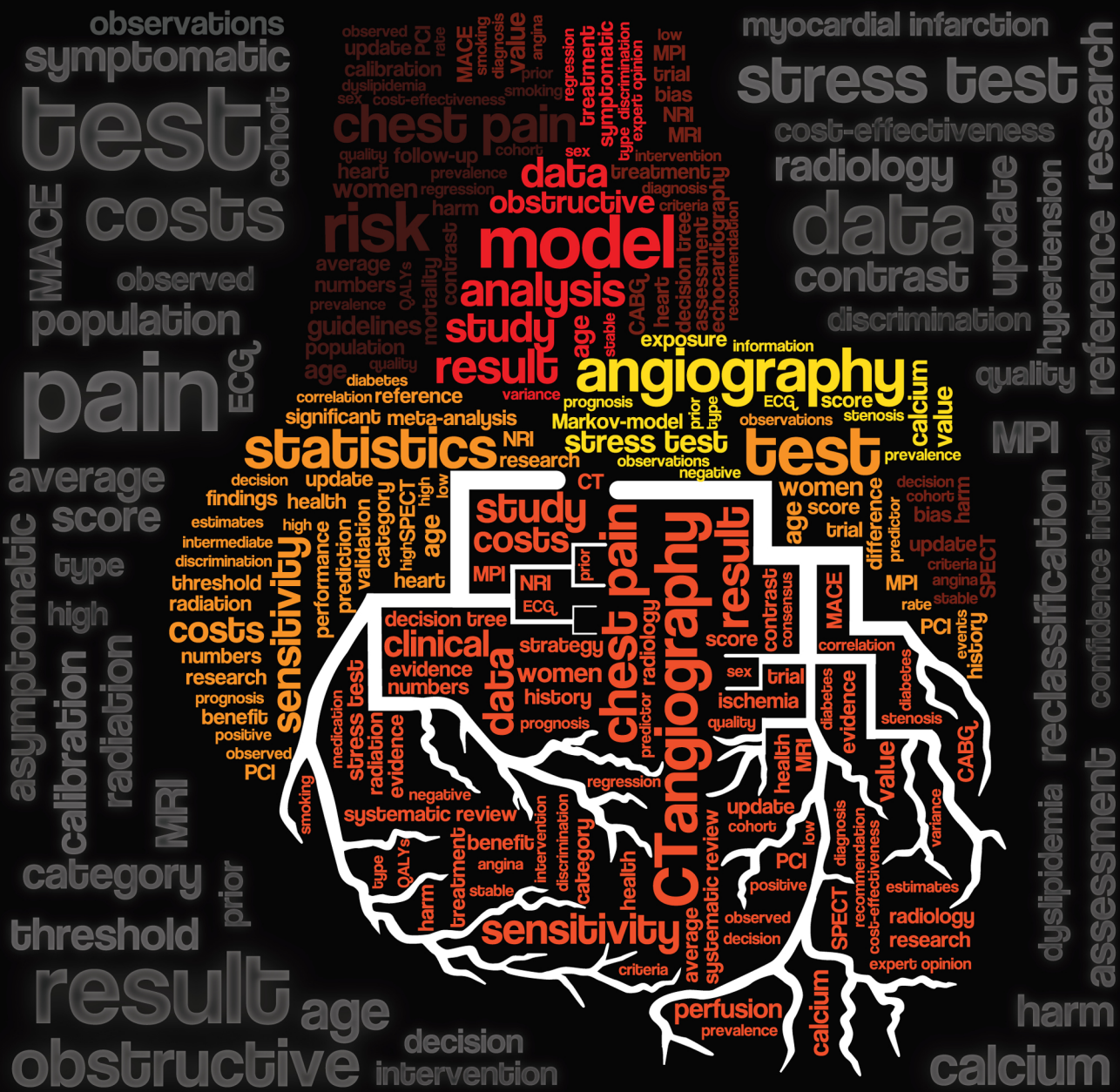
Diagnostic Imaging Strategies for Patients with Suspected Coronary Artery Disease



Tessa Genders

Diagnostic Imaging Strategies for Patients with Suspected Coronary Artery Disease

Tessa Sarah Suzanne Genders

Diagnostic Imaging Strategies for Patients with Suspected Coronary Artery Disease

Beeldvormende diagnostiek bij patiënten met mogelijk coronair lijden

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam op gezag van de Rector Magnificus

Prof. dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 29 juni 2012 om 9.30 uur

door

Tessa Sarah Suzanne Genders
geboren te Maastricht, Nederland

ERASMUS UNIVERSITEIT ROTTERDAM

Financial support for printing of this thesis was generously provided by the Departments of Radiology and Epidemiology of the Erasmus University Medical Center, Cardialysis, the J.E. Jurriaanse Stichting, and the Erasmus University Rotterdam.

Cover design: Summer Wood

Layout: A.W. Everaers

Printed by: Ipskamp Drukkers BV

ISBN: 978-90-9026566-7

Promotiecommissie

Promotor: Prof. dr. M.G.M. Hunink

Overige leden: Prof. dr. E.W. Steyerberg
Prof. dr. G.P. Krestin
Prof. dr. K.E. Fleischmann

Copromotor: Dr. K. Nieman

Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

Contents

Part I

Preface

Chapter 1

Introduction and Outline

Introduction

Coronary artery disease

The coronary arteries are the vessels responsible for supplying the heart muscle (myocardium) with blood and nutrients. The contraction of the myocardium depends on an adequate blood supply through the coronary circulation. Coronary artery disease (CAD) refers to coronary arteries affected by atherosclerosis, a slow but progressive pathological process that involves thickening of arterial walls. Arterial wall thickening (and subsequently lumen stenosis) is caused by formation of atherosclerotic plaque that consists of fatty materials with or without calcium depositions. Obstruction of the blood flow can impede a required increase in coronary blood flow during increased oxygen demands (e.g. during exercise), resulting in myocardial ischemia, dysfunction, and ultimately chest pain (i.e. "angina") (Figure 1). Advanced atherosclerotic lesions carry the risk of rupture and thrombus formation, resulting in an abrupt obstruction of a coronary artery – a life-threatening situation of an acute myocardial infarction (1).

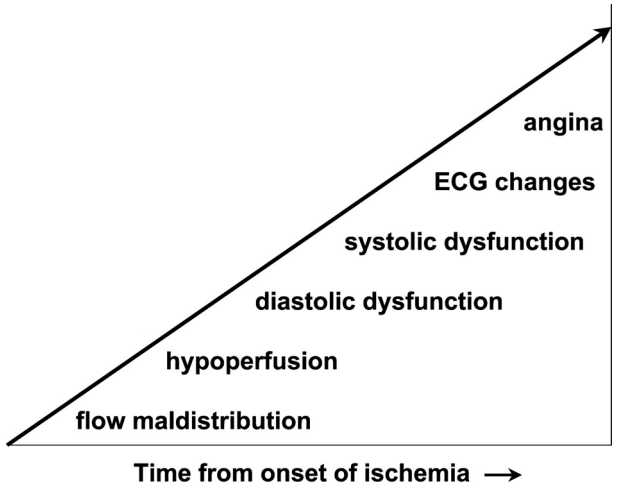


Figure 1. Schematic representation of the ischemic cascade. ECG = electrocardiography.

Coronary artery disease is a common condition, its prevalence (and incidence) increases with age. The lifetime risk at age 40 of developing CAD is estimated to be 49% in men, and 32% in women (2). Major risk factors for cardiovascular disease include hypertension, diabetes, dyslipidaemia, smoking, and obesity (3). CAD may start to develop early in life and usually remains subclinical (asymptomatic) until the 5th decade of life or longer. This relates to the fact that up to 50% of acute myocardial infarctions occur in patients without a history of symptomatic CAD (4). See Chapter 2 on the epidemiology of CAD.

Severity of CAD is often classified by the degree (percentage) of lumen diameter stenosis as measured by catheter-based coronary angiography (CAG). Commonly, ≥50% lumen diameter stenosis is considered 'significant' or 'obstructive', whereas a stenosis <50% is considered non-obstructive. However, an obstructive lesion does not always limit the coronary blood flow, and subsequently does not always result in myocardial ischemia. Therefore, not merely the degree of stenosis on angiography, but also the 'hemodynamic relevance' of a stenosis is a determinant of CAD severity.

Clinical presentation

Patients with stable CAD usually present to their physicians with chest pain on exertion, which is commonly referred to as "angina pectoris". Traditionally, angina pectoris is classified as typical, atypical, or non-anginal (or non-specific) chest pain. Typical chest pain is

defined as having (1) substernal chest pain or discomfort, that is (2) provoked by exertion or emotional stress and (3) relieved by rest and/or nitroglycerine (usually within minutes). Atypical chest pain is defined as having two of the before-mentioned criteria. If one or none of the criteria is present, a patient is classified as having non-specific chest pain. Patients often characterize their pain as a pressure, heaviness, tightness, or squeezing, and radiation may occur to the left jaw, shoulder, arm, and back. Patients with unstable CAD (acute coronary syndrome) presenting to the emergency department often present with persistent chest pain at rest.

Treatment for CAD

Treatment for CAD aims at relieving symptoms, preventing progression of atherosclerosis, and preventing plaque rupture (i.e. myocardial infarction) (5). Established treatments for the prevention of adverse events in patients with stable CAD include aspirin, statins, and ACE-inhibitors. Additionally, anti-ischemic therapies including long-acting nitrates, □ -blockers, and calcium antagonists can be considered for symptomatic relief (5). In patients with severe coronary disease (left main-, proximal left anterior descending-, or multi-vessel-disease) and documented myocardial ischemia, percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG) surgery are well-established treatments that improve quality-of-life and prevent future adverse events (6).

Pre-test probability of CAD

Since CAD is a serious condition, patients presenting with chest pain suggestive of CAD should be investigated. However, a definitive diagnosis cannot be made based on clinical presentation alone. A substantial proportion of patients presenting with chest pain does not suffer from CAD. The pre-test probability of CAD reflects the likelihood of the presence of obstructive CAD based on clinical presentation before a diagnostic test is performed. The pre-test probability is important when choosing which diagnostic test is optimal (see below) and interpreting diagnostic test results. Traditionally, the probability of CAD is based on age, sex, and type of chest pain according to the prediction model by Diamond & Forrester (7). Other models are available that also incorporate cardiovascular risk factors and electrocardiography findings for estimation of the pre-test probability (8-9).

Table 1. Diagnostic tests for the detection of coronary artery disease in patients with stable chest pain

Test	Detection of	Sensitivity (%)	Specificity (%)	Source
Exercise ECG	Ischemic ECG-changes (in women)	61	70	(11)
	Ischemic ECG-changes (without work-up bias)	50	90	(12)
Stress MRI	Perfusion defects	89	80	(13)
	Wall motion abnormalities	83	86	(14)
Stress SPECT	Perfusion defects / wall motion abnormalities	88	73	(15)
Stress ECHO	Wall motion abnormalities	79	87	(15)
Stress PET	Perfusion defects	92	85	(16)
CT coronary calcium	Presence of coronary calcium	98	40	(17)
Coronary CT angiography	≥50% lumen diameter stenosis	98	90	(18-21)

ECG = stress electrocardiography, MRI = magnetic resonance imaging, SPECT = single-photon emission computed tomography, PET = positron-emission computed tomography, CT = computed tomography.

Diagnosing CAD

For diagnosing obstructive CAD (commonly defined as ≥50% stenosis in ≥1 vessel), CAG is considered to be the reference standard. Because CAG is expensive and carries a small risk of complications and death, non-invasive testing is recommended to select patients who are most likely to benefit from CAG. The diagnostic performance of a non-invasive test can be expressed in terms of sensitivity and specificity as compared with the reference standard. Sensitivity is defined as the proportion of patients with the disease that will have a positive test, and specificity is defined as the proportion of patients without the disease that will have a negative test. Numerous non-invasive (imaging) tests are available for the diagnostic work-up of patients with stable chest pain (Table 1).

It is important to note that the different diagnostic modalities assess different manifestations of myocardial ischemia (Figure 1). For example, calcium can be deposited in atherosclerotic plaques, which may already be detectable in early (non-obstructive) stages of CAD. This explains that the presence of coronary calcium is very sensitive. Almost all patients with obstructive CAD have at lease some coronary calcium depositions. However, the presence of calcium does not necessarily imply that atherosclerotic plaque is obstructing ≥50% of the coronary lumen, hence its low specificity. Stress echocardiography on the other hand, images wall motion abnormalities, which occurs later in the cascade, and is therefore more specific. Lastly, exercise ECG investigates the events at the far end of the ischemic cascade, namely ischemic ECG-changes and angina during exercise. Consequently, the sensitivity of exercise ECG is low, whereas its specificity is high.

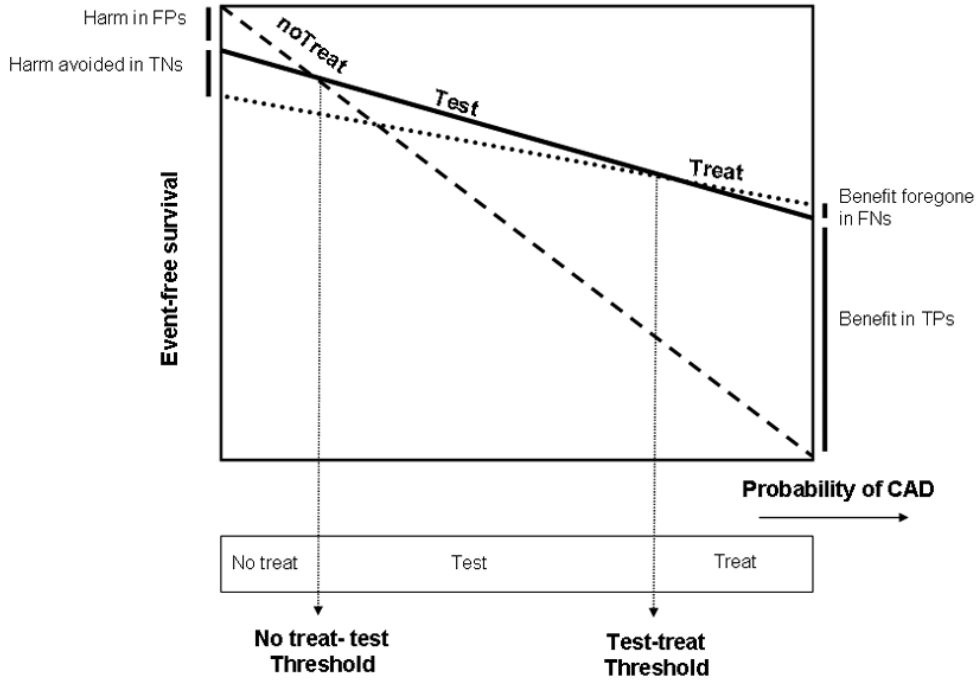


Figure 2. No treat – test – treat thresholds. The benefit of testing is in patients with an intermediate probability of coronary artery disease. The thresholds for diagnostic testing depend on the probability of CAD, test sensitivity, test specificity, and the benefits and harms of treating CAD. FP = false positive, TN = true negative, FN = false negative, TP = true positive.

Although test characteristics (sensitivity and specificity) indicate whether a test is accurate, the optimal diagnostic test for a particular patient also depends on the pre-test probability of disease in that patient, the potential benefit by making a correct diagnosis, and the potential harm caused by false positive test results (Figure 2) (10). The net benefit of a treatment is defined as the difference in outcome between a patient *with CAD* who receives the treatment, compared with the outcome of a patient *with CAD* without that treatment. The net harm, likewise, is defined as the difference in outcome between a patient *without CAD* without the treatment, compared with the outcome of a patient *without CAD* who receives the treatment (10).

If the pre-test probability is low, watchful waiting without any testing would be preferable, whereas if the pre-test probability is high, a direct invasive strategy with CAG would be optimal. The main benefit of non-invasive testing is in patients with an intermediate pre-test probability, in whom a negative test rules out the presence of coronary artery disease, whereas a positive test justifies treatment or further testing (10, 22), see Figure 2.

Outline of this thesis

This thesis aims to determine the optimal diagnostic strategy for patients suspected of having CAD. By performing systematic reviews and meta-analyses, we summarized the evidence concerning the *diagnostic accuracy* of diagnostic tests. Prediction models were validated, updated, and extended to improve the estimate of the *pre-test probability of CAD*. Finally, to estimate the long-term cost-effectiveness of various diagnostic strategies, we performed decision analyses that integrate the best-available evidence regarding the *benefits and harms* of treatment for CAD.

Chapter 2 describes the epidemiology of CAD, and its manifestations; angina pectoris and myocardial infarction.

In **Chapter 3**, we systematically review the literature for published guidelines, summarizing available recommendations regarding diagnostic strategies (e.g. CT calcium screening) for asymptomatic individuals.

Chapter 4 summarizes the available evidence on the diagnostic accuracy of stress myocardial perfusion imaging as compared with CAG. A systematic search strategy was used to identify studies that assessed magnetic resonance imaging, echocardiography, single-photon emission CT, or positron-emission tomography.

Management of patients presenting to the emergency department with acute chest pain – without diagnostic ECG-changes and without elevated cardiac biomarkers – is challenging, because the diagnosis of acute myocardial infarction cannot be excluded based on clinical findings alone. In **Chapter 5** we summarize the evidence on different imaging modalities for patients with acute chest pain presenting to the emergency department.

CT coronary calcium has been demonstrated to predict the occurrence of future cardiovascular events. In **Chapter 6** we explore the incremental diagnostic value of the CT coronary calcium score in the prediction of the presence of obstructive CAD for patients with stable chest pain.

Although multiple risk factors for CAD have been identified, the estimation of the pre-test probability in patients with stable chest pain is often based on age, sex, and symptoms only, according to the prediction model described by Diamond & Forrester in 1979. We assess the validity of the Diamond & Forrester model in **Chapter 7**, and update the model using modern statistical methods and individual patient data from 14 hospitals.

In **Chapter 8**, we extend the existing prediction model based on age, sex, and symptoms with established cardiovascular risk factors such as diabetes, hypertension, dyslipidemia and smoking. Furthermore, we assess the incremental value of the CT coronary calcium score in the prediction of the presence of CAD.

In **Chapter 9** we study the cost-effectiveness of coronary CT angiography as a non-invasive triage test prior to performing CAG using a decision tree and Markov-model, and determine the threshold probability below which coronary CT angiography would be cost-effective.

The cost-effectiveness of CT coronary calcium scoring with or without subsequent coronary CT angiography as compared with standard-of-care (i.e. stress electrocardiography) as a first-line diagnostic approach is studied in **Chapter 10**. We developed a decision tree and Markov-model based on a clinical cohort (with follow-up) combined with the best-available evidence.

Next, we evaluated the comparative costs and effectiveness of diagnostic strategies involving coronary CT angiography and myocardial perfusion imaging by stress cardiac magnetic resonance imaging, in **Chapter 11**.

In **Chapter 12**, we describe the follow-up results of a cohort of patients with stable chest pain who underwent both coronary CT angiography and stress electrocardiography. We compared their respective prognostic values regarding the occurrence of major adverse cardiac events.

Calculations of sensitivity and specificity of diagnostic tests commonly involve multiple observations per patient, which implies that the data are clustered. **Chapter 13** represents a tutorial article that describes the most commonly used methods for the analysis of clustered data and provides statistical code.

In **Chapter 14**, results are summarized and discussed.

References

1. Fuster V, Badimon L, Badimon JJ, Chesebro JH. The pathogenesis of coronary artery disease and the acute coronary syndromes (1). N Engl J Med 1992; 326:242-250.
2. Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. Lancet 1999; 353:89-92.
3. Framingham Heart Study. National Heart Lung and Blood Institute. http://www.framinghamheartstudy.org/. Accessed October 13
4. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, Tolonen H, Ruokokoski E, Amouyel P. Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease. Lancet 1999; 353:1547-1557.
5. Simoons ML, Windecker S. Controversies in cardiovascular medicine: Chronic stable coronary artery disease: drugs vs. revascularization. Eur Heart J 2010; 31:530-541.
6. Bravata DM, Gienger AL, McDonald KM, et al. Systematic review: the comparative effectiveness of percutaneous coronary interventions and coronary artery bypass graft surgery. Ann Intern Med 2007; 147:703-716.
7. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med 1979; 300:1350-1358.
8. Pryor DB, Shaw L, McCants CB, et al. Value of the history and physical in identifying patients at increased risk for coronary artery disease. Ann Intern Med 1993; 118:81-90.
9. Morise AP, Haddad WJ, Beckner D. Development and validation of a clinical score to estimate the probability of coronary artery disease in men and women presenting with suspected coronary disease. Am J Med 1997; 102:350-356.
10. Hunink MGM, Glasziou PP, Siegel JE, et al. Decision making in health and medicine: Integrating evidence and values. Cambridge: Cambridge University Press, 2001.
11. Kwok Y, Kim C, Grady D, Segal M, Redberg R. Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol 1999; 83:660-666.
12. Morise AP, Diamond GA. Comparison of the sensitivity and specificity of exercise electrocardiography in biased and unbiased populations of men and women. Am Heart J 1995; 130:741-747.
13. Hamon M, Fau G, Nee G, Ehtisham J, Morello R. Meta-analysis of the diagnostic performance of stress perfusion cardiovascular magnetic resonance for detection of coronary artery disease. J Cardiovasc Magn Reson 2010; 12:29.
14. Nandalur KR, Dwamena BA, Choudhri AF, Nandalur MR, Carlos RC. Diagnostic performance of stress cardiac magnetic resonance imaging in the detection of coronary artery disease: a meta-analysis. J Am Coll Cardiol 2007; 50:1343-1353.
15. Heijenbrok-Kal MH, Fleischmann KE, Hunink MG. Stress echocardiography, stress single-photon-emission computed tomography and electron beam computed tomography for the assessment of coronary artery disease: a meta-analysis of diagnostic performance. Am Heart J 2007; 154:415-423.
16. Nandalur KR, Dwamena BA, Choudhri AF, Nandalur SR, Reddy P, Carlos RC. Diagnostic performance of positron emission tomography in the detection of coronary artery disease: a meta-analysis. Acad Radiol 2008; 15:444-451.
17. Sarwar A, Shaw LJ, Shapiro MD, et al. Diagnostic and prognostic value of absence of coronary artery calcification. JACC Cardiovasc Imaging 2009; 2:675-688.
18. Schuetz GM, Zacharopoulou NM, Schlattmann P, Dewey M. Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. Ann Intern Med 2010; 152:167-177.

19. Sun Z, Lin C, Davidson R, Dong C, Liao Y. Diagnostic value of 64-slice CT angiography in coronary artery disease: a systematic review. Eur J Radiol 2008; 67:78-84.
20. Vanhoenacker PK, Heijenbrok-Kal MH, Van Heste R, et al. Diagnostic performance of multidetector CT angiography for assessment of coronary artery disease: meta-analysis. Radiology 2007; 244:419-428.
21. von Ballmoos MW, Haring B, Juillerat P, Alkadhi H. Meta-analysis: Diagnostic Performance of Low-Radiation-Dose Coronary Computed Tomography Angiography. Ann Intern Med 2011; 154:413-420.
22. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med 1980; 302:1109-1117.

Chapter 2

The Epidemiology of Coronary Artery Disease

Clinical Applications of Cardiac CT – Springer-Verlag Italia 2012
Filippo Cademartiri, Giancarlo Casolo, and Massimo Midiri (eds.)

Reproduced with permission

Tessa S.S. Genders
M.G. Myriam Hunink

Coronary artery disease is defined as a pathologic process affecting the coronary arteries, although it is most often used to indicate atherosclerotic changes in the coronary arteries. This is also referred to as coronary heart disease (CHD). According to the American Heart Association (1), coronary heart disease includes the diagnoses of acute myocardial infarction, angina pectoris and all other forms of acute and chronic ischemic heart disease (ICD/10 I20-I25).

This chapter will discuss the epidemiology of CHD, including estimates on incidence, prevalence, mortality, and risk factors. Patients who suffer from CHD often present with either angina pectoris or an acute myocardial infarction (MI). Since cardiac computed tomography can play an important role in evaluating such patients, we will discuss the epidemiology of angina pectoris and acute MI separately as well.

Coronary Heart Disease

Incidence

Based on observations in the Framingham Heart Study, the lifetime risk at age 40 of developing CHD is estimated to be 49% in men, and 32% in women (2). For both men and women, the risk of CHD increases with age (3-4) (Table 1.1) The incidence in men is substantially higher compared to premenopausal women. However, the morbidity incidence in women increases dramatically beyond menopause (5), thereby reducing the differences between men and women. In support of this, a large Finnish cohort study of over 14 000 individuals found that the CHD incidence was 3 times higher in men compared to women. It was demonstrated that risk factor levels were more favorable in women, although the differences in the risk factors between men and women decreased with advancing age (6).

The National Health and Nutrition Examination Survey (NHANES) Epidemiologic Follow-up study in the United States (US) compared the CHD incidence in two historical cohorts. They found that the age-standardized incidence of CHD decreased from 133.3 per 10 000 person years for the 1971-1982 cohort to 113.5 per 10 000 person years for the 1982-1992 cohort (7).

Prevalence

According to the NHANES 2003-2006, the prevalence of CHD is 7.9% in US adults aged 20 and older. For men and women separately, the prevalence is 9.1% and 7.0% (1).

Mortality

Although total mortality from cardiovascular diseases (which includes stroke) has declined over the past years, it remains the most common cause of death both in Europe (8) and the US (1). CHD accounts for approximately 50% of all deaths due to cardiovascular diseases (1, 8). In Europe, 1.92 million CHD related deaths occur every year. The age-standardized mortality rate for men varies considerably across Europe (9), with lower mortality rates in Central, Southern, and Western Europe. The highest mortality rates are observed in Eastern Europe. A similar pattern was observed for women. On average, it is estimated that over one in six women and over one in five men will die from CHD (8).

Table 1. Incidence (cases / 1000 person years) of coronary heart disease* by age and sex based on the Framingham Heart Study, 1980-2003 (7)

Age	Men	Women
45-54	8.0	1.8
55-64	16.4	6.4
65-74	21.8	11.0
75-84	32.8	17.1
85-94	38.0	22.8

** MI, angina pectoris, coronary insufficiency, or fatal CHD*

In the US from 1980 through 2000, the age-adjusted death rate for CHD decreased from 543 to 267 deaths per 100 000 population among men. Among women, the rate decreased from 263 to 134 deaths per 100 000. It is estimated that approximately half of the decrease is attributable to improvements in diagnosis and treatment and the other half to changes in risk factors (10, 11).

Risk Factors

The Framingham Heart Study (12) has provided valuable insight into the epidemiology and risk factors of cardiovascular diseases. The most important modifiable risk factors associated with an increased cardiovascular disease risk include: smoking, dyslipidaemia, high blood pressure, diabetes, and obesity (13). Other factors proven to increase the risk of cardiovascular disease include: psychosocial factors (14), a low physical activity (15), a lack of daily consumption of fruit and vegetables (16) and a lack of regular alcohol intake (17). A recent case-control study across 52 countries demonstrated that over 90% of all acute MIs as initial coronary event can be attributed to currently established risk factors (17).

Angina Pectoris

Incidence

Based on data from the NHLBI ARIC (National Heart, Lung, and Blood Institute - Atherosclerosis Risk in Communities) cohort (1987-2001), the age-adjusted incidence of angina pectoris as determined by the Rose Questionnaire was 10.8 cases per 1000 person years in men, and 13.4 in women. The highest incidence (17.9 per 1000 person years) was observed in black women. A large Finnish prospective cohort study focused on the incidence of angina defined as a nitrate prescription or angina with positive diagnostic findings. The age-standardized annual incidence was 2.03 per 100 in men and 1.89 per 100 in women. In this study, no sex difference in prognosis (i.e. absolute coronary event rate at 4 years) was observed. The Euro Heart Survey studied over 3000 patients with a new clinical diagnosis by a cardiologist of stable angina and found that the rate of death and non-fatal MI at 1 year was 2.3 per 100 patient years (18). They also demonstrated that of those patients with a new clinical diagnosis of stable angina, women with confirmed coronary artery disease have a 2-fold increased rate of death and nonfatal MI within the first year as compared to men (19).

Prevalence

According to a systematic review of existing healthy population based studies, the prevalence of angina as assessed by the Rose Angina Questionnaire, is estimated to be 5.7% in men and 6.7% in women, which varied widely (0.7-15%) across countries (20).

Acute Myocardial Infarction

More than 50% of all acute coronary events occur in patients without any history of CHD (21) and only 18% is preceded by longstanding angina symptoms (1).

Incidence

An analysis of US patients hospitalized between 1999 and 2008 revealed an age- and sex-adjusted incidence rate of acute MI of 208 cases per 100 000 person years (22). Reports indicate that the incidence rate of ST elevation MI declined over the past decades (22, 23). However, results also suggest that the true (larger) decline in incidence rate may be masked by an increased detection by using increasingly sensitive biomarkers in the diagnosis of acute MI (23).

Few data are available on CHD incidence rates for the European region (8). Results from the WHO MONICA (World Health Organization - monitoring trends and determinants in cardiovascular disease) Project studied the trends over 10 years in CHD across 21 European countries, starting in the early 1980s. A coronary event included non-fatal events satisfying the criteria for definite MI, and fatal events classified as definite, possible, and unclassifiable coronary deaths. Observed annual event rates were 434 per 100 000 in men, and 103 per 100 000 in women. In this study, an annual decline in coronary events of approximately 2.1% for men and 1.4% for women was observed (24).

The Rotterdam study, a population-based cohort of 5148 men and women aged 55 and older in the Netherlands, revealed a documented MI incidence rate of 8.4 and 3.1 per 1000 person years for men and women, respectively. Furthermore, the rate of additional unrecognized infarctions was estimated to be 4.2 for men and 3.6 per 1000 person years for women, which implies that a substantial proportion (approximately 50%) of myocardial infarctions remain clinically unrecognized (25).

Prevalence

According to NHANES data, the prevalence of MI between 1988 and 1994 among US adults aged 35-54 was 2.5% for men and 0.7% for women (26). The prevalence among men decreased over time (2.5% to 2.1%) but the prevalence among women increased (0.7% to 1.0%), as observed in the period between 1999 and 2004.

Prognosis

It is well recognized that patients who experienced a coronary event are at increased risk for future cardiovascular events. The absolute risk however, depends largely on the clinical presentation and the presence of other cardiovascular risk factors.

In 2002, the Euro Heart Survey of acute coronary syndromes found a 30-day mortality rate of 8.4%, 3.5% and 13.3% for patients with ST elevation MI (STEMI), non ST elevation MI (NSTEMI) and undetermined electrocardiogram acute coronary syndrome, respectively (27). The OPERA-registry demonstrated a similar 30-day mortality rate for STEMI and NSTEMI, but an increased 1-year mortality rate for NSTEMI as compared to STEMI (11.6% vs. 9.0%) (28). The GUSTO-IIB trial investigated the prognosis of over 12 000 patients after an acute coronary syndrome. Results indicated that the 30-day mortality was higher for STEMI compared to NSTEMI patients (6.1% vs. 3.8%), whereas the cumulative mortality rate after one year was not significantly different (9.6% vs. 8.8%) (29). Furthermore, a US population- based study of nearly 3000 acute MI patients found that the rate of sudden cardiac death after MI was constant at 1.2% per year, and that the 5-year cumulative incidence was 6.9% (30).

The National Registry of Myocardial Infarction in the US evaluated nearly 2 million patients that were hospitalized for acute MI from 1994 to 2006 and showed that the overall in-hospital mortality decreased from 10.4% in 1994 to 6.3% in 2006 (31).

Since a large part of the myocardial infarctions remain clinically unrecognized, it is important to investigate the clinical significance of such an event. A US based Cardiovascular Health Study database of patients 65 years and older allowed for the comparison of the prognosis between patients with unrecognized and recognized MIs. It was revealed that the six-year mortality rate was not significantly different between the two groups (32).

Conclusions

The CHD incidence and mortality decreased over the past decades due to improved medical treatments and changes in risk factors. CHD accounts for approximately 50% of all deaths due to cardiovascular diseases. In spite of the declining incidence and mortality, CHD manifestations such as angina pectoris and acute myocardial infarction are common. A substantial proportion of acute myocardial infarctions remain clinically unrecognized, but carry a similar prognosis as compared to recognized acute myocardial infarctions. More than 50% of all acute myocardial infarctions occur in patients without any history of coronary heart disease.

References

1. American Heart Association (2010) Heart Disease and Stroke Statistics – 2010 Update. Dallas, Texas: American Heart Association
2. Lloyd-Jones DM, Larson MG, Beiser A, Levy D (1999) Lifetime risk of developing coronary heart disease. Lancet 353:89-92
3. Wilson PW, D'Agostino RB, Levy D et al (1998) Prediction of coronary heart disease using risk factor categories. Circulation 97:1837- 1847
4. National Heart, Lung, and Blood Institute (2006) Incidence and Prevalence: 2006 Chart Book on Cardiovascular and Lung Diseases. Bethesda, MD: National Heart, Lung, and Blood Institute
5. Gordon T, Kannel WB, Hjortland MC, McNamara PM (1978) Menopause and coronary heart disease. The Framingham Study. Ann Intern Med 89:157-161
6. Jousilahti P, Vartiainen E, Tuomilehto J, Puska P (1999) Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. Circulation 99:1165-1172
7. Ergin A, Muntner P, Sherwin R, He J (2004) Secular trends in cardiovascular disease mortality, incidence, and case fatality rates in adults in the United States. Am J Med 117:219-227
8. European Heart Network (2008) European cardiovascular disease statistics: University of Oxford. http://www.ehnheart.org/cdv-statistics. html. Accessed 12 October 2011
9. Muller-Nordhorn J, Binting S, Roll S, Willich SN (2008) An update on regional variation in cardiovascular mortality within Europe. Eur Heart J 29:1316-1326
10. Ford ES, Ajani UA, Croft JB et al (2007) Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. N Engl J Med 356:2388-2398
11. Wijeysundera HC, Machado M, Farahati F et al (2010) Association of temporal trends in risk factors and treatment uptake with coronary heart disease mortality, 1994-2005. JAMA 303:1841- 1847
12. National Heart Lung and Blood Institute (2010) Framingham Heart Study. National Heart Lung and Blood Institute. http://www.framinghamheartstudy. org/. Accessed 13 October 2011
13. D'Agostino RB Sr,Vasan RS, Pencina MJ et al (2008) General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 117:743-753
14. Haynes SG, Feinleib M, Levine S et al (1978) The relationship of psychosocial factors to coronary heart disease in the Framingham study. II. Prevalence of coronary heart disease. Am J Epidemiol 107:384-402
15. Kannel WB (1967) Habitual level of physical activity and risk of coronary heart disease: the Framingham study. Can Med Assoc J 96:811-812
16. Bazzano LA, He J, Ogden LG et al (2002) Fruit and vegetable intake and risk of cardiovascular disease in US adults: the first National Health and Nutrition Examination Survey Epidemiologic Follow- up Study. Am J Clin Nutr 76:93-99
17. Yusuf S, Hawken S, Ounpuu S et al (2004) Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 364:937-952
18. Daly CA, De Stavola B, Sendon JL et al (2006) Predicting prognosis in stable angina—results from the Euro heart survey of stable angina: prospective observational study. BMJ 332:262-267
19. Daly C, Clemens F, Lopez Sendon JL et al (2006) Gender differences in the management and clinical outcome of stable angina. Circulation 113:490-498
20. Hemingway H, Langenberg C, Damant J et al (2008) Prevalence of angina in women versus men: a sysematic review and meta-analysis of international variations across 31 countries. Circulation 117:1526-1536

21. Tunstall-Pedoe H, Morrison C,Woodward M et al (1996) Sex differences in myocardial infarction and coronary deaths in the Scottish MONICA population of Glasgow 1985 to 1991. Presentation, diagnosis, treatment, and 28-day case fatality of 3991 events in men and 1551 events in women. Circulation 93:1981-1992
22. Yeh RW, Sidney S, Chandra M et al (2010) Population trends in the incidence and outcomes of acute myocardial infarction. N Engl J Med 362:2155-2165
23. Parikh NI, Gona P, Larson MG et al (2009) Long-term trends in myocardial infarction incidence and case fatality in the National Heart, Lung, and Blood Institute's Framingham Heart study. Circulation 119:1203-1210
24. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M et al (1999) Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease. Lancet 353:1547-1557
25. de Torbal A, Boersma E, Kors JA et al (2006) Incidence of recognized and unrecognized myocardial infarction in men and women aged 55 and older: the Rotterdam Study. Eur Heart J 27:729-736
26. Towfighi A, Zheng L, Ovbiagele B (2009) Sex-specific trends in midlife coronary heart disease risk and prevalence. Arch Intern Med 169:1762-1766
27. Hasdai D, Behar S,Wallentin L et al (2002) A prospective survey of the characteristics, treatments and outcomes of patients with acute coronary syndromes in Europe and the Mediterranean basin; the Euro Heart Survey of Acute Coronary Syndromes (Euro Heart Survey ACS). Eur Heart J 23:1190-1201
28. Montalescot G, Dallongeville J, Van Belle E et al (2007) STEMI and NSTEMI: are they so different? 1 year outcomes in acute myocardial infarction as defined by the ESC/ACC definition (the OPERA registry). Eur Heart J 28:1409-1417
29. Armstrong PW, Fu Y, Chang WC et al (1998) Acute coronary syndromes in the GUSTO-IIb trial: prognostic insights and impact of recurrent ischemia. The GUSTO-IIb Investigators. Circulation 98:1860-1868
30. Adabag AS, Therneau TM, Gersh BJ et al (2008) Sudden death after myocardial infarction. JAMA 300:2022-2029
31. Rogers WJ, Frederick PD, Stoehr E et al (2008) Trends in presenting characteristics and hospital mortality among patients with ST elevation and non-ST elevation myocardial infarction in the National Registry of Myocardial Infarction from 1990 to 2006. Am Heart J 156:1026-1034
32. Sheifer SE, Gersh BJ,Yanez ND 3rd et al (2000) Prevalence, predisposing factors, and prognosis of clinically unrecognized myocardial infarction in the elderly. J Am Coll Cardiol 35:119-126

Part II

Systematic Reviews and Meta-Analyses

Chapter 3.1

Systematic Review of Guidelines on Imaging Tests for Asymptomatic Coronary Artery Disease

J Am Coll Cardiol. 2011 Apr 12;57(15):1591-600.

Bart S. Ferket
Tessa S.S. Genders
Ersen B. Colkesen
Jacob J. Visser
Sandra Spronk
Ewout W. Steyerberg
M.G. Myriam Hunink

Abstract

Objectives

The purpose of this study was to critically appraise guidelines on imaging of asymptomatic coronary artery disease (CAD).

Background

Various imaging tests exist to detect CAD in asymptomatic persons. Because randomized controlled trials are lacking, guidelines that address the use of CAD imaging tests may disagree.

Methods

Guidelines in English published between January 1, 2003, and February 26, 2010, were retrieved using MEDLINE, Cumulative Index to Nursing and Allied Health Literature, the National Guideline Clearinghouse, the National Library for Health, the Canadian Medication Association Infobase, and the Guidelines International Network International Guideline Library. Guidelines developed by national and international medical societies from Western countries, containing recommendations on imaging of asymptomatic CAD were included. Rigor of development was scored by 2 independent reviewers using the Appraisal of Guidelines Research and Evaluation (AGREE) instrument. One reviewer performed full extraction of recommendations, which was checked by a second reviewer.

Results

Of 2,415 titles identified, 14 guidelines met our inclusion criteria. Eleven of 14 guidelines reported relationship with industry. The AGREE scores varied across guidelines from 21% to 93%. Two guidelines considered cost-effectiveness. Eight guidelines recommended against or found insufficient evidence for testing of asymptomatic CAD. The other 6 guidelines recommended imaging patients at intermediate or high CAD risk based on the Framingham risk score, and 5 considered computed tomography calcium scoring useful for this purpose.

Conclusions

Guidelines on risk assessment by imaging of asymptomatic CAD contain conflicting recommendations. More research, including randomized controlled trials, evaluating the impact of imaging on clinical outcomes and costs is needed.

Introduction

As many as 50% of myocardial infarctions occur in persons without a known history of symptomatic coronary artery disease (CAD) (1). To diminish disease burden, primary prevention on the individual level is currently rendered by targeting high-risk subjects, who are identified by office based risk assessment using multiple traditional cardiovascular risk predictors: age, sex, smoking, lipid levels, and blood pressure. Screening using these traditional predictors, however, misses a considerable proportion of persons who will suffer from coronary events (2).

Because symptomatic CAD has a pre-clinical detectable phase (i.e., coronary atherosclerosis), early detection of CAD in apparently healthy persons may be an important substitute for or supplement to risk assessment based on the traditional risk factors. See page 1601 Because technical developments have created various imaging techniques to assess a patient's coronary condition, clinicians are faced with multiple options to choose from. Before a doctor decides to test for asymptomatic disease, the intervention should meet a set of specific screening criteria (3–5). Hence, clinicians and decision makers usually rely on clinical practice guidelines in which recommendations are made on the basis of these criteria. As opposed to cancer screening, few large randomized controlled trials (RCTs) studying the effect of early detection of CAD on event rates within an asymptomatic population have been performed. In absence of RCTs demonstrating a net health benefit of imaging, the weighing of harms and benefits is more likely to result in different judgments, and therefore, conflicting recommendations. Therefore, a critical appraisal of guidelines and review of the agreements and the differences among recommendations can serve as a guide for deciding which imaging tests to use in clinical practice.

For this purpose, we systematically reviewed guidelines containing recommendations on imaging of asymptomatic CAD within the general population.

Methods

Data sources and searches

To identify appropriate guidelines, the literature search used for a previous article on cardiovascular risk assessment (6) was updated and covered a period from January 1, 2003, to February 26, 2010. Briefly, MEDLINE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), and 4 guideline databases — the National Guideline Clearinghouse (United States), the National Library for Health (United Kingdom) on Guideline Finder, Canadian Medical Association InfoBase (Canada), and the Guidelines International Network (G-I-N) International Guideline Library — were searched. Searches were limited to guidelines from the United States, Canada, United Kingdom, Australia, and New Zealand, and international guidelines in the English language. A search on websites of guideline development organizations was performed for additional guidelines.

Study selection

Articles were considered if they met the Institute of Medicine definition for clinical practice guidelines. The Institute of Medicine defines clinical practice guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate health

care for specific clinical circumstances." If doubt existed whether a report met this definition or not, we verified eligibility by checking the inclusion of similar reports in the National Guideline Clearinghouse. This database also uses the Institute of Medicine definition. For this reason, we also considered American Heart Association (AHA) expert consensus documents and scientific statements, and American College of Cardiology Foundation (ACCF) appropriateness criteria reports. We included guidelines if they: 1) contained recommendations on imaging of asymptomatic CAD specifically aimed to prevent a first coronary event; 2) involved apparently healthy persons, that is adults without, for example, diabetes mellitus; and 3) were produced on behalf of a national or international medical specialty society. For completeness, we also included guidelines on electrocardiography and exercise tolerance tests, because these tests are traditionally used in the diagnosis of CAD.

The SRS 4.0 (Mobius Analytics, Ottawa, Ontario, Canada), a web-based software package developed for systematic review data management, was used. Review of titles and abstracts was performed independently by 2 reviewers (B.S.F. and E.B.C.). For a paper to be excluded, both reviewers had to agree that the article was ineligible. For abstracts, disagreements between the reviewers were discussed and resolved by consensus. The final selection based on the full text was performed by the first author.

Data extraction and quality assessment

One reviewer (B.S.F.) extracted all relevant recommendations from each included guideline. A second reviewer (T.S.S.G.) checked the results obtained for accuracy and completeness. Discrepancies were resolved by consensus. Each guideline could provide 1 or more relevant recommendations. Data extracted on a guideline level included the reported methodology for evidence synthesis, and formulating of recommendations. On the recommendation level, we extracted data on consideration of cost effectiveness, the target population, the strategy for delivery of the test, coronary atherosclerosis tests, intervention, and follow-up. In addition, the strength of the recommendation was classified as "for," "consider," "not for, not against," "insufficient evidence," or "against."

We assessed the quality of development for each included guideline using the 7-item Rigor of Development domain of the Appraisal of Guidelines Research and Evaluation (AGREE) instrument (7). This domain considers the reporting of: 1) methods to search for evidence; 2) criteria for selecting the evidence; 3) methods for formulating the recommendations; 4) consideration of health benefits, side effects, and risks; 5) supporting evidence; 6) procedures for external peer review; and 7) the update process. Each item was independently rated on a 4-point Likert scale by 2 reviewers (B.S.F. and T.S.S.G.). Websites of guideline developers were examined by both reviewers for additional information on the development processes. For each reviewer, AGREE scores were calculated as a percentage using the sum of the 7 items and the maximum possible score. If the total AGREE scores of the 2 reviewers differed >20%, a third independent reviewer (J.J.V.) also assessed the guideline. Final rigor scores were calculated by averaging the AGREE scores from all reviewers. Three guidelines (8–10) were rated by 3 reviewers. We ranked included guidelines according to their score. Editorial independence from funding body, external funding, proportion of guideline panel member-industry relationships, and disclosure of identities and relationships with industry of peer reviewers were assessed by 1 reviewer (B.S.F.) and checked by a second reviewer (T.S.S.G.). Discrepancies were resolved by consensus.

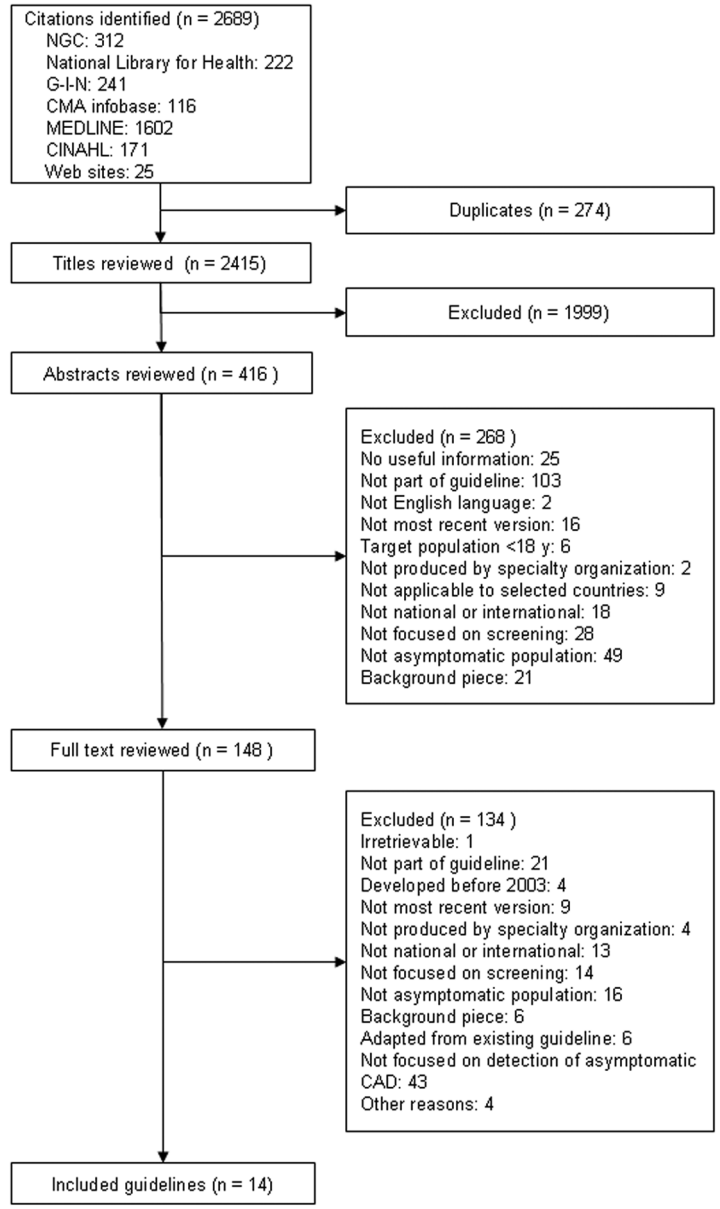


Figure 1. Literature Search and Selection Numbers of guidelines for each step of the process are indicated. Group totals may exceed the reported numbers for the excluded articles at abstract and full text level because several reasons for exclusion were allowed. CAD = coronary artery disease, CINAHL = Cumulative Index to Nursing and Allied Health Literature, CMA = Canadian Medical Association, G-I-N = Guidelines International Network, NGC = National Guideline Clearinghouse.

Data synthesis and analysis

A table for comparison of the recommendations from the selected guidelines was constructed. The table was divided into 1) methodology of guideline development; 2) consideration of cost effectiveness regarding the recommendation; 3) target group and delivery of early detection; 4) tests considered; and 5) thresholds for intervention and follow-up.

Agreement between reviewers on AGREE scores was assessed using the intraclass correlation coefficient. Given the limited number of guidelines, only explorative quantitative analyses were possible. We examined the correlation between the proportion of guideline panel members who reported relationships with industry and the AGREE score with guidelines as units of analysis. Furthermore, we examined whether the proportion of panel members with industry relationships and the AGREE score were associated with a positive recommendation ("consider" or "for") by logistic regression. Two guidelines that had no explicit statement on conflicts of interest of panel members were excluded from the analyses. An alpha level of 0.05 was used to indicate statistical significance. All analyses were performed using SPSS, version 15.0 (SPSS Inc., Chicago, Illinois).

Results

Selected guidelines

Fourteen guidelines (11–22) relevant to testing of asymptomatic CAD were eligible for full data extraction (Figure 1). Table 1 summarizes the selected guidelines, together with AGREE score and conflict of interest results. Most guidelines (10 of 14) were developed in the United States. The AGREE scores varied from 21% to 93%, with a median AGREE score of 57%. Reproducibility of the 2 reviewers' average AGREE scores was good, with an intraclass correlation coefficient of 0.76. Examples of low scoring guidelines are the ACCF appropriateness criteria reports (ACCF2-4) (20–22). These guidelines provided excellent information on the methods followed for achieving consensus and formulating recommendations, but did not contain detailed information on the search strategy used to identify the evidence. Although "a standardized literature review" was performed for these reports, key words used in the search strategy, and inclusion and exclusion criteria for selecting articles were not reported. In addition, these guidelines did not explicitly discuss benefits and harms of recommendations and methodology for guideline updating.

Twelve of the 14 guidelines contained disclosure of relationships with industry, and in 11, at least 1 panel member declared having a relevant financial relationship. In this limited set of 12 guidelines, no relationship between the AGREE score and the proportion of panel members with an industry relationship was observed (Pearson's correlation r = -0.205; P = 0.523).

General findings among the recommendations

The 14 included guidelines contained 26 recommendations on testing of asymptomatic CAD (Table 2). The following tests were considered: computed tomography (CT) calcium scoring, CT angiography, magnetic resonance (MR) angiography, single-photon emission computed tomography (SPECT), positron emission tomography (PET), stress echography, resting electrocardiography, and exercise tolerance testing. The majority of guidelines, except for the Canadian Cardiovascular Society (CCS) 2 guideline (19), were based on a comprehensive review including study quality assessment. Apart from the Canadian Association of Radiologists (CAR) (18) and CCS2 (19) guidelines, a grading system for assigning the level of evidence was used. Evaluation of cost-effectiveness of recommended tests was explicitly done in only 2 guidelines, the U.S. Preventive Services Task Force (USPSTF) 1 (11) and ACCF1 (15) guidelines, by reviewing decision modeling studies on exercise tolerance testing (23–25) and CT calcium scoring (26,27), respectively. However, both guideline

groups were unable to find a sufficient number of high-quality cost-effectiveness analyses on which to base their recommendations. In other guidelines (ACCF2 [20] and ACCF4 [22]), group members were requested to consider costs in their decision making as well, but this was not based on a review of cost-effectiveness studies or decision analyses.

Eight of the 14 guidelines recommended against or concluded that there is insufficient evidence for testing of asymptomatic CAD. In the remaining 6 guidelines (ACCF1 [15], AHA2 [8], National Cholesterol Education Program [NCEP] [16,17], CAR [18], CCS2 [19], and ACCF2 [20]), testing was only advocated for patients with an a priori elevated risk level based on absolute CAD risk or multiple risk factors. Generally, risk was determined by Framingham risk equations for estimation of a 10-year risk for coronary events (fatal and non-fatal) using the categories <10%, 10% to 20%, and >20% for, respectively, low, intermediate, and high risk. However, 2 guidelines (CAR [18] and CCS2 [19]) did not specify any criteria for low, intermediate, and high risk. None advocated a universal screening approach or screening based on an age criterion alone. Whether a guideline contained a recommendation that supports testing or not did not statistically significantly depend on AGREE score or proportion of panel members with industry relationships. Adjusted odds ratios per 10% increase were 0.73 (95% confidence interval: 0.41-1.33) and 0.68 (95% confidence interval: 0.35-1.32), respectively.

The indications for further testing and primary preventive measures were not described in much detail. Overall, in guidelines that recommended for consideration of testing of asymptomatic CAD within an intermediate-risk population (ACCF1 [15], AHA2 [8], NCEP [16,17], CAR [18], and CCS2 [19]), all (previous intermediate risk) subjects were marked as high risk after a positive test. None of these guidelines contained recommendations in which traditional prediction models were updated by including test results as covariate. In addition, none of the guidelines reported whether the tests should be performed once or periodically in case of a negative test result.

CT calcium scoring

Most guidelines (10 of 14) considered the CT calcium score as a test for improvement of total coronary risk assessment based on traditional risk factors. Among these 10 guidelines (USPSTF1 [11], USPSTF2 [12], New Zealand Guidelines Group [NZGG] [13], ACCF1 [15], CCS1 [10], AHA2 [8], NCEP [16,17], CAR [18], CCS2 [19], and ACCF4 [22]), 4 guidelines (ACCF1 [15], AHA2 [8], NCEP [16,17], and CCS2 [19]) concluded that there was sufficient evidence for consideration of its use, and 1 guideline (CAR) (18) recommended for its use. These guidelines recommended CT calcium scoring solely in an intermediate CAD risk population. In contrast, the USPSTF2 (12), NZGG (13), and ACCF4 (22) guidelines concluded that there is insufficient evidence for the intermediate-risk population. For low CAD risk persons and persons already known to be at high CAD risk, guidelines were unanimous in not advocating CT calcium scoring

Electrocardiography and exercise tolerance testing

The USPSTF1 guideline (11) recommended against performing electrocardiography testing in a low-risk population and found insufficient evidence for subjects at elevated risk. No other guidelines provided recommendations for this test. Exercise tolerance testing was considered in 4 guidelines (USPSTF1 [11], NCEP [16,17], CCS1 [10], and AHA3 [9]): 1 (NCEP [16,17]) recommended considering testing, and 3 (CCS1 [10], USPSTF1 [11], and AHA3 [9]) were inconclusive.

Myocardial perfusion imaging

Single-photon emission computed tomography was considered in 3 guidelines (AHA2 [8], NCEP [16,17], and ACCF2 [20]), of which 2 (AHA2 [8] and NCEP [16,17]) also considered PET. The AHA2 guideline (8) recommended against any use of myocardial perfusion imaging in asymptomatic subjects, whereas the NCEP (16,17) and ACCF2 (20) guidelines recommended its use for different target populations: either for intermediate-risk subjects (NCEP [16,17]) or solely for those at high risk (ACCF2 [20]).

CT angiography and MR angiography

The AHA1 (14), CAR (18), CCS2 (19), and ACCF4 (22) guidelines considered these tests for the asymptomatic population. None of these guidelines advocated their use. For subjects at high risk, insufficient evidence was found by the ACCF4 guideline (22).

Table 1. Characteristics of 14 Guidelines on Imaging of Asymptomatic Coronary Artery Disease

Guideline Identifier, Year (Ref. #)	Organization(s) Responsible for Guideline Development	Country Applied	AGREE Rigor Score, %	Conflicts of Interest	Reported Relations with Industry†
USPSTF1, 2004 (27)	U.S. Preventive Services Task Force	United States	93	EI, DIRp	-
USPSTF2, 2009 (28)	U.S. Preventive Services Task Force	United States	90	EI, SCI	0/23
NZGG, 2003 (29)	New Zealand Guidelines Group	New Zealand	79	EI, FPO, SCI*, DIRp	9/35
AHA1, 2008 (30)	American Heart Association	United States	76	SCI*, DIR, SCIR	8/11
ACCF1, 2007 (31)	American College of Cardiology Foundation, American Heart Association	United States	74	SCI*,DIR, SCIR*	4/14
CCS1, 2009 (10)	Canadian Cardiovascular Society	Canada	59	SCI*	20/23
AHA2, 2006 (8)	American Heart Association	United States	57	SCI*, DIR,SCIR*	1/4
AHA3, 2005 (9)	American Heart Association	United States	57	SCI*, DIR, SCIR*	6/12
NCEP, 2002, 2004 update (32,33)	National Heart, Lung, and Blood Institute, American College of Cardiology Foundation, and American Heart Association	United States	52	SCI*, DIR	6/28
CAR, 2009 (34)	Canadian Association of Radiologists	Canada	36	-	-
CCS2, 2009 (35)	Canadian Cardiovascular Society	Canada	31	SCI*	1/13
ACCF2, 2009 (36)	American College of Cardiology Foundation et al.	United States	24	SCI*, DIRp, SCIR*	13/29
ACCF3, 2008 (37)	American College of Cardiology Foundation et al.	United States	21	SCI*, DIRp, SCIR*	15/23
ACCF4, 2006 (38)	American College of Cardiology Foundation et al.	United States	21	SCI*, DIRp, SCIR*	11/25

ACCF = American College of Cardiology Foundation, AHA = American Heart Association, CAR = Canadian Association of Radiologists, CCS = Canadian Cardiovascular Society, DIR = disclosure of the identities of peer reviewers, DIRp = disclosure of the identities of peer reviewers for some parts of the guidelines, EI = editorial independence from funding organization declared, FPO = funding by external public organization reported, NCEP = National Cholesterol Education Program, NZGG = New Zealand Guidelines Group, SCI = statement about conflicts of interest of panel members present, SCIR = statement about conflicts of interest of external peer reviewers present.
** Relationship with industry reported by at least one person.*
† Proportion of panel members with reported industry relationships.

Stress echocardiography

Only 1 guideline (ACCF3 [21]) provided recommendations for stress echocardiography. For adults at high risk, insufficient evidence was found for its use; for the remaining asymptomatic population, stress echocardiography is not justified according to the ACCF3 guidelines (21).

Discussion

In summary, we identified 14 guidelines on testing of asymptomatic CAD. In the development of most guidelines, relationships with the industry were present. A considerable number of guidelines achieved a low AGREE score. Various inconsistencies were observed among the guidelines regarding interpretation of the value of early detection of CAD. Many guideline groups recommended against testing of asymptomatic CAD or concluded that there is insufficient evidence. The guidelines that contained recommendations to consider testing of asymptomatic CAD only reported benefit for those at elevated risk, that is, those who were either at intermediate or high absolute risk for having a CAD event. The majority of these guidelines supported consideration of CT calcium scoring in case of intermediate CAD risk.

Some possible limitations of this review deserve attention. First, only guidelines developed by national or international medical specialty organizations were reviewed. Hence, guidelines developed by local organizations, private organizations, and individual experts were not considered. An example of an often-cited guideline, therefore, not included is the Society for Heart Attack Prevention and Eradication (SHAPE) guideline (28). The SHAPE guideline recommends periodic measurement of coronary calcium or carotid intima-media thickness in all asymptomatic men ages 45 to 75 years and women ages 55 to 75 years except those defined at very low risk. Such a universal screening approach is, however, not advocated by any of the guidelines included in this systematic review. Second, we used the AGREE instrument, which provides an overall score of the construction process of guidelines, not components. Although we expect that the quality of development across the whole guideline influences the quality of individual recommendations, in theory, a solid recommendation could be created within a poorly developed guideline and vice versa. Third, the AGREE instrument only considers the reported information related to the development of the guideline. The actual quality of the guideline development can, therefore, not be fully captured. For example, guideline groups that performed a full search for evidence and that did not report detailed information on the search strategy followed, received a low AGREE score for this item. In reality, the search followed may be adequate for identifying solid evidence. Fourth, it was difficult to quantify the true degree of influence by industry relationships, also because guidelines did not report payment amounts. Fifth, the ability to detect statistically significant relationships in the quantitative analyses, such as an association between industry relationships and the likelihood of a positive recommendation, was limited owing to the small set of included guidelines.

The disagreements on the value of early detection of CAD across the guidelines could partly be explained by the paucity of experimental research. A search on ClinicalTrials.gov (29) up to March 22, 2010, using search terms "coronary artery disease" and "prevention" or "screening," provided 97 interventional studies. We found 5 RCTs on the effect of early detection of CAD versus current practice of risk assessment using traditional risk factors. Only 1 RCT (ClinicalTrials.gov identifier: NCT00927693) was conducted in an apparently healthy population, with CT calcium scoring as the intervention. The study's results on hard endpoints are, however, not yet published (30). One RCT (ClinicalTrials.gov identifier: NCT00769275) was performed in a population with diabetes and revealed no effect of screening by myocardial perfusion scans on cardiac event rates, although event rates in the screened and not-screened groups were low, and no standardized preventive

treatment strategy was used (31). Other RCTs (ClinicalTrials.gov identifiers: NCT00431977, NCT00488033, and NCT00547872), on CT angiography and exercise tolerance testing, were also conducted in diabetic patients, and are still ongoing.

Patients with subclinical atherosclerosis identified by accurate imaging tests can be expected to benefit from preventive treatment because they are at elevated risk for an event. Ideally, decision making as to whether imaging individual patients is beneficial should be based on RCTs comparing preventive measures guided by imaging versus not imaging and evaluating CAD event rates as outcome. Such RCTs are, however, expensive and time-consuming and not always feasible. In the absence of these RCTs, one would want to combine data from trials evaluating the effect of preventive measures with data from cohort studies reporting the association between imaging test results and CAD event rates. Qualitatively weighing and combining the relevant harms and benefits, as was done in the development of the reviewed guidelines, is difficult and may lead to different judgments about net health gains. Disagreements across guidelines can occur for other reasons, including different judgments about which research is relevant; risk of biases in selected research; the applicability of the research findings to the key questions; the relative importance of the anticipated costs; and also poor guideline development processes and conflicts of interests (32). We explored whether the latter 2 influenced the variation in recommendations, but found no evidence for this in the limited set of guidelines reviewed. Quantitatively, as opposed to qualitatively, weighing harms and benefits can be done using decision models that integrate the best-available evidence from multiple sources. Beneficial effects, adverse effects, and incurred costs of preventive treatment and follow-up can be summarized in an incremental cost-effectiveness ratio. In a few of the included guidelines, decision modeling studies were discussed; however, their quality was considered too low for policy making.

The recommended methods of refining CAD risk stratification using imaging test results can be improved by updating existing prediction models (33). None of the guidelines contained recommendations for the use of prediction models combining traditional risk factors and test results to calculate a new risk estimate. Instead, the Framingham-based intermediate risk (10% to 20% 10-year CAD risk) is reclassified to high risk (a 10-year CAD risk >20%), if the test result is positive, rather than updating the risk estimate. This approach has limitations. First, it requires consensus on these risk categories, which is not the case. Second, validity of the reclassified risk might become an issue. A positive test result may not elevate the predicted absolute CAD risk to the level of high risk if the subject was at the lower end of the intermediate risk distribution, for example, if the 10-year CAD risk was between 10% and 15% (34). Reported risk ratios of asymptomatic CAD adjusted for traditional risk factors, which might reclassify individuals, are usually derived from a comparison with a reference group without or with low indication for asymptomatic CAD (34–37). However, converting a risk ratio to absolute risk also depends on the distribution of the risk marker within the general population, which consists of subjects with and without this risk marker (38). Finally, a communication of a refined numerical risk theoretically offers a benefit in informing patients. Thus, we believe that future research should also focus on the value of updating traditional prediction models.

Conclusions

Guidelines on risk assessment by imaging of asymptomatic CAD contain conflicting recommendations. More research, including RCTs, evaluating the impact of imaging on clinical outcomes and costs is needed.

Table 2. Recommendations (n=26) in Guidelines (n=14) on Imaging of Asymptomatic CAD

	USPSTF1	USPSTF1	USPSTF2	NZGG	AHA1
AGREE rigor score, %	93%	93%	90%	79%	76%
Method to evaluate evidence	Systematic review * covering 1966 – June 2002	Systematic review * covering 1966 – June 2002	Systematic review * covering 1966 – July 2008; meta-analysis	Systematic review * covering 1989 – August 10, 2002; review of published systematic reviews, meta-analyses or guidelines	Standardized review † with MEDLINE search covering 1990 – 2006
Method to formulate recommendations	Expert consensus	Expert consensus	Expert consensus	Expert consensus	Expert consensus
Consideration of costs	Yes, a review of cost-effectiveness studies is performed	Yes, a review of cost-effectiveness studies is performed	No, because of limitations in the evidence of effectiveness, little information is available on cost-effectiveness	NR	NR
Target group	Adults at low CAD risk: <5-10% 10-year risk of CAD events	Adults at increased CAD risk: >15-20% 10-year risk of CAD events	Adults in the intermediate CAD risk category with 10-year CAD risk 10% - 20% (FRS)	Adults regardless of risk	Adults regardless of risk
Strategy	Opportunistic screening / case-finding	Opportunistic screening / case-finding	Opportunistic screening / case-finding	NR	NR
Strength of recommendation	Against	Insufficient evidence to make a recommendation	Insufficient evidence to make a recommendation	Insufficient evidence to make a recommendation	Against
Tests considered	CTCS; r-ECG; ETT	CTCS; r-ECG; ETT	CTCS	CTCS	CTA; MRA
Intervention(s) considered	More intensive risk factor modification or follow-up testing / ICA if: presence of calcium, r-ECG abnormalities, ST-segment depression ≥ 1 mm; CABG / PCI if severe CAS	More intensive risk factor modification or follow-up testing / ICA if: presence of calcium, r-ECG abnormalities, ST-segment depression ≥ 1 mm; CABG / PCI if: severe CAS	Agressive risk reduction if reclassified 10-year CAD risk > 20% using CAC score categories: none, 1-100, 101-300, and >300, no established norms for the general population	Statins, aspirin, and intensive lifestyle therapy if 5-year CVD risk ≥ 15%, cut-off values for CAC score not specified	NR
Screening intervals	NR	NR	NR	Traditional risk assessment: annually if 5-year CVD risk ≥15%, in 5 years if 5-year CVD risk 5 – 15%, in 10 years if 5-year CVD risk < 5%; NR for CTCS	NR

Continued on next page

Table 2. Continued.

