. Original 95% CIs, standard errors (SE), standard deviations (SD) or p-values of summary estimates of interest were extracted. (94, 95). If the Kaplan-Meier survival curve was available, this was noted to enable HR estimation. Study quality was rated using the Quality In Prognosis Studies (QUIPS) tool (96). The included studies were assessed in various aspects, such as study participation, attribution, measurement of prognostic factor, outcome measure and confounding factors.

3.3 Estimating missing hazard ratios from Kaplan-Meier Curves

Logistic regression is a statistically less powerful method than survival analysis and therefore emphasis was placed on obtaining HRs from survival analysis among the included studies. HRs were not reported even when Kaplan-Meier Curve was available. A method applicable to restricted mean survival time was used to extract HRs from Kaplan-Meier Curves (97). In brief, this method required manual tracking of each of the Kaplan-Meier Curves to generate outputs. These were subsequently input into STATA version 14.0 (StataCorp, College Station, Texas) to generate an estimated hazard ratio and their respective 95% confidence intervals (95% CIs). The essential information that enables this estimation method includes the number at risk at the start point of survival analysis and the choice of desired reference group.

3.4 Estimating confidence intervals of area under the receiver operating curve

The 95% CIs of Δ AUC were often missing but the relevant P-values were reported. CIs are considered more informative than P-values for significance tests (98, 99). Altman et al demonstrated CIs can be obtained when only the aggregate size effect and the P-value are reported (100). Δ AUC was considered as a mean difference and therefore the proposed method in this context is applicable. Most of the missing 95% CIs of Δ AUC have been extracted using this method. When the relevant P-values were not available, the 95% CIs from the SD were estimated according to the Cochrane Handbook (101). For studies where 95% CIs or SE were not reported, a correlation coefficient (*r*) of 0.3 between FRS and CACS/ CTCA was used to allow estimation of the 95% CIs for Δ AUC based on data from (102).

3.5 The calculation of net reclassification index using reclassification table

Some studies reported NRI but not with the 95% CIs. The reported NRIs were often the overall NRI, without the relevant components, specifically event and non-event NRIs. When a reclassification table is available, the 95% CIs and their components can be calculated. The individual components of NRI can be displayed as percentages or proportions. However, the overall NRI is unitless as it is not a proportion, but the sum of two proportions. This is important and it is frequently displayed and interpreted wrongly in the literature. The information required to calculate the overall and sub-components of NRI is shown in the formula below:

Calculation of event NRI:

Probability of up event = (Up events/ Total events)

Probability of down event = (Down events/ Total events)

Event NRI = Probability of up event – Probability of down event

Event SE = √ [(Probability of up-event – Probability of down-event)/ Total number of events]

Event NRI upper/ lower 95% CIs = Event NRI +/- (1.96 x Event NRI SE)

Calculation of non-event NRI:

Probability of up non-event = (Up non-events/ Total non-events)

Probability of non-event = (Down non-events/ Total non-events)

Non-event NRI = Probability of up non-event – Probability of down non-event

Non-event SE = √ (Probability of up-event – Probability of down-event)/ Total number of events)

Non-event NRI upper/ lower 95% CIs = Non-event NRI +/- (1.96 x Non-event NRI SE)

Calculation of overall NRI:

Overall NRI = Event NRI + Non-event NRI

Overall NRI SE =√ [(SE of event NRI) ^{2} + (SE of non-event NRI) ^{2}]

Overall NRI upper/ lower 95% CIs = Overall NRI +/- (1.96 x Overall NRI SE)

3.6 Evidence synthesis

There were multiple studies with multiple summary estimates and missing data. As a result, there were many potential groups eligible for meta-analysis. This led to the development of specific strategies to efficiently identify meaningful clinical groups qualitatively. According to the pre-defined systematic review protocol, the overall aim was to minimise both clinical and statistical heterogeneity. For each size effect that indicates incremental value, the overarching strategy was to minimise heterogeneity by matching the outcomes, FRS iterations, cut-point, baseline risk, symptoms and adjustment. For reclassification and discrimination measures, the prediction horizon of studies was matched.

3.7 Meta-analysis

Statistical pooling of relevant prognostic or time-to-event summary statistics was attempted. A weighted average of any relevant summary statistics across subgroups within a study was obtained (101). Statistical heterogeneity was assessed using the I^2 statistic (103). Pooling was attempted provided there was at most moderate statistical ($I^2 \leq 50\%$) and clinical heterogeneity. A random-effects meta-analysis model was used for pooling individual study estimates and the overall estimates were expressed as pooled summary statistics with 95% CIs (104). All statistical testing was two-sided and $p<0.05$ was considered statistically significant. In the secondary analysis, the effect size of the adjusted and unadjusted summary statistics used in the meta-analysis was compared. For example, comparison between adjusted and unadjusted hazard ratios. Meta-analyses were carried out using STATA version 14.0 (StataCorp, College Station, Texas).

3.8 Unplanned methodology work

During data extraction, considerable variation across studies was encountered, making direct comparison with different populations difficult. For example, the number of risk categories, iterations of FRS and thresholds in cardiovascular outcomes were all different. Although methodology work was not initially planned, the described differences prompted further investigation to examine the potential impact of this observed variation. In addition, there was also a variation of reporting practice when it came to summary statistics related to prognosis. The comparison of good and poor reporting practice was also evaluated.

Chapter 4 – Results & discussion of the systematic review & meta-analysis

Section 1: Search results, calculations & evidence synthesis

4.1 Literature search results

A total of 801 unique hits were screened, with 35 studies encompassing 206,663 patients (7, 16, 18, 21, 22, 102, 105-133) meeting the inclusion criteria (Figure 1). Appendix 2 shows the studies that investigated CACS or TACS and their individual inclusion and exclusion criteria. Appendix 3 shows the studies that investigated CTCA and the relevant criteria. To obtain unpublished publications, authors of conference abstracts were contacted up to a maximum of 3 times and a log was kept. Thirty-three authors from unpublished studies who might have additional full publications were identified. One conference abstract was associated with a full publication but did not fit the inclusion criteria. Fifteen authors were associated with relevant observational cohorts, for example MESA, and full publications were identified during the full-text search. We contacted 11 authors who might have additional full publications and received 4 replies which yielded 1 full publication (110) that was included. Six authors' emails could not be identified.

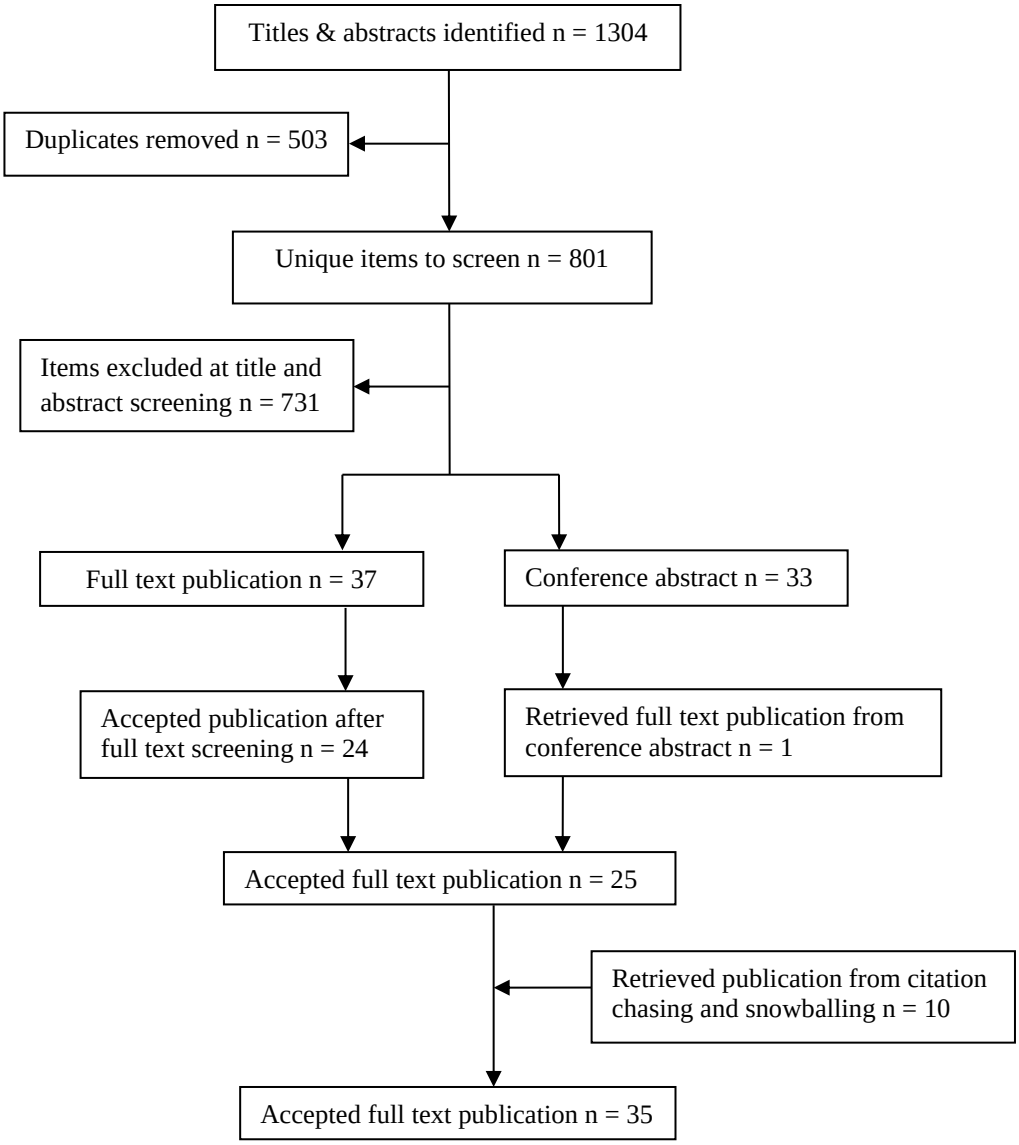

Figure 1. PRISMA-P flow chart

4.2 Assessment of bias within the included studies

Using the QUIPS bias assessment tool (96), the included studies were predominantly at low risk of bias concerning study participation, measurement of CACS/ CTCA/ TACS, outcome measure and confounding factors. There was low to moderate risk of bias for statistical analysis as several studies selectively reported results and/ were not clear about the process

of model building. Most studies were at high risk of attrition bias because there was a notable amount of missing data and the number of participants lost to follow-up was not accounted for. Figure 2 shows the overall quality assessment using the QUIPS bias assessment tool.

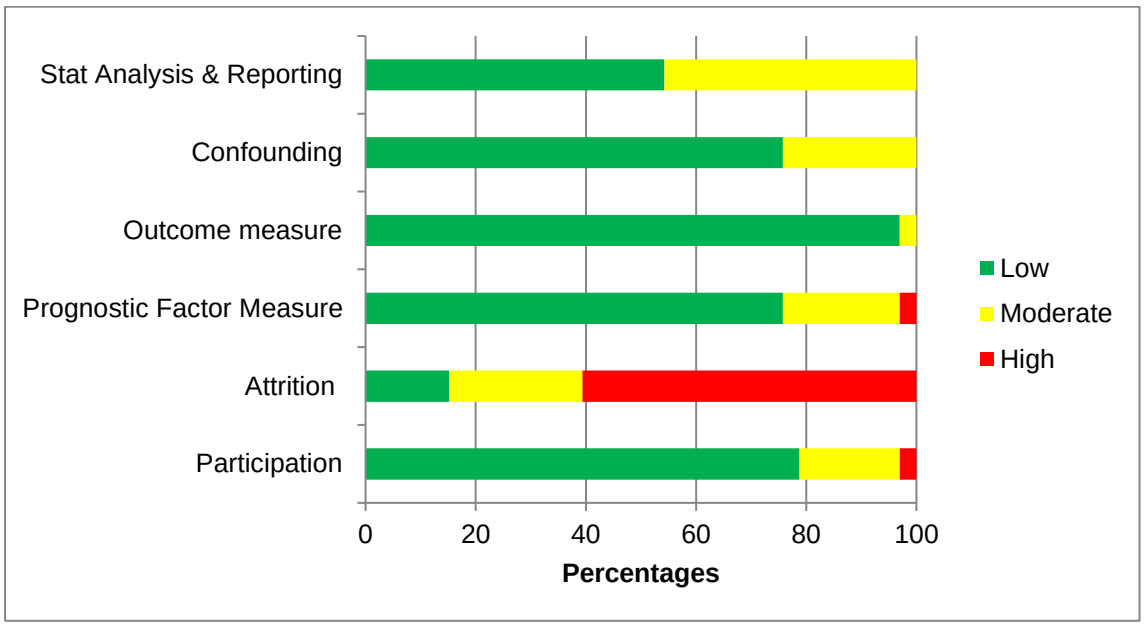

Figure 2. QUIPS bias assessment tool

4.3 Demographics of included cohort studies

Calcium & thoracic calcium score: Included studies' demographics

Thirty-one studies encompassing 165,861 patients met our inclusion criteria. The CONFIRM cohort had 2 studies (124, 132), the DHS cohort had 1 study (122), the EISNER cohort had 2 studies (21, 121), the HNR cohort had 5 studies (7, 112, 117, 118, 127), the MESA cohort had 5 studies (16, 109, 113, 115, 116, 119, 123, 126), the Rotterdam cohort had 2 studies (16, 109), the St Francis Heart cohort had 1 study (123) and there were 13 independent studies (18, 102, 105-108, 110, 111, 114, 120, 125, 131, 133). The clinical classification means that the meta-analysis of time-to-event measures was conducted according to the categorical or logarithmic classification of CACS/TACS.

Table 1 shows the demographic information of the included studies. The studies were conducted worldwide, predominantly using a prospective design. A total of 165,861 participants (men = 97085, 56%) either with or without symptoms at enrolment were evaluated with either CACS or TACS. None of the participants were known to have cardiovascular disease. The Wilson 1998 and Adult Treatment Panel III 2002 were the most popular models used. A variety of risk thresholds have been used with most studies adopting 3 or 4 cardiovascular risk categories.

Computed tomographic coronary angiogram: Included studies' demographics

There were 9 included studies of which the CONFIRM cohort had 4 studies (22, 124, 128, 132) and the remaining 5 were independent studies. The clinical classification means that the meta-analysis of the time-to-event measure was conducted according to the number of vessels that had obstructive coronary disease. Unspecified obstructive coronary disease was classed as 1-vessel disease. The included studies were conducted in North America, Europe and Asia from 2003 and 2010. Information regarding scanning parameters of CTCA and image quality was inadequate. Although there were a variety of CT scanners used, most were 64-slice multi-detector scanners.

A total of 60,687 participants (men = 33,193, 57%) either with or without symptoms at enrolment were evaluated with CTCA. No participants were known to have cardiovascular disease. The most commonly used FRS was Framingham/ Wilson 1998 (3). The included studies that solely had symptomatic participants had the highest crude mortality rate. Table 2 shows the other characteristics of the included studies.

4.4 Technology

In the period investigated, there was a transition from electron beam CT to multi-detector technology. All the studies used the Agatston method to measure both CACS and TACS (113). In addition, all 31 included studies reported CACS however only 3 studies reported TACS. A variety of CT scanners were used conducting CTCA using 64-slice multi-detector scanners (see information in subsequent chapters). All studies used the Agatston method to measure both CACS & TACS (113). Retrospective ECG-gating yielded the highest radiation dose. Information regarding scanning parameters of CACS can be found in Table 3. The mean body mass index (BMI) was between 24 and 29.1.

Studies	N	Cohort	Prospective	Symptoms	Age, years	Framingham Risk Score	Lost-to-follow-up, %	Follow Up, months	Horizon, years	Men, %	Event Rate, %
Agarwal et al, 2013	1443	DHS	Yes	No	61.4	n/a	22.2	88	7.4	45.9	0.1
Ahmadi et al, 2011	730	Independent	No	No	61.0	2002	n/a	48	10	87.1	7.3
Arad et al, 2005	4903	St Francis Heart	Yes	No	59.0	?2002	5.9	51.6^^	10	65.0	2.6
Budoff et al, 2007	25253	Independent	Yes	No	56.0	2002	0.0	81.6	10	54.1	2.0
Chang et al, 2015	988	Independent	Yes	Yes	57.5	2002	4.3	82.8^	10	75.3	11.2
Cho et al, 2012	7590	CONFIRM	Yes	No	58.0	1998	35.8	24	2.5	60.9	1.5
Cho et al, 2015	3217	CONFIRM	Yes	No	57.0	1998	n/a	24^	2.5	63.1	1.8
Elias-Smale et al, 2010	2040	Rotterdam	Yes	No	69.6	1998	0.6	110.4^	10	42.6	6.7
Elias-Smale et al, 2011	2196	HNR	Yes	No	69.2	2002	2.0	42^	5	45.2	5.1
Erbel et al, 2010	4487	HNR	Yes	No	59.3 or 62.9	1998 & 2002	8.0	61,2, 60^	10	47.3	2.3
Forouzandeh et al, 2013	760	Independent	No	Yes	54.4	1998 + BMI	n/a	39.6^	n/a	40.8	5.9
Gibson et al, 2014	6814	MESA	Yes	No	61.9 or 67.9	Framingham Stroke Risk Score	0.5	114	n/a	47.2	3.5

Greenland et al, 2004	1031	Independent	Yes	No	65.7	2002	0.2	75.9	10	90.1	8.2
Hadamitzky et al, 2010	451	Independent	No	No	58.7	1998	5.1	27.5^	n/a	74.1	2.2
Han et al, 2015	34386	Independent	No	No	53.8	2002	36.4	57.6^	10	77.1	0.9
Hermann et al, 2013	4180	HNR	Yes	No	59.2	1998	n/a	94.9^^^	n/a	47.1	2.2
Kavousi et al, 2012	5933	Rotterdam	Yes	No	69.1	1998	0.3	81.6^	10	40.5	5.8
Lau et al, 2012	151	Independent	Yes	No	60.9	1998**	0.0	61	10	40.4	11.3
Matsushita et al, 2015	6562	MESA	Yes	No	62.0	2008	0.1	100.8^	n/a	48.0	9.9
Mohlenkamp et al, 2011	2238	HNR	No	No	56.6	1998	13.6	61.2^	5	30.8	2.8
Mohlenkamp et al, 2011	4338	HNR	Yes	No	59.2 or 62.8	1998	8.6	60^	10	46.5	2.3
Park et al, 2013	7071	Independent	No	No	53.0	1998	26.7	48^	n/a	60.5	1.3
Polonsky et al, 2010	5931	MESA	Yes	No	62.0	2002	0.9	69.6^	5	46.4	3.6
Raggi et al, 2001	676	Independent	Yes	No	52 or 55	n/a	n/a	32	n/a	50.9	4.4
Raggi et al, 2004	10377	Independent	Yes	No	52 or 55	2002	n/a	60	5	59.6	2.4

Rana et al, 2012	1286	EISNER	Yes	No	58.6	2008	0.5	49.2	4	52.8	2.7
Valenti et al, 2015	9715	Independent	Yes	No	53.4	1998 & 2002	n/a	175.2	10	59.3	9.6
Versteylen et al, 2013	1556	Independent	Unknown	Yes	60.0	2008	19.9	26	n/a	57.7	1.8
Wong et al, 2009	2483	EISNER*	Yes	No	55.7	2002	7.2	52.8^	10	61.6	0.7, 1.8 or 2.0
Yeboah et al, 2012	1330	MESA	Yes	No	63.8	2002	n/a	90	7.5 for NRI	66.7	7.1 or 9.2
Yeboah et al, 2014	5745	MESA	Yes	No	62.1	1998	n/a	90^	10	46.1	4.4, 5.6 or 6.0

Follow-up period is expressed as the mean of the population unless otherwise specified. ^ = median; ^^ = as total; ^^^ = unclear whether the results represent median or mean;
* = This is consisted of a mixture of EISNER and a separate independent cohort ** = calibrated for the Chinese population

Table 1. The characteristics of the 31 included studies

Studies	N	Cohort	Prospective	Symptoms	Age, year	Male %	FRS type	Low %	Medium %	High %	FU, months	Event %	CT	Dose mSv	CMR %
Cho 2012 (Asia, Europe &	7590	CONFIRM	Yes	No	58	60.9	1998	48	33	19	24	1.5	MDCT	n/a	7.60

North America)													64		
Cho 2015 (Asia, Europe & North America)	3217	CONFIRM	Yes	No	57	63.1	1998	n/a	n/a	n/a	24	1.8	MDCT 64	*10.1	9.01
Chow 2010 (North America)	2172	n/a	Yes	Both	58	52.6	2002	8	64	28	16	2.9	MDCT 64	14.9	22.04
Chow 2011 (Asia, Europe & North America)	14064	CONFIRM	Yes	Yes	57	50.9	2002 + LVEF	24	58	18	22.5	1.9	MDCT 64	10.7	10.35
Hadamitzky 2010 (Europe)	451	n/a	No	No	59	74.1	1998	33	56	11	27.5	2.2	MDCT 16/64	7.9/5.7	9.68
Hadamitzky 2013 (Asia, Europe & North America)	21902	CONFIRM	No	Yes	57/58	53.0	1998 & 2002	51	32	17	27.6/18	1.8/1.2	MDCT 64	n/a	11.88
Lin 2011 (North America)	2664	n/a	Yes	Both	53	42.1	Modified FRS	Median = 11	n/a	n/a	37.2	2.2	MDCT 64	3-18	6.74
Park 2013 (Asia)	7071	n/a	No	No	53	60.5	1998	42	53	5	48	1.3	MDCT 64	12.5-13	3.14
Versteylen 2013 (Europe)	1556	n/a	n/a	Yes	60	57.7	2008	Median = 26	n/a	n/a	26	19.8	MDCT 64	5.7	8.11
Abbreviations: CACS = calcium score; CMR = crude mortality rate expressed in 1000 patients years; FRS = Framingham Risk Score; FU = follow-up; MDCT = multi-detector CT; n/a = not available; LVEF = left ventricular ejection fraction. *Conversion factor = 0.014															

Table 2. The design and participant characteristics of the 9 included studies

Studies, Year	CT Machine & Manufacturer	Scanner Type	Dose (mSv)
Agarwal et al, 2013	GE HiSpeed Lx with Smart Score Cardiac scan package	Single slice	n/a
Ahmadi et al, 2011	Siemens Definition dual-source 64-slice scanner	MDCT	n/a
Arad et al, 2005	GE Imatron C-150 XP	EBCT	n/a
Budoff et al, 2007	GE Imatron C-150 XP	EBCT	0.6
Chang et al, 2015	GE Imatron C-150	EBCT	n/a
Cho et al, 2012	64-slice CT	MDCT	n/a
Cho et al, 2015	64-slice CT with single or dual source	MDCT	*10.1
Elias-Smale et al, 2010	GE Imatron C-150	EBCT	n/a
Elias-Smale et al, 2011	SOMATOM Sensation 16-slice or 64-slice CT	EBCT	2.1
Erbel et al, 2010	GE Imatron C-100 or C-150	EBCT	n/a
Forouzandeh et al, 2013	Philips Prededence 16-slice	MDCT	n/a
Gibson et al, 2014	GE Imatron C150, Volume Zoom 4-detector row CT system, LightSpeed 4-detector row CT systems (LightSpeed QXi & LightSpeed Plus)	EBCT/MDCT	0.5-6.5
Greenland et al, 2004	n/a	n/a	n/a

Hadamitzky et al, 2010	Siemens Somatom Sensation 16-slice system, 64-slice singe & dual-source system	MDCT	n/a
Han et al, 2015	Philips Brillaince 256 iCT, Philips Brilliance 40 channel MDCT, Siemens 16-slice Sensation or GE 64-slice Lightspeed	MDCT	n/a
Hermann et al, 2013	GE C-150 Imatron	EBCT	n/a
Kavousi et al, 2012	GE Imatron C-150, Simens Somatom Sensation 16 or 64	EBCT	n/a
Lau et al, 2012	GE Lightspeed 64	MDCT	n/a
Matsushita et al, 2015	Same as Gibson 2014	MDCT	0.5-6.5
Mohlenkamp et al, 2011	GE Imatron C-150	EBCT/MDCT	n/a
Mohlenkamp et al, 2011	GE Imatron C-150	EBCT	n/a
Park et al, 2013	Philips Brilliance 64	EBCT	12.5-13
Polonsky et al, 2010	n/a	EBCT/MDCT	n/a
Raggi et al, 2001	GE Imatron C-100	EBCT	n/a
Raggi et al, 2004	GE Imatron C-100 or C-150	EBCT	n/a
Rana et al, 2012	GE Imatron or Siemens MDCT	EBCT	1-2
Valenti et al, 2015	C-100 or C-150 Ultrafast Imatron	EBCT	1
Versteylen et	Philips Brilliance 64 & Siemens Definition First generation	MDCT	5.7

al, 2013	dual-source		
Wong et al, 2009	GE C-150XP or Siemens 4-slice Somatom Volume Zoon	EBCT/MDCT	n/a
Yeboah et al, 2012	n/a	EBCT/MDCT	n/a
Yeboah et al, 2014	n/a	EBCT/MDCT	n/a
*Conversion factor = 0.014. Abbreviations: CT = computed tomography; EBCT = electron beam CT; GE = General Electric; MDCT = multi-detector CT; mSv = millisieverts; n/a = not available			

Table 3. The vendor and technology of the scanners to determine CACS

4.5 Reported outcomes & thresholds in Framingham Risk Score

To meaningfully summarise the results, the outcomes were divided into 4 categories: stroke events; all-cause mortality; composite cardiac events including and excluding all-cause mortality (Table 4). This classification deviated from the original protocol due to the abundance of composite cardiovascular outcomes. Further analysis was conducted regarding the use of these outcomes. Appendix 2 and 3 illustrate the definition of endpoints (as well as inclusion and exclusion criteria) used within the studies.

Predictor	**Outcome**	**Number of studies**
CACS	All-cause mortality only	5
	Composite cardiac events only (excluding all-cause mortality)	16
	Composite cardiac events (including all-cause mortality)	5
	Stroke events only	2
TACS	Composite cardiac events only (excluding all-cause mortality)	3
CTCA	All-cause mortality only	3
	Composite cardiac events only (excluding all-cause mortality)	3
	Composite cardiac events (including all-cause mortality)	3

Table 4. The reported outcomes of the 35 included studies

Most participants had estimated low or intermediate baseline risk, but the choice of cut-off was variable. All studies used 2 or 3 cut-offs for categorisation of cardiovascular risk but definitions differed (Table 5). The presence of at least 2 cut-offs for cardiovascular risk estimation meaning a straightforward correlation between category based NRIs and Δ AUCs was not possible.

Study, Year	Cut-offs
Agarwal et al, 2013	<7, 7-<20 & >20
Ahmadi et al, 2011	<10, 10-20 & >20
Arad et al, 2005	<10, 10-20 & >20
Budoff et al, 2007	n/a
Chang et al, 2015	<6, 6-20 & >20
Cho et al, 2012	<10, 10-15, 16-20 & >20
Cho et al, 2015	<10, 10-15, 16-20 & >20
Chow et al, 2010	<10, 10-20 & >20
Chow et al, 2011	<10, 10-20 & >20
Elias-Smale et al, 2010	<10, 10-20 & >20
Elias-Smale et al, 2011	<5, 5-10 & >10
Erbel et al, 2010	<6, <10, 6-20, 10-20 & >20
Forouzandeh et al, 2013	n/a
Gibson et al, 2014	n/a
Greenland et al, 2004	<9, 10-15, 16-20 & >21
Hadamitzky et al, 2010	<6, 6-20 & >20
Hadamitzky et al, 201	<10, 10-20 & >20
Han et al, 2015	<10, 10-15, 15-20 & >20
Hermann et al, 2013	<10, 10-20 & >20
Kavousi et al, 2012	<10, 10-20 & >20
Lau et al, 2012	n/a
Lin et al, 2011	<1.4, 1.4-1.8, >1.8
Matsushita et al, 2015	n/a
Mohlenkamp et al, 2011	<3, 3-10 & >10
Mohlenkamp et al, 2011	<10, 10-20 & >20
Park et al, 2013	<5, 5-15 & >15
Polonsky et al, 2010	<3, 3-10 & >10
Raggi et al, 2001	n/a

Raggi et al, 2004	<0.6, 0.6-2 & >2
Rana et al, 2012	<2.4, 2.4-8 & >8
Valenti et al, 2015	<10, 10-20 & >20
Versteylen et al, 2013	n/a
Wong et al, 2009	n/a
Yeboah et al, 2012	<5, 5-20 & >20
Yeboah et al, 2014	<5, 5-20 & >20

Table 5. Thresholds used for Framingham Risk Score

4.6 Modelling methods, prediction horizons, validation & calibration

Two-thirds of the included studies applied the traditional Cox proportional hazard model. Two studies used the stratified or stepwise Cox proportional hazard model (105, 106) whilst 2 studies used the Weibull hazard model (109, 112). Three studies adopted logistic regression (120, 122, 123) and 4 studies did not specify the model used (7, 116, 125, 132). The majority of studies predicted outcomes for 10 years. Other studies used a mixture of unspecified and specified time frames and some did not specify the time frame. Of note, some studies' prediction horizons correlate with their maximum follow-up period. Only one-third of the included studies reported information about validation and calibration (7, 22, 105, 109, 112, 113, 116-120, 124, 128, 130-132) (see Table 6).