	ACCF1	ACCF1	CCS1	AHA2	AHA2
AGREE rigor score, %	74%	74%	59%	57%	57%
Method to evaluate evidence	Standardized review † with MEDLINE search covering 1998 – early 2005; review of published systematic reviews, meta-analyses or guidelines	Standardized review † with MEDLINE search covering 1998 – early 2005; review of published systematic reviews, meta-analyses or guidelines	Systematic review * covering January 1, 2006 - February 1, 2009; review of published systematic reviews, meta-analyses or guidelines	Standardized review †	Standardized review †
Method to formulate recommendations	Expert consensus	Expert consensus	Expert consensus	Expert consensus	Expert consensus
Consideration of costs	Yes, a review of cost-effectiveness studies is performed	Yes, a review of cost-effectiveness studies is performed	NR	NR	NR
Target group	Adults at high CAD risk: 10-year risk of CAD events ≥ 20% or other high-risk diagnosis; adults at low CAD risk: 10-year risk of CAD events < 10%	Adults at intermediate CAD risk: 10-year risk of CAD events 10-20%	Men at least 40 years of age, women at least 50 years of age or postmenopausal, adults at any age and ≥ 1 cardiovascular risk factor (family history of premature CAD, smoking, obesity)	Clinically selected intermediate CAD risk patients (e.g., those with a 10-year CAD risk 10-20% FRS)	Adults regardless of risk
Strategy	NR	NR	NR	NR	NR
Strength of recommendation	Against	Consider	Insufficient evidence to make a recommendation	Consider	Against
Tests considered	CTCS	CTCS	CTCS; ETT	CTCS	h-SPECT-CT/h-PET-CT
Intervention(s) considered	Pharmacologic treatment according to NCEP guidelines if 10-year CAD risk ≥ 20% based on high CAC score (≥ 400)	Pharmacologic treatment according to NCEP guidelines if 10-year CAD risk ≥ 20% based on high CAC score (≥ 400)	Statins and lifestyle intervention if subclinical atherosclerosis	More aggressive target values for lipid-lowering therapies if high CAC score based on absolute plaque burden	NR
Screening intervals	NR	NR	NR	Serial imaging for assessment of progression of coronary calcification is not indicated at this time	NR

Continued on next page

Table 2. Continued.

	AHA3	NCEP	CAR	CAR	CAR
AGREE rigor score, %	56%	52%	36%	36%	36%
Method to evaluate evidence	Standardized review †; review of published guideline	Standardized review † of literature identified by the panel members and by a MEDLINE search; review of published systematic reviews, meta-analyses or guidelines;	Systematic review * covering 1966 - October 2008; review of published systematic reviews, meta-analyses or guidelines	Systematic review * covering 1966 - October 2008; review of published systematic reviews, meta-analyses or guidelines	Systematic review * covering 1966 - October 2008; review of published systematic reviews, meta-analyses or guidelines
Method to formulate recommendations	Expert consensus	Expert consensus	Expert consensus	Expert consensus	Expert consensus
Consideration of costs	NR	NR for this recommendation	NR	NR	NR
Target group	Adults regardless of risk	2002: multiple risk factors and 10-year CAD risk ≤ 20%, 0-1 risk factor and LDL-C 160-189 mg/dL after lifestyle changes; 2004 update: 10-year CAD risk 10-20% and LDL-C 100-129 mg/dL (FRS)	Adults at intermediate CAD risk	Adults at low CAD risk or high CAD risk	Adults regardless of risk
Strategy	NR	Opportunistic screening / case-finding	NR	NR	NR
Strength of recommendation	Insufficient evidence to make a recommendation	Consider	For	Against	Against
Screening tests considered	ETT	2002: CTCS; ETT; SPECT/PET; 2004 update: CTCS	CTCS	CTCS	CTA
Intervention(s) considered	NR	2002: Statins and lifestyle intervention if subclinical atherosclerosis; 2004 update: consider statins if CAC score ≥ 75th percentile for a person's age and sex to achieve LDL-C < 100 mg/dL	Calcium scoring using a traditional scoring system may influence the decision to intensify risk factor modification	Calcium scoring using a traditional scoring system may influence the decision to intensify risk factor modification	If CAS ≥ 50%, intervention(s) not further specified
Screening intervals	NR	Traditional risk assessment in 3 months - 1 year depending on LDL-C level, NR for recommended tests	NR	NR	NR

Continued on next page

Table 2. Continued.

	CCS2	CCS2	ACCF2	ACCF2	ACCF2
AGREE rigor score, %	31%	31%	24%	24%	24%
Method to evaluate evidence	Review ‡	Review ‡	Standardized review †	Standardized review †	Standardized review †
Method to formulate recommendations	Expert consensus	Expert consensus	Expert consensus: Delphi method	Expert consensus: Delphi method	Expert consensus: Delphi method
Consideration of costs	NR	NR	Cost were considered implicitly in the appropriateness determination	Cost were considered implicitly in the appropriateness determination	Cost were considered implicitly in the appropriateness determination
Target group	Adults at intermediate CAD risk	Adults regardless of risk	Adults at low CAD risk or at intermediate CAD risk with an interpretable ECG (FRS)	Adults at intermediate CAD risk with an uninterpretable ECG (FRS)	Adults at high CAD risk (FRS)
Strategy	NR	NR	NR	NR	NR
Strength of recommendation	Consider	Against	Against	Insufficient evidence to make a recommendation	Consider
Screening tests considered	CTCS	CTA	SPECT	SPECT	SPECT
Intervention(s) considered	NR	Optimal medical therapy; PCI; CABG if test results consistent with high-risk CAD	NR	NR	NR
Screening intervals	NR	NR	NR	NR	NR

Continued on next page

Table 2. Continued.

	ACCF3	ACCF3	ACCF4	ACCF4	ACCF4	ACCF4
AGREE rigor score, %	21%	21%	21%	21%	21%	21%
Method to evaluate evidence	Standardized review†	Standardized review†	Standardized review†	Standardized review†	Standardized review†	Standardized review†
Method to formulate recommendations	Expert consensus: Delphi method	Expert consensus: Delphi method	Expert consensus: Delphi method	Expert consensus: Delphi method	Expert consensus: Delphi method	Expert consensus: Delphi method
Consideration of costs	NR	NR	Cost were considered implicitly in the appropriateness determination	Cost were considered implicitly in the appropriateness determination	Cost were considered implicitly in the appropriateness determination	Cost were considered implicitly in the appropriateness determination
Target group	Adults at low CAD risk or intermediate CAD risk (FRS)	Adults at high CAD risk (FRS)	Adults at low CAD risk or intermediate CAD risk (FRS)	Adults at high CAD risk (FRS)	Adults at low CAD risk (FRS)	Adults at intermediate CAD risk or high CAD risk (FRS)
Strategy	NR	NR	NR	NR	NR	NR
Strength of recommendation	Against	Insufficient evidence to make a recommendation	Against	Insufficient evidence to make a recommendation	Against	Insufficient evidence to make a recommendation
Screening tests considered	SE	SE	CTA	CTA	CTCS	CTCS
Intervention(s) considered	NR	NR	NR	NR	NR	NR
Screening intervals	NR	NR	NR	NR	NR	NR

CABG = coronary artery bypass graft, CAD = coronary artery disease, CAS = coronary artery stenosis, CT = computed tomography, CTA = computed tomography angiography, CTCS = computed tomography calcium score, CVD = cardiovascular disease, ETT = exercise tolerance test, FRS = Framingham Risk Score, h-SPECT = hybrid single-photon emission computed tomography, h-PET = hybrid positron emission tomography, ICA = invasive coronary angiography, LDL-C = low-density lipoprotein cholesterol, MRA = magnetic resonance angiography, NR = not reported, PCI = percutaneous coronary intervention, PET = positron emission tomography, r-ECG = resting electrocardiography, SE = stress echocardiography, SPECT = single-photon emission computed tomography.

* *Comprehensive literature review, which includes a search strategy that covers multiple databases and other sources, study selection criteria, and quality assessment of the evidence.*

† *Comprehensive literature review, which includes quality assessment of the evidence.*

‡ *Comprehensive literature review without systematic methods.*

References

1. Tunstall-Pedoe H, Morrison C, Woodward M, Fitzpatrick B, Watt G. Sex differences in myocardial infarction and coronary deaths in the Scottish MONICA population of Glasgow 1985 to 1991. Presentation, diagnosis, treatment, and 28-day case fatality of 3991 events in men and 1551 events in women. Circulation 1996;93:1981–92.
2. Law M, Wald N, Morris J. Lowering blood pressure to prevent myocardial infarction and stroke: a new preventive strategy. Health Technol Assess 2003;7:1–94.
3. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med 2001;20:21–35.
4. UK National Screening Committee. Programme appraisal criteria. 2009. Available at: http://www.screening.nhs.uk/criteria/. Accessed March 22, 2010.
5. Wilson J, Jungner G. Principles and practice of screening for disease. Public Health Paper. Geneva: World Health Organization, 1968.
6. Ferket BS, Colkesen EB, Visser JJ, et al. Systematic review of guidelines on cardiovascular risk assessment: which recommendations should clinicians follow for a cardiovascular health check? Arch Intern Med 2010;170:27– 40.
7. AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003;12:18 –23.
8. Budoff MJ, Achenbach S, Blumenthal RS, et al. Assessment of coronary artery disease by cardiac computed tomography: a scientific statement from the American Heart Association Committee on Cardiovascular Imaging and Intervention, Council on Cardiovascular Radiology and Intervention, and Committee on Cardiac Imaging, Council on Clinical Cardiology. Circulation 2006;114: 1761–91.
9. Lauer M, Froelicher ES, Williams M, Kligfield P. Exercise testing in asymptomatic adults: a statement for professionals from the American Heart Association Council on Clinical Cardiology, Subcommittee on Exercise, Cardiac Rehabilitation, and Prevention. Circulation 2005; 112:771– 6.
10. Genest J, McPherson R, Frohlich J, et al. 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult— 2009 recommendations. Can J Cardiol 2009;25:567–79.
11. U.S. Preventive Services Task Force. Screening for coronary heart disease: recommendation statement. Ann Intern Med 2004;140: 569 –72.
12. U.S. Preventive Services Task Force. Using nontraditional risk factors in coronary heart disease risk assessment: recommendation statement. Ann Intern Med 2009;151:474–82.
13. New Zealand Guidelines Group (NZGG). Assessment and management of cardiovascular risk. July 6, 2003. Available at: http://www. nzgg.org.nz/guidelines/dsp_guideline_popup.cfm?guidelineID_35&guideline_popup_mode_newsDisplay&NewsID_21. Accessed March 22, 2010.
14. Bluemke DA, Achenbach S, Budoff M, et al. Noninvasive coronary artery imaging: magnetic resonance angiography and multidetector computed tomography angiography. A scientific statement from the American Heart Association Committee on Cardiovascular Imaging JACC Vol. 57, No. 15, 2011 Ferket et al. 1599 April 12, 2011:1591–600 Guidelines on Imaging of Asymptomatic CAD and Intervention of the Council on Cardiovascular Radiology and Intervention, and the Councils on Clinical Cardiology and Cardiovascular Disease in the Young. Circulation 2008;118:586–606.
15. Greenland P, Bonow RO, Brundage BH, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography). J Am Coll Cardiol 2007;49:378–403.

16. 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;106:3143– 421.
17. Grundy SM, Cleeman JI, Merz CN, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. Circulation 2004;110:227–39.
18. Dennie CJ, Leipsic J, Brydie A. Canadian Association of Radiologists: consensus guidelines and standards for cardiac CT. Can Assoc Radiol J 2009;60:19 –34.
19. Chow BJ, Larose E, Bilodeau S, et al. The "what, when, where, who and how?" of cardiac computed tomography in 2009: guidelines for the clinician. Can J Cardiol 2009;25:135–9.
20. Hendel RC, Berman DS, Di Carli MF, et al. ACCF/ASNC/ACR/ AHA/ASE/SCCT/SCMR/SNM 2009 appropriate use criteria for cardiac radionuclide imaging: a report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the American Society of Nuclear Cardiology, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the Society of Cardiovascular Computed Tomography, the Society for Cardiovascular Magnetic Resonance, and the Society of Nuclear Medicine. J Am Coll Cardiol 2009;53:2201–29.
21. Douglas PS, Khandheria B, Stainback RF, et al. ACCF/ASE/ACEP/ AHA/ASNC/SCAI/SCCT/SCMR 2008 appropriateness criteria for stress echocardiography: a report of the American College of Cardiology Foundation Appropriateness Criteria Task Force, American Society of Echocardiography, American College of Emergency Physicians, American Heart Association, American Society of Nuclear Cardiology, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance. J Am Coll Cardiol 2008;51: 1127–47.
22. Hendel RC, Patel MR, Kramer CM, Poon M. ACCF/ACR/SCCT/ SCMR/ASNC/NASCI/SCAI/SIR 2006 appropriateness criteria for cardiac computed tomography and cardiac magnetic resonance imaging: a report of the American College of Cardiology Foundation/ American College of Radiology, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance, American Society of Nuclear Cardiology, North American Society for Cardiac Imaging, Society for Cardiovascular Angiography and Interventions, and Society of Interventional Radiology. J Am Coll Cardiol 2006;48:1475–97.
23. Sox HC Jr., Littenberg B, Garber AM. The role of exercise testing in screening for coronary artery disease. Ann Intern Med 1989;110: 456 – 69.
24. Pilote L, Pashkow F, Thomas JD, et al. Clinical yield and cost of exercise treadmill testing to screen for coronary artery disease in asymptomatic adults. Am J Cardiol 1998;81:219 –24.
25. Stason WB, Fineberg HV. Implications of alternative strategies to diagnose coronary artery disease. Circulation 1982;66:III80–6.
26. O'Malley PG, Greenberg BA, Taylor AJ. Cost-effectiveness of using electron beam computed tomography to identify patients at risk for clinical coronary artery disease. Am Heart J 2004;148:106 –13.
27. Shaw LJ, Raggi P, Berman DS, Callister TQ. Cost-effectiveness of screening for cardiovascular disease with measures of coronary calcium. Prog Cardiovasc Dis 2003;46:171– 84.
28. Naghavi M, Falk E, Hecht HS, et al. From vulnerable plaque to vulnerable patient—Part III: executive summary of the Screening for Heart Attack Prevention and Education (SHAPE) Task Force report. Am J Cardiol 2006;98:2H–15H.
29. U.S. National Institutes of Health clinicaltrials.gov web site. 2010. Available at: http://clinicaltrials.gov/. Accessed March 22, 2010.
30. Shaw LJ, Min JK, Budoff M, et al. Induced cardiovascular procedural costs and resource consumption patterns after coronary artery calcium screening: results from the EISNER (Early Identification of Subclinical Atherosclerosis by Noninvasive Imaging Research) study. J Am Coll Cardiol 2009;54:1258–67.
31. Young LH, Wackers FJ, Chyun DA, et al. Cardiac outcomes after screening for asymptomatic coronary artery disease in patients with type 2 diabetes: the DIAD study. A randomized controlled trial. JAMA 2009;301:1547–55.

32. Oxman AD, Glasziou P, Williams JW Jr. What should clinicians do when faced with conflicting recommendations? BMJ 2008;337:a2530.
33. Steyerberg EW. Updating for a new setting. Clinical prediction models. A Practical Approach to Development, Validation, and Updating. New York, NY: Springer, 2009:361–90.
34. Greenland P, LaBree L, Azen SP, Doherty TM, Detrano RC. Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. JAMA 2004;291:210 –5.
35. Vliegenthart R, Oudkerk M, Hofman A, et al. Coronary calcification improves cardiovascular risk prediction in the elderly. Circulation 2005;112:572–7.
36. Detrano R, Guerci AD, Carr JJ, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. N Engl J Med 2008;358:1336–45.
37. Balady GJ, Larson MG, Vasan RS, Leip EP, O'Donnell CJ, Levy D. Usefulness of exercise testing in the prediction of coronary disease risk among asymptomatic persons as a function of the Framingham risk score. Circulation 2004;110:1920 –5.
38. Dupont WD, Plummer WD Jr. Understanding the relationship between relative and absolute risk. Cancer 1996;77:2193–9.

Chapter 3.2 (supplement)

Editorial: "Actually, It Is More of a Guideline Than a Rule"

J Am Coll Cardiol. 2011 Apr 12;57(15):1601-3.

Roger S. Blumenthal
Rani K. Hasan

In a classic scene from the iconic movie *Ghostbusters* (1984, directed by Ivan Reitman), Dr. Peter Venkman (played by Bill Murray) confesses to Dana Barrett (played by Sigourney Weaver), "I make it a rule never to get involved with possessed people." However, after Dana starts to seduce him, Dr. Venkman replies, "Actually, it is more of a guideline than a rule!"

Although 90% to 95% of the global population attributable risk for myocardial infarction has been ascribed to 9 modifiable risk factors (1), clinical risk prediction models for hard events (myocardial infarction, cardiac death) remain suboptimal. In the United States, 40% to 60% of myocardial infarctions and sudden death occur as unheralded first manifestations of atherosclerotic cardiovascular disease (ASCVD) (2).

Current U.S. guidelines for identifying and treating people at increased risk for ASCVD events with proven therapies, such as aggressive lipid lowering and aspirin, are based on the Framingham Risk Score (FRS), which is derived from several generations of long-suffering Caucasian Red Sox fans in Massachusetts—clearly, a unique group. While useful and widely accepted as an office-based risk assessment tool, the Adult Treatment Panel (ATP) III version of the FRS has demonstrated limitations in predicting the risk of a major atherosclerotic disease event, particularly among patients with a family history of premature ASCVD and metabolic syndrome phenotype. Misclassification of risk results in both under- and overtreatment of many persons on the basis of the current ATP III guidelines (3–5).

A number of noninvasive imaging modalities for assessing the degree of subclinical atherosclerosis have been proposed to improve cardiac risk prediction. Professional organizations have sought to evaluate the literature and distill the available data into a summary form that is succinct and conveniently accessible to the busy clinician: hence, the guideline.

In this issue of the *Journal*, Ferket *et al.* (6) present a systematic review of the guidelines on imaging of asymptomatic coronary artery disease (CAD) published by major professional organizations between 2003 and early 2010. The authors based their search on the Institute of Medicine definition for clinical practice guidelines, and they limited their selections to guidelines that included recommendations for imaging for primary prevention of CAD in presumably healthy non-diabetic populations. They also performed an assessment of the guideline generation process for each of the included guidelines using the 7-item Rigor of Development domain of the Appraisal of Guidelines Research and Evaluation (AGREE) instrument.

Fourteen guidelines published by the U.S. Preventive Services Task Force (USPSTF), the New Zealand Guidelines Group (NZGG), the American Heart Association (AHA), the American College of Cardiology Foundation (ACCF), the National Cholesterol Education Program (NCEP), the Canadian Cardiovascular Society (CCS), and the Canadian Society of Radiologists (CSR) were included in the review, with AGREE rigor scores ranging from 93% (most rigorous) to 21% (least rigorous). Imaging modalities considered among these guidelines included resting and exercise electrocardiography, stress echocardiography, myocardial perfusion imaging (single-positron emission computed tomography and positron emission tomography), computed tomography (CT) and magnetic resonance angiography, and CT coronary artery calcium scoring (CAC).

The authors found wide variability with regard to consideration of these testing modalities, with most guidelines recommending against or noting insufficient evidence for the majority of noninvasive imaging modalities, with the only positive recommendations made in reference to intermediate-risk or selected higher-risk populations. Logistic regression analysis suggested no relationships between the likelihood of a guideline recommending for or against testing and the AGREE rigor score or the proportion of guideline panelists with reported industry relationships. Hence, industry relationships did not appear to have any bearing on the directionality of guideline content in this exploratory analysis.

The only imaging modality that was addressed by a majority of the included guidelines was CAC, with 10 of the 14 guidelines making specific recommendations about this modality as an adjunct to current risk prediction. Among the intermediate-risk population, 1 guideline (CSR) made a favorable recommendation for use of CAC, 4 guidelines (ACCF, AHA, NCEP, CCS) recommended consideration of CAC, and 3 guidelines (USPSTF, NZGG, ACCF) found insufficient evidence to make a recommendation. The 10 guidelines were unanimous in recommending against CAC among very low- and high-risk subjects.

The authors note that while there were widespread inconsistencies and several low AGREE rigor scores among the included guidelines, there was general support for consideration of CAC among intermediate-risk subjects. After a cogent discussion of the limitations of the present review, Ferket *et al.* (6) explore the paucity of randomized controlled trial (RCT) data for early detection of CAD, and touch on the challenges in generating such data.

While we agree with the authors that RCT data on this topic is lacking, the challenges and potential pitfalls in pursuing such studies are substantial. These challenges include a large sample size and expense as well as a complicated study design. Most would advocate randomizing the study population to receive the screening test or not and then utilizing various intensities of lipid-lowering therapy based on how the screening test might influence perceived risk.

Should the basis for trial inclusion be the presence of multiple ASCVD risk factors or by FRS criteria? Should lipid-lowering therapy be mandated by the study protocol, and if so, what lipid-lowering algorithm should be used? Should such a study be placebo controlled as opposed to comparing various intensities of lipid-lowering therapy? Given that statin therapy is now generally used aggressively by many clinicians for intermediate-risk patients even without imaging tests, the power to show an incremental gain when adding an imaging test may prove to be limited.

It is likely that most men over the age of 55 years and women over the age of 65 years will have some degree of coronary calcification. Are we then obligated to treat them with at least a low-dose statin and aspirin if they have below average CAC for their age, and use a high dose of a potent statin only for those with Agatston scores >100 or for those with scores >75th percentile for their age?

How will costs and adverse effects of additional testing on quality of life be measured? How does one track the cost and emotional anguish surrounding incidental findings such as non-calcified lung nodules that must be followed up by 2 or more chest CT scans over the next 18 to 24 months to document stability or make a diagnosis of possible malignancy?

Conversely, how does one measure the potential increased motivation for lifestyle change and increased compliance with medication when patients can directly visualize advanced atherosclerosis in their coronary arteries? Appropriate design, interventions, and duration will prove critical and costly for such a trial, but necessary. Efforts by the National Heart, Lung, and Blood Institute to address this need are ongoing.

One note of caution is that a suboptimally designed or executed RCT could be very confusing to clinicians and to the public. Recently, the DIAD (Diagnostic Imaging in Asymptomatic Diabetics) trial assessed whether screening of diabetic patients with single-positron emission computed tomography myocardial perfusion imaging could enhance detection of those with high ASCVD risk and whether the detection of this risk was associated with an improvement in clinical outcomes (7).

Although 22% of subjects in this study had an abnormal myocardial perfusion imaging, only 6% of the defects were moderate or large, and there was no ability to detect persons with advanced subclinical atherosclerosis without overt ischemia. As a result, there was only a slightly higher rate of coronary angiography in the group that was screened, but there was no difference in the intensification of secondary prevention measures in the screening group. Ultimately, there was no difference in clinical events between the screened and unscreened arms. One wonders if CAC testing had been employed and identified persons with advanced subclinical atherosclerosis for their age, it would likely have resulted in more appropriate use of aggressive secondary prevention measures, as recently demonstrated by Nasir *et al.* (8) in a multiethnic population with elevated CAC scores.

In addition to its ability to identify persons with advanced subclinical atherosclerosis for their age, the absence of CAC has been associated with a very low risk of cardiac events over the next 5 years (9). That provides a rationale to emphasize lifestyle changes and scale back on expensive high-potency statins and focus on generic statin therapy if the low-density lipoprotein cholesterol is >130 mg/dl despite improved dietary and exercise habits. Clinicians may also decide to refrain from ordering stress imaging tests in the setting of atypical chest discomfort.

Some have suggested that one should restrict aggressive pharmacotherapy to patients with at least moderate subclinical atherosclerosis to lower the number-needed-to-treat for expensive pharmacotherapy. For example, in the ASCOT (Anglo-Scandinavian Cardiac Outcomes Trial), 93 adults would have needed to be treated for a mean of 3.3 years to prevent a single cardiac event (10). Could the use of atherosclerosis imaging have targeted those patients likely to benefit from aggressive pharmacotherapy and reduced the number-needed-to-treat?

No randomized, controlled, interventional outcome trial evaluating aggressive lipid and blood pressure lowering as well as aspirin therapy guided by the FRS exists. The FRS has become widely accepted as a reasonable approach to CAD risk prediction and stratification in guiding primary prevention on the basis of prospective, observational cohort data. This is certainly not the first instance in which a medical intervention or strategy is widely used or accepted in the absence of RCT evidence of mortality benefit: consider the examples of nitrates for myocardial ischemia, furosemide for decompensated heart failure, oxygen for hypoxemia, or parachutes for free fall (11).

Conservative guidelines often withhold support for a therapeutic or diagnostic intervention in the absence of robust RCT evidence. However, the absence of such evidence is not equivalent to evidence against selective use of a diagnostic test. In the case of imaging asymptomatic CAD, ample prospective cohort and observational data support the consideration of the use of CAC in improving risk prediction in appropriate groups of patients (12,13), as currently reflected by several of the guidelines reviewed by Ferket *et al.* (6).

In a recently published study of 5,878 participants in the MESA (Multi-Ethnic Study of Atherosclerosis) with 209 CAD events over 6 years of follow-up, CAC correctly reclassified one-half of the subjects deemed at intermediate risk into both higher- and lower-risk categories (14). The statistical methods in this study demonstrated that CAC had an additive effect on risk prediction, with accurate reclassification and improved estimation of risk, with the caveats that this study used the FRS modified to account for ethnicity for initial risk prediction and truncated 10-year risk categories for the 5-year follow-up.

Hence, selective use of CAC in intermediate-risk populations may prove beneficial in avoiding over-treatment of some persons while identifying others who previously would not have qualified for intensive pharmacologic and lifestyle prevention efforts based on FRS alone. There is generally a low risk with and low cost for aspirin and generic statin therapy, and there is a very low radiation risk of a single CT scan for CAC screening (~1 to 1.5 mSv). Withholding the potential benefits of selective screening of persons at indeterminate risk, such as those with family history of premature cardiovascular disease and at least a 6% risk of a myocardial infarction over the next decade in the absence of RCT evidence of improved mortality, seems overly dogmatic.

Guidelines are intended to help clinicians navigate issues and challenges in patient care by application of the best available evidence. However, they are not dicta to which all care decisions should perfunctorily adhere. In the case of screening for asymptomatic CAD, a double standard exists for what is the currently accepted standard—the FRS—and for novel strategies for risk prediction, as there are no clamors for an outcome-based RCT of FRS-guided interventions.

Given the difficulties inherent in generating robust RCT evidence, judicious application of the available evidence from well-executed prospective, observational cohort studies is needed to continue to improve risk prediction and primary prevention of CAD events. We must avoid the "cognitive dissonance" that often impedes forward progress and confines guidelines to the necessity of the RCT to optimize the care of our patients in light of the available evidence.

In summary, when there's something wrong in the neighborhood of cardiovascular risk prediction, who should we collectively call (upon)? The data in the preceding text argue for thoughtful interpretation of the available evidence as summarized in current guidelines rather than narrow adherence to rules. We think that Dr. Venkman would agree.

References

1. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364:937–52.
2. Gibbons RJ, Jones DW, Gardner TJ, et al. The American Heart Association's 2008 statement of principles for healthcare reform. Circulation 2008;118:2209 –18.
3. Schlendorf KH, Nasir K, Blumenthal RS. Limitations of the Framingham risk score are now much clearer. Prev Med 2009;48:115– 6.
4. DeMazumder D, Hasan RK, Blumenthal RS, et al. Should statin therapy be allocated on the basis of global risk or on the basis of randomized trial evidence? Am J Cardiol 2010;106:905–9.
5. Berger JS, Jordan CO, Lloyd-Jones D, Blumenthal RS. Screening for cardiovascular risk in asymptomatic patients. J Am Coll Cardiol 2010;55:1169 –77.
6. Ferket BS, Genders TSS, Colkesen EB, et al. Systematic review of guidelines on imaging of asymptomatic coronary artery disease. J Am Coll Cardiol 2011;57:1591– 600.
7. Wackers FJ, Young LH. Lessons learned from the Detection of Ischemia in Asymptomatic Diabetics (DIAD) study. J Nucl Cardiol 2009;16:855–9.
8. Nasir K, McClelland RL, Blumenthal RS, et al. Coronary artery calcium in relation to initiation and continuation of cardiovascular preventive medications: the Multi-Ethnic Study of Atherosclerosis (MESA). Circ Cardiovasc Qual Outcomes 2010;3:228 –35.
9. Blaha M, Budoff MJ, Shaw LJ, et al. Absence of coronary artery calcification and all-cause mortality. J Am Coll Cardiol Img 2009;2: 692–700.
10. Sever PS, Dahlof B, Poulter NE, et al. Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial–Lipid Lowering Arm (ASCOT-LLA): a multicentre randomized controlled trial. Lancet 2003;36:1149 –58.
11. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ 2003;327:1459–61.
12. Lakoski SG, Greenland P, Wong ND, et al. Coronary artery calcium scores and risk for cardiovascular events in women classified as "low risk" based on Framingham risk score: the Multi-Ethnic Study of Atherosclerosis (MESA). Arch Intern Med 2007;167:2437– 42.
13. Scheuner MT, Setodji CM, Pankow JS, et al. General cardiovascular risk profile identifies advanced coronary artery calcium and is improved by family history: the Multi-Ethnic Study of Atherosclerosis. Circ Cardiovasc Genet 2010;3:97–105.
14. Polonsky TS, McClelland RL, Jorgensen NW, et al. Coronary artery calcium score and risk classification for coronary heart disease prediction. JAMA 2010;303:1610–6.

Chapter 4

Diagnostic Performance of Stress Myocardial Perfusion Imaging for Coronary Artery Disease: A systematic Review and Meta-Analysis

European Radiology (In press)

Marcus de Jong
Tessa S.S. Genders
Adriaan Moelker
Robert-Jan M. van Geuns
M.G. Myriam Hunink

Abstract

Objectives

To determine and compare the diagnostic performance of stress myocardial perfusion imaging (MPI) for the diagnosis of obstructive coronary artery disease (CAD), using conventional coronary angiography (CCA) as the reference standard.

Methods

We searched Medline and Embase for literature that evaluated stress MPI for the diagnosis of obstructive CAD using magnetic resonance imaging (MRI), contrast-enhanced echocardiography (ECHO), single-photon emission computed tomography (SPECT) and positron emission tomography (PET).

Results

All pooled analyses were based on random-effects models. Articles on MRI yielded a total of 2970 patients from 28 studies, articles on ECHO yielded a sample size of 795 from 10 studies, articles on SPECT yielded 1323 from 13 studies. For CAD defined as either ≥50%, ≥70% or ≥75% lumen diameter reduction on CCA, the natural logarithms of the diagnostic odds ratio (lnDOR) for MRI (3.63, 95%CI: 3.26–4.00) was significantly higher compared to that of SPECT (2.76, 95%CI: 2.28–3.25, *P*=0.006) and that of ECHO (2.83, 95%CI: 2.29–3.37, *P*=0.02). There was no significant difference between the lnDOR of SPECT and ECHO (*P*=0.52)

Conclusion

Our results suggest that MRI is superior for the diagnosis of obstructive CAD compared to ECHO and SPECT. ECHO and SPECT demonstrated similar diagnostic performance.

Introduction

Coronary artery disease (CAD) is one of the major causes of mortality and morbidity throughout the world (1). The initial assessment of a patient with chest pain usually consists of a stress ECG (electrocardiogram). However, its diagnostic accuracy is low (2) compared to conventional coronary angiography (CCA), which is the reference standard for diagnosing CAD. CCA on the other hand, is an invasive technique and carries a small risk of complications (3, 4). Myocardial perfusion imaging (MPI) is a non-invasive technique that is used clinically as a gatekeeper test before CCA.

MPI can be conducted using stress magnetic resonance imaging (MRI), contrast-enhanced echocardiography (ECHO), single-photon emission computed tomography (SPECT), positron emission tomography (PET) and, under development, computed tomography (CT). The only available extensive study directly comparing two techniques is the MR-IMPACT study (5), a multicentre randomised trial which found that MRI is superior to SPECT. Systematic reviews and meta-analyses have been published for most of the techniques but none of these reviews compare MPI techniques (6-10). The comparability between these different meta-analyses is questionable mainly because of differences in publication period, searching the literature, selection of the evidence, and analysis of the data. Furthermore, studies with verification bias are often included in these reports which may have overestimated the sensitivity and underestimated the specificity of the tests considered. To overcome these problems a systematic review of different MPI techniques is required using the same selection criteria and methods of analysis for all techniques and excluding studies with (potential) verification bias, to make a fair comparison between these imaging tests.

The aim of this study was to determine and compare the diagnostic performance of stress MPI tests for the diagnosis of obstructive CAD, with conventional CCA as the reference standard. We performed the review according to the PRISMA statement for such reviews (11, 12).

Methods

Search strategy

We searched Medline and Embase for English language literature published between January 2000 and May 2011 evaluating the presence of obstructive CAD by stress perfusion imaging tests, namely MRI, contrast-enhanced ECHO, SPECT and PET. In this meta-analysis we focus on functional imaging tests evaluating perfusion as measure of haemodynamically significant myocardial ischemia as opposed to anatomical imaging tests, such as coronary CT angiography, which evaluates structural abnormalities of the coronary arteries. We limited the search to publications from 2000 onwards to include only studies that evaluated state-of-the art MPI techniques. This may have introduced a selection bias with respect to SPECT, because many SPECT studies were published before 2000. To deal with this problem we compare our results with a review of meta-analyses of SPECT studies by Heijenbrok-Kal *et al.* (13). CT was excluded because it is still being developed technically. Review articles were checked for potential additional studies. The search included keywords corresponding to the four index tests (MRI, ECHO, SPECT and PET), the reference test (CCA), the target condition (CAD) and diagnostic performance. We used numerous syn-

onyms including both 'text words' and MeSH (Medical Subject Headings) terms to maximise the sensitivity of our search. See Appendix A in the Electronic Supplementary Material for a detailed description of the search strategy.

Study selection

Two authors reviewed article titles and abstracts for eligibility. Discrepancies were resolved by consensus.

We included studies if they met all of the following criteria: 1) the study assessed diagnostic performance of stress perfusion MRI, stress perfusion contrast-enhanced ECHO, stress perfusion SPECT, or stress perfusion PET as diagnostic test for CAD, 2) a prospective study design was used, 3) the study population consisted of known

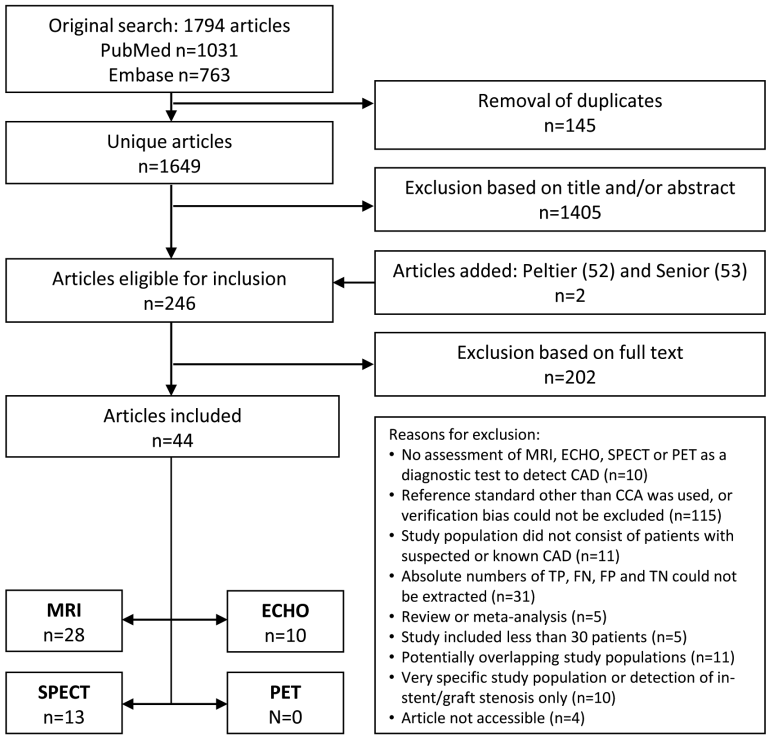


Figure 1. Flowchart of systematic literature search

MRI = Magnetic resonance imaging, ECHO = Echocardiography, SPECT = Single photon emission computed tomography, PET = Positron emission tomography, CAD = Coronary artery disease, CCA = conventional coronary angiography, TP = True positive, FN = False negative, FP = False positive, TN = True negative (previously diagnosed) or suspected adult CAD patients, 4) CCA was used as the reference standard test in all patients irrespective of the non-invasive test result, that is, selective verification was not present, 5) obstructive CAD was defined as ≥1 vessel with ≥50%, ≥70% or ≥75% lumen diameter reduction and, 6) absolute numbers of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) were available at the patient level or could be derived adequately.

Studies were excluded if they met one of the following criteria: 1) the article was a review or meta-analysis, 2) patients had (suspected) acute coronary syndrome (ACS), 3) normal healthy volunteers or asymptomatic patients were included, 4) less than 30 patients were

included (criterion to avoid TPs, FPs, TNs or FNs of zero), 5) (potentially) overlapping study populations were reported, 6) a very specific patient population (e.g. only patients with a heart transplant, left bundle branch block or aortic stenosis) was studied, 7) the study focused on in-stent or graft stenosis after percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG).

Data extraction

Two authors independently extracted data on: author, journal, year of publication, technique used, country, hospital type, number of patients, mean age, percentage male, patient selection, brand of imaging device, magnetic field strength, radiotracer, contrast agent used, type of assessment (qualitative or quantitative), stressor used, CAD definition and the numbers of TP, FP, TN and FN. Discrepancies were resolved by consensus.

If a study reported pairs of sensitivities and specificities at different cut-off points, we extracted the pair with the highest sensitivity. When studies reported data for multiple CAD definitions (e.g. ≥50% and ≥70% stenosis), the highest sensitivity was used to calculate the overall estimates. This also applied when studies reported sensitivities and specificities for different observers.

Quality assessment

We used a modified QUADAS checklist (quality assessment of studies of diagnostic performance included in systematic reviews) (14) to assess the quality of included studies. Two authors independently assessed the study quality of the included articles. Discrepancies were resolved by consensus.

Statistical analysis and data synthesis

We analysed the data at the patient level using a bivariate random effects regression model (15). The model assumes a binomial distribution of the within-study variability (variability between sensitivity and specificity within a study). The model furthermore assumes correlated normally distributed random effects between studies. The degree of correlation between the logit sensitivity and logit specificity corresponds to the inverse relation between sensitivity and specificity when the positivity criterion is varied. Additionally, meta-regression was performed to explore the effect of differences in patient selection and CAD disease definition, taking into account the possible interaction between differences in CAD disease definition and the techniques considered.

The data of each study was summarised in forest plots and summary estimates with a 95% confidence interval of sensitivity and specificity for each imaging technique. Additionally, we summarised these numbers in receiver operator characteristic (ROC) spaces showing the summary estimates with a 95% confidence region and a summary curve. To distinguish SPECT studies that used different protocols, we highlighted the studies that combined gated-SPECT with the use of ^{99m}Technetium as radiotracer (Figure 4). Similarly, MRI studies that included the assessment of delayed contrast enhancement were highlighted. Figures were created using Cochrane's Review Manager (version 5, Copenhagen, Denmark).

To estimate the clinical utility of each technique we calculated the positive and negative likelihood ratios (LR+ and LR-). The likelihood ratio is equivalent to the ratio of the likelihood of a certain test result in patients with the disease and the likelihood of the same test result

in those without the disease. LR+ (= sensitivity / (1 - specificity)) describes the likelihood when the test is positive and LR- (= (1 - sensitivity) / specificity) describes the likelihood when the test is negative. To illustrate the clinical utility, we used the LRs to calculate post-test probabilities across the range of possible pre-test probabilities (Figure 5).

Finally, we calculated the natural logarithm of the diagnostic odds ratio (lnDOR). The lnDOR represents an overall summary estimate of diagnostic performance. The diagnostic odds ratio (DOR) is the odds of positive test results in patients with disease compared to the odds of positive test results in those without disease which equals the ratio of the positive and negative likelihood ratios.

We also created funnel plots to assess the presence of publication bias. The funnel plot shows the DOR horizontally and the standard error of the log transformed DOR vertically. Publication bias usually occurs when negative publications (in our case studies with a low DOR) with a small sample size are not published. An asymmetric funnel plot, for example one with fewer studies in the lower left part of the graph, suggests the presence of publication bias.

The statistical software package SAS (Proc NLMIXED, SAS v9.2, Raleigh, NC, USA) was used for the analyses.

Results

Medline (PubMed) and Embase searches yielded 1649 unique studies (Figure 1). Based on title and abstract we excluded 1405 articles. Based on the full text, we excluded 202 for various reasons detailed in Figure1. Review of the study characteristics shows considerable differences between the included studies (Table 1).

Forty-four studies met the inclusion criteria. Articles on MRI yielded a total of 2970 patients from 28 studies, articles on ECHO yielded a sample size of 795 from 10 studies, articles on SPECT yielded 1323 from 13 studies. We could not include any PET studies, which is why PET was excluded from the analysis. The overview of the QUADAS checklist for all studies demonstrates some differences in terms of study quality (see Appendix B in the Electronic Supplementary Material). The funnel plots of MRI and SPECT suggest evidence for publication bias, whereas the funnel plot of ECHO shows no obvious evidence for publication bias (Figure 2). The sensitivities and specificities of each study vary across studies with sample sizes ranging from 30 to 823 (Tables 2 and 3). The forest plots show the sensitivities and specificities of each study with their 95% confidence intervals depicted as horizontal lines (Figure 3), grouped by CAD definition and study population and then sorted by sensitivity.

Summary estimates

Compared with coronary angiography the meta-analysis of the sensitivities and specificities of the different techniques (Table 3, Figure 3) resulted for MRI in a sensitivity of 0.91 (95%CI: 0.88–0.93) and a specificity of 0.80 (95%CI: 0.76–0.83). Perfusion ECHO showed a sensitivity of 0.87 (95%CI: 0.81–0.91) and a specificity of 0.72 (95%CI: 0.56–0.83). SPECT demonstrated a sensitivity of 0.83 (95%CI: 0.73–0.89) and a specificity of 0.77

Table 1. Study characteristics

Author	Journal	Year	Technique	Country	Type	Patients (n)	Mean age	Age SD	% male	Patients**	Brand	Tesla	Perfusion tests	Assessment	Stressor	CAD definition
Arnold, *et al.* (21)	JACC Cardiovasc Imaging	2010	MRI	UK	N	65	64	9	65	S&K	Siemens	3	rest, stress, DE	qualitative	adenosine	≥50%
Bernhardt, *et al.* (22)	JACC Cardiovasc Imaging	2009	MRI	Germany and Canada	A	823	64	12	76	S&K	Philips	1.5	stress, DE*	qualitative	adenosine	≥70%
Cheng, *et al.* (23)	J Am Coll Cardiol	2007	MRI	UK	A	61	64	8	75	S&K	Siemens	3	rest, stress	qualitative	adenosine	≥50%
Cury, *et al*. (24)	Radiology	2006	MRI	Brasil	N	46	63	5	81	S&K	GE	1.5	stress, DE	qualitative	dipyrdamole	≥70%
Donati, *et al.* (25)	Am J Roentgenol	2010	MRI	Switserland	A	65	64	9	81	S&K	Philips	1.5	rest, stress, DE	qualitative	adenosine	>50%
Doyle, *et al*. (26)	J Cardiovasc Magn Reson	2003	MRI	USA	A	184	59	11	0	NS	Philips	1.5	rest, stress	semi-quantitative	dipyridamole	≥70%
Gebker, *et al*. (27)	Radiology	2007	MRI	Germany	N	40	61	8	70	S&K	Philips	1.5	rest, stress, DE	qualitative	adenosine	≥50%
Gebker, *et al*. (28)	Radiology	2008	MRI	Germany	N	101	62	8	70	S&K	Philips	3	rest, stress, DE	qualitative	adenosine	≥50%
Gebker, *et al*. (29)	Int J Cardiol	2011	MRI	Germany	N	78	65	10	76	S&K	Philips	1.5	rest, stress, DE	qualitative	dobutamine/ atropine	≥70%
Giang, *et al*. (30)	Eur Heart J	2004	MRI	Switzerland	A	44	58	NA	78	S&K	GE	1.5	Stress	semi-quantitative	adenosine	≥50%
Kawase, *et al*. (31)	Osaka City Med J	2004	MRI	Japan	N	50	67	12	58	NS	Philips	1.5	rest, stress	qualitative	nicorandil	≥70%
Kitagawa, et. al. (32)	European radiology	2008	MRI	Japan	A	50	65	9	72	S&K	GE	1.5	rest, stress, DE	qualitative	ATP	≥50%
Klein, *et al.* (33)	J Cardiovasc Magn Reson	2008	MRI	Germany	N	51	60	10	65	NS	Philips	1.5	rest, stress, DE	qualitative	adenosine	>50%
Klein, *et al.* (34)	JACC Cardiovasc Imaging	2009	MRI	UK and Germany	A	78	66	8	90	K	Philips	1.5	Stress, DE	qualitative	adenosine	>50%
Klem, *et al.* (35)	J Am Coll Cardiol	2006	MRI	USA	A	92	58	12	49	S	Siemens	1.5	rest, stress, DE*	qualitative	adenosine	≥50% and ≥70% (≥50 LM)
Klumpp, *et al.* (36)	Eur Radiol	2010	MRI	Germany	A	57	62	11	82	S&K	Siemens	3	rest, stress, DE	qualitative	adenosine	>70%
Krittayaphong et. al. (37)	Int J Cardiovasc Imaging	2009	MRI	Thailand	A	66	61	12	58	S	Philips	1.5	rest, stress	semi-quantitative	adenosine	≥50%
Merkle, *et al*. (38)	Heart	2007	MRI	Germany	A	228	61	11	79	S&K	Philips	1.5	rest, stress, DE	qualitative	adenosine	>50% and >70%
Meyer, *et al.* (39)	Eur Radiol	2008	MRI	Germany	A	60	59	10	63	S&K	Philips	3	rest, stress, DE	qualitative	adenosine	≥70%

Nagel, *et al.* (40)	Circulation	2003	MRI	Germany	A	84	63	8	81	S	Philips	1.5	rest, stress	semi-quantitative	adenosine	≥75%‡
Paetsch, *et al.* (41)	Circulation	2004	MRI	Germany	N	79	61	9	66	S&K	Philips	1.5	rest, stress	qualitative	adenosine	>50%
Pilz, *et al.* (42)	Clin Res Cardiol	2006	MRI	Germany	A	171	62	12	63	S&K	GE	1.5	rest, stress, DE	qualitative	adenosine	>70%
Pingitore, *et al.* (43)	Am J Cardiol	2008	MRI	Italy	A	93	61	NA	70	S&K	GE	1.5	rest, stress	quantitative	dipyridamole	>50%
Plein, *et al.* (44)	Radiology	2005	MRI	UK	N	82	58	NA	74	S&K	Philips	1.5	rest, stress	semi-quantitative	adenosine	>70%
Plein, *et al.* (45)	Eur Heart J	2008	MRI	Switzerland	A	51	59	10	76	S&K	Philips	1.5	stress, DE	qualitative	adenosine	>50%
Plein, *et al.* (46)	Radiology	2008	MRI	Switzerland and UK	A	33	58	11	73	S&K	Philips	3	stress, DE	qualitative	adenosine	>50%
Stolzmann, *et al.* (47)	Int J Cardiovasc Imaging	2010	MRI	Switserland	A	65	64	10	87	NS	Philips	1.5	rest, stress, DE	semi-quantitative	adenosine	>50%
Takase, *et al.* (48)	Jpn Heart J	2004	MRI	Japan	N	102	66	9	83	S&K	GE	1.5	rest, stress, DE	qualitative	dipyridamole	>50%

Author	Journal	Year	Technique	Country	Type	Patients (n)	Mean age	Age SD	% male	Patients	Brand	Contrast agent	Perfusion tests	Assessment	Stressor	CAD definition
Aggeli, *et al.* (49)	Am J Hypertens	2007	ECHO	Greece	A	50	67	5	68	S	Philips	SonoVue	rest, stress	qualitative	adenosine	≥50%
Arnold, *et al.* (21)	JACC Cardiovasc Imaging	2010	ECHO	UK	N	65	64	9	65	S&K	Philips	Optison	rest, stress	qualitative	adenosine	≥50%
Chiou, *et al.* (50)	Can J Cardiol	2004	ECHO	Taiwan	N	132	67	11	75	S&K	Philips	PESDA	rest, stress	qualitative	dobutamine	≥50% (≥40% LM)
Jeetley, *et al.* (51)	J Am Coll Cardiol	2006	ECHO	UK	A/N	123	62	12	71	S&K	Philips	Sonazoid	rest, stress	semi-quantitative	dipyridamole	≥70%
Kowatsch, *et al.* (52)	J Am Soc Echocardiogr	2007	ECHO	Brasil	A	54	60	9	61	S&K	Philips	PESDA	rest, stress	quantitative	adenosine	>50%
Lipiec, *et al.* (53)	J Am Soc Echocardiogr	2008	ECHO	Poland	A	103	58	9	63	S&K	Siemens	Optison	rest, stress	qualitative	dipyridamole or atropine	≥70%
Miszalski-Jamka, *et al.* (54)	Int J Cardiol	2008	ECHO	Poland	A	103	58	NA	80	S&K	Philips	Sonovue	rest, stress	qualitative	exercise	≥50%
Moir, *et al.* (55)	J Am Soc Echocardiogr	2005	ECHO	Australia	A	79	56	NA	80	S&K	Philips	Definity	rest, stress	quantitative	dipyridamole, exercise	≥50%
Peltier *et al.* (56)	J Am Coll Cardiol	2004	ECHO	Belgium	A	35	62	10	71	S&K	Agilent Technologies	PESDA	rest, stress	quantitative	dipyridamole	>70%
Senior, et. al. (57)	American Heart Journal	2004	ECHO	UK, Germany, Belgium	N	53	61†	NA	82	NS	Philips	Sonazoid	rest, stress	qualitative	dipyridamole	>50%

Author	Journal	Year	Technique	Country	Type	Patients (n)	Mean age	Age SD	% male	Patients	Brand	Radiotracer	Perfusion tests	Assessment	Stressor	CAD definition
Aggeli, *et al.* (49)	Am J Hypertens	2007	SPECT	Greece	A	48	67	5	68	S	GE	^{201}Tl	rest, stress	qualitative and qualitative	adenosine	≥50%
Astarita, *et al.* (58)	J Hypertens	2001	SPECT	Italy	N	53	58	10	55	S	Picker	^{201}Tl	rest, stress	qualitative	exercise	≥50%
Budoff, *et al.* (59)	Acad Radiol	2007	SPECT	USA	A	30	54	9	70	NS	NS	^{99m}Tc-MIBI	rest, stress	qualitative	exercise	>70% (>50% LM)
Doyle, *et al.* (26)	J Cardiovasc Magn Reson	2003	SPECT	USA	A	184	59	11	0	NS	ADAC	^{99m}Tc-MIBI/^{201}Tl	rest, stress (gated)	qualitative	dipyridamole	≥70%
Gonzalez, *et al.* (60)	Rev Esp Med Nucl	2005	SPECT	Chile	A	145	60	12	68	S&K	GE	^{201}Tl	rest, stress	qualitative	exercise (n=63), dipyridamole (n=82)	≥50% and ≥75%
Jeetley, *et al.* (51)	J Am Coll Cardiol	2006	SPECT	UK	A/N	123	62	12	71	S&K	NS	^{99m}Tc-MIBI	rest, stress	semi-quantitative	dipyridamole	≥70%
Johansen, *et al.* (61)	J Nucl Cardiol	2005	SPECT	Denmark	A	357	57	9	54	S	Picker	^{201}Tl(rest) and ^{99m}Tc-MIBI (stress)	rest, stress (gated)	semi-quantitative	adenosine or dobutamine (n=180), exercise stress test (n=177)	≥50%
Lipiec, *et al.* (53)	J Am Soc Echocardiogr	2008	SPECT	Poland	A	103	58	9	63	S&K	GE	^{99m}Tc-MIBI	rest, stress	semi-quantitative	dipyridamole	≥70%
Peltier *et al.* (56)	J Am Coll Cardiol	2004	SPECT	Belgium	A	35	62	10	71	S&K	GE	^{99m}Tc-MIBI	rest, stress	qualitative	dipyridamole	>70%
Schepis, *et al.* (62)	J Nucl Med	2007	SPECT	Switserland	A	77	66	9	62	S	GE	^{99m}Tc-tetrofosmin	rest, stress (gated)	semiquantitative	adenosine	≥50%
Senior, et. al. (57)	American Heart Journal	2004	SPECT	UK, Germany, Belgium	N	53	61†	NA	82	NS	Amersham Health	^{99m}Tc-tetrofosmin	rest, stress	qualitative	dipyridamole	>50%
Yao, *et al.* (63)	Nucl Med Commun	2000	SPECT	China	N	64	51	NA	95	S&K	Siemens	^{99m}Tc-MIBI	rest, stress	qualitative	exercise	≥50%
Yeih, *et al.* (64)	J Formos Med Assoc	2007	SPECT	Taiwan	A	51	63	9	0	S&K	GE	^{201}Tl	rest, stress	qualitative	dobutamine	≥50%

A = academic, N = non-academic, NS = not specified, NA = not available, PESDA = perfluorocarbon-exposed sonicated dextrose albumin, S = suspected coronary artery disease only (and no history of myocardial infarction (MI), CABG or PCI), S&K = patients with either suspected or known coronary artery disease, K = known coronary artery disease, ^{99m}Tc = ^{99m}Technetium, -MIBI = Sestamibi, ^{201}Tl = ^{201}Thallium, DE = delayed enhancement, ATP = adenosine triphosphate, LM = left main coronary artery

** In Klem et al. (35) and Bernhardt et al. (22) the index test outcome was initially based on delayed enhancement (DE) images, considering perfusion images when DE was negative.*

† Median

‡ Area reduction

***If there was any uncertainty about the study population it was not classified as suspected and/or known CAD but rather as "not specified"*

(95%CI: 0.64–0.86). The receiver operating characteristic (ROC) spaces show the summary estimates for sensitivity and specificity of each technique two-dimensionally surrounded by its 95% confidence area (Figure 4). The sensitivity of MRI and SPECT differed significantly (P=0.03). In terms of specificity, no significant differences were found.

We found no effect of CAD definition on the sensitivities (Table 3, P=0.55). The disease definition >/≥70% stenosis compared to >/≥50% stenosis resulted in significantly lower specificities for SPECT (Table 3, P=0.045), but no significant differences for ECHO (P=0.39) and MRI (P=0.51). Furthermore, we found no effect of CAD definition on the lnDORs of MRI (P=0.24), ECHO (P=0.96) and SPECT (P=0.34) (Table 3).

Furthermore, MRI, ECHO and SPECT showed no significant differences in terms of sensitivity, specificity and lnDOR when comparing patients with suspected CAD without a prior history of CAD to patients with known or suspected CAD (all P-values >0.05, Table 3).

We did not observe an association between the use of gated-SPECT in combination with ^{99m}Technetium as radiotracer, and the diagnostic performance of SPECT (Figure 4). MRI studies that assessed delayed contrast enhancement were associated with high sensitivities albeit with a wide range of specificities (Figure 4).

The positive likelihood ratios (LR+) of MRI, ECHO and SPECT were 4.43 (95%CI: 3.64–5.23), 3.08 (95%CI: 1.65–4.50) and 3.56 (95%CI: 2.07–5.04) respectively (Table 3). The negative likelihood ratios (LR-) for MRI, ECHO, and SPECT were 0.12 (95%CI: 0.08–0.15), 0.18 (95%CI: 0.13–0.24) and 0.22 (95%CI: 0.14–0.31), respectively. Figure 5 illustrates the revised probability of CAD after a positive and negative test. The lnDORs of MRI, ECHO and SPECT were 3.63 (95%CI: 3.26–4.00), 2.83 (95%CI: 2.29–3.37) and 2.76 (95%CI: 2.28–3.25), respectively (Table 3). We found significantly higher lnDORs for MRI in comparison with SPECT (P=0.006) and ECHO (P=0.02). There was no significant difference between the lnDOR of SPECT and ECHO (P=0.52).

Discussion

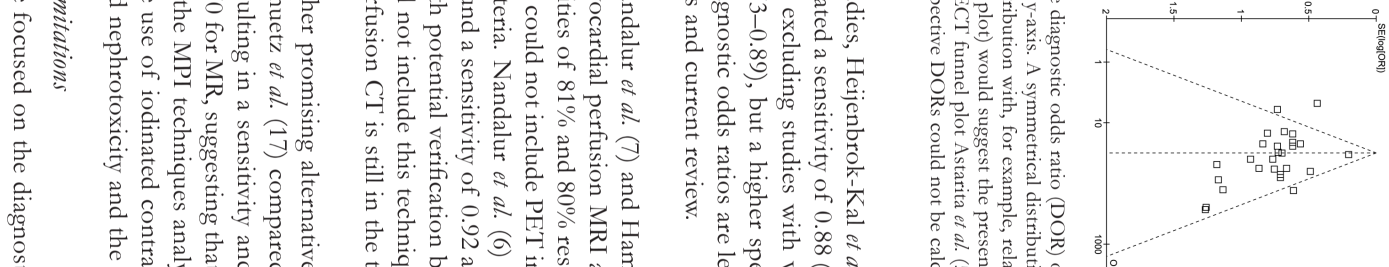
In this systematic review and meta-analysis we compared the diagnostic performance of different stress myocardial perfusion imaging techniques. MRI showed the best diagnostic performance with the narrowest confidence intervals; the latter is explained by the large number of patients studied with MRI. We found a significantly higher sensitivity for MRI compared to SPECT and a significantly higher lnDOR for MRI compared to both ECHO and SPECT. In contrast to previous meta-analyses (9), we compared the different imaging techniques using the same search strategy and methods of analysing the data. Furthermore, we only included studies without verification bias.

In our review we paid special attention to the issue of verification bias. Sensitivity may be overestimated and specificity underestimated if patients with a positive test result are more likely to be verified with the reference standard test. Diagnostic odds ratios are generally not, or only minimally, affected by verification bias (16). Underwood *et al.* (9) reviewed the diagnostic performance of SPECT and explained the overall low specificity (0.70–0.75 for high quality studies) of SPECT studies by verification bias. In their review of SPECT

Table 2. Source data for MRI, ECHO and SPECT

Author	Year	Technique	TP	FN	TN	FP	Sens.	Spec.	CAD definition
Arnold, *et al.* (21)	2010	MRI	37	4	17	4	90.2%	81.0%	≥50%
Bernhardt, *et al.* (22)	2009	MRI	274	39	421	89	87.5%	82.5%	≥70%
Cheng, *et al.* (23)	2007	MRI	39	1	16	5	97.5%	76.2%	≥50%
Cury, *et al.* (24)	2006	MRI	29	1	12	4	96.7%	75.0%	≥70%
Donati, *et al.* (25)	2010	MRI	30	3	14	0	90.9%	100%	>50%
Doyle, *et al.* (26)	2003	MRI	15	11	123	35	57.7%	77.8%	≥70%
Gebker, *et al.* (27)	2007	MRI	19	3	14	4	86.4%	77.8%	≥50%
Gebker, *et al.* (28)	2008	MRI	63	7	22	9	90.0%	71.0%	≥50%
Gebker, *et al.* (29)	2011	MRI	48	4	19	4	92.3%	82.6%	≥70%
Giang, *et al.* (30)	2004	MRI	26	2	12	4	92.9%	75.0%	≥50%
Kawase, *et al.* (31)	2004	MRI	31	2	16	1	93.9%	94.1%	≥70%
Kitagawa, et. al. (32)	2008	MRI	33	3	8	6	91.7%	57.1%	≥50%
Klein, *et al.* (33)	2008	MRI	22	3	23	3	88.0%	88.5%	>50%
Klein, *et al.* (34)	2009	MRI	36	18	21	3	66.7%	87.5%	>50%
Klem, *et al.* (35)	2006	MRI	34	10	42	6	77.3%	87.5%	≥50%
			33	4	48	7	89.2%	87.3%	≥70% (≥50 LM)
Klumpp, *et al.* (36)	2010	MRI	40	1	14	2	97.6%	87.5%	>70%
Krittayaphong et. al. (37)	2009	MRI	34	4	22	6	89.5%	78.6%	≥50%
Merkle, *et al.* (38)	2007	MRI	160	12	48	8	93.0%	85.7%	>50%
			147	6	54	21	96.1%	72.0%	>70%
Meyer, *et al.* (39)	2008	MRI	32	4	19	5	88.9%	79.2%	≥70%
Nagel, *et al.* (40)	2003	MRI	38	5	37	4	88.4%	90.2%	≥75%
Paetsch, *et al.* (41)	2004	MRI	48	5	16	10	90.6%	61.5%	>50%
Pilz, *et al.* (42)	2006	MRI	109	4	48	10	96.5%	82.8%	>70%
Pingitore, *et al.* (43)	2008	MRI	61	5	18	9	92.4%	66.7%	>50%
Plein, *et al.* (44)	2005	MRI	52	7	17	6	88.1%	73.9%	>70%
Plein, *et al.* (45)	2008	MRI	31	4	7	9	88.6%	43.8%	>50%
Plein, *et al.* (46)	2008	MRI	12	1	16	4	92.3%	80.0%	>50%
Stolzmann, *et al.* (47)	2010	MRI	28	8	21	3	77.8%	87.5%	>50%
Takase, *et al.* (48)	2004	MRI	71	5	22	4	93.4%	84.6%	>50%
Aggeli, *et al.* (49)	2007	ECHO	28	4	16	2	87.5%	88.9%	≥50%
Arnold. *et al.* (21)	2010	ECHO	35	6	16	5	85.4%	76.2%	≥50%
Chiou, *et al.* (50)	2004	ECHO	69	16	36	11	81.2%	76.6%	≥50% (≥40% LM)
Jeetley, *et al.* (51)	2006	ECHO	74	11	19	19	87.1%	50.0%	≥70%
Kowatsch, *et al.* (52)	2007	ECHO	22	3	21	8	88.0%	72.4%	>50%
Lipiec, *et al.* (53)	2008	ECHO	69	10	18	6	87.3%	75.0%	≥70%
Miszalski-Jamka, *et al.* (54)	2008	ECHO	65	9	25	4	87.8%	86.2%	≥50%
Moir, *et al.* (55)	2005	ECHO	35	5	20	19	87.5%	51.3%	≥50%
Peltier *et al.* (56)	2004	ECHO	22	0	10	3	100.0%	76.9%	>70%
Senior, *et al.* (57)	2004	ECHO	35	7	7	5	83.3%	58.3%	>50%
Aggeli, *et al.* (49)	2007	SPECT	24	6	17	1	80.0%	94.4%	≥50%
Astarita, *et al.* (58)	2001	SPECT	23	0	14	16	100.0%	46.7%	≥50%
Budoff, *et al.* (59)	2007	SPECT	17	4	7	2	81.0%	77.8%	>70% (>50% LM)
Doyle, *et al.* (26)	2003	SPECT	16	10	130	28	61.5%	82.3%	≥70%
Gonzalez, *et al.* (60)	2005	SPECT	102	15	16	12	87.2%	57.1%	≥50%
			91	7	24	23	92.9%	51.1%	≥75%
Jeetley, *et al.* (51)	2006	SPECT	73	12	19	19	85.9%	50.0%	≥70%
Johansen, *et al.* (61)	2005	SPECT	94	32	183	48	74.6%	79.2%	≥50%
Lipiec, *et al.* (53)	2008	SPECT	73	6	13	11	92.4%	54.2%	≥70%
Peltier *et al.* (56)	2004	SPECT	18	4	11	2	81.8%	84.6%	>70%
Schepis, *et al.* (62)	2007	SPECT	32	10	32	3	76.2%	91.4%	≥50%
Senior, et. al. (57)	2004	SPECT	20	21	11	1	48.8%	91.7%	>50%
Yao, *et al.* (63)	2000	SPECT	42	3	18	1	93.3%	94.7%	≥50%
Yeih, *et al.* (64)	2007	SPECT	20	8	20	3	71.4%	87.0%	≥50%

Figure 2. Funnel plots



The diagnostic odds ratio (DOR) on the x-axis is plotted against the standard error (SE) of the log(DOR) on the y-axis. A symmetrical distribution of studies indicates the absence of publication bias. An asymmetrical distribution with, for example, relatively more smaller studies with a positive result (in the lower right part of the plot) would suggest the presence of publication bias. In the ECHO funnel plot Peltier *et al.* (56), in the SPECT funnel plot Astarita *et al.* (58) and in the MRI funnel plot Donati *et al.* (25) are not included, because their respective DORs could not be calculated (0 false negatives or false positives). A: MRI, B: ECHO, C: SPECT

studies, Heijenbrok-Kal *et al.* (13) did not exclude studies with verification bias and demonstrated a sensitivity of 0.88 (95%CI: 0.87–0.90) and a specificity of 0.73 (95%CI: 0.69–0.74). By excluding studies with verification bias, we found a lower sensitivity of 0.83 (95%CI: 0.73–0.89), but a higher specificity of 0.77 (95%CI: 0.64–0.86). As pointed out above, the diagnostic odds ratios are less affected by verification bias and were the same for the previous and current review.

Nandalur *et al.* (7) and Hamon *et al.* (10) previously studied the diagnostic performance of myocardial perfusion MRI and found sensitivities of 91% and 89% respectively and specificities of 81% and 80% respectively, which is very similar to what we found. Unfortunately we could not include PET in the analysis, because no PET studies met our in- and exclusion criteria. Nandalur *et al.* (6) performed a meta-analysis of PET perfusion studies and they found a sensitivity of 0.92 and a specificity of 0.85. However, their analysis included studies with potential verification bias. Stress perfusion CT is an upcoming MPI technique, but we did not include this technique because of the low number of available studies and because perfusion CT is still in the technical development phase.

Other promising alternatives to CCA are non-invasive CT and MR coronary angiography. Schuetz *et al.* (17) compared CT and MR coronary angiography to CCA in a meta-analysis resulting in a sensitivity and specificity of respectively 0.97 and 0.87 for CT, and 0.87 and 0.70 for MR, suggesting that CT angiography has a better diagnostic performance compared to the MPI techniques analysed in this article. However, drawbacks of CT angiography are the use of iodinated contrast material which poses a small risk of idiosynchratic reactions and nephrotoxicity and the lack of functional information (18).

Limitations

We focused on the diagnostic performance of myocardial perfusion imaging. However, an MPI examination can yield functional information as well (e.g. left ventricular function, presence of wall motion abnormalities, presence of scar tissue), rather than perfusion images alone. Our analysis does not take into account the possible impact of these parameters

Table 3. Measures of diagnostic performance for MRI, ECHO and SPECT, estimated using the bivariate random effects model.

		Sensitivity	Specificity	LR+	LR-	DOR	lnDOR	CAD prevalence*
MRI	**Overall †**	0.91 (0.88 – 0.93)	0.80 (0.76 – 0.83)	4.43 (3.64 – 5.23)	0.12 (0.08 – 0.15)	37.69 (26.00 – 54.63)	3.63 (3.26 – 4.00)	54% (1603/2970)
	Suspected	0.90 (0.78 – 0.96)	0.86 (0.74 – 0.93)	6.61 (2.23 – 10.99)	0.12 (0.03 – 0.22)	54.70 (20.07 – 149.07)	4.00 (3.00 – 5.00)	49% (118/242)
	CAD 50	0.89 (0.86 – 0.92)	0.79 (0.73 – 0.84)	4.25 (3.15 – 5.35)	0.13 (0.09 – 0.17)	31.84 (20.96 – 48.37)	3.46 (3.04 – 3.88)	66% (882/1338)
	CAD 70	0.91 (0.87 – 0.94)	0.82 (0.75 – 0.87)	4.97 (3.47 – 6.47)	0.11 (0.07 – 0.15)	46.40 (28.90 – 74.49)	3.84 (3.36 – 4.31)	48% (937/1952)
ECHO	**Overall †**	0.87 (0.81 – 0.91)	0.72 (0.56 – 0.83)	3.08 (1.65 – 4.50)	0.18 (0.13 – 0.24)	16.94 (9.84 – 29.15)	2.83 (2.29 – 3.37)	66% (525/795)
	Suspected	0.88 (0.60 – 0.97)	0.89 (0.58 – 0.98)	8.35 (6.67 – 21.76)	0.13 (-0.05 – 0.32)	62.76 (7.37 – 534.54)	4.14 (2.00 – 6.28)	64% (32/50)
	CAD 50	0.86 (0.79 – 0.92)	0.74 (0.63 – 0.82)	3.28 (2.09 – 4.47)	0.19 (0.10 – 0.27)	17.59 (9.48 – 32.66)	2.87 (2.25 – 3.49)	63% (339/534)
	CAD 70	0.90 (0.80 – 0.96)	0.65 (0.46 – 0.80)	2.58 (1.32 – 3.84)	0.15 (0.04 – 0.26)	17.04 (6.60 – 44.04)	2.84 (1.89 – 3.79)	71% (186/261)
SPECT	**Overall †**	0.83 (0.73 – 0.89)	0.77 (0.64 – 0.86)	3.56 (2.07 – 5.04)	0.22 (0.14 – 0.31)	15.84 (9.74 – 25.77)	2.76 (2.28 – 3.25)	50% (666/1323)
	Suspected	0.83 (0.70 – 0.91)	0.79 (0.66 – 0.87)	3.88 (2.03 – 5.73)	0.21 (0.09 – 0.34)	18.15 (8.34 – 39.52)	2.90 (2.12 – 3.68)	41% (221/535)
	CAD 50	0.81 (0.72 – 0.87)	0.81 (0.72 – 0.87)	4.15 (2.55 – 5.75)	0.24 (0.15 – 0.33)	17.24 (9.67 – 30.73)	2.85 (2.27 – 3.43)	53% (452/848)
	CAD 70	0.85 (0.76 – 0.91)	0.66 (0.54 – 0.77)	2.53 (1.69 – 3.37)	0.22 (0.12 – 0.33)	11.42 (6.04 – 21.59)	2.44 (1.80 – 3.07)	53% (331/620)

CAD 50 corresponds to the studies that defined obstructive CAD either as >50% or ≥50% stenosis
CAD 70 corresponds to the studies that defined obstructive CAD either as >70%, ≥70% or ≥75% stenosis and studies that combined one of these with >50% (or ≥50%) stenosis in the left main coronary artery
"Suspected" refers to studies that only included patients with suspected CAD without a history of MI, PCI or CABG.
** The CAD prevalence defined by "CAD diagnosed by CCA" divided by the total sample size*
† When data was available for both CAD definitions (≥50% and ≥70%) the overall estimates only include data from CAD ≥70% stenosis

on the interpretation of the myocardial perfusion imaging test and the results of MRI are therefore likely to be even better than we estimated.

Also, it is important to note that in clinical practice a small proportion of patients will be unsuitable for MRI, either due to contra-indications or claustrophobia. Likewise, an echocardiography procedure relies on an adequate acoustic window. Often, unsuitable patients were excluded from the original studies, which in turn could have resulted in an overestimation of the diagnostic performance in our analysis. Unfortunately, the included studies did not report sufficient information to explore these issues.

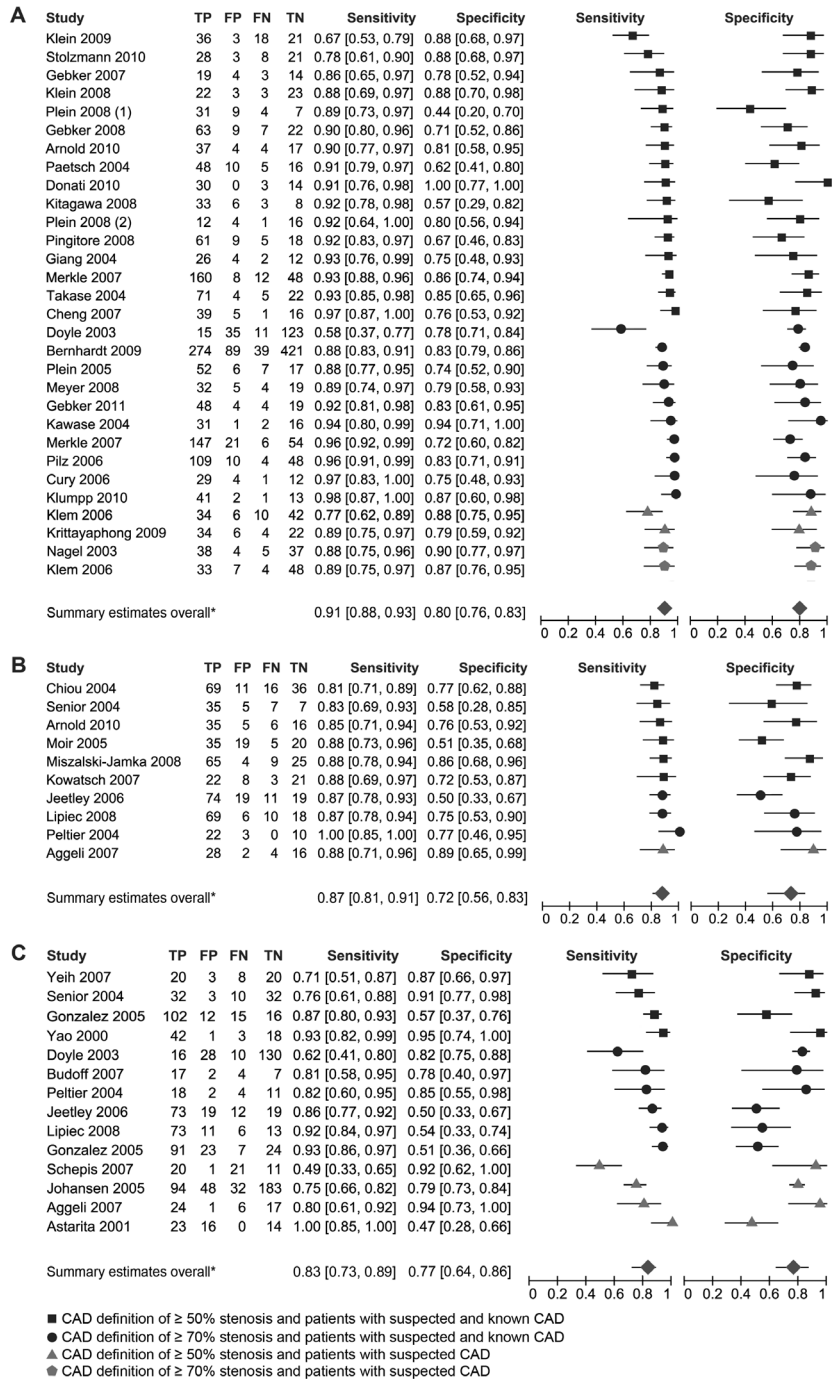
In the current review we included only studies that used the most advanced technology by searching for studies published from 2000 until 2011, which implies that some large landmark SPECT studies performed in the 1980s and 1990s were excluded from our analysis. A previously published comprehensive systematic review sheds light on the effect of this exclusion criterion (13). In the previous review 103 SPECT studies with a total of 11977 patients published between 1984 and 2002 were analysed. There is no overlap with the SPECT studies that we included. The diagnostic odds ratios for SPECT found in the previous review and in the current review are the same: they found an lnDOR of 2.8 (95%CI: 2.6–3.0) compared to our lnDOR of 2.8 (95%CI: 2.3–3.3).

The funnel plot for MRI and SPECT suggests that there is evidence of publication bias, which implies that our summary measures may be overestimated. Nevertheless, the overestimation applies to both MRI and SPECT. The funnel plot for ECHO does not suggest evidence of publication bias.

Heterogeneity across studies is a limitation of meta-analyses of diagnostic performance. Across studies differences exist with respect to imaging techniques, assessment methods, stressors, radiotracers, contrast media, CAD definition (lumen diameter reduction of ≥50%, ≥70% or ≥75%), CAD prevalence, percentage male patients, patient inclusion criteria, setting and country. Although we were able to analyse the effect of using different CAD definitions and patient inclusion criteria, sample size limitations did not allow us to do subset analyses for the other cross-study variations. Due to chance there will always be variability between studies, but there may also be different types of biases influencing the results. We used a random effects model which adjusts the estimates and confidence intervals to account for between-study variations. Nevertheless, heterogeneity across studies remains an important limitation. For calculation and precision purposes, we excluded studies with less than 30 patients. In this way, we minimized the number of studies with for example zero FPs or FNs. This exclusion criterion may have introduced a selection bias. Another limitation of meta-analyses is the dependence on the level of detail reported in the original papers. For example, data on the individual territories was generally not available. Furthermore, most studies included a mix of known and suspected CAD patients or did not report the test characteristics for the subgroup of patients with suspected CAD without a prior history of MI, PCI or CABG. Therefore, our subgroup analysis of suspected CAD was limited due to a small sample size. Nevertheless, our analysis did suggest that the diagnostic performance of MPI tests is not substantially affected by including patients with known CAD.

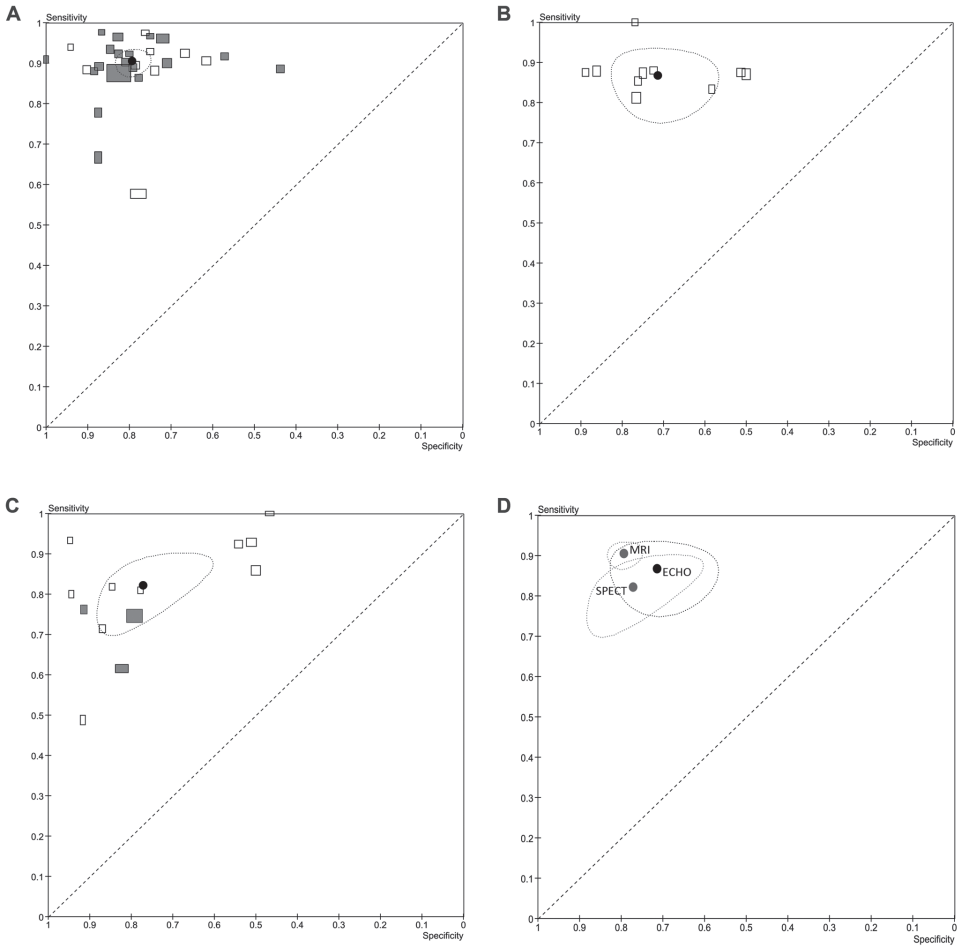
Although our results show all tests are reasonably accurate, the likelihood ratios suggest that neither one of them is suitable to rule out or rule in the presence of disease (19). This can

Figure 3. Forest plots



The data is sorted by suspected and known CAD versus suspected CAD and CAD definition of ≥50% versus ≥70% stenosis from lowest to highest sensitivity and data are reported at the patient level. A: MRI, B: ECHO, C: SPECT *When data was available for both CAD definitions (≥50% and ≥70%) the summary estimates only include data from CAD ≥70% stenosis

Figure 4. ROC space with summary estimates for each technique with 95% confidence areas



This figure shows the diagnostic performance of studies relative to each other with specificity (plotted in reverse) on the x-axis and sensitivity on the y-axis. Perfect diagnostic accuracy is in the upper left corner, where sensitivity and specificity are both 1. A: MRI, B: ECHO, C: SPECT, D: All three techniques. The grey rectangles in (A) refer to the studies using delayed contrast enhancement and in (C) they refer to the studies using gated SPECT with radiotracer ^{99m}Tc. The size of the rectangles corresponds with the inverse standard error of sensitivity and specificity, which correlates with the size of the study.

also be seen in Figure 5, where the post-test probability after a positive test rarely exceeds 90%, and the post-test probability of disease after a negative test may still be substantial. Since MPI is intended as a gatekeeper test, ruling out disease is more important than ruling in disease. MRI performs quite well in this respect with an LR- of 0.12 (0.08 – 0.15). SPECT and ECHO demonstrate less favourable LRs (Table 3).

The reference standard test for diagnosing coronary artery disease is CCA. Innovative technological developments in diagnosing CAD are most often compared with CCA. The limitation of CCA is that it evaluates the lumen diameter reduction of the coronary arteries, but for instance a 50% vessel diameter reduction does not always result in the same reduction in blood flow and does not necessarily lead to myocardial ischemia. There are alternative

techniques such as fractional flow reserve (FFR) that measure the pressure difference across a coronary stenosis. It is even possible that the imaging techniques we evaluated are better diagnostic tools than CCA to begin with, since they measure myocardial perfusion which is the physiological basis of myocardial function. Thus, the less than perfect sensitivity and specificity could in part be attributed to imperfections of CCA instead of the limitations of perfusion imaging.

Clinical implications

The results of our systematic review and meta-analysis suggest that MRI is superior to ECHO and SPECT in diagnosing CAD. This statement is strengthened firstly by the findings of the MR-IMPACT study (5) – a multicentre randomised trial – which suggested that MRI is superior to SPECT and secondly by the findings of the EuroCMR registry (20), which demonstrated that in patients who underwent stress MRI for the diagnostic workup of suspected CAD, invasive angiography could be avoided in nearly one-half of the patients. All-in-all, the results suggest that stress perfusion MRI is potentially useful as a gatekeeper test before CCA in patients with low to intermediate prior probability of CAD but this needs to be confirmed with a comparative cost-effectiveness analysis. Furthermore, more research of the diagnostic performance of stress perfusion ECHO, PET and CT are required to evaluate their clinical usefulness.

In conclusion, our results suggest that stress perfusion MRI is superior for the diagnosis of obstructive coronary artery disease compared to stress perfusion contrast-enhanced echocardiography and SPECT, and that echocardiography and SPECT are similar in terms of diagnostic performance.

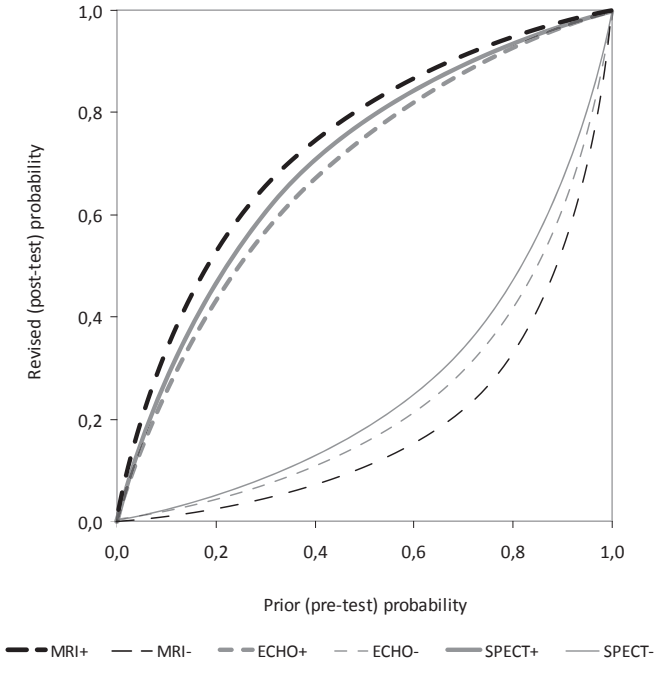


Figure 5. Revised probability of CAD

This figure shows the revised (post-test) probability of CAD (y-axis) as a function of prior (pre-test) probability (x-axis) of CAD for positive and negative MPI results, based on the likelihood ratios presented in Table 3 (overall analysis). MRI+, ECHO+ and SPECT+ represent the lines for a positive test result and MRI-, ECHO- and SPECT- represent the lines for a negative test result.

Appendix A. Medline and Embase search

Index test MRI (1)

"Magnetic Resonance Imaging"[Mesh] OR ("magnetic" AND "resonance"[tw]) OR NMR OR MRI[tw] OR MR[tw]

Index test PET (2)

"Positron-Emission Tomography"[Mesh] OR PET[tw] OR (((positron[tw] AND emission[tw]) OR positron-emitt*[tw]) AND tomograph*[tw]) OR rubidium[tw] OR nitrogen[tw] OR deoxyglucose[tw] OR FDG[tw] OR N-13[tw] OR N13[tw] OR 82rb[tw]

Index test SPECT (3)

("Tomography, Emission-Computed, Single-Photon"[Mesh] OR "Myocardial Perfusion Imaging"[Mesh] OR ((CT[tw] OR CAT[tw] OR (comput*[tw] AND tomograph*[tw])) AND single[tw] AND photon[tw]) OR SPECT[tw] OR SPET[tw] OR (myocard*[tw] AND perfusion[tw] AND imag*[tw]) OR (myocard*[tw] AND scintigraph*[tw]))

Index test ECHO (4)

("Ultrasonography"[Mesh] OR ultrasound[tw] OR ultrason*[tw] OR echocardio*[tw]) AND (perfusion[tw] OR "Perfusion Imaging"[Mesh])

Stressor – index tests (5)

dipyridamole[tw] OR dobutamine[tw] OR adenosine[tw] OR nicorandil[tw] OR stress*[tw] OR perfusion[tw] OR (blood[tw] AND flow[tw]) OR pharmacolog*[tw]

Target condition (6)

"Coronary Disease"[Mesh] OR (coronary[tw] AND (disease[tw] OR stenos*[tw])) OR "Angina Pectoris"[Mesh] OR angina[tw] OR (chest[tw] AND pain[tw]) OR "Myocardial Infarction"[Mesh] OR MI[tw] OR (myocardial[tw] AND infarction*[tw])

Reference test (7)

"Coronary Angiography"[Mesh] OR (coronary[tw] AND cathether*[tw]) OR angiograph*[tw] OR angiogram[tw] OR "Fractional Flow Reserve, Myocardial"[Mesh] OR FFR[tw] OR (fractional[tw] AND flow[tw] AND reserve[tw])

Diagnostic keywords (8)

"sensitivity and specificity"[Mesh] OR sensitivity[tw] OR specificity[tw] OR "predictive value of tests"[Mesh] OR "ROC Curve"[Mesh] OR roc*[tw] OR sroc[tw] OR "receiver operating characteristic"[tw] OR "receiver operator characteristic"[tw] OR "pre-test odds"[tw] OR "pretest odds"[tw] OR (pre-test[tw] AND probabilit*[tw]) OR (pretest[tw] AND probabilit*[tw]) OR "post-test odds"[tw] OR "posttest odds"[tw] OR ("post test"[tw] AND probabilit*[tw]) OR (posttest[tw] AND probabilit*[tw]) OR (likelihood[tw] AND ratio*[tw]) OR ("positive predictive"[tw] AND value*[tw]) OR ("negative predictive"[tw] AND value*[tw]) OR (false[tw] AND negative*[tw]) OR (false[tw] AND positive*[tw]) OR (true[tw] AND negative*[tw]) OR (true[tw] AND positive*[tw]) OR misdiagnosis[tw] OR misdiagnoses OR "Diagnostic Errors"[Mesh] OR (diagnost*[tw] AND (accuracy[tw] OR error*[tw] OR efficacy[tw]))

Limits (9)

(English[lang] NOT ((Animals[Mesh] NOT Humans[Mesh]) OR Editorial[Publication type] OR Comment[Publication type] OR Letter[Publication type] OR Case Reports[Publication type])

Final Medline search

(1 OR 2 OR 3 OR 4) AND 5 AND 6 AND 7 AND 8 AND 9

The Embase search was the same as the Medline search, but with only 'text words' and excluding Medline studies.

Appendix B. The quality checklist

Author	Year	1	2	3	4*	5	6	7	8	9	10	11	12	13	14	15	16
Aggeli, *et al.* (49)	2007	N	Y	Y	<1 month	Y	Y	Y	Y	Y	Y	U	N	Y	N	Y	N
Arnold, *et al.* (21)	2010	Y	Y	Y	<2 weeks	Y	Y	Y	Y	Y	Y	U	N	Y	Y	Y	N
Astarita, *et al.* (58)	2001	N	Y	Y	<15 days	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	N
Bernhardt, *et al.* (22)	2009	Y	Y	Y	<2 weeks (8±5 days)	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N
Budoff, *et al.* (59)	2007	Y	Y	Y	< 1 month	Y	Y	Y	Y	Y	Y	Y	U	N	N	U	N
Cheng, *et al.* (23)	2007	Y	Y	Y	<2 weeks	Y	Y	Y	Y	Y	Y	U	N	Y	Y	Y	N
Chiou, *et al.* (50)	2004	Y	Y	Y	<7 days	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	N
Cury, *et al.* (24)	2006	Y	Y	Y	<2 weeks	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y
Donati, *et al.* (25)	2010	Y	Y	Y	< median of 8 days	Y	Y	Y	Y	Y	Y	Y	N	U	Y	Y	N
Doyle, *et al.* (26)	2003	N	N	Y	NS	Y	Y	Y	Y	Y	U	U	U	Y	Y	U	N
Gebker, *et al.* (27)	2007	Y	Y	Y	<24 hours	Y	Y	Y	Y	Y	Y	U	N	Y	Y	Y	Y
Gebker, *et al.* (28)	2008	Y	Y	Y	<48 hours	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
Gebker, *et al.* (29)	2011	Y	Y	Y	<90 days	Y	Y	Y	Y	Y	Y	U	N	Y	Y	U	N
Giang, *et al.* (30)	2004	Y	Y	Y	<30 days	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N
Gonzalez, *et al.* (60)	2005	Y	N	Y	mean 27 days (range: 1-180)	Y	Y	Y	Y	Y	Y	U	N	N	N	Y	N
Jeetley, *et al.* (51)	2006	Y	N	Y	4 weeks	Y	Y	Y	U	Y	Y	Y	N	N	N	Y	N
Johansen, *et al.* (61)	2005	Y	Y	Y	<3 month	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N
Kawase, *et al.* (31)	2004	Y	Y	Y	<1 weeks	Y	Y	Y	Y	Y	Y	U	N	N	N	Y	N
Kitagawa, et. Al. (32)	2008	Y	Y	Y	<14 days	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y
Klein, *et al.* (33)	2008	Y	Y	Y	<24 hours	Y	Y	Y	Y	N	Y	Y	N	Y	N	Y	Y
Klein, *et al.* (34)	2009	N	Y	Y	<24 hours	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	N
Klem, *et al.* (35)	2006	Y	Y	Y	<24 hours	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N
Klumpp, *et al.* (36)	2010	Y	Y	Y	NS	Y	Y	Y	Y	Y	U	U	N	Y	N	Y	N
Kowatsch, *et al.* (52)	2007	Y	Y	Y	<1 month	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	N
Krittayaphong et. Al. (37)	2009	Y	Y	Y	<1 weeks	Y	Y	Y	Y	Y	Y	Y	U	N	Y	U	Y
Lipiec, *et al.* (53)	2008	Y	Y	Y	<14 days	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	N
Merkle, *et al.* (38)	2007	Y	Y	Y	<4 weeks	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N
Meyer, *et al.* (39)	2008	Y	Y	Y	<28 days	Y	Y	Y	Y	Y	Y	U	U	Y	N	U	N
Miszalski-Jamka, *et al.* (54)	2008	Y	N	Y	<15 days	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	N
Moir, *et al.* (55)	2005	Y	Y	Y	NS	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	N
Nagel, *et al.* (40)	2003	Y	Y	Y	NS	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	N
Paetsch, *et al.* (41)	2004	Y	Y	Y	NS	Y	Y	Y	Y	Y	Y	Y	U	Y	N	U	N
Peltier *et al.* (56)	2004	Y	Y	Y	NS	Y	Y	Y	Y	Y	Y	U	N	Y	N	Y	N
Pilz, *et al.* (42)	2006	Y	Y	Y	NS	Y	Y	Y	Y	Y	Y	U	U	Y	Y	U	N

Pingitore, *et al*. (43)	2008	Y	Y	Y	<15 days	Y	Y	Y	Y	Y	U	Y	U	Y	Y	U	N
Plein, *et al*. (44)	2005	Y	Y	Y	< mean of 4.3 days ±12	Y	Y	Y	Y	Y	Y	Y	N	U	Y	Y	Y
Plein, *et al*. (45)	2008	Y	Y	Y	<median of 4 days	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	U
Plein, *et al*. (46)	2008	Y	N	Y	<14 days	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
Schepis, *et al*. (62)	2007	Y	Y	Y	<2 weeks	Y	Y	Y	Y	Y	Y	Y	U	Y	N	U	N
Senior, et. al. (57)	2004	Y	Y	Y	<4 weeks	Y	Y	Y	Y	Y	Y	U	U	N	N	U	N
Stolzmann,*et al*. (47)	2010	Y	Y	Y	<14 days	Y	Y	Y	Y	Y	Y	Y	U	U	Y	U	N
Takase, *et al*. (48)	2004	Y	Y	Y	<1 month	Y	Y	Y	Y	Y	Y	U	N	N	N	Y	N
Yao, *et al*. (63)	2000	Y	N	Y	<3 weeks	Y	Y	Y	Y	Y	Y	U	U	N	N	U	N
Yeih, *et al*. (64)	2007	N	Y	Y	<2 weeks	Y	Y	Y	Y	Y	Y	U	U	N	N	U	N

Y= yes; N = no; U = unclear; NS = not specified
** The time period between the index and reference tests as reported is tabulated to allow the reader to make his/her own judgment on this issue.*

1: Was the spectrum of patients representative of the patients who will receive the test in practice?
2: Were selection criteria clearly described?
3: Is the reference standard likely to correctly classify the target condition?
4: Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?*
5: Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
6: Did patients receive the same reference standard regardless of the index test result?
7: Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
8: Was the execution of the index test described in sufficient detail to permit replication of the test?
9: Was the execution of the reference standard described in sufficient detail to permit its replication?
10: Were the index test results interpreted without knowledge of the results of the reference standard?
11: Were the reference standard results interpreted without knowledge of the results of the index test?
12: Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
13: Were uninterpretable/intermediate test results reported?
14: Were withdrawals from the study explained?
15: Were the scans read blind to clinical data?
16: Was the expertise of the clinician assessing the results of the diagnostic tests reported?

References

1. Murray CJ, Lopez AD (1997) Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. Lancet, 349:1436-1442.
2. Kwok Y, Kim C, Grady D, Segal M, Redberg R (1999) Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol, 83:660-666.
3. Noto TJ, Jr., Johnson LW, Krone R, et al. (1991) Cardiac catheterization 1990: a report of the Registry of the Society for Cardiac Angiography and Interventions (SCA&I). Cathet Cardiovasc Diagn, 24:75-83.
4. Scanlon PJ, Faxon DP, Audet AM, et al. (1999) ACC/AHA guidelines for coronary angiography. A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on Coronary Angiography). Developed in collaboration with the Society for Cardiac Angiography and Interventions. J Am Coll Cardiol, 33:1756-1824.
5. Schwitter J, Wacker CM, van Rossum AC, et al. (2008) MR-IMPACT: comparison of perfusion-cardiac magnetic resonance with single-photon emission computed tomography for the detection of coronary artery disease in a multicentre, multivendor, randomized trial. Eur Heart J, 29:480-489.
6. Nandalur KR, Dwamena BA, Choudhri AF, Nandalur SR, Reddy P, Carlos RC (2008) Diagnostic performance of positron emission tomography in the detection of coronary artery disease: a meta-analysis. Acad Radiol, 15:444-451.
7. Nandalur KR, Dwamena BA, Choudhri AF, Nandalur MR, Carlos RC (2007) Diagnostic performance of stress cardiac magnetic resonance imaging in the detection of coronary artery disease: a meta-analysis. J Am Coll Cardiol, 50:1343-1353.
8. Abdelmoneim SS, Dhoble A, Bernier M, et al. (2009) Quantitative myocardial contrast echocardiography during pharmacological stress for diagnosis of coronary artery disease: a systematic review and meta-analysis of diagnostic accuracy studies. Eur J Echocardiogr, 10:813-825.
9. Underwood SR, Anagnostopoulos C, Cerqueira M, et al. (2004) Myocardial perfusion scintigraphy: the evidence. Eur J Nucl Med Mol Imaging, 31:261-291.
10. Hamon M, Fau G, Nee G, Ehtisham J, Morello R (2010) Meta-analysis of the diagnostic performance of stress perfusion cardiovascular magnetic resonance for detection of coronary artery disease. J Cardiovasc Magn Reson, 12:29.
11. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ, 339:b2535.
12. Liberati A, Altman DG, Tetzlaff J, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ, 339:b2700.
13. Heijenbrok-Kal MH, Fleischmann KE, Hunink MG (2007) Stress echocardiography, stress single-photon-emission computed tomography and electron beam computed tomography for the assessment of coronary artery disease: a meta-analysis of diagnostic performance. Am Heart J, 154:415-423.
14. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol, 3:25.
15. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MG, Stijnen T (2008) Bivariate random effects meta-analysis of ROC curves. Med Decis Making, 28:621-638.
16. Knottnerus JA (1987) The effects of disease verification and referral on the relationship between symptoms and diseases. Med Decis Making, 7:139-148.
17. Schuetz GM, Zacharopoulou NM, Schlattmann P, Dewey M (2010) Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. Ann Intern Med, 152:167-177.

18. Einstein AJ, Henzlova MJ, Rajagopalan S (2007) Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. JAMA, 298:317-323.
19. Jaeschke R, Guyatt GH, Sackett DL (1994) Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA, 271:703-707.
20. Bruder O, Schneider S, Nothnagel D, et al. (2009) EuroCMR (European Cardiovascular Magnetic Resonance) registry: results of the German pilot phase. J Am Coll Cardiol, 54:1457-1466.
21. Arnold JR, Karamitsos TD, Pegg TJ, et al. (2010) Adenosine stress myocardial contrast echocardiography for the detection of coronary artery disease: a comparison with coronary angiography and cardiac magnetic resonance. JACC Cardiovasc Imaging, 3:934-943.
22. Bernhardt P, Spiess J, Levenson B, et al. (2009) Combined assessment of myocardial perfusion and late gadolinium enhancement in patients after percutaneous coronary intervention or bypass grafts: a multicenter study of an integrated cardiovascular magnetic resonance protocol. JACC Cardiovasc Imaging, 2:1292-1300.
23. Cheng AS, Pegg TJ, Karamitsos TD, et al. (2007) Cardiovascular magnetic resonance perfusion imaging at 3-tesla for the detection of coronary artery disease: a comparison with 1.5-tesla. J Am Coll Cardiol, 49:2440-2449.
24. Cury RC, Cattani CA, Gabure LA, et al. (2006) Diagnostic performance of stress perfusion and delayed-enhancement MR imaging in patients with coronary artery disease. Radiology, 240:39-45.
25. Donati OF, Scheffel H, Stolzmann P, et al. (2010) Combined cardiac CT and MRI for the comprehensive workup of hemodynamically relevant coronary stenoses. AJR Am J Roentgenol, 194:920-926.
26. Doyle M, Fuisz A, Kortright E, et al. (2003) The impact of myocardial flow reserve on the detection of coronary artery disease by perfusion imaging methods: an NHLBI WISE study. J Cardiovasc Magn Reson, 5:475-485.
27. Gebker R, Jahnke C, Paetsch I, et al. (2007) MR myocardial perfusion imaging with k-space and time broad-use linear acquisition speed-up technique: feasibility study. Radiology, 245:863-871.
28. Gebker R, Jahnke C, Paetsch I, et al. (2008) Diagnostic performance of myocardial perfusion MR at 3 T in patients with coronary artery disease. Radiology, 247:57-63.
29. Gebker R, Jahnke C, Manka R, et al. (2011) High spatial resolution myocardial perfusion imaging during high dose dobutamine/atropine stress magnetic resonance using k-t SENSE. Int J Cardiol. Doi:S0167-5273(11)00107-0 (pii)10.1016/j.ijcard.2011.01.060 (doi)
30. Giang TH, Nanz D, Coulden R, et al. (2004) Detection of coronary artery disease by magnetic resonance myocardial perfusion imaging with various contrast medium doses: first European multi-centre experience. Eur Heart J, 25:1657-1665.
31. Kawase Y, Nishimoto M, Hato K, Okajima K, Yoshikawa J (2004) Assessment of coronary artery disease with nicorandil stress magnetic resonance imaging. Osaka City Med J, 50:87-94.
32. Kitagawa K, Sakuma H, Nagata M, et al. (2008) Diagnostic accuracy of stress myocardial perfusion MRI and late gadolinium-enhanced MRI for detecting flow-limiting coronary artery disease: a multicenter study. Eur Radiol, 18:2808-2816.
33. Klein C, Gebker R, Kokocinski T, et al. (2008) Combined magnetic resonance coronary artery imaging, myocardial perfusion and late gadolinium enhancement in patients with suspected coronary artery disease. J Cardiovasc Magn Reson, 10:45.
34. Klein C, Nagel E, Gebker R, et al. (2009) Magnetic Resonance Adenosine Perfusion Imaging in Patients After Coronary Artery Bypass Graft Surgery. JACC Cardiovasc Imaging, 2:437-445.
35. Klem I, Heitner JF, Shah DJ, et al. (2006) Improved Detection of Coronary Artery Disease by Stress Perfusion Cardiovascular Magnetic Resonance With the Use of Delayed Enhancement Infarction Imaging. J Am Coll Cardiol, 47:1630-1638.

36. Klumpp BD, Seeger A, Doesch C, et al. (2010) High resolution myocardial magnetic resonance stress perfusion imaging at 3 T using a 1 M contrast agent. Eur Radiol., 20:533-541.
37. Krittayaphong R, Boonyasirinant T, Saiviroonporn P, et al. (2009) Myocardial perfusion cardiac magnetic resonance for the diagnosis of coronary artery disease: do we need rest images? Int J Cardiovasc Imaging, 25 Suppl 1:139-148.
38. Merkle N, Wohrle J, Grebe O, et al. (2007) Assessment of myocardial perfusion for detection of coronary artery stenoses by steady-state, free-precession magnetic resonance first-pass imaging. Heart, 93:1381-1385.
39. Meyer C, Strach K, Thomas D, et al. (2008) High-resolution myocardial stress perfusion at 3 T in patients with suspected coronary artery disease. Eur Radiol., 18:226-233.
40. Nagel E, Klein C, Paetsch I, et al. (2003) Magnetic resonance perfusion measurements for the noninvasive detection of coronary artery disease. Circulation, 108:432-437.
41. Paetsch I, Jahnke C, Wahl A, et al. (2004) Comparison of dobutamine stress magnetic resonance, adenosine stress magnetic resonance, and adenosine stress magnetic resonance perfusion. Circulation, 110:835-842.
42. Pilz G, Bernhardt P, Klos M, Ali E, Wild M, Hofling B (2006) Clinical implication of adenosine-stress cardiac magnetic resonance imaging as potential gatekeeper prior to invasive examination in patients with AHA/ACC class II indication for coronary angiography. Clin Res Cardiol, 95:531-538.
43. Pingitore A, Lombardi M, Scattini B, et al. (2008) Head to head comparison between perfusion and function during accelerated high-dose dipyridamole magnetic resonance stress for the detection of coronary artery disease. Am J Cardiol, 101:8-14.
44. Plein S, Radjenovic A, Ridgway JP, et al. (2005) Coronary artery disease: myocardial perfusion MR imaging with sensitivity encoding versus conventional angiography. Radiology, 235:423-430.
45. Plein S, Kozerke S, Suerder D, et al. (2008) High spatial resolution myocardial perfusion cardiac magnetic resonance for the detection of coronary artery disease. Eur Heart J, 29:2148-2155.
46. Plein S, Schwitter J, Suerder D, Greenwood JP, Boesiger P, Kozerke S (2008) k-Space and time sensitivity encoding-accelerated myocardial perfusion. Radiology, 249:493-500.
47. Stolzmann P, Alkadhi H, Scheffel H, et al. (2010) Combining cardiac magnetic resonance and computed tomography coronary calcium scoring: added value for the assessment of morphological coronary disease? Int J Cardiovasc Imaging, 27: 969-977.
48. Takase B, Nagata M, Kihara T, et al. (2004) Whole-heart dipyridamole stress first-pass myocardial perfusion MRI for the detection of coronary artery disease. Jpn Heart J, 45:475-486.
49. Aggeli C, Christoforatou E, Giannopoulos G, et al. (2007) The diagnostic value of adenosine stress-contrast echocardiography for diagnosis of coronary artery disease in hypertensive patients: comparison to Tl-201 single-photon emission computed tomography. Am J Hypertens, 20:533-538.
50. Chiou KR, Huang WC, Lin SL, et al. (2004) Real-time dobutamine stress myocardial contrast echocardiography for detecting coronary artery disease: correlating abnormal wall motion and disturbed perfusion. Can J Cardiol, 20:1237-1243.
51. Jeetley P, Hickman M, Kamp O, et al. (2006) Myocardial contrast echocardiography for the detection of coronary artery stenosis: a prospective multicenter study in comparison with single-photon emission computed tomography. J Am Coll Cardiol, 47:141-145.
52. Kowatsch I, Tsutsui JM, Osorio AF, et al. (2007) Head-to-head comparison of dobutamine and adenosine stress real-time myocardial perfusion echocardiography for the detection of coronary artery disease. J Am Soc Echocardiogr, 20:1109-1117.
53. Lipiec P, Wejner-Mik P, Krzeminska-Pakula M, et al. (2008) Accelerated stress real-time myocardial contrast echocardiography for the detection of coronary artery disease: comparison with 99mTc single photon emission computed tomography. J Am Soc Echocardiogr, 21:941-947.
54. Miszalski-Jamka T, Kuntz-Hehner S, Schmidt H, et al. (2009) Impact of previous myocardial infarction on the incremental value of myocardial contrast to two-dimensional supine bicycle stress echocardiography in evaluation of coronary artery disease. Int J Cardiol, 136:47-55.

55. Moir S, Haluska BA, Jenkins C, McNab D, Marwick TH (2005) Myocardial blood volume and perfusion reserve responses to combined dipyridamole and exercise stress: A quantitative approach to contrast stress echocardiography. J Am Soc Echocardiogr, 18:1187-1193.
56. Peltier M, Vancraeynest D, Pasquet A, et al. (2004) Assessment of the physiologic significance of coronary disease with dipyridamole real-time myocardial contrast echocardiography. Comparison with technetium-99m sestamibi single-photon emission computed tomography and quantitative coronary angiography. J Am Coll Cardiol, 43:257-264.
57. Senior R, Lepper W, Pasquet A, et al. (2004) Myocardial perfusion assessment in patients with medium probability of coronary artery disease and no prior myocardial infarction: Comparison of myocardial contrast echocardiography with 99mTc single-photon emission computed tomography. Am Heart J, 147:1100-1105.
58. Astarita C, Palinkas A, Nicolai E, Maresca FS, Varga A, Picano E (2001) Dipyridamole-atropine stress echocardiography versus exercise SPECT scintigraphy for detection of coronary artery disease in hypertensives with positive exercise test. J Hypertens, 19:495-502.
59. Budoff MJ, Rasouli ML, Shavelle DM, et al. (2007) Cardiac CT angiography (CTA) and nuclear myocardial perfusion imaging (MPI)-a comparison in detecting significant coronary artery disease. Acad Radiol, 14:252-257.
60. Gonzalez P, Massardo T, Jofre MJ, et al. (2005) 201Tl myocardial SPECT detects significant coronary artery disease between 50% and 75% angiogram stenosis. Rev Esp Med Nucl, 24:305-311.
61. Johansen A, Hoilund-Carlsen PF, Christensen HW, et al. (2005) Diagnostic accuracy of myocardial perfusion imaging in a study population without post-test referral bias. J Nucl Cardiol, 12:530-537.
62. Schepis T, Gaemperli O, Koepfli P, et al. (2007) Added value of coronary artery calcium score as an adjunct to gated SPECT for the evaluation of coronary artery disease in an intermediate-risk population. J Nucl Med, 48:1424-1430.
63. Yao Z, Liu XJ, Shi RF, et al. (2000) A comparison of 99Tcm-MIBI myocardial SPET and electron beam computed tomography in the assessment of coronary artery disease in two different age groups. Nucl Med Commun, 21:43-48.
64. Yeih DF, Huang PJ, Ho YL (2007) Enhanced diagnosis of coronary artery disease in women by dobutamine thallium-201 ST-segment/heart rate slope and thallium-201 myocardial SPECT. J Formos Med Assoc, 106:832-839.

Chapter 5

Imaging Strategies for Acute Chest Pain in the Emergency Department

American Journal of Roentgenology (In press)

Admir Dedic
Tessa S.S. Genders
Koen Nieman
M.G. Myriam Hunink

The imaging question

The entity acute coronary syndrome (ACS) encompasses the conditions unstable angina pectoris, and myocardial infarction with or without ST-segment elevation. Myocardial infarction is a diagnosis based on patient symptoms, electrocardiographic (ECG) changes, and markers of myocardial necrosis in the blood (1). Unstable angina pectoris, on the other hand, indicates myocardial ischemia without biochemical evidence of cardiac myocyte death (2).

Patients with ACS frequently present with atypical chest pain complaints, unremarkable physical examination and an ECG that is either poorly interpretable or has normalized at presentation (3). Cardiac biomarkers are often normal during the initial phase and patients with unstable angina may not show a rise of these markers at all. In these patients, ACS cannot be ruled out based on the initial assessment alone, which generally requires clinical observation and sequential testing. Optimal triage requires a quick noninvasive test that is cost-effective and readily available to identify all patients with ACS. Moreover, a triage test should also be able to accurately identify patients in whom significant coronary artery disease (CAD) can be excluded and who can thus safely be sent home. This article quantitatively examines existing evidence on the diagnostic performance of imaging tests in this setting.

Background and importance

Acute chest pain represents a common diagnostic dilemma in the emergency department (ED) and its impact on the health care system is substantial, with an estimated annual cost of several billions of dollars in the US (4). ED physicians and cardiologists are commonly required to make a decision whether or not to admit a patient with chest pain, based on little more than clinical judgment and a rough estimation of risk, but without conclusive evidence whether or not an ACS is developing. Most patients presenting to the ER with sudden chest pain do not suffer from an ACS, and many are in fact free of any coronary artery disease (CAD) (5). The vast majority of chest pain complaints can be explained by other causes, like gastro-esophageal diseases and chest-wall syndromes (6).

Nevertheless, the risk of overlooking an underlying ACS has major consequences, which is why most patients are hospitalized to undergo additional stress testing and even invasive catheter coronary angiography (CAG) to rule out ischemic heart disease. A large multicenter study demonstrated that while the majority of patients with suspected ACS were hospitalized for further evaluation, only 17% was ultimately diagnosed with ACS (7).

Despite this defensive approach, the literature suggests that an estimated 2-6% of the patients sent home were found to have an ACS (8-12). Patients mistakenly discharged from the emergency department generally have a worse prognosis than appropriately managed patients, partly because of their risk for sudden cardiac death but also because of the delay in implementing treatments that are known to be beneficial for ACS (8, 11). In the USA, an estimated 25% of the law suits concerning emergency care involves errors in the diagnosis of myocardial infarction (13).

Synopsis and Synthesis of Evidence

Echocardiography, radionuclide myocardial perfusion imaging (MPI) and coronary CT angiography (CTA) are the three main imaging techniques employed in clinical practice for the diagnosis of ACS. Echocardiography and MPI are functional imaging modalities used to assess wall motion abnormalities (WMA) and myocardial perfusion. Performed in rest, they are used to identify ACS. They can also be performed during or after stress to detect inducible ischemia. Coronary CTA is an anatomical technique that can visualize atherosclerotic plaque in the coronary tree, and provide information on its composition and degree of stenosis. A discussion of each technique follows, with a summary of these imaging techniques shown in Table 1.

Echocardiography

Echocardiography is a non-invasive, portable and relatively inexpensive bedside imaging technique that is available in most hospitals. It can assess both left ventricular systolic function and regional WMA and therefore provides valuable diagnostic as well as prognostic information. Because of the close relationship between wall motion and myocardial blood flow (MBF) echocardiography is a very useful tool in patients suspected of ACS. Development of WMA is preceded by considerable reductions in MBF or myocardial infarction.

Echocardiography is also useful in the assessment of many non-ischemic causes of acute chest pain such as perimyocarditis, valvular heart disease, cardiomyopathy, pulmonary embolism, or aortic dissection. It is the preferred imaging method for detecting complications of acute infarction including myocardial free wall rupture, acute ventricular septal defect, and mitral regurgitation secondary to papillary muscle rupture or ischemia.

Rest echocardiography

When echocardiography is performed shortly after a patient's arrival to the ED or during a chest pain episode, WMAs are detected in up to 90% of the cases (14). However, chest pain has subsided in the majority of patients at the time of evaluation and they may have a completely normal resting echocardiogram. A large study reported rest echocardiography findings of 901 patients with acute chest pain, but no clinical manifestations of acute MI, as part of protocol-driven care along with serial creatine kinase (CK-MB) measurement and continuous ECG-monitoring. Rest echocardiography was associated with high specificity (99%, 873/882) but unsatisfactory sensitivity (47%, 9/19) for adverse events within 30 days, including MI, revascularization or (un)stable angina (15). Smaller studies have shown similar results (16-17).

Di Pasquale *et al.* performed rest echocardiography in 280 patients, presenting with chest pain of suspected cardiac origin, but normal initial CK-MB levels and no evidence of ST-elevation or new left bundle branch block on their ECG (18). The presence of WMAs was used to predict significant stenosis (>50% left main coronary stenosis or >70% stenosis in other branches) on CAG. In this high risk population, significant CAD being present in 50%, the authors found once again high specificity of 91% (84/92), but also a sensitivity of 93% (170/182).

Contrast enhanced echocardiography

Microbubble contrast agents enhance delineation of endocardial borders, which facilitates wall motion assessment (Figure 1). It also provides important information on myocardial perfusion and viability (19). A prospective analysis by Kang *et al.* of 114 patients with cardiac chest pain and no clinical manifestations of acute MI, demonstrated that the addition of contrast improved the sensitivity of echocardiography from 49% (43/87) to 77% (67/87) while specificity was similar, 78% (74/95) vs. 73% (69/95), for the detection of ACS at the index visit (16). In a large observational study 1017 patients underwent contrast echocardiography in addition to routine clinical evaluation. For the detection of a composite endpoint of CAD within 48 hours, the authors demonstrated a high sensitivity of 89% (148/166) and a modest specificity of 57% (485/851) (20).

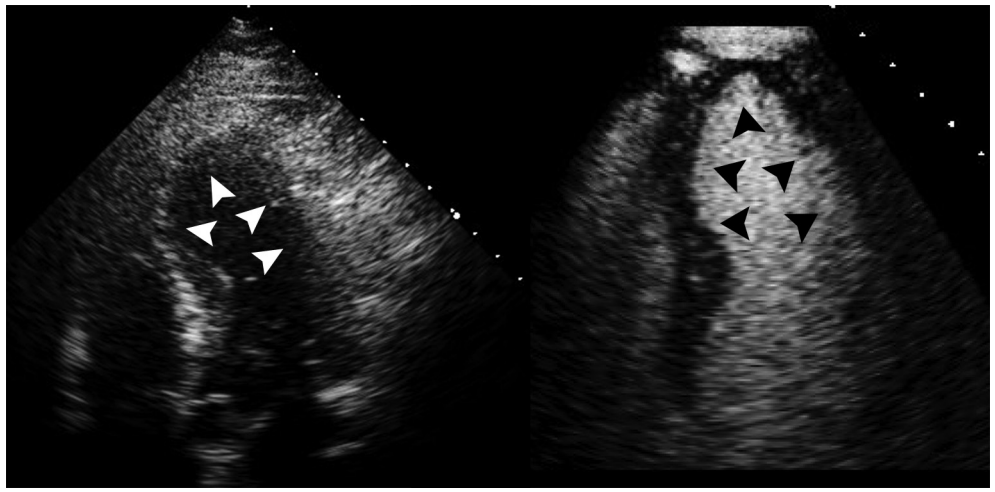


Figure 1. A 67 year-old male suspected of myocardial ischemia. Microbubble contrast agents are used to enhance delineation of endocardial borders (white arrowheads indicate endocardial borders on the non-enhanced echo, black arrowheads indicate improved delineation of endocardial borders by using contrast agents).

Stress echocardiography

Generally, excluding only myocardial infarction will not be sufficient to safely discharge patients. Patients may have myocardial ischemia with unstable angina pectoris, putting them at risk of adverse events. As soon as serial cardiac markers and rest imaging has excluded the presence of AMI, stress echocardiography, during pharmacological stress or exercise, may be used to visualize inducible ischemia.

Trippi and colleagues investigated the diagnostic performance of stress echocardiography in 163 patients with normal initial markers and normal resting echocardiogram. The authors demonstrated a sensitivity of 90% (17/19) and specificity of 89% (128/144) for the detection of AMI or significant (>50%) CAD on invasive angiography (21). Another study compared stress echocardiography to stress MPI in 503 patients without evidence of acute MI after 6 hours of observation and initial work-up (22). The authors used a composite endpoint consisting of ≥50% coronary stenosis on CAG or cardiac events during a follow-up of 6 months. Echocardiography had a sensitivity and specificity of 85% (65/77) and 95% (405/426), while MPI had a sensitivity and specificity of 86% (66/77) and 90% (383/426).

Limitations

There are some key concerns regarding the use of echocardiography in the ED. Evaluation of WMA remains a subjective and difficult skill to master. Interpreters should therefore be experienced readers. Secondly, a considerable number of patients have a poor thoracic imaging window resulting in indeterminate results in at least 10% (23). Discriminating between existing and newly developed WMA is difficult, and both conditions may even co-exist in the same patient. Acute MI should be excluded before performing stress echocardiography, by measuring serial cardiac markers or rest imaging. This results in a longer diagnostic work-up.

Myocardial perfusion imaging

Radionuclide myocardial perfusion imaging (MPI) provides a direct assessment of myocardial blood flow (MBF) and is used for identification of ischemia or infarction. The injected radionuclide agents are transported through the coronary vasculature and eventually accumulate in the myocardium. At the moment of MBF impairment, MPI shows perfusion defects which provide early detection of obstructive CAD. Pioneer work from three decades ago demonstrated that impaired myocardial uptake of thallium-201 (TI-201) could be visualized on planar images in patients with myocardial infarction (24). By now, TI-201 has largely been replaced by technetium-99m based agents, which are associated with less scatter and blurring. Images could be acquired over a longer time because of slow myocardial clearance, overcoming some logistic barriers (25). Replacements of planar imaging by single photon emission computed tomography resulted in improved visualization of the location as well as extent of disease (26). Lastly, introduction of gated reconstructions permitted regional and global ventricular function assessment (27).

Rest MPI

The value of rest MPI in the patient with acute chest pain has been studied extensively. Numerous studies have reported sensitivities of >90% for the detection of myocardial infarction, accompanied with specificities of 50-80% (28-30). One study conducted at the Medical College of Virginia in 620 patients with suspected ACS reported a sensitivity of 92% (54/59) and a specificity of 67% (376/561) for the detection of AMI. Among the 59 patients with AMI, 5 patients with an enzymatic small infarction had a normal rest MPI. The diagnostic performance to predict the need for revascularization was somewhat lower, sensitivity 81% (47/58) and specificity 74% (416/562).

In a randomized trial Udelson *et al.* demonstrated that the addition of rest MPI to standard care decreased the number of unnecessary admissions without increase in inadvertent discharge of patients with acute cardiac ischemia (31).

Stress MPI

Conti and colleagues evaluated the implementation of exercise MPI in the early triage of 306 patients suspected of ACS and normal initial work-up (Figure 2). The sensitivity and specificity to predict significant CAD or adverse events within 6 months were 94% (45/48) and 77% (198/258) (32). A large observational study of 805 patients with low to intermediate risk compared the diagnostic performance of rest MPI with stress MPI. The sensitivity and specificity for diagnosing events within 30 days (acute MI, revascularization, >70% stenosis on CAG not amenable for revascularization, life-threatening complication or cardiac death) of rest MPI, 71% (109/153) and 73% (476/652) was significantly lower than that

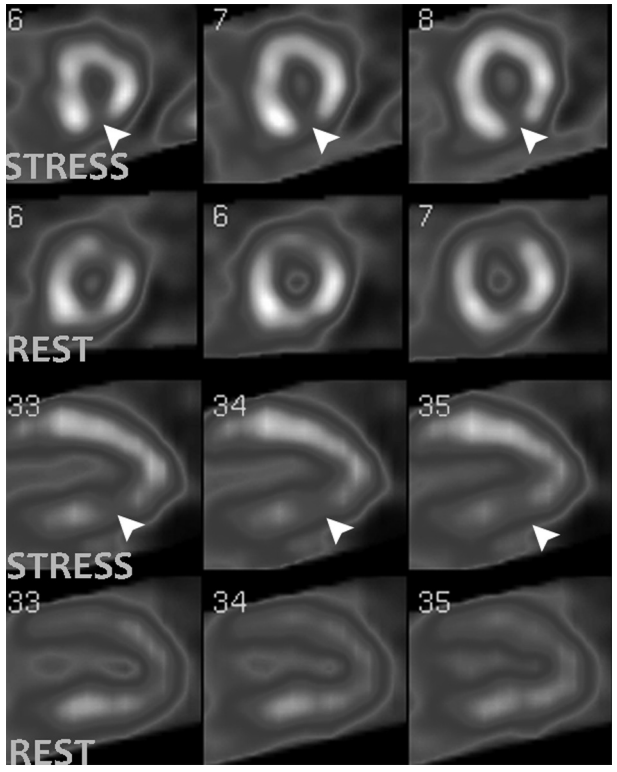


Figure 2. A 51-year-old female presented with acute chest pain in the emergency department. Short- and long-axis stress/rest nuclear perfusion scan with ischemia of the inferior left ventricular wall (arrowheads indicate diminished perfusion of the inferior wall during stress acquisition).

of stress MPI, 97% (148/153) and 88% (574/652) (33). In 2007 stress MPI was compared with CTA in a randomized trial in low risk patients without evidence of acute MI (34). The diagnostic performance of MPI and CTA was comparable, but there was a reduction in time-to-diagnosis and costs in the CTA group.

Limitations

Limited availability of nuclear facilities in some hospitals and transportation of patients to the nuclear medicine department may form a barrier to the use of MPI. Physicians should be aware of possible false negative results of rest images in patients with subsided chest pain or balanced ischemia caused by three-vessel disease (35). Acute MI should be excluded before performing stress MPI, by measuring serial cardiac markers or rest imaging. This results in a longer diagnostic work-up.

Finally, there is concern about the increasing radiation exposure from diagnostic imaging procedures and the potential risk of cancer (36).

Coronary CT angiography

Coronary CT angiography can provide high quality images of the heart and coronary vasculature, requiring minimal patient cooperation (Figure 3). Coronary CTA provides accurate information on the degree of stenosis as well as certain plaque characteristics, such as spotty calcification and low attenuation, which are associated with a higher risk of future ACS (37-38).

Image acquisition can be performed in a matter of minutes. Widely available dedicated software for automated post-processing makes interpretation quick and undemanding. In the past years, numerous papers on the diagnostic performance of coronary CTA have been published. Recent reviews demonstrated sensitivities of >95% and specificities of ≥90% (39-40). A large observational study by Hoffmann *et al.* in the ED setting, demonstrated high sensitivity of CTA for the detection of ACS, while absence of atherosclerosis was associated with an excellent 6-month outcome (41). This study confirmed the results of

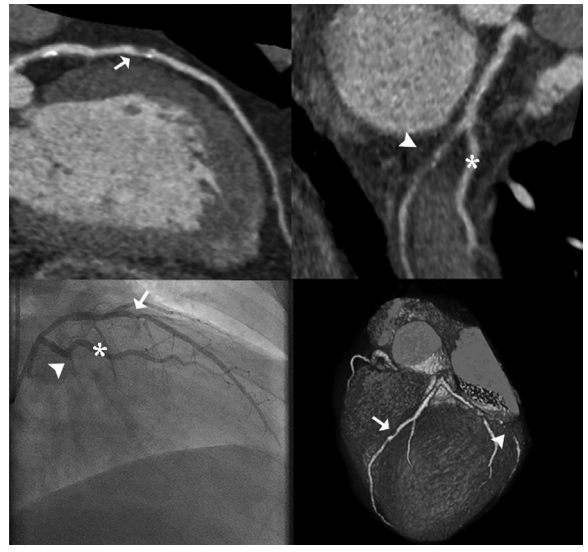


Figure 3. A 59 year-old male presented with acute chest pain. Coronary CTA was performed in the emergency department, which showed obstructive disease of the left anterior descending artery (arrow) and an occlusion of the left circumflex artery (arrowhead). On the catheter angiogram, the lesions in the left anterior descending and left circumflex artery is portrayed. The occlusion of the left circumflex is denoted by the abrupt filling of the contrast after the bifurcation of the first marginal branch (asterisk).

some earlier, smaller studies (42-44). Recently, a large multicenter randomized trial (CT-STAT) focused on the cost effectiveness of CTA vs. nuclear imaging. Low risk patients were randomly allocated to undergo CTA (n = 361) vs. rest/stress MPI (n = 338) (45). The CTA strategy reduced time-to-diagnosis by 54% and costs of care by 38%. The occurrence of major adverse cardiac events (MACE) was similar for CTA and nuclear imaging. It should be noted, however, that this study was powered to detect a difference in the primary endpoint, i.e. time-to-diagnosis. Subsequently, the study may be underpowered to detect a difference in MACE.

Additionally, coronary CTA allows for the evaluation of important non-coronary cardiac findings. Depending on the scan protocol, conditions such as acute aortic syndromes, pulmonary embolism, or oesophageal abnormalities may be detected. Incidental findings may increase downstream testing (e.g., CT for the work-up and follow-up of pulmonary nodules). Future studies that account for this additional testing are needed to evaluate the implications for cost effectiveness of coronary CTA.

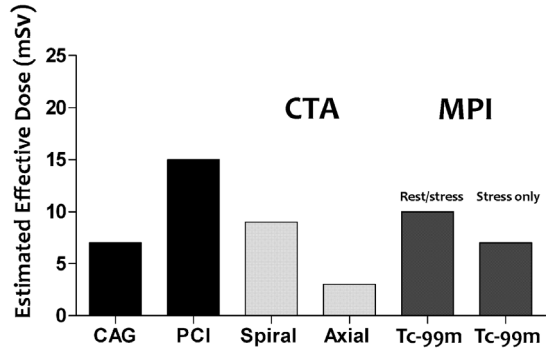


Figure 4. Estimated effective doses for CAG, PCI, coronary CTA and MPI. CAG = invasive angiography, PCI = percutaneous coronary intervention, CTA = computed tomography angiography; spiral and axial image acquisition both with tube current modulation, MPI = myocardial perfusion imaging with Technetium-99m using a rest/stress or a stress only protocol (49, 60).

Limitations

High-attenuation structures such as calcified plaques or stents appear enlarged (or bloomed) as a result of partial volume and beam hardening effects. They obscure the adjacent lumen

and may result in false positive results. In response to the concern about the potential risk related to radiation exposure, several radiation dose reduction protocols have recently been developed and validated, including ECG-dependent tube modulation, prospective ECG gating, adaptive iterative reconstruction and high-pitch spiral acquisition (46-48). When these techniques are applied, bearing in mind that prospective ECG gating may not be applicable to patients with high heart rates, the radiation dose of coronary CTA compares favorably to current rest/stress MPI protocols (40, 49) (Figure 4).

Systematic review

To summarize the available literature on this topic we formulated the following question: In patients with acute chest pain suggestive of ACS, what is the diagnostic performance of echocardiography, radionuclide myocardial perfusion imaging, and coronary CT angiography? A systematic search of the English language literature in PubMed was performed to identify studies that addressed this clinical question leading to 219 citations (search details are provided in the Appendix). Subsequently, all systematic reviews and meta-analyses that were found with the original search were reviewed to identify additional articles (Figure 5). One reviewer (A.D.) extracted all relevant data from each included study. A second reviewer (T.S.S.G.) checked the results obtained for accuracy and completeness. Discrepancies were resolved by consensus. A summary of the relevant articles categorized by imaging modality is included in Figure 6 and 7.

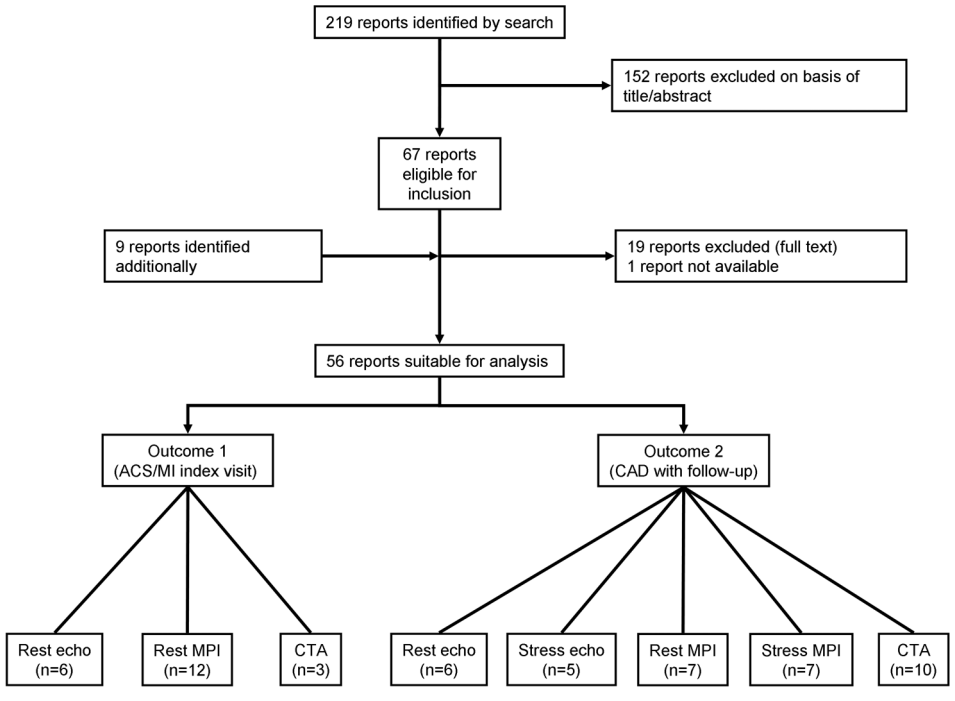


Figure 5. Flow diagram of the review process.

Statistical analysis

To obtain summary estimates of sensitivity and specificity and account for a possible correlation between sensitivities and specificities, a bivariate random-effects model was used to perform the meta-analysis (50). In the bivariate random-effects model, it is assumed that the logit-transformed true sensitivity and specificity in each study follow a bivariate normal distribution across studies, allowing a possible correlation between sensitivities and specificities. The random-effects model produces estimates of the mean logit-transformed sensitivity, logit-transformed specificity, log-transformed PPV, and log-transformed NPV with their standard errors. Sensitivity, specificity, PPV, and NPV estimates with their 95% confidence intervals were reported. All analyses were performed using SAS (version 9.2, SAS Institute Inc., North Carolina, USA). All studies were divided over two diagnosis groups according to the endpoint used as reference standard for calculating diagnostic performance. Studies appointed to diagnosis group 1 used MI or ACS during the index visit (based on clinical presentation, serial cardiac markers and ECGs) as the reference standard. Studies appointed to diagnosis group 2 used the composite endpoint "CAD" (a combination of proven ACS, angiographic evidence of obstructive CAD, obstructive CAD proven by non-invasive tests and information from follow-up) as the reference standard. For diagnosis group 1, we compared coronary CTA, rest echocardiography, and rest MPI. For diagnosis group 2, we compared coronary CTA, rest and stress echocardiography, and rest and stress MPI (Figure 5).