Investigation	Year	Author	Model Information	Note
CACS	2011	Ahmadi	Bootstrapping internal validation	Unspecified N
CACS	2012	Cho	Bootstrapping internal validation	Unspecified N
CACS	2010	Elias-Smale	Bootstrapping internal validation	N = 150
CACS	2011	Elias-Smale	Bootstrapping internal validation	N = 150
CACS	2010	Erbel	Calibration with Hosmer-	X^2 = 15.5, p =

			Lemeshaw test	0.05
CACS	2015	Matsushita	Calibration with modified Hosmer-Lemeshaw test	X^2 = <18, no p-value
CACS	2011	Mohlenkamp	Calibration with Hosmer-Lemeshaw test	X^2 = 1.3, p = 0.74
CACS	2011	Mohlenkamp	Calibration with Hosmer-Lemeshaw test	X^2 = 7.1, p = 0.53
CACS	2010	Polonsky	Calibration with Hosmer-Lemeshaw test	X^2 = 9.15, p = 0.24
CACS	2001	Raggi	Calibration with Hosmer-Lemeshaw test	Not specified
CACS	2014	Yeboah	Calibration with Hosmer-Lemeshaw test	X^2 = 8.42, p = 0.42
TACS	2011	Elias-Smale	Bootstrapping internal validation	N = 150
CTCA	2015	Cho	Bootstrapping internal validation	Unspecified N
CTCA	2011	Chow	Overfitting considered	Unspecified
CTCA	2013	Hadamitzky	Internal & external validation	Internal = splitting test samples; external = external data set
CTCA	2010	Chow	Overfitting considered	Sensitivity analysis
CTCA	2013	Park	Bootstrapping internal validation	N = 1000

Table 6. Information regarding validation and calibration

4.7 Estimation of size effect: Hazard ratio, area under the operating curve & net reclassification index

Estimating missing hazard ratios from Kaplan Meier curve

Five studies were included for meta-analysis after assessment of conceptual, clinical and statistical heterogeneity but had missing HRs (106, 108, 110, 111, 131). Using the described method in Chapter 3, the source of Kaplan-Meier Curves in the relevant publication and the reference group used is shown below:

- Using Figure 1a from Budoff 2007 (106), the following HRs and respective 95% CIs were extracted:
    - Reference group: CACS = 0, HR = 1
    - Estimated group 1: CACS = 1-10, HR = 2.26, (95% CIs 1.58 to 3.24)
    - Estimated group 2: CACS = 11-100, HR = 2.14 (95% CIs 1.86 to 2.47)
    - Estimated group 3: CACS = 101-299, HR = 1.86 (95% CIs 1.69 to 2.06)
    - Estimated group 4: CACS = 300-399, HR = 1.84 (95% CIs 1.69 to 2.00)
    - Estimated group 5: CACS = 400-699, HR = 1.68 (95% CIs 1.56 to 1.82)
    - Estimated group 6: CACS = 700-999, HR = 1.69 (95% CIs 1.56 to 1.72)
    - Estimated group 7: CACS = >1000, HR = 1.37 (95% CIs 1.29 to 1.46)
- Using Figure 1c from Chang 2015 (108), the following HRs and respective 95% CI were extracted:
    - Reference group: CACS = 0, HR = 1
    - Estimated group 1: CACS = 11-100, HR = 1.505 (95% CIs 0.682 to 3.319)
    - Estimated group 2: CACS = 101-400, HR = 1.440 (95% CIs 1.012 to 2.048)
    - Estimated group 3: CACS = >400, HR = 1.751 (95% CIs 1.408 to 2.177)
- Using Figure 4a from Foronzandeh 2013 (110), the following HRs and respective 95% CI were extracted:
    - Reference group: CACS = 0, HR = 1

    - Estimated group 1: CACS = 11-100, HR = 6.12 (95% CIs 2.15 to 17.36)
    - Estimated group 2: CACS = 101-400, HR = 3.92 (95% CIs 2.31 to 6.65)
    - Estimated group 3: CACS = >400, HR = 2.99 (95% CIs 2.14 to 4.16)
- Using Figure 2 from Hadamitzky 2010 (111), the following HRs and respective 95% CIs were extracted:
    - Reference group: Non-obstructive CAD, HR = 1
    - Estimated group 1: Obstructive CAD, HR = 23.45 (95% CIs 2.89 to 190.60)

As discussed, the choice of reference group influences the estimated HRs as illustrated below:

- Using Figure 3 from Park 2013 (131), the following HRs and respective 95% CI were extracted:
    - Reference group 1: degree of epicardial stenosis = 0%, HR = 1
    - Reference group 2: degree of epicardial stenosis = 1-49%, HR = 1
    - Estimated group 1 (using reference group 1): degree of epicardial stenosis 1-49%, HR = 2.83 (95% CIs 1.47 to 5.43)
    - Estimated group 2 (using reference group 1): degree of epicardial stenosis = 50-69% or left main stem stenosis 1-49%, HR = 3.24 (95% CIs 2.28 to 4.63)
    - Estimated group 2 (using reference group 2): degree of epicardial stenosis = 50-69% or left main stem stenosis 1-49%, HR = 2.03 (95% CIs 1.47 to 2.79)
    - Estimated group 3 (using reference group 1): degree of epicardial stenosis >/= 70% or left main stem stenosis >/= 50%, HR = 3.05 (95% CIs 2.47 to 3.76)
    - Estimated group 3 (using reference group 2): degree of epicardial stenosis >/= 70% or left main stem stenosis >/= 50%, HR = 1.69 (95% CIs 1.43 to 2.00)

Comparison of adjusted & unadjusted hazard ratios

Most reported HRs were not adjusted and therefore these results should be interpreted cautiously as overestimation is a concern. Two studies provided both adjusted and unadjusted HR. The pooled HR concerning the number of vessels with obstructive disease reduced after adjustment for FRS variables (124, 129). The following illustrates the difference:

- Subgroup 1: One-vessel disease
    - Unadjusted HR = 2.48 (95% CIs 1.30 to 3.66)
    - Adjusted HR = 1.42 (95% CIs 0.73 to 2.12)
- Subgroup 2: Two-vessel disease
    - Unadjusted HR = 4.16 (95% CIs 2.07 to 6.25)
    - Adjusted HR = 2.10 (95% CIs 0.98 to 3.22)
- Subgroup 3: Three-vessel disease
    - Unadjusted HR = 8.04 (95% CIs 4.06 to 12.07)
    - Adjusted HR = 2.91 (95% CIs 1.55 to 5.47)

Estimating confidence intervals of area under the receiver operating curve

Selective reporting of Δ AUC is prevalent; only 7 publications provided 95% CIs of Δ AUC (16, 102, 116, 127, 130, 132, 133). In addition to the description in the method section, four publications (102, 123-125) required estimation from a method that derives AUCs from the same population (94). In one publication (123) it was unclear whether the stated values represent SE or SD and the method described in (94) was used.

Weighted average

Cut points between studies were not comparable and therefore aggregate data was sought to enable meaningful comparison across studies. A weighted average of HRs or Δ AUC and their respective 95% CIs across subgroups within a study were obtained according to the Cochrane Handbook (101). This method used the number of participants and the SD of each subgroup. The 3 publications below required this method:

- Cho 2015 (132), effect estimate = Δ AUC
    - Subgroup 1: CACS 0-10, n=1818, Δ AUC = 0.07 (95% CIs -0.02 to 0.17, SD 2.07)
    - Subgroup 2: CACS 11-100, n=574, Δ AUC = 0.05 (95% CIs -0.05 to 0.14, SD 1.16)
    - Subgroup 3: CACS 101-400, n=472, Δ AUC = 0.13 (95% CIs 0.03 to 0.23, SD 1.11)
    - Subgroup 4: CACS 401-1000, n=217, Δ AUC = 0.17 (95% CIs -0.03 to 0.37, SD 1.50)
    - Subgroup 5: CACS >1000, n=136, Δ AUC = 0.23 (95% CIs 0.05 to 0.40, SD 1.04)
    - Weighted average: All CACS, n=3217, Δ AUC = 0.09 (95% CIs 0.03 to 0.15, SD 1.74)
- Raggi 2004 (107), effect estimate = Δ AUC
    - Subgroup 1: Men, n = 6186, Δ AUC = 0.07 (95% CIs 0.03 to 0.10, SD 1.32)
    - Subgroup 2: Women, n = 4191, Δ AUC = 0.05 (95% CIs 0.03 to 0.07, SD 0.79)
    - Weighted average: Men & women, n=10377, Δ AUC = 0.06 (95% CIs 0.04 to 0.08, SD 1.14)
- Budoff 2007 (106), effect estimate = HR
    - Subgroup 1: CACS 1-10 n=3567, HR = 2.26 (95% CIs 1.58 to 3.24, SD 25.29)

- Subgroup 2: CACS 11-100, n=5032, HR = 2.14 (95% CIs 1.86 to 2.47, SD 11.04)
- Weighted average: CACS 1-100, n=8599, HR = 2.19 (95% CIs 1.46 to 2.92, SD 34.5)
- Subgroup 3: CACS 101-299, n=2616, HR = 1.86 (95% CIs 1.69 to 2.06, SD 4.83)
- Subgroup 4: CACS 300-399, n=561, HR = 1.89 (95% CIs 1.69 to 2.00, SD 1.87)
- Weighted average: CACS 101-399, n=3177, HR = 1.87 (95% CIs 1.71 to 2.02, SD 4.45)
- Subgroup 5: CACS 400-699, n=955, HR = 1.68 (95% CIs 1.56 to 1.82, SD 2.05)
- Subgroup 6: CACS 700-999, n=514, HR = 1.69 (95% CIs 1.56 to 1.72, SD 0.93)
- Subgroup 7: CACS >1000, n=964, HR = 1.37 (95% CIs 1.29 to 1.46, SD 1.35)
- Weighted average: CACS >400, n=2433, HR = 1.56 (95% CIs 1.50 to 1.62, SD 1.60)

The calculation of net reclassification index using reclassification table

NRI was not recorded for TACS. Of the studies reporting CTCA only 1 recorded continuous NRI therefore pooling was not possible. For CACS, 11 of 31 studies reported NRIs but none shared the same baseline characteristics or outcomes so again pooling was not possible. However, 9 of 29 studies provided detailed reclassification tables (21-24, 27, 31, 33, 47, 48) that enabled the calculation of categorical NRIs

Each formula produces 3 main outputs, event NRI, non-event NRI and the overall NRI. Three of the 9 studies shared the same prediction horizon (5 years). They also had similar risk thresholds and subgroups, outlined below:

- Elias-Smale 2011; cut-point 1= FRS at 10%, cut-point 2 = FRS at 20%; subgroup 1 = FRS <5%, subgroup 2 = FRS 5-10%, subgroup 3 = FRS >10%, subgroup 4 = all subjects
- Mohlenkamp 2011; cut-point 1 = FRS at 3%, cut-point 2 = FRS at 10%; subgroup 1 = FRS <3%, subgroup 2 = FRS 3-10%, subgroup 3 = FRS > 10%, subgroup 4 = all subjects
- Polonsky 2010 is exactly the same as Mohlenkamp 2011.

Calculations were made for each of the 3 studies described above. Each contained 4 subgroups. As a result, there were 12 outputs for each study and in total 36 outputs for 3 studies. An example of the calculation of individual NRI components and overall NRI is as illustrated below:

Elias-Smale 2011 (overall group as one of the 4 subgroups)

Event NRI = (17/110) – (6/110) = 0.12

Event SE = 0.04

Event NRI 95% lower CI = 0.02

Event NRI 95% upper CI = 0.15

Non-event NRI = (138/2039) – (104/2039) = -0.02

Non-event SE = 0.01

Event NRI 95% lower CI = -0.03

Event NRI 95% upper CI = 0.00

Overall NRI = 0.08

Overall NRI SE = 0.04

Overall NRI 95% lower CI = -0.00

Overall NRI 95% upper CI = 0.17

4.8 Evidence synthesis

As previously discussed in Chapter 3, there were many potential groups eligible for meta-analysis. This led to the development of specific strategies to efficiently identify meaningful clinical groups. There were 2 stages of evidence synthesis, including the early (see Figure 3 & 4) and late stage. The process was conducted qualitatively. Appendix 4 provides an example. For both CACS and CTCA, the measures that indicate association were organised according to the clinical and their respective classifications. Specifically, CACS was organised according to widely agreed categorical cut-points (100 and 400) and continuously as logarithmic scale; CTCA was organised according to widely accepted categorical cut-points: 1-vessel, 2-vessls and 3-vessels diseases. There was one main difference between the early and late stages of evidence synthesis. In the late stage, the qualitative matching process was relaxed. The early strategy of evidence synthesis was robust but led to too many subgroups in CACS (see Figure 5) due to the self-imposed stringent qualitative matching process. The presence of multiple subgroups led to the limited conclusion and therefore the strategy was revised. Both all-cause mortality and composite cardiac outcomes

were treated the same. For CTCA, the concerning size effect was considered to be 1 vessel disease when it is not specified. The matching of symptoms and baseline risk was also relaxed. The difference between symptomatic and asymptomatic participants was explored further within in the sensitivity analysis.

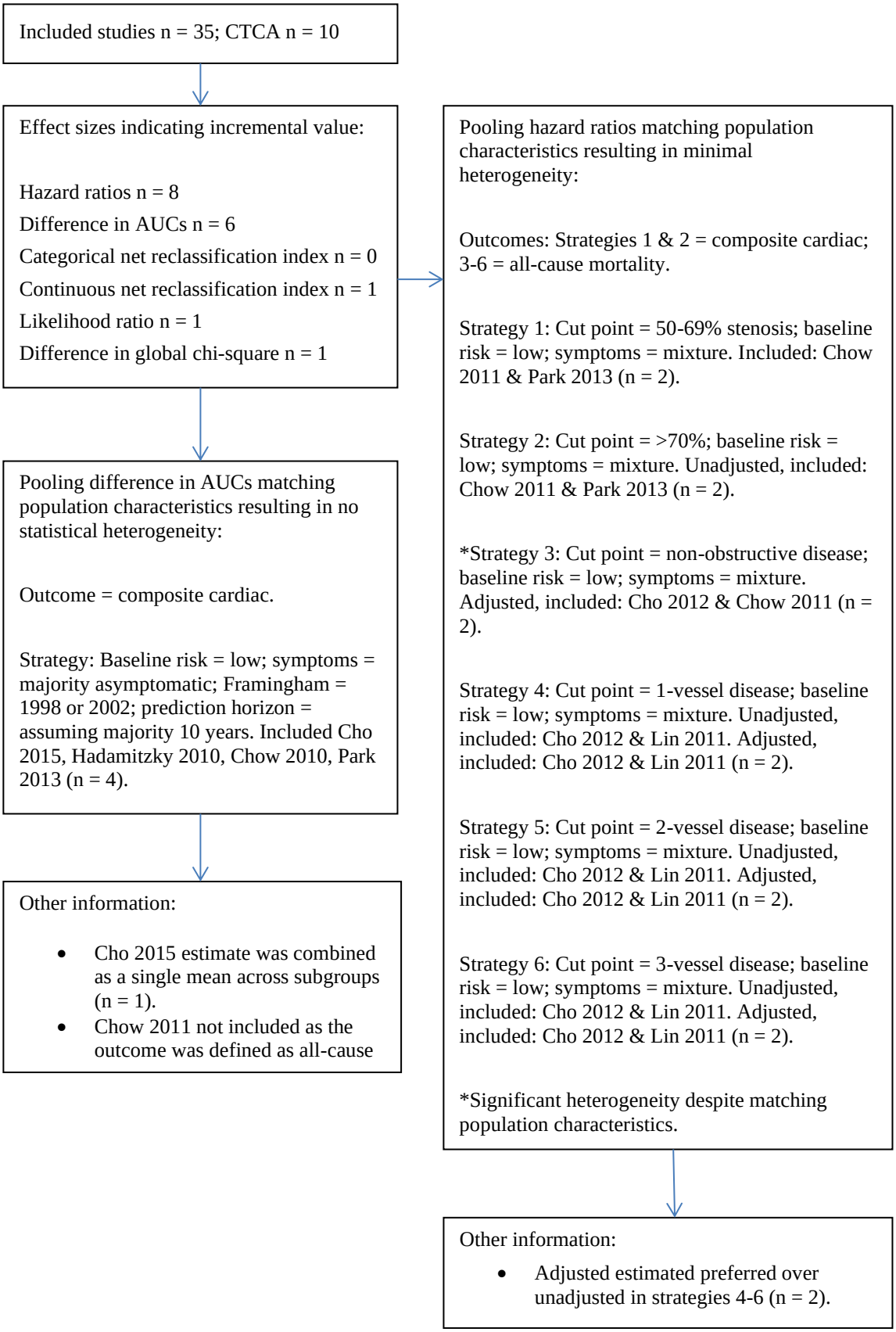

Figure 3. Early strategy of evidence synthesis - CTCA

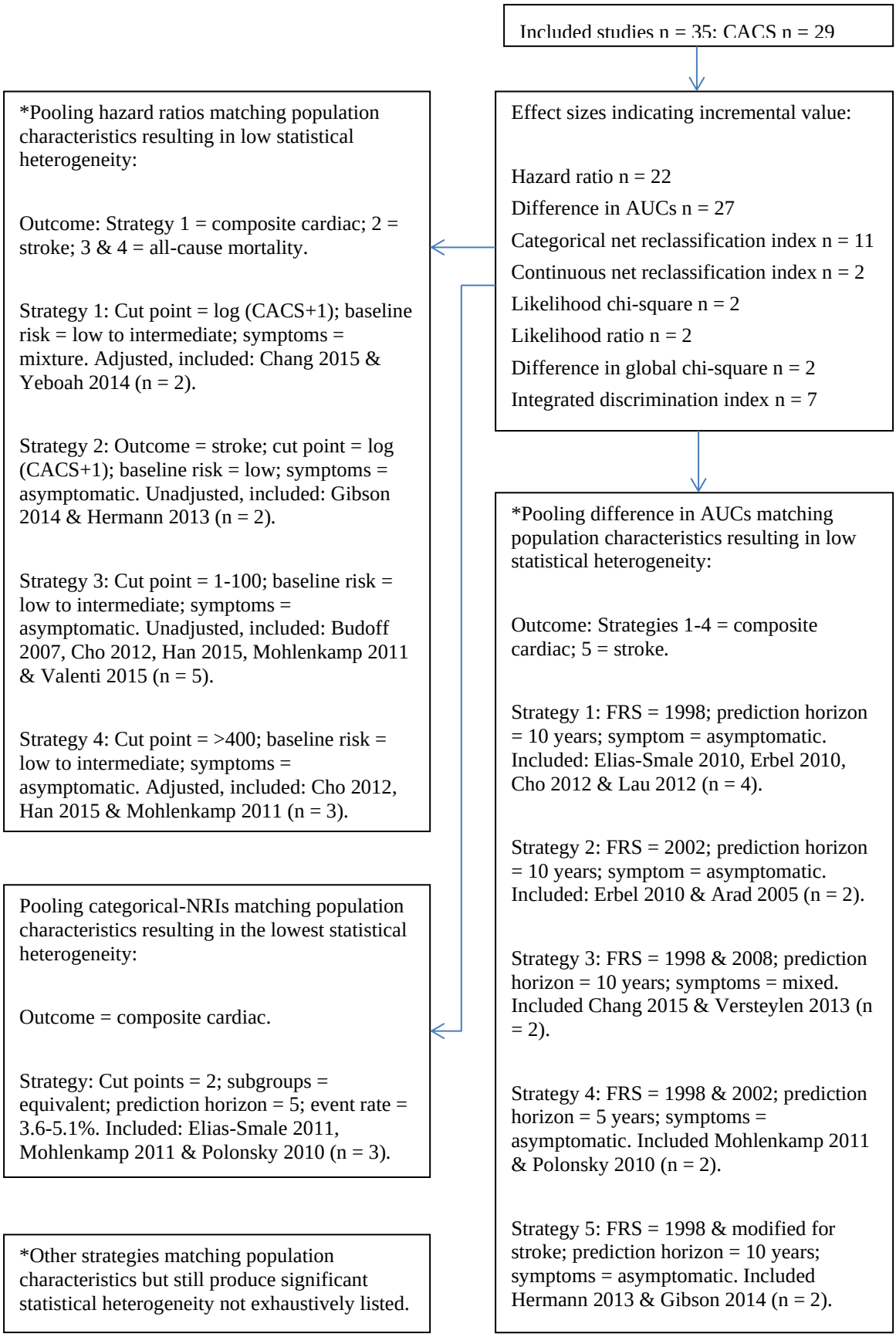

Figure 4. Early strategy of evidence synthesis - CACS

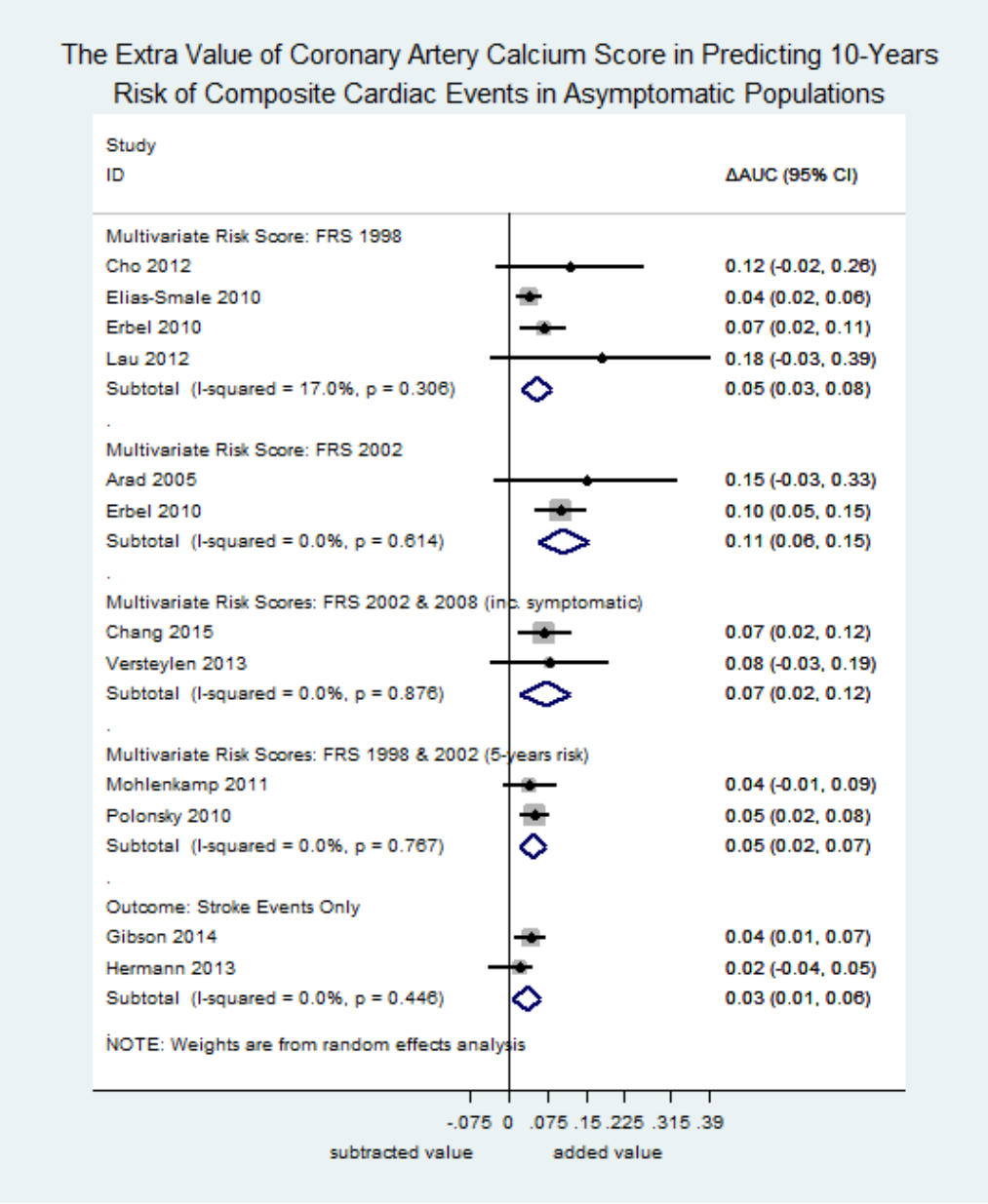

Figure 5. Initial evidence synthesis for CACS

Section 2: Meta-analysis & discussion

4.9 Thoracic calcium score: Association

Three studies described limited information about TACS (21, 112, 113). TACS was associated with both coronary and cardiovascular events (unadjusted HR 1.18, 95% CI 1.13-1.23, I^2 = 0%, n = 2; unadjusted HR 1.17, 95% CI 1.10 to 1.24, I^2 = 31.9%, n = 2) (21, 113).

4.10 Coronary calcium score: Association, discrimination & reclassification

Ten studies described the association of CACS with different outcomes (18, 21, 106, 108, 113, 118, 124, 126, 127, 133) (see Figure 6). One study had symptoms at baseline (108). For the outcome all-cause mortality, CACS predicted events in CACS subgroups 1-100 (adjusted HR 2.09, 95% CI 1.81-2.36, I^2 = 0%, n = 5) (18, 106, 118, 124, 133) and >400 (unadjusted HR 2.48, 95% CI 1.56-3.39, I^2 = 4.6%, n = 3) (118, 124, 133). CACS also predicted composite cardiac events (adjusted HR 1.29, 95% CI 1.24-1.35, I^2 = 0%, n = 2) and stroke events (adjusted HR 1.14, 95% CI 1.07-1.20, I^2 = 0%, n = 2).

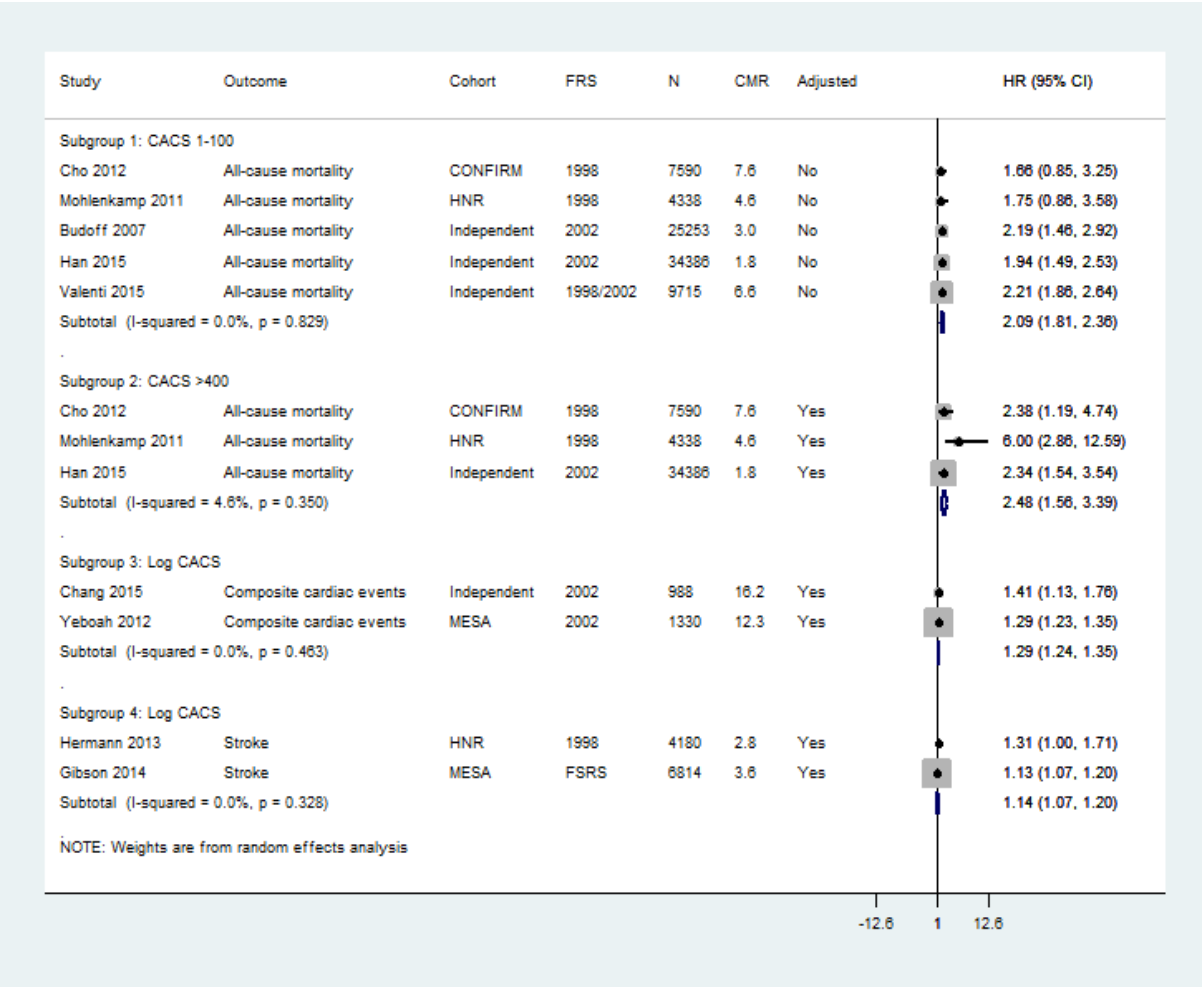

Figure 6. Association between cardiovascular events & CACS

Fifteen studies reported that CACS added discrimination in addition to the Framingham Risk Score (7, 102, 105, 106, 108, 109, 117-119, 123-127, 134) (see Figure 7). Two studies had symptoms at baseline (108, 125). Baseline AUC ranges from 0.59-0.72 across all FRS iterations. Using FRS 1998 as a baseline risk model, CACS added discrimination (unadjusted Δ AUC = 0.05, 95% CI 0.03-0.06, I^2 = 5.7%, n = 5). Similarly, adopting FRS 2002 as a baseline risk model, CACS also added discrimination (unadjusted Δ AUC = 0.09, 95% CI 0.06-0.13, I^2 = 63.1%, n = 8). Sensitivity analysis showed a difference in added discrimination between symptomatic (unadjusted Δ AUC = 0.07, 95% CI 0.09-0.12, I^2 = 0%, n = 2) and asymptomatic (unadjusted Δ AUC = 0.11, 95% CI 0.06-0.16, I^2 = 73.6%, n = 6)

participants. In subgroup analysis, a shorter prediction horizon of 5 years (rather than 10 years) (unadjusted Δ AUC = 0.05, 95% CI 0.02-0.07, I^2 = 0%, n = 2) and the use of the outcome cerebrovascular disease (unadjusted Δ AUC = 0.03, 95% CI 0.01-0.06, I^2 = 0%, n = 2) also demonstrated added discrimination.