Results (diagnostic performance)

For the detection of MI or ACS during the index visit (diagnosis group 1) the pooled sensitivities of coronary CTA, rest echocardiography and rest MPI were: 0.94 (95% CI: 0.74–0.99), 0.86 (0.72–0.93) and 0.91 (0.85–0.95), respectively. The pooled specificity was

Echo

Study	Age (mean)	Male (%)	Outcome	Follow-up	Patients (n)	Sensitivity	Specificity	Sensitivity	Specificity
Atar 2004	50	80	Not defined	Index	70	1.00 [0.69, 1.00]	0.93 [0.84, 0.98]		
Hickman 2004	64	78	MI (CKMB)	Index	80	0.77 [0.46, 0.95]	0.43 [0.31, 0.56]		
Loh 1982	62	37	MI (CKMB)	Index	30	0.83 [0.52, 0.98]	1.00 [0.81, 1.00]		
Mohler 1998	56	62	MI (Troponin)	Index	92	0.71 [0.52, 0.86]	0.92 [0.82, 0.97]		
Peels 1990	53	74	MI (undefined)	Index	43	0.92 [0.64, 1.00]	0.53 [0.34, 0.72]		
Sasaki 1986	*	*	MI (undefined)	Index	46	0.93 [0.68, 1.00]	0.87 [0.70, 0.96]		

MPI

Study	Age (mean)	Male (%)	Outcome	Follow-up	Patients (n)	Sensitivity	Specificity	Sensitivity	Specificity
Berman 1977	64	72	MI (undefined markers+ECG)	Index	235	0.93 [0.86, 0.97]	0.69 [0.60, 0.77]		
Codini 1979	61	71	MI (undefined markers+ECG)	Index	436	0.68 [0.59, 0.76]	0.95 [0.91, 0.98]		
Cowley 1977	*	*	MI (undefined)	Index	203	0.94 [0.86, 0.98]	0.86 [0.79, 0.92]		
Duca 1999	59	40	MI (undefined)	Index	75	1.00 [0.66, 1.00]	0.73 [0.60, 0.83]		
Forberg 2009	55	50	MI (undefined)	Index	40	1.00 [0.16, 1.00]	0.71 [0.54, 0.85]		
Heller 1998	*	56	MI (undefined)	Index	357	0.90 [0.68, 0.99]	0.60 [0.54, 0.65]		
Henneman 1992	56	55	MI (undefined)	Index	47	0.75 [0.19, 0.99]	0.42 [0.27, 0.58]		
Hilton 1994	57	59	MI (undefined)	Index	102	1.00 [0.74, 1.00]	0.78 [0.68, 0.86]		
Kan 1977	*	*	MI (undefined markers+ECG)	Index	91	0.81 [0.67, 0.92]	0.94 [0.83, 0.99]		
Kontos 1997	55	45	MACE	Index	532	0.93 [0.76, 0.99]	0.71 [0.67, 0.75]		
Kontos 1999	56	46	MI (undefined)	Index	620	0.92 [0.81, 0.97]	0.67 [0.63, 0.71]		
Wieken 1983	58	59	MI (undefined)	Index	149	0.97 [0.87, 1.00]	0.71 [0.61, 0.79]		

CTA

Study	Age (mean)	Male (%)	Outcome	Follow-up	Patients (n)	Sensitivity	Specificity	Sensitivity	Specificity
Hansen 2010	56	63	MI (Troponin)	Index	89	0.75 [0.19, 0.99]	0.86 [0.77, 0.92]		
Hoffmann 2006	57	53	ACS (clinical consensus)	Index	40	1.00 [0.48, 1.00]	0.74 [0.57, 0.88]		
Hoffmann 2009	53	61	ACS (clinical consensus)	Index	368	1.00 [0.89, 1.00]	0.54 [0.49, 0.60]		

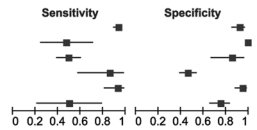
Figure 6. Relevant articles evaluating the diagnostic performance of rest echocardiography, rest MPI and coronary CTA for the detection of ACS or MI during index visit (diagnosis group 1). MI = Myocardial infarction, MACE = Major Adverse Cardiac Event (death, myocardial infarction, coronary revascularization), ACS = Acute coronary syndrome. * = Data not reported. Sensitivity and specificity reported with their 95% confidence intervals in brackets.

0.73 (0.46–0.90), 0.82 (0.65–0.91) and 0.76 (0.64-0.85), respectively. Summary receiver operating characterises (SROC)-curve for group 1 is displayed in Figure 8.

For the detection of CAD (diagnosis group 2), rest echocardiography and rest MPI demonstrated a pooled sensitivity of 0.76 (0.58–0.89) and 0.80 (0.64–0.90), respectively. The pooled specificity was 0.89 (0.78–0.95) and 0.83 (0.70–0.91), respectively. Stress echocardiography and stress MPI demonstrated a pooled sensitivity of 0.78 (0.57—0.90) and 0.92 (0.84–0.97). Their corresponding pooled specificity was 0.92 (0.83-0.96) and 0.88 (0.78–0.94), respectively. Coronary CTA demonstrated a pooled sensitivity of 0.93 (0.84-0.97),

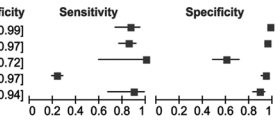
Echo (rest)

Study	Age (mean)	Male (%)	Outcome	Follow-up	Patients (n)	Sensitivity	Specificity
Di Pasquale 2004	60	65	>70% or >50% LM CAG	Index	280	0.93 [0.89, 0.97]	0.92 [0.85, 0.96]
Gibler 1995	*	*	MACE	30 days	901	0.47 [0.24, 0.71]	0.99 [0.98, 1.00]
Kang 2005	60	64	ACS / Revasc	Index	114	0.49 [0.39, 0.60]	0.85 [0.66, 0.96]
Kontos 1998	*	51	MI / Revasc	Index	260	0.86 [0.57, 0.98]	0.46 [0.39, 0.54]
Muscholl 2002	56	69	MACE	30 days	132	0.93 [0.81, 0.99]	0.94 [0.88, 0.98]
Weston 2004	54	58	Death / MI / Myocardial ischemia	Index	108	0.50 [0.21, 0.79]	0.75 [0.65, 0.83]



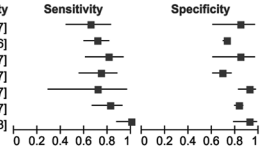
Echo (stress)

Study	Age (mean)	Male (%)	Outcome	Follow-up	Patients (n)	Sensitivity	Specificity
Bedetti 2005	58	58	Cardiac death / ACS	13 months	552	0.87 [0.73, 0.95]	0.98 [0.96, 0.99]
Conti 2005	60	64	CAG or MACE + UAP	6 months	503	0.85 [0.76, 0.92]	0.95 [0.93, 0.97]
Geleijnse 2000	57	64	MACE	Index	80	1.00 [0.59, 1.00]	0.60 [0.48, 0.72]
Iglesias 2005	64	56	CAG > 50%	Index	487	0.24 [0.19, 0.29]	0.94 [0.90, 0.97]
Trippi 1997	50	52	MI or CAG	30 days	163	0.89 [0.67, 0.99]	0.89 [0.83, 0.94]



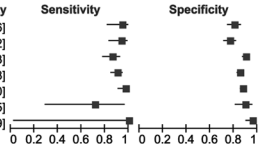
MPI (rest)

Study	Age (mean)	Male (%)	Outcome	Follow-up	Patients (n)	Sensitivity	Specificity
Bilodeau 1991	58	53	CAG > 50%	Index	45	0.65 [0.44, 0.83]	0.84 [0.60, 0.97]
Fesmire (1) 2001	54	54	MACE	30 days	805	0.71 [0.59, 0.81]	0.73 [0.69, 0.76]
Gregoire 1990	58	53	CAG > 50%	Index	45	0.81 [0.61, 0.93]	0.84 [0.60, 0.97]
Kaul 2004	62	68	MACE	Index	203	0.74 [0.55, 0.88]	0.69 [0.60, 0.77]
Kosnik 1999	56	43	MACE	12 months	69	0.71 [0.29, 0.96]	0.92 [0.82, 0.97]
Tatum 1997	51	49	MACE	Index	438	0.82 [0.66, 0.92]	0.83 [0.79, 0.87]
Varetto 1993	58	55	MI / CAG / SPECT	Index	64	1.00 [0.87, 1.00]	0.92 [0.78, 0.98]



MPI (stress)

Study	Age (mean)	Male (%)	Outcome	Follow-up	Patients (n)	Sensitivity	Specificity
Conti 2001	59	68	MI / CAG / SPECT	6 months	231	0.94 [0.81, 0.99]	0.81 [0.74, 0.86]
Conti 2003	60	65	MI / CAG / SPECT	6 months	306	0.94 [0.83, 0.99]	0.77 [0.71, 0.82]
Conti 2005	60	64	CAG or MACE + UAP	6 months	503	0.86 [0.78, 0.92]	0.90 [0.87, 0.93]
Conti 2008	62	60	CAG or MACE + UAP	12 months	798	0.90 [0.84, 0.95]	0.85 [0.82, 0.88]
Fesmire (2) 200	54	54	MACE	30 days	805	0.97 [0.90, 1.00]	0.88 [0.85, 0.90]
Gallagher 2007	49	53	CAG / SPECT	30 days	85	0.71 [0.29, 0.96]	0.90 [0.81, 0.95]
Goldstein 2007	48	42	CAG / CTA / SPECT	6 months	99	1.00 [0.03, 1.00]	0.96 [0.90, 0.99]



CTA

Study	Age (mean)	Male (%)	Outcome	Follow-up	Patients (n)	Sensitivity	Specificity
Chow 2010	53	58	CAG / SPECT	Index	55	0.86 [0.57, 0.98]	0.93 [0.80, 0.98]
Gallagher 2007	49	53	CAG / SPECT	30 days	85	0.86 [0.42, 1.00]	0.90 [0.81, 0.96]
Goldstein 2007	48	42	CAG / CTA / SPECT	6 months	99	1.00 [0.63, 1.00]	0.74 [0.63, 0.82]
Johnson 2008	64	72	CAG > 50%	6 months	109	1.00 [0.78, 1.00]	0.96 [0.89, 0.99]
Kim 2010	58	49	Clinical judgement cardiologist	30 days	296	0.89 [0.78, 0.95]	0.85 [0.80, 0.89]
Olivetti 2006	59	61	CAG > 50%	Index	31	0.83 [0.59, 0.96]	1.00 [0.75, 1.00]
Rubinshtein 2007	56	64	MI / CAG / SPECT	15 months	58	1.00 [0.83, 1.00]	0.92 [0.79, 0.98]
Tsai 2007	61	71	CAG > 50%	Index	78	0.98 [0.90, 1.00]	0.81 [0.62, 0.94]
Ueno 2009	66	53	CAG / SPECT	30 days	36	0.92 [0.62, 1.00]	0.83 [0.63, 0.95]
White 2005	51	51	CAG / SPECT / Stress echo / CTA	30 days	69	0.83 [0.52, 0.98]	0.96 [0.88, 1.00]

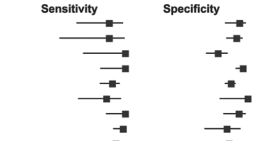


Figure 7. Relevant articles evaluating the diagnostic performance of rest and stress echocardiography, rest and stress MPI and coronary CTA for the detection of CAD with or without follow-up (diagnosis group 2). LM = Left main coronary artery, MACE = Death, myocardial infarction, coronary revascularization, ACS = Acute coronary syndrome, Revasc = coronary revascularization, MI = Myocardial infarction, UAP = unstable angina pectoris. * = Data not reported. Sensitivity and specificity reported with their 95% confidence intervals in brackets.

and a pooled specificity of 0.90 (0.83-0.95), which was comparable to its diagnostic performance for the detection of MI or ACS (diagnosis group 1). SROC-curve for group 2 is displayed in Figure 9.

Evidence-Based Guidelines

Current guidelines from the 2011 American College of Radiology Appropriateness Criteria consider combined rest/stress MPI as the preferred test in intermediate to high risk patients. A reasonable alternative is rest/stress echocardiography, especially because it does not require radiation exposure. Combined rest/stress protocols are considered more appropriate than single protocols. Coronary CTA is regarded as an appropriate alternative in low to intermediate risk patients (51). Those with high likelihood of CAD would benefit most from invasive angiography. The current analysis incorporated several recent publications with coronary CTA, reporting excellent diagnostic performance. Compared to the guidelines, our findings are more in favor of using coronary CTA.

Outstanding issues that warrant Research

Heterogeneity

The diagnostic performance of the three modalities showed large variation across studies. This is in part due to differences in patient populations, but mainly because different clinical endpoints were used as reference standard to calculate diagnostic performance. These endpoints ranged from cardiac markers, to angiographic results or adverse events during follow-up. Some studies used combined endpoints, which may be misleading, because they may not truly reflect the ability to correctly risk-stratify patients with suspected ACS.

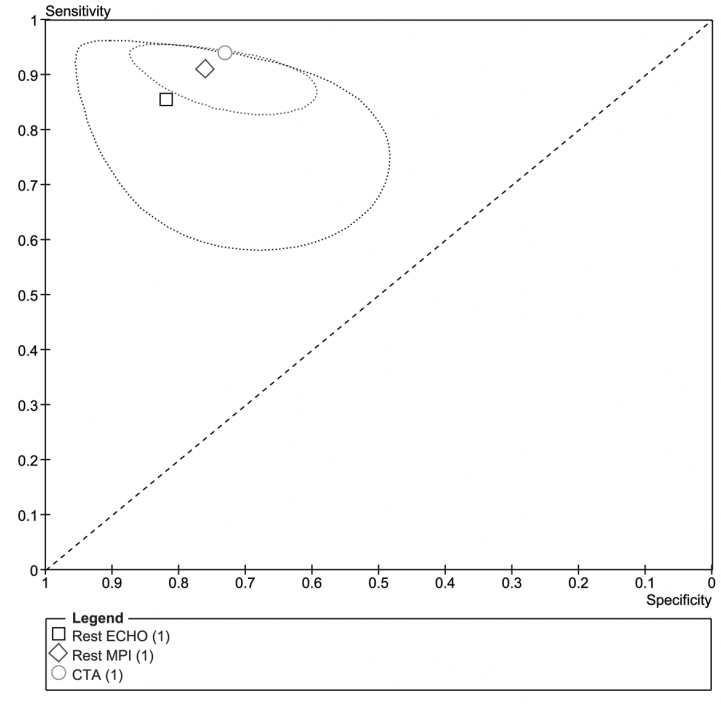


Figure 8. Summary receiver operating curves for diagnosis group 1. Rest echo (square), rest MPI (diamond) and CTA (circle) encircled by their 95% confidence intervals. No 95% confidence interval could be calculated for CTA because of the small number of studies.

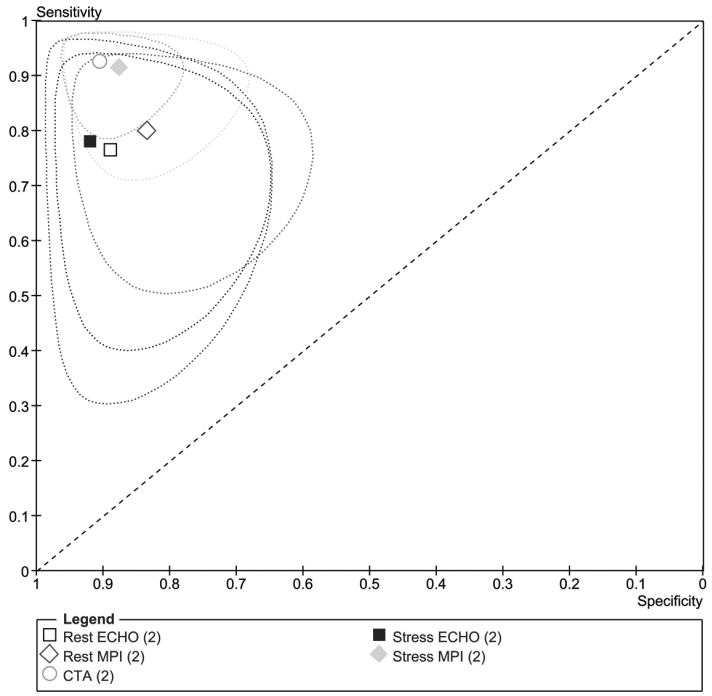


Figure 9. Summary receiver operating curves for diagnosis group 2. Rest echo (square), stress echo (blue filled square), rest MPI (diamond), stress MPI (green filled diamond) and CTA (circle) encircled by their 95% confidence intervals.

Technical advancements

Three-dimensional volumetric imaging is a promising application of echocardiography, which shortens acquisition time, reduces operator dependency and provides more accurate assessment of ventricular volumes. However, for assessment of WMA this technique seems less applicable and its use is further held back because of lower spatial and temporal resolutions compared to conventional two-dimensional echocardiography, as well as multiple beat acquisition artefacts (52-53). Three-dimensional echocardiography does not preclude conventional two-dimensional imaging and these techniques can be used side by side to reinforce each other.

In nuclear perfusion imaging, reduced imaging time and better diagnostic information will be possible because of improved camera designs, iterative reconstruction methods and new acquisition protocols (stress-only acquisitions, dynamic imaging etc.) (54). Also, radionuclide agents with "ischemic memory" could help identify patients with acute chest pain and subsided ischemia in the future (55).

In search of the functional significance of plaques seen on coronary CTA, new techniques like CT myocardial perfusion and computational fluid dynamics are being investigated and show very promising results (56-57).

Summary

Recommendations for best practice

Our systematic search of the medical literature demonstrated no significant difference between the three modalities for the detection of both 1) ACS and 2) CAD with or without follow-up. There was a slight, non-significant positive trend favouring coronary CTA in both analyses. It is important to keep in mind that the optimal imaging strategy is not only determined by the diagnostic performance of a modality, but also by local practice, expertise with imaging techniques, medical facilities and individual patient characteristics. Given the absence of large differences in diagnostic performance, these practical aspects may be even more important.

Since the diagnostic performance of CTA is excellent and the test can be performed quickly, we advocate performing coronary CTA in low to intermediate risk patients provided the test and expertise to interpret the results are readily available in emergency situations. For institutions with nuclear imaging expertise and rapid access to nuclear facilities, MPI will generally be the first choice for low to intermediate risk patients, despite the fact that recent trials suggest it is more costly and time-consuming. Rapid screening with echocardiography for the presence of many non-ischemic conditions as well as assessment of hemodynamic status remains a vital part of the standard work-up of all patients with acute chest pain. However, extensive imaging seems less suitable for diagnosing ACS or CAD. In concordance with the current guidelines, we believe high risk patients would benefit most from an invasive strategy (58-59).

Recommendations for future research

The lack of robust evidence to identify the optimal test is partially caused by substantial heterogeneity across studies. The studies identified by our systematic search used various clinical endpoints to calculate diagnostic performance. At this moment, there seems to be no general agreement which clinical endpoint (cardiac markers, angiographic results or combined with follow-up data) should be used for validation. In addition, there is still a limited number of comparative studies between imaging strategies. For better understanding of the optimal imaging strategy for patients with acute chest pain, there is a need for large randomized trials with general accepted clinical endpoints.

Appendix

Studies were included if they met all of the following criteria: 1) Study population were patients with acute chest pain suggestive of acute coronary syndrome (ACS), 2) Study reported operating characteristics of echocardiography, nuclear myocardial perfusion imaging or coronary CTA, 3) The reference standard was the presence of ACS based on cardiac markers or ECG; clinical consensus based on invasive angiography or proven obstructive coronary artery disease on additional tests (not consisting of the index test); cardiac events during follow-up.

Studies were excluded if they met one of the following criteria: 1) Article was a review, guideline, or cost-effectiveness analysis, or 2) Study population consisted of patients with stable angina, asymptomatic patients, or patients with proven myocardial infarction at baseline, 3) Study population consisted of patients with ST-elevation myocardial infarction, 4) Study population overlapped with previous publications.

Articles not accessible or not written in English were excluded from the analysis.

Search strategy

("Myocardial Infarction"(MeSH Terms) OR "Myocardial Ischemia"(MeSH Terms) OR "Angina, Unstable"(MeSH Terms) OR "Acute Coronary Syndrome"(MeSH Terms) OR "Coronary Disease"(MeSH Terms) OR "Coronary Artery Disease"(MeSH Terms) or "Chest Pain"(Mesh))

AND

("acute disease"(Mesh) OR "Emergency Service, Hospital"(Mesh) OR "Emergency Medical Services"(Mesh) OR "Emergency Treatment"(Mesh) OR "Emergencies"(Mesh) OR "Emergency Medicine"(Mesh))

AND

("Echocardiography"(Mesh) OR "Radionuclide Imaging"(Mesh) OR "Technetium Tc 99m Sestamibi"(Mesh) OR "Tomography, X-ray Computed"(MeSH Terms))

AND

("Coronary Angiography"(Mesh) OR "consensus"(MeSH Terms) OR "Acute Coronary Syndrome"(Mesh) OR "Angina, Unstable"(MeSH Terms) OR "Myocardial Infarction"(MeSH Terms) OR "Myocardial Ischemia"(MeSH Terms))

AND

("sensitivity and specificity"(Mesh) OR sensitivity(tw) OR specificity(tw) OR "predictive value of tests"(Mesh) OR "ROC Curve"(Mesh) OR ("positive predictive"(tw) AND value*(tw)) OR ("negative predictive"(tw) AND value*(tw)) OR (false(tw) AND negative*(tw)) OR (false(tw) AND positive*(tw)) OR (true(tw) AND negative*(tw)) OR (true(tw) AND positive*(tw)) OR misdiagnosis(tw) OR misdiagnoses OR "Diagnostic Errors"(Mesh) OR (diagnost*(tw) AND (accuracy(tw) OR error*(tw) OR efficacy(tw))))

AND

(English(lang) NOT ((Animals(Mesh) NOT Humans(Mesh)) OR Editorial(Publication type) OR Comment(Publication type) OR Letter(Publication type) OR Case Reports(Publication type)))

References

1. Thygesen K, Alpert JS, White HD, Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. Journal of the American College of Cardiology 2007;50:2173-2195
2. Luepker RV, Apple FS, Christenson RH, et al. Case definitions for acute coronary heart disease in epidemiology and clinical research studies: a statement from the AHA Council on Epidemiology and Prevention; AHA Statistics Committee; World Heart Federation Council on Epidemiology and Prevention; the European Society of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease Control and Prevention; and the National Heart, Lung, and Blood Institute. Circulation 2003;108:2543-2549
3. Lev EI, Battler A, Behar S, et al. Frequency, characteristics, and outcome of patients hospitalized with acute coronary syndromes with undetermined electrocardiographic patterns. Am J Cardiol 2003;91:224-227
4. Roberts RR, Zalenski RJ, Mensah EK, et al. Costs of an emergency department-based accelerated diagnostic protocol vs hospitalization in patients with chest pain: a randomized controlled trial. JAMA 1997;278:1670-1676
5. Brown DW, Xie J, Mensah GA. Electrocardiographic recording and timeliness of clinician evaluation in the emergency department in patients presenting with chest pain. Am J Cardiol 2007;99:1115-1118
6. Fruergaard P, Launbjerg J, Hesse B, et al. The diagnoses of patients admitted with acute chest pain but without myocardial infarction. Eur Heart J 1996;17:1028-1034
7. Selker HP, Beshansky JR, Griffith JL, et al. Use of the acute cardiac ischemia time-insensitive predictive instrument (ACI-TIPI) to assist with triage of patients with chest pain or other symptoms suggestive of acute cardiac ischemia. A multicenter, controlled clinical trial. Annals of internal medicine 1998;129:845-855
8. Pope JH, Aufderheide TP, Ruthazer R, et al. Missed diagnoses of acute cardiac ischemia in the emergency department. N Engl J Med 2000;342:1163-1170
9. Christenson J, Innes G, McKnight D, et al. Safety and efficiency of emergency department assessment of chest discomfort. Cmaj 2004;170:1803-1807
10. McCarthy BD, Beshansky JR, D'Agostino RB, Selker HP. Missed diagnoses of acute myocardial infarction in the emergency department: results from a multicenter study. Ann Emerg Med 1993;22:579-582
11. Lee TH, Rouan GW, Weisberg MC, et al. Clinical characteristics and natural history of patients with acute myocardial infarction sent home from the emergency room. Am J Cardiol 1987;60:219-224
12. Schor S, Behar S, Modan B, Barell V, Drory J, Kariv I. Disposition of presumed coronary patients from an emergency room. A follow-up study. Jama 1976;236:941-943
13. Oetgen WJ, Parikh PD, Cacchione JG, et al. Characteristics of Medical Professional Liability Claims in Patients With Cardiovascular Diseases. Am J Cardiol;105:745-752
14. Zabalgoitia M, Ismaeil M. Diagnostic and prognostic use of stress echo in acute coronary syndromes including emergency department imaging. Echocardiography 2000;17:479-493
15. Gibler WB, Runyon JP, Levy RC, et al. A rapid diagnostic and treatment center for patients with chest pain in the emergency department. Ann Emerg Med 1995;25:1-8
16. Kang DH, Kang SJ, Song JM, et al. Efficacy of myocardial contrast echocardiography in the diagnosis and risk stratification of acute coronary syndrome. Am J Cardiol 2005;96:1498-1502
17. Mohler ER, 3rd, Ryan T, Segar DS, et al. Clinical utility of troponin T levels and echocardiography in the emergency department. Am Heart J 1998;135:253-260
18. Di Pasquale P, Cannizzaro S, Scalzo S, et al. Sensitivity, specificity and predictive value of the echocardiography and troponin-T test combination in patients with non-ST elevation acute coronary syndromes. Int J Cardiovasc Imaging 2004;20:37-46
19. Mulvagh SL, Rakowski H, Vannan MA, et al. American Society of Echocardiography Consensus Statement on the Clinical Applications of Ultrasonic Contrast Agents in Echocardiography. J Am Soc Echocardiogr 2008;21:1179-1201; quiz 1281
20. Rinkevich D, Kaul S, Wang XQ, et al. Regional left ventricular perfusion and function in patients presenting to the emergency department with chest pain and no ST-segment elevation. Eur Heart J 2005;26:1606-1611
21. Trippi JA, Lee KS, Kopp G, Nelson DR, Yee KG, Cordell WH. Dobutamine stress tele-echocardiography for evaluation of emergency department patients with chest pain. J Am Coll Cardiol 1997;30:627-632
22. Conti A, Sammicheli L, Gallini C, Costanzo EN, Antoniucci D, Barletta G. Assessment of patients with low-risk chest pain in the emergency department: Head-to-head comparison of exercise stress echocardiography and exercise myocardial SPECT. Am Heart J 2005;149:894-901
23. Weston P, Alexander JH, Patel MR, Maynard C, Crawford L, Wagner GS. Hand-held echocardiographic examination of patients with symptoms of acute coronary syndromes in the emergency department: the 30-day outcome associated with normal left ventricular wall motion. Am Heart J 2004;148:1096-1101
24. Wackers FJ, Lie KI, Liem KL, et al. Potential value of thallium-201 scintigraphy as a means of selecting patients for the coronary care unit. Br Heart J 1979;41:111-117
25. Berman DS, Kiat H, Van Train KF, Friedman J, Garcia EV, Maddahi J. Comparison of SPECT using technetium-99m agents and thallium-201 and PET for the assessment of myocardial perfusion and viability. Am J Cardiol 1990;66:72E-79E
26. Fintel DJ, Links JM, Brinker JA, Frank TL, Parker M, Becker LC. Improved diagnostic performance of exercise thallium-201 single photon emission computed tomography over planar imaging in the diagnosis of coronary artery disease: a receiver operating characteristic analysis. Journal of the American College of Cardiology 1989;13:600-612
27. DePuey EG, Rozanski A. Using gated technetium-99m-sestamibi SPECT to characterize fixed myocardial defects as infarct or artifact. J Nucl Med 1995;36:952-955
28. Kontos MC, Jesse RL, Anderson FP, Schmidt KL, Ornato JP, Tatum JL. Comparison of myocardial perfusion imaging and cardiac troponin I in patients admitted to the emergency department with chest pain. Circulation 1999;99:2073-2078
29. Heller GV, Stowers SA, Hendel RC, et al. Clinical value of acute rest technetium-99m tetrofosmin tomographic myocardial perfusion imaging in patients with acute chest pain and nondiagnostic electrocardiograms. Journal of the American College of Cardiology 1998;31:1011-1017
30. Hilton TC, Thompson RC, Williams HJ, Saylors R, Fulmer H, Stowers SA. Technetium-99m sestamibi myocardial perfusion imaging in the emergency room evaluation of chest pain. Journal of the American College of Cardiology 1994;23:1016-1022
31. Udelson JE, Beshansky JR, Ballin DS, et al. Myocardial perfusion imaging for evaluation and triage of patients with suspected acute cardiac ischemia: a randomized controlled trial. Jama 2002;288:2693-2700
32. Conti A, Zanobetti M, Grifoni S, et al. Implementation of myocardial perfusion imaging in the early triage of patients with suspected acute coronary syndromes. Nucl Med Commun 2003;24:1055-1060
33. Fesmire FM, Hughes AD, Stout PK, Wojcik JF, Wharton DR. Selective dual nuclear scanning in low-risk patients with chest pain to reliably identify and exclude acute coronary syndromes. Ann Emerg Med 2001;38:207-215
34. Goldstein JA, Gallagher MJ, O'Neill WW, Ross MA, O'Neil BJ, Raff GL. A randomized controlled trial of multi-slice coronary computed tomography for evaluation of acute chest pain. J Am Coll Cardiol 2007;49:863-871
35. Fram DB, Azar RR, Ahlberg AW, et al. Duration of abnormal SPECT myocardial perfusion imaging following resolution of acute ischemia: an angioplasty model. J Am Coll Cardiol 2003;41:452-459
36. Chen J, Einstein AJ, Fazel R, et al. Cumulative exposure to ionizing radiation from diagnostic and therapeutic cardiac imaging procedures: a population-based analysis. Journal of the American College of Cardiology 2010;56:702-711

37. Motoyama S, Sarai M, Harigaya H, et al. Computed tomographic angiography characteristics of atherosclerotic plaques subsequently resulting in acute coronary syndrome. Journal of the American College of Cardiology 2009;54:49-57
38. Pundziute G, Schuijf JD, Jukema JW, et al. Evaluation of plaque characteristics in acute coronary syndromes: non-invasive assessment with multi-slice computed tomography and invasive evaluation with intravascular ultrasound radiofrequency data analysis. Eur Heart J 2008;29:2373-2381
39. Meijer AB, O YL, Geleijns J, Kroft LJ. Meta-analysis of 40- and 64-MDCT angiography for assessing coronary artery stenosis. AJR Am J Roentgenol 2008;191:1667-1675
40. von Ballmoos MW, Haring B, Juillerat P, Alkadhi H. Meta-analysis: diagnostic performance of low-radiation-dose coronary computed tomography angiography. Annals of internal medicine 2011;154:413-420
41. Hoffmann U, Bamberg F, Chae CU, et al. Coronary computed tomography angiography for early triage of patients with acute chest pain: the ROMICAT (Rule Out Myocardial Infarction using Computer Assisted Tomography) trial. J Am Coll Cardiol 2009;53:1642-1650
42. Tsai IC, Lee T, Lee WL, et al. Use of 40-detector row computed tomography before catheter coronary angiography to select early conservative versus early invasive treatment for patients with low-risk acute coronary syndrome. J Comput Assist Tomogr 2007;31:258-264
43. White CS, Kuo D, Kelemen M, et al. Chest pain evaluation in the emergency department: can MDCT provide a comprehensive evaluation? AJR Am J Roentgenol 2005;185:533-540
44. Rubinshtein R, Halon DA, Gaspar T, et al. Usefulness of 64-slice multidetector computed tomography in diagnostic triage of patients with chest pain and negative or nondiagnostic exercise treadmill test result. Am J Cardiol 2007;99:925-929
45. Goldstein JA, Chinnaiyan KM, Abidov A, et al. The CT-STAT (Coronary Computed Tomographic Angiography for Systematic Triage of Acute Chest Pain Patients to Treatment) Trial. Journal of the American College of Cardiology 2011;58:1414-1422
46. Abada HT, Larchez C, Daoud B, Sigal-Cinqualbre A, Paul JF. MDCT of the coronary arteries: feasibility of low-dose CT with ECG-pulsed tube current modulation to reduce radiation dose. AJR Am J Roentgenol 2006;186:S387-39
47. Leipsic J, Labounty TM, Heilbron B, et al. Estimated radiation dose reduction using adaptive statistical iterative reconstruction in coronary CT angiography: the ERASIR study. AJR Am J Roentgenol 2010;195:655-660
48. Maruyama T, Takada M, Hasuike T, Yoshikawa A, Namimatsu E, Yoshizumi T. Radiation dose reduction and coronary assessability of prospective electrocardiogram-gated computed tomography coronary angiography: comparison with retrospective electrocardiogram-gated helical scan. Journal of the American College of Cardiology 2008;52:1450-1455
49. Einstein AJ. Radiation risk from coronary artery disease imaging: how do different diagnostic tests compare? Heart (British Cardiac Society) 2008;94:1519-1521
50. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in medicine 2002;21:589-624
51. Mammen L, White RD, Woodard PK, et al. ACR Appropriateness Criteria(R) on chest pain, suggestive of acute coronary syndrome. J Am Coll Radiol 2011;8:12-18
52. Matsumura Y, Hozumi T, Arai K, et al. Non-invasive assessment of myocardial ischaemia using new real-time three-dimensional dobutamine stress echocardiography: comparison with conventional two-dimensional methods. Eur Heart J 2005;26:1625-1632
53. Sawada SG, Thomaides A. Three-dimensional stress echocardiography: the promise and limitations of volumetric imaging. Current opinion in cardiology 2009;24:426-432
54. Slomka PJ, Berman DS, Germano G. New Imaging Protocols for New Single Photon Emission CT Technologies. Curr Cardiovasc Imaging Rep 2010;3:162-170

55. Tamaki N, Yoshinaga K. Novel iodinated tracers, MIBG and BMIPP, for nuclear cardiology. J Nucl Cardiol 2011;18:135-143
56. Min JK, Berman DS, Budoff MJ, et al. Rationale and design of the DeFACTO (Determination of Fractional Flow Reserve by Anatomic Computed Tomographic AngiOgraphy) study. J Cardiovasc Comput Tomogr 2011;5:301-309
57. Blankstein R, Shturman LD, Rogers IS, et al. Adenosine-induced stress myocardial perfusion imaging using dual-source cardiac computed tomography. Journal of the American College of Cardiology 2009;54:1072-1084
58. Hamm CW, Bassand JP, et al. ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes (ACS) in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J 2011
59. Wright RS, Anderson JL, Adams CD, et al. 2011 ACCF/AHA focused update incorporated into the ACC/AHA 2007 Guidelines for the Management of Patients with Unstable Angina/Non-ST-Elevation Myocardial Infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines developed in collaboration with the American Academy of Family Physicians, Society for Cardiovascular Angiography and Interventions, and the Society of Thoracic Surgeons. Journal of the American College of Cardiology 2011;57:e215-367
60. Gerber TC, Carr JJ, Arai AE, et al. Ionizing radiation in cardiac imaging: a science advisory from the American Heart Association Committee on Cardiac Imaging of the Council on Clinical Cardiology and Committee on Cardiovascular Imaging and Intervention of the Council on Cardiovascular Radiology and Intervention. Circulation 2009;119:1056-1065.

Part III

Prediction Models

Chapter 6

Incremental Value of the CT Coronary Calcium Score for the Prediction of Coronary Artery Disease

Eur Radiol. 2010;20(10):2331-40.

Tessa S. S. Genders
Francesca Pugliese
Nico R. Mollet
W. Bob Meijboom
Annick C. Weustink
Carlos A. G. van Mieghem
Pim J. de Feyter
M. G. Myriam Hunink

Abstract

Objectives

To validate published prediction models for the presence of obstructive coronary artery disease (CAD) in patients with new onset stable typical or atypical angina pectoris and to assess the incremental value of the CT coronary calcium score (CCS).

Methods

We searched the literature for clinical prediction rules for the diagnosis of obstructive CAD, defined as ≥50% stenosis in at least one vessel on conventional coronary angiography. Significant variables were re-analyzed in our dataset of 254 patients with logistic regression. CCS was subsequently included in the models. The area under the receiver operating characteristic curve (AUC) was calculated to assess diagnostic performance.

Results

Re-analyzing the variables used by Diamond & Forrester yielded an AUC of 0.798, which increased to 0.890 by adding CCS. For Pryor, Morise 1994, Morise 1997 and Shaw the AUC increased from 0.838 to 0.901, 0.831 to 0.899, 0.840 to 0.898 and 0.833 to 0.899. CCS significantly improved model performance in each model.

Conclusions

Validation demonstrated good diagnostic performance across all models. CCS improves the prediction of the presence of obstructive CAD, independent of clinical predictors, and should be considered in its diagnostic work-up.

Introduction

The CT coronary calcium score (CCS) is used in both the diagnosis of coronary artery disease (CAD) (1-6) and the prediction of cardiovascular events (7-13). Although substantial evidence is available on the incremental value of CCS in predicting future cardiovascular events and mortality in asymptomatic individuals (10-13), the diagnostic value in symptomatic patients is less clear. A meta-analysis revealed that the presence of coronary calcium had a sensitivity and specificity of 98% and 40%, respectively, in detecting significant stenoses (8). Furthermore, numerous studies have reported on the value of CCS in the prediction of the probability of obstructive CAD (1–4). However, the incremental value (i.e. in addition to all known clinical predictors of CAD) of the CCS as a continuous predictor of prevalent obstructive CAD is less well studied. The purpose of this study was to validate previously published clinical prediction models and to determine the incremental value of CCS for the prediction of prevalent obstructive CAD in patients with new onset stable typical or atypical angina pectoris.

Materials and methods

Study population

The study population was derived from an existing database, which consisted of 402 patients with chest pain suggestive of stable angina pectoris and suspected of having CAD. All patients were prospectively included in a large study evaluating 64-slice coronary CT angiography (CCTA) at our institution. All patients were referred for catheter-based coronary angiography (CAG) based on their presentation or functional testing that suggested the presence of ischemia and all patients underwent multi-detector CT angiography within a week before CAG. Inclusion criteria for this study were: informed consent, sinus heart rhythm and the ability to hold their breath for 15 s. Patients with a history of percutaneous coronary intervention or coronary artery bypass surgery, impaired renal function (serum creatinine >120 μmol/L) or a known intolerance to iodinated contrast medium were excluded. The Institutional Review Board approved the study and all patients signed informed consent. As this paper focuses on patients with new onset stable chest pain, we also excluded patients with acute coronary syndromes and patients with a previous myocardial infarction (Figure 1).

CT coronary calcium images

Metoprolol (100 mg, Selokeen, AstraZeneca, London, UK) was administered orally 1 h before CT in patients with heart rates >65 beats per minute. A 64-slice single source CT system (Sensation 64; Siemens, Forchheim, Germany) with a gantry rotation time of 330 ms, acquisition time of 165 ms and voxel size of 0.4 mm3 was used to acquire standard spiral low-dose and ECG-gated coronary calcium CT images. CT parameters were 32×2 slices per rotation, individual detector width of 0.6 mm, 3.8-mm/rotation table feed, 120-kV tube voltage, 150-mAs tube current, with activated prospective x-ray tube modulation. Overlapping slices were reconstructed at 65% of the R–R interval using the B35f convolution kernel. Reconstructed slice thickness was 3.0 mm with an increment of 1.5 mm. The radiation exposure, estimated using dedicated software (ImPACT, version 0.99x, St. George's Hospital, Tooting, London, UK), was 1.4 mSv in men and 1.8 mSv in women. One observer (with more than 3 years' experience), who was blinded to the CAG and clinical data, measured the coronary calcium.

Conventional coronary angiography

The CCS and CAG were carried out within 1 week. Coronary segments were assessed on CAG following a 17-segment modified American Heart Association (AHA) classification model (14) by a single observer (with more than 10 years' experience), who was blinded to the CT and clinical data. A mean luminal narrowing of ≥50% was considered to be a significant stenosis. Validated quantitative coronary angiography software (CAAS II®, Pie Medical, Maastricht, the Netherlands) was used.

Clinical variables and outcome

All patients were interviewed at enrollment in the prospective cohort study. Clinical parameters recorded were: age (years), sex (male/female), type of chest pain (atypical vs. typical), body mass index (BMI) (defined as weight/height^2 in kg/m^2), smoking status (past or current smoker, yes/no), hypertension (present/absent), dyslipidaemia (serum cholesterol >200 mg/dL or 5.18 mmol/L, present/absent), diabetes (plasma glucose ≥126 mg/dL or 7.0 mmol/L, present/absent) and family history of CAD (present/absent). The CCS was measured by the Agatston method (15) using dedicated software (syngo Calcium Scoring VE31H, Siemens, Germany). The outcome of interest was the presence of obstructive CAD defined as ≥50% stenosis in at least one vessel (present/absent) on CAG.

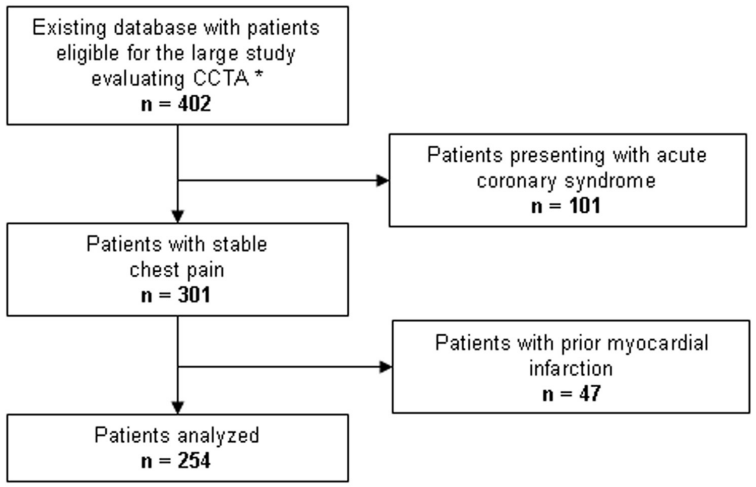


Figure 1. Flow chart of patients in the study. CCTA = coronary computed tomography angiography. *Data from an existing database were used. All patients were referred for conventional coronary angiography based on their presentation or functional testing that suggested the presence of cardiac ischemia. See Materials and methods.

Sample size

As a general rule, 10 patients with the condition of interest per analyzed variable are required for regression analysis. In our dataset (n=254), 123 patients were identified as having obstructive CAD on CAG. This allowed for the analysis of 12 variables. Our sample meets the required number of cases and non-cases that has been suggested for external validation of prediction models (16).

Systematic literature search

We searched the English-language medical literature in PubMed up to October 14, 2009 for diagnostic prediction models. See the Appendix for a detailed description of the search strategy. From the included articles, clinical variables that were identified as significant predictors of CAD were extracted.

Data analysis

Age was analyzed as a continuous variable. To account for the skewed distribution of the coronary calcium scoring, CCS was transformed by taking the natural logarithm of CCS+1. All other variables were dichotomous. Estrogen status was not available in our dataset. Therefore, we assumed women below the age of 50 to be estrogen positive, women of 50 years and above to be estrogen negative and all men to be estrogen neutral. Obesity was considered in the model by Morise (1997) only. We defined obesity as a BMI > 27 kg/m2, corresponding to their definition (17).

The extracted sets of clinical variables were analyzed with multivariate logistic regression analysis, fitting new regression coefficients. No attempt was made to validate original regression coefficients, as such coefficients were often not reported. CCS was subsequently included in each of the models. Models without CCS were compared with corresponding models including CCS using the likelihood ratio test. The level of significance was defined at a P-value less than 0.05.

Diagnostic performance was assessed by calculating the area under the receiver operating characteristic (ROC) curve, the c-index. The c-index is a measure of discrimination and is interpreted as being the probability that a randomly chosen patient with CAD will have a higher predicted probability of disease than a randomly chosen patient without CAD (18). An area under the ROC curve (AUC) of 0.5 corresponds to a model that provides no diagnostic information, whereas an AUC of 1.0 corresponds to a perfect diagnostic model.

STATA statistical analysis software v10.0 (StataCorp, Texas, USA) was used for logistic regression analysis.

Next, we quantified the effect of adding CCS to the model on the classification of patients into probability categories of CAD. Four probability categories were defined: < 30%, ≥30–50%, ≥50–70% and ≥70%. Reclassification tables were constructed for the Diamond & Forrester model and the Pryor model (see Tables 4 and 5) (19). We computed the reclassification calibration statistic (RCS) (20) which is equivalent to the Hosmer– Lemeshow statistic, applied to the cross-classified cells of the reclassification table with at least 20 observations. A significant result indicates a lack of fit.

Furthermore, the following reclassification measures were calculated for each model: the overall (correct) percentage of reclassification, the net reclassification improvement (NRI) (21) and the integrated discrimination improvement (IDI) (20). The NRI is the difference in proportions reclassifying to higher and lower probability categories among cases and non-cases. It is interpreted as the percentage reclassified, adjusted for the reclassification direction. A significant NRI indicates that classification improves when CCS is included. The IDI compares the difference in the average regression slope of cases and non-cases among the models with and without CCS. A significant IDI indicates that the new model performs better in discriminating cases and non-cases.

Reclassification computations were executed by using syntax made available by Cook and Ridker (20) in SAS Enterprise Guide v3 (SAS Inc, North Carolina, USA).

Results

Study population

During a 24-month period, 402 patients were enrolled. Patients with acute coronary syndrome or a history of myocardial infarction were excluded and 254 patients were left for data analysis (Table 1). Of these, 123 (48%) patients had obstructive CAD on CAG. CCS ranged from 0 to 3,839 with a median of 4 for patients without obstructive CAD and a median of 337 for patients with obstructive CAD. Of 131 patients without obstructive CAD, 44 (33.6%) patients had no coronary calcification; whereas of 123 patients diagnosed with obstructive CAD, 3 (2.4%) patients did not have any coronary calcification.

Systematic literature search

We obtained 649 articles in our literature search of which 632 articles were excluded based on title or abstract. After excluding 11 articles based on the full text, 6 studies were left for analysis.

Table 1. Patient characteristics

Variable	Total (n=254) no.	Patients without CAD (n=131) no.	Patients with CAD (n=123) no.	*P*-value
Mean Age (SD)	59 (11)	56 (12)	62 (10)	<0.001
Male sex	171 (67)	78 (60)	93 (76)	0.01
Typical chest pain	118 (46)	33 (25)	85 (69)	< 0.001
Mean BMI * (SD)	27 (4)	27 (4)	28 (5)	0.11
Smoking †	63 (25)	30 (23)	33 (27)	0.47
Hypertension	140 (55)	59 (45)	80 (66)	<0.001
Dyslipidaemia ‡	136 (54)	47 (36)	89 (72)	<0.001
Diabetes §	32 (13)	10 (8)	22 (18)	0.01
Family history	126 (50)	57 (44)	69 (56)	0.05
Mean calcium score || (SD)	346 (572)	132 (320)	574 (685)	<0.001
Median calcium score ||	138	4	337	-
CAD ¶ on CAG	123 (49)	0 (0)	123 (100)	

Unless otherwise specified, data are numbers of patients, with percentages in parentheses. CAD = obstructive coronary artery disease, CAG = catheter-based coronary angiography, SD = standard deviation.
** Body Mass Index, defined as weight/height^2 (in kg/m^2).*
† Past or current.
‡ Serum cholesterol >200 mg/dL or 5.18 mmol/L.
§ Plasma glucose ≥126 mg/dL or 7.0 mmol.
|| Measured according to Agatston (15).
¶ Defined as ≥50% stenosis in at least one vessel.

Multivariate logistic regression analysis

Five published prediction rules were validated (Tables 2 and 3): Diamond & Forrester (22), Pryor *et al.* (23), Morise *et al.* 1994 (24), Morise *et al.* 1997 (17) and Shaw *et al.* (25). The Diamond & Forrester prediction rule includes age, sex and type of chest pain, all of which were

significant predictors of obstructive CAD in our dataset and with an area under the ROC curve (AUC) of 0.798. Including CCS increased the AUC to 0.890, which was a statistically significant improvement (P < 0.001). In the expanded model age and sex were no longer significant predictors (Table 2).

Pryor *et al.* analyzed age, sex, type of chest pain, smoking, dyslipidaemia, diabetes and the interaction between age and smoking, age and dyslipidaemia, sex and smoking, and age and sex, of which type of chest pain and the presence of diabetes were significant predictors. This model resulted in an AUC of 0.838. After including CCS, the AUC increased to 0.901 which was a statistically significant improvement (P < 0.001). In the expanded model diabetes was no longer a significant predictor (Table 2).

Morise *et al.* (1994) included diabetes and dyslipidaemia in addition to the variables used by Diamond & Forrester and resulted in an AUC of 0.831. All variables were significant predictors of the presence of obstructive CAD. After including CCS, chest pain was the only variable that remained significant. After inclusion of CCS the AUC increased to 0.899, which was a statistically significant improvement (P < 0.001) (Table 2).

Morise *et al.* (1997) assessed age, sex, type of chest pain, smoking, dyslipidaemia, diabetes, estrogen status, hypertension, family history, obesity, BMI and the interaction between dyslipidaemia and family history. This model resulted in an AUC of 0.840. Age, sex, type of chest pain and dyslipidaemia were significant predictors. After including CCS, the AUC increased to 0.898, which was a significant model improvement (P < 0.001). Age, sex and dyslipidaemia were no longer significant predictors after the addition of CCS (Table 3).

Shaw *et al.* considered age, sex, typical chest pain, smoking, dyslipidaemia and diabetes and resulted in an AUC of 0.833. After including CCS, only type of chest pain remained a significant predictor and the AUC increased to 0.899 which was a statistically significant improvement (P < 0.001) (Table 3).

Reclassification tables for the Diamond & Forrester model and the Pryor model are presented in Tables 4 and 5. The addition of CCS to Diamond & Forrester resulted in reclassification of 47.2% of patients of whom 73.3% were correctly reclassified. The reclassification calibration statistic (RCS) indicated a strong lack of fit for the Diamond & Forrester model (P < 0.00001) which decreased substantially when CCS was added to the model (P < 0.01). The NRI (net reclassification improvement) was 33.6% (P < 0.0001) and the IDI (integrated discrimination improvement) was also statistically significant (18.8%, P < 0.001) indicating improvement in the classification of cases and non-cases in probability categories and improvement in discrimination between cases and non-cases.

For the model by Pryor *et al.*, 36.2% of the patients were reclassified, of whom 54.3% were correctly classified. The RCS indicated a lack of fit (P = 0.01), which decreased when CCS was added to the model (P = 0.03). The NRI was 24.0% (P < 0.0001) and the IDI was 14.8% (P < 0.001).

The reclassification measures for all models are presented in Table 6.

Table 2. Comparison of multivariate logistic regression models

	Model 1: Diamond & Forrester 1979				Model 2: Pryor *et al.* 1993				Model 3: Morise *et al.* 1994			
	No Calcium score		Calcium score		No Calcium score		Calcium score		No Calcium score		Calcium score	
Variables	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
ln CCS *			1.97	1.60-2.41			1.93	1.55-2.41			1.87	1.52-2.29
Age	1.05	1.02-1.08	0.98	0.95-1.02	1.02	0.95-1.09	0.93	0.86-1.01	1.05	1.01-1.08	0.98	0.95-1.02
Male sex	2.97	1.57-5.61	1.64	0.79-3.42	0.19	0.00-12.20	0.02	0.00-1.91	3.48	1.77-6.83	2.05	0.94-4.45
Typical chest pain	6.61	3.67-11.89	5.34	2.70-10.56	5.44	2.90-10.18	5.22	2.52-10.81	5.44	2.94-10.05	4.91	2.43-9.91
Smoking					0.08	0.00-9.40	0.06	0.00-11.40				
Dyslipidaemia					10.90	0.19-611	2.43	0.03-196	3.03	1.63-5.61	1.94	0.96-3.94
Diabetes					2.81	1.07-7.38	2.14	0.71-6.42	2.61	1.02-6.73	2.22	0.74-6.65
Age*Smoking					1.04	0.97-1.13	1.03	0.95-1.12				
Age*Dyslipidaemia					0.98	0.92-1.05	1.00	0.93-1.07				
Sex*Smoking					1.77	0.36-8.84	3.82	0.66-22.09				
Age*Sex					1.05	0.98-1.12	1.07	1.00-1.15				
Estrogen												
Hypertension												
Family History												
Dyslipidaemia*Family History												
Obesity												
BMI												
ROC curve †	0.798	0.742-0.854	0.890	0.851-0.930	0.838	0.789-0.887	0.901	0.863-0.938	0.831	0.780-0.881	0.899	0.861-0.937
LR test ‡	$P < 0.001$				$P < 0.001$				$P < 0.001$			

Odds ratios (ORs) in bold typeface are statistically significant.
* *Natural logarithm of CCS+1.*
† *Area under the receiver operating characteristic curve.*
‡ *Likelihood ratio test comparing model without CCS and model including CCS.*

Table 3. Comparison of multivariate logistic regression models

	Model 4: Morise *et al.* 1997				Model 5: Shaw *et al.* 1998			
	No Calcium score		Calcium score		No Calcium score		Calcium score	
Variables	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
ln CCS *			1.87	1.50-2.31			1.86	1.51-2.30
Age	1.05	1.01-1.09	0.99	0.95-1.03	1.05	1.02-1.08	0.98	0.95-1.02
Male sex	3.15	1.29-7.70	1.37	0.48-3.95	3.42	1.74-6.74	2.05	0.94-4.45
Typical pain	5.56	2.94-10.51	4.82	2.36-9.86	5.50	2.96-10.21	4.91	2.44-9.90
Smoking	1.53	0.72-3.24	1.03	0.44-2.39	1.63	0.80-3.30	1.04	0.47-2.27
Dyslipidaemia	3.20	1.29-7.95	1.80	0.62-5.22	3.04	1.63-5.66	1.95	0.96-3.94
Diabetes	2.66	0.98-7.26	2.01	0.61-6.61	2.85	1.10-7.39	2.24	0.74-6.75
Age*Smoking								
Age*Dyslipidaemia								
Sex*Smoking								
Age*Sex								
Estrogen	0.78	0.33-1.86	0.53	0.19-1.48				
Hypertension	1.83	0.93-3.60	1.31	0.62-2.79				
Family History	2.02	0.77-5.29	1.14	0.38-3.39				
Dyslipidaemia*Family History	0.74	0.21-2.60	1.15	0.28-4.82				
Obesity	0.88	0.31-2.48	0.65	0.20-2.12				
BMI	0.99	0.88-1.13	1.05	0.91-1.22				
ROC curve †	0.840	0.792-0.889	0.898	0.859-0.936	0.833	0.783-0.883	0.899	0.861-0.937
LR test ‡	$P < 0.001$				$P < 0.001$			

Odds ratios (ORs) in bold typeface are statistically significant
* *Natural logarithm of CCS+1.*
† *Area under the receiver operating characteristic curve.*
‡ *Likelihood ratio test comparing model without CCS and model including CCS*

Discussion

We analyzed the incremental value of CCS in the prediction of prevalent obstructive CAD. We showed CCS to be a significant predictor, independent of other variables included in the model. Furthermore, we confirm that the prediction of prevalent obstructive CAD is mainly determined by age, sex and type of chest pain. In four of the five models age and sex were significant predictors before the addition of CCS. However, after including CCS, both age and sex were no longer significant predictors in each of those models. Type of chest pain was a significant predictor of obstructive CAD, independent of other variables included in the model and independent of whether CCS was included or not. CCS proved to be an excellent and significant predictor of CAD with adjusted odds ratios close to 2. Apart from age and sex, most risk factors did not result in significant odds ratios which suggests that these risk factors are of minor importance in the prediction of prevalent obstructive CAD. Furthermore, the risk factors that were significant in the models without CCS lost their significance after including CCS.

Analysis of reclassification showed that adding CCS yields reclassification of 34–47% of patients, most of which was correct. The reclassification calibration statistic demonstrated a lack of fit for all models, which decreased when CCS was added in all models except for Shaw's.

Numerous studies have previously reported on the incremental value of CCS in the prediction of cardiovascular events and mortality in asymptomatic individuals (10–13). Some studies have reported on the incremental value of CCS in the prediction of prevalent obstructive CAD. Both Guerci *et al.* (1) and Kennedy *et al.* (2) studied the relationship between obstructive CAD and CCS, adjusting for various risk factors. Both studies found highly significant odds ratios for the CCS in predicting the presence of obstructive CAD, but the odds ratios were lower compared to what we found. Budoff *et al.* (4) evaluated the value of CCS in diagnosing CAD and showed that the addition of CCS increased the AUC from 0.672 to 0.842. However, the models included only age and sex as clinical variables. Considering the fact that the type of chest pain was not taken into account, an increase in AUC with CCS can be expected. However, the resulting area under the ROC curve for the model including CCS was similar to what we found. Schmermund *et al.* (3) previously studied the value of CCS in predicting the extent of CAD. They showed an independent and incremental value for CCS in multiple linear regression in predicting the total number of segments per patient with ≥50% stenosis.

Study limitations

Our study assessed the prediction of ≥50% stenosis in at least one vessel. One could argue that physicians are primarily interested in diagnosing severe CAD, as these patients would be eligible for revascularization whereas others can be adequately treated medically. Likewise, physicians might be primarily interested in predicting future cardiovascular adverse events. However, we did not consider prognosis in this analysis.

All patients in our study were referred for CAG based on their presentation or functional testing that suggested the presence of cardiac ischemia. In this way a high-risk population was selected, which could have biased our results. Unfortunately this limitation is inherent to the study design. Further research is necessary to determine the value of CCS in other (e.g. lower risk) populations.

Table 4. Probability of CAD: reclassification table after addition of CCS to Diamond and Forrester (model 1)

	Probability category based on model 1 + CCS				
Probability category based on model 1	<30%	≥30 – 50%	≥50 – 70%	≥70 – 100%	Total, n (%)
<30%					
N (%)	51 (67.1)	19 (25.0%)	4 (5.2)	2 (2.6)	76 (29.9)
Observed probability, %	11.8	36.8	50.0	100.0	22.4
≥30 – 50%					
N (%)	25 (39.7)	13 (20.6)	19 (30.2)	6 (9.5)	63 (28.4)
Observed probability, %	0.0	23.1	68.4	50.0	30.2
≥50 – 70%					
N (%)	4 (11.1)	5 (13.9)	6 (16.7)	21 (58.3)	36 (14.2)
Observed probability, %	25	80.0	50.0	66.7	61.1
≥70 – 100%					
N (%)	5 (6.3)	2 (2.5)	8 (10.1)	64 (81.0)	79 (31.1)
Observed probability, %	0.0	50.0	37.5	95.3	82.3
Total					
N (%)	85 (33.5)	39 (15.4)	37 (14.6)	93 (36.6)	254 (100)
Observed probability, %	8.2	38.5	56.8	86.0	48.4

CCS = CT coronary calcium score

Table 5. Probability of CAD: reclassification table after addition of CCS to the model by Pryor *et al.* (model 2)

	Probability category based on model 1 + CCS				
Probability category based on model 1	<30%	≥30 – 50%	≥50 – 70%	≥70 – 100%	Total, n (%)
<30%					
N (%)	76 (85.4)	10 (11.2)	2 (2.3)	1 (1.1)	89 (35.0)
Observed probability, %	9.2	40.0	100.0	100.0	15.7
≥30 – 50%					
N (%)	10 (27.0)	10 (27.0)	12 (32.4)	5 (13.5)	37 (14.6)
Observed probability, %	10.0	20.0	50.0	100.0	37.8
≥50 – 70%					
N (%)	7 (13.5)	6 (11.5)	15 (28.9)	24 (46.2)	52 (20.5)
Observed probability, %	0.0	50.0	73.3	66.7	57.7
≥70 – 100%					
N (%)	3 (4.0)	3 (4.0)	9 (11.8)	61 (80.3)	76 (29.9)
Observed probability, %	33.3	0.0	66.7	95.0	85.5
Total					
N (%)	96 (37.8)	29 (11.4)	38 (15.0)	91 (35.8)	254 (100)
Observed probability, %	9.4	31.0	65.8	87.9	48.4

CCS = CT coronary calcium score

Also, risk factors such as type of chest pain, smoking status and family history of CAD were obtained by interviewing the patient. Potentially, this method underestimates their predictive effects as compared to the predictive effect of the CCS, which was directly measured.

Tables 4 and 5 illustrate how CCS influences the classification of patients in probability categories. However, the limitations of reclassification measures in the context of this research should be taken into account. Our sample size was too small to reliably assess reclassification. For example, the RCS only uses the cross-classified cells containing at least 20 observations. In Table 4 only four cells contain 20 or more observations, implying that a substantial amount of (correctly) reclassified patients are ignored. Thus, in our study the reclassification percentages, NRI and IDI indices are more reliable than the RCS.

Ideally, the probability categories should be based on clinically relevant cutoffs. However, no well-established clinically relevant probability threshold exists. The probability of CAD is commonly defined as low (<30%), intermediate (≥30–70%) and high (≥70%) (26). In our view the intermediate category is rather wide, which is why we divided this category into low–intermediate (≥30–50%) and intermediate–high (≥50–70%). It should be noted that the overall percentage of reclassification is highly dependent on the choice and number of probability categories.

Clinical implications

Our results demonstrate that the estimation of the probability of obstructive CAD can be improved by including CCS. This implies that clinicians can make better decisions as to whether a particular patient would benefit from further testing, for example CCTA or CAG. In low-risk patients, a CCS of zero could exclude CAD and avoid further testing using CCTA. Hereby, one also avoids the intravenous administration of contrast agent, the extra radiation exposure, and extra scan time and costs associated with CCTA. In patients with a low CCS, CAG can be avoided and further non-invasive testing would be preferred. In patients with an intermediate CCS, a CCTA might be the optimal next step. In patients with a high CCS, direct CAG might be justified because of the high probability of CAD. All in all, CCS could be useful as a triage test for patients who are suspected of having CAD.

We confirmed that the prediction of significant CAD is primarily driven by the patient's symptoms. A detailed history of the patient's symptoms remains most important in the diagnostic work-up of patients with suspected CAD. However, history taking is difficult and subjective, therefore limiting our ability to accurately predict the presence of CAD. Hence, further diagnostic testing will be important, even in patients with a low to intermediate probability of CAD.

On the other hand, the harms and costs of obtaining CCS should be considered. Kim (27) studied the radiation dose and cancer risk of CT calcium screening (every 5 years) in asymptomatic individuals. They concluded that the excess lifetime cancer risk was 42 (62) per 100 000 men (women). It is important to note that our study assessed the value of a single CCS in symptomatic patients, for whom the excess lifetime cancer risk will be lower and small compared with the risk of missing a CAD diagnosis. Moreover, CCS could reduce the use of additional testing in patients with a low CCS and a low probability of CAD, thereby reducing the total radiation exposure.

Although performing a CCS measurement is a fast, low dose and relatively inexpensive procedure, the harms and benefits should be considered in a cost-effectiveness analysis.

Conclusion

Our results suggest that CCS significantly improves the prediction of prevalent obstructive CAD, independent of other clinical variables. Therefore, CCS should be considered in the diagnostic work-up of CAD.

Table 6. Reclassification measures obtained by adding CCS to the existing prediction models

Model	Reclassification Percentages													
	Overall	Reclassified from ≥30 -50%	Reclassified from ≥50 – 70%	% Correct *	Chi-squared excluding CCS †	*P*-value	Chi-squared including CCS ‡	*P*-value	NRI, %	*P*-value for NRI	Reclassification Improvement (cases) §	Reclassification Improvement (non-cases) §	IDI, %	*P*-value for IDI
1	47.2	79.4	83.3	73.3	25.01	<0.00001	10.91	<0.01	33.6	<0.0001	<0.0001	0.23	18.8	<0.001
2	36.2	73.0	71.2	54.3	6.57	0.01	4.70	0.03	24.0	0.001	<0.001	0.31	14.8	<0.001
3	38.2	77.5	75.5	63.9	4.14	0.04	3.31	0.07	21.8	0.005	<0.001	0.58	14.9	<0.001
4	34.3	74.4	74.0	83.9	3.82	0.05	1.58	0.21	24.9	<0.001	<0.001	0.29	13.3	<0.001
5	37.8	77.5	75.9	69.8	3.65	0.06	5.14	0.02	22.6	0.003	<0.001	0.58	14.4	<0.001

CCS = computed tomography coronary calcium score, NRI = net reclassification improvement, IDI = integrated discrimination improvement.

** If the predicted probability of obstructive CAD of the model including CCS was closer to the observed probability of CAD compared with the prediction of the model without CCS, the reclassification was considered to be correct.*

† Reclassification calibration (Hosmer–Lemeshow) statistic for model without CCS, using cells from the reclassification table with at least 20 observations.

‡ Reclassification calibration statistic for models including CCS, using cells from the reclassification table with at least 20 observations.

§ The reclassification improvement is defined as the difference in proportions of patients moving up and down for cases and non-cases separately.

Appendix

Studies were included if they met all of the following criteria: (1) Study population were patients with chest pain suggestive of stable angina pectoris, (2) study reported a new multivariate prediction rule that included clinical variables, (3) outcome was the presence of significant CAD defined as ≥50% stenosis in at least one vessel. Studies were excluded if they met one of the following criteria: (1) Article was a review, guideline, or cost-effectiveness analysis, or (2) study did not report a prediction model, or (3) the outcome of interest was prognostic (e.g. event rate after treatment) or the authors did not use ≥50% stenosis in at least one vessel as their outcome, or (4) study population consisted of unstable patients, asymptomatic patients, patients with known CAD or patients with a specific co-morbidity, or (5) main topic was a diagnostic (imaging) test, or (6) a reference standard other than conventional coronary angiography was used, or (7) authors did not use the traditional classification of typical and atypical chest pain, or (8) the study focused on the association between one particular risk factor and CAD. Articles not accessible were excluded from the analysis.

Search strategy

1. Coronary artery disease [MeSH]
2. Coronary heart disease [MeSH]
3. Coronary stenosis [MeSH]
4. Coronary disease
5. #1 OR #2 OR 3# OR 4
6. Chest pain [MeSH]
7. Angina pectoris [MeSH]
8. Suspected [title/abstract]
9. #6 OR #7 OR #8
10. Logistic models [MeSH]
11. Probability [MeSH]
12. Risk [MeSH]
13. Models, statistical [MeSH]
14. ROC curve [MeSH]
15. #10 OR #11 OR #12 OR #13 OR #14
16. Medical history taking [MeSH]
17. Physical examination [MeSH]
18. Clinical [title/abstract]
19. #16 OR #17 OR #18
20. #5 AND #9 AND #15 AND #19
21. Animals [MeSH] NOT humans [MeSH]
22. Editorial [publication type]
23. Comment [publication type]
24. Letter [publication type]
25. Meta-analysis [publication type]
26. Case reports [publication type])
27. #22 OR #23 OR #24 OR #25 OR #26
28. Acute [title/abstract]
29. #20 NOT #21 NOT #27 NOT #28

References

1. Guerci AD, Spadaro LA, Goodman KJ, Lledo-Perez A, Newstein D, Lerner G, Arad Y (1998) Comparison of electron beam computed tomography scanning and conventional risk factor assessment for the prediction of angiographic coronary artery disease. J Am Coll Cardiol 32:673–679
2. Kennedy J, Shavelle R, Wang S, Budoff M, Detrano RC (1998) Coronary calcium and standard risk factors in symptomatic patients referred for coronary angiography. Am Heart J 135:696–702
3. Schmermund A, Denktas AE, Rumberger JA, Christian TF, Sheedy PF 2nd, Bailey KR, Schwartz RS (1999) Independent and incremental value of coronary artery calcium for predicting the extent of angiographic coronary artery disease: comparison with cardiac risk factors and radionuclide perfusion imaging. J Am Coll Cardiol 34:777–786
4. Budoff MJ, Diamond GA, Raggi P, Arad Y, Guerci AD, Callister TQ, Berman D (2002) Continuous probabilistic prediction of angiographically significant coronary artery disease using electron beam tomography. Circulation 105: 1791–1796
5. Knez A, Becker A, Leber A, White C, Becker CR, Reiser MF, Steinbeck G, Boekstegers P (2004) Relation of coronary calcium scores by electron beam tomography to obstructive disease in 2, 115 symptomatic patients. Am J Cardiol 93:1150–1152
6. Haberl R, Becker A, Leber A, Knez A, Becker C, Lang C, Bruning R, Reiser M, Steinbeck G (2001) Correlation of coronary calcification and angiographically documented stenoses in patients with suspected coronary artery disease: results of 1,764 patients. J Am Coll Cardiol 37:451–457
7. Greenland P, Bonow RO, Brundage BH, Budoff MJ, Eisenberg MJ, Grundy SM, Lauer MS, Post WS, Raggi P, Redberg RF, Rodgers GP, Shaw LJ, Taylor AJ, Weintraub WS, American College of Cardiology Foundation Clinical Expert Consensus Task Force, Society of Atherosclerosis Imaging and Prevention, Society of Cardiovascular Computed Tomography (2007) ACCF/ AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/ AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography) developed in collaboration with the Society of Atherosclerosis Imaging and Prevention and the Society of Cardiovascular Computed Tomography. J Am Coll Cardiol 49:378–402
8. Sarwar A, Shaw LJ, Shapiro MD, Blankstein R, Hoffman U, Cury RC, Abbara S, Brady TJ, Budoff MJ, Blumenthal RS, Nasir K (2009) Diagnostic and prognostic value of absence of coronary artery calcification. J Am Coll Cardiol Cardiovasc Imaging 2:675–688
9. Shareghi S, Ahmadi N, Young E, Gopal A, Liu ST, Budoff MJ (2007) Prognostic significance of zero coronary calcium scores on cardiac computed tomography. J Cardiovasc Comput Tomogr 1:155–159
10. Arad Y, Goodman KJ, Roth M, Newstein D, Guerci AD (2005) Coronary calcification, coronary disease risk factors, C-reactive protein, and atherosclerotic cardiovascular disease events: the St. Francis Heart Study. J Am Coll Cardiol 46:158–165
11. Greenland P, LaBree L, Azen SP, Doherty TM, Detrano RC (2004) Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. JAMA 291:210–215
12. Raggi P, Cooil B, Callister TQ (2001) Use of electron beam tomography data to develop models for prediction of hard coronary events. Am Heart J 141: 375–382
13. Shaw LJ, Raggi P, Schisterman E, Berman DS, Callister TQ (2003) Prognostic value of cardiac risk factors and coronary artery calcium screening for all-cause mortality. Radiology 228:826–833
14. Austen WG, Edwards JE, Frye RL, Gensini GG, Gott VL, Griffith LS, McGoon DC, Murphy ML, Roe BB (1975) 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 51:5–40

15. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M Jr, Detrano R (1990) Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol 15: 827–832
16. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD (2005) Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 58:475–483
17. Morise AP, HaddadWJ, Beckner D (1997) Development and validation of a clinical score to estimate the probability of coronary artery disease in men and women presenting with suspected coronary disease. Am J Med 102:350–356
18. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36
19. Janes H, Pepe MS, Gu W (2008) Assessing the value of risk predictions by using risk stratification tables. Ann Intern Med 149:751–760
20. Cook NR, Ridker PM (2009) Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med 150:795–802
21. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 27:157–172, discussion 207–112
22. Diamond GA, Forrester JS (1979) Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med 300:1350–1358
23. Pryor DB, Shaw L, McCants CB, Lee KL, Mark DB, Harrell FE Jr, Muhlbaier LH, Califf RM (1993) Value of the history and physical in identifying patients at increased risk for coronary artery disease. Ann Intern Med 118:81–90
24. Morise AP, Bobbio M, Detrano R, Duval RD (1994) Incremental evaluation of exercise capacity as an independent predictor of coronary artery disease presence and extent. Am Heart J 127:32–38
25. Shaw LJ, Peterson ED, Shaw LK, Kesler KL, DeLong ER, Harrell FE Jr, Muhlbaier LH, Mark DB (1998) Use of a prognostic treadmill score in identifying diagnostic coronary disease subgroups. Circulation 98:1622–1630
26. Meijboom WB, van Mieghem CA, Mollet NR, Pugliese F, Weustink AC, van Pelt N, Cademartiri F, Nieman K, Boersma E, de Jaegere P, Krestin GP, de Feyter PJ (2007) 64-slice computed tomography coronary angiography in patients with high, intermediate, or low pretest probability of significant coronary artery disease. J Am Coll Cardiol 50:1469–1475
27. Kim KP, Einstein AJ, Berrington de Gonzalez A (2009) Coronary artery calcification screening: estimated radiation dose and cancer risk. Arch Intern Med 169:1188–1194.

Chapter 7.1

A Clinical Prediction Rule for the Diagnosis of Coronary Artery Disease: Validation, Updating, and Extension

Eur Heart J. 2011;32(11):1316-30.

Tessa S.S. Genders
Ewout W. Steyerberg
M.G. Myriam Hunink

On behalf of the CAD consortium collaborators (see page 339)

Abstract

Aims

The aim was to validate, update, and extend the Diamond–Forrester model for estimating the probability of obstructive coronary artery disease (CAD) in a contemporary cohort.

Methods and results

Prospectively collected data from 14 hospitals on patients with chest pain without a history of CAD and referred for catheter-based coronary angiography (CAG) were used. Primary outcome was obstructive CAD, defined as ≥50% stenosis in one or more vessels on CAG. The validity of the Diamond–Forrester model was assessed using calibration plots, calibration-in-the-large, and recalibration in logistic regression. The model was subsequently updated and extended by revising the predictive value of age, sex, and type of chest pain. Diagnostic performance was assessed by calculating the area under the receiver operating characteristic curve (c-statistic) and reclassification was determined. We included 2260 patients, of whom 1319 had obstructive CAD on CAG. Validation demonstrated an overestimation of the CAD probability, especially in women. The updated and extended models demonstrated a c-statistic of 0.79 (95% CI: 0.77–0.81) and 0.82 (95% CI: 0.80–0.84), respectively. Sixteen per cent of men and 64% of women were correctly reclassified. The predicted probability of obstructive CAD ranged from 10% for 50-year-old females with non-specific chest pain to 91% for 80-year-old males with typical chest pain. Predictions varied across hospitals due to differences in disease prevalence.

Conclusion

Our results suggest that the Diamond–Forrester model overestimates the probability of CAD especially in women. We updated the predictive effects of age, sex, type of chest pain, and hospital setting which improved model performance and we extended it to include patients of 70 years and older.

Introduction

In patients presenting with chest pain suggestive of stable angina pectoris, numerous diagnostic strategies can be used. The reference standard for diagnosing coronary artery disease (CAD) is catheter-based coronary angiography (CAG). However, CAG is expensive and involves a small risk of complications and death (1). Therefore, non-invasive testing is recommended to select patients who will benefit from CAG (2-4). The clinical value of non-invasive diagnostic tests depends on the test sensitivity, the specificity, the potential gain from making the correct diagnosis, the potential harm caused by false-positive test results, and the pre-test (prior) probability of the suspected disease (5-7). In choosing the appropriate test for a particular patient with chest pain suggestive of CAD, the pre-test probability of CAD is crucial (7).

Diamond and Forrester (8) demonstrated the importance of the pre-test probability on interpreting test results in their classic paper in 1979. Using estimates from autopsy and cross-sectional studies, they developed a simple but elegant model that considers age, sex, and type of chest pain to estimate the probability of obstructive CAD in patients between 30 and 70 years old. In spite of its limitations, the Diamond–Forrester model is still used in current guidelines (2-4). Although other cardiovascular risk factors such as diabetes, smoking and dyslipidaemia have been included in, e.g. the Duke Clinical Score (9,10) the predictive effects of other risk factors in diagnostic models are often small compared with the predictive effects of age, sex, and type of chest pain. Furthermore, complicated models are less likely to be used by physicians in clinical practice especially since non-invasive diagnostic tests are commonly ordered immediately at the first visit. The Diamond–Forrester model allows the immediate calculation of an estimate of the patients' pre-test risk of CAD, without the need to wait for laboratory findings or exercise test results.

Since the Diamond–Forrester model was developed >30 years ago and based on data from the USA only, our aims were to study the validity of the Diamond–Forrester predictions in estimating the probability of obstructive CAD, to update the predictions using recently collected data, and to extend the model for patients beyond the age of 70, using data from contemporary cohorts.

Methods

The CAD Consortium

The CAD Consortium is part of the European network for the Assessment of Imaging in Medicine (EuroAIM), which is an initiative of the European Institute of Biomedical Imaging Research (EIBIR) (11). The main goal of EuroAIM is to perform pooled analyses of existing prospectively collected data, which will improve power of the studies and increase generalizability of the results obtained.

Data collection

A consortium of researchers from various countries in Europe and the USA was formed. An existing database with prospectively collected data on 80 or more eligible patients was required for participation in this consortium. All patients included in this consortium had to be enrolled in single-centre studies, for which local approval from the Institutional Review

Board and signed informed consent had been obtained. Participation in the consortium did not involve any financial incentives.

Patient population

Our patient population consists of patients with chest pain, suggestive of stable angina pectoris. Patients were eligible for the analysis if they presented with stable chest pain (typical, atypical, or non-specific chest pain) and if CAG was performed. Patients were excluded if they met one of the following criteria: (i) acute coronary syndrome or unstable chest pain, (ii) history of myocardial infarction or previous revascularization (percutaneous coronary intervention or coronary artery bypass graft surgery), and (iii) no informed consent.

Only patients who underwent CAG were eligible for this analysis, implying that our population was highly selected. To explore the effect of selection bias (i.e. verification bias) in this study population, we performed a separate analysis using inverse probability weighting on each patient in the data set. Based on an independent registry data set (not included in our consortium) consisting of unselected outpatients presenting to the cardiologist who were intended to undergo both stress-ECG and coronary CT angiography (CCTA) for the evaluation of chest pain (12,13) a logistic regression analysis was performed to calculate the probability of undergoing CAG (i.e. the probability of verifying the presence of CAD with CAG) depending on age, sex, and type of chest pain. Each patient was subsequently weighted with the inverse of the probability of verification which corrects for verification bias (Appendix, Correction for verification bias) (14,15).

Clinical definitions

Data on age, sex, type of chest pain, and the presence of CAD were collected. Type of chest pain was classified as being typical, atypical, or non-specific. Typical chest pain was defined as having (i) substernal chest pain or discomfort, that is (ii) provoked by exertion or emotional stress and (iii) relieved by rest and/or nitroglycerine. Atypical chest pain was defined as having two of the before-mentioned criteria. If one or none of the criteria was present, the patient was classified as having non-specific chest pain (4,16). The presence of obstructive CAD was defined as one or more vessels with ≥50% lumen diameter reduction on CAG. As we used existing databases, CAG was performed at each institution according to local protocols; both visual assessment and quantitative assessment were allowed for interpretation of the CAG. Indicator variables for hospital were used to allow adjustment for hospital.

Data analysis

Validation and updating of the Diamond–Forrester model was performed using state-of-the-art methods (17-19) by one of the authors (T.S.S.G.). See Appendix for a more detailed description of the methods.

Validation

The Diamond–Forrester model takes into account age, sex, and type of chest pain and was developed for patients between 30 and 70 years. For validation, we therefore excluded both patients below the age of 30 and patients above the age of 69.

The observed frequency of CAD in our data set was calculated stratifying for 10-year age category, sex, and type of chest pain (i.e. equivalent to the Diamond–Forrester categories).

Table 1. Patient characteristics, ordered according to the prevalence of coronary artery disease

Hospital, country	Medical University of South Carolina, Charleston, U.S.A.	Turku University Hospital, Finland	Innsbruck Medical University, Austria	Maasstad Ziekenhuis, Rotterdam, ·The Netherlands	Leiden University Medical Center, The Netherlands	Erasmus University Medical Center, The Netherlands	Charité Medical School, Berlin, Germany	University Medical Center Utrecht, The Netherlands	Papworth Hospital NHS Foundation Trust, Cambridge, UK
N	99	98	101	90	99	289	186	85	83
Mean age (SD)	60.3 (10.4)	64.4 (6.8)	61.2 (8.6)	59.0 (9.9)	62.1 (10.9)	59.4 (10.6)	62.7 (9.4)	60.5 (5.4)	65.5 (10.8)
Male sex (%)	49.5	58.2	58.4	58.9	56.6	68.5	71.0	67.1	65.1
Typical CP (%)	76.8	50.0	21.8	10.0	62.6	50.2	57.0	55.3	41.0
Atypical CP (%)	20.2	42.9	78.2	27.8	35.4	26.0	29.0	18.8	36.1
Non-specific CP (%)	3.0	7.1	0	62.2	2.0	23.9	14.0	25.9	22.9
CAD on CAG (%)	39.4	42.9	45.5	50.0	51.5	52.2	54.8	57.6	59.0

Hospital, country	University Hospital Parma, Italy	The Essex Cardiothoracic Center, UK	University Hospital Leuven. Belgium	University Hospital Zurich, Switzerland	Federal Center for Medicine and Rehabilitation – Moscow, Russia	Total (CAD+)	Total (CAD-)	Total
N	277	110	100	549	106	1325	947	2272
Mean age (SD)	59.9 (11.4)	63.2 (9.5)	64.0 (10.7)	65.6 (10.9)	59.7 (8.4)	63.8	60.3 (10.9)	62.3 (10.4)
Male sex (%)	67.9	68.2	59.0	74.5	76.4	78.0	52.2	67.2
Typical CP (%)	65.0	52.7	58.0	53.4	56.6	69.3	30.2	53.0
Atypical CP (%)	35.0	24.5	28.0	10.9	24.5	19.1	37.3	26.7
Non-specific CP (%)	0.0	22.7	14.0	35.7	18.9	11.6	32.5	20.3
CAD on CAG (%)	60.0	61.8	63.0	68.1	75.5	100	0	58.3

SD = standard deviation, CP = chest pain, CAD = coronary artery disease, CAG = catheter-based coronary angiography.

Observed frequencies of CAD were compared with the estimates according to Diamond and Forrester and subsequently tabulated and presented by means of calibration plots.

Calibration-in-the-large – When assessing the validity of a prediction model, the first step is to check whether the average prediction is equal or close to the average observed outcome. This concept is referred to as 'calibration-in-the-large'. Hereto, we compared the mean observed frequency of CAD with the mean prediction according to Diamond and Forrester.

Recalibration – The second step is to test whether the overall effect of the predictors in the Diamond–Forrester model is valid for the consortium data.

Re-estimation – The third step is to re-estimate the predictor effects in the consortium data and to compare the effects with the original effects according to Diamond and Forrester. Subsequently, we calculated the difference between the re-estimated and the original effects (Appendix, Table A1).

Updating and extension

For updating the Diamond–Forrester model, we used all data including patients <30 and >69 years. Age, sex, and type of chest pain were entered simultaneously in a logistic regression model. All analyses were performed both unadjusted and adjusted for hospital.

Separately, we extended the model with a random intercept to allow for heterogeneity in CAD prevalence across different hospitals (Appendix, Table A3). Because of potential differences in symptom classification across hospitals, we also tested random effects for type of chest pain between different hospitals.

Diagnostic performance was quantified by calculating the area under the receiver operating curve (c-statistic). Calibration was assessed using the Hosmer–Lemeshow goodness-of-fit test and a calibration plot (Appendix, Figure A2).

Reclassification was assessed by cross-tabulating the probability classification of patients according to the Diamond–Forrester model vs. the updated model. We analyzed reclassification with the commonly used cutoffs of 30 and 70% (Appendix, Table A4, A6) (20). In a second reclassification analysis, we analyzed the cutoff 40% (Appendix, Table A5, A7), which is the threshold below which CCTA is cost-effective (7). Furthermore, we determined the reclassification calibration statistic and the net reclassification improvement (Appendix, Table A8) (21,22).

Validation of the updated model

As mentioned, the CAD Consortium population is highly selected based on referral to CAG. To assess the performance of the updated model in a low-risk population, we validated the updated model in an independent data set (not included in our consortium) consisting of outpatients presenting to the cardiologist who were intended to undergo both stress-ECG and CCTA for the evaluation of chest pain (12,13). These data were also used to perform the 'correction for verification bias'. We followed the step-wise approach as described above. Since not all patients in this population underwent CAG, the CCTA results were used as a proxy for the reference standard in patients who did not undergo CAG.

A P-value < 0.05 was considered statistically significant. Statistical analyses were performed using Stata/SE 10.1 (StataCorp, TX, USA). Reclassification was assessed by using syntax made available by Cook *et al.* (22) in SAS Enterprise Guide v3 (SAS, Inc., NC, USA).

Results

Data collection and study population

Existing databases with prospectively collected data were retrieved from 14 hospitals (Table 1). After excluding 12 cases because of missing values, the total study population consisted of 2260 patients (1521 men, 739 women, mean age 62 [range 21–93, IQR 55–70]), of whom 1319 (58.4%) were found to have obstructive CAD on CAG.

For validation, we excluded patients below the age of 30 (n = 7) and above the age of 69 (n = 570). Therefore, 1683 patients (1159 [68.9%] men, 524 women) and 937 (55.7%) with obstructive CAD were left for validation.

Data analysis

Validation

Table 2 and Figures 1 and 2 show the average observed frequency of CAD in men and women, respectively, as observed in our data set, compared with the prediction according to Diamond and Forrester. In men, we observed a slight overestimation of the probability of CAD by Diamond and Forrester for patients with atypical and typical chest pain. In women, the overestimation was more pronounced. The validation of predictions for men <40 and women <50 years old was less reliable, due to the limited number of patients in these subgroups.

Calibration-in-the-large – We found that the average predicted probability according to Diamond and Forrester was higher compared with the CAD frequency in the consortium data ($P < 0.001$).

Recalibration – Recalibration demonstrated that the overall effect of the predictors in the Diamond–Forrester model was higher compared with the overall effect of the predictors in the consortium data ($P < 0.001$).

Re-estimation – The effects of age ($P < 0.001$) and type of chest pain ($P < 0.001$) were significantly larger in the Diamond–Forrester model compared with the effects in the consortium data. We conclude that these predictors require updating.

Updating and extension

The updated model (Table 3) showed highly significant effects for age, sex, and type of chest pain, with similar effect sizes after adjustment for hospital (not shown). The area under the receiver operating characteristic curve demonstrated good performance (c-statistic, 0.79 [95% CI: 0.77–0.81]). After correction for verification bias, the regression coefficients were similar (Appendix, Table A3).

Interactions between the main effects did not show statistical significance (not shown). We found significant interactions between 'typical chest pain' and hospital (not shown),

implying that the effect of having typical chest pain on the predicted probability is different across hospitals.

Table 2. Head-to-head comparison of the pre-test probability according to Diamond and Forrester, the observed frequency of CAD, and the prediction according to the updated model.

	Men			Women		
	Diamond and Forrester	Observed frequency of CAD	Updated model *	Diamond and Forrester	Observed frequency of CAD	Updated model *
Typical chest pain						
30-39	69.7	90.0	**59.1**	25.8	†	27.5
40-49	87.3	**73.2**	**68.9**	55.2	**41.4**	**36.7**
50-59	92.0	**82.0**	**77.3**	79.4	**38.5**	**47.1**
60-69	94.3	**86.1**	**83.9**	90.6	**56.9**	**57.7**
70-79	–	88.6	88.9	–	70.8	67.7
>80	–	86.2	92.5	–	92.3	76.3
Atypical chest pain						
30-39	21.8	**16.7**	28.9	4.2	†	9.6
40-49	46.1	**28.9**	**38.4**	13.3	19.2	14.0
50-59	58.9	**45.9**	**48.9**	32.4	**24.2**	**20.0**
60-69	67.1	**53.9**	**59.4**	54.4	**30.1**	**27.7**
70-79	–	63.6	69.2	–	35.6	37.0
>80	–	83.3	77.5	–	†	47.4
Non-anginal chest pain						
30-39	5.2	14.3	17.7	0.8	†	5.3
40-49	14.1	20.9	24.8	2.8	**0.0**	8.0
50-59	21.5	28.6	33.6	8.4	22.2	11.7
60-69	28.1	50.5	43.7	18.6	**18.0**	**16.9**
70-79	–	62.1	54.4	–	22.0	23.8
>80	–	45.0	64.6	–	20.0	32.3

Subgroup estimates that are overestimated by the Diamond– Forrester model are printed in boldface.
** Probabilities shown reflect the estimates for patients aged 35, 45, 55, 65, 75, and 85 years.*
† Data on subgroups with less than five observations are not shown.

Finally, a random effects logistic regression model showed that there was substantial heterogeneity in disease prevalence across hospitals. The model could not be improved by adding a random effect for atypical chest pain, whereas a random effect for typical chest pain showed statistical significance (Appendix, Table A3). The random effects model is available for online use (Figure 3). The test for a non-linear age effect was not statistically significant, which is why we assumed a linear effect for age in all models.

Analysis of reclassification in 30–69-year-old patients demonstrated that 16% of men and 64% of women reclassified correctly, when using the probability categories ,30, ≥30–70, and ≥70% (Figures 4 and 5). The net reclassification index was negative for both men and women, which is explained by the fact that the updated model predicts less high probabilities compared with Diamond and Forrester, resulting in down classification of patients among both cases and non-cases (Appendix, Table A8).

Table 3. Updating and extension of the Diamond– Forrester model (random effects model)

	Coefficient	Odds ratio	95% CI lower limit	95% CI upper limit
Age	0.04	1.04	1.03	1.05
Male sex	1.34	3.82	3.08	4.74
Typical chest pain	1.91	6.72	3.97	11.37
Atypical chest pain	0.64	1.89	1.38	2.59
Non-specific chest pain *	–	–		
Intercept	-4.37	–		
c-Statistic (95% CI †)	0.82		0.80	0.84

** Reference category.*
† Estimated by bootstrapping with 1000 repetitions.

Validation of the updated model

Calibration-in-the-large – The average predicted probability according to our updated model was significantly higher compared with the CAD frequency in the independent data set ($P < 0.001$).

Recalibration – Recalibration demonstrated that the overall effect of the predictors in the updated model is similar to the overall effect of the predictors in the independent registry of outpatients ($P = 0.79$). From this, we conclude that the predictor effects are valid and do not require updating (Appendix, Table A9).

The model, adjusted for the lower disease prevalence, is available for online use (Figure 3).

Figure 1. Predicted probability of obstructive coronary artery disease in men (triangles) for the Diamond–Forrester age categories a: 30–39, b: 40–49, c: 50–59, and d: 60–69, vs. the observed frequency of obstructive coronary artery disease in our data. The legend provides the number of patients per age category for each type of chest pain. The bars indicate the 95% confidence interval of the observed frequency.

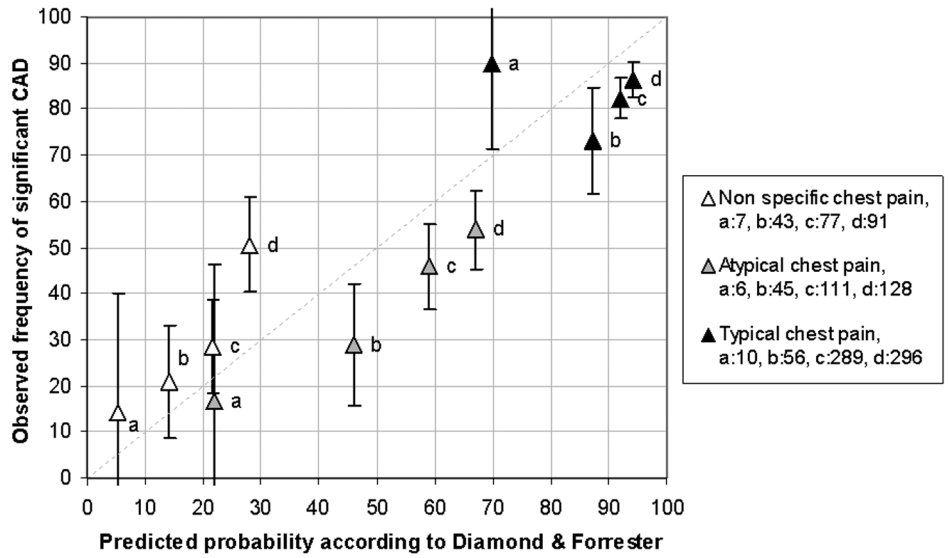
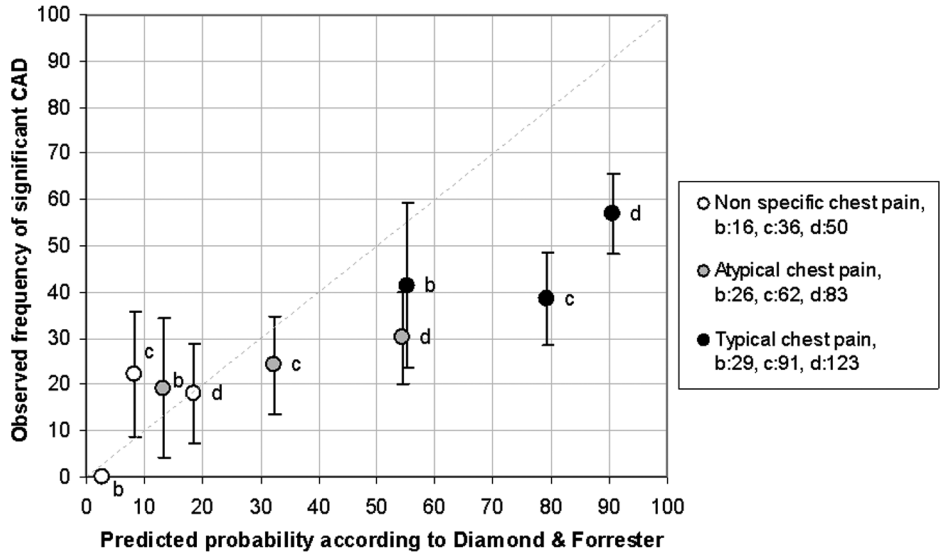


Figure 2. Predicted probability of obstructive coronary artery disease in women (circles) for the Diamond–Forrester age categories a: 30– 39, b: 40–49, c: 50–59, and d: 60–69, vs. the observed frequency of obstructive coronary artery disease in our data. The legend provides the number of patients per age category for each type of chest pain. Data on subgroups with less than five observations are not shown. The bars indicate the 95% confidence interval of the observed frequency.



Discussion

Using recently collected data and modern statistical methods, we assessed whether the Diamond–Forrester model is valid in a contemporary mainly European cohort (Box 1). Furthermore, we updated and extended the model by re-estimating the predictive effects of age, sex, and type of chest pain. We observed that the prevalence of CAD was different across hospitals and adjusted for this difference.

In short, we validated the Diamond–Forrester predictions in patients between 30 and 70 years old. By comparing the Diamond–Forrester predicted probabilities with the average observed frequencies of obstructive CAD in our data using calibration plots, we demonstrated that there is a tendency for the Diamond–Forrester model to overestimate the prevalence of CAD in a contemporary cohort. We showed that the Diamond–Forrester model needs to be updated according to the overall disease prevalence for the current European situation. Furthermore, we demonstrated that the predictor effects in the consortium data are not as extreme as the model by Diamond and Forrester suggests.

It is important to note that the Diamond–Forrester model is based on 30–40-year-old data from the USA. To some extent, the differences we demonstrated are explained by changes in the risk factor distributions over the past decades, as well as by the differences between populations from the USA and Europe.

All-in-all, we demonstrated that the validity of the Diamond– Forrester model for current practice is limited, which justifies updating. Furthermore, Diamond and Forrester did not

provide any estimates for patients above the age of 69. Nowadays, a substantial proportion of patients with chest pain will be at least 70 years old, which motivated the extension of the Diamond–Forrester model. Therefore, we re-estimated the predictive effect of age, sex, and type of chest pain in all patients and confirmed the importance of these in the prediction of obstructive CAD. We also demonstrated that predictions may vary across different hospitals. The performance of the updated model improved discrimination and calibration, as indicated by the c-statistic and the non-significant Hosmer–Lemeshow test. Measures of reclassification indicated correct reclassification by the updated model of a substantial proportion of patients, especially in women.

Box 1. Study strengths and weaknesses

Strengths
- A large contemporary cohort was studied.
- Multicentre collaboration of 14 hospitals.
- Modern statistical methods were used for validation, updating, and extension of the Diamond–Forrester model.
- In contrast to the Diamond–Forrester model, the updated model uses age as a continuous predictor and also predicts probabilities for patients 70 years and older.
- An easy-to-use online probability calculator was developed.

Weaknesses
- Only age, sex, and type of chest pain were considered predictors.
- The primary outcome was limited to obstructive CAD vs. no obstructive CAD.
- Existing databases designed for other research objectives were combined.
- A high-risk population was selected by including only patients referred for coronary angiography.
- Heterogeneity across hospitals with respect to the assessment of chest pain, adherence to guidelines, and interpretation of the coronary angiographies.

Updated predicted probabilities of obstructive CAD ranged from 10% for a 50-year-old female with non-specific chest pain to 91% for an 80-year-old man with typical chest pain. Prior studies assessed the performance of the Diamond–Forrester model (23) or developed new models to estimate the pre-test probability of CAD (9,10,24). However, in spite of efforts to develop new prediction models, the Diamond–Forrester model remains a common method for estimating the probability of obstructive CAD (2,4).

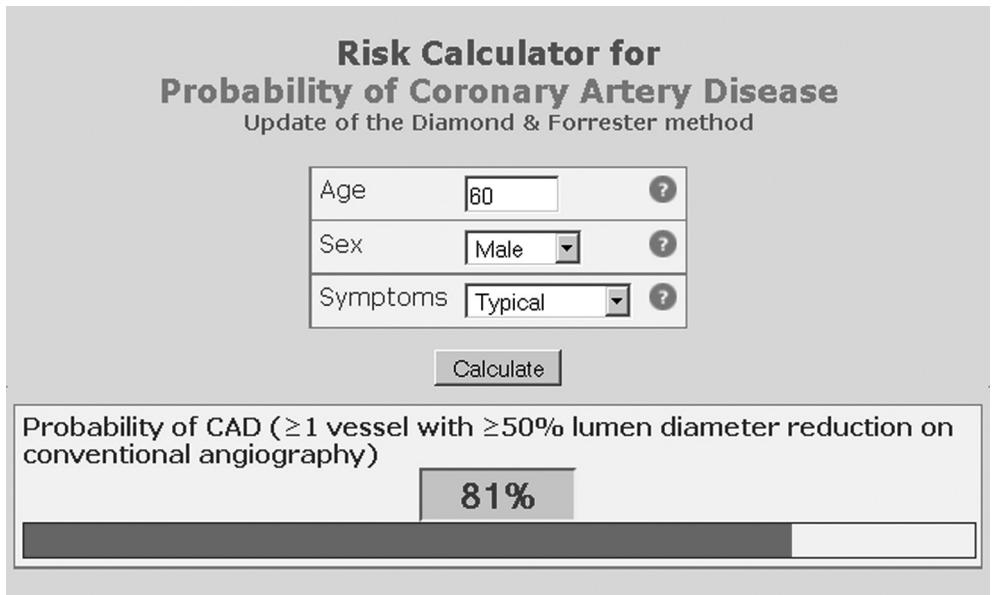
Limitations

The study population was derived from existing databases, some of which were designed for other research objectives (e.g. to investigate the diagnostic accuracy of non-invasive imaging tests for CAD). In some studies, all patients underwent the reference standard test, whereas in other studies, patients were selected for CAG based on the results of an index non-invasive test. When evaluating the diagnostic performance of the index non-invasive test, selection based on the test results may lead to 'verification bias' or 'referral bias' which

may bias estimates of sensitivity and specificity of such a test. In the current study, however, we did not assess the diagnostic performance of a non-invasive test, making bias less likely. To explore the potential effect of verification bias, we used inverse probability weighting on each individual with the probability of verification based on age, sex, and type of chest pain. The analyses with or without correction for verification bias yielded similar results, indicating that our results are unlikely to be biased by selected verification.

The fact that we only included patients who underwent CAG indicates the selection process that occurred in this study. Some of the patients will have had prior positive stress testing or another non-invasive imaging test. However, the association between age, sex, type of chest pain, and the presence of obstructive CAD on CAG does not depend on whether another diagnostic imaging test was performed and whether it was positive or not. The elderly, males, and patients with typical chest pain are more likely to undergo stress testing. Such patients are therefore overrepresented in our population. Incorporating the test results from the non-invasive test in the prediction model would influence the prediction of the probability of obstructive CAD. However, our aim was to predict the presence of obstructive CAD on initial presentation prior to diagnostic testing to provide decision support for the decision to test. The presence of obstructive CAD was determined by CAG. As existing databases were used, CAGs were carried out at each individual hospital according to local protocols. Some institutions used quantitative coronary angiography to determine the degree of stenosis, whereas others used visual assessment. All in all, heterogeneity due to differences between protocols and guidelines across hospitals could have influenced our results.

Figure 3. Online calculator example. Here, the probability of obstructive coronary artery disease is calculated for a 60-year-old male with typical chest pain. The calculator is based on the random effects model (Table 3) and is available for online use via http://rcc.simpal.com/tgenders_CAD consortium/. The recalibrated model for the low-risk populations is available via http://rcc. simpal.com/7TO293.



One could argue that our study population does not represent the target population of the Diamond–Forrester model because of the high prevalence of CAD. It might be more reasonable to apply the model in patients without a clinical indication for CAG and a lower probability of disease. However, despite the overall high prevalence of disease in our study population, we showed that the Diamond–Forrester model tends to overestimate the probability of CAD. The overestimation will be even larger, if the model would be applied to lower risk populations. As expected, validation of the updated model in the independent data consisting of outpatients presenting with chest pain demonstrated an overestimation of the probability of CAD, although to a lesser extent than the overestimation by Diamond and Forrester would be. Recalibration results did not justify updating of the predictor effects, suggesting that after adjusting the intercept, our updated model would be valid for the estimation of the pre-test probability of CAD in this low-risk population.

Furthermore, the Diamond–Forrester model predicts the probability of the presence vs. the absence of obstructive CAD. In validation and updating, we focused on this dichotomous outcome. This is a limitation of our study design, since detecting severe CAD (e.g. ≥70% stenosis, multi-vessel disease, left main disease) would have different clinical implications, as those patients would be eligible for revascularization.

Although our total sample size was large, some analyses involved stratification for several variables. This resulted in small numbers of patients in the lower age categories, especially for women because women represented only 30% of the total population.

Clinical implications

According to our analysis, the Diamond–Forrester model overestimates the probability of obstructive CAD in a contemporary cohort. Thus, the pre-test probability for today's patients can be estimated with more precision if the updated model is used.

The clinical value of a diagnostic test depends largely on the pre-test probability. A better estimate of the pre-test probability will therefore help clinicians make better decisions as to whether and which diagnostic test is indicated in a particular patient and to decide on further management based on the results of such tests. In patients with a (very) low pre-test probability of disease, a wait-and-see strategy without any testing is preferable. In patients with a high pre-test probability of disease, a direct invasive strategy is optimal. The main benefit of testing is in patients with an intermediate pre-test probability, in which a negative test rules out the presence of obstructive CAD, whereas a positive test justifies further testing.

Since the updated model predicts less high probabilities compared with the Diamond–Forrester model, using the updated model could lead to decreased referral to CAG, a higher yield of angiography, and increased use of non-invasive testing for risk stratification. This would be a welcome response to the issue brought to light by Patel *et al.* (25) who reported that the diagnostic yield of elective coronary angiography in the USA was only 41% (i.e. only 41% was found to have obstructive CAD) and concluded that better risk-stratification tools are needed.

Finally, the revised model is user-friendly and requires only three inputs from the physician. The model can be used via the website, or it could be implemented in electronic patient records or electronic order entry systems.

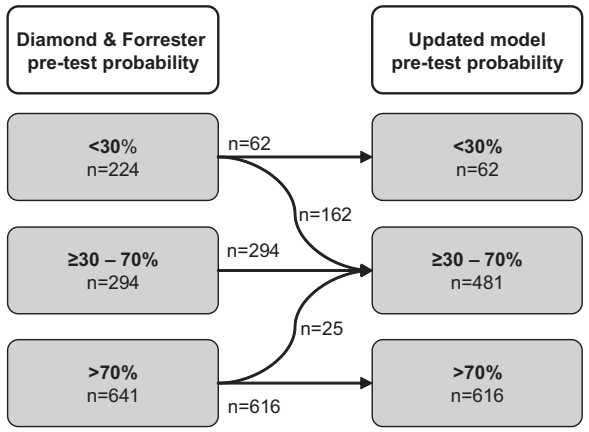


Figure 4. Reclassification flow chart for men between 30 and 70 years old. See Appendix, Table A8 for reclassification statistics.

Future research

Our study focused on the prediction of obstructive CAD according to age, sex, type of chest pain, and hospital only. Other risk factors such as smoking, diabetes, hypertension and dyslipidaemia have previously been demonstrated to be associated with the presence of CAD (26). Therefore, other known cardiovascular risk factors should be considered in future prediction models. On the other hand, the predictive effects of cardiovascular risk factors (in diagnostic models) are probably small in comparison with the predictive effects of age, sex, and type of chest pain, and care should be taken to minimize the number of variables, because simple prediction tools are more likely to be used by physicians in clinical practice.

Prospective data should be collected for the development and validation of prediction models, including a more heterogeneous study population. For example, including populations with a lower overall probability of obstructive CAD would improve the generalizability of the results.

Conclusion

All-in-all, we updated and extended the predictive effects of age, sex, and type of chest pain, based on a contemporary cohort and using modern statistical methods. We demonstrated that the Diamond–Forrester model can be improved for the current European situation. The updated model is available online.

Funding:

This work was supported by a Health Care Efficiency grant from the Erasmus University Medical Center, Rotterdam. The authors' work was independent of the funding organizations. The funding organizations had no involvement in the design or conduct of this study; data management and analysis; or manuscript preparation and review or authorization for submission.

Conflict of interest:

M.D.: Research Grants—European Regional Development Fund, German Heart Foundation/German Foundation of Heart Research, GE Healthcare (Amersham), Bracco, Guerbet, and Toshiba Medical Systems. Speakers Bureau—Toshiba Medical Systems, Guerbet, and Bayer-Schering. Other—Cardiac CT Workshops: www.ct-kurs.de. Book Authorship: Coronary CT Angiography, Springer, 2008. Institutional research collaboration—Siemens Medical Solutions, Philips Medical Systems, and Toshiba Medical Systems. U.J.S. is a consultant for and receives research support from Bayer-Schering, Bracco, General Electric, Medrad, and Siemens. J.J.B. has research grants from Biotronik (Berlin, Germany), BMS Medical Imaging (North Billerica, MA, USA), Boston Scientific (Natick, MA, USA), Edwards Lifesciences (Irvine, CA, USA), GE Healthcare (Buckinghamshire, UK), Medtronic (Minneapolis, MN, USA), and St Jude Medical (St Paul, MN, USA). F.R.G.: No conflicts of interests. F.R.G. is co-supported by the Dutch Technology Foundation STW (Utrecht, The Netherlands), applied science division of NWO, and the Technology Program of the Ministry of Economic Affairs, grant no. 10084. F.P.: This work forms part of the research themes contributing to the translational research portfolio of Barts and The London Cardiovascular Biomedical Research Unit, which is supported and funded by the National Institute for Health Research. All other authors do not have any potential conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript. The principal investigator, M.G.M.H., had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

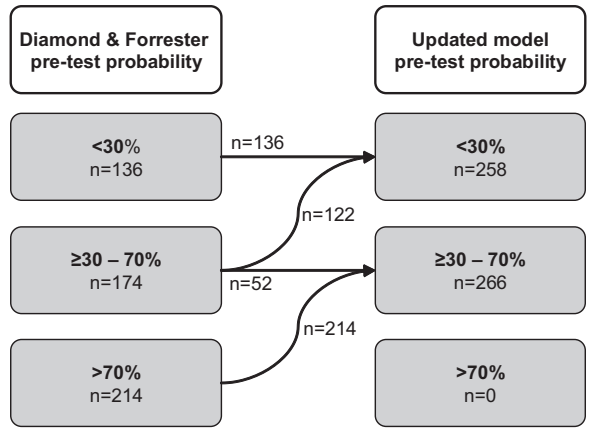


Figure 5. Reclassification flow chart for women between 30 and 70 years old. See Appendix, Table A8 for reclassification statistics.

Appendix

Validation and updating of the Diamond–Forrester model was performed by T.S.S.G., using state-of-the-art methods (17-19).

Validation

The Diamond–Forrester model takes into account age, sex, and type of chest pain and was designed for patients between 30 and 70 years only. For the validation of the Diamond–Forrester model, we therefore excluded both patients below the age of 30 and patients above the age of 69.

'Original' Diamond–Forrester coefficients

The Diamond–Forrester model8 is based on a Bayesian analysis. Therefore, it does not provide regression coefficients for the effects of age, sex, and type of chest pain on the probability of CAD. To unravel the implicit coefficients of the predictors in this model, we performed a weighted linear regression on the log odds of the Diamond–Forrester predictions per subgroup. The weights were proportional to the inverse of the standard errors reported by Diamond and Forrester. The coefficients obtained by the weighted linear regression are further referred to as the 'original coefficients' of Diamond and Forrester.

The observed frequency of CAD in our data set was calculated stratifying for 10-year age category, sex, and type of chest pain (i.e. equivalent to the Diamond–Forrester categories). Observed frequencies of CAD were compared with the predicted probabilities of CAD according to Diamond and Forrester and were subsequently tabulated (Table 2) and presented by means of calibration plots (Figures 1 and 2).

Calibration-in-the-large

In the consortium data, patients aged 30–69 (Diamond–Forrester age range) were assigned the log odds of obstructive CAD as estimated by the weighted linear regression model, i.e. the linear predictor. A model (Equation 1 in Figure A1) was fitted to calculate the log odds of obstructive CAD as observed in our data set. The linear predictor of Diamond and Forrester ($lp_{D\&F}$) was included in the model and its coefficient was fixed at unity. In this way, the original absolute and relative effects of the original Diamond–Forrester model are maintained. The intercept (new) is the only free parameter in the model, which allows us to quantify the 'calibration-in-the-large' and adjust for difference in disease prevalence.

The intercept ($_{new}$) can be interpreted as the difference in log odds between the mean observed outcome and the mean predicted probability of CAD according to Diamond and Forrester (19). In other words, we assessed calibration-in-the-large by comparing the mean observed frequency of CAD with the mean of the predicted probabilities according to Diamond and Forrester in a logistic regression model. We tested whether the difference (i.e. the intercept) was significantly different from zero.

Table A1. Validation of the Diamond–Forrester model

Model	Diamond and Forrester (weighted linear regression *) 'original' regression coefficients			δ coefficients † (Equation 3 in Figure A1) (in patients aged 30-69, n=1683)		
		Coefficient	*P*-value		Coefficient	*P*-value
Baseline analysis	α	-7.52	–	δ α	2.97	<0.001
	Age	0.09	–	δ Age	-0.04	<0.001
	Sex	1.35	–	δ Male sex	0.05	0.67
	Typical CP	3.77	–	δ Typical CP	-1.78	<0.001
	Atypical CP	1.70	–	δ Atypical CP	-1.20	<0.001
C-statistic (95% CI ‡)	0.78 (0.76-0.79)			0.78 (0.76-0.81)		
HL *P*-value	<0.001			0.36		
Analyses adjusted for hospital§				δ α	2.05	<0.001
				δ Age	-0.04	<0.001
				δ Male sex	-0.02	0.88
				δ Typical CP	-1.21	<0.001
				δ Atypical CP	-0.96	<0.001
C-statistic (95% CI ‡)				0.81 (0.79-0.83)		
HL *P*-value				0.06		

CP = chest pain, HL = Hosmer–Lemeshow goodness-of-fit test (significant result indicates lack of model fit, poor calibration), α = intercept.
** Regression coefficients were approximated by weighted linear regression (see Appendix, Original Diamond-Forrester coefficients).*
† Difference between Diamond–Forrester coefficients and refitted coefficients.
‡ Estimated by bootstrapping with 1000 repetitions.
§ Reference category is Erasmus University Medical Center.

Recalibration

Next, we recalibrated the model by comparing the average regression slope of Diamond and Forrester with the average regression slope in the consortium data. A second linear predictor variable was added to the model (while maintaining both the previous linear predictor as offset variable and the new intercept), and its coefficient $\beta_{miscalibration}$ (Equation 2 in Figure A1) was estimated. This coefficient reflects the miscalibration of the Diamond–Forrester predictor effects when compared with the predictor effects in the consortium data (19). We tested whether $\beta_{miscalibration} = 0$, corresponding to the hypothesis that the Diamond–Forrester prediction (adjusted for calibration-in-the-large) fits the data well. If significant, we conclude that the overall effects of age, sex, and type of chest pain together are different in our data and that model revision is justified.

1. $Logit(pCAD)_{data} = \alpha_{new} + \text{offset}(lp_{D\&F})$
2. $Logit(pCAD)_{data} = \alpha_{new} + \beta_{miscalibration} \cdot lp_{D\&F} + \text{offset}(lp_{D\&F})$
3. $Logit(pCAD)_{data} = \alpha_{new} + \delta_{age} \cdot X_{age} + \delta_{sex} \cdot X_{sex} + \delta_{typical} \cdot X_{typical} + \delta_{atypical} \cdot X_{atypical} + \text{offset}(lp_{D\&F})$
4. $Logit(pCAD)_{data} = \alpha_{new} + \beta_{age} \cdot X_{age} + \beta_{sex} \cdot X_{sex} + \beta_{typical} \cdot X_{typical} + \beta_{atypical} \cdot X_{atypical}$
5. $Logit(pCAD)_{data} = \alpha^{*}_{new} + \beta_{age} \cdot X_{age} + \beta_{sex} \cdot X_{sex} + \beta_{typical} \cdot X_{typical} + \beta_{atypical} \cdot X_{atypical}$
6. $Logit(pCAD)_{data} = \alpha^{*}_{new} + \beta_{age} \cdot X_{age} + \beta_{sex} \cdot X_{sex} + \beta_{typical}^{*} \cdot X_{typical} + \beta_{atypical}^{*} \cdot X_{atypical}$

Figure A1. Equations. Logit = natural log odds of the probability, pCAD = probability of obstructive coronary artery disease, = intercept of logistic regression model, offset = regression coefficient fixed at unity, = regression coefficient, ∂ = difference between β_{new} and $\beta_{original}$ (see Table A1 for 'original' coefficients), $lp_{D\&F}$ = linear predictor of Diamond and Forrester, * = random effect.

Re-estimation

Finally, we re-estimated the predictor effects in a model including the linear predictor as offset and the new intercept (Equation 3 in Figure A1). The coefficients from this analysis refer to the difference between the re-estimated and the original coefficients (i.e. ∂-coefficients). We tested whether these differences were significantly different from zero. From these analyses, we can judge which predictor effects are different in our data compared with the model according to Diamond and Forrester.

All analyses were performed both unadjusted and adjusted for hospital.

Updating and extension

To update the Diamond–Forrester model, we performed a logistic regression analysis (Equation 4 in Figure A1), using all data, including patients below the age of 30 and above the age of 69. All variables (age, sex, and type of chest pain) were entered simultaneously in the model ('baseline analysis') (Table A3). To judge whether interaction terms should be considered, we performed an overall test for interaction using second-order interactions of the main effects (i.e. age, sex, and type of chest pain). Significant interaction terms were tested one-by-one and omitted if non-significant. Similarly, interactions between main effects and hospital were tested.

Separately, we extended the model with a random intercept to allow for heterogeneity in CAD prevalence across different hospitals (Equation 5 in Figure A1). Because of potential

differences in symptom classification across hospitals, we also tested random effects for type of chest pain across different hospitals (Equation 6 in Figure A1). The linearity assumption for the continuous variable age was checked graphically and tested statistically by including a restricted cubic spline function with three knots (2 d.f.).

Diagnostic performance was quantified by calculating the area under the receiver operating curve (c-statistic), and confidence intervals were obtained by bootstrapping with 1000 repetitions. Calibration was estimated by the Hosmer–Lemeshow goodness-of-fit test and by constructing a calibration plot (Figure A2).

A P-value < 0.05 was considered statistically significant. Statistical analyses were performed using Stata/SE 10.1, StataCorp, TX, USA.

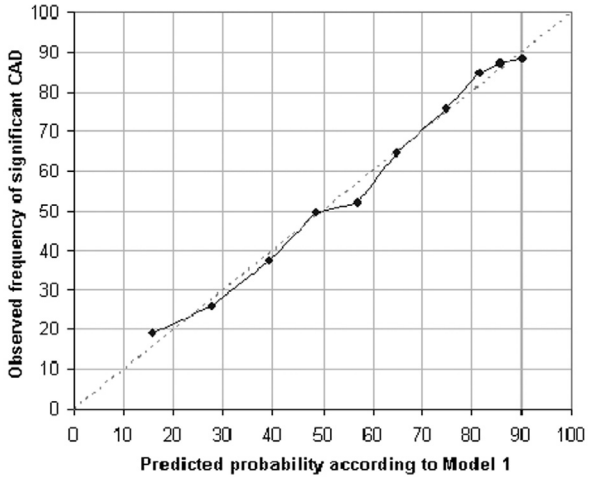


Figure A2. Calibration plot of the updated model (See Table A3, Model 1)

Table A2. Logistic regression predicting the probability of verification with conventional coronary angiography

Variable	Odds ratio	95% CI lower limit	95% CI upper limit	*P*-value
Age	1.04	1.02	1.07	<0.01
Male sex	2.18	1.34	3.55	<0.01
Typical chest pain	3.09	1.47	6.50	<0.01
Atypical chest pain	1.23	0.58	2.61	0.58
Non-specific chest pain	reference			

Table A3. Updating and extension of the Diamond– Forrester model

	Coefficient	Odds ratio	95% CI lower limit	95% CI upper limit
Model 1: Baseline analysis				
Age	0.04	1.04	1.03	1.05
Sex	1.34	3.82	3.11	4.70
Typical chest pain	1.99	7.33	5.69	9.43
Atypical chest pain	0.50	1.65	1.26	2.16
Non-specific chest pain *	–	–		
Intercept	-4.32	–		
c-Statistic (95% CI †)	0.79		0.77	0.81
HL *P*-value	0.52			
Model 2: Corrected for verification bias				
Age	0.04	1.04	1.02	1.06
Sex	1.16	3.20	2.47	4.15
Typical chest pain	1.99	7.31	5.62	9.52
Atypical chest pain	0.63	1.87	1.40	2.51
Non-specific chest pain *	–	–		
Intercept	-4.09	–		
c-Statistic (95% CI †)	0.78		0.76	0.80
Model 3: Random intercept + random slope 'typical chest pain' ‡				
Age	0.04	1.04	1.03	1.05
Sex	1.34	3.82	3.08	4.74
Typical chest pain	1.91	6.72	3.97	11.37
Atypical chest pain	0.64	1.89	1.38	2.59
Non-specific chest pain *	–	–		
Intercept	-4.37	–		
c-Statistic (95% CI †)	0.82		0.80	0.84

** Reference category.*
† Estimated by bootstrapping with 1000 repetitions.
‡ Group variable is 'hospital'.

Correction for verification bias

To correct for verification bias in our study population, we performed inverse probability weighting on each individual in our data set, using the probability of verification. For this purpose, we used registry data (12,13) of 471 outpatients who presented to the cardiologist for the evaluation of chest pain. All patients in this registry were intended to undergo both stress-ECG and coronary CT angiography. Of the 471 patients, 98 were referred for CAG based on clinical parameters and the non-invasive test results.

A logistic regression analysis was performed, predicting the probability of undergoing CAG based on age, sex and type of chest pain (Table A2). Subsequently, the probability of verification was calculated for each individual in the CAD Consortium. The inverse of the probability of verification was used to weigh each individual, which corrects for verification bias.

Reclassification

To assess the clinical utility of the updated Diamond–Forrester model, we assessed reclassification of patients between the ages of 30 and 69 for men (Table A4, A5) and women (Table A6, A7) separately, comparing the original Diamond– Forrester probability predic-

tion (i.e. according to age category, sex, and type of chest pain) with the updated model in the online calculator (i.e. according to age [continuous], sex and type of chest pain; Table A3, Model 3).

Validation of the updated model

The CAD Consortium population is highly selected based on referral to CAG. To assess the performance of the updated model in a lower risk population, we attempted to validate the model in an independent registry data set (not included in our consortium) consisting of unselected outpatients presenting to the cardiologist who were intended to undergo both stress-ECG and CCTA for the evaluation of chest pain (12,13). (This data set was also used to perform the correction for verification bias.)

Table A4. Reclassification table using probability categories <30, ≥30–70, and ≥70% (men)

Probability category based on Diamond and Forrester	Probability category based on updated model			
	0 – 30%	30 – 50%	70 – 100%	Total
0 – 30%				
n (%)	62 (27.7)	162 (72.3)	–	224 (19.3)
Observed probability, %	17.7	42.0	–	35.3
30 – 70%				
n (%)	–	294 (100.0)	–	294 (25.4)
Observed probability, %	–	48.3	–	48.3
70 – 100%				
n (%)	–	25 (3.9)	616 (96.1)	641 (55.3)
Observed probability, %	–	64.0	84.3	83.5
Total				
n (%)	62 (5.4)	481 (41.5)	616 (53.2)	1159 (100)
Observed probability, %	17.7	47.0	84.3	65.2

Table A5. Reclassification table using probability categories <40 and ≥40% (men)

Probability category based on Diamond and Forrester	Probability category based on updated model		
	0 – 40%	40 – 100%	Total
0 – 40%			
n (%)	150 (67.0)	74 (33.0)	224 (19.3)
Observed probability, %	26.0	54.1	35.5
40 – 100%			
n (%)	22 (2.4)	913 (97.6)	935 (80.7)
Observed probability, %	27.3	73.5	72.4
Total			
n (%)	172 (14.8)	987 (85.2)	1159 (100)
Observed probability, %	26.2	72.0	65.2

For this purpose, we followed the step-wise approach as described above (Table A9). Since not all patients in this population underwent CAG, the CCTA results were used as proxy for the reference standard in patients who did not undergo CAG. Out of 471 patients, 17 did not undergo either CCTA or CAG and were therefore excluded from the analysis.

Table A6. Reclassification table using probability categories <30, ≥30–70, and ≥70% (women)

Probability category based on Diamond and Forrester	Probability category based on updated model			
	0 – 30%	30 – 50%	70 – 100%	Total
0 – 30%				
n (%)	136 (100.0)	–	–	136 (25.9)
Observed probability, %	17.6	–	–	17.6
30 – 70%				
n (%)	122 (70.1)	52 (29.9)	–	174 (33.2)
Observed probability, %	27.0	36.5	–	29.9
70 – 100%				
n (%)	–	214 (100)	–	214 (40.8)
Observed probability, %	–	49.1	–	49.1
Total				
n (%)	258 (49.2)	266 (50.8)	–	524 (100)
Observed probability, %	22.1	46.6	–	34.5

Table A7. Reclassification table using probability categories <40 and ≥40% (women)

Probability category based on Diamond and Forrester	Probability category based on updated model		
	0 – 40%	40 – 100%	Total
0 – 40%			
n (%)	198 (100.0)	–	198 (37.8)
Observed probability, %	19.7	–	19.7
40 – 100%			
n (%)	105 (32.2)	221 (67.8)	326 (62.2)
Observed probability, %	32.4	48.9	43.6
Total			
n (%)	303 (57.8)	221 (42.2)	524 (100.0)
Observed probability, %	24.1	48.9	34.5

Table A8. Reclassification statistics

Cutoffs		Overall Reclassification Percentage	% Correct *	X^2 model 1 †	*P*-value	X^2 model 2 ‡	*P*-value	NRI §, %	*P*-value for NRI
<30%, ≥30-70%, ≥70%	Men	16.1	100.0	125.5	0	10.9	0.01	-14	<0.001
	Women	64.1	100.0	261.7	0	4.25	0.12	-19	0.02
<40%, ≥40%	Men	8.3	100.0	99.1	0	6.98	0.03	0	0.99
	Women	20.0	100.0	244	0	4.1	0.04	2	0.64

NRI =net reclassification improvement.
** If the predicted probability of obstructive CAD of the updated model was closer to the observed probability of CAD compared with the prediction of the original model, the reclassification was considered to be correct.*
† Reclassification calibration (Hosmer–Lemeshow) statistic for the original model, using cells from the reclassification table with at least 20 observations. The significant P-value indicates poor calibration of the original model.
‡ Reclassification calibration statistic for the updated model, using cells from the reclassification table with at least 20 observations. The higher P-value indicates better fit of the updated model compared with the original model.
§ The net reclassification improvement is defined as the difference in proportions of patients moving up and down for cases and non-cases separately and it is interpreted as the percentage reclassified, adjusted for the reclassification direction. Here, the net reclassification index is negative because, on average, most individuals are being down classified by the updated model (i.e. less overestimation by the updated model compared with the Diamond–Forrester model), irrespective or their disease status

Table A9. Validation of the updated model

	Calibration-in-the-large		Recalibration	
	Coefficient	*P*-value	Coefficient	*P*-value
α	-0.92	< 0.001	-0.92	< 0.001
$lp_{consortium}$	1	–	-0.03	0.79
$lp_{consortium}$			1	–
c-Statistic (95% CI *)	0.76 (0.71-0.81)		0.76 (0.71-0.81)	

α = intercept, $lp_{consortium}$ = linear predictor of the updated model (random effects model, see Table 3).
** Estimated by bootstrapping with 1000 repetitions.*

References

1. Noto TJ Jr, Johnson LW, Krone R, Weaver WF, Clark DA, Kramer JR Jr, Vetrovec GW. Cardiac catheterization 1990: a report of the Registry of the Society for Cardiac Angiography and Interventions (SCA&I). Cathet Cardiovasc Diagn 1991;24:75–83.
2. Gibbons RJ, Balady GJ, Bricker JT, Chaitman BR, Fletcher GF, Froelicher VF, Mark DB, McCallister BD, Mooss AN, O'Reilly MG, Winters WL Jr, Gibbons RJ, Antman EM, Alpert JS, Faxon DP, Fuster V, Gregoratos G, Hiratzka LF, Jacobs AK, Russell RO, Smith SC Jr. ACC/AHA 2002 guideline update for exercise testing: summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines). Circulation 2002; 106:1883–1892.
3. Hendel RC, Berman DS, Di Carli MF, Heidenreich PA, Henkin RE, Pellikka PA, Pohost GM, Williams KA, Wolk MJ, Alagona P Jr, Bateman TM, Cerqueira MD, Corbett JR, Dean AJ, Dehmer GJ, Goldbach P, Gordon L, Kushner FG, Kwong RY, Min J, Quinones MA, Ward RP, Yang SH, Allen J, Brindis RG, Douglas PS, Patel M, Peterson E. ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/ SNM 2009 appropriate use criteria for cardiac radionuclide imaging: a report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the American Society of Nuclear Cardiology, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the Society of Cardiovascular Computed Tomography, the Society for Cardiovascular Magnetic Resonance, and the Society of Nuclear Medicine: endorsed by the American College of Emergency Physicians. Circulation 2009; 119:e561–e587.
4. Hendel RC, Patel MR, Kramer CM, Poon M, Hendel RC, Carr JC, Gerstad NA, Gillam LD, Hodgson JM, Kim RJ, Kramer CM, Lesser JR, Martin ET, Messer JV, Redberg RF, Rubin GD, Rumsfeld JS, Taylor AJ, Weigold WG, Woodard PK, Brindis RG, Hendel RC, Douglas PS, Peterson ED, Wolk MJ, Allen JM, Patel MR. ACCF/ACR/SCCT/SCMR/ASNC/NASCI/SCAI/SIR 2006 appropriateness criteria for cardiac computed tomography and cardiac magnetic resonance imaging: a report of the American College of Cardiology Foundation Quality Strategic Directions Committee Appropriateness Criteria Working Group, American College of Radiology, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance, American Society of Nuclear Cardiology, North American Society for Cardiac Imaging, Society for Cardiovascular Angiography and Interventions, and Society of Interventional Radiology. J Am Coll Cardiol 2006;48:1475–1497.
5. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med 1980;302:1109–1117.
6. Hunink MGM, Glasziou PP, Siegel JE, Weeks J, Pliskin J, Elstein A, Weinstein M. Decision Making in Health and Medicine: Integrating Evidence and Values. Cambridge: Cambridge University Press; 2001.
7. Genders TS, Meijboom WB, Meijs MF, Schuijf JD, Mollet NR, Weustink AC, Pugliese F, Bax JJ, Cramer MJ, Krestin GP, de Feyter PJ, Hunink MG. CT coronary angiography in patients suspected of having coronary artery disease: decision making from various perspectives in the face of uncertainty. Radiology 2009; 253:734–744.
8. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med 1979;300:1350–1358.
9. Pryor DB, Harrell FE Jr, Lee KL, Califf RM, Rosati RA. Estimating the likelihood of significant coronary artery disease. Am J Med 1983;75:771–780.
10. Pryor DB, Shaw L, McCants CB, Lee KL, Mark DB, Harrell FE Jr, Muhlbaier LH, Califf RM. Value of the history and physical in identifying patients at increased risk for coronary artery disease. Ann Intern Med 1993;118:81–90.
11. EIBIR. European Institute for Biomedical Imaging Research. http://www.eibir.org/ cms/website.php.

12. Nieman K, Galema T, Weustink A, Neefjes L, Moelker A, Musters P, de Visser R, Mollet N, Boersma H, de Feijter PJ. Computed tomography versus exercise electrocardiography in patients with stable chest complaints: real-world experiences from a fast-track chest pain clinic. Heart 2009;95:1669–1675.
13. Nieman K, Galema TW, Neefjes LA, Weustink AC, Musters P, Moelker AD, Mollet NR, de Visser R, Boersma E, de Feijter PJ. Comparison of the value of coronary calcium detection to computed tomographic angiography and exercise testing in patients with chest pain. Am J Cardiol 2009;104: 1499–1504.
14. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983;39:207–215.
15. Hunink MG, Polak JF, Barlan MM, O'Leary DH. Detection and quantification of carotid artery stenosis: efficacy of various Doppler velocity parameters. AJR Am J Roentgenol 1993;160:619–625.
16. Diamond GA. A clinically relevant classification of chest discomfort. J Am Coll Cardiol 1983;1(2 Pt 1):574–575.
17. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer; 2001.
18. 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;23:2567–2586.
19. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation and Updating. New York: Springer; 2008.
20. Meijboom WB, van Mieghem CA, Mollet NR, Pugliese F, Weustink AC, van Pelt N, Cademartiri F, Nieman K, Boersma E, de Jaegere P, Krestin GP, de Feyter PJ. 64-slice computed tomography coronary angiography in patients with high, intermediate, or low pretest probability of significant coronary artery disease. J Am Coll Cardiol 2007;50:1469–1475.
21. 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;27:157–172; Discussion 207–212.
22. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med 2009;150: 795–802.
23. Morise AP, Haddad WJ, Beckner D. Development and validation of a clinical score to estimate the probability of coronary artery disease in men and women presenting with suspected coronary disease. Am J Med 1997;102:350–356.
24. Morise AP, Bobbio M, Detrano R, Duval RD. Incremental evaluation of exercise capacity as an independent predictor of coronary artery disease presence and extent. Am Heart J 1994;127:32–38.
25. Patel MR, Peterson ED, Dai D, Brennan JM, Redberg RF, Anderson HV, Brindis RG, Douglas PS. Low diagnostic yield of elective coronary angiography. N Engl J Med 2010;362:886–895.
26. Chun AA, McGee SR. Bedside diagnosis of coronary artery disease: a systematic review. Am J Med 2004;117:334–343.

Chapter 7.2 (supplement)

Editorial: Diagnosing Coronary Artery Disease—the Diamond and Forrester Model Revisited

Eur Heart J 2011; 32:1311-1312.

Ian M. Graham

'Stop annoying me with this probability stuff. Does my patient have coronary artery disease or not?'

What does a doctor do? Well, we teach our students to take a meticulous history, perform a, hopefully, skilled examination, make a provisional diagnosis, arrange appropriate tests, establish a diagnosis, and initiate a management plan. David Sackett, however, taught us that, in practice, we shortcut this process by making an immediate diagnosis based on probabilities and then adjust this probability as more information comes in (1). Thus, intuitively, we adopt a Bayesian (2) approach to diagnosis—our first diagnosis is the 'prior' or 'pre-test' one, followed by a 'post-test' probability that is refined by additional information. Even if you hate probabilities, such an approach gives a figure to aid logical decision making.

What does a cardiologist do? Most clinical cardiologists who have not been seduced into the comforting womb of the cardiac catheterization laboratory spend their days applying the above principles to people with chest pain, breathlessness, palpitation, dizzy spells, or syncope.

Thirty years ago, Diamond and Forrester (3) applied Bayesian principles in a classical paper that is a model of clarity. Using data from 4952 subjects undergoing coronary arteriography, they treated chest pain as a diagnostic test. They tabulated:

I the prevalence of coronary artery disease (CAD) in subjects with non-anginal chest pain, atypical angina, and typical angina, then

II the pre-test likelihood of CAD by age, sex, and type of chest pain, and

III the post-exercise test likelihood of CAD based on the degree of ST-segment depression on exercise testing—with probabilities varying from 0.1% (asymptomatic female aged 30–39 with no or minimal ST changes) to 99.8% (typical angina in a man aged 60–69 with .2.5 mm ST depression).

They also noted that additional tests such as stress scintigraphy or coronary artery calcification can refine probability estimates until the point of diminishing returns is reached, paving the way for evidence-based management algorithms.

Implicit in all of this research is the observation that the maximum yield from an additional test occurs when the pre-test probability is ~50% – just where you need help. If the pre-test probability is very high, a normal test result is likely to be wrong (false negative) and if the pre-test probability is very low, an abnormal test result is likely to be wrong (false positive).

Diamond and Forrester were well aware of the limitations of their heterogeneous database '. . . the data presented in the tables should not be considered as absolute standards but, rather, as preliminary estimates that will require some modification as more precise data become available' (3).

So this, after a modest delay of 30 years, is where the paper of Genders and colleagues (4) comes in. Authorship has grown a little to 35, representing epidemiology, public health, clinical cardiology, and imaging. While the principles of Bayesian theory remain solid, considerable methodological advances have occurred. The data were corrected for selection or

verification bias, additional measures of diagnostic performance were used, and the percentage of patients whose probabilities of CAD were derived from Diamond and Forrester's estimates that would be correctly re-classified using the current data set was calculated.

Two primary questions were addressed. First, can the likelihood of obstructive coronary artery disease (OCD) be quantified based on age, sex, and type of chest pain (typical, atypical, and non-anginal), and, secondly, can the Diamond and Forrester estimates be updated using a more modern data set? The answer to both appears to be yes.