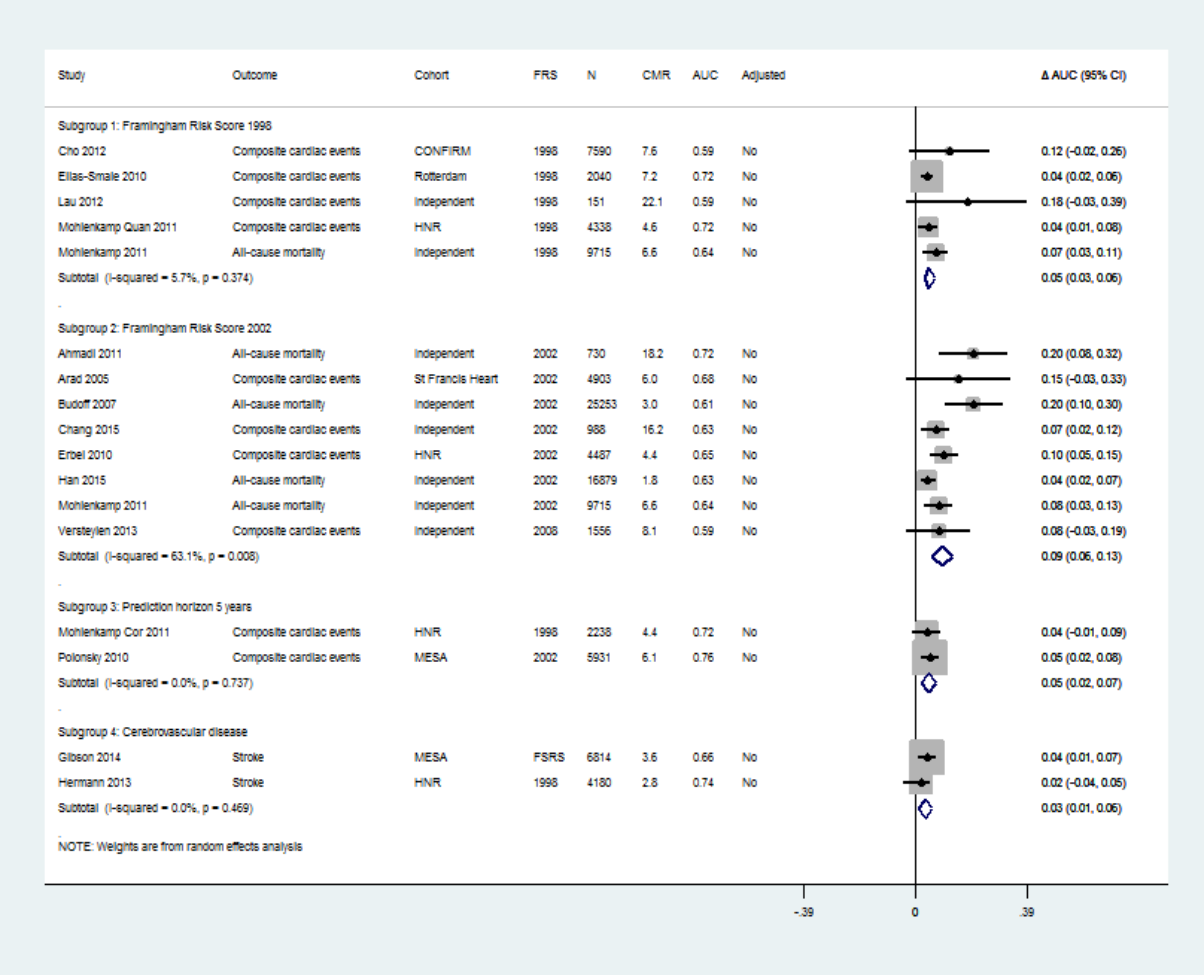

Figure 7. Incremental discrimination of CACS in addition to FRS

Eleven of 31 studies reported NRIs but none shared the same baseline characteristics or outcomes and pooling was not done. However, 9 of the 29 studies provided detailed reclassification tables (16, 112, 117, 119, 121, 122, 124, 128, 135) that enabled calculation of NRIs. Three of those 9 studies shared similar characteristics which enabled pooling of category-based event, non-event and combined NRI (112, 117, 119) (see Figure 8).

Statistical pooling showed that CACS had a differing impact on different components of NRI within the overall risk category (category -based event NRI = 0.22, 95% CI 0.15-0.28, I^2 = 0%, n = 3; non-event NRI = 0, 95% CI -0.03-0.03, I^2 = 89%, n = 3; combined NRI = 0.22, 95% CI 0.16-0.28, I^2 = 0%) (112, 117, 119). The comparison of unadjusted and adjusted NRIs (using Kaplan-Meier event rate) was made (136). NRIs in 1 study showed a difference ranging from 0.022 to 0.035 (119).

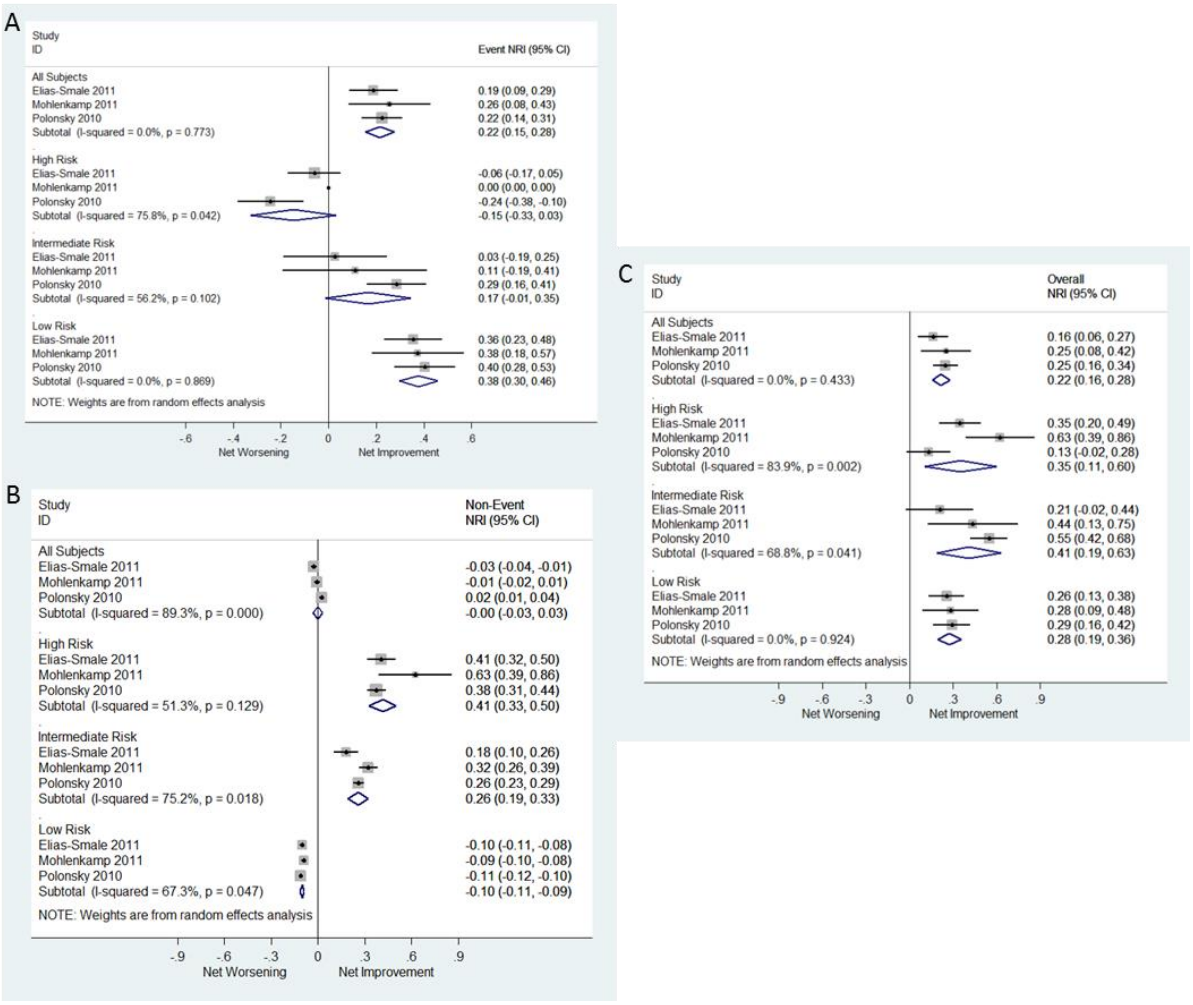

Figure 8. The category-based NRI (A) Event NRI. (B) Non-event NRI. (C) Combined NRI.

Incremental values displayed as IDI or rIDI (7, 108, 116-119, 121, 127) and continuous NRI (127, 132, 135) did not enable statistical pooling.

4.11 Computed tomographic coronary angiogram: Association & discrimination

During a mean follow-up of 2.9 years (range 1.3-4.0 years, 124,680 patient years) there were 1238 composite cardiac events reported. The overall pooled incident composite cardiac event rate was 1.41 per 1000 patient years. Most studies reported the overall events but not according to a patient with no coronary disease or obstructive coronary disease. Pooling of crude mortality ratios was not done due to heterogeneity. CTCA demonstrated incremental prognostic value in addition to FRS in terms of association and discrimination but no relevant measures of overall performance, reclassification or clinical usefulness were seen.

Obstructive coronary disease (luminal stenosis > 50%) detected on CTCA and the number of vessels with obstructive disease was associated with cardiac events and/ or cardiovascular comorbidities as indicated by time-to-event measures (Figure 9). There was evidence that individuals with 1-vessel obstructive disease compared to no coronary artery disease had a higher hazard for cardiac events or cardiovascular comorbidities (HR 1.75, 95% CI 1.42 to 2.07, I^2=0%, n=4). Two of the included studies did not specify whether the summary estimates were unadjusted or adjusted (130, 131). Sensitivity analyses showed that studies (124, 129) that explicitly adjusted for FRS risk factors resulted in a lower hazard ratio for cardiac events or cardiovascular comorbidities (HR 1.42, 95% CI 0.73 to 2.12, I^2=0%, n=2), compared to studies (130, 131) that did not explicitly state adjustment (HR 2.01, 95% CI 0.76 to 3.26, I^2=15%, n=2). Individuals with 2-vessel obstructive disease compared to no coronary artery disease had a higher hazard for cardiac events and/ or cardiovascular comorbidities (HR 2.10, 95% CI 0.98 to 3.21, I^2=0%, n=2). Individuals with 3-vessel

obstructive disease compared to no coronary artery disease had a higher hazard for cardiac events or cardiovascular comorbidities (HR 3.32 95% CI 1.55 to 5.08, I^2=0%, n=3). In the 3-vessel disease category, sensitivity analysis showed that studies that explicitly adjusted for FRS risk factors led to small changes (HR 3.22, 95% CI 1.44 to 5.01, I^2=0%, n=2). Overall, 3-vessel obstructive coronary disease was the most hazardous compared to no coronary disease.

There was evidence for the added prognostic value of CTCA in predicting composite cardiac events (Δ AUC = 0.07, 95% CI 0.04 to 0.09, I^2 = 0%, n = 4) (Figure 10) (111, 130-132). Subgroup analysis including only asymptomatic participants still showed incremental value (Δ AUC = 0.06, 95% CI 0.04 to 0.08, I^2 = 0%, n = 3) (111, 131, 132). CTCA showed added prognostic value when both coronary calcium score and FRS were utilised (Δ AUC = 0.04, 95% CI 0.01 to 0.07, I^2 = 6.8%, n = 2) (111, 124). Two studies from the CONFIRM cohort recorded summary estimates apart from HR and Δ AUC.

A 2012 publication (124) recorded categorical NRI = 0.43, 95% CI 0.25-0.64 and likelihood ratio = 1128.35, p<0.001. A 2015 publication (132) recorded continuous NRI and likelihood ratios according to calcium score subgroups (results not shown here).

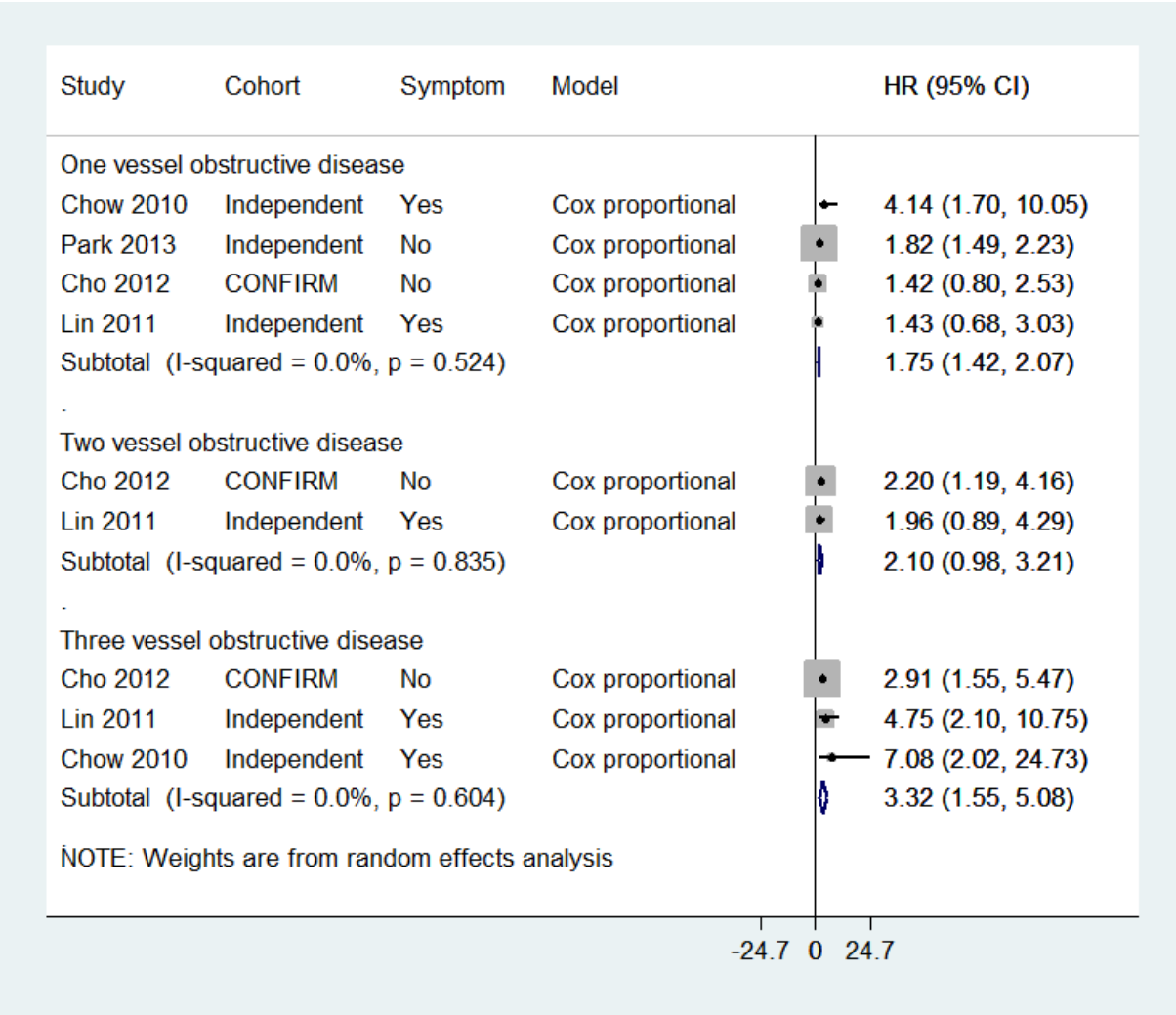

Figure 9. Association between obstructive coronary disease & cardiovascular disease

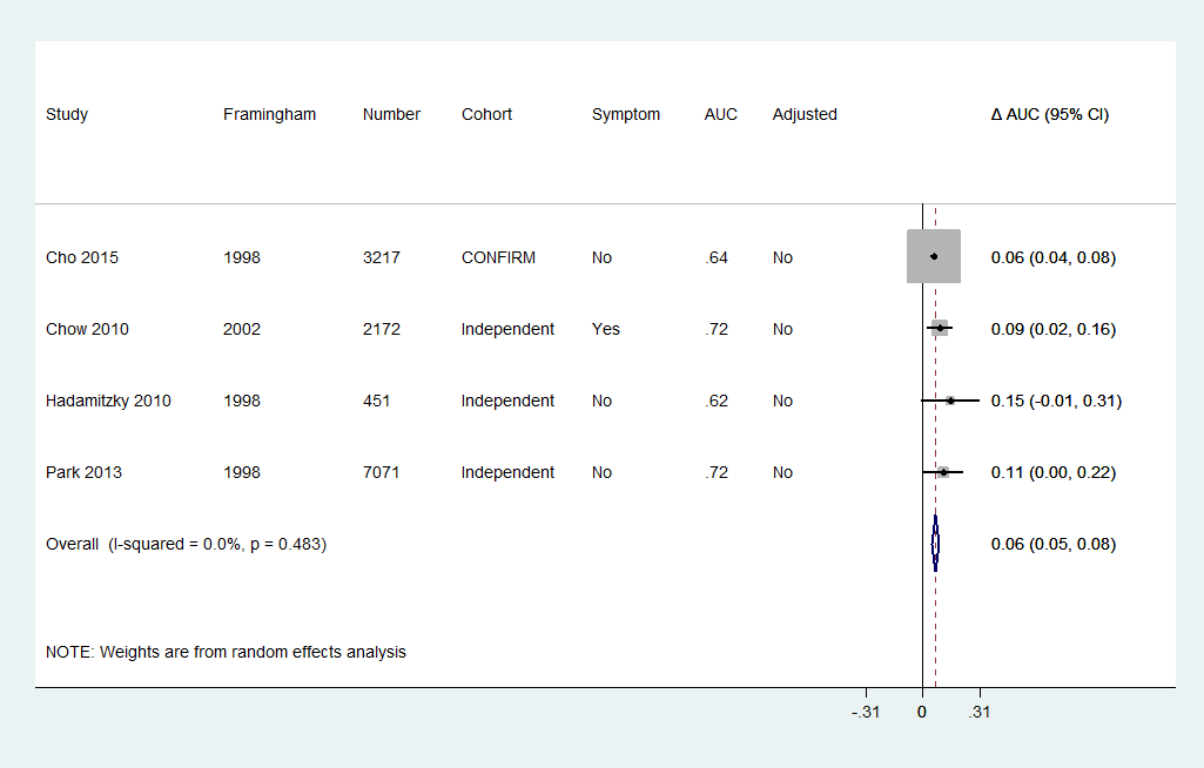

Figure 10. The added discrimination of computed tomographic coronary angiogram in predicting composite cardiac events in addition to Framingham Risk Score

4.12 Discussion

This review provides more information about the discriminatory abilities of adding CACS and CTCA to the well-documented Framingham multivariate model. However, the magnitude of the improved reclassification above FRS as a baseline model deserves further scrutiny. There is no available data on the incremental value of TACS. Regarding prediction, there was evidence that the higher the CACS, the higher the risk of CVD (27). The results showed an overlap between CACS categories. The number of vessels with obstructive coronary disease detected on CTCA positively correlated with the occurrence of endpoints. The survival data was independent of invasive coronary angiography.

Based on Δ AUCs, there was some evidence that the addition of CACS or CTCA modestly improved discrimination when compared to FRS alone. This small improvement should not be discounted; accepted risk factors, such as smoking, have only marginal impact on AUC but are capable of more accurate reclassification (38). Of the studies that demonstrated incremental value to FRS, there was often a lack of statistical significance but the general direction indicated added explanatory power (88, 89). The performance of baseline FRS had an impact on the amount of extra discrimination with the addition of CACS or CTCA (see Figure 10), which is known in the literature (31, 137). Overall, the magnitude of improvement in Δ AUC is currently difficult to relate to clinical practice. There was relatively more prognostic evidence here associated with CACS compared to CTCA, though the prime focus was not head-to-head comparison. For the purpose of CVD risk stratification in asymptomatic patients, there is an argument to initially risk stratify based solely on CACS and not perform CTCA. CACS has become routine clinical practice when patients are investigated for chest pain. This was not available at the time when these studies were conducted. With the widespread routine documentation of FRS risk factors and the increasing availability of calcium score, using the calcium score as an adjunct is a realistic possibility. Based on the added value demonstrated, there may be a role for CACS in existing clinical pathways. Clinical scenarios are more straightforward when FRS and CACS correlate. When both are low, patients can be assumed to be low risk. When both are high, patients can be assumed to be high risk. While there is an overlap between FRS risk factors and CACS, discrepancy between the two can be problematic. Clinicians will probably be left wondering whether to leave the patient alone, attempt to serially monitor CACS, empirically treat or refer patients for onward investigation.

It was difficult to draw firm conclusions based on direct comparisons of category-based NRIs because they have been derived from different populations (138). All of the included studies used at least 2 cut-offs for cardiovascular risk estimation meaning a straightforward

correlation between category-based NRIs and Δ AUCs was not possible. Overall, the combined category-based NRIs of CACS suggested net improvement of reclassification. From our meta-analysis, the combined NRIs were predominantly driven by event NRIs. This illustrates the potential inflation of the ability of CACS to reclassify events whilst downplaying its inability to reclassify non-events. In addition, the impact of non-event NRIs was diluted as unweighted combined NRIs did not take prevalence into account. The results suggest that when CACS-led strategy is applied to a cohort of participants, it will correctly identify those who experience events. However, this comes at a cost of classifying individuals as higher risk who will not actually experience any event.

The available evidence needs to be cautiously interpreted as the methodology is complicated. Clinicians should be aware that reclassification measures depend on demographic factors. The pooling of the Δ AUC may be insensitive to detect (small) improvements in the performance of FRS when a new marker is added. This is because FRS already included important predictors. Careful consideration of costs and benefits, test trade-offs and calibrations are required to make decisions on meaningful changes in AUC (38, 139). Similarly, the key to interpreting NRI depends on reporting transparency. A detailed reclassification table would promote interpretation of the individual components of NRI. To fundamentally address these issues, the ideal study to determine whether CACS or CTCA and FRS together are better than FRS alone would be a randomised controlled trial. If a person is classified as high risk based on FRS or develops chest pain, it would not be ethical to not actively manage the person with onward treatment or investigation. It is also important to remember that FRS was developed within an asymptomatic population. Future studies would probably continue to be of an observational cohort design and likely conducted within an asymptomatic population. The effect of treatment during the inception of any future cohort should be considered, which is not currently the case. The current lack of reporting on calibration and validation of the models reflects previous studies (5). Focus

should be shifted towards external validation, rather than remaining at the stage of model development.

Chapter 5 - Use of association, discrimination and reclassification to improve prediction: An update using studies reporting calcium score and computed tomographic coronary angiogram in addition to Framingham Risk Score

5.1 Link with the systematic review

Amongst CACS, there was a variety of baseline AUC ranges (0.59-0.72) across all FRS iterations. Variation in patient demographics may explain this. The baseline performance impacts on the incremental value as that is the starting point from which any additional benefit is measured. To examine the variable baseline performance further, the included studies from Chapter 3 were scrutinised to look for any particular reporting practice that may bias the magnitude of incremental value measure. The results from Chapter 3 were analysed using a binary classification – adequate or inadequate reporting practice. The main findings are below.

5.2 Methods, data extraction & analysis

The methods (including the screening process and study selection) were the same as in the previously described systematic review. Please refer to Chapter 2 for details. The data extraction and analysis were different and have been outlined below.

The first author, journal, publication year, outcome assessed, population evaluated and their inferences on whether the additional predictor improves prediction beyond the FRS were recorded. Publications were classified whenever possible as defined by Framingham/ Wilson 1998, Framingham/ Adult Treatment Panel III (ATP) 2002 and Framingham/ D'Agnostino 2008 (91-93). Original 95% CI, SE, SD or P-values of summary estimates of interest were

extracted (94, 95). If the Kaplan-Meier survival curve was available, any missing hazard ratio was estimated (140). Specifically, the standard of reporting effect sizes that signalled incremental prognostic value was evaluated. The use of composite and the components of composite outcomes were examined due to their impact on cardiovascular trials (141). The outcomes were divided into groups of importance and it was noted as to whether CHD was assessed. The correlation between these groups and the effect sizes was explored. The choice of optimal cut points/ thresholds were also examined, particularly in relation to size effects that indicate association (142).

Various methods of quantifying the incremental prognostic value of an additional test have been described (44). This review focussed on the reporting characteristics of multivariable regression, calibration, discrimination and reclassification (31). For the documentation of multivariable regression, adequacy was determined based on the availability of information on whether an additional predictor was significant at $p<0.05$ level or the use of tests that penalise the inclusion of an additional predictor. For discrimination, the documentation of the baseline area under the receiver operating curve (AUC) of FRS and the difference in AUC (Δ AUC) as a result of an additional predictor of interest was assessed. The adequacy of reporting baseline AUC relies on accurate documentation of the FRS as originally published (31). In brief, the calculation of FRS could be threatened by the addition, deletion or modification of the original FRS items. Other aspects include whether CHD was measured and whether the measured population was similar to the original FRS population. For reclassification, all publications were searched for NRI calculation or results. The type of NRI was verified. Established categories (e.g. <10%, 10-20%, >20%), or any justified use of categories as appropriate to relevant data set, were considered. The recommendation for reporting reclassification was taken from (138).

For studies where the 95% CI or standard errors were not reported, a correlation coefficient of 0.3 between FRS and CACS/ CTCA was used to allow estimation of the 95% CI for Δ AUC based on data from (102). Numbers were displayed as exact numbers, median or percentages. The alteration of the risk factors used to calculate FRS was assessed based on previously published items (31) with modifications. The items were scored ordinally as either yes, no or unclear. FRS model of the 1998 iteration was scored against 18 items. FRS model of the 2002 and 2008 iterations were scored against 15 items (3 diabetes related items discounted). The summation of the individual item score indicates the overall level of alteration which was dichotomised into a binary variable: minor and major alterations. The threshold for dichotomisation was based on the median number of items altered among the included studies. The summary of NRIs and AUCs were displayed as medians and interquartile ranges. NRIs and AUCs were subsequently split into two groups depending on the practice of reporting being either adequate or inadequate, or as equivalent binary groups. The aspects of reporting were based on previously published work on AUC (31) and NRI (138) with adaptations. The respective groups were then compared using the Wilcoxon sign rank test at significance level $p<0.05$. Specifically, we were looking for any particular practice of reporting AUC or NRI that lead to excessive claims of an additional predictor. All statistical analysis was carried out using STATA version 14.0 (StataCorp, College Station, Texas).

5.3 Results

Included studies

The reviewers (kappa 0.593) had moderate agreement at the screening stage and good agreement (kappa 0.818) at the full-text review stage. Eight-hundred and one unique hits were screened, leading to 35 studies encompassing 206,663 patients (men = 118,114, 55.1%) (7, 16, 18, 21, 22, 102, 105-133). All publications concluded that at least 1 imaging biomarker indicated either independent association with composite endpoints, improved discrimination or classification beyond traditional risk factors. However, there were reservations about TACS (21, 113) and some argued against the reclassification properties of CACS and CTCA (117, 124).

The types and calculation of Framingham Risk Score

Eleven studies (31.4%) adopted FRS 1998 (16, 102, 109, 111, 113, 117, 118, 124, 127, 131, 132), 13 studies (37.1%) adopted FRS 2002 (21, 105-108, 112, 114, 115, 119, 123, 128, 130, 133) and 3 studies adopted FRS 2008 (116, 121, 125). Three studies used both FRS 1998 and 2002 (7, 18, 22). Some studies made fundamental changes to the baseline FRS model, including the addition of body mass index (110) or left ventricular ejection fraction (128), adaptation for stroke (126), unspecified modification (129) and calibration for a different ethnic group (102). Two studies did not specify the iteration of FRS used (120, 122). According to previously published criteria (31) additions, deletions and modifications of risk factors are shown in Table 7. Six studies (17.1%) did not provide any mean estimate or a breakdown of different categories of FRS (106, 116, 120, 123, 132, 143). Twenty studies (57.1%) used categorical FRS estimates (7, 18, 22, 105, 109, 111-114, 117-119, 121, 122, 124, 128, 130-133, 144); eight studies (22.9%) used continuous FRS mean estimates (21, 102, 110, 115, 125-127, 129); and one study had documented both categorical and

continuous FRS estimates (108). Four studies excluded participants from the calculation based on certain FRS components (102, 115, 123, 134).

	No. of Studies (n = 35)		
	Ordinal outcomes		
Items of Alteration (n = 18)	Yes	No	Unclear
Addition			
Item 1. Antihypertensive	16	15	4
Item 2. Weight related measures, e.g. BMI	0	35	0
Item 3. Race/ ethnic groups	2	33	0
Item 4. Triglycerides	2	33	0
Item 5. Alcohol	0	35	0
Item 6. Previous cardiovascular disease	1	34	0
Item 7. Others (family history, PVD & stroke)	5	30	0
Deletion			
Item 8. Diastolic blood pressure	9	21	5
Item 9. ^Diabetes	0	19	0
Item 10. HDL cholesterol	4	28	3
Modification			
Blood pressure			
Item 11. Systolic blood pressure	3	23	9
Item 12. History of hypertension/ self-reported hypertension	7	23	5
Item 13. Other blood pressure definition modification	3	31	1
Lipid levels			
Item 14. History of hyperlipidaemia/ self-reported hyperlipidaemia	14	14	7
Diabetes			
Item 15. ^Fasting glucose >126 mg/dL or 7.8 mmol/L	0	19	0
Item 16. ^Self-reported diabetes/ use diabetic medication	3	10	6
Smoking			
Item 17. Pack years of smoking	0	34	1
Item 18. Use of ex-smoker category	11	23	1
Abbreviations: BMI = body mass index; HDL = high density lipoprotein; PVD = peripheral vascular disease			
^13 studies used FRS 2002 and 3 studies used FRS 2008 and diabetes related items were discounted.			

Table 7. Alteration of the risk factors used for the calculation of Framingham Risk Score in 35 eligible studies compared to the Framingham Risk Score 1998, 2002 and 2008

The median number of items altered was 3. Using that as the threshold, twelve studies (34.3%) had major alterations (7, 18, 102, 106-109, 111, 116, 124, 132, 133) and 23 studies (65.7%) had minor alterations (16, 21, 22, 105, 110, 112-115, 117-123, 125-131). Five of 23 studies that had minor alterations did not have any components of FRS altered (114, 120, 121, 125, 131). Of those 5 studies, two studies did not provide any information about the components of FRS (121, 125) and were given the benefit of the doubt, however, findings should be interpreted with caution. Sensitivity analysis was conducted using more stringent criteria (if "unclear" was classed as alteration). This increased the median number of items altered to 4.

Outcomes, thresholds and reporting of association

Nine studies (25.7%) did not examine CHD as an outcome (18, 22, 106, 107, 122, 126, 127, 129, 133). All studies used composite outcomes and the most frequently used outcome was all-cause mortality. In total, there were 61 composite outcomes with 115 components classified according to their importance (141). Forty eight were classified as death (41.7%), five were critical (4.3%), thirty nine were major (33.9%), thirteen were moderate (11.3%) and ten were minor (8.7%). Analysis using a meta-analytical approach or comparison of medians was not suitable for demonstrating trends.

Odds ratio, relative risk, c-index and hazard ratio were used to indicate association of the imaging biomarker with outcomes. There was selective reporting of subgroups and p-values among the reported subgroups (Table 8). Chapter 4 displays the details regarding the thresholds for different types of investigation. There were 29 studies that recorded CACS as the investigation of choice in addition to FRS. Various different categories and cut points were used for different effect sizes, including hazard (n=23) and odds ratios (n=3) and relative risk (n=3). Some were comparable with other studies whilst others were not.

	Reference groups, n	Subgroups, n	Reported subgroups, n (%)	Missing p-value, n (%)
All effect sizes	92	381	328 (86)	85 (26)
OR	4	27	27 (100)	9(33)
RR	8	37	37(100)	18(49)
C-index	0	22	22(100)	0(0)
HR	80	295	242(82)	58(24)
Abbreviations: OR = odds ratio; RR = relative risk; C-index = index of concordance; HR = hazard ratio				

Table 8. Selective reporting of association

Appendix 5 shows all the different thresholds and subgroups of CACS, TACS and CTCA.