Probability estimates were based on 2260 subjects undergoing coronary arteriography in 14 hospitals in six countries. Of these, 58.45% had obstructive coronary disease, with marked regional variations, from 39.4% (the USA) to 75.5% (Russia). In contrast to Diamond and Forrester's study, which presented data on subjects up to age 69, subjects aged up to and exceeding 80 years were included.

For the clinician, the relevant results are contained in Table 2 of the paper by Genders *et al.*, which shows the probability of OCD in subjects with typical, atypical, and non-anginal chest pain. Figure 1 summarizes the results from their present study. Diamond and Forrester's estimates overpredicted the observed prevalences of OCD, particularly in women with typical and atypical pain, although not in those with non-anginal pain. The range of predicted prevalences was narrower than in Diamond and Forrester, and some of the observed disease prevalences are unexpected, being apparently U-shaped in men with typical angina and not increasing above age 60. Overall, the updated model performed reasonably well, with a c-statistic of 0.79 in men and 0.82 in women. More women than men were re-classified and the overall net re-classification index was negative because of the higher probability estimates given by the Diamond and Forrester model. The limitations of the study are well discussed, including selection effects, variability in disease prevalences in different institutions, and the lack of standardization of methods of coronary arteriography and reporting. In addition, it is not known whether American physicians of 30 years ago interpreted the categories of chest pain used in the same way as the physicians of today.

In some ways, Diamond and Forrester went further than the study of Genders *et al.* One of the beauties of the earlier paper was the presentation of pre- and post-test probabilities of OCD before and after stress testing, stratified by the degree of ST depression. They also gave some indication of the effect of additional testing in further refining probability estimates and when further tests will add little or nothing.

The paper by Genders *et al.* is timely and relevant to everyday clinical practice, updating our ability to treat chest pain as a diagnostic test and documenting the likelihood of OCD based on age, sex, and the type of chest pain. This powerful group could, one hopes, extend this work to study larger and more heterogeneous groups of patients prospectively and with standardized methods. The addition of stress test data would widen the range of probabilities and facilitate the decision to proceed, or not, to coronary arteriography. Ideally the effects of additional investigations on serial post-test probabilities could also be refined as a further compliment and complement to the work of Diamond and Forrester.

Conflict of interest: none declared.

References

1. Sackett DL, Haynes RB, Tugwell P, Guyatt GH. Clinical Epidemiology: A Basic Science for Clinical medicine, 2nd edn. Boston: Little Brown, 1991.
2. Bayes T. Essay towards solving a problem in the doctrine of chances. Philos Trans R Soc 1764 (submitted posthumously by Richard Price).
3. Diamond GA, Forrester JS. Analysis of probability in the clinical diagnosis of coronary-artery disease. N Engl J Med 1979;300:1350–1358.
4. Genders TSS, Steyerberg EW, Alkadhi H, Leschka S, Desbiolles L, Nieman K, Galema TW, Meijboom WB, Mollet NR, de Feyter PJ, Cademartiri F, Maffei E, Dewey M, Zimmermann E, Laule M, Pugliese F, Barbagallo R, Sinitsyn V, Bogaert J, Goetschalckx K, Schoepf UJ, Rowe GW, Schuijf JD, Bax JJ, de Graaf FR, Knuuti J, Sami K, van Mieghem CAG, Meijs MFL, Cramer MJ, Gopalan D, Feuchtner G, Friedrich Krestin GP, Hunink MGM. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating and extension. Eur Heart J; doi:10.1093/eurheartj/ehr014. Published online ahead of print 2 March 2011.

Chapter 8

A Prediction Model for the Presence of Coronary Artery Disease: a pooled analysis of existing cohorts

British Medical Journal (In press)

Tessa S.S. Genders
Ewout W. Steyerberg
M.G. Myriam Hunink

On behalf of the CAD consortium collaborators (see page 339)

Abstract

Objective

To develop prediction models for the presence of coronary artery disease (CAD) and to improve the estimate of the pre-test probability of CAD.

Design, Setting and Participants

We analyzed individual patient data from 18 hospitals. Patients with stable chest pain without evidence for prior CAD were eligible if they were referred for coronary CT angiography (CCTA) (low-prevalence setting) or catheter-based coronary angiography (CCA) (high-prevalence setting).

Main outcome measures

Outcome was obstructive CAD defined as ≥50% diameter stenosis in ≥1 vessel on CCA. Multiple imputation was performed to account for missing predictors and outcomes, exploiting the strong correlation between CCTA and CCA. A basic model included age, sex, symptoms, and setting. A clinical model also included diabetes, hypertension, dyslipidaemia, and smoking, while an extended model added the coronary calcium score (CCS). Cross-validation (discrimination [c-statistic], calibration, and continuous net reclassification improvement [NRI]) was assessed for the four largest low-prevalence datasets separately and the smaller low-prevalence datasets combined.

Results

We included 5677 patients (3283 men, 2394 women), of whom 1654 had obstructive CAD on CCA. All potential predictors were significantly associated with the presence of CAD in univariable and multivariable analyses. The clinical model improved the prediction compared to the basic model (cross-validated c-statistic from 0.77 to 0.79, cross-validated NRI = 35%). The CCS was a major predictor, increasing the cross-validated c-statistic to 0.88 (cross-validated NRI = 102%). Calibration for low-prevalence datasets was satisfactory.

Conclusions

Updated prediction models including age, sex, symptoms, and cardiovascular risk factors allow for accurate estimation of the likelihood of CAD in low-prevalence populations. Adding CCS improves the prediction.

Introduction

In the United States approximately 10.2 million individuals suffer from chest pain complaints (1) and over 1.1 million inpatient diagnostic catheter-based coronary angiographies (CCA) are performed each year (2). A recent report based on the American College of Cardiology (ACC) National Cardiovascular Data Registry (3) demonstrated that only 41% of patients undergoing elective CCA are diagnosed with obstructive coronary artery disease (CAD). The authors concluded that better risk-stratification is needed, which is underlined by decision analyses demonstrating that the choice of further diagnostic work-up in patients with chest pain depends primarily on the pre-test probability of CAD (4-6).

Current guidelines from the American College of Cardiology/American Heart Association (ACC/AHA) (7-8) European Society of Cardiology (ESC) (9), and the United Kingdom (UK)(10) recommend the Diamond & Forrester model (11) or the Duke Clinical Score (DCS) (12-13) to estimate the pre-test probability of CAD in patients with chest pain. The Diamond & Forrester model tends to overestimate the probability of CAD (defined as ≥50% stenosis) and a revised version has recently been published (14). The Duke Clinical Score (12-13) estimates the probability of CAD (≥75% stenosis) which, to our knowledge, has not been validated in populations outside the United States. Whereas the ACC/AHA and ESC recommend exercise electrocardiography to select patients for further diagnostic work-up, the UK guidelines recommend the CT coronary calcium score (CCS) in patients with a low to intermediate pre-test probability (10-29%).

We perceived a need for an updated and stepwise approach to estimating the probability of CAD in patients who present with new-onset chest pain in a low-prevalence population as clinical information and test results become available, in particular because implementation of the guidelines requires the calculation of the pre-test probability. Purpose of this study was therefore to estimate the probability of obstructive CAD based on clinical presentation and cardiovascular risk factors and to determine the incremental diagnostic value of exercise electrocardiography and the coronary calcium score.

Methods

Design overview

A consortium of researchers from Europe and the United States of America was formed. An existing database with data on ≥80 eligible patients was required for participation. Participation did not involve any financial incentives. All patients had to be enrolled in single-centre studies, and local approval from the Institutional Review Board for the original research objectives was required.

The CAD consortium is part of the European network for the Assessment of Imaging in Medicine (EuroAIM) which is an initiative of the European Institute of Biomedical Imaging Research (EIBIR) (15). One of the goals of EuroAIM is to perform pooled analyses of existing prospectively collected data, improving power and increasing generalizability of results.

Participants

Patients were eligible for the analysis if they presented with stable chest pain and if they were referred for CCA or coronary CT angiography (CCTA) (≥64-slice). Patients were not eligible if they met one of the following criteria: i) acute coronary syndrome or unstable chest pain, ii) history of myocardial infarction (MI) or previous revascularisation (percutaneous coronary intervention or coronary artery bypass graft surgery), iii) no informed consent. For diagnosing CAD, CCA is considered to be the reference standard, which is an expensive and invasive procedure with a risk of complications. Non-invasive testing is generally recommended to select patients who may benefit from CCA. CCTA is a less invasive and less expensive test for CAD with a high sensitivity and specificity as compared to CCA (16-18). A negative CCTA virtually excludes the presence of obstructive CAD, whereas a positive CCTA may need confirmation by CCA.

Fourteen datasets consisted of consecutive patients who were enrolled in a prospective study for other research objectives. Four datasets consisted of patients who were retrospectively identified as eligible through electronic radiology reporting systems. In- and exclusion criteria were evaluated by experienced physicians and missing information was obtained via patient records. See Appendix Table 1 for more details.

Definitions

Data on age, sex, symptoms, cardiovascular risk factors, test results, and the presence of CAD were collected. Symptoms were classified as typical, atypical, or non-specific chest pain. Typical chest pain was defined as i) substernal chest pain or discomfort, that is ii) provoked by exertion or emotional stress and iii) relieved by rest and/or nitro-glycerine. Atypical chest pain was defined as two of the before mentioned criteria. If one or none of the criteria was present, symptoms were classified as non-specific (19).

Definitions for hypertension, diabetes, dyslipidaemia, and smoking were slightly different across hospitals (Appendix Table 1). The most common definitions are listed below Table 1. CCS was determined using the Agatston method (20) and log-transformed (natural logarithm of CCS+1) to account for its skewed distribution.

Outcomes

Primary outcome was obstructive CAD defined as ≥1 vessel with ≥50% diameter stenosis on CCA. As existing databases from different hospitals were combined, CCTA and CCA were performed at each institution according to local protocols; both visual assessment and quantitative assessment were allowed for interpretation of CCTA and CCA.

Statistical Analysis

We assumed missing data occurred at random conditional on the clinical variables and the CCTA result and performed multiple imputations using chained equations (MICE) (21). Missing values were predicted based on all other predictors considered, the CCTA result, as well as the outcome (22-23). Twenty datasets were created with identical known information, but differences in imputed values reflecting the uncertainty associated with imputations. In total, 667 (2%) clinical data items were imputed.

Only a minority of patients (n=2062, 36%) underwent CCA. An analysis restricted to patients who underwent CCA could be influenced by verification bias (24). We therefore imputed CCA using CCTA as an auxiliary variable in addition to all other predictors (25). CCTA results correlate well with CCA results especially for negative CCTA results (16-18), and therefore CCTA is a good proxy for CCA. This strong correlation was confirmed in the patients (n = 1609) who underwent both CCTA and CCA (Pearson r = 0.72). Since CCTA was used for imputing CCA, CCTA was **not** included as a predictor in the prediction models. Our approach was similar to using the CCTA result as outcome variable when CCA was not performed (which was explored in a sensitivity analysis) but is more sophisticated in that it takes into account other predictors and the uncertainty around the imputed values. In total, 3615 CCA outcome values (64%) were imputed. Multiple imputations were performed using Stata/SE 11 (StataCorp, Texas, USA).

External Validation of the Duke Clinical Score (DCS)

To evaluate the performance of the DCS, we calculated the predicted probability based on published coefficients (13) for predicting ≥75% stenosis. Since patients with evidence of prior CAD were excluded we assumed all had a normal resting electrocardiogram. Predictions were compared with the observed proportion of severe CAD (≥70% stenosis or ≥50% left main stenosis) in a calibration plot.

Development of new prediction models

Three models were defined: i) a basic model with age, sex, symptoms, and setting; ii) a clinical model with age, sex, symptoms, setting, diabetes, hypertension, dyslipidaemia, smoking, and body mass index (BMI); and iii) an extended model with clinical variables + CCS. Since all clinical parameters are known to be associated with CAD (26), all predictors were entered simultaneously in a multivariable random-effects logistic regression model. Hospital was included as a random effect to account for clustering of patients within hospitals. Because of the limited availability of exercise electrocardiography data, its incremental predictive value was explored but presented only in the Appendix. Statistically non-significant predictors with small effects (odds ratio <1.01) were omitted.

Setting variable

To account for differences in patient selection across datasets (based on referral to CCA, high-prevalence vs. referral to CCTA, low-prevalence), a dummy variable for setting was created. This variable was coded "0" (low-prevalence setting) if a patient originated from a database that was created by selecting patients who underwent CCTA (of whom only a proportion underwent CCA in addition to CCTA), and coded as "1" (high-prevalence setting), if a patient originated from a database that was created by selecting patients who underwent CCA (of whom a proportion also underwent CCTA). The intended clinical application of our prediction model is for low-prevalence populations, because for those patients the optimal diagnostic work-up should be determined based on the estimated pre-test probability (10). In contrast, all patients in the high-prevalence setting had a clinical indication for CCA, and estimating the pre-test probability is not relevant for them. Because it is inefficient to derive a prediction model only in a low-prevalence population (since the majority of patients will not undergo the reference standard), we also included databases with patients referred for CCA. These data provided valuable information on the correlations between clinical presentation, risk factors, CCTA, and CCA, which was essential for reliable imputation of co-variables and outcomes in the low-prevalence populations. By including the setting variable we could derive the model using all available data, and adjust for differences in patient selection. When applying the model for new patients with chest pain, the setting variable is set to zero.

Predictor effects may be different across the low- and high-prevalence setting, which was tested by interaction terms between setting and all other variables. Furthermore, interactions between symptoms and sex, symptoms and age, and symptoms and diabetes were tested. Linear effects of age and the log-transformed CCS were tested by including a restricted cubic spline function with 3 knots (2 *df*) (27-28).

Diagnostic performance was quantified by calculating the area under the receiver–operating-characteristic-curve (c-statistic). Reclassification was assessed using the continuous Net Reclassification Improvement (NRI) (Table 2, see Appendix) (29). A *P*-value <0.05 was considered statistically significant. Analyses were performed using Stata/SE 11 (StataCorp, Texas, USA).

Validation

The validity of the clinical model was assessed in a cross-validation procedure. The five largest datasets with sufficient numbers for reliable validation (30) and the remaining databases combined were each in turn removed from the model development sample. Subsequently, we validated each model using the database that was omitted during model development. We calculated the c-statistic and validated the model according to the following steps (23, 27-28) (see Appendix for more details on the validation steps):

1. Calibration-in-the-large

When assessing the validity of a prediction model, the first step is to check whether the average prediction approximates the average observed outcome. This concept is referred to as "calibration-in-the-large". Hereto, we compared the mean observed frequency of CAD in the validation data to the mean prediction.

2. Logistic recalibration

The second step is to test whether the overall effect of the predictors in the model is valid for the validation data.

3. Re-estimation

The third step is to re-estimate the predictor effects in the validation data and to calculate the difference (∂-coefficients) with the estimated predictor effects of the prediction model (23). Insignificant differences (high *P*-values) indicate no difference in predictor effects, supporting the validity of the model.

Results

Data collection and study population

Databases were retrieved from 18 hospitals (Table 1 and Appendix Table 2). The study population consisted of 5677 patients (3283 men, 2394 women, mean age 58, 60). Nearly all patients (5190, 91%) underwent CCTA, of which 1634 (31%) had obstructive CAD on CCTA. Among the 1634 patients with obstructive CAD on CCTA, 1083 (66%) underwent CCA which was positive in 886 (82%). Among the 3556 patients without obstructive CAD on CCTA, 526 (15%) underwent CCA which was negative in 498 (95%). Overall, 2062 patients (36%) underwent CCA, with 1176 (57%) patients diagnosed with obstructive CAD.

Missing values occurred in; age (0.1%), symptoms (0.1%), hypertension (2.2%), diabetes (3.4%), dyslipidaemia (3.3%), smoking (2.7%), BMI (6.2%), and CCS (14%).

Among the 3556 patients without obstructive CAD on CCTA, 3030 (85%) patients did not undergo CCA. CCAs imputed for these patients were mostly negative (range of 97-98.4% across the multiple imputations), which is in accordance with the high negative predictive value of CCTA. Among the 1634 patients with obstructive CAD on CCTA, 551 (34%) patients did not undergo subsequent CCA. For these patients, a positive CCA was imputed in 65-77% (range across imputations), which is in accordance with a somewhat lower positive predictive value of CCTA.

External validation of the DCS

External validation of the DCS demonstrated an overestimation of the probability of severe CAD as observed in our dataset (Figure 1 and Appendix Table 3).

Development of new prediction models

Table 2 summarizes the results of the random-effects logistic regression analysis and the continuous NRI. See Appendix Table 4 for more details. The models are available as online calculator (Figure 2). In the clinical model, all predictors except BMI were significantly associated with obstructive CAD. The clinical model improved the prediction compared to the basic model (cross-validated c-statistic improved from 0.77 to 0.79, see Table 2). Whereas an abnormal exercise electrocardiography had limited predictive value in the multivariable prediction model (see Appendix Table 5), the CCS was a major predictor (Table 2) which increased the cross-validated c-statistic from 0.79 to 0.88.

Most predictor effects decreased after addition of CCS; dyslipidaemia and smoking were no longer significant. We obtained similar results when using CCTA as outcome in patients who did not undergo CCA.

Validation

Figure 3 and Appendix Table 6 show the cross-validation results for the clinical model. The c-statistic ranged from 0.78 to 0.81. The continuous NRI is a measure of the relative increase (and decrease) in the observed proportion when the predicted probability goes up (and down) (see Appendix), which was most favourable (102%) for the Clinical+CCS model compared to the Clinical model (Table 2).

1. Calibration-in-the-large

Assessment of calibration-in-the-large demonstrated a significant difference between the average observed outcome and the predicted probability (clinical model), for Azienda Ospedaliero Universitaria Parma, Rotterdam, and the combined low-prevalence hospitals (Figure 3).

2. Logistic recalibration

Recalibration showed no significant differences between the overall hospital-specific effects of the predictors compared to the overall effects of the predictors in the clinical model (Figure 3).

Table 1. Patient characteristics. Most commonly used risk factor definitions are given. See appendix for patient characteristics and risk factor definitions by hospital.

	Low-prevalence setting N=4426 (10 hospitals)			High-prevalence setting N=1251 (8 hospitals)		
Characteristic	Available	Measure / category	Value	Available	Measure / category	Value
n per hospital		Range/ mean	80-1241 / 443		Range/ mean	85-549 / 156
Age, years	4422 (99.9%)	Mean (SD) IQR Range	57.2 (12) 49-66 18-92	1251 (100%)	Mean (SD) IQR Range	63.6 (10) 57-70 18-93
Male sex	4426 (100%)		2406 (54%)	1251 (100%)		877 (46%)
Chest pain *	4424 (99.9%)	Typical Atypical Non specific	759 (17%) 2699 (61%) 966 (22%)	1247 (99.7%)	Typical Atypical Non specific	656 (53%) 278 (22%) 313 (25%)
Diabetes †	4238 (96%)		622 (15%)	1250 (99.9%)		229 (18%)
Hypertension ‡	4300 (97%)		2475 (58%)	1251 (100%)		840 (67%)
Dyslipidaemia §	4255 (96%)		2194 (52%)	1235 (99%)		801 (65%)
Smoking ‖	4273 (97%)		1231 (29%)	1249 (99.8%)		454 (36%)
Body mass index ¶	4117 (93%)	Mean (median)	28 (27)	1206 (96%)	Mean (median)	28 (27)
Family history of CAD	3938 (89%)		1720 (44%)	265 (21%)		136 (51%)
History of cerebrovascular disease††	2531 (57%)		78 (3%)	269 (22%)		25 (9%)
History of renal artery disease	3351 (76%)		43 (1%)	316 (25%)		2 (1%)
History of peripheral arterial disease	3356 (76%)		79 (2%)	369 (29%)		10 (3%)
Exercise electrocardiography	1612 (36%)	Normal Abnormal Non diagnostic	671 (42%) 443 (27%) 498 (31%)	547 (44%)	Normal Abnormal Non diagnostic	166 (30%) 336 (61%) 45 (8%)
Coronary calcium score ‡‡	4009 (91%)	mean (SD) / median	160 (399) / 4	858 (69%)	mean (SD) / median	442 (643) / 182
		0 0-<10 ≥10-<100 ≥100-<400 ≥400	1777 (44%) 402 (10%) 749 (19%) 606 (15%) 475 (12%)		0 0-<10 ≥10-<100 ≥100-<400 ≥400	155 (18%) 44 (5%) 154 (18%) 208 (24%) 297 (35%)
Coronary computed tomography angiography §§	4287 (97%)	No obstructive CAD Moderate CAD Severe CAD	3232 (75%) 505 (12%) 550 (13%)	903 (72%)	No obstructive CAD Moderate CAD Severe CAD	324 (36%) 501 (55%) 78 (9%)
Catheter-based coronary angiography ‖‖	848 (19%)	No obstructive CAD Moderate CAD Severe CAD	406 (48%) 177 (21%) 265 (31%)	1214 (97%)	No obstructive CAD Moderate CAD Severe CAD	480 (40%) 541 (45%) 193 (26%)
Coronary computed tomography angiography §§ & Catheter-based coronary angiography ‖‖	742 (17%)	No obstructive CAD Moderate CAD Severe CAD	230 (31%) 277 (37%) 235 (32%)	867 (69%)	No obstructive CAD Moderate CAD Severe CAD	296 (34%) 495 (57%) 76 (9%)
	742 (17%)	No obstructive CAD Moderate CAD Severe CAD	356 (48%) 172 (23%) 214 (29%)	867 (69%)	No obstructive CAD Moderate CAD Severe CAD	339 (39%) 436 (50%) 92 (11%)

IQR = interquartile range, SD = standard deviation, CAD = coronary artery disease, moderate CAD = ≥50-70% stenosis, severe CAD = ≥70% stenosis or ≥50% left main stenosis. All values are given as number (percentage of listed total) unless indicated otherwise

* *According to the traditional chest pain classification (19), see Methods.*
† *Defined as fasting glucose levels of ≥126 mg/dL (≥7 mmol/L) or treatment with either diet intervention, oral glucose lowering agent or insulin.*
‡ *Defined as a blood pressure ≥140/90 mmHg or the use of antihypertensive medication.*
§ *Defined as a total cholesterol ≥200 mg/dL (≥5.2 mmol/L) or treatment with lipid-lowering drugs.*
‖ *Current or past smoking.*
¶ *Defined as weight in kilograms divided by the squared height in meters (kg/m2).*
** *Presence of CAD in a first-degree female (<65 years) or male (<55 years) relative.*
†† *History of carotid artery disease, stroke or transient ischemic attack.*
‡‡ *Agatston score (20) as measured by computed tomography.*
§§ *Some hospitals (no. 2, 3, and 18 [see Appendix Table 1]) only considered obstructive CAD (≥50% stenosis) vs. no obstructive CAD, e.g. did not consider the severe CAD category for coronary computed tomography angiography. One hospital (no. 6 [see Appendix Table 1]) did not categorize patients with ≥50% left main stenosis in the severe CAD category.*
‖‖ *Two hospitals (no. 2 & 18 [see Appendix Table 1]) only considered obstructive CAD (≥50% stenosis) vs. no obstructive CAD, e.g. did not consider the severe CAD category for catheter-based coronary angiography.*

3. Re-estimation

The differences between the predictor effects when re-estimated in specific datasets as compared to the predictor effects (clinical model) were small and insignificant, except for the effect of typical chest pain for Azienda Ospedaliero Universitaria Parma. The results indicate that predictor effects are similar across datasets (Appendix Table 6).

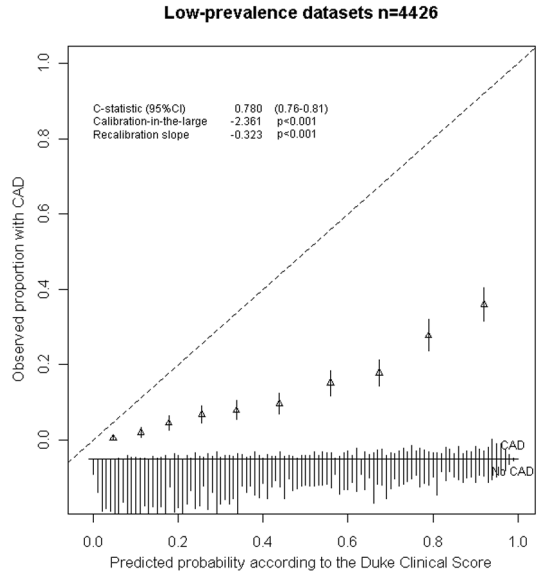


Figure 1. Calibration plot of the Duke Clinical Score (DCS). The distribution of predicted probabilities is shown at the bottom of the graphs, separately for those with and without severe coronary artery disease (CAD) (≥70% stenosis or ≥50% left main stenosis). The triangles indicate the observed proportion of severe CAD by deciles of the predicted probability according to the DCS. The DCS overestimates the probability of severe CAD. We assumed normal resting ECG findings for all patients since patients with prior CAD were excluded. If resting ECG findings were available and taken into account, the overestimation would increase further. See Appendix Table 3 for more details on the external validation of the DCS.

Calibration-in-the-large: A negative value implies that the average observed proportion in the validation data is lower than the average predicted probability. Conversely, a positive value implies that the average observed proportion in the validation data is higher than the average predicted probability. A significant result indicates significant miscalibration, whereas a non-significant result supports validity of the model.

Recalibration slope: The overall predictor effects in the validation data were re-estimated and compared with the reference overall predictor effects in the prediction model (reference is set to zero). A negative value implies that the overall effects of the predictors in the validation data are lower than in the prediction model. Conversely, a positive value implies that the overall effects of the predictors in the validation data are stronger. A significant result indicates significant miscalibration of the predictor effects, whereas a non-significant result supports validity of the model.

Discussion

Summary of key findings

Using recently collected data and modern statistical methods, we developed a well-performing prediction model for estimating the probability of CAD. The need for an updated model was demonstrated by showing that the Duke Clinical Score significantly overestimated the probability of CAD. Age, sex, symptoms, and CCS were strong predictors. The clinical model predicts probabilities between 2% for a 50-year-old female with non-specific chest pain without any risk factors, and 91% for an 80-year-old male with typical chest pain and multiple risk factors.

Previous publications

In 1979, Diamond & Forrester (11) demonstrated the importance of age, sex, and symptoms in the prediction of CAD. In spite of its limitations, its use is still recommended in current guidelines (7-8). As we previously demonstrated, the Diamond & Forrester model

tends to overestimate the probability of CAD, mainly in women (14), and does not take into account cardiovascular risk factors associated with the presence of CAD.

In 1993, Pryor *et al.* (12-13) developed the Duke Clinical Score (DCS) in a large cohort that underwent CCA. The model predicts the presence of ≥75% stenosis based on age, sex, symptoms, history of MI, smoking, dyslipidaemia, diabetes, and resting electrocardiography (ECG) findings. We demonstrated that the DCS also overestimates the probability of CAD.

Table 2. Random-effects logistic regression results*, cross-validated c-statistic, and cross-validated continuous net reclassification improvement. A random intercept for hospital was used to account for clustering of individuals within hospital. The analysis was adjusted for setting. Odds ratios presented here apply to the low-prevalence setting.

	Basic		Clinical		Clinical +CCS	
	OR	95% CI	OR	95% CI	OR	95% CI
Age (per 10 years)	**1.89**	1.74-2.04	**1.85**	1.70-2.02	1.11	0.99-1.25
Male sex	**3.89**	3.24-4.66	**3.79**	3.13-4.58	**2.19**	1.75-2.75
Atypical chest pain†	**1.93**	1.48-2.52	**1.88**	1.44-2.46	**2.05**	1.50-2.80
Typical chest pain† *(if diabetes is absent)*	**7.21**‡	5.64-9.22	**7.36**	5.64-9.61	**7.57**	5.56-10.3
Typical chest pain† *(if diabetes is present)*			**4.91**	3.16-7.63	**3.46**	2.12-5.63
Diabetes			**2.29**	1.72-3.04	**1.93**	1.41-2.65
Hypertension			**1.40**	1.18-1.67	**1.26**	1.04-1.54
Dyslipidaemia			**1.53**	1.25-1.86	1.20	0.95-1.53
Smoking			**1.59**	1.30-1.93	1.23	0.97-1.55
Log transformed CCS § (per SD)					**4.69**	3.76-5.84
CCS >0-<10 ||					**2.23**	1.34-3.74
CCS ≥10-<100 ||					**5.04**	3.38-7.52
CCS ≥100-<400 ||					**15.3**	9.96-23.5
CCS ≥400 ||					**35.9**	22.6-56.9
Cross-validated c-statistic (mean ¶)	0.77		0.79		0.88	
Cross-validated NRI ** (mean ¶)	-		35%		102%	

Bold odds ratios indicate significant association (p-value < 0.05). CAD = coronary artery disease; Coef. = beta-coefficient; OR = odds ratio; CCS = coronary calcium score; NRI = Net Reclassification Improvement

* A random effect for hospital was included to account for clustering of patients within hospitals. Body mass index was omitted from all analyses because the odds ratio was <1.01 and non-significant. Setting and the interaction between diabetes and typical chest pain, and between setting and CCS were predictive and were included in all models. All other interactions were not significant. The test for a non-linear effect of age was not significant. There was evidence for an additinonal non-linear effect of CCS beyond the log-transformation, which was considered not clinically important, and omitted for simplicity.

† Reference category is 'Non specific chest pain'

‡ Effect of typical chest pain regardless of diabetic status, since the basic model does not include diabetes

§ Natural logarithm of CCS+1.

|| Separate analysis using CCS as a categorical variable, adjusted for all other predictors in the model, reference category is CCS=0

¶ Mean of the cross-validation procedures using the four largest low-prevalence datasets and the remaining low-prevalence datasets combined.

** The continuous net reclassification improvement[29] was calculated compared to the previous model on the left and is defined as the weighted sum of the observed proportion increase among the individuals for whom the predicted probability goes up and observed proportion decrease among those for whom the predicted probability goes down (see Appendix).

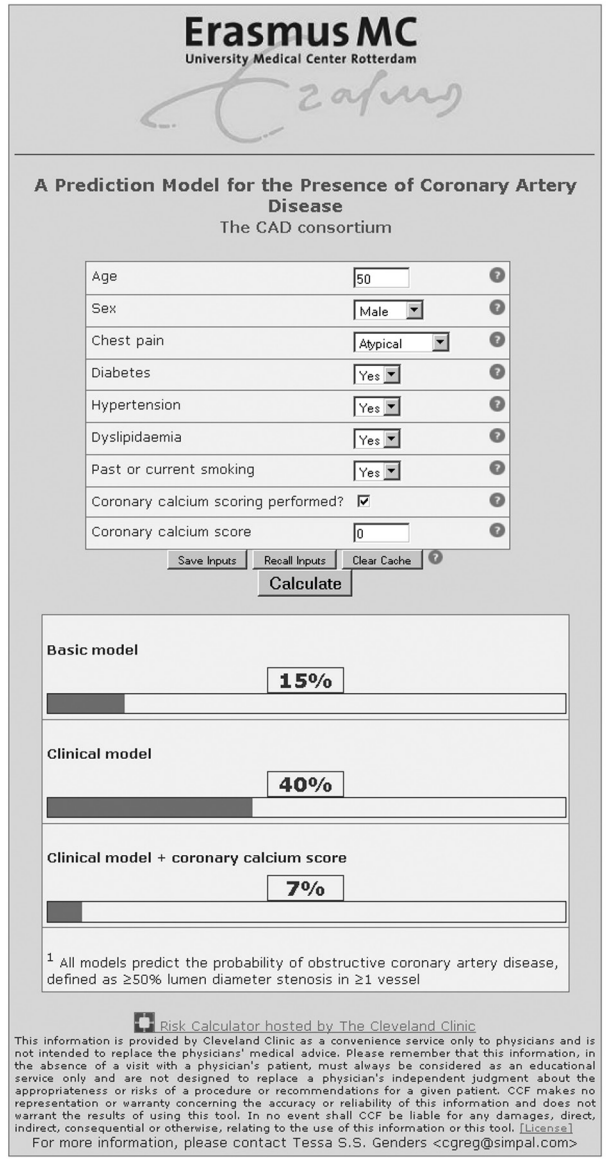


Figure 2. Online probability calculator, available through http://rcc.simpal.com/NpfpV5

Limitations

The study population was derived from existing databases, some of which were designed for other research objectives (e.g., to investigate the diagnostic accuracy of CCTA for CAD). In some studies, all patients underwent the reference standard test whereas in other studies, patients were selected for CCA based on the results of the CCTA. When evaluating the diagnostic performance of CCTA, selection based on the test results may lead to 'verification bias' or 'selective referral bias' which may bias estimates of sensitivity and specificity of such a test. In the current study, however, we did not assess the diagnostic value of CCTA, making bias less likely. Moreover, we also considered patients who did not undergo CCA, but only CCTA. We selected patients who underwent either CCTA, CCA, or both. It should

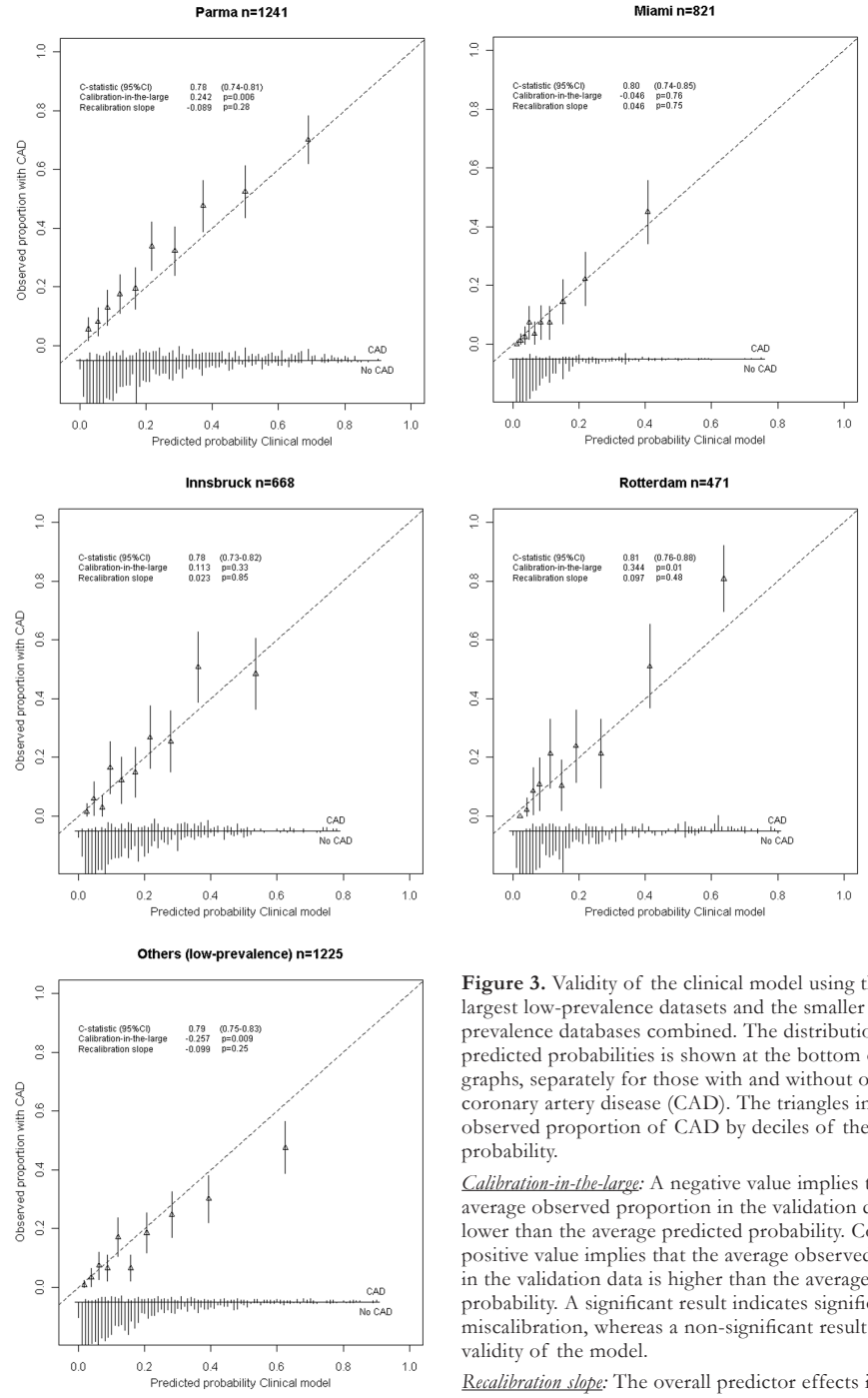


Figure 3. Validity of the clinical model using the four largest low-prevalence datasets and the smaller low-prevalence databases combined. The distribution of predicted probabilities is shown at the bottom of the graphs, separately for those with and without obstructive coronary artery disease (CAD). The triangles indicate the observed proportion of CAD by deciles of the predicted probability.

Calibration-in-the-large: A negative value implies that the average observed proportion in the validation data is lower than the average predicted probability. Conversely, a positive value implies that the average observed proportion in the validation data is higher than the average predicted probability. A significant result indicates significant miscalibration, whereas a non-significant result supports validity of the model.

Recalibration slope: The overall predictor effects in the validation data were re-estimated and compared with the reference overall predictor effects in the prediction model (reference is set to zero). A negative value implies that the overall effects of the predictors in the validation data are lower than in the prediction model. Conversely, a positive value implies that the overall effects of the predictors in the validation data are stronger. A significant result indicates significant miscalibration of the predictor effects, whereas a non-significant result supports validity of the model.

be noted, however, that not all patients presenting with chest pain will be referred for CCTA or CCA, which limits the generalizability of our results.

To also consider patients who did not undergo CCA, we multiply imputed CCA results based on CCTA results and all other covariables. Although CCA was imputed for a large proportion of patients, we believe that the high sensitivity and specificity of CCTA justifies the imputation of CCA, also because we did not consider CCTA as an explanatory variable in the prediction models. The limitation of this approach is that CCTA may overestimate the severity of disease, which in turn could have caused an overestimation of imputed severity of disease on CCA. If present, the bias in our models would tend to advocate further diagnostic workup rather than missing the diagnosis.

We combined existing data from a number of different hospitals. Since the current analysis was not the main purpose of the data collection, selection of patients, availability of data, and predictor definitions differed across hospitals (Appendix Table 1). Furthermore, some hospitals used quantitative coronary angiography to determine the degree of stenosis, whereas others used visual assessment. All-in-all, heterogeneity due to differences between protocols, level of physician experience, and guideline adherence across hospitals could have influenced our results. Despite these limitations, the models presented had generally good discrimination (c-statistic) and calibration.

Because the intended application of our model is in low-prevalence populations, cross-validation was performed using only the low-prevalence datasets. The cross-validation results in the data from Parma, Rotterdam, and the smaller hospitals combined was less favourable in terms of calibration-in-the-large, possibly explained by heterogeneity. In general, however, calibration assessed graphically (Figure 3) could be considered satisfactory, suggesting that the model is generalizable to other settings. However, further external validation of our model in other populations is needed.

Our study focused on the prediction of obstructive CAD according to age, sex, type of chest pain, cardiovascular risk factors, setting, and CCS. Unfortunately, no data were available to assess the predictive value of findings based on other imaging tests (e.g. nuclear perfusion imaging, perfusion MRI, echocardiography). Furthermore, we were unable to demonstrate incremental predictive value of exercise electrocardiography, which may be explained by the limited availability of data.

Finally, our analysis focused on the diagnostic prediction of the presence of CAD defined as ≥50% diameter stenosis in at least one vessel on CCA. Predicting severe CAD (e.g. ≥70% stenosis in the left anterior descending coronary artery, three vessel disease, or left main coronary artery disease) would be of interest as a tool to help select patients for revascularization. However, the main purpose of the current analysis is the development of a set of prediction models which can help physicians to select patients who would benefit from further testing.

Future directions

Other non-invasive imaging tests (e.g. nuclear perfusion imaging, perfusion MRI, echocardiography) for the evaluation of patients with chest pain, should also be considered as predictors for the presence of CAD. Furthermore, cost-effectiveness analyses should be

performed to establish appropriate thresholds for diagnostic testing taking into account the long term benefits and harms (4, 31).

Clinical implications

We demonstrated that the Duke Clinical Score (DCS) overestimates the likelihood of CAD, and we believe that our model improves the estimate of the pre-test probability. In contrast to the DCS, our model was developed and validated using data from different hospitals, settings, countries, and included hypertension as a predictor. Finally, our model does not require that resting ECG findings are available, which may be convenient when the model is used in primary practice.

A refined estimate of the probability allows clinicians to make better decisions as to whether and which diagnostic test is indicated in a particular patient according to the NICE guidelines and decide on further management based on the results of such tests. Our stepwise models can be used to evaluate the added value of performing CCS either prior to performing the test or after obtaining the calcium score. The results of our analysis demonstrate that CCS significantly improves the estimate of the probability of CAD, suggesting that CCS should be considered for patients with chest pain. CCS is not routinely recommended by the ACC/AHA (32) or ESC (9) guidelines, whereas the UK guidelines recommend CCS if the pre-test probability is 10-29% (10). However, triage strategies using CCS have been proposed (32) and are used in several centres. In this context, our calculator can be used both to determine whether performing CCS is clinically useful by checking whether CCS alters the probability such that clinical management would change and to determine the revised probability of CAD based on the CCS result.

Our findings also suggest that the diagnostic value of exercise electrocardiography is limited (see Appendix), consistent with its low sensitivity and specificity for detecting CAD (33), and explicit recommendations in the UK **not** to use exercise electrocardiography to diagnose or exclude CAD in patients with chest pain (10). However, many physicians argue that the prognostic information obtained by exercise electrocardiography remains important for clinical practice. On the other hand, CCS provides both diagnostic (34-36) and prognostic information (37-40).

Prediction tools are only useful when they are easily accessible at the point-of-care, which is why we designed an online calculator (Figure 2). The calculator could be implemented in electronic patient records, electronic order entry systems, or smartphone/smarttablet applications.

All-in-all, prediction models including age, sex, symptoms, and risk factors allow for accurate estimation of the likelihood of CAD in low-prevalence populations. Adding CCS improves the prediction of CAD. Implementation of the prediction models may improve clinical outcomes, however, this needs further evaluation.

Acknowledgements

The authors thank C. Greg Hagerty and Michael W. Kattan, PhD – from the Cleveland Clinic Lerner Research Institute in Cleveland, Ohio, United States of America – for developing, customizing, and improving the Cleveland Clinic Risk Calculator Constructor (http://rcc.simpal.com/).

References

1. American Heart Association. Heart Disease and Stroke Statistics – 2010 Update. Dallas, Texas: American Heart Association, 2010.
2. DeFrances CJ, Lucas CA, Buie VC, Golosinskiy A. 2006 National Hospital Discharge Survey. Natl Health Stat Report 2008:1-20.
3. Patel MR, Peterson ED, Dai D, et al. Low diagnostic yield of elective coronary angiography. N Engl J Med 2010; 362:886-895.
4. Genders TS, Meijboom WB, Meijs MF, et al. CT coronary angiography in patients suspected of having coronary artery disease: decision making from various perspectives in the face of uncertainty. Radiology 2009; 253:734-744.
5. Ladapo JA, Jaffer FA, Hoffmann U, et al. Clinical outcomes and cost-effectiveness of coronary computed tomography angiography in the evaluation of patients with chest pain. J Am Coll Cardiol 2009; 54:2409-2422.
6. Min JK, Gilmore A, Budoff MJ, Berman DS, O'Day K. Cost-effectiveness of coronary CT angiography versus myocardial perfusion SPECT for evaluation of patients with chest pain and no known coronary artery disease. Radiology 2010; 254:801-808.
7. Gibbons RJ, Balady GJ, Bricker JT, et al. ACC/AHA 2002 guideline update for exercise testing: summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines). Circulation 2002; 106:1883-1892.
8. Fraker TD, Jr., Fihn SD, Chronic Stable Angina Writing C, et al. 2007 chronic angina focused update of the ACC/AHA 2002 guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines Writing Group to develop the focused update of the 2002 guidelines for the management of patients with chronic stable angina. J Am Coll Cardiol 2007; 50:2264-2274.
9. Fox K, Garcia MA, Ardissino D, et al. Guidelines on the management of stable angina pectoris: executive summary - the Task Force on the Management of Stable Angina Pectoris of the European Society of Cardiology. Eur Heart J 2006; 27:1341-1381.
10. Cooper A, Calvert N, Skinner J, et al. Chest pain of recent onset: Assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin. London: National Clinical Guideline Centre for Acute and Chronic Conditions, 2010.
11. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med 1979; 300:1350-1358.
12. Pryor DB, Harrell FE, Jr., Lee KL, Califf RM, Rosati RA. Estimating the likelihood of significant coronary artery disease. Am J Med 1983; 75:771-780.
13. Pryor DB, Shaw L, McCants CB, et al. Value of the history and physical in identifying patients at increased risk for coronary artery disease. Ann Intern Med 1993; 118:81-90.
14. Genders TS, Steyerberg EW, Alkadhi H, et al. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating, and extension. Eur Heart J 2011; 32:1316-1330.
15. EIBIR. European Institute for Biomedical Imaging Research. http://www.eibir.org/cms/website.php. http://www.eibir.org/cms/website.php. Accessed 2009
16. Schuetz GM, Zacharopoulou NM, Schlattmann P, Dewey M. Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. Ann Intern Med 2010; 152:167-177.
17. Mowatt G, Cook JA, Hillis GS, et al. 64-Slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. Heart 2008; 94:1386-1393.

18. von Ballmoos MW, Haring B, Juillerat P, Alkadhi H. Meta-analysis: Diagnostic Performance of Low-Radiation-Dose Coronary Computed Tomography Angiography. Ann Intern Med 2011; 154:413-420.
19. Diamond GA. A clinically relevant classification of chest discomfort. J Am Coll Cardiol 1983; 1:574-575.
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; 15:827-832.
21. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999; 18:681-694.
22. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991; 10:585-598.
23. 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; 23:2567-2586.
24. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983; 39:207-215.
25. de Groot JA, Bossuyt PM, Reitsma JB, et al. Verification problems in diagnostic accuracy studies: consequences and solutions. BMJ 2011; 343:d4770.
26. Yamada H, Do D, Morise A, Atwood JE, Froelicher V. Review of studies using multivariable analysis of clinical and exercise test data to predict angiographic coronary artery disease. Prog Cardiovasc Dis 1997; 39:457-481.
27. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer, 2001.
28. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation and Updating. New York: Springer, 2008.
29. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011; 30:11-21.
30. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 2005; 58:475-483.
31. Hunink MGM, Glasziou PP, Siegel JE, et al. Decision making in health and medicine: Integrating evidence and values. Cambridge: Cambridge University Press, 2001.
32. Greenland P, Bonow RO, Brundage BH, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography) developed in collaboration with the Society of Atherosclerosis Imaging and Prevention and the Society of Cardiovascular Computed Tomography. J Am Coll Cardiol 2007; 49:378-402.
33. Kwok Y, Kim C, Grady D, Segal M, Redberg R. Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol 1999; 83:660-666.
34. Budoff MJ, Diamond GA, Raggi P, et al. Continuous probabilistic prediction of angiographically significant coronary artery disease using electron beam tomography. Circulation 2002; 105:1791-1796.
35. Haberl R, Becker A, Leber A, et al. Correlation of coronary calcification and angiographically documented stenoses in patients with suspected coronary artery disease: results of 1,764 patients. J Am Coll Cardiol 2001; 37:451-457.
36. Knez A, Becker A, Leber A, et al. Relation of coronary calcium scores by electron beam tomography to obstructive disease in 2,115 symptomatic patients. Am J Cardiol 2004; 93:1150-1152.
37. Budoff MJ, Shaw LJ, Liu ST, et al. Long-term prognosis associated with coronary calcification: observations from a registry of 25,253 patients. J Am Coll Cardiol 2007; 49:1860-1870.

38. Dedic A, Genders TS, Ferket BS, Galema TW, Mollet NR, Moelker A, et al. Stable angina pectoris: head-to-head comparison of prognostic value of cardiac CT and exercise testing. Radiology 2011;261(2):428-36.
39. Elias-Smale SE, Proenca RV, Koller MT, et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. J Am Coll Cardiol 2010; 56:1407-1414.
40. Hadamitzky M, Distler R, Meyer T, et al. Prognostic value of coronary computed tomographic angiography in comparison with calcium scoring and clinical risk scores. Circ Cardiovasc Imaging 2011; 4:16-23.
41. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 2003; 26 Suppl 1:S5-20.
42. European Society of Hypertension - European Society of Cardiology Guidelines Committee. 2003 European Society of Hypertension-European Society of Cardiology guidelines for the management of arterial hypertension. J Hypertens 2003; 21:1011-1053.
43. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP). JAMA 2001; 285:2486-2497.

Appendix

Validation methods

Validation was performed using state-of-the-art methods (23, 27-28).

1. Calibration-in-the-large

For each patient in the validation dataset, the predicted probability according to the clinical model was calculated. A logit transformation (natural logarithm of the odds) was performed to obtain the "linear prediction". A validation model (Equation 1 in Figure A1) was fitted to calculate the log odds of obstructive CAD as observed in the validation data. The linear predictor of the clinical model was included in the validation model and its coefficient was fixed at unity. In this way, the absolute and relative effects of the clinical model are maintained. The intercept (β_{new}) is the only free parameter in the model, which allows us to quantify the "calibration-in-the-large" and adjust for a difference in disease prevalence. The intercept (β_{new}) can be interpreted as the difference in log odds between the mean observed outcome in the validation data and the mean predicted probability according to the clinical model (28). In other words, we assessed calibration-in-the-large by comparing the mean observed frequency of CAD in the validation data with the mean of the predicted probabilities according the clinical model in a logistic regression model (validation model). We tested whether the difference (i.e. the intercept) was significantly different from zero.

2. Recalibration

Next, we recalibrated the model by comparing the average regression slope of the clinical model with the average regression slope in the validation data. A second linear predictor variable was added to the model (while maintaining both the previous linear predictor as offset variable and the new intercept) and its coefficient, the $\beta_{miscalibration}$ (Equation 2 in Figure A1) was estimated. This coefficient reflects the miscalibration of the predictor effects in the clinical model when compared to the predictor effects in the validation data (28). We tested whether $\beta_{miscalibration} = 0$, corresponding to the hypothesis that the prediction of the clinical model (adjusted for calibration-in-the-large) fits the data well. If significant, we conclude that the overall effects of all predictors together are different in the validation data and that the model should be revised.

3. Re-estimation

Finally, we re-estimated the predictor effects in a validation model including the linear predictor of the clinical model as offset and the new intercept (Equation 3 in Figure A1). The coefficients from this analysis refer to the difference between the re-estimated (validation coefficients) and the coefficients from the clinical model (i.e. β-coefficients). We tested whether these differences were significantly different from zero. From these analyses we can judge which predictor effects are different in the validation data as compared to the clinical model which was based on the total cohort.

Reclassification

To quantify the clinical impact of adding risk factors and CCS to the model, we calculated the Net Reclassification Improvement (NRI), which is a measure of correctly reclassified subjects penalized for those incorrectly classified. Clinically relevant cut-off values for the probability of CAD (i.e. the thresholds of probability where reclassification to another category would influence clinical management) are absent. Thus we used an extension to the traditional NRI which was recently introduced by Pencina *et al.* (29) The so-called 'continuous NRI' (or category-free NRI) does not depend on the existence of probability categories. The continuous NRI is the weighted sum of the observed proportion increase among the individuals for whom the predicted probability goes up and observed proportion decrease among those for whom the predicted probability goes down. The weights are proportional to the probability of an increase, respectively decrease, in predicted probability.

Predictive value of exercise electrocardiography (exercise tolerance testing, ETT)

ETT results were classified in four categories: 1) not performed or missing, 2) normal, 3) abnormal, and 4) non-diagnostic. Unfortunately, ETT was not performed or missing in 62% of patients. To assess its incremental diagnostic value, ETT was added to the clinical model using three approaches: i) using data from patients with ETT results available (n=2159), ii) assuming all missing ETT results were equivalent to a non-diagnostic ETT, and iii) using the imputed values for ETT (See Appendix Table 5). The three methods yielded similar results. Although in univariable analysis the ETT result was significantly associated with the presence of CAD, in the multivariable prediction model ETT did not significantly improve the cross-validated c-statistic, and the cross-validated NRI was relatively small. Moreover, details on the type of exercise test, test protocols, and criteria for abnormality are likely to be different across hospitals (data which is not available), which further limits the generalizability. We therefore omitted ETT as a predictor from our main analysis.

Appendix Table 1. Details on data collection and risk factor definitions according to hospital

Country	Hospital	Patients presenting to	Cohort	Data collection	Definition of diabetes	Definition of hypertension	Definition of dyslipidaemia	Definition of smoking
The Netherlands	Erasmus University Medical Centre, Rotterdam	Cardiologist	Consecutive patients presenting with chest pain and a pre-test probability >5% according to Diamond & Forrester	Prospective	Fasting glucose levels over 7 mmol/l or treatment currently with either diet, oral glucose lowering agents or insulin (41).	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥ 5 mmol/l	Current or past smoking
Switzerland	University Hospital Zurich	Cardiologist: 38% Cardiologist after referral by general practitioner: 23%, Cardiovascular surgery: 41%	Consecutive patients during weekdays from 8 AM-6 PM	Prospective	Self reported; impaired fasting glucose in the medical records; drug treatment for elevated glucose.	Self reported; known hypertensive with blood pressure of > 130/85 on the medical records; undergoing drug treatment for hypertension.	Self reported; increased serum triglycerides, total cholesterol and LDL cholesterol levels; undergoing lipid lowering therapy	Current or active smoker, smoking at least 1 cigarette/day in the past year
The Netherlands	University Medical Centre Utrecht	Cardiologist, after referral by general practitioner	Consecutive series of patients with stable and unstable anginal symptoms who underwent CCA. Patients with unstable angina were excluded.	Prospective	Treatment with oral antidiabetic medication or insulin.	Blood pressure ≥140/90 or use of blood pressure lowering medication for the purpose of blood pressure control (42).	Total cholesterol of ≥180 mg/dl or treatment for hypercholesterolemia.	Current smoking or smoking in past 6 months
Finland	Turku University Hospital	Cardiologist, after referral by general practitioner	Consecutive series of patients with stable angina and a pre-test probability between 30% and 70% according to clinical presentation and exercise tests (11)	Prospective	Fasting glucose levels over 7 mmol/l or oral glucose tolerance test ≥ 11 mmol/l (= measurement at 2 hours)	Systolic blood pressure ≥140 and/or diastolic blood pressure ≥ 90	Total cholesterol level of ≥ 5 mmol/l	Current or past smoking during the last year
Belgium	University Hospital Leuven	Cardiologist	Consecutive patients	Retrospective	Current treatment with diet intervention, oral glucose-lowering agent or insulin.	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥190 mg/dL or treatment with lipid-lowering drugs	Current or past smoking
Italy	Azienda Ospedaliero-Universitaria, Parma	Cardiologist	Consecutive patients	Prospective	Fasting plasma glucose level of ≥126 mg/dL treated currently with either diet intervention, oral glucose-lowering agent or insulin (41).	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥200 mg/dL or treatment with lipid-lowering drugs (43).	Current or past smoking
Switzerland	Kantonsspital St.Gallen	Cardiologist or specialist in internal medicine	Consecutive patients	Prospective, but missing data was retrospectively obtained	Self reported; impaired fasting glucose in the medical records; drug treatment for elevated glucose	Self reported; known hypertensive with blood pressure of > 130/85 mm Hg on the medical records; undergoing drug treatment for hypertension.	Self reported; increased serum triglycerides, total cholesterol and LDL cholesterol levels; undergoing lipid lowering therapy	Current or active smoker, smoking at least 1 cigarette/day in the past year
Austria	Innsbruck Medical University	Cardiologist	Consecutive patients	Prospective	Fasting plasma glucose level of ≥126 mg/dL treated currently with diet intervention, oral glucose-lowering agent or insulin (41).	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥200 mg/dL or treatment with lipid-lowering drugs (43).	Current or past smoking
United Kingdom	Barts and The London NHS Trust / London Chest Hospital	Cardiologist	Consecutive patients that underwent CCTA	Retrospective	Fasting plasma glucose level of ≥126 mg/dL treated currently with either diet intervention, oral glucose-lowering agent or insulin (41).	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥200 mg/dL or treatment with lipid-lowering drugs (43).	Current or past smoking
United States	Medical University of South Carolina, Charleston	Cardiologist	Not consecutive, patients were enrolled when study associates were available.	Prospective, but missing data was retrospectively obtained	Fasting glucose levels of ≥126 mg/dL	Blood pressure of ≥140/90 or use of blood pressure medication.	Total cholesterol level of ≥200 mg/dL	Current or past smoking
Belgium	OLV Hospital Aalst	Cardiologist	Consecutive patients that underwent CCTA	Prospective	Fasting plasma glucose level of ≥126 mg/dL treated currently with either diet intervention, oral glucose-lowering agent or insulin (41).	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥200 mg/dL or treatment with lipid-lowering drugs (43).	Current smoking or smoking in past 6 months
Russia	Federal Center of Medicine and Rehabilitation, Moscow	Cardiologist	Consecutive patients	Retrospective	Fasting glucose ≥7 mmol/L in repeated blood tests	Blood pressure of ≥140/90 mmHg	Total cholesterol level of ≥ 5.5 mmol/l	Current smoking (≥1 year)
United Kingdom	Papworth Hospital NHS Foundation Trust, Cambridge	Cardiologist	Not consecutive, patients were identified by going through the Radiology reporting system over a period of 6 months	Retrospective	Treated currently with oral glucose-lowering agent or insulin.	Hypertension in medical records, using antihypertensive drugs.	Treated currently with lipid-lowering drugs.	Current or past smoking
Germany	University Hospitals Munich, Grosshadern Campus	Cardiologist (± 2/3rd), other from within the hospital (± 1/3rd)	Consecutive patients	Prospective	Fasting plasma glucose level of ≥126 mg/dL or treatment with oral glucose-lowering agent or insulin (41).	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥200 mg/dL or treatment with lipid-lowering drugs (43).	At least one cigarette per day in the year prior study.
United States	Baptist Hospital of Miami and Baptist Cardiac and Vascular Institute, Miami, Florida	Cardiologist	Patients were enrolled during times the investigators or study associates were available	Prospective, but missing data was retrospectively obtained	Self-reported.	Self-reported.	Self-reported.	Self-reported.
Hungary	Heart Centre, Semmelweis University, Budapest	Cardiologist	Consecutive patients	Prospective, but missing data was retrospectively obtained	Fasting plasma glucose level of ≥7.0 mmol/L; 2-hour plasma glucose of ≥11.1 mmol/L or treatment with oral glucose-lowering agent or insulin (41).	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥5.2 mmol/L; LDL-cholesterol ≥3.5 mmol/L or treatment with lipid-lowering drugs.	Current or past smoking
Germany	German Heart Centre, Munich	Cardiologist	Consecutive patients	Prospective	Fasting blood glucose level >7mml/l or treatment with oral glucose-lowering agent or insulin (41).	Systolic blood pressure >140 mmHg regardless of antihypertensive therapy.	Total cholesterol level >380 mg/dL or treatment with lipid-lowering drugs.	Current or past smoking
Germany	Charité Medical School, Humboldt University, Berlin	Cardiologist, after referral by general practitioner	Consecutive patients	Prospective	Fasting plasma glucose level of ≥126 mg/dL or treated currently with either diet intervention, oral glucose-lowering agent or insulin (41).	Blood pressure of ≥140/90 mmHg or the use of anti hypertension medication (42).	Total cholesterol level of ≥200 mg/dL or treatment with lipid-lowering drugs (43).	Current or past smoking

Appendix Table 2. Patient characteristics by hospital *

Country	The Netherlands	Switzerland	The Netherlands	Finland	Belgium	Italy	Switzerland	Austria	United Kingdom
Hospital	**1. Erasmus University Medical Centre, Rotterdam**	**2. University Hospital Zurich**	**3. University Medical Centre Utrecht**	**4. Turku University Hospital**	**5. University Hospital Leuven**	**6. Azienda Ospedaliero Universitaria, Parma**	**7. Kantonsspital St.Gallen**	**8. Innsbruck Medical University**	**9. Barts and The London NHS Trust / London Chest Hospital**
Setting	Low	High	High	High	High	Low	Low	Low	Low
N	471	549	85	85	100	1241	294	668	185
Mean age (SD)	56 (10)	66 (11)	61 (5)	64 (7)	64 (10)	61 (11)	55 (11)	58 (11)	51 (11)
Male sex (%)	52	74	67	62	57	57	57	56	55
Diabetes (%)	14	17	24	18	13	12	12	16	21
Hypertension (%)	50	66	62	44	63	57	42	67	46
Dyslipidaemia (%)	59	63	64	31	73	47	28	71	65
Smoking (%)	29	41	25	23	47	27	26	46	37
Body mass index †	27	26	27	27	27	26	28	26	N/A
Family history (%)	45	N/A	N/A	N/A	N/A	44	N/A	39	35
Chest pain type									
Typical CP (%)	31	53	55	52	56	31	12	6	11
Atypical CP (%)	53	11	19	44	31	49	49	84	49
Non-specific CP (%)	16	36	26	4	13	20	39	10	40
Exercise electrocardiography available (%)	90	47	59	100	37	18	22	51	20
Normal (%)	45	36	20	13	32	44	38	27	35
Non-diagnostic (%)	33	0	16	0	30	30	45	33	22
Abnormal (%)	22	64	64	87	38	26	17	40	43
CCTA available (%)	96	100	100	100	<10 §	100	100	99	100
No obstructive CAD (%)	69	32	21	52	N/A	65	90	70	78
Moderate CAD (%)	26	68 ‡	79 ‡	27	N/A	15	10	6	12
Severe CAD (%)	5	N/A	N/A	21	N/A	20	0	24	10
CCS available (%)	98	100	100	98	<10 §	100	100	99	88
Mean / median	206 (15)	492 (263)	462 (216)	385 (136)	N/A	212 (15)	117 (3)	143 (4)	95 (0)
CCA available (%)	26	100	100	100	98	19	22	28	18
No obstructive CAD (%)	41	32	42	52	39	51	53	52	79
Moderate CAD (%)	27	68 ‡	39	13	6	28	47	12	12
Severe CAD (%)	32	N/A	19	35	55	21	0	36	9

Country	United States	Belgium	Russia	United Kingdom	Germany	United States	Hungary	Germany	Germany
Hospital	10. Medical University of South Carolina	11. OLV Hospital Aalst	12. Federal Centre of Medicine and Rehabilitation	13. Papworth Hospital NHS Foundation Trust, Cambridge	14. University Hospitals Munich, Grosshadern Campus	15. Baptist Hospital of Miami	16. Heart Centre, Semmelweis University, Budapest	17. German Heart Center, Munich	18. Charité Medical School, Berlin
Setting	High	Low	High	Low	High	Low	Low	Low	High
N	95	91	90	107	80	821	297	251	167
Mean age (SD)	59 (10)	59 (10)	60 (8)	58 (12)	63 (13)	52 (12)	59 (13)	60 (12)	63 (9)
Male sex (%)	77	42	64	52	48	49	56	57	70
Diabetes (%)	40	13	6	15	18	18	21	6	17
Hypertension (%)	77	45	81	61	71	53	78	74	73
Dyslipidaemia (%)	82	58	89	50	58	38	70	57	62
Smoking (%)	46	21	49	63	23	19	27	22	20
Body mass index †	32	27	29	29	27	29	28	26	26
Family history (%)	58	27	69	60	24	50	36	38	N/A
Chest pain type									
Typical CP (%)	63	8	64	37	1	4	17	3	59
Atypical CP (%)	36	42	27	54	43	77	24	96	26
Non-specific CP (%)	1	50	9	9	56	19	59	1	15

Exercise electrocardiography available (%)	0	71	<10 §	56	68	0 §	74	<10 §	68
Normal (%)	N/A	63	N/A	18	35	N/A	45	N/A	35
Non-diagnostic (%)	N/A	32	N/A	40	22	N/A	16	N/A	22
Abnormal (%)	N/A	5	N/A	42	42	N/A	39	N/A	42
CCTA available (%)	100	95	<10 §	100	0	100	68	100	0
No obstructive CAD (%)	52	78	N/A	80	N/A	41	83	88	N/A
Moderate CAD (%)	14	16	N/A	7	N/A	25	7	6	N/A
Severe CAD (%)	34	6	N/A	13	N/A	34	10	6	N/A
CCS available (%)	95	96	<10 §	100	0	59	0	92	0
Mean / median	371 (45)	200 (21)	N/A	254 (4)	N/A	68 (0)	N/A	94 (0)	N/A
CCA available (%)	98	25	92	9	100	68	34	5	100
No obstructive CAD (%)	77	44	32	30	45	24	48	12	45
Moderate CAD (%)	5	39	10	10	54 ‡	24	5	11	54 ‡
Severe CAD (%)	18	17	58	60	N/A	52	47	77	N/A

SD = standard deviation, CP = chest pain, CCTA = coronary computed tomography angiography, CAD = coronary artery disease, CCS = coronary calcium score, CCA = conventional coronary angiography, N/A = not availabl
** Means, medians and proportions were calculated amongst individuals with non-missing variable*
† Median body mass Index, defined as Weight (kg) / Height (meters)²
‡ No distinction between ≥50% and ≥70% stenosis was made, includes any obstructive CAD
§ Variable was recorded in less than 10% of patients, estimates not shown.