Comparison of different categories of CACS was not possible in some instances when the categories were fundamentally different, for example continuous data could not compared with categorical. For categorical data, CACS was most commonly subdivided into groups including 0, 1-99, 100-399 and >400. The cut-points often varied by a CACS of 1 amongst publications. Other subgroups included >0, 400-699, 700-999, >100. A CACS of zero was often used as the reference group but other reference groups were used: CACS <40; CACS <10. The reference group was not specified in some included studies. Transformation of CACS into a logarithmic scale was encountered. Analysis according to deciles, quartiles, baseline FRS risk categories and disease status was also encountered. Of the 29 studies that reported CACS, 10 studies did not report the pre-specified groups and 8 studies did not report p-values relevant to the size effects. Of the 10 studies that did not report the pre-specified groups, 3 provided a Kaplan-Meier curve that enabled estimation of hazard ratio (106, 108, 110).

Nine studies recorded CTCA in addition to FRS and 3 studies reported both CTCA and CACS (124, 131, 132). Of the 9 studies that reported CTCA, 2 only reported incremental values that were not association (125, 132). Various different categories and cut points were

used for different effect sizes, including hazard (n=5) and odds ratios (n=1), and c-index (n=2). The reference group was unclear in 3 studies. When the reference group was specified, no CAD or 0% stenosis was used as the reference group in 3 groups and non-obstructive diseases was used in 2 groups. The use of reference group affects the effect size of hazard ratio. One study did not report p-values relevant to hazard ratio. Two studies did not report pre-specified groups. Two studies reported all the pre-specified groups but extra information (hazard ratio) was extracted from the Kaplan Meier curve (111, 131).

Three studies recorded TACS in addition to FRS (21, 112, 113). None of them had overlapping groups. There was selective reporting however, there were too few studies to infer anything.

Intended population for Framingham Risk Score

Four studies (11.4%) had an exclusively Caucasian population (16, 21, 117, 118). Eleven studies (31.4%) had greater than 10% non-Caucasian population (106, 113-116, 119, 122, 123, 126, 131, 133). Twenty two studies (62.9%) did not record ethnicity as a variable (7, 18, 22, 105, 107-112, 117, 118, 120, 121, 123-125, 127-130, 132). Five studies (14.3%) had documented CHD at baseline (110, 122, 128, 130, 132). Considering all of the above information, only 4 studies (11.4%) were identified as similar to the original Framingham population (16, 21, 117, 118). Sensitivity analysis was performed, which considered all unreported ethnicity to have a less than 10% non-Caucasian population, increasing the number of studies to only twenty (57.1%) similar to the original Framingham population (7, 16, 18, 21, 22, 105, 107-109, 111, 112, 117, 118, 120, 121, 123-125, 127, 129).

Documentation of regression, discrimination & AUC analysis

Of the 35 studies, the majority appropriately reported multivariable regression (74.3%). Thirty-three studies reported AUC estimates for both FRS alone and the FRS with additional CT biomarkers with data on 76 such pairings. Appropriate documentation of AUC was not common practice (36.4%). The method used to compare ROC curves were not always described (39.4%). Only eight studies reported calibration (22.9%) (7, 113, 116-120, 133). Table 9 shows the reporting of regression and discrimination. The AUC of FRS alone ranged from 0.53 to 0.77 (median = 0.68). The Δ AUC ranged from -0.07 to 0.24 (median = 0.06). There was strong inverse correlation between the Δ AUC and the baseline FRS AUC (Spearman correlation coefficient, -0.46, p<0.0001). When the baseline FRS AUC performed well, the Δ AUC was relatively lower with the addition of a CT biomarker (see Figure 11).

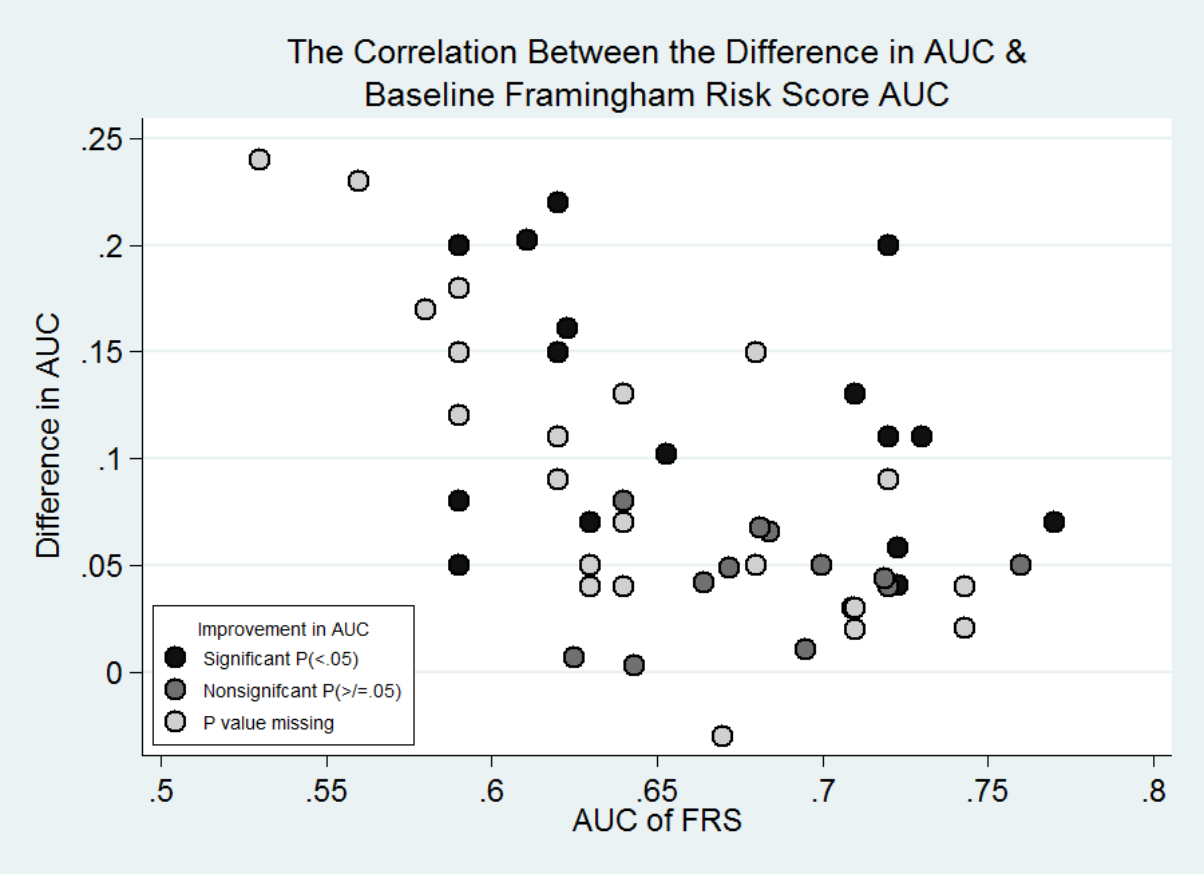

Figure 11. The correlation between the difference in AUC & baseline FRS AUC

Part 1. Documentation of multivariable regression (n = 35)	No.	(%)
a. Information on whether additional predictor is significant at <.05 level	24	68.6
b. Results of a test that penalises for the inclusion of additional predictor	8	22.9
Appropriate documentation (1a or 1b)	26	74.3
Part 2. Documentation of AUC in ROC analysis (n = 33)		
a. Described method used to compare ROC curves	13	39.4
b. Presented the AUC values with and without the additional predictor	31	93.9
c. Presented CIs of AUC values with and without additional predictor	9	27.3
d. Presented p-value for comparison	26	78.8
f. Availability or enable calculation of Δ AUC CIs	30	90.9
Appropriate documentation 1 (2a and 2b and [2c or 2d])	11	33.3
Appropriate documentation 2 (2a and 2b and [2c or [2d or 2f])	12	36.4
Part 3. Documentation of calibration (n = 35)		
Documentation of Hosmer-Lemeshaw test (n = 7) or Schoenfeld residuals (n = 1)	8	22.9
Part 4. Documentation of reclassification analysis (n = 35)		
Report using table or text		
Not reported	19	54.3
Partial	5	14.3
Complete	11	31.4
Standard of reporting of reclassification analysis (n = 16)		
a. Use of standard categories of risk	11	68.8
b. Justified use of other categories of risk	15	93.8
c. Reported the number of patients changing categories	9	56.3
Appropriate documentation ([4a or 4b] and 4c)	9	56.3
Inadequate	7	43.8
Part 5. Documentation of NRI (n = 23)		
Type of NRIs		
Continuous/ category-free NRI	4	17.4
Categorical NRI	16	69.6
Reported both continuous & categorical NRIs	1	4.3
Reported relative NRI	1	4.3
Unclear	3	13.0
Standard of reporting of categorical NRI (n = 16)		
a. Report censor handling	5	31.3
b. No extrapolation	7	43.8
c. Categorical NRI reference available	14	87.5
d. Justification of risk categories	14	87.5

e. Report NRI components	5	31.3
f. Availability of reclassification table showing event and non-event	8	50.0
g. Reclassification table enables the calculation of NRI components	7	43.8
h. Combined NRI reported as a sum not a percentage	8	50.0
i. The proportion of correctly reclassified subjects available	7	43.8
j. Reported NRI not used to construct strong summary	5	31.3
Adequate reporting of categorical NRI (>5 items listed 5a-j)^	11	68.8
^The threshold is the median number of items reported in a skewed sample.		

Table 9. Documentation of multivariable regression, calibration, discrimination & reclassification

Table 10 shows the median AUC values and the Δ AUC when the data was classified according to features of design and analysis (31). The baseline FRS AUC performs better with minor alterations compared with those with major alterations of the Framingham model (p = 0.0006). The improvement in AUC was greater in those with major alterations of the Framingham model (p = 0.015). Other factors that significantly affected the performance of AUC include the exploration of data analysis, reporting of calibration and validation, multivariable and AUC documentation (all p<0.05). The types of incremental value reported were associated with a difference in AUC performance, but only significant when a threshold of 2 was chosen. In the sample population, measurement of CHD as an outcome or whether the population was similar to the original Framingham cohort did not significantly alter the AUC performance. Although reclassification analysis did not significantly affect AUC performance, there was a difference between inadequate reporting and those that did not report (p <0.0443).

	No.	AUC FRS (median)	IQR	P-value	No.	AUC FRS + CT (median)	IQR	P-value	No.	Δ AUC (median)	IQR	P-value
1. Alteration of Framingham model												
Major	31	0.64	0.62-0.68		31	0.74	0.71-0.77		30	0.07	0.05-0.15	
Minor	42	0.7	0.64-0.74	0.0006	45	0.76	0.68-0.79	0.7271	46	0.05	0.02-0.09	0.015
2. Coronary heart disease measured												
Yes	58	0.68	0.62-0.72		61	0.75	0.71-0.78		61	0.06	0.04-0.11	
No	15	0.66	0.64-0.68	0.5208	15	0.72	0.68-0.75	0.2452	15	0.05	0.04-0.08	0.5393
3. Explore analysis model												
Yes	13	0.75	0.72-0.76		16	0.77	0.71-0.80		16	0.05	0.03-0.06	
No	60	0.65	0.62-0.71	<0.0001	60	0.74	0.71-0.78	0.4559	60	0.07	0.04-0.13	0.0274
4. Population as intended for Framingham												
Yes	45	0.68	0.64-0.72		48	0.75	0.71-0.77		48	0.06	0.04-0.11	
No	28	0.64	0.63-0.71	0.1841	28	0.74	0.68-0.78	0.5901	28	0.05	0.04-0.12	0.8546
5. Calibration reporting												
Yes	19	0.69	0.64-		19	0.74	0.67-		19	0.04	0.01-	

			0.72				0.77				0.06	
No	54	0.67	0.62-0.72	0.1427	57	0.75	0.71-0.78	0.1465	57	0.07	0.05-0.12	0.0007
6. Validation reporting												
Yes	22	0.65	0.62-0.70		22	0.74	0.71-0.76		22	0.08	0.05-0.13	
No	51	0.68	0.36-0.74	0.0433	54	0.76	0.68-0.78	0.7267	54	0.06	0.04-0.09	0.1231
7. Multivariable documentation												
Adequate	52	0.64	0.62-0.71		52	0.74	0.71-0.78		52	0.07	0.04-0.13	
Inadequate	21	0.72	0.68-0.75	0.003	24	0.76	0.69-0.78	0.6588	24	0.05	0.03-0.08	0.1002
8. AUC documentation												
Adequate	28	0.72	0.64-0.75		28	0.77	0.69-0.80		28	0.05	0.01-0.08	
Inadequate	45	0.66	0.62-0.70	0.0018	48	0.74	0.71-0.77	0.3431	48	0.07	0.05-0.13	0.016
9. Reclassification analysis documentation 1												
Adequate (reference)	14	0.69	0.62-0.72		17	0.76	0.74-0.78		17	0.06	0.05-0.11	
Inadequate or not reported	59	0.67	0.63-0.72	0.3924	59	0.74	0.70-0.74	0.2539	59	0.06	0.03-0.11	0.2032

Inadequate	17	0.64	0.63-0.67	0.095	17	0.73	0.70-0.76	0.1678	17	0.07	0.05-0.11	0.9035
Not reported	42	0.68	0.63-0.72	0.7189	42	0.74	0.71-0.78	0.3885	42	0.05	0.02-0.11	0.0772
10. Reclassification analysis documentation 2												
Inadequate	17	0.64	0.63-0.67		17	0.73	0.70-0.76		17	0.07	0.05-0.11	
Not reported	42	0.68	0.63-0.72	0.0443	42	0.74	0.71-0.78	0.6452	42	0.05	0.02-0.11	0.0877
11. Types of incremental value threshold												
>2	40	0.68	0.63-0.72		43	0.74	0.68-0.76		43	0.05	0.03-0.08	
<2	33	0.67	0.62-0.75	0.731	33	0.77	0.73-0.83	0.0013	33	0.08	0.05-0.15	0.0034
12. Types of incremental value threshold 2												
>3	12	0.72	0.69-0.74		12	0.76	0.74-0.78		12	0.05	0.04-0.07	
<3	61	0.66	0.62-0.71	0.0158	64	0.74	0.69-0.78	0.2884	64	0.06	0.04-0.11	0.192
*AUC = area under the operating curve; ΔAUC = difference in AUC; CT = CT biomarkers; FRS = Framingham model; IQR = interquartile range												
P-values generated using Wilcoxan ranksum test.												

Table 10. Median AUC values and ΔAUC according to different aspects of design and analysis

Documentation of reclassification & NRI analysis

Twenty three studies reported NRI estimates and all have at least 2 cut-offs, with those that had 3 cut-offs making up the biggest NRIs (124, 132). The number of thresholds influenced the value of NRI (145). The most commonly used type of NRI was categorical NRI (69.6%). The studies that reported calibration were the same as those documented in the last section. Table 11 shows the reporting of reclassification analysis and NRI. Complete reporting of reclassification analysis using a reclassification table or text was not common practice (31.4%). When reclassification analysis was done, only half was considered appropriate (56.3%). The actual number of patients being up or down classified to a different risk group was not documented. In conjunction with the documentation of NRI, the proportion of subjects being correctly reclassified was not always available (43.8%). Most studies subjectively drew strong conclusions from the NRI calculated (68.7%). The individual components of events and non-events and also their respective NRI components were not always available (at least 43.8%). Fifteen studies reported categorical NRI with 46 data points. The values of categorical combined NRIs ranged from -0.083 to 0.785 (median, 0.249). None of the aspects of reporting (138) was significantly related to the difference in the values of categorical combined NRIs (Table 5).

	No.	NRI (median)	IQR	P-value
1. Reporting of censor handling				
Yes	13	0.18	0.14-0.43	
No	33	0.26	0.19-0.35	0.4869
2. No extrapolation				
Yes	20	0.28	0.14-0.49	
No	26	0.23	0.16-0.34	0.4186
3. Category NRI reference*				
Quoted	38	0.25	0.18-0.39	
Not quoted	8	0.18	0.00-0.38	0.1922
4. Justification of NRI categories*				
Yes	40	0.25	0.15-0.41	
No	6	0.21	0.18-0.29	0.7691
5. Reporting NRI components				
Yes	22	0.24	0.13-0.34	
No	24	0.28	0.19-0.48	0.1907
6. Reclassification table showing the number of events and non-events				
Yes	32	0.27	0.18-0.37	
No	14	0.21	0.14-0.43	0.3396
7. The availability of reclassification table with sufficient information to enable the calculation of event and non-event NRI				
Yes	30	0.25	0.18-0.35	
No	16	0.23	0.14-	0.9265

			0.48	
8. Describing the combined NRI as sum not a percentage				
Yes	25	0.28	0.14-0.47	
No	21	0.22	0.18-0.30	0.256
9. Provide any indication of the proportion of correctly reclassified				
Yes	15	0.35	0.14-0.53	
No	31	0.23	0.16-0.34	0.0916
10. Strong conclusion based on the reporting of NRI				
No	20	0.29	0.19-0.48	No
Yes	26	0.23	0.13-0.34	0.08
11. NRI adequate documented				
Adequate	39	0.25	0.14-0.39	
Inadequate	7	0.24	0.20-0.49	0.7949
12. Types of incremental value				
>3	18	0.28	0.2-0.44	
<3	28	0.23	0.13-0.37	0.2464
Abbreviations: IQR = interquartile range; NRI = net reclassification index				
P-values generated using Wilcoxon ranksum test.				
*One group has less then 10 studies and therefore the comparison is limited.				

Table 11. Median NRI values according to different aspects of design and analysis

5.4 Discussion

The majority of studies claimed improved discrimination and reclassification of the outlined CT biomarkers over the established Framingham model. For association, hazard ratio was the most commonly used size effect but the variation in reporting practice hindered evidence synthesis. Although all studies used similar baseline model for AUC analysis, the performance of FRS varied. There was a clear negative correlation between improved discrimination and baseline performance of FRS. In contrast, despite the poor reporting, there was no difference in the magnitude of categorical NRI between adequate and inadequate reporting practice.

Selective reporting of association is well known and remains an issue here (146). We found that non-standardisation of thresholds and reference groups across studies prohibited future meta-analysis. The following is an example to substantiate this claim; Chow et al used non-obstructive coronary disease as the reference group for 3 different composite outcomes (130). The same author then used no coronary artery disease as the reference group in the CONFIRM cohort (128). Does this matter? In (131), we estimated hazard ratio using Kaplan-Meier curve (140). When non-obstructive coronary disease was the reference group, the estimated association of obstructive disease with composite outcome was smaller (HR = 2.03, 95% CI 1.47 – 2.79), compared to when no coronary disease/ normal was the reference, the estimated association was bigger (HR = 3.24, 95% CI 2.28 – 4.63) (131). This adds to the known issues on publication bias (147).

NRI records the change in a person's predicted risk from one category to another category after the introduction of an additional test (148). However, it was only meaningful when information about risk thresholds was available. The change in predicted risk could be correct or incorrect. In this population, the concept that the subject could be wrongly

reclassified with an additional test was not clearly outlined. The combined NRI could have been driven by predominantly event NRI, leading to overestimation. This could, however, have been clarified by reporting the components of NRI but this was not standard practice despite its recommendation (149). Another recommendation was the regular reporting of calibration derived from concern about miscalibration (39, 150, 151), with graphical plot being the best assessment (152). Calibration, however, was regularly overlooked. To counteract the issues with missing data, we have used solutions such as Weibull extrapolation (16, 109, 112) and adjustment of risk cut-offs by the ratio of actual follow up. These strategies translate to the fact that a significant proportion of the included studies used non-standardised risk categories but almost all managed to justify. A more definitive solution would be a move towards decision curve analysis (153). Only 1 study provided information to allow adjustment using Kaplan-Meier estimates (119). This is on a background of insufficient reporting on the handling of censoring. This adjustment should receive more attention especially when censoring happened early on during follow-up (136). There is currently no consensus on what is a large enough NRI. Overall, considering the uncertainty in NRI (149) and the small values of NRI, strong conclusions should not be drawn from the use of NRI alone. Given the popularity of NRI in cardiovascular research (39), a framework of reporting NRI should be followed, for example in (138).

This investigation indicates that the previously raised issues about establishing incremental value of any predictor beyond the Framingham model persists 10 years on (31). Discrimination, as measured by AUC analysis, is an established method of measuring incremental value (154). It was reported in almost all the studies but adequate documentation was not common practice. Reporting of calibration, validation and AUC documentation all influenced the values of AUCs and inadequate reporting practices were associated with inflated estimates. It is evident from the linguistic spin from the included articles that there is ongoing hype regarding the utilisation of additional CT biomarkers in

CVD risk prediction. However, big improvement in AUC was predominantly seen in cases where Framingham models performed badly. Questions have been raised as to whether this perceived poor performance is due to illegitimate alterations of this extensively tested Framingham models. As eluded to in (31), this phenomenon is similar to when a new drug is only effective when compared to an ineffective comparator drug (155).

There are other complex issues that need to be considered here. There is an assumption that any positive incremental value effect estimate equates to clinical benefit and perhaps different effect estimates relate to one another. Neither a small increase in AUC after addition of a predictive marker, nor an effect estimate that indicates very strong association (e.g. HR >10) between a predictive marker and an outcome, equate to definitive incremental value (139). In fact, it is not possible to decide which one is more informative. A gain in discrimination or a strong association does not meaningfully translate to better clinical performance. These uncertainties can potentially be solved by a decision analytic approach (139). Even when there is a single well-designed cohort study, there are issues regarding compatibility of the design with existing literature, such as incompatible cut offs in risk categories and prediction horizon (156). Finally, none of these perceived improvements in AUCs has been formally evaluated in a systematic review or meta-analysis. Overall, one should be cautious regarding any claim of importance of a new predictor as risk assessment with traditional CVD risk factors, as outlined in the Framingham model, works well despite its imperfections. It is simple, cheap and easily understood by both patients and clinicians, and it leads to logical treatment strategies (157).

Limitations

Our investigations on AUC (31) and NRI (149) were not empirical and can only serve as an update in a different population. The assessment on thresholds was minimal compared with

previous investigations (149). The harm of imaging using CT (radiation burden) was not explored; the focus was solely on the potential benefit. Studies that indicated association or only had a reclassification table were excluded because we focused on studies that had at least 1 summary estimate that indicated incremental value. The impact of using multiple incremental value measures was not been fully investigated. We were unable to assess publication bias as articles that showed no or worsening predictions were not published.

5.5 Summary

Association on its own is insufficient to substantiate incremental value (37) and large values are infrequent in biomarker research (34). AUC analysis is seen as a good starting point and reclassification should follow rather than replace AUC analysis (154). However, AUC analysis is not without its flaws (34). Transparent reporting of NRI should be compulsory, for example by using a reclassification table (138), and the reader should be aware of the controversies surrounding NRI (42, 43). The co-existence of a lack of increase in AUC and a positive NRI should alarm readers (158). In general, the reporting in prognosis studies needs to be more robust (159-162). Data should be made available to allow individual patient data meta-analysis.

Inconsistent thresholds, reference groups and selective reporting prohibit future evidence synthesis of association. Inadequate documentation of discrimination, calibration and validation were widespread. The variable baseline performance and other aspects of reporting discrimination inflate potential incremental values. Reporting of reclassification is also insufficient but a significant difference between adequate and inadequate reporting practice was not identified.

Chapter 6 Feasibility study

6.1 Abstract

Objective: To evaluate the image quality of thorax computed tomographic (CT) imaging by using prospectively electrocardiogram (ECG) gated technique.

Methods: This prospective study included 80 patients who were having follow-up. Patients received 2 successive CTs which were assessed for artefacts on a 4-point scale by two blinded radiologists. The image quality of the coronary arteries was also assessed using a 15-segment model.

Results: ECG-gating resulted in an improvement in subjective image quality (ungated versus gated mean scores, 4.23 versus 4.93 respectively, $p < 0.001$) with reduced motion artefact, especially of the heart borders (ungated vs gated mean scores, 3.31 vs 3.94 respectively, $p < 0.001$). Step artefact did not affect the diagnostic acceptability (3.79, SD 0.5). Diagnostic images of the coronary arteries were obtained in 84.5% of patients. The degree of diagnostic segments was heart rate dependent (<60 bpm = 100%; >60-<70 bpm = 87.9%, 95% CI 83.6-91.4%; >70-<80 bpm = 86.3%, 95% CI 82.0-90.0%; >80-<90 bpm = 73.8%, 95% CI 67.0%-79.9%, >90 bpm = 68.2%, 95% CI 58.6-76.7%).

Conclusion:

Prospectively ECG-gated CT thorax provides excellent image quality of the lungs with the addition of coronary evaluation. The heart rate threshold for a diagnostic scan is 85 bpm without beta blockers.

6.2 Background & introduction

At Plymouth Hospitals NHS Trust, there is a cardiac enabled research scanner which is optimised for cardiac CT but is also used to perform routine CT scanning, predominantly for oncology. Taking into account the locality of the work, the findings of the systematic review and the risk of exposure to ionising radiation, a primary study was carried out to look at the technical feasibility of obtaining information about the coronary arteries during a routine follow-up CT scan for cancer staging.

CT scanning is central to the diagnosis and management of cancer and scanning of the thorax is standard practice for most cancers. Even though the heart is inevitably scanned as part of the staging process, information about it is limited due to movement blur. However, it is known that elderly patients who have had a newly diagnosed cancer have a high prevalence of coronary disease (163). In addition, men with testicular cancer have an elevated risk of cardiac events among long-term survivors (164) and breast cancer patients have a high prevalence of cardiac disease if they receive radiotherapy (165). There are also cancer drugs that have been associated with myocardial ischaemia and thromboembolic events, including thalidomide, docetaxel, fluorouracil, paclitaxel and cisplatin (166). From a wider perspective, there seems to be a temporal relationship between the development of cancer and coronary events. There is a significant increase in coronary events in the 2 years before a cancer diagnosis (167) and it is known that most cancers are associated with an increased risk of coronary disease during the first six months after diagnosis (168). Coronary disease and cancer share common risk factors, such as smoking, and there is a moderately increased risk of tobacco-related cancers among survivors of myocardial infarction (169).

Computed tomographic coronary angiography (CTCA) is a variation in CT scanning where information regarding the heart rate is obtained via ECG monitoring and scanning of the

patient is performed at times of relative cardiostasis (170). Major technical advances have been made in the technique and it is now the investigation of choice in all patients with chest pain of recent onset (171). Its use has been shown to be associated with better patient outcomes (172) and gives important prognostic information both in symptomatic and asymptomatic individuals (171). In contrast to oncology scans, which are normally performed with a helical acquisition, CTCA is typically performed with an axial technique to reduce radiation dose and improve image quality. In addition, patients usually undergo heart rate control (typically with the use of intravenous beta-blockade) to improve image quality (173). However, adopting CTCA techniques could improve the image quality of staging studies in other specific ways. Cardiac-related motion artefact can be minimised with gating leading to more accurate measurement of lung nodules (62). Similarly, peripheral pulmonary arteries move considerably with cardiac motion which could be reduced by gating (63, 64). In addition, by modification of CTCA technique, information could be obtained about the heart for “free” (that is without any additional radiation burden, or procedural time or cost) at the time of staging as part of the oncology scan, which could be potentially useful for oncology patients.

To assess the feasibility and the usefulness of ECG gating in the oncology setting, a prospective study of patients undergoing follow-up oncology scanning, where the patients received both a gated and non-gated study of the thorax, was performed. The scans were then evaluated for cardiac-related motion artefact within the lungs and radiation dose. We also evaluated the coronary arteries to see how many coronary artery segments had acceptable image quality.

6.3 Materials & methods

This prospective single centre study was performed with regional ethics committee approval in 80 patients who provided signed informed consent. The study period occurred between April and October 2015. The participants were consecutively recruited until the sample size had been reached. If the patient declined to take part in the study, the information related to their group was not recorded.

Inclusion criteria was: patients undergoing routine follow-up for malignancy requiring CT thorax abdomen and pelvis scan; age 40 years of age or greater at the time of scan; able to provide informed written consent; able to hold their breath for at least 10 seconds; in sinus rhythm; able to follow verbal commands; able to lie supine for the scan. Participants were excluded if their estimated Glomerular Filtration Rate was less than 30, had a known contrast reaction or if pregnancy could not be excluded.

Patient demographic details (weight, height, and body mass index [BMI]), malignancy status and cardiovascular history were taken at the time of consenting and also from medical records.

Imaging protocol

All exams were performed with a 64-row CT scanner (Discovery 750 HD, GE Healthcare) with and without iodine contrast agent. Participants first had a modified standard of care CT thorax helical scan (number 1, non-contrast), followed by the research gated step-and-shoot CT thorax (number 2, contrast) (Table 12). The modification was necessary to minimise contrast burden.

Technique	No. 1 = STD	No. 2 = Research contrast CT chest	
Contrast Type	Non-contrast	Optiray 350	
Volume (ml)	NA	100-120	
Flow Rate (ml/ sec)	NA	5.5 to 6.5 ml/ min (according to kVp)	
Start Location	Lung Apices	Lung Apices	Above diaphragm
End Location	Below diaphragm	Below diaphragm	Iliac Crest
Respiration	Inspiratory	Inspiratory	Inspiratory
Scan Start Delay (sec)	NA	Smart prep	+40 seconds
Scan Type	Helical full	Cine	Helical full
Helical Collimation (mm)	40	40	40
Coverage per rotation (mm)	40	40	31.74
Rotation Time (sec)	0.5	0.35	0.5
Image slice width (scan)	1.25	0.625	1.25
ASiR	30%	30%	30%
Noise index	39.68	NA	31.74
Pitch Factor	1.375:1	ECG linked	1.375:1
Speed (mm/ rotation)	55	ECG linked	55
KV	120	80 -120 (according to BMI)	120
mA	100 – 750	250 – 420 (according to BMI)	100-750
Automated mA	Yes	NA	Yes
DFOV	Large body	Cardiac large	Large body
Reconstruction Algorithm	Standard	Standard	Standard
Abbreviations: AsiR = adaptive statistical iterative reconstruction; DFOV = display field of field;			
ECG = electrocardiogram; KVp = peak kilovoltage; NA = not applicable; min = minute; mm = millimetre;			
mA = tube current; ml = millilitre; sec = seconds; STD = standard of care			

Table 12. The chest CT protocols of standard of care and research scans

The research study had varied tube current and peak voltage settings according to BMI as standard practice for CTCA (174). Diastolic trigger (75% of the R-R interval) was used if the heart rate was less than 65 bpm and systolic trigger was used otherwise (45% of R-R interval). No oral or IV beta-blocker was given. A bolus of 125 ml iodinate IV contrast

(ioversol; Optiray 350, Tyco Health Care UK, Gosport, UK) was injected at 5.5 ml/s. A bolus-tracking method (Smartprep, GE Healthcare) was used to time the scan trigger whilst imaging at the level of the pulmonary artery bifurcation with a region of interest in the ascending aorta. This was followed by a 50 ml saline bolus "chaser", injected between 4.5 and 5.5 ml/s via a dual pump injector (Tyco Health Care UK). The standard of care CT abdomen and pelvis scan was subsequently acquired using the same volume of contrast immediately after the two successive CT thorax at the appropriate timings, adhering to the departmental protocol.