Appendix Table 3. External validation of the Duke Clinical Score (DCS)* (12-13)

	Hospitals (n=10) with data on severe CAD† n=4426	
c-statistic (95%CI)	**0.78 (0.76-0.81)**	
Validation step:‡	**Coefficient**	**p-value**
1. Calibration-in-the-large		
∂ Intercept	-2.361	<0.001
2. Logistic recalibration		
$\beta_{miscalibration}$	-0.323	<0.001
3. Re-estimation		
∂ Intercept	-1.341	0.006
∂ Age (per year)	-0.013	0.06
∂ Female sex	0.822	<0.001
∂ Atypical chest pain†	-0.655	0.001
∂ Typical chest pain†	-1.103	<0.001
∂ Diabetes	0.344	0.03
∂ Hypertension	Not considered by the DCS	
∂ Dyslipidaemia	0.298	0.02
∂ Smoking	0.045	0.73

** Assuming normal resting ECG findings for all patients since patients with prior CAD were excluded. Including potential resting ECG abnormalities would increase the predicted probability according to the DCS and increase the observed overestimation.*
† Severe CAD was defined as ≥70% stenosis or ≥50% left main stenosis. One hospital (no. 6 [see Appendix Table 1]) did not categorize patients with ≥50% left main stenosis in the severe CAD category.
‡ See Appendix

Appendix Table 4. Random-effects logistic regression analysis reporting coefficients

	Basic			Clinical			Clinical + CCS		
	Coef.	**OR**	**95% CI**	**Coef.**	**OR**	**95% CI**	**Coef.**	**OR**	**95% CI**
Intercept	-6.917	-	-	-7.539	-	-	-5.975	-	-
Age	0.063	**1.07**	1.06-1.07	0.062	**1.06**	1.06-1.07	0.011	1.01	1.00-1.02
Male sex	1.358	**3.89**	3.24-4.66	1.332	**3.79**	3.13-4.58	0.786	**2.19**	1.75-2.75
Atypical chest pain*	0.658	**1.93**	1.48-2.52	0.633	**1.88**	1.44-2.46	0.718	**2.05**	1.50-2.80
Typical chest pain*	1.975	**7.21**	5.64-9.22	1.998	**7.37**	5.64-9.63	2.024	**7.57**	5.56-10.3
Diabetes				0.828	**2.29**	1.73-3.04	0.658	**1.93**	1.41-2.64
Hypertension				0.338	**1.40**	1.18-1.67	0.235	**1.26**	1.04-1.54
Dyslipidaemia				0.422	**1.53**	1.25-1.86	0.185	1.20	0.95-1.53
Smoking				0.461	**1.59**	1.30-1.94	0.207	1.23	0.97-1.55
Log transformed CCS †							0.577	**1.78**	1.64-1.93
Setting ‡	1.065	**2.90**	1.69-4.99	1.049	**2.85**	1.60-5.10	1.566	**4.79**	2.19-10.5
Setting x log transformed CCS							-0.157	**0.86**	0.77-0.95
Diabetes x typical chest pain				-0.402	**0.67**	0.42-1.06	-0.780	**0.46**	0.27-0.77

Bold odds ratios indicate significant association (p-value < 0.05).
Coef. = beta-coefficient, OR = odds ratio, SE = standard error, CCS = coronary calcium score
** Reference category is 'Non specific chest pain'*
† Natural logarithm of CCS+1.
‡ High-prevalence vs. low-prevalence (reference)

Appendix Table 5. Predictive value of exercise electrocardiography.

	Method i (complete cases) n=2159			**Method ii (missing = non-diagnostic)** n=5677			**Method iii (imputed ETT values)** n=5677		
	Coef.	**OR**	**95% CI**	**Coef.**	**OR**	**95% CI**	**Coef.**	**OR**	**95% CI**
Intercept	-8.511	-	-	-8.021	-	-	-8.031	-	-
Age	0.065	**1.07**	1.05-1.08	0.062	**1.06**	1.06-1.07	0.064	**1.07**	1.06-1.08
Male sex	1.511	**4.53**	3.40-6.04	1.349	**3.85**	3.18-4.67	1.391	**4.02**	3.28-4.92
Atypical chest pain*	0.700	**2.01**	1.29-3.14	0.612	**1.84**	1.41-2.42	0.588	**1.80**	1.36-2.39
Typical chest pain*	1.800	**6.05**	3.88-9.44	1.953	**7.05**	5.37-9.24	1.895	**6.66**	5.02-8.82
Diabetes	0.984	**2.68**	1.76-4.07	0.823	**2.28**	1.72-3.03	0.796	**2.22**	1.67-2.94
Hypertension	0.555	**1.74**	1.31-2.32	0.339	**1.40**	1.18-1.67	0.332	**1.39**	1.17-1.66
Dyslipidaemia	0.372	**1.45**	1.06-1.98	0.423	**1.53**	1.25-1.86	0.410	**1.51**	1.22-1.86
Smoking	0.552	**1.74**	1.30-2.33	0.457	**1.58**	1.29-1.93	0.437	**1.55**	1.25-1.92
ETT non-diagnostic ‡	0.485	**1.62**	1.08-2.44	0.473	**1.61**	1.21-2.13	0.389	1.48	0.98-2.23
ETT abnormal ‡	0.979	**2.66**	1.88-3.77	0.946	**2.58**	1.84-3.61	0.965	**2.63**	1.76-3.91
Setting §	1.221	**3.39**	1.93-5.97	0.971	**2.64**	1.47-4.75	0.859	**2.36**	1.30-4.28
Diabetes x typical chest pain	-0.479	0.62	0.30-1.26	-0.397	0.67	0.42-1.07	-0.401	0.67	0.42-1.07
Cross-validated c-statistic (mean ||)		N/A			0.79			0.80	
Cross-validated NRI ¶ (mean ||)		N/A			18%			33%	

Bold odds ratios indicate significant association (p-value < 0.05). Coef. = beta-coefficient, OR = odds ratio, SE = standard error, ETT = exercise electrocardiography, N/A = not available
** Reference category is 'Non specific chest pain'*
† Natural logarithm of CCS+1.
‡ Reference category is 'ETT normal'
§ High-prevalence vs. low-prevalence (reference)
|| Mean of the cross-validation procedures using the five largest datasets and the remaining hospitals combined.
¶ The continuous net reclassification improvement(29) was calculated compared to the 'Clinical Model' and is defined as the weighted sum of the observed probability increase among the individuals for whom the predicted probability goes up and observed probability decrease among those for whom the predicted probability goes down (Appendix).

Figure A1. Equations

1 $Logit(pCAD)_{validation} = \alpha_{new} + \mathbf{offset}(lp)$

2 $Logit(pCAD)_{validation} = \alpha_{new} + \beta_{miscalibration} \cdot lp + \mathbf{offset}(lp)$

3 $Logit(pCAD)_{validation} = \alpha_{new} + \delta_{age} \cdot X_{age} + \delta_{sex} \cdot X_{sex} + \delta_{typical} \cdot X_{typical} + \delta_{atypical} \cdot X_{atypical} + \mathbf{offset}(lp)$

Logit = natural log odds of the probability, pCAD = probability of obstructive coronary artery disease, □ = intercept of logistic regression model, offset = regression coefficient fixed at unity, □ = regression coefficient, □ = difference between □ *validation* and □ *development*, *lp* = linear predictor of the clinical model

Appendix Table 6. Validation of the clinical model in a cross-validation procedure.

	Parma n=1241		**Miami** n=821		**Innsbruck** n=668		**Rotterdam** n=471		**Others combined** n=1225	
c-statistic (95%CI)	0.78 (0.74-0.81)		0.80 (0.74-0.85)		0.78 (0.73-0.82)		0.81 (0.76-0.88)		0.79 (0.75-0.83)	
Validation step*:	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
1. Calibration-in-the-larg										
∂ Intercept	**0.334**	0.001	-0.046	0.76	0.113	0.33	**0.344**	0.01	**-0.257**	0.009
2. Logistic recalibration										
βmiscalibration	-0.117	0.14	0.046	0.75	0.023	0.85	0.097	0.48	-0.099	0.25
3. Re-estimation										
∂ Intercept	1.010	0.14	-0.292	0.77	-0.355	0.72	-0.420	0.72	0.044	0.95
∂ Age (per year)	-0.003	0.72	0.015	0.29	0.018	0.18	0.013	0.43	-0.000	0.97
∂ Male sex	-0.015	0.94	0.199	0.56	0.184	0.50	0.161	0.59	-0.405	0.05
∂ Atypical chest pain †	-0.470	0.06	-0.782	0.05	-0.569	0.17	-0.311	0.53	-0.050	0.84
∂ Typical chest pain †	**-0.615**	0.01	-0.485	0.44	-0.839	0.14	-0.194	0.70	-0.148	0.61
∂ Diabetes	0.241	0.32	0.088	0.79	-0.043	0.88	0.038	0.92	0.214	0.37
∂ Hypertension	0.096	0.59	-0.433	0.18	-0.274	0.30	0.219	0.45	0.321	0.20
∂ Dyslipidaemia	-0.131	0.43	0.008	0.98	-0.042	0.89	-0.117	0.70	-0.166	0.46
∂ Smoking	-0.121	0.57	0.449	0.77	0.022	0.93	0.222	0.50	-0.386	0.08

Reported ∂-coefficients denote the difference between the hospital-specific predictor effects and the predictor effects as estimated in the development data (see Methods). Insignificant differences (high P-values) indicate no difference, supporting the validity of the model.
OR = odds ratio, SE = standard error
** See Appendix*
† Reference category is 'Non specific chest pain'

Part IV

Cost-Effectiveness Analyses

Chapter 9

Coronary CT Angiography in Patients Suspected of Having Coronary Artery Disease: Decision Making from Various Perspectives in the Face of Uncertainty

Radiology. 2009;253(3):734-44.

Tessa S. S. Genders
W. Bob Meijboom
Matthijs F. L. Meijs
Joanne D. Schuijf
Nico R.A. Mollet
Annick C. Weustink
Francesca Pugliese
Jeroen J. Bax
Maarten J. Cramer
Gabriel P. Krestin
Pim J. de Feyter
M. G. Myriam Hunink

Abstract

Purpose

To determine the cost-effectiveness of computed tomography (CT) coronary angiography as a triage test, performed prior to conventional coronary angiography, by using a Markov model.

Materials and Methods

A Markov model was used to analyze the cost-effectiveness of coronary CT angiography performed as a triage test prior to conventional coronary angiography from the perspective of the patient, physician, hospital, health care system, and society by using recommendations from the United Kingdom, the United States, and the Netherlands for cost-effectiveness analyses. For coronary CT angiography, a range of sensitivities (79%–100%) and specificities (63%–94%) were used to help diagnose significant coronary artery disease (CAD). Optimization criteria (i.e., outcomes considered) were: revised posttest probability of CAD, life-years, quality-adjusted life-years (QALYs), costs, and incremental cost-effectiveness ratios (ICERs). Extensive sensitivity analysis was performed.

Results

For a prior probability of CAD of less than 40%, the probability of CAD after coronary CT angiography with negative results was less than 1%. The Markov model calculations from the patient/physician perspective suggest that coronary CT angiography maximizes life-years respectively in 60-year-old men and women at a prior probability of less than 38% and 24% and maximizes QALYs at a prior probability of less than 17% and 11%. From the hospital/health care perspective, coronary CT angiography helps reduce health care and direct non-health care–related costs (according to UK/US recommendations), regardless of prior probability, and lowers all costs, including production losses (Netherlands recommendations) at a prior probability of less than 87%–92%. Analysis performed from a societal perspective by using a willingness-to-pay threshold level of €80 000/QALY suggests that coronary CT angiography is cost-effective when the prior probability is lower than 44% and 37% in men and women, respectively. Sensitivity analyses showed that results changed across the reported range of sensitivity of coronary CT angiography.

Conclusion

The optimal diagnostic work-up depends on the optimization criterion, prior probability of CAD, and the diagnostic performance of coronary CT angiography.

Introduction

Patients with chest pain who are suspected of having coronary artery disease (CAD) usually undergo conventional coronary angiography to help diagnose CAD. These patients may be imaged non-invasively with computed tomography (CT) coronary angiography and avoid invasive conventional coronary angiography. Systematic reviews and meta-analyses have shown that coronary CT angiography is accurate in helping diagnose CAD with a patient-level sensitivity of 96%–99% and a specificity of 74%–94% (1–3). Although coronary CT angiography is rapidly being introduced in clinical practice as a triage test performed prior to conventional coronary angiography, its effect on patient outcome and cost-effectiveness has not yet been determined. Every year, approximately 4.5 people per 1000 visit a doctor with chest pain (4); more than 2 million conventional coronary angiograms are performed in Europe (5) and approximately 1.7 million are performed in the United States (6). The use of coronary CT angiography as an initial triage test could reduce costs and minimize discomfort for patients. However, a tradeoff must be made between the benefits and disadvantages of coronary CT angiography. Current guidelines recommend the use of coronary CT angiography in patients with a low to intermediate prior probability of CAD who are unable to exercise or who have inconclusive functional test results (7). However, what constitutes a low to intermediate prior probability remains to be elucidated. The purpose of this study was to determine the cost-effectiveness of coronary CT angiography performed as a triage test prior to conventional coronary angiography in patients with suspected CAD.

Materials and Methods

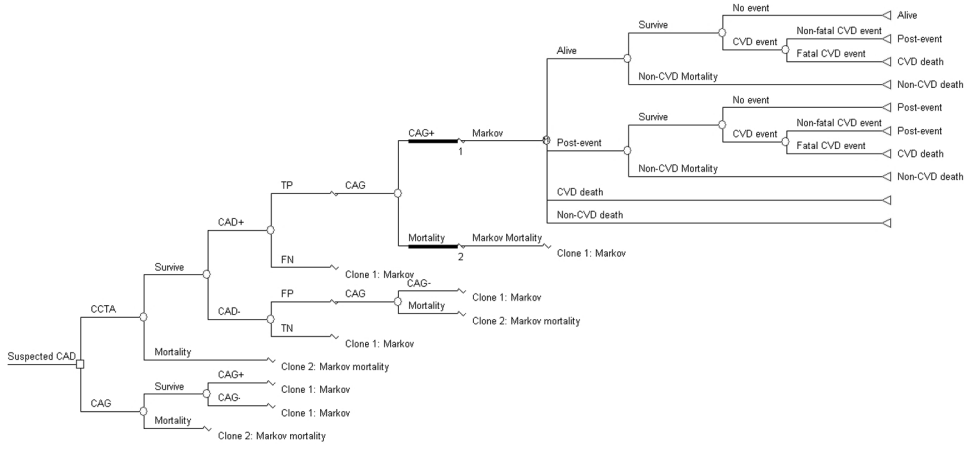
One author (J.J.B.) is the recipient of research grants from Edwards Lifesciences (Nyon, Switzerland), Biotronik (Tilburg, the Netherlands), GE Healthcare (Brussels, Belgium), BMS Medical Imaging (North Billerica, Mass), St Jude (St Paul, Minn), and Medtronic (Maastricht, the Netherlands). All other authors had full control over the inclusion of any data and information that might have represented a conflict of interest.

Decision Model

We developed a decision model (in DATA Pro, 2009 Suite; TreeAge Software, Williamstown, Mass) to evaluate the use of 64-section coronary CT angiography (new strategy) as an initial imaging test, followed by conventional coronary angiography if coronary CT angiographic results were positive when compared with conventional coronary angiography only (current practice) (Figure 1). Short-term outcomes related to the diagnostic imaging tests were modeled with a decision tree and a Markov model (cycle length, 1 year) was used to model long-term outcomes. We modeled whether a patient was alive or dead and whether a cardiovascular event occurred. Quality of life was modeled on the chance of successful relief from angina by means of treatment. Costs were estimated for diagnostic tests, treatment (percutaneous coronary intervention [PCI], coronary artery bypass graft [CABG], medication), and events during follow-up. The decision was analyzed from the perspectives of the physician, patient, hospital, health care system, and society by using various optimization criteria (8) and by taking into account the uncertainty involved. Data Sources and Assumptions We searched the literature for input data and for data to be used in sensitivity analyses (Appendix E1, [http://radiology.rsna.org/content/253/3/734/suppl/ DC1]). All variables were entered in the model as distributions. Recent studies and meta-analyses were used to

derive a range of per-patient specificities for coronary CT angiography (64%–93%) (1,3,10). To account for the inverse relationship between sensitivity and specificity, we modeled the sensitivity (79%–100%) as a function of specificity and the diagnostic odds ratio and its reported range (1). Specificity and the diagnostic odds ratio were assumed to be independent of age, sex, risk factors, and presentation. Significant CAD was defined as a reduction of lumen diameter of 50% or more, which is appropriate in the setting of selecting patients for some form of treatment or further diagnostic testing. The mortality rate for coronary CT angiography was assumed to be equivalent to that for intravenous contrast materials (11). Conventional coronary angiography was used as the reference standard, assuming 100% sensitivity and specificity. Sex-specific probabilities of 60-year-old patients for having CAD were determined by using the method of Diamond and Forrester (9), which, although published in 1979, is still commonly used to estimate the prior probability of CAD on the basis of age, sex, and type of chest pain. A wide range was used to account for the uncertainty in the prior probability of CAD (mean, 79% for men and 65% for women). Cardiovascular event rates in patients without CAD (patients with true-negative or false-positive test results) were assumed to be equivalent to first cardiovascular event rates in the general population and were calculated with a recently published prediction model by D'Agostino *et al.* (15) on the basis of the Framingham Heart Study, which was the most representative study population available. To account for the higher risk in our target population, event rates were modeled for the average risk factor profiles of patients as observed in a cohort of patients presenting to our institution with chest pain. Cardiovascular events included coronary death, myocardial infarction, coronary insufficiency, angina, stroke, cardiac arrest, peripheral arterial disease, and heart failure. We estimated that symptomatic patients diagnosed with CAD who are subsequently treated (those with true-positive test results) will have a 1.5-fold (95% confidence interval: 1.3-1.7) cardiovascular event rate compared with patients without symptomatic CAD on the basis of a multivariable Cox model reported from the EUROPA study (16). Following a first cardiovascular event, the cardiovascular recurrence rate will be 1.44-fold (95% confidence interval: 1.25-1.66) higher than that for a patient with CAD and no previous cardiovascular event (16). We assumed that the cardiovascular event rate is

Figure 1. Decision tree for treatment shows decision node (□), chance nodes (□), and Markov nodes (M). Catheter-based coronary angiography (CAG) performed after coronary CT angiography (CCTA) reveals positive results. Clone indicates structure identical to another sub tree, which is marked with black line and corresponding number. CVD = cardiovascular disease, FN = false negative, FP = false positive, TN = true negative, TP = true positive.



reduced by means of treatment (hazard rate ratio, 0.63; range 0.44– 0.88) (17,18), modeled with a combined weighted average effectiveness of CABG, PCI, and treatment with medication (28,29). Missed CAD patients (those with false negative results) forego the benefit of treatment, implying reduced quality of life and a 2.4-fold (range, 1.5–3.2) cardiovascular event rate compared with a patient without CAD (17,18). After a cardiovascular event during the follow-up in patients with an initial negative test result, CAD is diagnosed and treated and the patient will be subject to a higher recurrence rate from then on. Cardiovascular disease-related 1-year mortality (including in-hospital mortality) following a cardiovascular event was assumed to be 17% (range, 10%–25%) (18,19,20). Age- and sex-specific risks of radiation-induced fatal cancer associated with performing CT or conventional coronary angiography were based on reported estimates of lifetime-attributable cancer incidence (21) and adjusted to reflect mortality given the BEIR VII report (22). Age- and sex-specific noncardiovascular mortality rates were obtained from the Dutch Central Bureau for Statistics (11). Technical details and assumptions are clarified in the Appendix.

Quality of Life

Quality of life estimates following treatment were a pooled weighted average taking into account that 51% of patients diagnosed with symptomatic CAD undergo PCI, 25% undergo CABG, and 24% will be on medication only (36). Five years after treatment, 15% of CABG treated patients and 16% of PCI-treated patients still have angina (23); the quality of life weight for angina is 0.74 (range, 0.71–0.77) and without angina is 0.87 (range, 0.86–0.88) (23). Patients in whom the diagnosis was missed (those with false-negative test results) were all assumed to have angina during follow-up until a cardiovascular event occurred, after which they would be diagnosed and treated (23). A disutility of 0.04 (range, 0.02–0.07) QALYs was modeled for a cardiovascular event (such as myocardial infarction) during follow-up (24).

Costs

Costs for both CCTA and CAG were determined with a cost analysis and included direct health care costs (personnel, materials, equipment), indirect health care costs (housing, overhead), direct non-health care– related costs (patient travel and time costs), and indirect non-health care– related costs (production losses). Costs for CABG and PCI were determined on the basis of estimates from the National Health Care Authority (27). Annual costs for medical therapy for diagnosed CAD patients were determined on the basis of treatment with aspirin, nitrates, statins, and angiotensin-converting enzyme inhibitors and included one follow-up visit per year (25). Costs for cardiovascular events were estimated to range from €8000– €18 000 (mean, €13 000), which is consistent with previously published data (38). Costs for reinterventions were taken into account by using weighted averages of reintervention rates for the treatment options. Non-cardiovascular disease–related health care costs arising from increased longevity (inducing costs associated with increased life expectancy) were not taken into account to avoid a financial advantage of reduced longevity (33–35). All costs were converted to year 2007 rates, given Dutch consumer price indices, and reported in euro's (14). In 2007, €1.00 was equivalent to US $1.37. All costs were represented by gamma distributions.

Data Analysis

To analyze the decision from various perspectives, we used several optimization criteria (i.e., the revised posttest probability of CAD, life-years, quality-adjusted life-years (QALYs), costs, and incremental cost-effectiveness ratios (ICERs).

Patient and physician perspectives

To reflect the physician and patient perspectives, we determined the revised (posttest) probability for positive and negative coronary CT angiographic results depending on the prior (pretest) probability. The probability of having CAD after a coronary CT angiogram with positive results is equivalent to the positive predictive value. The probability of CAD after a coronary CT angiogram with negative results is equivalent to 1 minus the negative predictive value. Posttest probabilities were calculated by using the per-patient sensitivities and specificities as reported in the literature (1,3,10). Next, we determined the strategy that maximized life-years and QALYs and calculated the prior probability threshold level below which coronary CT angiography would be preferred.

Hospital and health care perspectives

In an analysis from the hospital perspective, we calculated the diagnostic costs and determined the prior probability threshold level below which coronary CT angiography would reduce cost. The analysis from the health care perspective considered QALYs and health care costs and was performed according to UK recommendations, discounting both future costs and effectiveness at 3.5% (31,32).

Societal perspective

A cost-effectiveness analysis from the societal perspective was performed according to US recommendations, which considered QALYs, health care costs, and direct non-health care–related costs (patient time and travel costs), and discounted both future costs and effectiveness at 3% (33– 35). Subsequently, an analysis from the societal perspective was performed according to Dutch recommendations, which, in addition to the above, also took productivity losses (friction costs) into account and discounted future costs and effectiveness at 4% and 1.5%, respectively (26). A willingness-to-pay (WTP) threshold level of €80 000/QALY, as recommended by the Dutch Council for Public Health (39), was used to assess cost-effectiveness. If the ICER (difference in costs divided by the difference in effectiveness, of strategy A compared with strategy B) is lower than the societal WTP threshold level, we conclude that strategy A is a cost-effective alternative to strategy B.

By using one- and two-way sensitivity analyses, we assessed the effect of varying each parameter across its distribution. Probabilistic sensitivity analysis was performed by drawing from all variable distributions (Table E1 [http://radiology.rsna.org/content/253/3/734/suppl/DC1]) by using a cohort Monte Carlo simulation of 100 000 samples (one-level). We calculated the probability that performing coronary CT angiography as the initial test was cost-effective compared with conventional coronary angiography for varying WTP threshold levels and present acceptability curves. Expected value of perfect information (EVPI; simulation with 100 000 samples) was calculated to assess the value of performing further research and partial EVPI calculations (two-level simulation performed with 1000 x 1000 samples) identified the parameters that were the major sources of uncertainty (40–42).

Specific Scenario: Cohort Study

To determine the cost-effectiveness of coronary CT angiography at our own institution, we modeled a specific scenario in which we reanalyzed the model on the basis of a study that evaluated 64-section coronary CT angiography (10) in our institution by using the per-patient sensitivity and specificity and the observed sex-specific prior probabilities of disease from this study. The study population in this cohort comprised 233 stable patients suspected of having CAD who presented with chest pain suggestive of angina. In this study, all patients were referred for conventional coronary angiography on the basis of their history or functional test results that suggested the presence of cardiac ischemia and all patients underwent coronary CT angiography prior to conventional coronary angiography. This study was approved by the institutional review board and all patients signed informed consent. All cost-effectiveness analyses were recalculated for the specific scenario.

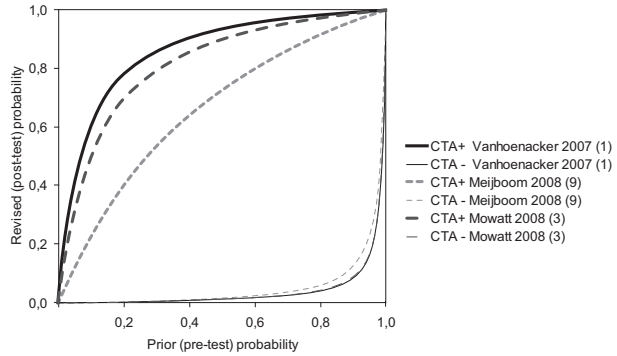


Figure 2. Revised (posttest) probability of CAD, plotted as function of prior (pretest) probability of CAD for positive and negative coronary CT angiography (CCTA) results by using test characteristics found in cohort study and meta-analyses (1,3,10). Note that for low to moderate prior probability of disease (prior probability < 40%), Coronary CT angiography with negative results virtually excludes CAD (revised probability < 1%), regardless of whether cohort study or meta-analysis results are used.

Results

Cohort Study

Data from 156 men and 77 women with stable angina in the cohort were analyzed. Of these, 113 (72.4%) men and 33 (42.9%) women had significant CAD seen at conventional coronary angiography. Mean patient age was 60 years (range, 49-74 years), 151 (64.8%) presented with typical chest pain, 75 (32.2%) smoked, 149 (63.9%) had hypertension, 47 (20.2%) had diabetes mellitus, and 36 (15.5%) had experienced prior myocardial infarction. Given these risk factors, the average annual hazard rate for a cardiovascular event was calculated by using the Framingham Heart Study, resulting in a rate of 0.024 (range, 0.014 – 0.077) for men and 0.014 (0.008–0.053) for women (15). Details are provided in the Appendix (http://radiology.rsna.org/content/253/3/734/suppl/DC1).

Reference Case Analysis

Patient and physician perspectives

In the setting of a prior probability of disease of less than 40%, the revised probability of CAD after coronary CT angiography with negative results is less than 1% (Figure 2), regardless of whether the test characteristics were based on the meta-analysis or the cohort study. In contrast, the probability of CAD after coronary CT angiography with positive results varies over a wide range, depending on the prior probability (Figure 2).

The analysis of life-years demonstrated that below a prior probability threshold level of 38% in men and 24% in women, patients would, on average, benefit from coronary CT angiography performed as the initial imaging test (Figure 3). Coronary CT angiography maximizes QALYs at a prior probability of less than 17% in men and less than 11% in women (Figure 3).

Hospital and health care perspectives

Coronary CT angiography lowered diagnostic costs below a prior probability of disease of 84% in men and women when compared with conventional coronary angiography. By using the UK recommendations for cost-effectiveness analysis, Coronary CT angiography lowered health care costs across all prior probabilities (Figure 3). For men, there was a small gain in QALYs of 0.037 with conventional coronary angiography, a small increment in cost of €589, and an ICER of €15 915/QALY gained when compared with coronary CT angiography. For women, there was a QALY gain of 0.036, a cost increment of €714, and an ICER of €19 913/QALY (Table). In both men and women, performing conventional coronary angiography alone increased the net health benefit compared with performing CT followed by conventional coronary angiography by 0.03 QALY equivalents (Table).

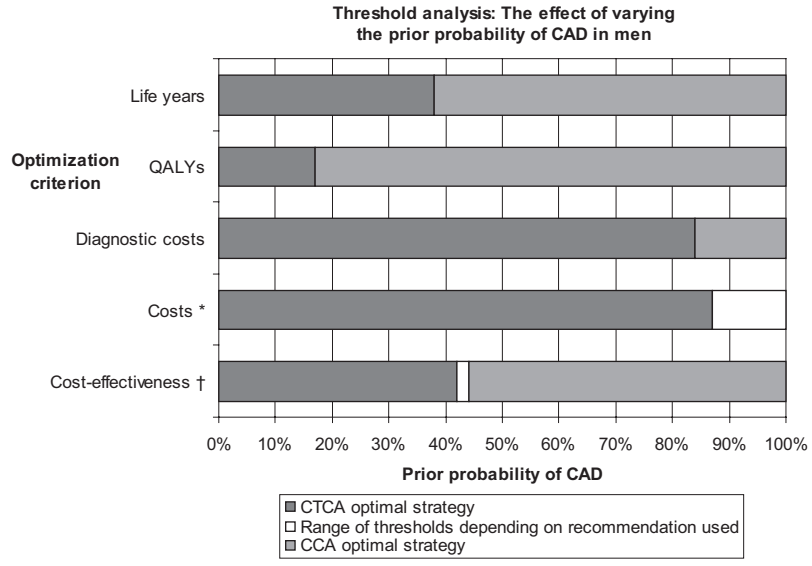
Societal perspective

By using the US recommendations for cost-effectiveness analysis, coronary CT angiography increases savings for health care and direct non-health care–related costs regardless of the prior probability (Figure 3). For men, there was a small gain in QALYs of 0.039, with a small increment in cost of €643, and an ICER of €16 509/QALY gained for conventional compared with coronary CT angiography. For women, there was a gain in QALYs of 0.038, a cost increment of €775, and a €20 360/QALY gained (Table). Performing conventional

Figure 3. Sensitivity analysis for prior probability of CAD in 60-year-old men (upper) and women (lower). Threshold level for prior probability of CAD below which coronary CT angiography (CCTA, followed by catheter-based coronary angiography [CAG] in case of positive results) is preferred to CAG only, from perspectives of patient, physician, hospital, health care system, and society. Above this threshold level, CAG optimizes criterion used.

(a) For 60-year-old men, cost-effectiveness analysis was performed by using recommendations from UK (health care perspective), US (societal perspective), and Dutch (societal perspective) recommendations by using WTP threshold level of €80 000/QALY.

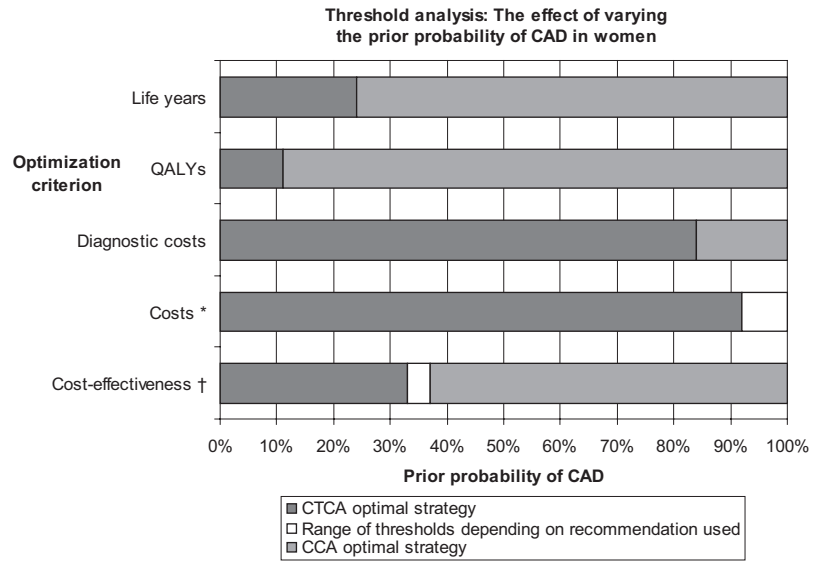
A



* Based on recommendations used, threshold level below which CCTA is optimal strategy (for costs) varies from 87% to 100%.

† Given recommendations used (US, UK, NL), threshold level below which CCTA is cost-effective varies from 42% to 44%.

B



(b) For 60-year-old women, cost-effectiveness analysis was performed by using recommendations from UK (health care perspective), US (societal perspective), and Dutch (societal perspective) recommendations by using WTP threshold level of €80 000/QALY.

* Based on recommendations used, threshold level below which CCTA is optimal strategy (when optimizing costs) varies from 92% to 100%.

† Depending on recommendations used, threshold level below which CCTA is cost-effective varies from 33% to 37%.

coronary angiography without prior coronary CT angiography increased net health benefit compared with initial coronary CT angiography by 0.03 QALY equivalents (Table).

By using the Dutch recommendations for cost-effectiveness analysis, health care costs and direct non-health care–related costs, including production losses, were reduced to a prior disease probability of less than 87% in men and less than 92% in women for coronary CT angiography when compared with conventional coronary angiography (Figure 3). For men, there was a small gain in QALYs of 0.046, with a small increment in cost of €182, and an ICER of €4095/QALY gained for conventional compared with coronary CT angiography. For women, there was a QALY gain of 0.047, a cost increment of €485, and an ICER of €10 383/QALY (Table). Performing conventional coronary angiography increased net health benefit by 0.04 QALY equivalents when compared with initial coronary CT angiography (Table).

Sensitivity Analyses

The prior probability threshold levels were not sensitive to changes across plausible ranges of all parameter inputs, with one exception. Varying the sensitivity of coronary CT angiography (independently from specificity) from 80% to 100%, the prior probability threshold level below which coronary CT angiography maximizes QALYs for women ranged from 2% to 44% and cost-effectiveness (UK recommendations) was optimized at 8% to 72%. Varying specificity (independently of sensitivity) of coronary CT angiography and test costs had little effect (Figure 4). Varying the disutility incurred by a cardiovascular event and the fatality rate associated with a cardiovascular event did not alter the results.

For probabilistic sensitivity analyses, the probability that coronary CT angiography is cost-effective when compared with conventional coronary angiography was 2% in men and 13% in women for a threshold level WTP of €80 000/QALY (Figure 5). Value of information analysis showed an EVPI for further research of €3 per man and €46 per woman, which, for the European Union population (500 million, annual incidence 4.5 per 1000) over a period of 5 years (discounted at 3.5%) amounts to approximately €0.38 billion, and for the US population (300 million) over a period of 5 years (discounted at 3%) amounts to €0.23 billion. Partial EVPI calculations demonstrated that the expected value of information for women was mainly a result of uncertainty in the prior probability of CAD, radiation risk, and test characteristics (combined EVPI, €44). For both men and women, the uncertainty in the parameters related to long-term outcome, quality of life, and costs had a negligible partial EVPI.

Specific Scenario

The Table shows the results from the cost-effectiveness analyses for the specific scenario. Conventional coronary angiography was the preferred strategy in all cases, except for women analyzed according to the UK and US recommendations. The prior probabilities observed in our cohort study were consistent with the estimates derived from Diamond and Forrester (9) for all patients, except for women with typical angina. The prior probability of CAD for women in the cohort study with typical angina was 43%, whereas this probability would be estimated as 85%, according to Diamond and Forrester. However, this discrepancy does not alter the decision given our threshold levels for cost-effectiveness.

Discussion

In this study, we evaluated the comparative effectiveness, costs, and cost-effectiveness of coronary CT angiography performed as an initial test (followed by conventional coronary angiography if positive results were obtained) compared with conventional coronary angiography alone in patients with stable angina and functional test results suggestive of ischemia. Our results demonstrate that in the setting of a low to moderate prior probability of disease, coronary CT angiography with negative results virtually excludes CAD. In contrast, the revised probability of CAD after coronary CT angiography with positive results probability varies over a wide range, implying that a positive coronary CT angiogram needs to be confirmed with conventional coronary angiography. Although coronary CT angiography is less costly and less invasive compared with conventional coronary angiography, the radiation risk is higher by using both studies, and false-negative coronary CT angiography results can occur, in which case patients forego the benefit of treatment. We showed that coronary CT angiography can be a cost saving technique in that it helps avoid unnecessary angiograms but comes with the disadvantage of a slight decrement in patient outcomes. Optimization criteria developed on the basis of QALYs favors the use of conventional coronary angiography because it will identify all patients with CAD and the potential long-term benefit of treating CAD outweighs the small risk involved. When considering costs only, coronary CT angiography is preferred because it is less expensive and can help avoid performing unnecessary angiograms in a substantial proportion of patients. When considering disadvantages, benefits, and costs together, our results suggest that the use of coronary CT angiography as an initial test is cost-effective below an average prior probability threshold level of 40%. Above this threshold level, conventional coronary angiography remains the

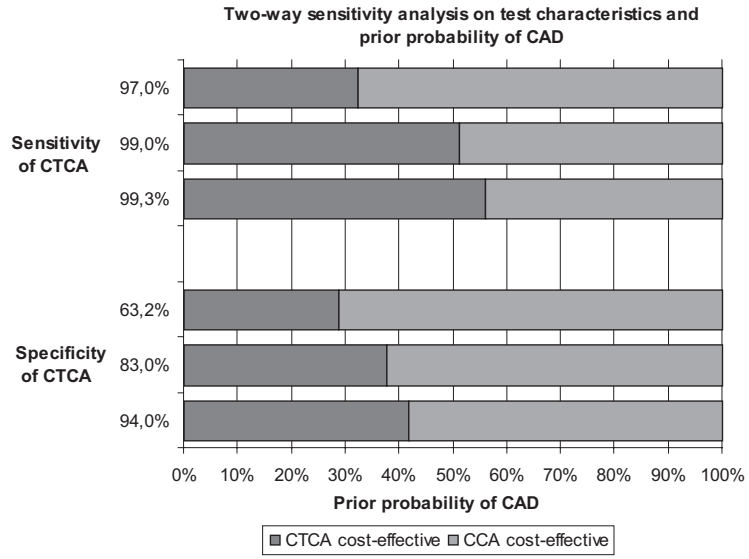
preferred strategy. The use of UK, US, and Dutch recommendations for cost-effectiveness analyses did not substantially influence the results.

It is important to note that a lower threshold level exists, one that is not addressed in our study. Below this threshold level, the net gain of performing coronary CT angiography is too small and it would therefore not be cost-effective when compared with either not testing or performing another less-invasive, less costly test. However, including additional strategies was beyond the scope of this paper.

In our cohort study, coronary CT angiography was a cost-effective strategy for women in the UK and US analyses. The difference in cost-effectiveness between the reference case analysis and ours is driven by the difference in prior probability (women had a lower prior probability in the study) and test characteristics.

In our cost-effectiveness analysis, we used a WTP threshold level of €80 000/QALY (39). Had we used a WTP threshold level of €50 000/ QALY, our conclusions would be the same because all ICERs for conventional coronary angiography were less than €50 000 in the reference case analysis. For our cohort study, the results only changed for women who were analyzed according to the Dutch recommendations for whom coronary CT angiography would be cost-effective at the €50 000 threshold level, whereas conventional coronary angiography would be cost-effective at the €80 000 threshold level.

Figure 4. Two-way sensitivity analyses for test characteristics and prior probability of CAD in women. Graph shows influence of varying test characteristics (independently from each other) on prior probability threshold level above which catheter-based coronary angiography (CAG) would be more cost-effective. Upper and lower bars are range in sensitivity and specificity, respectively, as observed in the Mowatt *et al.* (3) and Vanhoenacker *et al.* (1). Other bars are sensitivity and specificity as observed in cohort study (Appendix [http://radiology.rsna.org/ content/253/3/734/suppl/DC1]). Note that effect of varying sensitivity on prior probability threshold level is substantial. Varying specificity does not alter prior probability threshold level as much as does varying sensitivity. Analysis performed by using UK recommendations for cost-effectiveness analysis (health care perspective). Changes in probability threshold level were similar in men: by varying sensitivity from 97%–99.3%, threshold level for cost-effectiveness ranged from 39%–66%. Varying specificity resulted in threshold levels ranging from 40%–52%. CCTA = coronary CT angiography.



In our study, a stenosis of 50% or more in at least one vessel was considered as significant, whereas a stenosis of 70% or more is commonly considered as hemodynamically significant. Only hemodynamically significant stenoses are eligible for revascularization, which makes a threshold level of 70% or more relevant. However, our aim was to select patients that require some form of treatment. We modeled medication-based therapy, PCI, and CABG as treatment strategies, which is why we used a threshold level of a stenosis of 50% or more for significant CAD. One could argue that physicians are primarily interested in diagnosing severe CAD, as these patients would be eligible for revascularization, whereas others can be adequately treated by using medication alone.

Although the model uses a dichotomized definition of CAD, the model does allow for differences in treatment effects incurred by differences in disease location and severity. This was carried out by including weighted averages of treatment effects, quality-adjusted life estimates, and costs. Additionally, the uncertainty in such parameters was taken into account by using distributions.

Table. Cost-effectiveness Analysis of Conventional Coronary Angiography Only Compared with CT Followed by Conventional Coronary Angiography

Strategy	Cost (€)	Effectiveness (QALY)	ICER (€/QALY)	Incremental NHB of CAG vs. CCTA
Reference case analysis				
Men (prior probability of CAD: 72%)				
UK				
CCTA	31 506	11.578		
CAG	32 095	11.615	15 915	0.0296
US				
CCTA	34 154	12.180		
CAG	34 797	12.219	16 509	0.0309
NL				
CCTA	386 640	14.360		
CAG	386 822	14.406	4 095	0.0435
Women (prior probability of CAD: 65%)				
UK				
CCTA	26 020	13.263		
CAG	26 734	13.299	19 913	0.0269
US				
CCTA	28 307	14.050		
CAG	29 082	14.088	20 360	0.0284
NL				
CCTA	252 455	16.941		
CAG	252 940	16.988	10 383	0.0407
Cohort Study				
Men (prior probability of CAD: 72%)				
UK				
CCTA	30 377	11.620		
CAG	30 531	11.626	25 014	0.0042
US				
CCTA	32 843	12.228		
CAG	32 998	12.234	23 681	0.0046
NL				
CCTA	383 971	14.415		
CAG	384 063	14.423	11 413	0.0069
Women (prior probability of CAD: 43%)				
UK				
CCTA	21 007	13.409		
CAG	21 384	13.413	95 602	-0.0008
US				
CCTA	22 788	14.206		
CAG	23 183	14.210	87 804	-0.0004
NL				
CCTA	245 313	17.132		
CAG	245 691	17.139	56 117	0.0020

This analysis was performed according to recommendations in the UK, the United States, and the Netherlands for 60-year-old men and women. The results indicate that at a willingness-to-pay threshold level of €80 000/QALY, conventional coronary angiography is optimal in men and women. In our cohort study, coronary CT angiography is optimal in women only if UK or US recommendations are used. A strategy is dominated if another strategy is equally as effective or more effective and less costly. The incremental net health benefit (NHB) is calculated as $NHB_{CAG} – NHB_{CCTA}$ with NHB = QALY – (cost/€80 000) and expressed in units of QALY equivalents. CAG = catheter-based coronary angiography, CCTA = 64-detector coronary CT angiography.

The additional information that is provided by coronary CT angiography (e.g., assessment of plaque burden) could potentially improve the management of CAD patients. Currently however, too little evidence is available. Future studies should investigate the added value of assessing plaque burden and the effectiveness of decision-making on the basis of such findings.

Our analysis focused on a 64-section CT scanner, which implies that our results are only applicable to patients undergoing 64-section CT. However, with the rapid rate of advancement in technology, newer generations of CT scanners are expected to be more accurate in helping diagnose significant CAD, owing to a higher temporal and spatial resolution. In addition, new techniques are being developed to minimize radiation dose. Such improvements will increase the cost-effective application of coronary CT angiography.

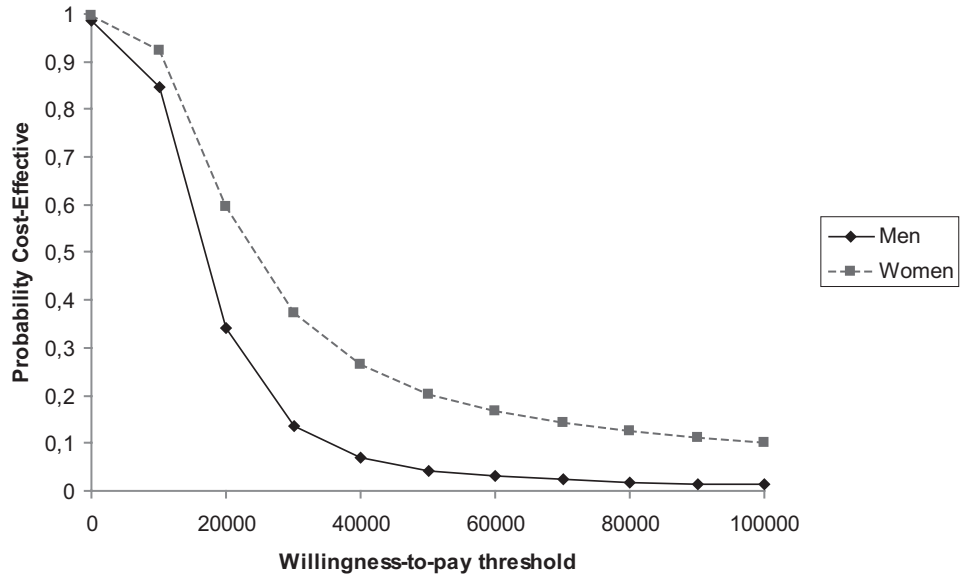
Our methods were different from a previously published cost-effectiveness analysis by Dewey and Hamm (43), who used costs per correctly identified CAD patient as a measure of cost-effectiveness, did not consider costs of subsequent treatment and ignored the benefit of correct exclusion of CAD. They found a threshold level of 60% prior probability of CAD below which coronary CT angiography is indicated. Kuntz *et al.* (44) examined the cost-effectiveness of several non-invasive functional (imaging and non-imaging) tests. When compared with exercise single photon emission CT and exercise echocardiography, conventional coronary angiography had an ICER of $32 600 and $35 200 respectively, for 50–59-year-old men with mild chest pain. This is slightly higher compared with what we found for conventional coronary angiography compared with coronary CT angiography for men by using the US recommendations.

More recently, Khare *et al.* (45) and Ladapo *et al.* (46) studied the cost-effectiveness of coronary CT angiography in low-risk patients with acute chest pain. Our analysis focused on patients with stable chest pain and much higher disease prevalence. Therefore, our analysis adds new information to the current knowledge about cost-effective applications of coronary CT angiography.

One limitation of our study was the use of meta-analyses for the diagnostic performance of coronary CT angiography that were published in 2007 and 2008. Alternative data sources for diagnostic performance, such as Miller *et al.* (47), recently studied the diagnostic accuracy of coronary CT angiography. Miller *et al.* reported a per-patient sensitivity and specificity for diagnosing significant (≥50% stenosis) CAD of 85% and 90%, respectively. They excluded patients with Agatston coronary calcium scores of 600 or higher, which makes their results relevant to a diagnostic strategy by using coronary CT calcium scoring as a triage test prior to performing coronary CT angiography. However, such a diagnostic strategy was not considered in our study. Although their study population was different from our

target population, the ranges of sensitivity and specificity we used also included the results reported by Miller *et al.*

Figure 5. Acceptability curves for coronary CT angiography (CCTA) strategy plots probability that strategy is cost-effective given particular WTP threshold level. We ran 100 000 Monte Carlo probabilistic simulations that used random draws from distributions that represent uncertainty around parameter estimates. For men and women, UK recommendations were used for cost-effectiveness analysis (health care perspective) and WTP threshold level of €80 000/QALY. In 2% and 13% of simulations for men and women, respectively, CCTA was cost-effective compared with conventional coronary angiography, which corresponds to 2% and 13% probability of cost-effectiveness.



Our aim was to design a lucid decision model that would be easy to interpret and that can help guide decision making. Consequently, we had to make several assumptions. First, parameter estimates were obtained from the literature by using the best available published evidence. Second, for the purpose of estimating costs and disutility of a cardiovascular event, we assumed that cardiovascular events were mainly myocardial infarctions and that costs and disutility of other cardiovascular events were similar to that of myocardial infarctions. Third, the quality- of-life estimates were derived from Hlatky *et al.* (23), which was a study on multi-vessel CAD. This may be an underestimate of the quality of life of patients with single-vessel disease, which therefore may have overestimated the gain in effectiveness with treatment, which, in turn, would have created bias in favor of conventional coronary angiography. Fourth, work-up for chest pain following a coronary CT angiogram with negative results was not modeled and may have created bias in favor of coronary CT angiography. Fifth, we assumed the sensitivity and diagnostic odds ratio to be independent of age, sex, risk factors, and presentation. Sixth, we did not consider other non-invasive tests but rather considered only patients referred for conventional coronary angiography for whom either the history or functional test results have suggested the presence of cardiac ischemia.

Furthermore, it is important to realize that all costs were based on European estimates. In the cost-effectiveness analysis, we were interested in evaluating how the different perspectives (health care system vs. societal) and the different (UK vs. US vs. Dutch) recommendations would affect the results and therefore chose to use the same cost estimates for these comparisons. Reported costs for CT and conventional coronary angiography in the United States range from $630–$3000 and $1750–$5176, respectively (45,46). In addition, a US health care cost database provides an estimate of $10 000 for performing conventional coronary angiography (48). Because the costs for conventional coronary angiography in the United States are relatively high compared with the costs for coronary CT angiography, coronary CT angiography might be a more cost-effective approach in the US. However, all costs in the US are generally higher, as is the WTP threshold level, which could lead to different results and merits further study.

Finally, it is important to note that the differences between the two strategies in terms of costs, QALYs, and net health benefits were rather small, which is why we performed extensive (probabilistic) sensitivity analysis and value-of-information analysis. We used modern techniques to evaluate whether further research is necessary and to inform the choice of a future study. Value-of-information analysis showed a rather high expected value of further research for both Europe and the United States and indicated that future research should focus on test characteristics, the risk of radiation, and prediction rules for the diagnosis of CAD.

In conclusion, the optimal diagnostic strategy depends on the optimization criterion, prior probability of CAD, and test characteristics. Analysis of our cost-effectiveness model suggests that coronary CT angiography performed as a triage test prior to conventional coronary angiography is cost-effective in men with a prior probability of CAD of less than 44% and in women with a prior probability of CAD of less than 37%. Above this threshold level, conventional coronary angiography remains the most cost-effective strategy. To maximize patient outcomes, a lower threshold level applies and to lower costs, a higher threshold level should be used.

References

1. Vanhoenacker PK, Heijenbrok-Kal MH, Van Heste R, et al. Diagnostic performance of multidetector CT angiography for assessment of coronary artery disease: meta-analysis. Radiology 2007;244:419–428.
2. Hamon M, Biondi-Zoccai GG, Malagutti P, 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;48:1896–1910.
3. Mowatt G, Cook JA, Hillis GS, et al. 64-slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. Heart 2008;94:1386–1393.
4. RIVM. Dutch National Institute for Public Health and the Environment. http://www.rivm.nl/vtv /object_ document/o3171n17964.html. Published June 15, 2006. Accessed March 2008.
5. Cook S, Walker A, Hugli O, Togni M, Meier B. Percutaneous coronary interventions in Europe: prevalence, numerical estimates, and projections based on data up to 2004. Clin Res Cardiol 2007;96:375–382.
6. Merrill CE, Elixhauser A. Procedures in U.S. hospitals. Rockville, Md: Agency for Healthcare Research and Quality http://www .ahrq.gov/data/hcup/factbk7/factbk7b.htm. Published 2005. Accessed March 2008 .
7. Fox K, Garcia MA, Ardissino D, et al. Guidelines on the management of stable angina pectoris: executive summary—the Task Force on the Management of Stable Angina Pectoris of the European Society of Cardiology. Eur Heart J 2006;27:1341–1381.
8. Hunink MG. Cost-effectiveness analysis: some clarifications. Radiology 2008;249: 753–755.
9. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med 1979; 300:1350–1358.
10. Meijboom WB, Meijs MF, Schuijf JD, et al. Diagnostic accuracy of 64-slice computed tomography coronary angiography: a prospective, multicenter, multivendor study. J Am Coll Cardiol 2008;52:2135–2144.
11. Katayama H, Yamaguchi K, Kozuka T, Takashima T, Seez P, Matsuura K. Adverse reactions to ionic and nonionic contrast media— a report from the Japanese Committee on the Safety of Contrast Media. Radiology 1990;175:621–628.
12. Scanlon PJ, Faxon DP, Audet AM, et al. ACC/AHA guidelines for coronary angiography: a report of the American College of Cardiology/ American Heart Association Task Force on practice guidelines (Committee on Coronary Angiography)—developed in collaboration with the Society for Cardiac Angiography and Interventions. J Am Coll Cardiol 1999; 33:1756–1824.
13. Noto TJ Jr, Johnson LW, Krone R, et al. Cardiac catheterization 1990: a report of the Registry of the Society for Cardiac Angiography and Interventions (SCA&I). Cathet Cardiovasc Diagn 1991;24:75–83.
14. CBS. Statline of Central Bureau of Statistics, The Netherlands. 2008.
15. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008;117:743–753.
16. Deckers JW, Goedhart DM, Boersma E, et al. Treatment benefit by perindopril in patients with stable coronary artery disease at different levels of risk. Eur Heart J 2006; 27:796–801.
17. Berger JS, Brown DL, Becker RC. Low-dose aspirin in patients with stable cardiovascular disease: a meta-analysis. Am J Med 2008; 121:43–49.
18. Dagenais GR, Pogue J, Fox K, Simoons ML, Yusuf S. Angiotensin-converting-enzyme inhibitors in stable vascular disease without left ventricular systolic dysfunction or heart failure: a combined analysis of three trials. Lancet 2006;368:581–588.
19. Schulz H, Sinn R, Wolf R. Cardiac risk in men with angiographically normal coronary arteries or minimal coronary arteriosclerosis [in German]. Z Kardiol 2003;92:245– 253.
20. Adabag AS, Therneau TM, Gersh BJ, Weston SA, Roger VL. Sudden death after myocardial infarction. JAMA 2008;300: 2022–2029.
21. Einstein AJ, Henzlova MJ, Rajagopalan S. Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. JAMA 2007;298:317–323.

22. Board on Radiation Effects Research. Health risks from exposure to low levels of ionizing radiation. In: BEIR VII phase 2. Washington, DC: The National Academies Press, 2006.
23. Hlatky MA, Boothroyd DB, Melsop KA, et al. Medical costs and quality of life 10 to 12 years after randomization to angioplasty or bypass surgery for multivessel coronary artery disease. Circulation 2004;110:1960– 1966.
24. Sullivan PW, Ghushchyan V. Preference- Based EQ-5D index scores for chronic conditions in the United States. Med Decis Making 2006;26:410–420.
25. Van Loenen A. Pharmacotherapeutic compass. Dutch Health Care Insurance Board. http://fk.cvz.nl. Accessed March 2008.
26. Oostenbrink J, Bouwmans CAM, Koopmanschap MA, Rutten FFH. Dutch manual for cost-analyses [in Dutch]. Amstelveen, the Netherlands: Dutch Health Care Insurance Board, 2004.
27. NZA. Dutch National Authority of Health Care. http://ctg.bit-ic.nl/Nzatarieven/top.do. Published 2008. Accessed March 2008.
28. Bravata DM, Gienger AL, McDonald KM, et al. Systematic review: the comparative effectiveness of percutaneous coronary interventions and coronary artery bypass graft surgery. Ann Intern Med 2007;147:703–716.
29. Katritsis DG, Ioannidis JP. Percutaneous coronary intervention versus conservative therapy in nonacute coronary artery disease: a meta-analysis. Circulation 2005; 111:2906 –2912.
30. Einstein AJ, Moser KW, Thompson RC, Cerqueira MD, Henzlova MJ. Radiation dose to patients from cardiac diagnostic imaging. Circulation 2007;116:1290 –1305.
31. Brouwer WB, Niessen LW, Postma MJ, Rutten FF. Need for differential discounting of costs and health effects in cost effectiveness analyses. BMJ 2005;331:446–448.
32. National Institute for Clinical Excellence (NICE). Guide to the methods of technology appraisal. Oxford, England: Radeliff Medical Press, 2004.
33. Gold MR, Siegel JE, Russell LB, et al. Costeffectiveness in health and medicine. J Ment Health Policy Econ 1996;2:91–92.
34. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses: panel on cost-effectiveness in health and medicine. JAMA 1996; 276:1339–1341.
35. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. recommendations of the panel on cost-effectiveness in health and medicine. JAMA 1996;276:1253–1258.
36. Lenzen MJ, Boersma E, Bertrand ME, et al. Management and outcome of patients with established coronary artery disease: the Euro Heart Survey on coronary revascularization. Eur Heart J 2005;26:1169–1179.
37. Hunink MGM, Glasziou Paul P. Decision making in health and medicine: integrating evidence and values: Cambridge University Press, 2006.
38. Russell MW, Huse DM, Drowns S, Hamel EC, Hartz SC. Direct medical costs of coronary artery disease in the United States. Am J Cardiol 1998;81:1110–1115.
39. Dutch Council for Public Health and Health Care. Sensible and sustainable care. http: //www.rvz.net/ data/download/Engelse_ vertaling_samenvatting.doc. Published 2006. Accessed January 2009.
40. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). Lancet 2002;360:711–715.
41. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. Med Decis Making 2004; 24:207–227.
42. Groot Koerkamp B, Hunink MG, Stijnen T, Hammitt JK, Kuntz KM, Weinstein MC. Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. Med Decis Making 2007;27:101– 111.
43. Dewey M, Hamm B. Cost effectiveness of coronary angiography and calcium scoring using CT and stress MRI for diagnosis of coronary artery disease. Eur Radiol 2007;17: 1301–1309.

44. Kuntz KM, Fleischmann KE, Hunink MG, Douglas PS. Cost-effectiveness of diagnostic strategies for patients with chest pain. Ann Intern Med 1999;130:709–718.
45. Khare RK, Courtney DM, Powell ES, Venkatesh AK, Lee TA. Sixty-four-slice computed tomography of the coronary arteries: cost-effectiveness analysis of patients presenting to the emergency department with low-risk chest pain. Acad Emerg Med 2008; 15:623–632.
46. Ladapo JA, Hoffmann U, Bamberg F, et al. Cost-effectiveness of coronary MDCT in the triage of patients with acute chest pain. AJR Am J Roentgenol 2008;191:455–463.
47. Miller JM, Rochitte CE, Dewey M, et al. Diagnostic performance of coronary angiography by 64-row CT. N Engl J Med 2008; 359:2324–2336.
48. HCUPnet. Healthcare Cost and Utilization Project (HCUP). http://hcupnet.ahrq.gov/. Rockville, Md: Agency for Healthcare Research and Quality, 2006

Technical appendix

I. Model

We developed a decision model (in DATA Pro, 2008 Suite; TreeAge Software, Williamstown, Mass) to evaluate the use of coronary CT angiography as initial imaging test, if positive followed by conventional coronary angiography (new strategy) compared with direct conventional coronary angiography (current practice) (Figure 1). Short-term outcomes related to the diagnostic imaging tests were modeled with a decision tree and a Markov model (cycle length of 1 year) was used to model long-term outcomes. We modeled whether a patient was alive or dead and whether a cardiovascular event occurred.

In our model, the variable "Male" coded the sex of the patients. Male = 1 refers to males, Male = 0 refers to females.

age = 60
ageMKV = age + _stage

II. Input data

See Table E1.

III. Data Sources

The input data were based on the best available evidence in the literature. Event rates were based on average risk factor profiles as observed in our cohort study (Table E3) (10). In this study, 233 patients were included who had new-onset chest pain and were referred for conventional coronary angiography. All patients underwent coronary CT angiography prior to conventional coronary angiography. A specific scenario was analyzed to reflect the decision on the basis of prior probabilities and test characteristics, as observed in this cohort study (Table E2).

Table E2. Prior Probabilities of CAD according to Age, Sex, and Chest Pain Type

	Chest pain in men (n=156)		Chest pain in women (n=77)	
Parameter	**Typical (n=104)**	**Atypical (n=52)**	**Typical (n=47)**	**Atypical (n=30)**
Mean age (years)	59	60	61	61
Prior probability of CAD (%)	82	54	43	43

Prior probabilities determined given the cohort study (10).

IV. Assumptions

Prior probability of CAD

The prior probability was entered in the model for men and women separately by using beta distributions. The □ and □ parameters of the beta distributions were estimated in such a way that the 95% confidence interval covered the range of probabilities as reported by Diamond and Forrester (9). The alternate scenario used the prior probabilities for men and women to have CAD, as observed in the cohort study (10).

Table E1. Parameter Estimates, Distributions and Reported Ranges, and Data Sources

Parameter	Point estimate	Distribution	95% CI / Range	Source
Prevalence of CAD				
Men	0.79	Beta	0.59 – 0.94	(9)
Women	0.65	Beta	0.32 – 0.91	(9)
CCTA				
Sensitivity * (patient-level)	0.98	-	-	-
Specificity (patient-level)	0.81	Beta	0.64 – 0.93	(1, 3, 10)
Natural log of the diagnostic odds ratio	5.00	Log normal	3.89 – 6.11	(1)
Mortality	0.000006	Beta	0 – 0.000016	(11)
Costs in Euro's	198	Gamma	128 – 283	Cost-analysis
Time cost for patients (min)	90	Gamma	58 – 129	Cost-analysis
CAG				
Sensitivity	1	-	-	Assumption
Specificity	1	-	-	Assumption
Risk of MI	0.0005	Beta	0.0003 – 0.0007	(12)
Mortality	0.0011	Beta	0.0009 – 0.0014	(13)
Costs in Euro's	1360	Gamma	880 – 1942	Cost-analysis
Time cost for patients (min)	362	Gamma	243 – 516	Cost-analysis
Event rates and hazard rate ratios				
Annual non-CVD mortality † ‡	-	-	-	Dutch Life tables (14)
CVD event rate (no CAD)	0.024	Uniform	0.014 – 0.077	(15)
Men	0.014	Uniform	0.008 – 0.053	(15)
Women				
HRR CVD events, CAD vs. no CAD	1.50	Log-normal	1.32 – 1.70	(16)
HRR CVD events, Meds vs. no Meds	0.64	Log-normal	0.45 – 0.88	(17, 18)
HRR CVD events, prior MI vs. no prior MI	1.44	Log-normal	1.25 – 1.66	(16)
Proportion Fatal Event	0.17	Beta	0.10 – 0.25	(18-20)
Risk of cancer due to radiation exposure				
LAR in men	0.0008	Beta	0.0005 – 0.0011	(21)
Proportion fatal	0.65	-	-	(22)
LAR in women	0.0021	Beta	0.0014 – 0.0029	(21)
Proportion fatal	0.70	-	-	(22)
Quality of life weights				
no CAD	0.87	Beta	0.85 – 0.89	(23)
Treated CAD (TP) §	0.85	Beta	0.82 – 0.87	(23)
Untreated CAD (FN)	0.74	Beta	0.71 – 0.77	(23)
Disutility CVD event	0.04	Beta	0.02 – 0.07	(24)
Mean costs ||				
Treatment	7512	Gamma	4858 – 10730	Standard Rates
Medication (per year)	364	Gamma	236 – 520	(25)
Myocardial infarction	13000	Gamma	8407 – 18582	Estimation
Travelling to hospital (per visit)	5	Gamma	3 – 6	(26)
Time cost (per hour)	18	Gamma	11 – 25	(26)
Reintervention ||	754	Gamma	488-1076	(27-29)
Friction period (weeks)	22	-	-	(26)
Friction cost (per hour)				
Men	51	Gamma	33 - 72	(26)
Women	39	Gamma	25 – 55	(26)

CVD = cardiovascular disease, LAR = lifetime attributable risk, MI = myocardial infarction.

* *Sensitivity modeled as a function of specificity and the diagnostic odds ratio to account for the inverse relationship of sensitivity and specificity.*

† *Age- and sex-specific.*

‡ *Treatment breakdown for patient population: 51% PCI, 25% CABG, 24% medication.*

§ *All costs converted to 2007-rate euro's (€1.00 equivalent to US $1.37); gamma distributions were approximated by assuming a standard deviation of 20% of the mean costs.*

|| *Treatment breakdown for patient population: 70% PCI, 30% CABG; average yearly costs.*

Modeling the Joint Distribution of Sensitivity and Specificity

The sensitivity and specificity of a diagnostic test are inversely related. To take into account the correlation between sensitivity and specificity, sensitivity was defined as a function of the specificity and the diagnostic odds ratio (DOR).

$$DOR = \frac{TP}{FN} \cdot \frac{TN}{FP}$$

$$DOR = \frac{sens}{1-sens} \cdot \frac{spec}{1-spec}$$

$$\frac{1-sens}{sens} = \frac{spec}{(1-spec)\cdot DOR}$$

$$1-sens = \frac{spec}{spec + (1-spec)\cdot DOR}$$

$$sens = 1 - \frac{spec}{spec + (1-spec)\cdot DOR}$$

where TP = true positive, FN = false negative, TN = true negative, and FP = false positive.

The range of specificities reported in recently published literature (1,3,10) was used to estimate a distribution for specificity. The diagnostic odds ratio was obtained from a meta-analysis (1) and modeled by using a lognormal distribution. In the alternate scenario, the specificity and diagnostic odds ratio from the cohort study were used.

Annual Probability of Dying

Non-cardiovascular disease–related mortality was calculated by using data from the Dutch Bureau of Statistics (14). By using the absolute numbers of deaths from all causes and from cardiovascular causes in 2006, we calculated the proportion of non-cardiovascular deaths for 10-year age categories for men and women separately. We obtained age- and sex-specific mortality rates of the general population from Dutch life table data and multiplied those numbers by the corresponding proportion of non-cardiovascular deaths.

Table E3. Risk Factor Profiles Used to Estimate Cardiovascular Event Rates

	Risk factors in men			Risk factors in women		
Parameter	Low	Mean	High	Low	Mean	High
Age (y)	59	60	61	60	61	63
Cholesterol (mg/dL) *	180	191	201	183	200	220
High-density lipoproteins (mg/dL) *	56	52	47	63.8	57	50
Systolic blood pressure (mmHg)	144	148	151	143	149	155
Smoking (%)	0	0.32	1	0	0.21	1
Diabetes (%)	0	0.20	1	0	0.19	1

Profiles determined given the cohort study (10).

* *For Systéme International units (millimoles per liter), multiply data by 0.02586.*

Cardiovascular event rates (coronary death, myocardial infarction, coronary insufficiency, angina, stroke, cardiac arrest, peripheral arterial disease, heart failure) in patients without known CAD (true negatives and false positives) were assumed to be equivalent to first CAD event rates in the general population and were therefore calculated with the Framingham Heart Study (15) and modeled specific for the average risk factor profiles of patients in the cohort study. Table E3 summarizes the average risk factor profiles as observed in the cohort study. Given the average values of risk factors, we calculated 10-year cardiovascular event rates, which were transformed to annual cardiovascular event rates. Low- and high-risk factor profiles were used to create a range of event rates. The Framingham Heart Study model was based on the Framingham Heart Study, with 3969 men and 4522 women who had a mean age of 49 years (15). We estimated that patients presenting with symptomatic CAD who are subsequently treated (true-positives) will have a 1.5-fold (95% confidence interval: 1.3-1.7) cardiovascular event rate compared with patients without symptomatic CAD based on a multivariable Cox model reported from the EUROPA study (16). Following a first cardiovascular disease event, the cardiovascular disease reoccurrence rate will be a 1.44-fold (95% confidence interval: 1.25-1.66) higher risk compared with that in a patient with CAD and no previous cardiovascular disease event (16). We assumed that the cardiovascular event rate is reduced by treatment (hazard rate ratio, 0.63; range, 0.44–0.88) (17,18) modeled with a combined weighted average effectiveness of CABG, PCI, and medical treatment (28,29). Missed CAD patients (false-negatives) forgo the benefit of treatment, implying reduced quality of life and a higher cardiovascular event rate (Table E1) (17,18). Following a cardiovascular event during follow-up in patients with an initial negative test result, CAD will be diagnosed and treated.