Assessment of Subjective Image Quality

Images were evaluated using the PACS workstation (Centricity, GE Healthcare). For image quality score, if the reviewers differed by only 1 after their separate reading sessions, a mean score of the two reviewers was obtained and used for statistical analysis. A greater difference in score was treated as a discrepancy and a consensus was sought.

Two series of images (standard of care, taken as a reference, and gated study) were randomly and anonymously presented to two chest radiologists (V.R. and C.R., with 7 and 25 years of experience respectively). Anonymisation was achieved by removing annotations and images were displayed in a standard lung window/ level (1600/−400) to blind the presence of contrast to reviewers. The order of the two series of images was chosen randomly. Calibration was carried out using 10 separate random cases where the reviewers discussed the scoring. A total of 160 images (80 patients who each underwent two types of examinations) were interpreted.

Subjective scoring was carried out in the upper, lingula/ middle and lower lobes. These are representative areas of the entire lungs which were chosen based on an image quality study (61). For individual segments of the lungs, a 4-point Likert scale was used and the reviewers subjectively scored motion artefact (including edge blurring artefact and double-line artefact) in each lobe on the axial images and stair-step artefact was scored on the sagittal images (only for research scans). We devised a scale based on the one described by Shuman et al and added modifications relating each score to diagnostic confidence (61). Blurring of the edges of bronchovascular structures was defined as indistinctness of the interface with the adjacent lung. Double line artefact was defined as parallel duplication of a single linear structure. Stair-step artefact was defined as an abrupt linear structure extending across the image on sagittal images (scored by C.L.P and T.P, radiology residents), which was only scored on the research axial scans. The described artefacts were subjectively scored as “1 = severe and uninterpretable”, “2 = moderate and significant impact on diagnostic confidence”, “3 = minor and mild impact on diagnostic confidence” and “4 = none”.

All gated CT thorax images were reviewed for coronary artery image quality by two level 3 BSCCT certified cardiac imagers (C.R and G.M.H, both with 12 years of experience). All gated CT thorax raw data was reconstructed to a 20 cm field-of-view using standard reconstruction for evaluation. The reviewers were blinded to patient demographic details and clinical information. Segments of the three main coronary arteries and their major side branches down to a minimum diameter of 1.5 mm were defined according to the 15-segment American Heart Association model (AHA) (175) (Figure 12). The AHA model was used, instead of the 18-segment Society of Cardiovascular Computed Tomography model as it has been used in many observational cohort studies in the past. Segments smaller than 1.5 mm were defined as absent. Coronary artery segment image quality was scored on a 4-point Likert scale similar to scales used by previous authors investigating coronary artery image quality in cardiac CT (176, 177). Image quality was classified for each segment as being “1 =

non-diagnostic" (lack of vessel wall definition due to marked motion artefact, poor vessel opacification, prominent structural discontinuity, or blurring related to high image noise preventing evaluation), "2 = poor" (moderate motion artefact or noise-related blurring, fair vessel opacification, or minimal structural discontinuity), "3 = moderate" (minor motion artefact or noise-related blurring, good vessel opacification, and no structural discontinuity) or "4 = good" (absence of motion artefact and noise-related blurring, excellent vessel opacification, and no structural discontinuity).

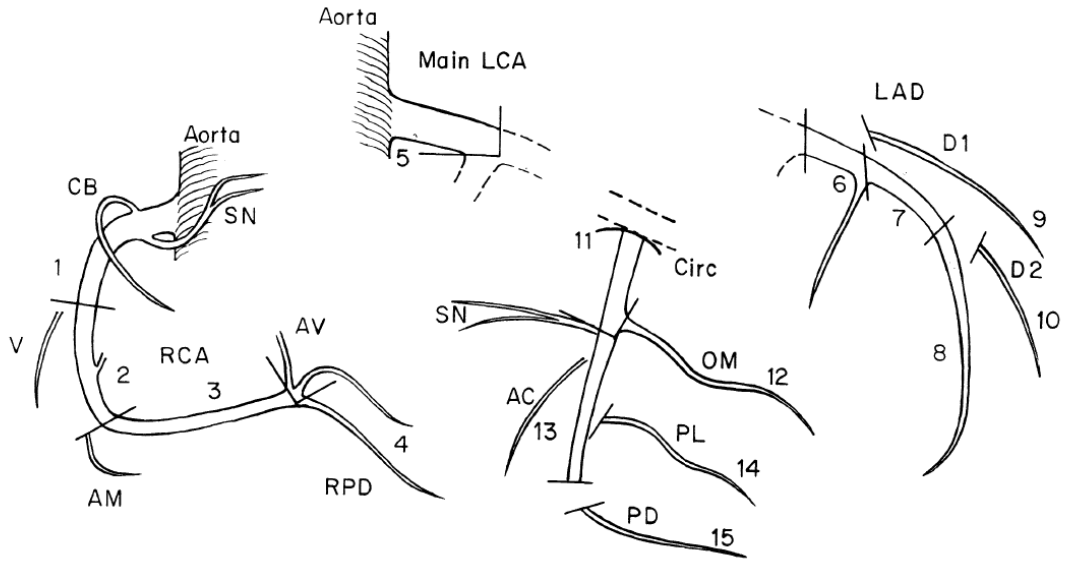

Figure12. AHA 15-segment model adopted from (175).

Quantitative Analysis

Researchers measured CT attenuation and image noise by placing circular region of interests (ROIs) in a homogenous anatomical area on axial planes using the functions on ADW Workstation and AW Server (GE Healthcare) on 1.25 mm slice thickness (for lungs). Image noise was taken as the SD in an ROI. Three contiguous ROIs were used for any anatomical area. The CT attenuation and image noise of the three ROIs was recorded to estimate mean values. All scans were loaded at the same time and attention was paid to

placing ROIs at similar locations for both scans. There was no homogenous area for the lung parenchyma. A method of simply displaying the CT attenuation and image noise was adapted from the literature (178). ROIs measuring at least 10 by 10 mm were placed in the pectoralis minor, pectoralis major, paraspinal muscle, sub-axillary fat and also in the air anterior to the participants (Figure 13). For the coronary arteries, the analysis method was modified from the previously described technique (179). Three contiguous ROIs were placed in the right lobe of the liver, ascending aorta at the origin of the left main stem and at the interventricular septum. ROIs measuring at least 14 by 14 mm were used in the liver and ROIs measuring at least 5 by 5 mm were used in the ascending aorta. In the event of partial hepatectomy, the remaining liver was used instead.

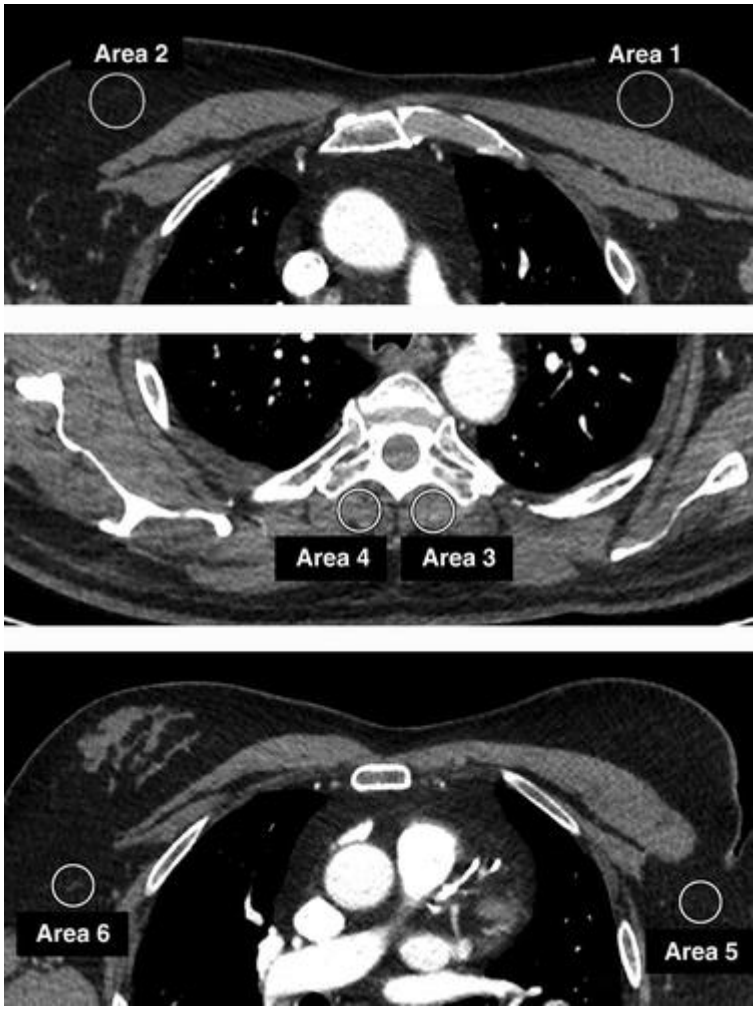

Figure 13. Representative example of where anatomical areas were chosen for quantitative analysis. Area 1 & 2 = anterior thoracic wall subcutaneous fat; area 3 & 4 = paraspinal muscles; area 5 & 6 sub-axillary fat. Adapted from (178) and performed at Plymouth Hospitals NHS Trust.

Radiation Dose

The research scan used prescribed tube potential and current based on BMI and previous work (180-183). There was no automatic tube current modulation employed and therefore the radiation dose the scanner delivers for a given BMI is known. The dose–length product (DLP) displayed by the CT scanner for both the modified standard of care and gated thorax scan was recorded. A conversion factor of 0.014 was used for the calculation of effective dose (184).

Statistical analysis

We conducted statistical analysis using STATA version 14.0 (StataCorp, College Station, Texas). A p-value of <0.05 was considered statistically significant. Bonferroni correction was used for multiple comparisons. In this non-inferiority study the sample size of 80 was based on 95% agreement in lung image quality being acceptable between the research and standard of care CT thorax, and the lower limit of non-inferiority between the two tests was 90%. Variables were presented as numbers and percentages or means and SD. The subjective and quantitative assessments of image quality were compared by using Wilcoxon signed rank test and paired t-test as appropriate. Cohen's kappa and proportion of concordant cases was used to assess inter-observer variability (<0.2 negligible; 0.2-0.4, weak; 0.4-0.6, moderate; 0.6-0.8, good; and >0.8, substantial) (185). Image quality scores of 3 and 4 were assumed to be acceptable. A linear weighted kappa was used to reflect that.

6.4 Results

We included 80 participants, the mean age was 63.9 years +/- 10.3 (range 43.1-92.2 years). There were 33 men (41%; mean age 67.1 years +/- 10.4 & range 43.1-92.2 years) and 47 women (59%; mean age 61.2 years +/- 9.7 & range 46.5-84.2 years). Patients with a wide range of body types were included in our investigation and patients' weights ranged from 45 to 114 kg (mean weight 78.2 kg +/- 15.2) with calculated BMIs ranging from 18.4 to 35.5 kg/m^2 (mean, BMI 27.5 kg/m2 +/- 4.1). A wide range of primary malignancies were present, with colorectal and anal cancers being the most common type (31.3%). Most participants had no previous coronary events, investigations or interventions. Other cardiovascular risk factors are shown in Table 13.

Characteristic	Results
No. of men^	33 (41)
Mean age (y)*	63.9 +/- 10.3 (43.1-92.2)
Weight (kg)*	78.2 +/- 15.2 (45.4-114.0)
Height (cm)*	168.1 +/- 10.4 (147.3-190.5)
Mean BMI (kg/m2)*	27.5 +/- 4.1 (18.4-35.5)
No. of patients with BMI </= 18.5: thin/ malnutrition^	1 (1.3)
No. of patients with BMI >18.5 - </=24.9: normal^	17 (21.3)
No. of patients with BMI >24.9 - </= 29.9: overweight^	39 (48.8)
No. of patients with BMI >29.9 - </= 34.9: moderately obese^	21 (26.3)
No. of patients with BMI >34.9: severely obese^	2 (2.5)
Primary malignancy types^	
Gynaecological cancers	11 (13.8)
Colorectal & anal cancers	25 (31.3)
Urinary tract cancers	15 (18.8)
Upper gastrointestinal tract, liver & pancreatic cancers	8 (10)
Neuroendocrine cancers	3 (3.8)
Lung cancers	2 (2.5)
Lymphomas	3 (3.8)
Breast cancers	7 (8.8)
Two or more primary cancers	4 (5)
Unknown primary cancer & others	2 (2.5)
Baseline macroscopic metastasis based on previous imaging^	
Yes	43 (53.8)

No	37 (46.3)
Progression of disease on standard of care scan^	
Yes	21 (26.3)
No	59 (73.8)
Lung nodules > 4mm on standard of care scan^	
Yes	31 (38.8)
No	49 (61.3)
History of documented history of hypertension in primary care/ on at least one antihypertensives^	
Yes	20 (25)
No	60 (75)
History of smoking^	
Never smoked	40 (50.0)
Current smoker	11 (13.8)
Ex-smoker	29 (36.3)
History of diabetes mellitus^	
Yes	5 (6.3)
No	75 (93.4)
Family history of premature cardiac death^	
Yes	4 (5)
No	76 (95)
History of previous myocardial infarction^	
Yes	4 (5)
No	76 (95)
Presence of coronary stent or coronary arterial bypass graft^	
Yes	4 (5)
No	76 (95)
Previous CTCA or invasive coronary angiogram^	
Yes	9 (11.3)
No	71 (88.8)
*Data are of the mean +/- SD; data in parentheses are the range	
^Data in parentheses are the percentages	

Table 13. Baseline characteristics of the population (n=80).

Objectively, there was a slight difference between the standard of care and the research study due to the presence of contrast in the research study (Table 14). A total of 480 lobes of lungs (the presence of 2 artefacts in the upper, middle/ lingula and lower lobes) were subjectively scored to assess the lung image quality. Figure 14 shows the difference in image quality score between the two techniques.

	No. 1 = STD	No.2 = Research	Mean Difference	95% CI	P value
Average HU					
Pectoralis major muscles	34.81 +/- 14.73	40.88 +/- 17.12	*-6.04	*-3.66 to -8.48	<0.001
Pectoralis minor muscles	32.63+/-10.73	42.53 +/- 14.4	*-9.9	*-12.7 to -7.11	<0.001
Paraspinal muscles	17.02 +/- 15.88	30.94 +/-14.62	*-13.91	*-17.16 to -10.66	<0.001
Sub-axillary fat	*-105.57+/-14.09	*-104.47 +/- 15.53	*-1.1	*-3.12 to -0.92	2.810
Air anterior to the chest wall	*-999.13 +/- 5.83	*-997.40 +/- 6.61	*-1.73	*-2.60 to -0.87	0.001
Image noise					
Pectoralis major muscles	22.49 +/- 7.66	38.27 +/- 9.81	*-15.78	*-18.81 to -12.76	<0.001
Pectoralis minor muscles	22.12 +/- 5.97	36.99 +/- 8.31	*-14.87	*-16.72 to -13.02	<0.001
Paraspinal muscles	35.3 +/- 7.71	54.08 +/- 11.42	*-18.78	*-21.4 to -16.2	<0.001
Sub-axillary fat	20.42 +/- 15.71	31.48 +/- 7.52	*-11.06	*-14.69 to -7.43	<0.001
Air anterior to the chest wall	*-0.19 +/- 113.73	15.69 +/- 7.73	*-15.88	*-41.07 to -9.32	2.135
Dosimetry					
Scan range (mm)	280.47 +/- 24.42	272.52 +/- 36.75	7.95	1.56-14.34	0.015
DLP (mGy.cm)	297.98 +/- 134.98	154.46 +/-60.71	143.53	121.04-166.02	<0.001
CTDIVol (mGy)	**8.63 +/- 3.88	5.64 +/- 2.1	2.95	2.34-3.56	<0.001

Effective dose ICRP 103 criteria	8.05 +/- 3.64	4.17 +/- 1.64	3.88	3.27-4.48	<0.001
Abbreviations: CI = confidence interval; HU = Housefield unit; STD = standard of care; DLP = dose length product;					
CTDIVol = CT dose index. ^Paired t-test used to estimate the 95% confidence intervals and p values.					
*Bonferroni correction is used for multiple comparisons. **There is apparent 23% deviation in scanner reported to calculated					
CTDIVol is a result of an additional and unaccounted 6cm over scan for the helical model for the stated protocol					

Table 14. Objective difference and dosimetry of standard of acre and research CT (n = 80).

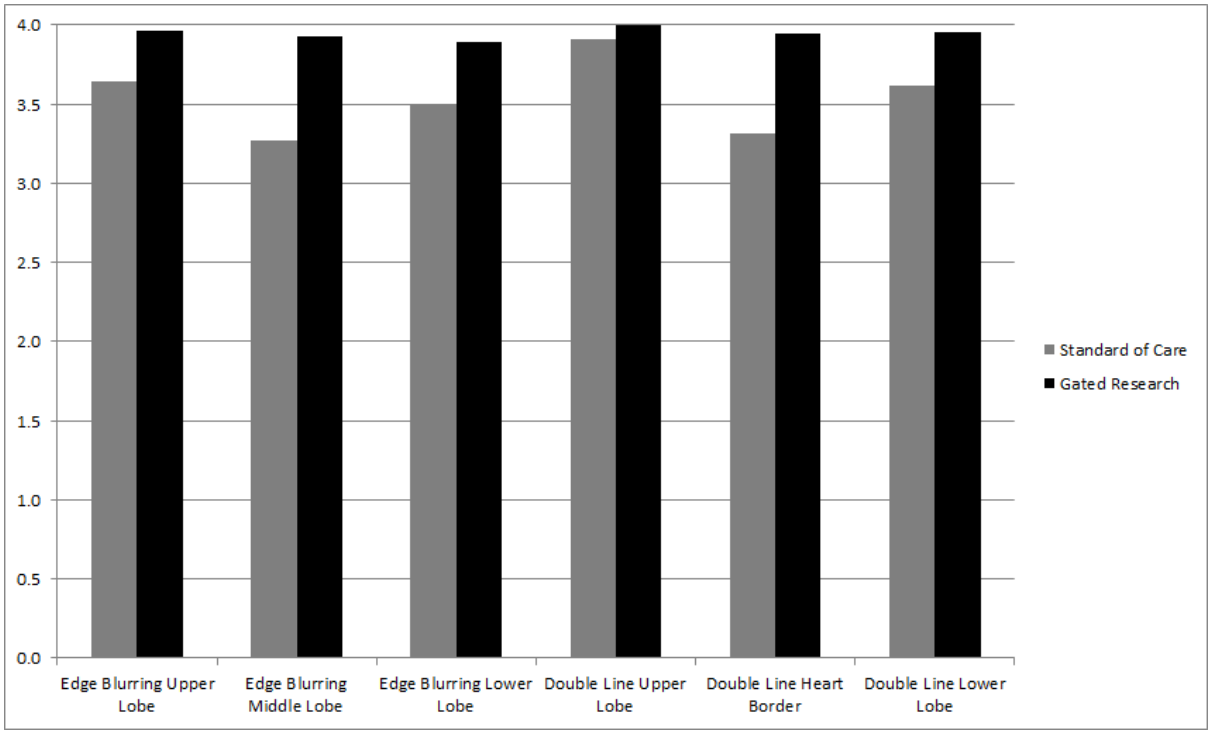

Figure 14. Comparison between standard of care and gated scans: The difference in subjective score between chest CTs.

Image quality was equivalent in the research CT thorax compared to the standard of care across all categories. The maximum subjective difference in image quality was observed in the middle lobe (STD = 3.27+/-0.73; Research = 3.93+/-0.26, $p<0.001$) and heart border (STD = 3.31+/-0.70; Research = 3.94+/-0.23, $p<0.001$). Agreement between assessors improved with the research gated study. Step artefact in the research gated study did not significantly affect the acceptability (3.79+/-0.5), where 2.50% had severe or moderate, 21.3% had minor and 76.3% had no artefacts. There was moderate inter-observer agreement for the lung image quality assessment of both the research and standard of care CT thorax (weighted kappa for overall image quality 0.59 and 0.52 respectively). The proportion of concordant cases improved with the gated technique. The following demonstrate the image quality between the standard of care and research scans (Figure 15, 16, 17).

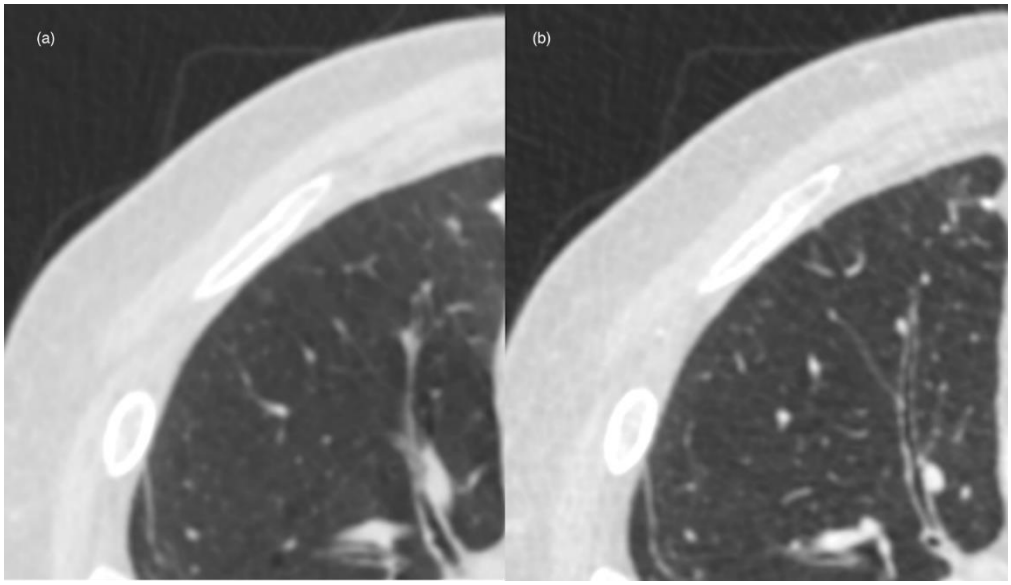

Figure 15. 71 year old man, BMI 30.3, received regular CT surveillance for prostate cancer, classed as never smoked. Images displayed in lung windows. (a) The standard of care scan showed edge blurring artefact affecting diagnostic confidence in the middle lobe. (b) The equivalent research scan showed no edge blurring artefact. The DLP of the standard of care scan was 515.7 mGy-cm compared to the DLP of the research scan was 270.3 mGy-cm.

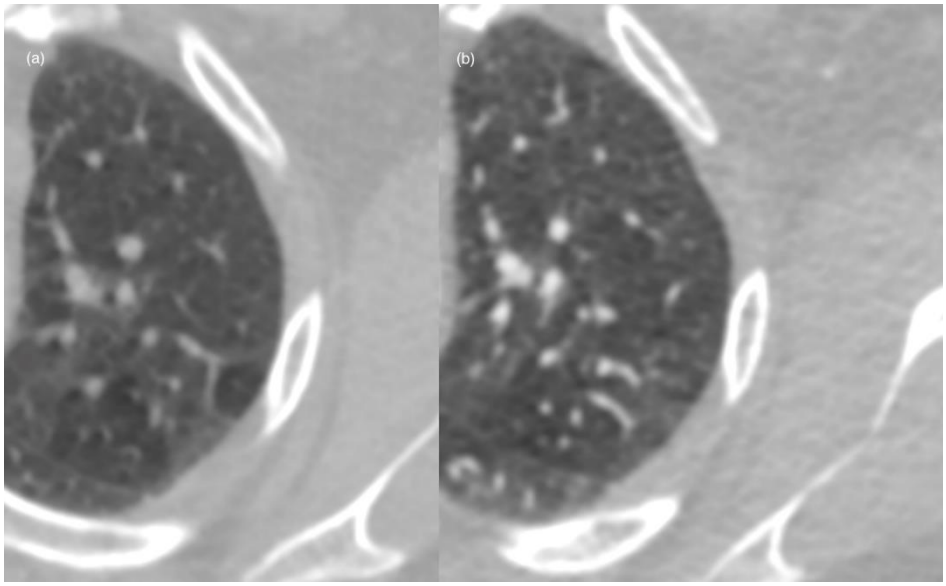

Figure 16. 48 year old woman, BMI 34.9, received regular CT surveillance for endometrial cancer, classed as ex-smoker. Images displayed in lung windows. (a) The standard of care scan showed edge blurring artefact in the upper lobes. (b) The equivalent research scan

showed no edge blurring artefact. The DLP of the standard of care scan was 452.2 mGy-cm compared to the DLP of the research scan was 180.2 mGy-cm.

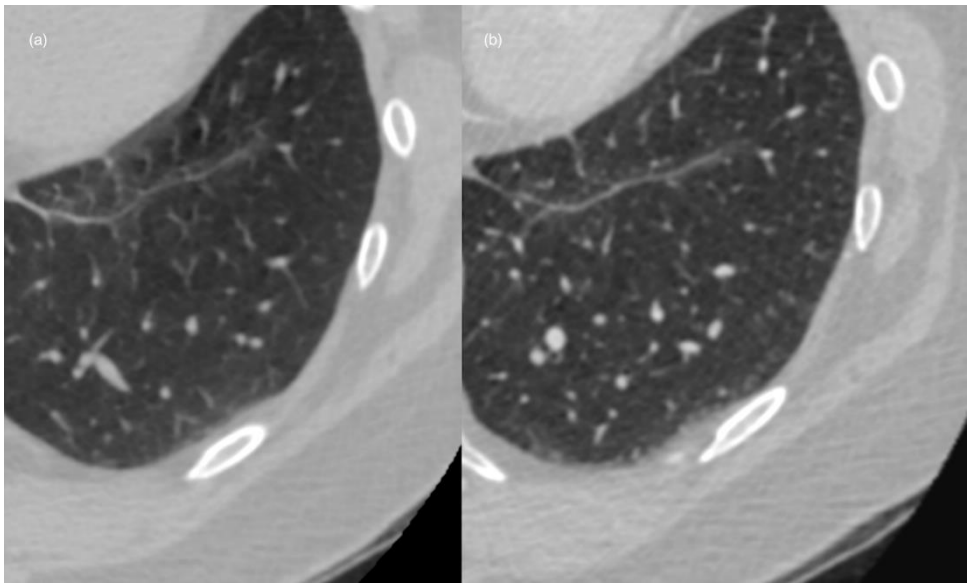

Figure 17. 65 year old male, BMI 33.4, received regular CT surveillance for renal cancer, classed as ex-smoker. Images displayed in lung windows. (a) The standard of care scan showed double line artefact in the heart border affecting diagnostic confidence. (b) The equivalent research scan showed minor double line artefact. The DLP of the standard of care scan was 611.9 mGy-cm compared to the DLP of the research scan was 240.3 mGy-cm.

Additional information regarding the coronary arteries was gained in the process of adopting the gated technique. In terms of coronary artery image quality, 5 of a total of 80 cases (75 coronary segments) were excluded from the analysis due to poor contrast timing during acquisition and were classed as non-diagnostic. These occurred early on during the study. A potential of 1125 coronary segments were subjectively evaluated for their image qualities. Eighty coronary segments (7.1%) were not evaluable as they were classed as absent. The overall impression score, independent of the individual segments, was 3.3+/-0.83 with good

agreement (kappa 0.62); the average image quality of the coronary arteries as a function of the individual segments was 3.43+/-0.68 with moderate agreement (kappa 0.51). Using a heart rate of 85 as a threshold, the proportion of acceptable coronary segments dropped off considerably above that (Table 15). There were 3 participants with coronary stents and 2 participants had coronary artery bypass grafts (including LIMA, LAD, RCA & diagonal grafts) (see an example in Figure 18). All these cases all had acceptable quality images and were satisfactorily evaluated (see another example in Figure 19).

Heart Rate	Participant	Potential Diagnostic Segments	Not Present Segments	Actual No. of Diagnostic Segments	Proportion of Diagnostic Segments	Lower 95% CI	Upper 95% CI
</=60	10	150	2	148	100.0%	n/a	n/a
>60-</=70	21	315	26	254	87.9%	83.6%	91.4%
>70-</=80	22	330	23	265	86.3%	82.0%	90.0%
>80-</=85	9	135	12	102	82.9%	75.1%	89.1%
>85-</=90	5	75	7	39	57.4%	44.8%	69.3%
>90-</=110	8	120	10	75	68.2%	58.6%	76.7%
Total	75	1125	80	883	84.5%	82.2%	86.6%
Percentage of diagnostic segments displayed as mean and 95% confidence intervals. Abbreviations: n/a = not applicable; CI = confidence intervals							

Table 15. The correlation between heart rate and the number of diagnostic coronary segments.

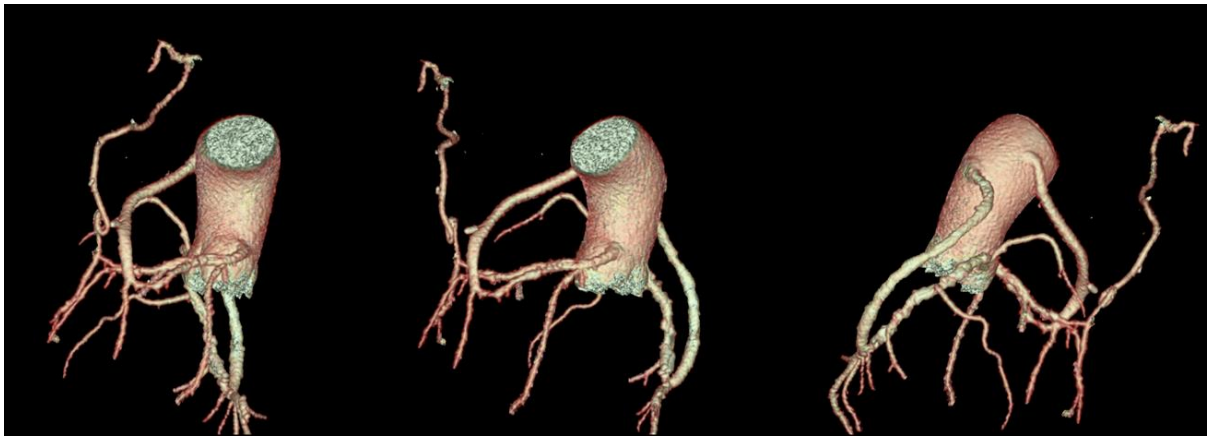

Figure 18. 81 year old man, BMI 31.8, received regular CT surveillance for rectal cancer, classed as ex-smoker and was in sinus rhythm (acquisition heart rate 41-44 bpm). A prospectively gated axial CT scan with diastolic triggering allowed evaluation of the LIMA (left image), venous to diagonal (middle image) and venous to right coronary artery grafts (right image). The DLP of the axial scan was 240.3 mGy-cm compared to the DLP of the helical scan was 465.7 mGy-cm.

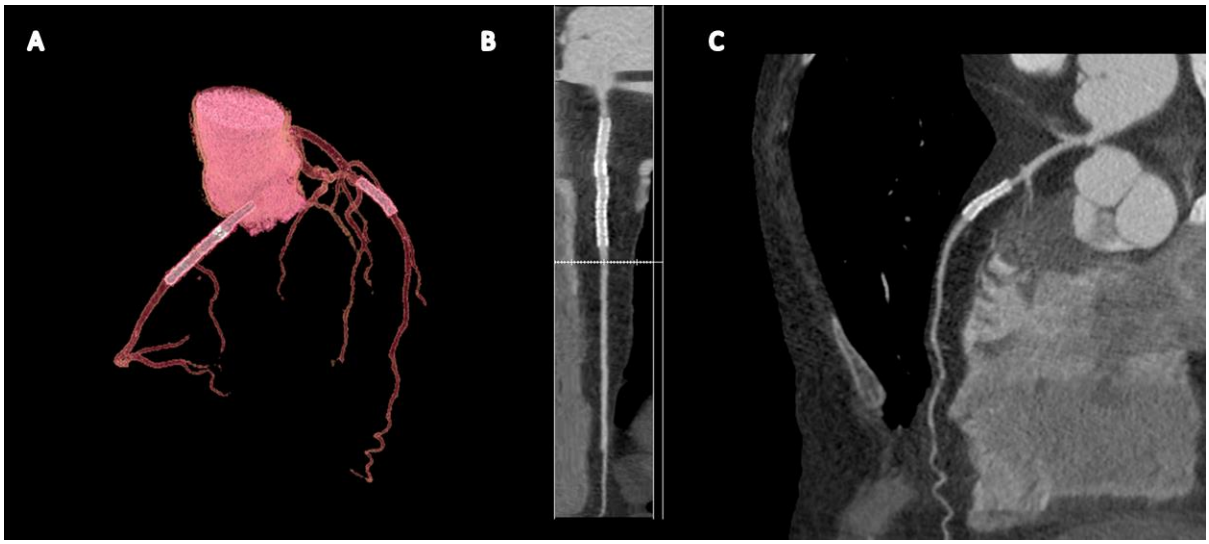

Figure 19. 58 year old woman, BMI 27.1, received regular CT surveillance for vulvar cancer, had previous coronary investigations. (A) A prospectively gated axial CT scan with diastolic triggering (acquisition heart rate 58-62 bpm) had good image quality that enabled reconstruction with volume rendering of the coronary tree and stents within the left anterior descending and right coronary artery. (B & C) Curved multiplane reformat images allowed further scrutiny of the right coronary stent.

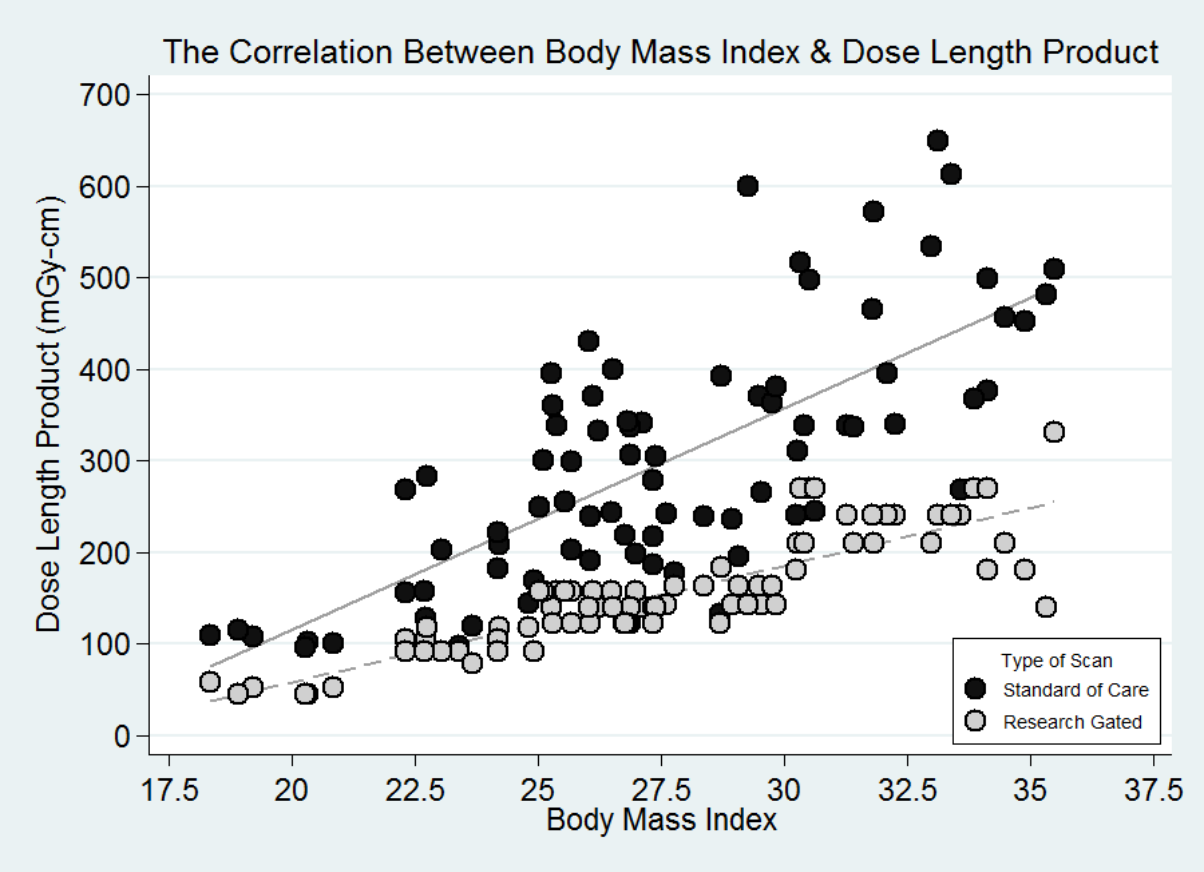

Figure 20. The correlation between body mass index and dose length product in both the helical (ungated) and axial (ECG gated) scans.

For the research CT thorax, the mean DLP was 152.9 mGy-cm +/- 63 and the effective dose was 2.2 mSv +/- 0.8. For the standard of care CT thorax, the mean DLP was 298.0 mGy-cm +/- 135.0 and the effective dose was 4.2 mSv +/- 1.9. Figure 20 displays the correlation between BMI and DLP.

6.5 Discussion

This study was performed to see whether ECG gating of the thorax with 64-slice CT would allow simultaneous assessment of the coronary arteries. For this to be feasible, it was important that there was no detriment in the image quality of the lungs. The image quality in

the parts of the lungs susceptible to movement blur improved with the prospectively gated studies when compared to the matched ungated studies (net improvement, $p<0.001$). These results echoed previous findings where cardiac related motion artefact improved with ECG gating (61, 66-68, 186). One potential disadvantage of axial scanning is artefact due to the prolonged breath hold (6-8 seconds for a typical helical scan versus 12 seconds for the same coverage with ECG gating). This concern was not realised, however. ECG gating technique with 64 slice CT has no detriment on the image quality of the lungs.

CT scanning of the chest is performed routinely as part of the staging process for the majority of patients with cancer. Our study shows that minimal modification of existing techniques makes visualisation of the coronaries possible at no extra cost or additional radiation exposure. Whilst this is currently not a routine practice, our study demonstrates that coronary evaluation before and during cancer treatment is possible in the majority of patients with heart rate less than 85 and during a routine appointment (see Table 14).

Our study has shown that imaging of the coronary arteries at the time of oncology follow-up is feasible and has multiple potential uses. Dedicated CTCA has proven itself to be a powerful prognostic tool. Prognostic information indicated by the presence/ severity of calcific atheroma burden, plaque characteristics and luminal stenosis is now available on CTCA (187). Patients without risk factors and without obstructive disease on CTCA show low-rate all-cause mortality at 5-years (188). In contrast, both all-cause and cardiac-related mortality increase with increasing atheromatous burden and luminal coronary narrowing (125, 187, 189). Therefore, consideration should be given as to whether full chest gated CT with additional coronary information is useful in the oncology setting. With prospective ECG gating, risk stratification of cardiovascular events in those at risk can be achieved before or during cancer treatment with no radiation penalty whilst maintaining image quality. Also, the

use of this technique would allow investigation as to whether coronary artery disease is an independent risk factor for prognosis in a variety of common cancers (190).

This additional information may be useful in perioperative/ peri-treatment risk assessment (102, 105, 111, 115, 128, 130). Perioperative cardiac events are a leading cause of death in the 234 million major surgical procedures performed worldwide annually (191, 192). Providing information about perioperative risk helps identify patients who could benefit from receiving more intensive treatment of their cardiac condition, or choosing a less invasive surgical procedure (193). Clinical risk indices are the common method of assessment (194-196) but they underestimate the risk of major perioperative complications (197). CTCA, however, allows identification of obstructive and non-obstructive coronary artery disease and, compared to the Revised Cardiac Risk Index, can appropriately improve risk estimation amongst patients who will experience perioperative cardiovascular events (198).

There are limitations to this study. This is a single-centre study with a relatively small number of patients. Currently, the number of expert readers in CTCA is small and reproducibility could be an issue. The pulmonary vessels were not assessed as the evaluation of lungs and coronary vessels was our objective. Whole chest coronary and pulmonary angiography can be achieved with a small adjustment of contrast delivery timing. Venous thromboembolism is another important cause of mortality and morbidity, especially in those with metastatic disease. We were unable to match the noise indices of the research and standard of care investigation without affecting imaging quality locally. The perceived dose reduction as indicated by a drop in DLP observed using the research protocol is probably exaggerated as the noise index of the standard of care scan is high relative to the research scan. We have only included patients with sinus rhythm and able to breath hold, and therefore the image quality may be worst in a more general population.

Conclusion

Prospectively ECG gated axial CT thorax resulted in equivalent image quality compared to standard ungated CT. Simultaneous whole body staging and coronary imaging is possible both before and during cancer treatment, without an additional time of acquisition or radiation burden. This technique could become a useful investigation in patients with cancer.

6.6 Learning experience

Conducting this primary study has been a valuable experience with many lessons learnt. It was with regret that the ethics committee at the time denied the usage of routine intravenous beta-blockers despite it becoming routine practice during CTCA acquisition. The reason for rejection was due to the perceived invasive nature of intravenous administration of beta-blockade. The desire to use beta blockers was so strong that we returned to the ethics committee a second time, however this request was still denied. The line of enquiry of the study was innovative, however, it did not fit with any study design and hence it is difficult to prove its worthiness. Image quality is a relatively weak outcome and the coronary findings from the research scan cannot be verified due to the design of the study. In addition, there were fundamental methodological flaws in attempting to evaluate the potential of incidental coronary assessment on routine CT whilst investigating lung image quality.

A better proof of concept study would be a diagnostic test accuracy study comparing the proposed new test with an existing test in the pre-operative setting. For example, the coronary findings from the research scan can be an index test compared to the maximum amount of oxygen an individual can consume (VO2 max) in cardiopulmonary exercise tolerance test as a reference test. Cardiopulmonary exercise test is the gold standard test for the evaluation of aerobic exercise in the pre-operative setting. In fact, this is a subsequent

study that was carried out in the same institution, albeit with a cohort design. However, a cohort design allows for incorporation of more powerful patient outcomes, rather than surrogate markers, such as image quality.

Chapter 7 – Conclusion

7.1 Summary of secondary research

Firstly, it is crucial to reflect on the existing tool in CVD risk prediction in asymptomatic individuals. The Framingham model and its iterations (Wilson 1998, ATP III 2002 and D'Agostino 2008) has been extensively validated and has a wealth of evidence to support it (5). However, when applied to subsequent populations, the evidence is sometimes misrepresented. The main issue is variation in implementation of these models, which has been highlighted in the methodology review section and also investigated previously (31). More transparency on key information, including thresholds used for baseline risk categories, the number of thresholds and prediction horizon, is required. Much emphasis has been placed on how an additional test can clarify the uncertainty for those within an intermediate risk group. Due to the large variation in thresholds used to define risk categories, it is not possible to meaningfully compare those labelled intermediate risk between different populations. This makes interpretation of any additional testing difficult, regardless of whether it adds or lacks of incremental value. A lesser known problem is the model building process of any FRS model during the investigation of incremental value. There is evidence that when the baseline model does not perform well (a low AUC initially), which can be due to alteration of the baseline model, the incremental value (as indicated by Δ AUC), will be falsely elevated. There should be more transparency during a multivariate model building process to ensure that it is built for its intended use/ implementation.

Almost thirty decades of research has led to an abundance of technical and observational information about CACS. During this period, a transition of technology from electron beam CT to multidetector technology has taken place. CACS is, however, consistently measured by the Agatston method. The concern regarding the different configurations of scanners and variability was not substantiated (199). However, the consensus is that standardisation of

calcium score measurements is paramount, especially when serial scoring is considered. Using FRS as a baseline model to investigate incremental value of CACS, there have been multiple cohorts that looked into this potential, particularly MESA, St Francis Heart, ESINER, HNR, Rotterdam, DHS and CONFIRM. In general, an association between all-cause mortality and development of CVD events has been observed. Despite the belief that higher CACS indicates a higher risk of developing an event, this piece of work showed an overlap between CACS subgroups. For example, the HR of developing all-cause mortality overlapped between CACS score of 1-100 and >400. CACS yielded additional discrimination in both the FRS 1998 and 2002 groups, with small improved discrimination as indicated by an increase in AUC. In contrast, the reclassification ability of CACS was mixed. In addition, there are ongoing controversies surrounding the nature of NRI and the reporting standard of NRI remains a concern.

A relative lack of information is available to describe the incremental value of TACS in addition to FRS. There were only 3 studies that described an association between TACS and CVD events. When compared with CACS, there is only a relatively minor association between TACS and the development of CVD events. The results adjusted for FRS risk factors were not available within the included studies. One might surmise that the association is unlikely to be significant after adjusting for FRS risk factors. Unless there is further convincing evidence, TACS is unlikely to be a useful biomarker in CVD risk prediction and resources should be focused elsewhere.

Approximately 15 years of research has led to plenty of technical information on CTCA, but observational data is currently limited. Imaging of coronary vessels was not widely available before the advent of MDCT prior to 2005. CTCA as a research tool has emerged to become a routine examination for predominantly low risk chest pain. This is due to comparable

diagnostic test accuracy between CTCA and ICA. However, there is little data on CVD prediction. Regarding incremental values, the largest cohort derives from CONFIRM which dwarfs the other small independent cohorts. The longest published follow-up is currently 5 years with results for longer follow-up still pending. The available data for meta-analysis is up to 2.5 years with promising results indicating association between obstructive coronary disease and development of CVD events. To slightly confuse matters, this association differs depending on whether the reference group is no coronary disease or non-obstructive coronary disease. Based on a relatively short prediction horizon, CTCA initially yielded a small amount of additional discrimination in addition to FRS, however longer term follow-up is required to draw more definitive conclusions.

Conducting a systematic review and meta-analysis on the topic of incremental value of CT biomarkers in addition to FRS has been challenging. At the early stages of the review process, guidance within the literature was limited, which led to a trial and error approach. The meta-analysis of the included studies was mainly hindered by the availability of aggregate data. The main limitation of the review was missing data. The required data was often either not available, selectively reported or crucial information related to the aggregate date, such confidence interval, was not provided. These limit how different strata and subgroups could be constructed. Other key information that determined the allocation of subgroups, including prediction horizon and the iteration of the FRS model, was missing. Finally, there was ongoing debate as to how these cohort studies were conducted and whether meta-analysis summarising incremental value should be attempted.

Going forward, one way to solve the issues outlined above would be individual participant data meta-analysis. However, this relies on the research community contributing the necessary data to allow this. If this is done, it would enable a much bigger sample of data

and could also potentially eliminate the issue of different publications using different thresholds to analyse their data, which makes subsequent meta-analysis using aggregate data difficult. Association on its own is insufficient to substantiate incremental value (37). The variable baseline performance and other aspects of reporting discrimination inflate potential incremental values. AUC analysis remains a good starting point for the investigation of incremental values. Reclassification should follow rather than replace AUC analysis (154). Reporting of reclassification is also insufficient but significant differences between adequate and inadequate reporting practices have not been identified. More rigorous reporting within prognostic cohort studies is desperately needed. There remains no current consensus as to what is the single best measure, however transparent reporting is key (46). From a wider perspective, reporting in prognosis studies needs to be more robust (159-162).

All of the included studies were at prediction model development stage but clinical decision making cannot not be based on a model in development. A few of the included studies looked at internal validation however external validation should be the immediate focus. In short, external validation assesses the calibration of the model in development and corrects any optimism. External validation would require a different population to the one that the prediction model development was based on. A lot more resources would be required to set up new studies. Alternatively, different cohorts with similar characteristics can share information and externally validate one another, which may be a cheaper and more time efficient alternative.

There is little cost-effectiveness information available. One argument for CACS screening is that the cost of conducting a CT scan to establish CACS is $400, whereas the cost of statin therapy is $1000. However, the patent on statins has since expired and generic statins are significantly cheaper. Subsequently, the Prospective Army Coronary Calcium project

recruited young volunteers (aged between 40 and 50) from the United States army and evaluated the cost effectiveness of screening young asymptomatic individuals using CACS (200). It estimated adding CACS to FRS was associated with $11,500 to over $1,000,000 per QALY gained. The calculation was based on relative risk reduction provided the CACS information was used to decide on initiating primary prevention, such as statins. Even among healthy individuals, the huge range indicates the uncertain benefit. Comparing CACS screening, current practice, guidelines and statins for all individuals, the Rotterdam study found that the improvement of QALY was the most beneficial in the intermediate CHD risk group in men but not in women (201). In another population, there was no difference in the long term cardiac events between different racial groups (202, 203). Clearly, not the entire population is going to benefit from the use of additional CT biomarkers. Considering the described variation in clinical and cost effectiveness, one solution is to ask the general public to decide. There is currently no published patient public involvement work in the field of cardiovascular prediction (5).

Radiation dose is seldom an outcome in cardiovascular prediction research. The assumption is often that an extra test will impact positively. The improvement of CT technology means that the radiation burden generated by additional CT scans has been reduced. According to simulation studies, sixty-four slice CTCA is unlikely to increase the number of cancers as estimated using lifetime attributable risk (204). Although the radiation of one scan is small, repeat scanning in young people will lead to cumulative radiation exposure. The starting age of CT scanning is a concern. Progression of CAS is not linked with risk factors and there is not currently sufficient evidence to support the use of CACS as a biomarker for treatment efficacy (205). Serial and repeated interval scans are therefore not justified. Other concerns include the lack of evaluation of the economic and psychosocial consequences associated with incidental findings on cardiac CT. One study reported that incidental findings on cardiac CT are common but not often clinically significant (206). There is also the associated cost of downstream testing regarding clarification of incidental findings.

Finally, it is difficult to translate any of incremental value summary estimate into definite improvement in clinical performance. More than just a positive incremental value measure is required to prove any CT biomarker is useful in clinical practice. Outcomes of individuals in a high CACS group were evaluated in the St. Francis Heart Study randomised controlled trial of atorvastatin and vitamins C and E versus placebo (207). Importantly, there was no significant difference in the occurrence of cardiac events between the treatment and placebo group and there was no change of CACS. Screening for abdominal aortic aneurysm was only implemented after trials showed improved outcomes (208). No other randomised controlled trials demonstrating clinical benefit of treating a positive CACS or CTCA have been identified. The issue with carrying out such a trial is that the event rate among asymptomatic participants is very low, substantially increasing the cost and duration of such a trial. Until such evidence is available, the incremental value of CT biomarkers will be challenging to interpret as better identification of coronary disease needs to be paired with a treatment that improves outcome.

7.2 Summary of primary research

The feasibility study has been carried out to investigate whether it is technically possible to assess coronary arteries in patients having follow-up whole body CT scans for cancer staging. These patients were asymptomatic and did not coronary angiography to assess the coronary arteries. Although it is technically possible to opportunistically examine the coronary arteries, the study has several shortcomings as discussed in the learning experience section of Chapter 6. The take home message is that coronary evaluation is technically possible during routine oncology CT, however the clinical benefit is not proven. Most importantly, there is currently no evidence that incidental coronary disease is linked to a worse outcome. There is also no randomised controlled trial to investigate whether treating

incidental coronary disease improves outcomes. Based on the lack of evidence, the conclusion here is that there is currently no justification of screening.

From the systematic review conducted, there is ongoing interest with regards to CTCA as a CVD risk prediction tool; this has been discussed in the introduction and discussion sections of Chapter 6. The studies reviewed typically look at long term outcomes (classically 10-year risk of developing CVD events) that are relevant to this type of long-term cohort design, such as all-cause mortality and composite cardiac outcomes. If CTCA is applied to CVD risk prediction in the oncology setting, relevant outcomes that are specific to the situation need to be developed first. For instance, outcomes that reflect cardiotoxicity in radiotherapy appears attractive, which can be roughly divided into short-term or late effects. The potential predictive value of CTCA in the setting of cardiotoxicity will, however, be a new hypothesis generating excise as its application implies underlying coronary disease predisposes patients to cardiotoxicity. To my knowledge, there is currently biological evidence to support this hypothesis, which opens to door for potential future research.

7.3 Personal experience & reflections

Acquiring the generic skills relevant to research has been a rewarding experience but the initial learning curve was very steep. Learning to use the statistical package (STATA), searching the literature and extracting information for the systematic review and meta-analysis was particularly challenging. However, the time and effort spent to acquire these skills was necessary to establish a foot hold in research. Relevant to the field of prognostic research, The Keele Prognosis Course provided me with the basics of prognosis methodology. Along with previous experience in the diagnostic test accuracy literature, the most precious lesson I learnt was that designing either a primary or secondary study requires time, reading, drafting and re-drafting. Most importantly, making an informed

decision about the study design at the outset is paramount. To reflect critically, I would say that initial inexperience contributed to the many shortcomings of the feasibility study. With more experience now, and having learnt the hard way, these shortcomings will be avoided in all future research performed.

As a result of the research process, I have learnt a lot about bias assessment and data extraction in the field of CVD risk prediction. With additional learning and some guidance, my next step will be to perform statistical analysis on an observational dataset in order to understand some of the selective reporting practices that were encountered. To overcome the lack of aggregate data, learning how to perform individual participant data analysis appears to be the next logical step. Prior to that, I may need to enhance my base level of statistical knowledge, perhaps undertaking a higher degree in statistics, or learning from the relevant institutions that have experience with this unique type of data analysis.

The knowledge gained about prognosis research is relevant to my own clinical practice in radiology, which is intertwined with emerging technology and analysis. With the rise of the newer techniques, such as textural analysis in radiomics and artificial intelligence, there are a lot of hypes and promises about their potential diagnostic and prognostic benefits. The risk is that the research community can become over optimistic with new technology or biased by those who have a vested interest in promoting these new technologies. Following on from this research process, I am now better equipped to assess new technology and establish whether it meaningfully adds to existing clinical practise or is prognostically significant. With the current scarcity of systematic reviewers performing prognostic type reviews, my experience will hopefully allow me to collate and present relevant evidence on new and emerging technologies to the imaging community.

Chapter 8 – Appendices

8.1 Appendix

Appendix 1 Search strategy

1. (CCTA or CTCA or CAC).tw.
2. "ct coronary angiogra*".tw.
3. "comput* tomogra* coronary angiogra*".tw.
4. "coronary comput* tomogra* angiogra*".tw.
5. "Computed tomography of the heart".tw.
6. cardiac imaging techniques/
7. ((arter* or coronary) adj1 calci*).tw.
8. ((Agatston or calcium) adj1 score*).tw.
9. or/1-8
10. (cardio* adj2 (disorder* or event* or disease* or condition* or isch?emia)).tw.
11. ((cardiac or heart or coronary or acute coronary or stable or unstable) adj2 (angina or disease* or condition* or event* or disorder* or syndrome* or isch?emia)).tw.
12. ((myocardial or coronary) adj1 (infarct* or isch?emia)).tw.
13. ((ST segment* elevat* or non ST elevat* or non ST segment* elevat* or ST elevat*) adj1 myocardial adj1 (infarct* or isch?emia)).tw.
14. (STEMI or NSTEMI or "Non STEMI").tw.
15. exp Cardiovascular Disease/ or exp Myocardial Ischemia/ or exp Myocardial Infarction/ or exp Coronary Artery Disease/ or exp Coronary Disease/ or exp Acute Coronary Syndrome/ or exp Angina Pectoris/ or exp Angina, Unstable/ or exp Angina, Stable/
16. stroke*.tw.
17. (cereb* adj1 (disease* or accident* or h?emorrhage or incident* or isch?emia)).tw.
18. ((ish* or thrombotic or h?emorrhagic) adj1 stroke*).tw.
19. exp stroke/ or exp cerebrovascular disorders/

20. or/10-19

21. "add* value*".tw.

22. "progno*".tw.

23. "reclassif*".tw.

24. "risk stratification*".tw.

25. or/21-24

26. "all-cause mortalit*".tw.

27. ((non fatal or nonfatal) adj1 myocardial adj1 (infarction* or disease* or isch?emia)).tw.

28. (mortalit* or morbidit*).tw.

29. (cardi* adj1 (event* or incident* or mortalit* or morbidit*)).tw.

30. (CABG* or OPCAB* or OPCABG*).tw.

31. (bypass adj1 (surger* or graft* or graft surger*)).tw.

32. ((aortocoronary or heart or cardiac or triple or quadruple or coronary or coronary artery or coronary bypass or coronary artery byapss or off pump or surgical or non surgical or nonsurgical) adj1 (coronary bypass* or bypass surger* or bypass graft* or bypass operation* or bypass* or graft surger* or revasculari?ation*)).tw.

33. (CABG* adj1 (surger* or operation*)).tw.

34. exp Coronary Artery Bypass/

35. ((coronary or coronary artery or percutaneous transluminal coronary or transluminal coronary or percutaneous transluminal or percutaneous or balloon) adj1 angioplast*).tw.

36. (angiograph* adj5 peripheral vascular disease*).tw.

37. "peripheral vascular disease*".tw.

38. exp Peripheral Vascular Diseases/

39. ((hospital* or admission* or inpatient admission* or inpatient care or inpatient procedure* or inpatient or initiat* or first line or second line or third line or treatment* or therap* or preventive treatment* or preventative treatment* or prophylactic treatment* or clinical treatment* or med* or medication* or medical therap* or agent* or drug* or statin* or cholestrol reduct* or cholestrol lower* or HMG-CoA reductase inhibitor* or

Hydroxymethylglutaryl-coenzyme A reductase inhibitor* or antihypertensive* or angiotensin converting enzyme* or ACE or beta* or beta adrenoreceptor antagonist* or calcium channel blocker* or diuretic* or antiplatelet* or clopidogrel* or aspirin* or acetylsalicylic acid or ASA or anticoagulant* or warfarin* or coumadin* or metformin or insulin*) adj3 ((cardiac or heart or coronary or acute coronary or cereb* or stable or unstable) adj2 (angina or disease* or condition* or event* or disorder* or accident* or h?emorrhage or incident* or syndrome*))).tw.

40. or/26-39

41. "framingham risk* score*".tw.

42. "Framingham General Cardiovascular Risk Score".tw.

43. (framingham adj1 (CHD or cardiovascular)).tw.

44. (FRS or GFRS or ATP III or framingham).tw.

45. "adult treatment panel III".tw.

46. ((ATP III or framin*) adj4 (risk* or score* or criteri* or ind* or categor* or profile* or stratif*)).tw.

47. or/41-46

48. 9 and 20 and (25 or 40) and 47

Appendix 2 Inclusion & exclusion criteria of studies investigated CACS and/or TACS

Year First Author	**Inclusion Criteria**	**Exclusion Criteria**	**Endpoint**
2013 Agarwal	Type 2 diabetes after age 34, with/ without CVD	Diabetic ketoacidosis or serious health condition (e.g. advanced nephropathy)	Composite = 1) all-cause mortality 2) MI, cardiac arrest, arrhythmia, PVD, stroke or HF
2011 Ahmadi	Veterans presented with suspected CHD	Subjects with established CVD, stroke, diabetic retinopathy, end-stage renal disease, Raynaud syndrome, infection, cancer, immunosuppression, systemic inflammation status, or end-stage liver disease	Single = all-cause mortality
2005 Arad	Age 50-70, without any history, symptoms & signs of CVD	Extensive exclusion criteria relating to any diagnosis, medication or biomarkers suggesting underlying CVD	Composite = 1) all-cause mortality 2) non-fatal MI/ coronary death, coronary revascularisation (bypass surgery or percutaneous angioplasty), stroke or PVD surgery
2007 Budoff	n/a	n/a	Single = all-cause mortality

2015 Chang	Any cardiovascular risk factors	Previous CHD	Composite = cardiac death, non-fatal MI or need for coronary revascularisation
2015 Cho	Age >18, > 64-row CT, interpretable image quality & prospective CHD risk factors collection	Experienced chest pain, unknown symptoms status, a previous history of MI, coronary revascularisation or cardiac transplant	Composite = 1) all-cause mortality 2) non-fatal MI
2012 Cho	Age >18, > 64-row CT, interpretable image quality, clinical indication for CHD evaluation & prospective CHD risk factors collection	Chest pain, unknown symptom status, previous MI, coronary revascularisation, cardiac transplant, withouth follow-up data of mortality or CTCA findings	Composite = 1) all-cause mortality 2) non-fatal MI

2010 Elias-Smale	Age 55-85 & asymptomatic	Symptomatic candidates defined by presence of MI, CABG or angioplasty at time of scanning, not living independently, lost to follow-up, inadequate image quality, refuse to participate or logistical reason leading to not able to attend follow-up	Composite = fatal & non-fatal CVD endpoints
2011 Elias-Smale	age >55	Symptoms defined by percutaneous coronary intervention, CABG, MI, stroke	Composite = classified according to ICD-10, including MI, CHD mortality, TIA or ischaemic strokes
2010 Erbel	Age 45-75	Known CHD at baseline	Composite = 1) all-cause mortality 2) non-fatal MI, cardiac death, major cerebrovascular, surgical or non-surgical coronary revascularisation, angiographically defined new-onset PVD, hospitalisation for cardiac disease or initiation of medical therapy for cardiac disease

2013 Forouzandeh	Age >18, chest pain within the previous 24 hours suggestive of ischaemia &admission under observational status for stress SPECT	Non-cardiac chest pain, elevated troponin, new/ presumably new ST-segment elevation or depression on baseline ECG, haemodynamic instability, previous CABG, previous angioplasty, women of child bearing potential with known or suspected pregnancy or inability to provide consent	Composite = MACE were defined as cardiac death, non-fatal MI or unstable angina
2014 Gibson	Age 45-84, free of clinical CVD at baseline, four racial ethnic groups from six US communities (African-American, Hispanic, Asian predominantly Chinese descent, White)	Physician-diagnosed MI, angina, HF, stroke, TIA, CABG, angioplasty, valve replacement, pacemaker placement or other vascular surgeries (see original MESA exclusion criteria)	Composite = fatal or non-fatal strokes due to haemorrhage infarct or TIA
2004 Greenland	Age >45 & at least 1 coronary risk	Diabetes, symptomatic, coronary events prior to CT scan	Composite = non-fatal MI or CHD death

	factor		
2010 Hadamitzky	n/a	Known CHD, angina, non-anginal chest pain defined by Diamond Forrester or other symptoms caused by cardiac disease	Composite = cardiac death, nonfatal MI, unstable angina requiring hospitalisation or late revascularisation (>90 days after CTCA)
2013 Hermann	Age 45-75, negative history of previous stroke, CHD & MI	Overt CVD, inability to give informed consent to participate in the study, conditions that preclude follow-up for 5 years, severe psychiatric disorders or illegal substance or pregnancy	Composite = stroke events both ischaemic & haemorrhagic, defined as focal neurologic deficit of presumed cerebrovascular origin >24 hours
2012 Kavousi	Age >55	CHD defined as clinically manifest MI, CABG or angioplasty	Composite = fatal and non-fatal CHD
2012 Lau	Type 2 diabetes, Chinese & FRS <20%	Patients calculated to be at high risk (>20%) of developing a cardiovascular event within 10 years based on the recalibrated FRS for Chinese, prior history of prior ACS, ischaemic stroke, acute limb ischaemia, stable angina, symptomatic PVD, creatinine	Composite = ACS, ischaemic stroke, new onset symptomatic PVD, death due to ACS or ischaemic stroke or symptom driven revascularisation procedures of the carotid, coronary or peripheral

		level >220, severe hepatic disease, malignancy or connective tissue diseases	arteries
2015 Matsushita	Same as Gibson 2014	Same as Gibson 2014	Composite = first incidence of CHD (ranges from angina to death), stroke, HF or PVD
2011 Mohlenkamp	Age 45-75	MI, coronary revascularisation or baseline indication for statin therapy according to Canadian Cardiovascular Society guidelines	Composite = fatal & non-fatal CHD events, CVD mortality, stroke or coronary revascularisation
2011 Mohlenkamp	Age 45-75	Subjects with hsCRP > 10 mg/l suggesting acute inflammation or history of MI or coronary revascularisation	Composite = CHD events
2013 Park	Self referred screening	Prior history of revascularisation or known CHD, poor image quality due to blooming artefact or patients who had	Composite = cardiac death, MI, unstable angina or stroke

		revascularisation procedure within 90 days of CTCA	
2010 Polonsky	Same as Gibson 2014	Original MESA exclusion criteria & diabetes	Composite = MI, death due to CHD, resuscitated cardiac arrest, definite or probable angina followed by coronary revascularisation or definite angina not followed by coronary revascularisation
2001 Raggi	n/a	n/a	Composite = cardiac events
2004 Raggi	At least one cardiac risk factor	Presence of CHD	Single = all-cause mortality
2012 Rana	Volunteers	Previous CVD or symptoms, >80 years old, pregnancy, significant medical co-morbidity, prior coronary catheterization or prior CACS	Composite = first CVD event defined by cardiac death, MI, stroke or late revascularisation >90 days after CACS
2003 Shaw	Referred by primary care due to	History of admission to hospital due to chest pain, ACS, MI, prior coronary	Single = all-cause mortality

	presence of cardiac risk factors	angiography & previous revascularisation	
2013 Versteylen	Had both CACS & CTCA at baseline	Unstable angina, haemodynamic instability, pregnancy, renal insufficiency, severe iodine allergy, history of coronary artery disease, inconclusive CTCA & clinical data missing	Composite = 1) single = all-cause mortality 2) ACS (troponin T elevation, ST segment elevation/ depression of >1 mm, or at least 2 of these symptoms together with invasive angiographic confirmation of culprit lesion), MI, unstable angina, revascularisation
2009 Wong	Had both CACS & TACS at baseline	Atrial fibrillation & history of CVD	Composite = 1) Hard CHD events = MI or cardiac death 2) Total CHD events = hard CHD events plus late revascularisations (>90 days) 3) Total CVD events = total CHD events plus stroke

2012 Yeboah	Same as Gibson 2014 + without diabetes, FRS of between >5% and <20%, complete data on all 6 of novel risk markers	Original MESA exclusion criteria & diabetes	Composite = 1) Incident CHD 2) Incident CVD = incident CHD, stroke or CVD death
2014 Yeboah	Same as Gibson 2014	Original MESA exclusion criteria & diabetes	Composite = 1) single = all-cause mortality 2) Incident CHD event = MI, death due to CHD, resuscitated cardiac arrest, definite or probable angina followed by coronary revascularisation, and definite angina not followed by coronary revascularisation 3) Incident CVD event = incident CHD + stroke, stroke death or other CVD death defined by MESA protocol

Appendix 3 Inclusion and exclusion criteria of studies investigated CTCA

Year **First Author**	**Inclusion criteria**	**Exclusion criteria**	**Endpoint**
2015 Cho	Age> 18, > 64-row CT, interpretable image quality & available CHD risk factors	Experienced chest pain, unknown symptoms status, a previous history of MI, coronary revascularisation, cardiac transplant	Composite = 1) single =all-cause mortality 2) non-fatal MI
2012 Cho	Age >18, > 64-row CT, interpretable image quality, clinical indication for CHD evaluation & prospective CHD risk factors collection	Same as Cho 2015 + without follow-up data of mortality or CTCA findings	Composite = 1) single =all-cause mortality 2) non-fatal MI
2011 Chow	Age> 18, > 64-row CT, interpretable image quality & available CHD risk factors	MI, coronary revascularisation or cardiac transplant	Single = all-cause mortality
2010 Hadamitzky	n/a	Known CHD, angina, non-anginal chest pain defined by Diamond Forrester or	Composite = cardiac death, nonfatal MI, unstable angina requiring hospitalisation and late revascularisation (>90 days after

		other symptoms caused by cardiac disease	CCTA)
2013 Hadamitzky	Suspected but not proven CHD, assessment of both luminal stenosis as well as presence & composition of plaque in CTCA & a follow-up of at least 90 days	Known CHD including self-reported MI; coronary revascularisation, stents/ CABG on CT	Single = time to death from any cause
2010 Chow	Primary indications for CTCA were chest pain or dyspnoea	History of coronary revascularisation, heart transplantation or congenital heart disease	Composite = 1) all-cause mortality 2) all MACE (cardiac death & non-fatal MI)
2011 Lin	Normal sinus rhythm & capable of breath-hold	Known CHD, prior coronary revascularisation, CTCA identified obstructive CHD (>50%) or non-verifiable Social Security numbers	Single = all-cause mortality

2013 Park	Self referred screening	Prior history of revascularisation or known CHD, poor image quality due to blooming artefact or patients who had revascularisation procedure within 90 days of CTCA	Composite = cardiac death, MI, unstable angina or stroke
2013 Versteylen	Had both CACS & CTCA	Unstable angina, haemodynamic instability, pregnancy, renal insufficiency, severe iodine allergy, history of CHD or inconclusive CCTA, clinical data missing	Composite = ACS (troponin T elevation, ST segment elevation/ depression of >1 mm, or at least 2 of these symptoms together with invasive angiographic confirmation of culprit lesion), MI, unstable angina, revascularisation + all-cause mortality

Appendix 4 An example of qualitative documentation for and against pooling aggregate data

Outcome of concern: The added value of CTCA in addition to CACS & FRS

Publications of concern:

1. Cho 2012, CCTA and risk of all-cause mortality and nonfatal myocardial infarction in subjects without chest pain syndrome from the CONFIRM Registry
2. Hadamitzky 2010, Prognostic value of coronary CT angiography
3. Versteylen 2013, Additive value of semi-automated quantification of coronary artery disease using cardiac computed tomographic angiography to predict future acute coronary syndrome

Potential meta-analysis size effects and groups:

Difference in AUC

- When compared with risk factors
    - Cho 2012
    - Hadamitzky 2010
    - Versteylen 2013
- When compared with risk factors and CACS
    - Cho 2012
    - Hadamitzky 2010
    - Versteylen 2013

Evidence for meta-analysis:

- Comparable endpoints – all recorded as composite cardiovascular outcomes, Versteylen 2013 and Cho 2012 included all-cause mortality as composite outcomes while Hadamitzky 2010 did not

- Different patient cohorts
- Overlapping regions
- Comparable inclusion and exclusion criteria
- Comparable mean ages
- Comparable follow-up time
- Comparable CT techniques
- Same year risk prediction – unknown for Versteylen 2013

Evidence against meta-analysis:

- Different designs – prospective, retrospective and unclear
- Different symptoms
    - Asymptomatic in Hadamitzky 2010 and Cho 2012
    - Symptomatic in Versteylen 2013
- Different inclusion and exclusion criteria – mainly Versteylen 2013 include patient with chest pain but not the other two
- Different hazard models – not stated in 2 of the studies
- Different number of participants
- Different FRS models
    - FRS 1998 = Cho 2012, Hadamitzky 2010
    - FRS 2008 = Versteylen
- Different FRS risk categories
- Different event rate 2.2 – 22.3%

Overall verdict for meta-analysis is probably justified provided that Versteylen 2013 is treated separately given that this is a group has higher pre-test probability at baseline, different FRS category and much higher event rate.

Potential meta-analysis groups:

The extra value of CTCA in addition to FRS 1998 and CACS as indicated by difference in AUC in predicting composite cardiovascular outcome

- When compared with risk factors only
    - Cho 2012 (asymptomatic)
        - Subgroup event definition = all-cause mortality + non-fatal MI
        - Adjustment = none
        - Effect size = 0.74 (0.66-0.81) - 0.59 (0.52-0.67) = 0.15
        - Comment = contact author for CI/ SE and/ variance
    - Hadamitzky 2010 (asymptomatic)
        - Subgroup event definition = composite cardiovascular outcome including cardiac death, nonfatal MI, unstable angina requiring hospitalisation and late revascularisation (>90 days after CCTA)
        - Adjustment = none
        - Effect size = 0.22, no CI
        - Comment = author contacted on 07/02/16
    - Versteylen 2013 (symptomatic, as a separate group)
        - Subgroup event definition = Composite = ACS (troponin T elevation, CT segment elevation/ depression of >1 mm, or at least 2 of these symptoms together with invasive angiographic confirmation of culprit lesion), MI, unstable angina, revascularisation + all-cause mortality
        - Adjustment = none
        - Effect size = 0.64, CI 0.52 to 0.76, - 0.59, CI 0.45 to 0.73 = 0.05
        - Comment = contact author for CI/ SE and/ variance
- When compared with both risk factors and CACS
    - Cho 2012 (asymptomatic)
        - Subgroup event definition = all-cause mortality + non-fatal MI

        - Adjustment = none
        - Effect size = 0.74 (0.66-0.81) - 0.71 (0.64-0.78) = 0.03
        - Comment = contact author for CI/ SE and/ variance
    - Hadamitzky 2010 (asymptomatic)
        - Subgroup event definition = composite cardiovascular outcome including cardiac death, nonfatal MI, unstable angina requiring hospitalisation and late revascularisation (>90 days after CCTA)
        - Adjustment = none
        - Effect size = 0.07, no CI
        - Comment = author contacted on 07/02/16
    - Versteylen 2013 (symptomatic, as a separate group)
        - Subgroup event definition = Composite = ACS (troponin T elevation, CT segment elevation/ depression of >1 mm, or at least 2 of these symptoms together with invasive angiographic confirmation of culprit lesion), MI, unstable angina, revascularisation + all-cause mortality
        - Adjustment = none
        - Effect size = 0.64, CI 0.52 to 0.76, - 0.67, CI 0.54 to 0.79 = -0.03
        - Comment = contact author for CI/ SE and/ variance

Appendix 5 Threshold information of CACS, TACS & CTCA

Coronary calcium score						
Author	Year	Reference Group	Group 1	Group 2	Group 3	Other Groups
Agarwal	2013	<10	10-99.	100-299	300-999, >1000	log, ordinal
Ahmadi	2011	Discordant low risk	Discordant high risk			Discordant high risk & 4 adjustments, CAC in discordant low risk
Arad	2005	0	1-99.	100-399	>/=400	n/a
Budoff	2007	0, absence of vessel >100, 3 groups from cumulative survival curve	1-10, 11-100	101-399	400-699, 999, >1000	1-vessel >100, 2-vessel >100, 3-vessel >100, any vessel >100 unadjusted, any vessel >100 adjusted
Budoff	2007	0 (3 subgroups)	1-10, 11-100	101-299, 300-399	400-699, 700-999, >1000	7 other subgroups
Chang	2015	0	</=10, 11-100	101-400	>400	log, 3 other adjusted log
Cho	2012	0 for 2 outcomes & 2 adjusted groups	1-100 (for 2 outcomes)	101-400 ((for 2 outcomes)	>400 (for 2 outcomes)	6 adjusted groups
Cho	2015	?</=10	</=100	100-400	400-1000, >1000	n/a
Elias-Smale	2010	0	>0	n/a	n/a	log
Elias-Smale	2011	1st tertiles for 2 outcomes & 3 adjustments	2nd tertiles (for 3 adjustments)	3rd tertiles (for 3 adjustments)	n/a	n/a
Erbel	2010	0 for unadjusted & adjusted, 1st quartile for unadjusted & adjusted	1-99, 2nd quartile. Both for unadjusted & adjusted.	100-399, 3rd quartile. Both for unadjusted & adjusted.	400-999, >/=400, >/=1000, 4th quartile. All for unadjusted & adjusted.	log unadjusted & adjusted

Forouzandeh	2013	0 for unadjusted & 3 adjustments	1-100	100-400	>400	>0 for unadjusted & 3 adjustments
Gibson	2014	0 for 3 outcomes & unadjusted & adjusted	0-100	100-400	>400	log for 3 outcomes & unadjusted & adjusted, >0 for 3 outcomes & unadjusted & adjusted
Greenland	2004	0, 0 & FRS 0-9, 0 & FRS 10-15, 1-100 & FRS 0-9, 101-300 & FRS 0-9	1-100, 1-100 & FRS categories 1-4	101-300, 101-300 & FRS categories 1-4	>/=301, >/=301 & FRS categories 1-4	trend, per SD increase
Hermann	2013	?0	1-99.	100-399	>/=400	log (including unadjusted & 3 adjustments), men log (unadjusted & 2 adjustments), women log (unadjusted & 1 adjustment)
Kavousi	2012	n/a	n/a	n/a	n/a	log, men log, women log
Lau	2012	<40	>40	n/a	n/a	n/a
Matsushita	2015	non-CKD quartile 2, CKD quartile 2	non-CKD quartile 3, CKD quartile 3	non-CKD quartile 4, CKD quartile 4	n/a	n/a
Mohlenkamp	2011 Cor	0 for unadjusted & adjusted	1-99 for unadjusted & adjusted	100-399 for unadjusted & adjusted	>/=400 for unadjusted & adjusted	n/a
Mohlenkamp	2011 Quan	0 (for 2 different outcomes, unadjusted and 2	1-99 (for 2 different outcomes, unadjusted and 2	100-399 (for 2 different outcomes, unadjusted and 2 adjustments)	>/=400 (for 2 different outcomes, unadjusted and 2 adjustments)	n/a

		adjustments)	adjustments)			
Park	2013	0	0-100	100-400	>400	n/a
Polonsky	2010	n/a	n/a	n/a	n/a	log
Raggi	2001	0 decile	>0, percentile univariate, percentile multivariate, 10th, 20th, 30th decile	40th, 50th, 60th decile	70th, 80th, 90th decile	log, >0
Raggi	2004	<10 (for men & women)	11-100 (for men & women)	101-400 (for men & women)	401-1000 , >1000 (for men & women)	n/a
Rana	2012	n/a	n/a	n/a	n/a	log
Wong	2009	<10 (for 3 outcomes)	10 to 99 (for 2 outcomes)	100-399 (for 3 outcomes)	>/=400 (for 3 outcomes)	log (for 3 outcomes)
Yeboah	2009	n/a	n/a	n/a	n/a	univariate log, multivariate log (adjusted only)
Yeboah	2014	n/a	n/a	n/a	n/a	univariate log (for 3 outcomes), multivariate log (for 3 outcomes)
Han	2015	0 (unadjusted & adjusted)	1-100 (unadjusted & adjusted)	101-400 (unadjusted & adjusted)	>400 (unadjusted & adjusted)	log (unadjusted & adjusted)
Valenti	2015	0 (unadjusted & 2 adjustments)	10-99 (unadjusted & 2 adjustments)	100-399 (unadjusted & 2 adjustments)	400-999 (unadjusted & 2 adjustments), >1000 (unadjusted & 2 adjustments)	>0 (unadjusted & adjusted)
Computed tomographic coronary angiogram						

Cho	2012	none/ non-obstructive (unadjusted & adjusted), none/ normal (unadjusted & 2 adjustments)	non-obstructive (unadjusted & 2 adjustments)	obstructive CAD (unadjusted & adjusted)	1-vessel disease (unadjusted & 2 adjustments), 2-vessel disease (unadjusted & 2 adjustments), 3-vessel disease/ LMS (unadjusted & 2 adjustments)	n/a
Chow	2011	no CAD (3 adjustments)	non-obstructive (3 adjustments)	obstructive low-risk (3 adjustments)	obstructive high risk (3 adjustments)	CAD severity (3 adjustments)
Cho	2015	?	non-obstructive	obstructive	1-vessel disease, 2-vessel disease, 3-vessel disease/ LMS	n/a
Hadamitzky	2010	n/a	n/a	n/a	n/a	presence of CAD, most severe stenosis, number of arteries narrowed, LMS/ proximal LAD stenosis
Hadamitzky	2010	non-obstructive	obstructive	n/a	n/a	n/a
Hadamitzky	2013	n/a	no. of segments with any plaque or stenosis (c-index unadjusted & adjusted)	no. of segments with stenosis >50% (HR, c-index unadjusted & adjusted), no. of proximal segments with >50% stenosis (c-index unadjusted & adjusted)	no. of segments with stenosis >70% (HR, c-index unadjusted & adjusted)	no. of segments with non-calcified plaques, no. of segments with mixed plaques, no. of segments with calcified plaques (c-index unadjusted & adjusted), no. of segments with mixed or calcified or plaques (c-index unadjusted & adjusted), no. of

						proximal segments with calcified or mixed plaques (c-index unadjusted & adjusted)
Hadamitzky	2013	no. of segments with non-calcified plaques	no. of segments with any plaque or stenosis	no. of segments with stenosis >50%, no. of proximal segments with >50% stenosis	no. of segments with stenosis >70%	no. of segments with mixed plaques, no. of segments with calcified plaques , no. of segments with mixed or calcified or plaques, no. of proximal segments with calcified or mixed plaques
Chow	2010	non-obstructive CAD (for 3 outcomes)	n/a	obstructive >50% (for 3 outcomes), obstructive but not high risk (for 3 outcomes)	obstructive >70% (for 3 outcomes), obstructive high risk (for 3 outcomes)	no CAD (for 3 outcomes), CAD severity (for 3 outcomes)
Lin	2011	?	n/a	n/a	1-vessel disease (unadjusted & 2 adjustments), 2-vessel disease (unadjusted & 2 adjustments), 3-vessel disease (unadjusted & 2 adjustments)	1-4 segments (unadjusted & 2 adjustments), >/=5 segments (unadjusted & 2 adjustments), non-calcified plaque (unadjusted & 2 adjustments), mixed plaque (unadjusted & 2 adjustments),

						calcified plaque (unadjusted & 2 adjustments)
Park	2013	0% stenosis	1-49% stenosis	50-69% stenosis/ LMS 1-49%	>/=70%/ LMS 50%	degree of stenosis (unadjusted & adjusted)
Park	2013	0% stenosis	1-49% stenosis	50-69% stenosis/ LMS 1-49%	>/=70%/ LMS 50%	n/a
Vertseylen	2013	n/a	n/a	n/a	n/a	n/a
Thoracic aortic calcium score						
Elias-Smale	2011	1st tertile (unadjusted & 2 adjustments)	2nd tertile (unadjusted & 2 adjustments)	3rd tertile (unadjusted & 2 adjustments)	n/a	n/a
Wong	2009	<10 for 3 different outcomes (hard CHD, total CHD & total CVD, for adjusted & unadjusted)	10-99 (hard CHD, adjusted & unadjusted)	100-399 (total CHD & total CVD, adjusted & unadjusted)	n/a	adjusted & unadjusted log for hard CHD, total CHD & total CVD
Yeboah	2014	n/a	n/a	n/a	n/a	Univariate & multivariate log for 3 different outcomes (incident CVD, CAD & all-cause mortality)
Abbreviations: CAC = coronary calcium score; CHD = coronary heart disease; CKD = chronic kidney disease;						
c-index = concordance index; CVD = cardiovascular disease; FRS = Framingham Risk Score;						
SD = standard deviation; HR = hazard ratio; LAD = left anterior descending; LMS = left main stem;						

8.2 Figure, Table & Appendix captions

Figure captions

Figure 14. Comparison between standard of care and gated scans: The difference in subjective score between chest CTs.

Figure 15. 71 year old man, BMI 30.3, received regular CT surveillance for prostate cancer, classed as never smoked. Images displayed in lung windows. (a) The standard of care scan showed edge blurring artefact affecting diagnostic confidence in the middle lobe. (b) The equivalent research scan showed no edge blurring artefact. The DLP of the standard of care scan was 515.7 mGy-cm compared to the DLP of the research scan was 270.3 mGy-cm.

Figure 16. 48 year old woman, BMI 34.9, received regular CT surveillance for endometrial cancer, classed as ex-smoker. Images displayed in lung windows. (a) The standard of care scan showed edge blurring artefact in the upper lobes. (b) The equivalent research scan showed no edge blurring artefact. The DLP of the standard of care scan was 452.2 mGy-cm compared to the DLP of the research scan was 180.2 mGy-cm.

Figure 17. 65 year old male, BMI 33.4, received regular CT surveillance for renal cancer, classed as ex-smoker. Images displayed in lung windows. (a) The standard of care scan showed double line artefact in the heart border affecting diagnostic confidence. (b) The equivalent research scan showed minor double line artefact. The DLP of the standard of care scan was 611.9 mGy-cm compared to the DLP of the research scan was 240.3 mGy-cm.

Figure 18. 81 year old man, BMI 31.8, received regular CT surveillance for rectal cancer, classed as ex-smoker and was in sinus rhythm (acquisition heart rate 41-44 bpm). A prospectively gated axial CT scan with diastolic triggering allowed evaluation of the LIMA (left image), venous to diagonal (middle image) and venous to right coronary artery grafts (right image). The DLP of the axial scan was 240.3mGy-cm compared to the DLP of the helical scan was 465.7 mGy-cm.

Figure 19. 58 year old woman, BMI 27.1, received regular CT surveillance for vulvar cancer had previous coronary investigations. (A) A prospectively gated axial CT scan with diastolic triggering (acquisition heart rate 58-62 bpm) had good image quality that enabled reconstruction with volume rendering of the coronary tree and stents within the left anterior descending and right coronary artery. (B & C) Curved multiplane reformat images allowed further scrutiny of the right coronary stent.

Figure 20. The correlation between body mass index and dose length product in both the helical (ungated) and axial (ECG gated) scans.

Table captions

Table 1. The characteristics of the 31 included studies

Table 2. The design and participant characteristics of the 9 included studies

Table 3. The vendor and technology of the scanners to determine CACS

Table 4. The reported outcomes of the 35 included studies

Table 5. Thresholds used for Framingham Risk Score

Table 6. Information regarding validation and calibration

Table 7. Alteration of the risk factors used for the calculation of Framingham Risk Score in 35 eligible studies compared to the Framingham Risk Score 1998, 2002 and 2008

Table 8. Selective reporting of association

Table 9. Documentation of multivariable regression, calibration, discrimination & reclassification

Appendix captions

8.3 References

1. Lloyd-Jones DM, Leip EP, Larson MG, D'Agostino RB, Beiser A, Wilson PW, et al. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. Circulation. 2006;113(6):791-8.
2. Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. Lancet. 1999;353(9147):89-92.
3. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97(18):1837-47.
4. D'Agostino RB, Sr., Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. Jama. 2001;286(2):180-7.
5. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016;353:i2416.
6. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012;9(5):1-12.
7. Erbel R, Mohlenkamp S, Moebus S, Schmermund A, Lehmann N, Stang A, et al. Coronary risk stratification, discrimination, and reclassification improvement based on quantification of subclinical coronary atherosclerosis: the Heinz Nixdorf Recall study. J Am Coll Cardiol. 2010;56(17):1397-406.
8. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 2002;106(25):3143-421.
9. D'Agostino RB, Sr., Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117(6):743-53.
10. Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. 2003;24(11):987-1003.
11. 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.
12. Collins GS, Omar O, Shanyinde M, Yu LM. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol. 2013;66(3):268-77.
13. Collins GS, Reitsma JB, Altman DG, Moons KM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): The tripod statement. Ann Intern Med. 2015;162(1):55-63.
14. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995;48(12):1503-10.
15. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.
16. Kavousi M, Elias-Smale S, Rutten JH, Leening MJ, Vliegenthart R, Verwoert GC, et al. Evaluation of newer risk markers for coronary heart disease risk classification: a cohort study. Ann Intern Med. 2012;156(6):438-44.
17. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. Circulation. 2009;119(17):2408-16.

18. Valenti V, B OH, Heo R, Cho I, Schulman-Marcus J, Gransar H, et al. A 15-Year Warranty Period for Asymptomatic Individuals Without Coronary Artery Calcium: A Prospective Follow-Up of 9,715 Individuals. JACC Cardiovascular imaging. 2015;8(8):900-9.
19. Oudkerk M, Stillman AE, Halliburton SS, Kalender WA, Möhlenkamp S, McCollough CH, et al. Coronary artery calcium screening: current status and recommendations from the European Society of Cardiac Radiology and North American Society for Cardiovascular Imaging. The International Journal of Cardiovascular Imaging. 2008;24(6):645-71.
20. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Jr., Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol. 1990;15(4):827-32.
21. Wong ND, Gransar H, Shaw L, Polk D, Moon JH, Miranda-Peats R, et al. Thoracic Aortic Calcium Versus Coronary Artery Calcium for the Prediction of Coronary Heart Disease and Cardiovascular Disease Events. JACC-Cardiovasc Imag. 2009;2(3):319-26.
22. Hadamitzky M, Achenbach S, Al-Mallah M, Berman D, Budoff M, Cademartiri F, et al. Optimized prognostic score for coronary computed tomographic angiography: results from the CONFIRM registry (COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter Registry). J Am Coll Cardiol. 2013;62(5):468-76.
23. Roberts WC, Jones AA. Quantitation of coronary arterial narrowing at necropsy in sudden coronary death. The American Journal of Cardiology. 1979;44(1):39-45.
24. Rumberger JA, Brundage BH, Rader DJ, Kondos G. Electron beam computed tomographic coronary calcium scanning: a review and guidelines for use in asymptomatic persons. Mayo Clin Proc. 1999;74(3):243-52.
25. Taylor F, Huffman MD, Macedo AF, Moore TH, Burke M, Davey Smith G, et al. Statins for the primary prevention of cardiovascular disease. The Cochrane database of systematic reviews. 2013(1):Cd004816.
26. Sanz J. Coronary Calcium Score and the New GuidelinesBack to Square One?∗. J Am Coll Cardiol. 2015;66(15):1669-71.
27. Waugh N, Black C, Walker S, McIntyre L, Cummins E, Hillis G. The effectiveness and cost-effectiveness of computed tomography screening for coronary artery disease: systematic review. Health Technol Assess. 2006;10(39):iii-iv, ix-x, 1-41.
28. Moons KG, van Es GA, Michel BC, Buller HR, Habbema JD, Grobbee DE. Redundancy of single diagnostic test evaluation. Epidemiology. 1999;10(3):276-81.
29. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem. 2004;50(3):473-6.
30. Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM. Quantifying the added value of a diagnostic test or marker. Clin Chem. 2012;58(10):1408-17.
31. Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. Jama. 2009;302(21):2345-52.
32. Ioannidis JP. Expectations, validity, and reality in omics. J Clin Epidemiol. 2010;63(9):945-9.
33. Ioannidis JPA. Limits to forecasting in personalized medicine: An overview. International Journal of Forecasting. 2009;25(4):773-83.
34. Moons KGM. Criteria for Scientific Evaluation of Novel Markers: A Perspective. Clinical Chemistry. 2010;56(4):537.
35. Bell KJL, Irwig L, Craig JC, Macaskill P. Use of randomised trials to decide when to monitor response to new treatment. BMJ. 2008;336(7640):361.
36. Lassere MN, Johnson KR, Boers M, Tugwell P, Brooks P, Simon L, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. The Journal of rheumatology. 2007;34(3):607-15.
37. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159(9):882-90.
38. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007;115(7):928-35.

39. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27(2):157-72; discussion 207-12.
40. Coppus SF, van der Veen F, Opmeer BC, Mol BW, Bossuyt PM. Evaluating prediction models in reproductive medicine. Human reproduction (Oxford, England). 2009;24(8):1774-8.
41. Pencina KM, Pencina MJ, D'Agostino RB, Sr. What to expect from net reclassification improvement with three categories. Stat Med. 2014;33(28):4975-87.
42. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology. 2014;25(1):114-21.
43. Pepe MS, Fan J, Feng Z, Gerds T, Hilden J. The Net Reclassification Index (NRI): A Misleading Measure of Prediction Improvement Even with Independent Test Data Sets. Statistics in Biosciences. 2015;7(2):282-95.
44. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. Am J Epidemiol. 2012;176(6):473-81.
45. Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. Eur J Clin Invest. 2012;42(2):216-28.
46. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55-63.
47. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. Am Heart J. 1991;121(1 Pt 2):293-8.
48. Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profile. A statement for health professionals. Circulation. 1991;83(1):356-62.
49. Yeboah J, Polonsky TS, Young R, McClelland RL, Delaney JC, Dawood F, et al. Utility of Non-Traditional Risk Markers in Individuals Ineligible for Statin Therapy According to the 2013 ACC/AHA Cholesterol Guidelines. Circulation. 2015.
50. Nasir K, Bittencourt MS, Blaha MJ, Blankstein R, Agatson AS, Rivera JJ, et al. Implications of Coronary Artery Calcium Testing Among Statin Candidates According to American College of Cardiology/American Heart Association Cholesterol Management Guidelines: MESA (Multi-Ethnic Study of Atherosclerosis). J Am Coll Cardiol. 2015;66(15):1657-68.
51. Abdulla J, Abildstrom SZ, Gotzsche O, Christensen E, Kober L, Torp-Pedersen C. 64-multislice detector computed tomography coronary angiography as potential alternative to conventional coronary angiography: a systematic review and meta-analysis. Eur Heart J. 2007;28(24):3042-50.
52. Hamon M, Biondi-Zoccai GG, Malagutti P, Agostoni P, Morello R, Valgimigli M, et al. Diagnostic performance of multislice spiral computed tomography of coronary arteries as compared with conventional invasive coronary angiography: a meta-analysis. J Am Coll Cardiol. 2006;48(9):1896-910.
53. Mowatt G, Cook JA, Hillis GS, Walker S, Fraser C, Jia X, et al. 64-Slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. Heart. 2008;94(11):1386-93.
54. Budoff MJ, Dowe D, Jollis JG, Gitter M, Sutherland J, Halamert E, et al. Diagnostic performance of 64-multidetector row coronary computed tomographic angiography for evaluation of coronary artery stenosis in individuals without known coronary artery disease: results from the prospective multicenter ACCURACY (Assessment by Coronary Computed Tomographic Angiography of Individuals Undergoing Invasive Coronary Angiography) trial. J Am Coll Cardiol. 2008;52(21):1724-32.
55. Meijboom WB, Meijs MF, Schuijf JD, Cramer MJ, Mollet NR, van Mieghem CA, et al. Diagnostic accuracy of 64-slice computed tomography coronary angiography: a prospective, multicenter, multivendor study. J Am Coll Cardiol. 2008;52(25):2135-44.

56. Miller JM, Rochitte CE, Dewey M, Arbab-Zadeh A, Niinuma H, Gottlieb I, et al. Diagnostic Performance of Coronary Angiography by 64-Row CT. New England Journal of Medicine. 2008;359(22):2324-36.
57. Roobottom C, Mitchell G, Iyengar S. The role of non-invasive imaging in patients with suspected acute coronary syndrome. The British Journal of Radiology. 2011;84(Spec Iss 3):S269-S79.
58. Menke J, Unterberg-Buchwald C, Staab W, Sohns JM, Seif Amir Hosseini A, Schwarz A. Head-to-head comparison of prospectively triggered vs retrospectively gated coronary computed tomography angiography: Meta-analysis of diagnostic accuracy, image quality, and radiation dose. Am Heart J. 2013;165(2):154-63.e3.
59. Achenbach S, Manolopoulos M, Schuhback A, Ropers D, Rixe J, Schneider C, et al. Influence of heart rate and phase of the cardiac cycle on the occurrence of motion artifact in dual-source CT angiography of the coronary arteries. J Cardiovasc Comput Tomogr. 2012;6(2):91-8.
60. Seifarth H, Wienbeck S, Pusken M, Juergens KU, Maintz D, Vahlhaus C, et al. Optimal systolic and diastolic reconstruction windows for coronary CT angiography using dual-source CT. AJR Am J Roentgenol. 2007;189(6):1317-23.
61. Shuman WP, Leipsic JA, Busey JM, Green DE, Pipavath SN, Hague CJ, et al. Prospectively ECG gated CT pulmonary angiography versus helical ungated CT pulmonary angiography: impact on cardiac related motion artifacts and patient radiation dose. Eur J Radiol. 2012;81(9):2444-9.
62. Boll DT, Gilkeson RC, Fleiter TR, Blackham KA, Duerk JL, Lewin JS. Volumetric assessment of pulmonary nodules with ECG-gated MDCT. AJR Am J Roentgenol. 2004;183(5):1217-23.
63. Schoepf UJ, Becker CR, Bruening RD, Helmberger T, Staebler A, Leimeister P, et al. Electrocardiographically gated thin-section CT of the lung. Radiology. 1999;212(3):649-54.
64. Marten K, Engelke C, Funke M, Obenauer S, Baum F, Grabbe E. ECG-gated multislice spiral CT for diagnosis of acute pulmonary embolism. Clinical radiology. 2003;58(11):862-8.
65. Boehm T, Willmann JK, Hilfiker PR, Weishaupt D, Seifert B, Crook DW, et al. Thin-Section CT of the Lung: Does Electrocardiographic Triggering Influence Diagnosis? Radiology. 2003;229(2):483-91.
66. Vikgren J, Bath M, Johnsson AA, Flinck A, Milde H, Thilander-Klang A, et al. High-resolution computed tomography with single-slice computed tomography and 16-channel multidetector computed tomography: a comparison regarding visibility and motion artifacts. Acta radiologica (Stockholm, Sweden : 1987). 2007;48(9):956-61.
67. Hofmann LK, Zou KH, Costello P, Schoepf UJ. Electrocardiographically gated 16-section CT of the thorax: cardiac motion suppression. Radiology. 2004;233(3):927-33.
68. Roos JE, Willmann JK, Weishaupt D, Lachat M, Marincek B, Hilfiker PR. Thoracic aorta: motion artifact reduction with retrospective and prospective electrocardiography-assisted multi-detector row CT. Radiology. 2002;222(1):271-7.
69. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. Bmj. 2009;338:b606.
70. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004;23(16):2567-86.
71. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD, et al. Third universal definition of myocardial infarction. J Am Coll Cardiol. 2012;60(16):1581-98.
72. Braunwald E, Cannon CP, McCabe CH. Use of composite endpoints in thrombolysis trials of acute myocardial infarction. Am J Cardiol. 1993;72(19):3g-12g.
73. Cannon CP. Clinical perspectives on the use of composite endpoints. Controlled clinical trials. 1997;18(6):517-29; discussion 46-9.
74. Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in end point selection for heart failure trials: composite end points. J Card Fail. 2005;11(8):567-75.

75. Lim E, Brown A, Helmy A, Mussa S, Altman DG. Composite outcomes in cardiovascular research: a survey of randomized trials. Ann Intern Med. 2008;149(9):612-7.
76. Gomez G, Gomez-Mateu M, Dafni U. Informed choice of composite end points in cardiovascular trials. Circulation Cardiovascular quality and outcomes. 2014;7(1):170-8.
77. Pitt B, Remme W, Zannad F, Neaton J, Martinez F, Roniker B, et al. Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. N Engl J Med. 2003;348(14):1309-21.
78. Yusuf S, Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJ, et al. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. Lancet. 2003;362(9386):777-81.
79. Elston J, Taylor RS. Use of surrogate outcomes in cost-effectiveness models: a review of United Kingdom health technology assessment reports. International journal of technology assessment in health care. 2009;25(1):6-13.
80. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharmaceutical statistics. 2006;5(3):173-86.
81. Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. The Lancet Oncology. 2006;7(9):741-6.
82. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.
83. Pang CL, Peters J, Hyde C, Roobottom C. The added value of computed tomography coronary angiogram in predicting future cardiovascular events in a low risk population: comparison with Framingham Risk Score. PROSPERO: International prospective register for systematic reviews. 2015:CRD42015023795.
84. British Library e-theses online service. http://ethos.bl.uk. Accessed September 03, 2015.
85. System for Information on Grey Literature in Europe. http://www.opengrey.eu. Accessed September 03, 2015.
86. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.
87. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net Reclassification Indices for Evaluating Risk-Prediction Instruments: A Critical Review. Epidemiology (Cambridge, Mass). 2014;25(1):114-21.
88. Pencina MJ, D'Agostino RB, Sr., Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med. 2012;31(2):101-13.
89. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011;30(1):11-21.
90. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744.
91. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. Circulation. 1998;97(18):1837-47.
92. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report. Circulation. 2002;106(25):3143.
93. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study. Circulation. 2008;117(6):743-53.

94. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148(3):839-43.
95. Altman DG, Bland JM. How to obtain the P value from a confidence interval. BMJ. 2011;343.
96. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280-6.
97. Wei Y, Royston P, Tierney JF, Parmar MK. Meta-analysis of time-to-event outcomes from randomized trials using restricted mean survival time: application to individual participant data. Stat Med. 2015;34(21):2881-98.
98. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. British Medical Journal (Clinical research ed). 1986;292(6522):746-50.
99. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340.
100. Altman DG, Bland JM. How to obtain the confidence interval from a P value. BMJ : British Medical Journal. 2011;343.
101. Higgins JPT, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.
102. Lau KK, Wong YK, Chan YH, Yiu KH, Teo KC, Li LS, et al. Prognostic implications of surrogate markers of atherosclerosis in low to intermediate risk patients with Type 2 Diabetes. Cardiovasc Diabetol. 2012;11(101).
103. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60.
104. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled clinical trials. 1986;7(3):177-88.
105. Ahmadi N, Hajsadeghi F, Blumenthal RS, Budoff MJ, Stone GW, Ebrahimi R. Mortality in Individuals Without Known Coronary Artery Disease but With Discordance Between the Framingham Risk Score and Coronary Artery Calcium. Am J Cardiol. 2011;107(6):799-804.
106. Budoff MJ, Shaw LJ, Liu ST, Weinstein SR, Mosler TP, Tseng PH, et al. Long-term prognosis associated with coronary calcification: observations from a registry of 25,253 patients. J Am Coll Cardiol. 2007;49(18):1860-70.
107. Raggi P, Shaw LJ, Berman DS, Callister TQ. Gender-based differences in the prognostic value of coronary calcification. J Womens Health. 2004;13(3):273-83.
108. Chang SM, Nabi F, Xu J, Pratt CM, Mahmarian AC, Frias ME, et al. Value of CACS Compared With ETT and Myocardial Perfusion Imaging for Predicting Long-Term Cardiac Outcome in Asymptomatic and Symptomatic Patients at Low Risk for Coronary Disease Clinical Implications in a Multimodality Imaging World. JACC-Cardiovasc Imag. 2015;8(2):134-44.
109. Elias-Smale SE, Proenca RV, Koller MT, Kavousi M, van Rooij FJ, Hunink MG, et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. J Am Coll Cardiol. 2010;56(17):1407-14.
110. Forouzandeh F, Chang SM, Muhyieddeen K, Zaid RR, Trevino AR, Xu J, et al. Does quantifying epicardial and intrathoracic fat with noncontrast computed tomography improve risk stratification beyond calcium scoring alone? Circ Cardiovasc Imaging. 2013;6(1):58-66.
111. Hadamitzky M, Meyer T, Hein F, Bischoff B, Martinoff S, Schomig A, et al. Prognostic value of coronary computed tomographic angiography in asymptomatic patients. Am J Cardiol. 2010;105(12):1746-51.
112. Elias-Smale SE, Wieberdink RG, Odink AE, Hofman A, Hunink MG, Koudstaal PJ, et al. Burden of atherosclerosis improves the prediction of coronary heart disease but not cerebrovascular events: the Rotterdam Study. Eur Heart J. 2011;32(16):2050-8.
113. Yeboah J, Carr JJ, Terry JG, Ding J, Zeb I, Liu S, et al. Computed tomography-derived cardiovascular risk markers, incident cardiovascular events, and all-cause mortality

in nondiabetics: The Multi-Ethnic Study of Atherosclerosis. Eur J Prev Cardiol. 2014;21(10):1233-41.
114. Greenland P, LaBree L, Azen SP, Doherty TM, Detrano RC. Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals.[Erratum appears in JAMA. 2004 Feb 4;291(5):563]. Jama. 2004;291(2):210-5.
115. Yeboah J, McClelland RL, Polonsky TS, Burke GL, Sibley CT, O'Leary D, et al. Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals. JAMA - Journal of the American Medical Association. 2012;308(8):788-95.
116. Matsushita K, Sang YY, Ballew SH, Shlipak M, Katz R, Rosas SE, et al. Subclinical Atherosclerosis Measures for Cardiovascular Prediction in CKD. J Am Soc Nephrol. 2015;26(2):439-47.
117. Mohlenkamp S, Lehmann N, Greenland P, Moebus S, Kalsch H, Schmermund A, et al. Coronary artery calcium score improves cardiovascular risk prediction in persons without indication for statin therapy. Atherosclerosis. 2011;215(1):229-36.
118. Mohlenkamp S, Lehmann N, Moebus S, Schmermund A, Dragano N, Stang A, et al. Quantification of coronary atherosclerosis and inflammation to predict coronary events and all-cause mortality. J Am Coll Cardiol. 2011;57(13):1455-64.
119. Polonsky TS, McClelland RL, Jorgensen NW, Bild DE, Burke GL, Guerci AD, et al. Coronary artery calcium score and risk classification for coronary heart disease prediction. Jama. 2010;303(16):1610-6.
120. Raggi P, Cooil B, Callister TQ. Use of electron beam tomography data to develop models for prediction of hard coronary events. Am Heart J. 2001;141(3):375-82.
121. Rana JS, Gransar H, Wong ND, Shaw L, Pencina M, Nasir K, et al. Comparative Value of Coronary Artery Calcium and Multiple Blood Biomarkers for Prognostication of Cardiovascular Events. Am J Cardiol. 2012;109(10):1449-53.
122. Agarwal S, Cox AJ, Herrington DM, Jorgensen NW, Xu J, Freedman BI, et al. Coronary calcium score predicts cardiovascular mortality in diabetes: Diabetes heart study. Diabetes Care. 2013;36(4):972-7.
123. Arad Y, Goodman KJ, Roth M, Newstein D, Guerci AD. Coronary calcification, coronary disease risk factors, C-reactive protein, and atherosclerotic cardiovascular disease events: The St. Francis heart study. J Am Coll Cardiol. 2005;46(1):158-65.
124. Cho I, Chang HJ, Sung JM, Pencina MJ, Lin FY, Dunning AM, et al. Coronary computed tomographic angiography and risk of all-cause mortality and nonfatal myocardial infarction in subjects without chest pain syndrome from the CONFIRM Registry (coronary CT angiography evaluation for clinical outcomes: an international multicenter registry). Circulation. 2012;126(3):304-13.
125. Versteylen MO, Kietselaer BL, Dagnelie PC, Joosen IA, Dedic A, Raaijmakers RH, et al. Additive Value of Semiautomated Quantification of Coronary Artery Disease Using Cardiac Computed Tomographic Angiography to Predict Future Acute Coronary Syndrome. J Am Coll Cardiol. 2013;61(22):2296-305.
126. Gibson AO, Blaha MJ, Arnan MK, Sacco RL, Szklo M, Herrington DM, et al. Coronary Artery Calcium and Incident Cerebrovascular Events in an Asymptomatic Cohort The MESA Study. JACC-Cardiovasc Imag. 2014;7(11):1108-15.
127. Hermann DM, Gronewold J, Lehmann N, Moebus S, Jockel KH, Bauer M, et al. Coronary artery calcification is an independent stroke predictor in the general population. Stroke. 2013;44(4):1008-13.
128. Chow BJ, Small G, Yam Y, Chen L, Achenbach S, Al-Mallah M, et al. Incremental prognostic value of cardiac computed tomography in coronary artery disease using CONFIRM: COroNary computed tomography angiography evaluation for clinical outcomes: an InteRnational Multicenter registry. Circ Cardiovasc Imaging. 2011;4(5):463-72.
129. Lin FY, Shaw LJ, Dunning AM, LaBounty TM, Choi JH, Weinsaft JW, et al. Mortality Risk in Symptomatic Patients With Nonobstructive Coronary Artery Disease A Prospective 2-Center Study of 2,583 Patients Undergoing 64-Detector Row Coronary Computed Tomographic Angiography. J Am Coll Cardiol. 2011;58(5):510-9.

130. Chow BJ, Wells GA, Chen L, Yam Y, Galiwango P, Abraham A, et al. Prognostic value of 64-slice cardiac computed tomography severity of coronary artery disease, coronary atherosclerosis, and left ventricular ejection fraction. J Am Coll Cardiol. 2010;55(10):1017-28.
131. Park HE, Chun EJ, Choi SI, Lee SP, Yoon CH, Kim HK, et al. Clinical and imaging parameters to predict cardiovascular outcome in asymptomatic subjects Int J Cardiovasc Imaging. 2013;29(7):1595-602.
132. Cho I, Chang HJ, Hartaigh BO, Shin S, Sung JM, Lin FY, et al. Incremental prognostic utility of coronary CT angiography for asymptomatic patients based upon extent and severity of coronary artery calcium: Results from the COronary CT Angiography EvaluatioN For Clinical Outcomes InteRnational Multicenter (CONFIRM) Study. Eur Heart J. 2015;36(8):501-8.
133. Han D, B OH, Gransar H, Yoon JH, Kim KJ, Kim MK, et al. Incremental benefit of coronary artery calcium score above traditional risk factors for all-cause mortality in asymptomatic Korean adults. Circulation journal : official journal of the Japanese Circulation Society. 2015;79(11):2445-51.
134. Han DH, Yoon JH, Kim KJ, Kim MK, Choi SY, Sung JD, et al. Incremental benefit of coronary artery calcium score over traditional risk factors for all-cause mortality in asian population from KOICA registry (Korea initiatives on coronary artery calcification: Multicenter registry). J Am Coll Cardiol. 2015;1):A1065.
135. Matsushita K, Sang Y, Ballew SH, Shlipak M, Katz R, Rosas SE, et al. Subclinical atherosclerosis measures for cardiovascular prediction in CKD. J Am Soc Nephrol. 2015;26(2):439-47.
136. Steyerberg EW, Pencina MJ. Reclassification calculations for persons with incomplete follow-up. Ann Intern Med. 2010;152(3):195-6; author reply 6-7.
137. Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. Biostatistics. 2005;6(2):227-39.
138. Leening MJG, Vedder MM, Witteman JCM, Pencina MJ, Steyerberg EW. Net Reclassification Improvement: Computation, Interpretation, and ControversiesA Literature Review and Clinician's Guide. Ann Intern Med. 2014;160(2):122-31.
139. Baker SG, Schuit E, Steyerberg EW, Pencina MJ, Vickers A, Moons KG, et al. How to interpret a small increase in AUC with an additional risk prediction marker: decision analysis comes through. Stat Med. 2014;33(22):3946-59.
140. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology. 2012;12(1):9.
141. Ferreira-González I, Permanyer-Miralda G, Domingo-Salvany A, Busse JW, Heels-Ansdell D, Montori VM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ. 2007;334(7597):786.
142. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem. 2008;54(4):729-37.
143. Raggi P, Gongora MC, Gopal A, Callister TQ, Budoff M, Shaw LJ. Coronary artery calcium to predict all-cause mortality in elderly men and women. J Am Coll Cardiol. 2008;52(1):17-23.
144. Kavousi M, Elias-Smale S, Rutten JH, Leening MJ, Vliegenthart R, Verwoert GC, et al. Evaluation of newer risk markers for coronary heart disease risk classification: a cohort study. Ann Intern Med. 2012;156(6):438-44.
145. Muhlenbruch K, Heraclides A, Steyerberg EW, Joost HG, Boeing H, Schulze MB. Assessing improvement in disease prediction using net reclassification improvement: impact of risk cut-offs and number of risk categories. European journal of epidemiology. 2013;28(1):25-33.
146. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. Journal of the National Cancer Institute. 2005;97(14):1043-55.

147. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report statistically significant results. European journal of cancer (Oxford, England : 1990). 2007;43(17):2559-79.
148. Pencina MJ, D' Agostino RB, D' Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Statistics in Medicine. 2008;27(2):157-72.
149. Tzoulaki I, Liberopoulos G, Ioannidis JP. Use of reclassification for assessment of improved prediction: an empirical evaluation. International journal of epidemiology. 2011;40(4):1094-105.
150. Pepe MS, Feng Z, Gu JW. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med. 2008;27(2):173-81.
151. Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index. Stat Med. 2014;33(19):3405-14.
152. Cox DR. Two Further Applications of a Model for Binary Regression. Biometrika. 1958;45(3/4):562-5.
153. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Medical decision making : an international journal of the Society for Medical Decision Making. 2006;26(6):565-74.
154. Janssens ACJW, Khoury MJ. Assessment of Improved Prediction Beyond Traditional Risk Factors. Circulation: Cardiovascular Genetics. 2010;3(1):3.
155. Bero L, Oostvogel F, Bacchetti P, Lee K. Factors Associated with Findings of Published Trials of Drug–Drug Comparisons: Why Some Statins Appear More Efficacious than Others. PLOS Medicine. 2007;4(6):e184.
156. Ioannidis JP, Tzoulaki I. What makes a good predictor?: the evidence applied to coronary artery calcium score. Jama. 2010;303(16):1646-7.
157. Wilson PW. Challenges to improve coronary heart disease risk assessment. Jama. 2009;302(21):2369-70.
158. Mihaescu R, van Zitteren M, van Hoek M, Sijbrands EJ, Uitterlinden AG, Witteman JC, et al. Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. Am J Epidemiol. 2010;172(3):353-61.
159. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. BMJ : British Medical Journal. 2013;346.
160. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. PLOS Medicine. 2013;10(2):e1001380.
161. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLOS Medicine. 2013;10(2):e1001381.
162. Hingorani AD, Windt DAvd, Riley RD, Abrams K, Moons KGM, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. BMJ : British Medical Journal. 2013;346.
163. Coebergh JW, Janssen-Heijnen ML, Post PN, Razenberg PP. Serious co-morbidity among unselected cancer patients newly diagnosed in the southeastern part of The Netherlands in 1993-1996. J Clin Epidemiol. 1999;52(12):1131-6.
164. Meinardi MT, Gietema JA, van der Graaf WTA, van Veldhuisen DJ, Runne MA, Sluiter WJ, et al. Cardiovascular Morbidity in Long-Term Survivors of Metastatic Testicular Cancer. Journal of Clinical Oncology. 2000;18(8):1725-32.
165. Clarke M, Collins R, Darby S, Davies C, Elphinstone P, Evans V, et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;366(9503):2087-106.

166. Bovelli D, Plataniotis G, Roila F. Cardiotoxicity of chemotherapeutic agents and radiotherapy-related heart disease: ESMO Clinical Practice Guidelines. Annals of oncology : official journal of the European Society for Medical Oncology. 2010;21 Suppl 5:v277-82.
167. Naschitz JE, Yeshurun D, Abrahamson J, Eldar S, Chouri H, Kedar S, et al. Ischemic heart disease precipitated by occult cancer. Cancer. 1992;69(11):2712-20.
168. Zoller B, Ji J, Sundquist J, Sundquist K. Risk of coronary heart disease in patients with cancer: a nationwide follow-up study from Sweden. European journal of cancer (Oxford, England : 1990). 2012;48(1):121-8.
169. Dreyer L, Olsen JH. Cancer risk of patients discharged with acute myocardial infarct. Epidemiology. 1998;9(2):178-83.
170. Roobottom CA, Mitchell G, Morgan-Hughes G. Radiation-reduction strategies in cardiac computed tomographic angiography. Clinical radiology. 2010;65(11):859-67.
171. Chest pain of recent onset: assessment and diagnosis. National Insitute of Clinical Excellence. 2010 (updated November 2016).
172. Williams MC, Hunter A, Shah AS, Assi V, Lewis S, Smith J, et al. Use of Coronary Computed Tomographic Angiography to Guide Management of Patients With Coronary Disease. J Am Coll Cardiol. 2016;67(15):1759-68.
173. Clayton B, Raju V, Roobottom C, Morgan-Hughes G. Safety of intravenous β-adrenoceptor blockers for computed tomographic coronary angiography. British Journal of Clinical Pharmacology. 2015;79(3):533-6.
174. Hausleiter J, Meyer T, Hermann F, et al. EStimated radiation dose associated with cardiac ct angiography. Jama. 2009;301(5):500-7.
175. Austen WG, Edwards JE, Frye RL, Gensini GG, Gott VL, Griffith LS, 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.
176. Brodoefel H, Reimann A, Burgstahler C, Schumacher F, Herberts T, Tsiflikas I, et al. Noninvasive coronary angiography using 64-slice spiral computed tomography in an unselected patient collective: effect of heart rate, heart rate variability and coronary calcifications on image quality and diagnostic accuracy. Eur J Radiol. 2008;66(1):134-41.
177. Matt D, Scheffel H, Leschka S, Flohr TG, Marincek B, Kaufmann PA, et al. Dual-source CT coronary angiography: image quality, mean heart rate, and heart rate variability. AJR Am J Roentgenol. 2007;189(3):567-73.
178. Vardhanabhuti V, Loader RJ, Mitchell GR, Riordan RD, Roobottom CA. Image quality assessment of standard- and low-dose chest CT using filtered back projection, adaptive statistical iterative reconstruction, and novel model-based iterative reconstruction algorithms. AJR Am J Roentgenol. 2013;200(3):545-52.
179. Williams MC, Golay SK, Hunter A, Weir-McCall JR, Mlynska L, Dweck MR, et al. Observer variability in the assessment of CT coronary angiography and coronary artery calcium score: substudy of the Scottish COmputed Tomography of the HEART (SCOT-HEART) trial. Open heart. 2015;2(1):e000234.
180. Gosling O, Morgan-Hughes G, Iyengar S, Strain W, Loader R, Shore A, et al. Computed tomography to diagnose coronary artery disease: a reduction in radiation dose increases applicability. Clinical radiology. 2013;68(4):340-5.
181. Loader RJ, Gosling O, Roobottom C, Morgan-Hughes G, Rowles N. Practical dosimetry methods for the determination of effective skin and breast dose for a modern CT system, incorporating partial irradiation and prospective cardiac gating. Br J Radiol. 2012;85(1011):237-48.
182. Gosling O, Loader R, Venables P, Rowles N, Morgan-Hughes G, Roobottom C. Cardiac CT: are we underestimating the dose? A radiation dose study utilizing the 2007 ICRP tissue weighting factors and a cardiac specific scan volume. Clinical radiology. 2010;65(12):1013-7.
183. Gosling O, Loader R, Venables P, Roobottom C, Rowles N, Bellenger N, et al. A comparison of radiation doses between state-of-the-art multislice CT coronary angiography

with iterative reconstruction, multislice CT coronary angiography with standard filtered back-projection and invasive diagnostic coronary angiography. Heart. 2010;96(12):922-6.
184. Huda W, Magill D, He W. CT effective dose per dose length product using ICRP 103 weighting factors. Medical physics. 2011;38(3):1261-5.
185. Ashby D. Practical statistics for medical research. Douglas G. Altman, Chapman and Hall, London, 1991. No. of pages: 611. Price: £32.00. Statistics in Medicine. 1991;10(10):1635-6.
186. Ardley ND, Lau KK, Buchan K, Paul E, Troupis JM. Effects of electrocardiogram gating on CT pulmonary angiography image quality. Journal of Medical Imaging and Radiation Oncology. 2014;58(3):303-11.
187. Min JK, Shaw LJ, Devereux RB, Okin PM, Weinsaft JW, Russo DJ, et al. Prognostic value of multidetector coronary computed tomographic angiography for prediction of all-cause mortality. J Am Coll Cardiol. 2007;50(12):1161-70.
188. Cheruvu C, Precious B, Naoum C, Blanke P, Ahmadi A, Soon J, et al. Long term prognostic utility of coronary CT angiography in patients with no modifiable coronary artery disease risk factors: Results from the 5 year follow-up of the CONFIRM International Multicenter Registry. J Cardiovasc Comput Tomogr. 2016;10(1):22-7.
189. Mushtaq S, De Araujo Goncalves P, Garcia-Garcia HM, Pontone G, Bartorelli AL, Bertella E, et al. Long-term prognostic effect of coronary atherosclerotic burden: validation of the computed tomography-Leaman score. Circ Cardiovasc Imaging. 2015;8(2):e002332.
190. Rinde LB, Smabrekke B, Hald EM, Brodin EE, Njolstad I, Mathiesen EB, et al. Myocardial infarction and future risk of cancer in the general population-the Tromso Study. European journal of epidemiology. 2017.
191. Ford MK, Beattie WS, Wijeysundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. Ann Intern Med. 2010;152(1):26-35.
192. Gualandro DM, Campos CA, Calderaro D, Yu PC, Marques AC, Pastana AF, et al. Coronary plaque rupture in patients with myocardial infarction after noncardiac surgery: frequent and dangerous. Atherosclerosis. 2012;222(1):191-5.
193. Dunkelgrun M, Boersma E, Schouten O, Koopman-van Gemert AW, van Poorten F, Bax JJ, et al. Bisoprolol and fluvastatin for the reduction of perioperative cardiac mortality and myocardial infarction in intermediate-risk patients undergoing noncardiovascular surgery: a randomized controlled trial (DECREASE-IV). Annals of surgery. 2009;249(6):921-6.
194. Kristensen SD, Knuuti J, Saraste A, Anker S, Botker HE, De Hert S, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). European journal of anaesthesiology. 2014;31(10):517-73.
195. Fleisher LA, Fleischmann KE, Auerbach AD, Barnason SA, Beckman JA, Bozkurt B, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. J Am Coll Cardiol. 2014;64(22):e77-137.
196. Poldermans D, Bax JJ, Boersma E, De Hert S, Eeckhout E, Fowkes G, et al. Guidelines for pre-operative cardiac risk assessment and perioperative cardiac management in non-cardiac surgery. Eur Heart J. 2009;30(22):2769-812.
197. A C, D L, J T. The Goldman and Detsky cardiac-risk indices: Do they work in patients undergoing hip-fracture surgery? . Ann R Coll Physicians Surg Can. 1999(32):337-41.
198. Sheth T, Chan M, Butler C, Chow B, Tandon V, Nagele P, et al. Prognostic capabilities of coronary computed tomographic angiography before non-cardiac surgery: prospective cohort study. BMJ. 2015;350.
199. Detrano RC, Anderson M, Nelson J, Wong ND, Carr JJ, McNitt-Gray M, et al. Coronary calcium measurements: effect of CT scanner type and calcium measure on rescan reproducibility--MESA study. Radiology. 2005;236(2):477-84.

200. Taylor AJ, Bindeman J, Feuerstein I, Cao F, Brazaitis M, O'Malley PG. Coronary calcium independently predicts incident premature coronary heart disease over measured cardiovascular risk factors: mean three-year outcomes in the Prospective Army Coronary Calcium (PACC) project. J Am Coll Cardiol. 2005;46(5):807-14.
201. van Kempen BJ, Spronk S, Koller MT, Elias-Smale SE, Fleischmann KE, Ikram MA, et al. Comparative effectiveness and cost-effectiveness of computed tomography screening for coronary artery calcium in asymptomatic individuals. J Am Coll Cardiol. 2011;58(16):1690-701.
202. Budoff MJ, Nasir K, McClelland RL, Detrano R, Wong N, Blumenthal RS, et al. Coronary calcium predicts events better with absolute calcium scores than age-sex-race/ethnicity percentiles: MESA (Multi-Ethnic Study of Atherosclerosis). J Am Coll Cardiol. 2009;53(4):345-52.
203. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. N Engl J Med. 2008;358(13):1336-45.
204. Einstein AJ, Henzlova MJ, Rajagopalan S. Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. Jama. 2007;298(3):317-23.
205. Hecht HS. Coronary Artery Calcium Scanning Past, Present, and Future. JACC-Cardiovasc Imag. 2015;8(5):580-96.
206. Machaalany J, Yam Y, Ruddy TD, Abraham A, Chen L, Beanlands RS, et al. Potential clinical and economic consequences of noncardiac incidental findings on cardiac computed tomography. J Am Coll Cardiol. 2009;54(16):1533-41.
207. Arad Y, Spadaro LA, Roth M, Newstein D, Guerci AD. Treatment of asymptomatic adults with elevated coronary calcium scores with atorvastatin, vitamin C, and vitamin E: the St. Francis Heart Study randomized clinical trial. J Am Coll Cardiol. 2005;46(1):166-72.
208. Ashton HA, Buxton MJ, Day NE, Kim LG, Marteau TM, Scott RA, et al. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. Lancet. 2002;360(9345):1531-9.