Radiation Risk (Tables E4, E5)

Radiation risk of coronary CT angiography was modeled with age- and sex-specific lifetime attributable risks (21). We assumed that the first radiation-induced fatal cancer would only occur 12 years after exposure. Therefore, the lifetime attributable risk was divided by the probability of surviving 12 years beyond exposure, which results in the lifetime probability of having a radiation-induced cancer, conditional on the fact that the individual survived up to the moment that such a cancer could occur. Next, the lifetime attributable risks were converted to 1-year probabilities by dividing them by the remaining average life expectancy. To reflect mortality, the risk was multiplied by the fraction of fatal cancers. This fraction was calculated by using the radiation-induced incidence and mortality data from de BEIR VII report (22). No adjustment for reduced quality of life after the diagnosis of a radiation-induced cancer was made.

Table E4. Lifetime Attributable Risk of Cancer in 60-year-old Patient Due to a CT Angiogram

	Point estimate	Range
Men	0.000797	0.00052 – 0.00107
Women	0.00215	0.00141 – 0.00288

Table E5. Proportions of Radiation-induced Cancers Resulting in Death

Age at exposure	Proportion of fatal cancers (men)	Proportion of fatal cancers (women)
20	0.523029683	0.462940462
30	0.555393586	0.508920188
40	0.581790123	0.572234763
50	0.609137056	0.633783784
60	0.652351738	0.697952218
70	0.728862974	0.775061125
80	0.879310345	0.887850467

For missing values, linear interpolation was used by Data TreeAge Pro.

The risk of cancer owing to conventional coronary angiography was assumed to be one-half of the risk of cancer owing to coronary CT angiography, considering the fact that the radiation dose is approximately one-half the dose of coronary CT angiography (30). The successive use of conventional coronary angiography after coronary CT angiography was assumed to have an additive effect on the risk of radiation-induced cancer.

pMortRadiation_CCTA =Male*(Dist_LAR_Male/0.790/12.3)*Prop_CancerMortality[age;1]
+(1–Male)*(Dist_LAR_Female/0.858/14.6)*Prop_CancerMortality[age;2]
pMortRadiation_CAG = 0.5*pMortRadiation_CCTA

V. Perspectives

The decision was analyzed from the perspective of the physician, patient, hospital, health care system, and society by using three recommendations for cost-effectiveness analysis (Table E6).

Table E6. Cost-effectiveness Analyses

Recommendation	Perspective (outcomes included)	Discount rate for costs	Discount rate for effectiveness
UK	Health care (QALYs, health care costs)	0.035	0.035
US	Society (QALYs, health care costs, patient time and travel costs)	0.030	0.030
NL (Dutch)	Society (QALYs, health care costs, patient time, travel and friction costs)	0.040	0.015

Items in parentheses are outcomes for each perspective.

In an analysis from the hospital perspective, we calculated the diagnostic costs and determined the prior probability threshold level below which coronary CT angiography would be cost saving. The analysis from the health care perspective considered QALYs and health care costs and was performed according to U.K. recommendations, discounting both future costs and effectiveness at 3.5% (31,32). Subsequently, we performed the cost-effectiveness analysis from the societal perspective according to US recommendations, which considered

QALYs, health care costs, and direct non-health care–related costs (patient time and travel costs) and discounted both future costs and effectiveness at 3% (33–35). Finally, an analysis from the societal perspective was performed according to Dutch recommendations, which, in addition to the above, also took productivity losses (friction costs) into account and discounted future costs and effectiveness at 4% and 1.5%, respectively (26).

Friction Costs

Friction costs are defined as costs of productivity losses owing to death or absence from work. The friction period (e.g., the time for replacing an employee after death) has been established at 22 weeks (26). The friction costs per hour (cFriction_hour) were entered in the model for 60-year-old men and women separately. The friction costs for surviving patients were calculated by multiplying the duration of absence from work by sex-specific friction costs per hour for diagnosis and treatment separately (26). We assumed PCI and CABG would result in 3- and 8-day admissions, respectively. A myocardial infarction was assumed to cost 14 days of productivity. The duration of diagnostic procedures was estimated in our cost analysis. According to the standard 1540 workable hours per year, we calculated the loss of workable hours.

TimePCI = 3/365*1540
TimeCABG = 8/365*1540
TimeMI = 14/365*1540
TimeCCTA = Dist_cTime_CCTA
TimeCAG = Dist_cTime_CAG
ProportionCABG = 0.25
ProportionPCI = 0.51
ProportionMedTx = 0.24
cFriction_hour = Male*Dist_cFriction_Men+(1–Male)*Dist_cFriction_Women
cFriction_dead = cFriction_hour*(22/52*1540)
cFrictionTreatment = proportionPCI*cFriction_hour*TimePCI+proportionCABG*Friction_hour*TimeCABG+proportionMedTx*cFriction_hour*0
cFrictionDiagnosis = cFrictionDiagnosis+cFriction_hour*Time_CAG
cFrictionDiagnosis = cFrictionDiagnosis+cFriction_hour*Time_CCTA
cFrictionMI = Male*Dist_cFriction_Men*TimeMI+(1–Male)*Dist_cFriction_Women*TimeMI

Patient Time Costs

Patient time costs consisted of transportation costs to the hospital and time costs of being absent from work (cPatient_Treatment and cPatient_Diagnosis) multiplied by the average income. The average income (cTime_patient_hour) assuming 67% men and, on average, 60-year-old patients (10), was calculated according to data from the Central Bureau of Statistics. Travel costs were based on standard costs of a hospital visit by car from the Dutch costing manual:

cTravel_per_test = Dist_cTravel
cTime_patient_hour = Dist_cTime_patient_hour
cPatient_Diagnosis = cPatient_Diagnosis+cTime_patient_hour*Time_CAG+cTravel_per_test
cPatient_Treatment = cPatient_Treatment+(0.51*cTime_patient_hour*3*24+0.25*cTime_patient_hour*8*24+0.24*cFriction_hour*0)

VI. Cost Estimates

Costs were estimated for diagnostic tests, treatment (PCI, CABG, medication), and events during follow-up. Costs for CT and conventional coronary angiography were determined with a cost analysis and included direct health care costs (personnel, materials, equipment), indirect health care costs (housing, overhead), direct non-health care–related costs (patient travel and time costs), and indirect non-health care–related costs (production losses) (Table E1).

Because our decision model was developed to analyze cohorts of patients, the cost estimates in our model represent mean costs. According to the central limit theorem, the distribution of the means is normal. However, as normal distributions are inappropriate for modeling costs, we used gamma distributions, which are commonly used to reflect cost distributions. We assumed a relatively small standard error of 20%*mean cost for all gamma distributions, which results in gamma distributions that are similar to normal distributions.

Costs for CABG and PCI were based on estimates from the National Health Care Authority (Table E1) (27). Low and high values of the distributions were based on the different costs of different types of CABG or PCI procedures; for example, a one-vessel procedure as a lower boundary and a two-vessel procedure as an upper boundary. Annual costs for medical therapy were based on treatment with aspirin, nitrates, statins, and angiotensin-converting enzyme inhibitors and included one follow-up visit per year (Table E1) (25). The mean value included aspirin, statins, and nitrates. The lower bound included aspirin only. The upper bound included aspirin, nitrates, statins, and ACE-inhibitors.

A weighted average for the cost of treatment was calculated and entered in the model. Costs for reinterventions were taken into account using weighted averages of reintervention rates for the different treatment options. Reintervention rates were derived using by 5-year freedom-of-intervention rates (28). Reinterventions were assumed to be 70% PCI and 30% CABG.

cReintervention = cPCI*0.70+cCABG*0.30
Probability of reintervention per year =
Proportion free of reintervention at 5 years after PCI = propFOR_PCI
Proportion free of reintervention at 5 years after CABG = propFOR_CABG
Proportion free of reintervention at 5 years after MedTx = propFOR_MedTx
ReinterventionRatePCI = – (1/5)*LN (propFOR_PCI)
ReinterventionRateCABG = – (1/5)*LN (propFOR_CABG)
ReinterventionRateMedTx = – (1/5)*LN (propFOR_MedTx)
pReintervention = 1–exp (ReinterventionRatePCI*proportionCABG+ReinterventionRateCABG*proportionCABG+ReinterventionRateMedTx*proportionMedTx)
cReinterventions = pReintervention*cReintervention

All costs were converted to year 2007 rates (€1.00 was equivalent to US $1.37) based on Dutch consumer price indices and reported in euro's (14).

VII. Quality of Life Estimates

Quality of life was modeled on the chance of successful relief of angina through treatment. Quality of life weights following treatment were a pooled weighted average taking into account that 51% of patients with diagnosed symptomatic CAD undergo PCI, 25% undergo CABG, and 24% will receive medication only (36); 5 years after CABG 15% and after PCI 16% of patients still have angina, and the quality of life weight for angina is 0.74 (range, 0.71–0.77) and without angina 0.87 (range, 0.86–0.88) (23). Patients in whom the diagnosis was missed (false-negatives) were assumed to all have angina during follow-up until a cardiovascular event occurred, after which they would be diagnosed and treated. A disutility of 0.04 (range, 0.02–0.07) was modeled for a cardiovascular event (myocardial infarction) during follow-up (24).

VIII. Outcome Measures

The net health benefit (NHB) was calculated by using the following formula: NHB = E – C/G, where E = effectiveness in terms of QALYs, C = total cost of the strategy concerned, and G = threshold WTP. Net health benefits are measures of effectiveness, adjusted for the costs and taking into account the WTP per QALY (37). Given a particular WTP, the strategy with the highest net health benefit is the most cost-effective.

The WTP reflects the amount of money that society is willing to pay for one quality-adjusted life year. A threshold level WTP of €80 000/QALY was used.

IX. Analyses

To assess uncertainty, probabilistic sensitivity analysis was performed drawing from all variable distributions (Table E1) by using Monte Carlo simulation of 100 000 samples. Random draws from variable distributions were single draws, which means they were equal for both strategies.

Chapter 10

Coronary Computed Tomography Versus Exercise Testing in Patients with Stable Chest Pain: Comparative Effectiveness and Costs

International Journal of Cardiology (In press)

Tessa S. S. Genders
Bart S. Ferket
Admir Dedic
Tjebbe W. Galema
Nico R. Mollet
Pim J. de Feyter
Kirsten E. Fleischmann
Koen Nieman
M. G. Myriam Hunink

Abstract

Background

To determine the comparative effectiveness and costs of a CT-strategy and a stresselectrocardiography- based strategy (standard-of-care; SOC-strategy) for diagnosing coronary artery disease (CAD).

Methods

A decision analysis was performed based on a well-documented prospective cohort of 471 outpatients with stable chest pain with follow-up combined with best-available evidence from the literature. Outcomes were correct classification of patients as CAD–(no obstructive CAD), CAD+(obstructive CAD without revascularization) and indication for Revascularization(using a combination reference standard), diagnostic costs, lifetime health care costs, and quality-adjusted life years (QALY). Parameter uncertainty was analyzed using probabilistic sensitivity analysis.

Results

For men (and women), diagnostic cost savings were €245 (€252) for the CT-strategy as compared to the SOC-strategy. The CT-strategy classified 82% (88%) of simulated men (women) in the appropriate disease category, whereas 83% (85%) were correctly classified by the SOC-strategy. The long-term cost-effectiveness analysis showed that the SOC-strategy was dominated by the CT-strategy, which was less expensive (-€229 in men,-€444 in women) and more effective (+0.002 QALY in men,+0.005 in women). The CT-strategy was cost-saving (-€231) but also less effective compared to SOC (-0.003 QALY) in men with a pre-test probability of ≥70%. The CT-strategy was cost-effective in 100% of simulations, except for men with a pre-test probability ≥70% in which case it was 59%.

Conclusions

The results suggest that a CT-based strategy is less expensive and equally effective compared to SOC in all women and in men with a pre-test probability <70%.

Introduction

The current guideline still recommends stress electrocardiography(X-ECG) as first line diagnostic test for patients with stable chest pain (1). However, the diagnostic accuracy of X-ECG is limited (2).

Coronary CT angiography (CCTA) is an alternative modality for diagnosing coronary artery disease (CAD). Its diagnostic accuracy compared to catheter-based coronary angiography (CAG) in highly selected patients has been studied extensively (3-7), demonstrating that CCTA is reliable in ruling out CAD (sensitivity 95-100%). Furthermore, previously published decision analyses indicate that CCTA as triage test in patients referred for CAG is cost-effective in patients with a low-intermediate probability of disease (8-10).

Recently, results from a Dutch outpatient chest pain clinic were published (11-12). Patients with stable chest pain were evaluated by X-ECG, CT calcium scoring, and CCTA. Results suggested that CT calcium scoring, selectively followed by CCTA could replace X-ECG as first line diagnostic test. However, long-term effectiveness and costs of CCTA compared to standard-of-care (SOC) in outpatients presenting with chest pain remain unclear.

Ideally, a large randomized controlled trial (RCT) comparing a CT-based strategy to SOC should be performed to evaluate comparative effectiveness and costs. Exploration of diagnostic strategies and preliminary estimates of outcomes can help design such a trial and can justify the investment of research resources. Furthermore, trial results will only be available after several years and in the meantime diagnostic testing decisions have to be made. A decision-analytic approach summarizing the evidence can be helpful in such situations.

Aim of this study was to determine the comparative effectiveness and costs of a hypothetical CT-strategy compared with SOC using a decision-analytic approach combining data from a well-documented prospective patient cohort with the best-available evidence from the literature.

Materials & Methods

Patient population

The model was based on a prospective clinical cohort (11) of 471 patients who presented with stable chest pain and no history of CAD. All patients were scheduled for X-ECG and CCTA (Table 1). During a mean follow-up (complete in 90%) of 2.6 years, 44 major adverse cardiovascular events (MACE: cardiac death, myocardial infarction, unstable angina requiring hospitalization, and revascularization) occurred in 30 patients (13) (Table A1). The study complied with the Declaration of Helsinki and the ethical committee at our institution approved the study. Informed consent was obtained from all patients.

Decision model

We developed a decision model (in DATA Pro 2009 Suite, TreeAge Software Inc, Williamstown, MA, USA) to evaluate the comparative effectiveness and costs of a hypothetical CT-based strategy compared to an X-ECG-based strategy (reflecting standard-of-care; SOC-strategy). Short-term diagnostic outcomes were modeled with a decision tree (Figure 1, 2).

Long-term prognosis (lifetime) was modeled using a Markov-Model (Figure A1). Model parameters were based on the clinical cohort with follow-up combined with best-available evidence from the literature (Table A2, A3). See supplementary material for a detailed model description.

Model probabilities for diagnostic test results were based on the clinical cohort and conditional on the "underlying truth", sex, and the pre-test probability. To model the "underlying truth", a disease category was assigned to all patients in the cohort: No obstructive CAD(CAD–), Obstructive CAD(CAD+) or Revascularization(Revasc) (Figure 3), which represents the true disease status at baseline. This was based on CAG(if performed), the treatment initiated and CCTA otherwise and included 6-month follow-up information. For example, if a patient was initially treated with medication only, but electively revascularized within 6 months, the patient was labeled as Revascularization. The modified reference standard was used in all analyses.

The diagnostic model classifies patients in one of the disease categories. Classification is correct if the classified category matches the underlying truth, and incorrect when the classified category does not match the underlying truth. Underlined categories refer to the underlying 'true' disease category, whereas *italic* categories refer to the disease category as

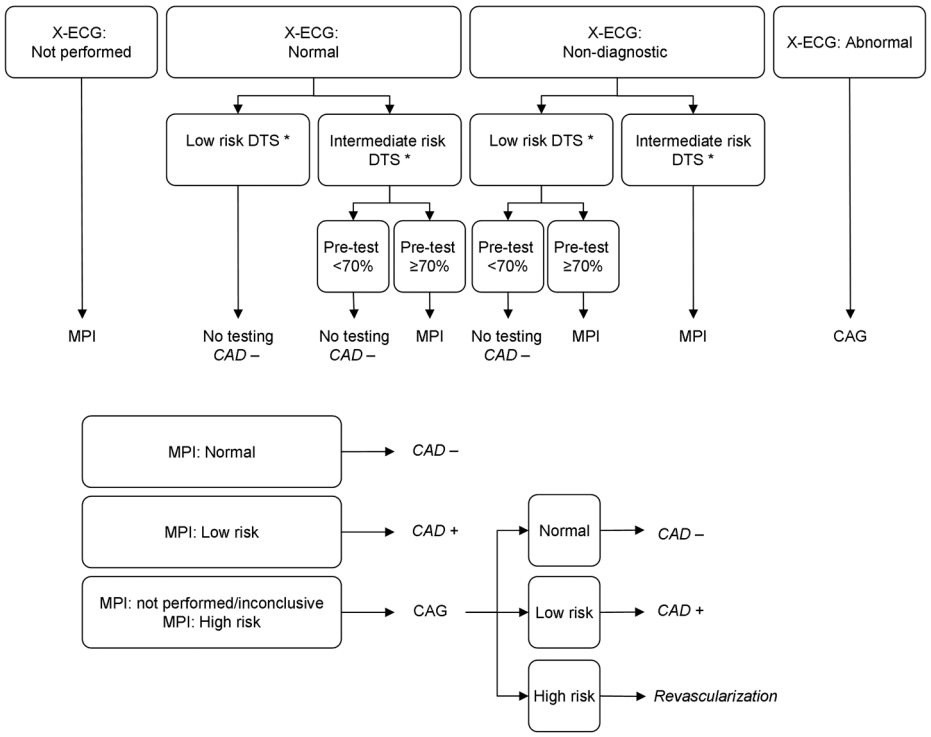


Figure 1. Further testing algorithm after X-ECG. Italic disease categories refers to classified category according to the diagnostic work-up.*Low risk DTS = DTS ≥5, intermediate risk DTS = DTS -10 to 5, high risk DTS = DTS ≤-11. DTS: Duke Treadmill Score.

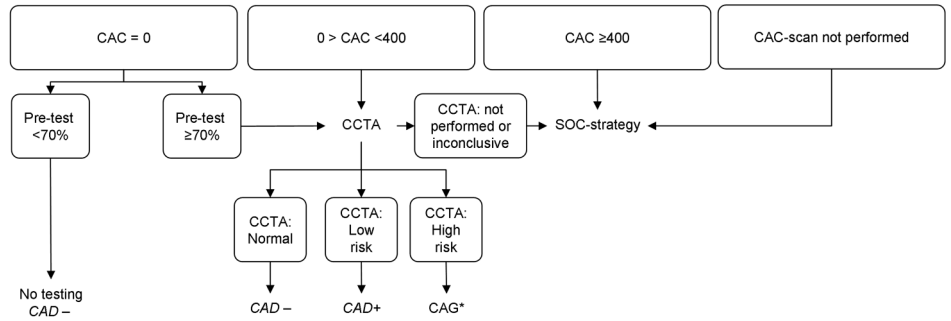


Figure 2. Further testing algorithm after CAC. Italic disease categories reflect the classified category according to the diagnostic work-up.

classified by the diagnostic work-up. Individuals classified as *CAD–* by the diagnostic strategy, who are CAD+ or Revascularization according to their underlying truth are "under-classified". Patients classified as *CAD+* who are Revascularization according to the underlying truth are "under-classified". Individuals classified as *CAD+*, who are CAD– according to the underlying truth, are "over-classified". The next paragraph explains how patients are classified by the diagnostic work-up.

Short-term decision tree

The SOC-strategy consists of initial evaluation with X-ECG according to the guideline (1) (Figure 1). Non-diagnostic X-ECGs are common (~25% (33))– which warrants further testing with pharmacological stress myocardial perfusion imaging (MPI) using single photon emission CT (SPECT). Patients unable to exercise are evaluated by MPI. We assume that a CAG classifies patients in the correct category.

The CT-strategy starts with a coronary artery calcification (CAC) scan in every patient and a CCTA in patients with a CAC>0 and <400 (Figure 2). Patients with CAC=0 and a pre-test probability <70% do not undergo CCTA, because obstructive CAD is unlikely to be present (34). This cutoff was chosen to capture the high-risk patients with typical presentation (14), which is consistent with clinical practice at our institution. Thus, a patient with zero calcium and a pre-test probability ≥70% will undergo CCTA (Figure 2). Based on evidence

CAD– (No obstructive CAD)	No obstructive CAD on CCTA or CAG within the first 6 months.
CAD+ (Obstructive CAD)	Obstructive CAD (≥50% diameter stenosis) determined in 78% by CCTA and 22% by CAG, without revascularization within the first 6 months. Patients are assumed to managed with medical therapy only.
Revascularization	Obstructive CAD treated with revascularization, either as initial treatment strategy, or performed within the first 6 months.

Figure 3. Definition of the disease category, as determined for each patient in the clinical cohort. Underlined refers to underlying "true" disease category, italic refers to classified category according to diagnostic test strategy.

that revascularization does not always improve survival beyond optimal medical treatment in patients with moderate disease (35), the CT-strategy consists of medical treatment for patients with moderate disease on CCTA and referral to CAG only if the CCTA shows severe CAD (left main-, three vessel-, or proximal left anterior descending artery disease).

Long-term Markov model

We used a Cox proportional hazards model to estimate the sex-specific rates of MACE for CAD–, CAD+, and Revascularization patients in the clinical cohort. Prognosis after the diagnostic work-up in the model depended on the correct vs. incorrect classification. Correctly classified individuals in the model were assigned the adjusted event rate as observed in the cohort. Under-classified (and under-treated) individuals experienced a higher event rate because of the forgone benefit of treatment (hazard rate ratio (HRR) based on the combined effectiveness of statins (29) and aspirin (27)). Over-classification only occurs when a CAD- patient is classified as *CAD+* and we assumed that medical treatment does not alter the event rate in these patients. See supplementary material for more details.

To mimic clinical follow-up of patients with chest pain, we assumed that every under-classified patient will be diagnosed with the correct disease category **within** the first year. We assumed that those patients remain symptomatic prior to the correct diagnosis because they are under-treated for a short period. As in clinical practice, patients with persistent angina are re-evaluated by the cardiologist. This implies that our model assumes that the benefit in terms of better outcomes of a diagnostic strategy can only be obtained in the first year after the initial assessment. In contrast, individuals who are over-classified are assumed **not** to reclassify to the CAD– category, but to remain in *CAD+*. The negative implications of overestimating the severity of disease in a CAD– patient consists of extra costs for medication and a (slightly) lower quality-of-life.

We modeled the risk of dying from non-cardiac causes based on age- and sex-specific mortality rates from the Dutch Central Bureau for Statistics (25).

Costs

Costs were based on a previous cost analysis (8) and expert opinion (Table 1), expressed in 2009 euros. We used the health care perspective according to recommendations for cost-effectiveness analyses (36), and a willingness-to-pay threshold (WTP) of €80,000/QALY (37). Medication costs were obtained via a registry provided by the Dutch Health Care Insurance board (16). Medication use was based on self-reported cardiovascular disease-related medication at the time of the last patient contact during follow-up and assumed to be constant over time.

Quality-of-life

Age and sex-specific utilities of the general population (30) were used to model the quality of life for CAD– patients. For CAD+ and Revascularization patients, the mean reduction in quality-of-life as compared to the general population was assumed to be 5% and 10%, respectively. Furthermore, under-classification (and under-treatment) was assumed to result in symptoms of angina due to the forgone benefit of anti-ischemic therapy. Based on reported relative reductions in utility due to anginal symptoms, the reduction in quality-of-life was estimated to be 10% (30) and 15% (32) if under classification occurred by 1 or 2 categories, respectively. The quality-of-life of CAD– patients who are classified as having *CAD+* was adjusted to reflect the disutility of taking medication.

Data analysis

All variables were entered in the model as distributions (Table A2, A3). Outcomes were calculated as the mean results from probabilistic sensitivity analysis, drawing random values from the parameter distributions(10000 samples).

Short-term outcomes included diagnostic costs, radiation exposure, and correct classification. Long-term outcomes included health care costs and quality-adjusted life years(QALYs). Both future costs and effectiveness were discounted at 3.5% (36).

Sensitivity analysis

The probability that a strategy was cost-effective was determined by the proportion of simulations in the probabilistic sensitivity analysis that demonstrated cost-effectiveness for that strategy (38). Value of information analysis was performed to determine the value of future research (39).

For the patients in the clinical cohort who did not undergo CAG, disease severity may have been overestimated by CCTA, which in turn could have caused a bias in favor of CCTA (since these data were used to determine the probability of correct classification, which would turn out high for the CT-strategy). To explore this limitation, we re-analyzed the model assuming that 40% of CAD+ men and women (randomly selected) who did not undergo CAG would actually be CAD– patients. Furthermore, we re-analyzed the model assuming that a proportion of patients with abnormal X-ECG would not be referred for CAG.

Results

Short-term analysis

Analysis of the short-term model revealed that the average diagnostic costs for the SOC-strategy were €739 (95%CI:547-978) and €526 (95%CI:€395-684) for men and women, respectively. The CT-strategy cost €494 (95%CI:€375-641) and €274 (95%CI:€205-356) for men and women, respectively (Table 2,3). The SOC-strategy classified 83% (95%CI:80-87%) of men correctly, whereas the CT-strategy classified 82% (95%CI:77-85%) correctly. The SOC-strategy classified 85% (95%CI:82 -88%) of women correctly, whereas the CT-strategy classified 88% (95%CI:85-92%) correctly. The SOC-strategy classified 85% (95%CI:82 -88%) of women correctly, whereas the CT-strategy classified 88% (95%CI:85-92%) correctly.

Subgroup analysis

For men with a pre-test probability of <70% and ≥70%, diagnostic cost-savings for the CT-strategy as compared to SOC were -€211 and -€312, respectively. In men with a pre-test probability ≥70%, the percentage correctly classified by CT was 11% lower compared to SOC (Table 2).

When re-analyzing women with a pre-test probability of <70% and ≥70%, diagnostic cost-savings for the CT-strategy as compared to SOC were -€234 and -€317, respectively (Table 3).

Table 1. Baseline characteristics*, diagnostic test results*, cost estimates and radiation exposure

Baseline characteristics		Value
Age(years), mean(SD)		56 (10)
Female: male		227:224 (0.48:0.52)
Risk profile		
	Nicotine use	138 (0.29)
	Hypertension	233 (0.49)
	Diabetes	68 (0.14)
	Dyslipidaemia	28 (0.59)
	Family history of cardiovascular disease	214 (0.45)
Chest pain (40)		
	Typical	146 (0.31)
	Atypical	251 (0.53)
	Non-anginal chest pain	74 (0.16)
Catheter-based coronary angiography		98 (0.21)
	≥50% stenosis, any vessel	57/98 (0.58)
	≥70% stenosis, any vessel	29/98 (0.30)
Percutaneous coronary intervention		46 (0.10)
Coronary bypass graft surgery		13 (0.03)
X-ECG Not performed		48/471 (0.10)
	Normal	190/423 (0.45)
	Non diagnostic	140/423 (0.33)
	Abnormal	93/423 (0.22)
CCS Not performed		8/471 (0.02)
	Mean CCS (median)	206 (15)
	Range	0-4817
	Interquartile range	0-145
CCTA Not performed		16/471 (0.03)
	Non-diagnostic	3/471 (0.01)
	No obstructive CAD	311/471 (0.66)
	Obstructive CAD (≥50%)	141/471 (0.34)
	Severe CAD (3VD, LM, prox. LAD)	48/141
CAG performed		121/471 (0.26)
	≥50% stenosis, any vessel	71/121 (0.59)
	≥70% stenosis, any vessel	34/121 (0.28)
Percutaneous coronary intervention		53/471 (0.11)
Coronary artery bypass graft surgery		18/471 (0.04)

	Cost estimates (euros)	Radiation exposure (mSv)
Exercise tolerance test (Expert opinion)	106	-
Coronary calcium score (8)	64	0.8
CT coronary angiography (8)	206	4.7
Single photon emission CT (Expert opinion)	545	12
Catheter-based coronary angiography (8)	1394	7.0
Percutaneous coronary intervention (24)	5000	15
Coronary bypass graft surgery (24)	14000	-

Results are shown as numbers (proportion of total) unless stated otherwise. X-ECG = exercise electrocardiography, CCTA = coronary computed tomography angiography, CAG = catheter-based coronary angiography
** Modified with permission from (11).*

Table 2. Short-term results (men).

	SOC-strategy			CT-strategy			Difference(CT vs. SOC)	
Disease category*	*CAD–*	*CAD+*	*Revasc.*	*CAD–*	*CAD+*	*Revasc.*	Mean	95%CI†
CAD–	**0.55**	0.05	-	**0.56**	0.04	-	**+0.01** ‡	-0.01,+0.03
CAD+	0.07	**0.10**	-	0.04	**0.14**	-	**+0.03** ‡	+0.01,+0.06
Revascularization	0.04	0.01	**0.18**	0.04	0.06	**0.12**	**-0.06** ‡	-0.09,-0.03
		Mean	95%CI†		Mean	95%CI†	Mean	95%CI†
Diagnostic costs(euros)		**739**	547,978		**494**	375,641	**-245**	-560,-117
Pre-test<70%		509	372,681		298	223,390	-211	-357,-87
Pre-test≥70%		1206	873,1617		894	654,1197	-312	-516,-146
% Correctly classified§		**0.83**	0.80,0.87		**0.82**	0.78,0.85	**-0.01**	-0.06,+0.02
Pre-test<70%		0.82	0.77,0.87		0.85	0.80,0.89	+0.03	-0.02,+0.07
Pre-test≥70%		0.85	0.79,0.89		0.74	0.66,0.82	-0.11	-0.19,-0.03
Radiation exposure‖(mSv)		**6.2**	4.5,8.3		**5.7**	4.4,7.1		
Pre-test<70%		4.1	2.9,5.6		4.1	3.1,5.3		
Pre-test≥70%		10.4	7.5,13.9		8.9	6.9,11.2		

SOC = standard-of-care, CT = computed tomography, Revasc. = revascularization, CI = confidence interval
** See Figure 4*
† Based on a probabilistic sensitivity analysis (10000 samples)
‡ Difference applies to correct classified cell(grey shaded)
§ Correct classification by the diagnostic strategy
‖ Radiation exposure related to diagnostic imaging

Table 3. Short-term results (women).

	SOC-strategy			CT-strategy			Difference(CT vs. SOC)	
Disease category*	*CAD–*	*CAD+*	*Revasc.*	*CAD–*	*CAD+*	*Revasc.*	Mean	95%CI†
CAD–	**0.77**	0.07	-	**0.80**	0.04	-	**+0.03**‡	+0.01,+0.05
CAD+	0.05	**0.03**	-	0.02	**0.05**	-	**+0.02**‡	+0.01,+0.04
Revascularization	0.03	0.00	**0.05**	0.02	0.04	**0.03**	**-0.02**‡	-0.04,-0.01
		Mean	95%CI†		Mean	95%CI†	Mean	95%CI†
Diagnostic costs(euros)		**526**	395,684		**274**	205,356	-252	-398,-127
Pre-test<70%		447	329,596		213	157,280	-234	-368,-113
Pre-test≥70%		796	581,1057		480	340,652	-317	-542,-123
% Correctly classified§		**0.85**	0.82,0.88		**0.88**	0.85,0.92	+0.03	-0.00,+0.06
Pre-test<70%		0.87	0.83,0.90		0.89	0.85,0.93	+0.02	-0.01,+0.06
Pre-test≥70%		0.81	0.75,0.85		0.86	0.78,0.92	+0.05	-0.03,+0.12
Radiation exposure‖(mSv)		5.1	3.6,7.0		4.1	3.1,5.3		
Pre-test<70%		3.9	2.7,5.4		3.3	2.4-4.3		
Pre-test≥70%		9.3	6.4-12.8		6.7	5.1-8.6		

SOC = standard-of-care, CT = computed tomography, Revasc. = revascularization, CI = confidence interval
** See Figure 4*
† Based on a probabilistic sensitivity analysis (10000 samples)
‡ Difference applies to correct classified cell (grey shaded)
§ Correct classification by the diagnostic strategy
‖ Radiation exposure related to diagnostic imaging

Long-term analysis

Analysis of the long-term model demonstrated a small gain in average quality-adjusted life years (QALYs)(+0.002,+0.004) and a decrease in health care costs(-€229,-€444) for the CT-strategy as compared with SOC, for men and women respectively (Table 4,5). Therefore, the CT-strategy is superior to the SOC-strategy (more effective and less expensive, SOC is dominated).

Subgroup analysis

For men with a pre-test probability of <70%, the difference in health care costs and effectiveness for the CT-strategy compared with SOC was -€227 and +0.004 QALY, respectively. For men with a pre-test probability of ≥70%, this difference was -€231 and -0.003 QALY, respectively (Table 4,Figure 4).

For women with a pre-test probability of <70%, the difference in health care costs and effectiveness for the CT-strategy compared with SOC was -€444 and +0.004 QALY, respectively. For women with a pre-test probability of ≥70%, this difference was -€782 and +0.006 QALY, respectively (Table 5, Figure 5).

Sensitivity analysis

In probabilistic sensitivity analysis the probability that the CT-strategy was cost-effective was 100% in all subgroups, except for men with a pre-test probability ≥70% in which case it was 59% Value of information analysis suggested no value for future research, except for men with a pre-test probability ≥70%.

The short-term diagnostic costs were insensitive to changes in underlying disease status based on CT(not shown). In men, the long-term cost-savings were reduced from -€229 (reference case) to -€135, and there was no longer a difference in QALYs. For men with a pre-test probability <70%, long-term cost-savings were reduced from -€227 to -€95, and for men with a pre-test probability ≥70% from -€231 to -€217. In women, the long-term cost-savings were reduced from -€444(base case) to -€296 and the net gain in QALYs was reduced from +0.004(reference case) to +0.003.

Short-term diagnostic costs for SOC were lowered when a proportion of patients with abnormal X-ECG would not undergo CAG, which reduced cost savings for CT. However, long-term costs were increased for SOC (due to over treatment in patients with false-positive X-ECG and follow-up testing), which was in favor of CT.

Discussion

Summary

We explored a hypothetical CT-strategy for its potential effectiveness and costs compared to SOC based on current guidelines for patients with stable chest pain.

Short-term results suggest that the CT-strategy is less expensive compared to SOC. This is explained by the fact that fewer patients undergo subsequent MPI or CAG, which are costly. Simultaneously, our results suggest that CT is more effective in correctly classifying patients, except for men with a pre-test probability ≥70%. Men with a pre-test probability

≥70% are more often correctly classified using SOC, because patients in the SOC-strategy are more often referred for CAG immediately after an abnormal test (which results in correct classification).

Long-term analyses demonstrated that the CT-strategy was slightly more effective and less costly compared to the SOC-strategy. Results were altered when the (potential) degree of disease severity overestimation by CCTA was taken into account. Because cost savings were

Table 4. Long-term results (men)

	SOC-strategy		CT-strategy		Difference (CT vs. SOC)	
	Mean	95%CI*	Mean	95%CI*	Mean	95%CI*
Health care costs (euros)	12969	10170,17764	12740	9957,17486	-229	-554,-84
Pre-test<70%	9691	7670,12588	9464	7457,12368	-227	-399,-56
Pre-test≥70%	19740	14439,30323	19509	14242,29952	-231	-448,-41
Effectiveness(QALYs)	11.671	11.079,12.158	11.672	11.078,12.160	+0.002	-0.002,+0.004
Pre-test<70%	11.980	11.505,12.403	11.984	11.510,12.406	+0.004	0.001,+0.007
Pre-test≥70%	11.025	9.935,11.703	11.022	9.926,11.703	-0.003	-0.011,+0.002
Radiation(mSv)	9.5	7.1,12.3	8.4	6.6,10.4		
Pre-test<70%	6.5	4.7,8.8	5.5	4.2,7.0		
Pre-test≥70%	15.6	11.8,20.1	14.2	11.1,17.9		
ICER(mean)	**Dominated**		**Superior**			
Pre-test<70%	Dominated		Superior			
Pre-test≥70%	30406		-			
Probability cost-effective (%)†	1		99			
Pre-test<70%	0		100			
Pre-test≥70%	41		59			

SOC = standard-of-care, CT = computed tomography, CI = confidence interval, ICER = incremental cost-effectiveness ratio
** Based on a probabilistic sensitivity analysis (10 000 samples)*
† Proportion of simulations that showed the CT-strategy to be cost-effective, using a willingness-to-pay threshold of €80.000/QALY

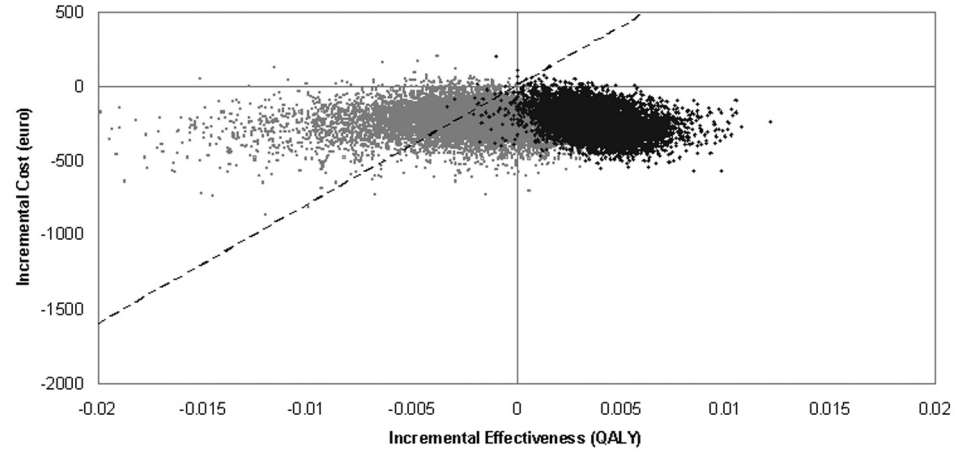


Figure 4. Incremental costs and effectiveness of the CT-strategy as compared with SOC; results from a probabilistic sensitivity analysis (10,000 samples) in men. Blue squares (right cloud) indicate results for men with a pre-test probability <70% and pink squares (left cloud) indicate men with a pre-test probability ≥70%. The dotted line represents the willingness-to-pay threshold of €80.000/QALY.

Table 5. Long-term results (women)

	SOC-strategy		CT-strategy		Difference(CT vs. SOC)	
	Mean	95%CI*	Mean	95%CI*	Mean	95%CI
Health care costs (euros)	8513	6977,10574	8068	6520,10134	-444	-696,-219
Pre-test<70%	7808	6303,9792	7464	5964,9457	-344	-582,-119
Pre-test≥70%	10896	8292,14736	10112	7474,13986	-782	-1319,-327
Effectiveness(QALYs)	12.684	12.269,13.097	12.689	12.273,13.101	+0.004	+0.002,+0.007
Pre-test<70%	12.727	12.306,13.139	12.731	12.310,13.144	+0.004	+0.002,+0.006
Pre-test≥70%	12.537	12.018,12.998	12.543	12.023,13.004	+0.006	+0.000,+0.012
Radiation(mSv)	6.8	4.9,9.1	5.2	4.0,6.7		
Pre-test<70%	5.6	3.9,7.8	4.3	3.2,5.6		
Pre-test≥70%	11.0	7.7,15.0	8.4	6.2,10.9		
ICER(mean)	**Dominated**		**Superior**			
Probability cost-effective (%)†	0		100			

SOC = standard-of-care, CT = computed tomography, CI = confidence interval, ICER = incremental cost-effectiveness ratio
** Based on a probabilistic sensitivity analysis (10000 samples)*
† Proportion of simulations that showed CT-strategy to be cost-effective, using a willingness-to-pay threshold of €80.000/QALY

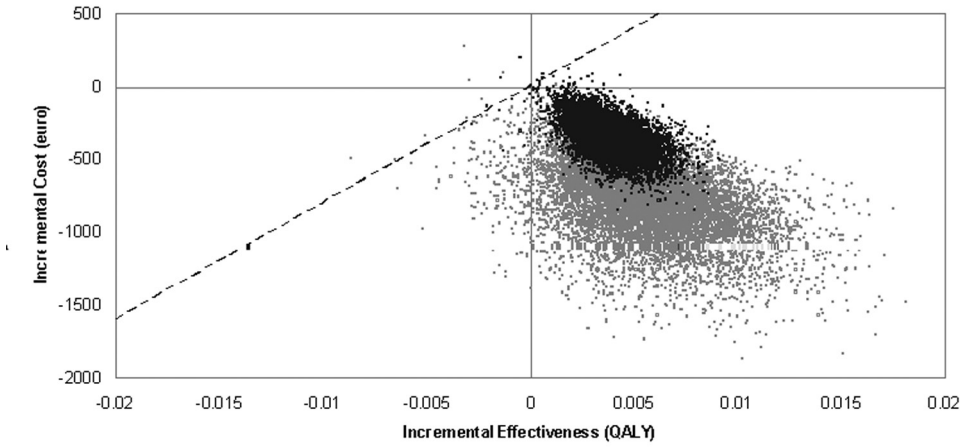


Figure 5. Incremental costs and effectiveness of the CT-strategy as compared with SOC; results from a probabilistic sensitivity analysis (10,000 samples) in women. Blue squares (upper cloud) indicate results for women with a pre-test probability <70%, pink squares (lower cloud) indicate women with a pre-test probability ≥70%. The dotted line represents the willingness-to-pay threshold of €80.000/QALY.

robust, the CT-strategy remained favorable even when the CT-strategy resulted in fewer QALYs, for example in men with a pre-test probability ≥70%. Results for CT were more favorable in women, which is explained by the lower prevalence of disease in women and the higher prevalence of zero calcium.

As expected, the gain in QALYs for the CT-strategy is small, since patients with persistent complaints will return to their physician until symptoms are treated adequately. The model assumes that within one year, all patients who are under–treated become appropriately treated. A benefit was gained from avoiding lifelong medication (over-treatment) in a substantial proportion of cases but this mainly affects costs. Nevertheless, even if the

gain in QALYs is very small or close to zero, the CT-strategy remains optimal because it is less costly. Furthermore, several expected additional benefits of the CT-strategy were not incorporated in the model, such as a reduced total time to final diagnosis and a reduction in additional downstream health care costs through a more expedient work-up. In addition, since the negative predictive value of CCTA is higher compared to X-ECG, physicians can be more confident in reassuring a patient after a negative CCTA.

Previous publications

Previous reports based on patient-level data from the US indicated that CCTA compared with SPECT reduces 1-year CAD-related health care expenditures (based on administrative databases and Medicare reimbursements) by 26% in a low risk population (41-42). No differences in clinical outcomes were observed. These reports analyzed 1-year outcomes of patients who underwent CCTA and who were matched to a cohort that underwent SPECT, whereas the current study analyzed the long-term outcomes of a pre-specified diagnostic protocol for patients presenting with chest pain. Furthermore, the current analysis is based on a cohort of patients who underwent both CCTA and X-ECG. In spite of these differences, the main conclusion is the same, namely that CT is cost-saving and equally effective as compared to SOC. However, other reports suggest that using CCTA increases costs as compared with MPI (43-44).

Lastly, a recent cost-effectiveness analysis which compared several CCTA-based strategies with myocardial perfusion SPECT and direct CAG found that the CCTA-based strategies were optimal up to a prevalence of CAD of 80% (45).

Limitations

Firstly, not all patients in the cohort underwent CAG. CCTA may have caused an overestimation of disease in these patients. To overcome this limitation, we included 6-months follow-up data in our reference standard in determining the disease category. Furthermore, we performed a sensitivity analysis to explore the magnitude of possible bias due to overestimation of disease by CCTA.

Secondly, we only observed the prognosis of patients who underwent both CCTA and X-ECG. Treatment decisions were based on the findings of both CCTA and X-ECG. We assumed that correctly classified patients would have a similar prognosis to that observed within follow-up of the clinical cohort. For patients incorrectly classified, however, prognosis was not observed. Therefore, several assumptions regarding unnecessary treatment and benefit of treatment forgone were made to estimate the prognosis of incorrectly classified patients.

Thirdly, model parameters were based on the patient cohort where possible, whereas best-available evidence from the literature was used otherwise. For example, we modeled the effect of SPECT using sensitivity and specificity as reported in a meta-analysis (3).

Fourthly, CCTA involves the possibility of incidental findings, which can occur in up to 28% of CCTAs (46-47). As of today, it is unclear whether it is useful or cost-effective to follow up on incidental findings. Moreover, the associated ethical and legal issues are difficult (if not impossible) to incorporate in a decision model. Also, although we estimated the radiation exposure, we did not model the harmful effects. Since the difference in radia-

tion exposure between the two strategies was small, this is unlikely to have an effect on the optimal decision.

Generalizability

Our analysis was based on a real-world Dutch population and Dutch cost estimates, which limits the generalizability. Nevertheless, in probabilistic sensitivity analysis we explored the effect of the uncertainty around our parameter inputs and found that our results were robust for all women and for men with a prior probability <70%. Furthermore, we compared only two strategies that reflect current practice at our institution. Other hospitals may have a different standard-of-care, which could alter the conclusion about the comparative effectiveness and costs.

Future research

Our analysis suggests that the CT-strategy is superior to SOC. However, the data in the model was based on a non-randomized observational study in which patients were prospectively recruited to undergo both tests. We did not directly observe the prognosis of patients who underwent X-ECG or CCTA only but instead estimated their prognosis with a decision model. A RCT would give valuable insight regarding outcomes and costs for both diagnostic strategies separately. In lieu of such a trial, this study provides preliminary estimates of the outcomes for a CT-strategy as compared with SOC. Our results can be used to make decisions regarding CT for patients presenting with stable chest pain, as long as RCTs with long-term follow-up are on-going. Furthermore, our results suggest that future research would mainly be beneficial for the decision regarding men with a pre-test probability ≥70%.

Conclusion

Analysis of our model suggests that a diagnostic strategy using initial evaluation with CT is less expensive and equally effective as compared to SOC, which was most pronounced for men with atypical symptoms and all women irrespective of their presenting symptoms. Although the results were robust, randomized controlled trials with long-term follow-up are needed to confirm our results.

Supplementary material

I Study population

Follow up was complete for 424 of 471 patients (90%). Cardiovascular events (cardiac death, myocardial infarction, unstable angina requiring hospitalization, and revascularization procedures) and CAGs during follow up were recorded. Revascularizations performed beyond 6 months or targeting lesions other than the lesions documented at the initial visit were considered as cardiovascular events. During a mean follow up of 2.6 years, 44 cardiovascular events occurred in 30 patients (Table A1) (13).

II Short term model (Decision tree)

Classification in disease categories

For the SOC-strategy, X-ECG result (normal vs. non-diagnostic vs. abnormal) probabilities were calculated in subgroups by 'true' disease category (CAD– vs. CAD+ vs. Revascularization), stratified for sex and pre-test probability (<70% vs. ≥70%) (14). Probabilities for a low vs. a high Duke Treadmill Score were calculated stratified by X-ECG result, disease category and pre-test probability. The CAG result (normal vs. low risk vs. high risk) was assumed to always reflect the 'true' disease category (CAD-, CAD+, Revascularization).

For the CT-strategy, probabilities for CAC scores (zero vs. >0-400 vs. ≥400) were calculated in subgroups by 'true' CAD risk category (CAD– vs. CAD+ vs. Revascularization), stratified for sex and the pre-test probability (<70% vs. ≥70%) (14). Subsequently, the CCTA test result probabilities were calculated for subgroups, stratified by the disease category, pre-test probability and CAC score. Again, the CAG result (normal vs. low risk vs. high risk) was assumed to always reflect the 'true' disease category (CAD, CAD+, Revascularization).

Myocardial perfusion imaging (MPI)

The effect of SPECT was modeled through its sensitivity and specificity reported in a recent meta-analysis (3) (Table A2). We assumed that all 'false positive' SPECT studies would occur in CAD– patients, resulting in misclassification of CAD– patients in the *CAD+* group. The sensitivity for detecting stenosis was assumed to be equal for both the CAD+ and the Revascularization group. False negative perfusion scans occur both in patients with CAD+ (classified as having *CAD–*) because of moderate disease and in Revascularization patients (classified as having *CAD+*) because of balanced ischemia when three vessel disease is present. We also assumed that MPI will never show severe ischemia in a CAD– patient.

III Long term analysis (Markov model)

Long term prognosis was modeled using a Markov model. A simplified state transition diagram is presented in Figure A1. Fatal and non fatal major adverse cardiac events (MACE) were based on the follow up data of the clinical patients. Since the event rate is likely to be higher in the first year, we calculated a baseline hazard for the first year and a baseline hazard for the subsequent years. A Cox proportional hazards model was used to determine the

sex-specific rates of MACE for CAD–, CAD+, and Revascularization patients (see Table A3). Based on our observation that 4 out of 44 MACE were fatal events, we assumed a constant proportion of fatal events of 9% (95%CI: 3-19%).

IV Cost estimates

Medication cost

Self-reported medication use was recorded during follow up contacts by telephone or via mailed questionnaires. The medication was grouped according to the Anatomical Therapeutic Chemical (ATC) classification system, developed by the World Health Organization. Only cardiovascular disease related medication was taken into account. As recommended by the Dutch Manual for cost analysis (15), we used an online cost registry provided by the Dutch Health Care Insurance board (16). From this registry, we obtained the average cost per Defined Daily Dosage (DDD) for each ATC code. For each patient, the average daily cost of medication was calculated by summing the cost per DDD for each self-reported medication. Subsequently, the sex-specific average daily cost across all patients within each disease category was calculated. Next, the average daily cost was multiplied by 365 to reflect annual costs (Table A3).

Costs for diagnostic work-up and revascularization during follow-up

Our model assumes that patients who are under classified will have persistent symptoms and will reclassify into the appropriate disease category within the first year after the initial assessment. To account for differences in follow-up diagnostic testing we assumed the following costs:

X-ECG strategy

From	To	Costs
CAD+, classified as CAD–	CAD+	MPI
Revascularization, classified as CAD–	Revascularization	MPI + CAG + Revascularization
Revascularization, classified as CAD+	Revascularization	CAG + Revascularization

CT strategy

From	To	Costs
Low risk CAD, classified as CAD–	CAD+	CAG
Revascularization, classified as CAD–	Revascularization	CAG + Revascularization
Revascularization, classified as CAD+	Revascularization	CAG + Revascularization

Assumptions & limitations

- All revascularizations performed were appropriate, no unnecessary interventions were performed. This assumption was explored in a sensitivity analysis.
- We assumed that different diagnostic findings (on CAG) obtained beyond the first 6 months are attributable to progress of disease as opposed to misdiagnosis by the initial work-up.
- The probability of a diagnostic test not being performed/ or being non-diagnostic, is independent of other variables in the model.

- Based on the cohort, we assumed that 77% of revascularizations are PCIs, and 23% are CABGs.
- All false positive SPECT studies occur in patients with 'Normal risk'.
- Sensitivity of SPECT for detecting CAD in CAD+ and Revascularization patients is assumed to be equal because false negative scans can occur in Revascularization patients if three vessel disease is causing balanced ischemia.
- After 1 year, all patients (except CAD– patients classified as *CAD+*) will be reclassified to the correct category. Once correctly classified, a patient cannot reclassify anymore.
- Two thirds of elective PCI procedures are performed in the same session as the diagnostic angiography. Costs, radiation exposure and mortality were adjusted accordingly.
- We did not model incidental findings.
- We did not model the harmful effects of radiation exposure.

Table A1. Follow up data

Characteristics		Value
Follow up completed		424/471 (0.90)
Follow-up duration (years)	Mean	2.6
	Median	2.6
	Interquartile range	2.1-3.2
All cause mortality		8/424 (0.02)
	Cardiac death	4/8
	Non-cardiac death	4/8
Stroke/ TIA		7/424
Cardiovascular events		30/424
	Cardiac death	4/424
	Non fatal MI	6/424
	Unstable angina, requiring hospitalization	10/424
	Percutaneous coronary intervention *	18/424
	Coronary artery bypass graft surgery *	4/424
Catheter-based angiography *		25/424
	Normal risk	13/25
	Low risk CAD	6/25
	High risk CAD	6/25

Results are shown as numbers (proportion of total) unless stated otherwise
MI = myocardial infarction, TIA = transient ischemic attack
** Excludes the procedures that were performed both within 6 months and were based on findings obtained during the initial visit to the chest pain clinic*

Table A2. Model Parameter Estimates, Distributions, and Data Sources

Parameter(s)	Point Estimate*	Range	Distribution type	Data source
Proportion of patients in disease category/ pre-test probability in men (women) †				
CAD–	0.59 (0.84)	based on number of individuals in subgroup	Dirichlet	Clinical cohort (11)
<70%	0.84 (0.79)			
≥70%	0.16 (0.21)			
CAD+	0.18 (0.08)	based on number of individuals in subgroup	Dirichlet	Clinical cohort (11)
<70%	0.62 (0.65)			
≥70%	0.38 (0.35)			
Revascularization	0.23 (0.08)	based on number of individuals in subgroup	Dirichlet	Clinical cohort (11)
<70%	0.27 (0.61)			
≥70%	0.73 (0.39)			
X-ECG				
Not performed	0.10	0.08-0.13	Beta	Clinical cohort (11)
Mortality	0.00001	0-0.0001	Beta	(17)
Cost ‡ (€)	106	69-151	Gamma	Expert opinion
DTS § | X-ECG normal				(18)
Low risk (high DTS)	0.69	based on number of individuals in subgroup	Dirichlet	Clinical cohort (11)
Intermediate risk	0.31			
High risk (low DTS)	0			
DTS § | X-ECG non-diagnostic				(18)
Low risk (high DTS)	0.54	based on number of individuals in subgroup	Dirichlet	Clinical cohort (11)
Intermediate risk	0.46			
High risk (low DTS)	0			
CAC scan				
Not performed	0.017	0.007-0.03	Beta	Clinical cohort (11)
Inconclusive result	0	-	-	Clinical cohort (11)
Mortality	0	-	-	Assumption
Cost ‡ (€)	64	41-91	Gamma	Cost analysis
Radiation exposure (mSv)	0.8	0.5-1.1	Gamma	(12)
Coronary CT angiography				
Not performed	0.017	0.008-0.03	Beta	Clinical cohort (11)
Inconclusive result	0.007	0.001-0.02	Beta	Clinical cohort (11)
Mortality	0.000006	0-0.00002	Beta	(19)
Cost ‡ (€)	206	110-322	Gamma	(8)
Radiation exposure ‡ (mSv)	4.7	3.0-6.7	Gamma	Random sample ¶
CAG				
Mortality	0.0011	0.0009-0.0014	Beta	(20)
Cost ‡ (€)	1394	900-1993	Gamma	(8)
Radiation exposure ‡ (mSv)	7	5-10	Gamma	(21)
MPI (SPECT)				
Not performed	0.017	0.008-0.03	Beta	As for CCTA
Inconclusive result	0.007	0.001-0.02	Beta	As for CCTA
Mortality	0.000006	0-0.00002	Beta	As for CCTA
Sensitivity	0.88	0.86-0.90	Beta	(22)
Specificity	0.73	0.69-0.77	Beta	(22)
Cost ‡ (€)	545	352-777	Gamma	Expert opinion
Radiation exposure ‡ (mSv)	12	8-17	Gamma	(21)
Percutaneous intervention				
Proportion of procedures	0.78	0.67-0.87	Dirichlet	Clinical cohort (11)
Mortality ||	0.011	0.008-0.015	Beta	(23)
Cost ‡ (€)	5000	3241-7152	Gamma	(24)
Radiation exposure ‡ (mSv)	15	10-21	Gamma	(21)
Bypass graft surgery				
Proportion of procedures	0.22	0.13-0.33	Dirichlet	Clinical cohort (11)
Mortality ||	0.018	0.014-0.022	Beta	(23)
Cost ‡ (€)	14000	9076-19991	Gamma	(24)

CAD = coronary artery disease, X-ECG = stress electrocardiography, DTS = Duke Treadmill Score, CAC = coronary artery calcification, CAG = catheter-based coronary angiography, MPI = myocardial perfusion imaging, SPECT = single photon emission computed tomography
** Point estimates are probabilities, unless stated otherwise*
† Diamond & Forrester (14)
‡ Gamma distributions were created by assuming a standard error of 20% of the mean
§ Low risk = DTS ≥ 5, Intermediate risk = DTS, -10 to <5, High risk = DTS ≤-11
|| 30-day mortality was used
¶ Based on 100 scans using a 128 slice dual source CT scanner

Table A3. Markov Model Parameter Estimates, Distributions, and Data Sources

Parameter(s)	Point Estimate	Range	Distribution type	Data source
Annual rate of MACE events*				
Overall (1st year)	0.038	0.023-0.061	Beta	Clinical cohort follow up
Overall (>1st year)	0.014	0.009-0.024	Beta	Clinical cohort follow up
Hazard rate ratios				
Male sex	1.40	0.60-3.22	Beta	Clinical cohort follow up
CAD–	-	-	-	
CAD+	7.81	2.7-22.5	Beta	
Revascularization	10.54	3.97-28.0	Beta	
Proportion fatal | MACE				
	0.09	0.03-0.19	Beta	Clinical cohort follow up
Probability of non cardiac death				
Dutch life tables (see Appendix)	Age, sex-specific			(25)
Hazard rate ratios (event rate)				
Post MACE vs. no previous MACE	1.45	1.25-1.66	Log normal	(26)
CAD– , classified as *CAD+*	1.00	0.83-1.20	Log normal	Assumption
CAD+, classified as *CAD–*	1.60	1.26-2.00	Log normal	(27-29)
Revascularization, classified as *CAD+*	1.60	1.26-2.00	‡	(27-29)
Revascularization, classified as *CAD–*	1.70	1.36-2.10	‡	Assumption

Quality of life estimates (SF-36) †					
Men (general population)					
45-54		0.790	0.77-0.81	Beta	(30)
55-64		0.778	0.76-0.80	Beta	(30)
65-74		0.746	0.72-0.77	Beta	(30)
75-84		0.688	0.67-0.71	Beta	(30)
85+		0.731	0.71-0.72	Beta	(30)
Women (general population)					
45-54		0.779	0.76-0.80	Beta	(30)
55-64		0.755	0.73-0.78	Beta	(30)
65-74		0.731	0.71-0.75	Beta	(30)
75-84		0.671	0.65-0.69	Beta	(30)
85+		0.698	0.68-0.72	Beta	(30)
Quality of life decrements (%, relative to general population)					
CAD–, classified as *CAD–*		0	-	-	
CAD+, classified as *CAD+*		5	0-7	Triangular	Estimation
Revascularization, classified as *Revascularization*		10	7-14	‡	Estimation
Quality of life decrements (%, relative to underlying disease status)					
CAD–, classified as *CAD+*		0.5	0-1	Triangular	Range based on (31)
CAD+, classified as *CAD–*		10	5-15.	Triangular	“Angina” (30)
Revascularization, classified as *CAD+*		10	5-15.	‡	“Angina” (30)
Revascularization, classified as *CAD–*		15	10-20.	‡	(32)
Costs (€)					
Annual cost of medication use §					
Men	*CAD–*	268	207-336	Gamma	Clinical cohort follow up / Dutch medication cost database (16)
	CAD+	540	434-659	Gamma	
	Revascularization	631	503-774	Gamma	
Women	*CAD–*	216	177-259	Gamma	Clinical cohort follow up / Dutch medication cost database (16)
	CAD+	636	473-821	Gamma	
	Revascularization	818	609-1058	Gamma	
Cardiovascular event		10000	6475-14265	Gamma	Estimation
Diagnostic work-up		||			see Appendix

CAD = coronary artery disease, HRR = hazard rate ratio
** MACE included cardiac death, myocardial infarction, unstable angina requiring hospitalization, and coronary revascularization (percutaneous coronary intervention, coronary artery bypass graft surgery)*
† SF-36, we assumed equal standard errors for all subgroups (SE=0.015)
‡ Similarly shaped and positively correlated with the previous distribution
§ See Appendix for medication cost calculations
|| Varies based on previous diagnostic testing in model, see Appendix

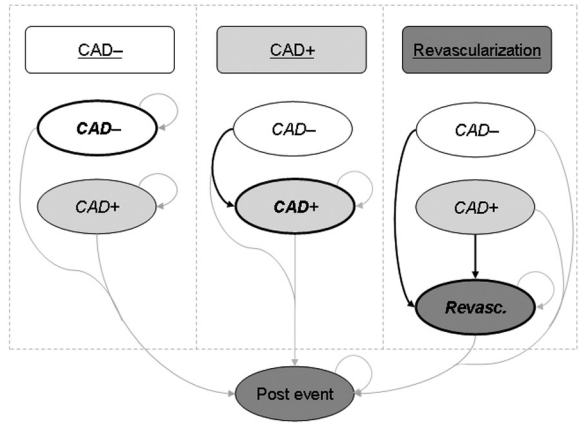


Figure A1. *Markov Model*: State transition diagram (i.e. schematic representation) of the Markov model that was used to model long term costs and quality adjusted life expectancy. Rectangular boxes refer to the underlying disease category (see Figure 4). Ellipses represent Markov health states. Bold arrows indicate the reclassifications (that can only occur during the first year) whereas light grey arrows indicate transitions that can occur during the entire follow-up. CAD– patients classified as *CAD+* do not reclassify which is reflected by the absence of a bold arrow. Patients are at risk of dying at any time and in every health state (Dead state not shown). Progression of disease was modeled through the cardiovascular event rates.

References

1. Fox K, Garcia MA, Ardissino D, et al. Guidelines on the management of stable angina pectoris: executive summary - the Task Force on the Management of Stable Angina Pectoris of the European Society of Cardiology. Eur Heart J 2006; 27:1341-1381.
2. Kwok Y, Kim C, Grady D, Segal M, Redberg R. Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol 1999; 83:660-666.
3. Vanhoenacker PK, Heijenbrok-Kal MH, Van Heste R, et al. Diagnostic performance of multidetector CT angiography for assessment of coronary artery disease: meta-analysis. Radiology 2007; 244:419-428.
4. Sun Z, Lin C, Davidson R, Dong C, Liao Y. Diagnostic value of 64-slice CT angiography in coronary artery disease: a systematic review. Eur J Radiol 2008; 67:78-84.
5. Schuetz GM, Zacharopoulou NM, Schlattmann P, Dewey M. Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. Ann Intern Med 2010; 152:167-177.
6. von Ballmoos MW, Haring B, Juillerat P, Alkadhi H. Meta-analysis: Diagnostic Performance of Low-Radiation-Dose Coronary Computed Tomography Angiography. Ann Intern Med 2011; 154:413-420.
7. Kristensen TS, Engstrom T, Kelbaek H, von der Recke P, Nielsen MB, Kofoed KF. Correlation between coronary computed tomographic angiography and fractional flow reserve. Int J Cardiol 2010; 144:200-205.
8. Genders TS, Meijboom WB, Meijs MF, et al. CT coronary angiography in patients suspected of having coronary artery disease: decision making from various perspectives in the face of uncertainty. Radiology 2009; 253:734-744.
9. Halpern EJ, Savage MP, Fischman DL, Levin DC. Cost-effectiveness of coronary CT angiography in evaluation of patients without symptoms who have positive stress test results. AJR Am J Roentgenol 2010; 194:1257-1262.
10. Pontone G, Andreini D, Bartorelli AL, et al. Comparison between low-dose multidetector computed coronary angiography and myocardial perfusion imaging test in patients with intermediate pre-test likelihood of coronary artery disease. Int J Cardiol 2011; 147:454-457.
11. Nieman K, Galema T, Weustink A, et al. Computed tomography versus exercise electrocardiography in patients with stable chest complaints: real-world experiences from a fast-track chest pain clinic. Heart 2009; 95:1669-1675.

12. Nieman K, Galema TW, Neefjes LA, et al. Comparison of the value of coronary calcium detection to computed tomographic angiography and exercise testing in patients with chest pain. Am J Cardiol 2009; 104:1499-1504.
13. Dedic A, Genders TS, Ferket BS, et al. Stable angina pectoris: head-to-head comparison of prognostic value of cardiac CT and exercise testing. Radiology 2011; 261:428-436.
14. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med 1979; 300:1350-1358.
15. Oostenbrink J, Bouwmans C, Koopmanschap MA, Rutten FFH. Dutch Manual for Cost-Analyses [in Dutch]. Amstelveen, The Netherlands: Dutch Health Care Insurance Board, 2004.
16. GIP databank. Dutch Health Care Insurance Board. http://www.gipdatabank.nl/. Accessed January 13, 2012
17. Gibbons L, Blair SN, Kohl HW, Cooper K. The safety of maximal exercise testing. Circulation 1989; 80:846-852.
18. Johnson GG, Decker WW, Lobl JK, et al. Risk stratification of patients in an emergency department chest pain unit: prognostic value of exercise treadmill testing using the Duke score. Int J Emerg Med 2008; 1:91-95.
19. Katayama H, Yamaguchi K, Kozuka T, Takashima T, Seez P, Matsuura K. Adverse reactions to ionic and nonionic contrast media. A report from the Japanese Committee on the Safety of Contrast Media. Radiology 1990; 175:621-628.
20. Noto TJ, Jr., Johnson LW, Krone R, et al. Cardiac catheterization 1990: a report of the Registry of the Society for Cardiac Angiography and Interventions (SCA&I). Cathet Cardiovasc Diagn 1991; 24:75-83.
21. Mettler FA, Jr., Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. Radiology 2008; 248:254-263.
22. Heijenbrok-Kal MH, Fleischmann KE, Hunink MG. Stress echocardiography, stress single-photon-emission computed tomography and electron beam computed tomography for the assessment of coronary artery disease: a meta-analysis of diagnostic performance. Am Heart J 2007; 154:415-423.
23. Bravata DM, Gienger AL, McDonald KM, et al. Systematic review: the comparative effectiveness of percutaneous coronary interventions and coronary artery bypass graft surgery. Ann Intern Med 2007; 147:703-716.
24. NZA. Dutch National Authority of Health Care. http://ctg.bit-ic.nl/Nzatarieven/top.do. Accessed January 2011
25. Dutch Central Bureau of Statistics. http://statline.cbs.nl. Accessed 10 January 2011
26. Deckers JW, Goedhart DM, Boersma E, et al. Treatment benefit by perindopril in patients with stable coronary artery disease at different levels of risk. Eur Heart J 2006; 27:796-801.
27. Berger JS, Brown DL, Becker RC. Low-dose aspirin in patients with stable cardiovascular disease: a meta-analysis. Am J Med 2008; 121:43-49.
28. Dagenais GR, Pogue J, Fox K, Simoons ML, Yusuf S. Angiotensin-converting-enzyme inhibitors in stable vascular disease without left ventricular systolic dysfunction or heart failure: a combined analysis of three trials. Lancet 2006; 368:581-588.
29. Baigent C, Keech A, Kearney PM, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 2005; 366:1267-1278.
30. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making 1993; 13:89-102.
31. Pletcher MJ, Lazar L, Bibbins-Domingo K, et al. Comparing impact and cost-effectiveness of primary prevention strategies for lipid-lowering. Ann Intern Med 2009; 150:243-254.

32. Hlatky MA, Boothroyd DB, Melsop KA, et al. Medical costs and quality of life 10 to 12 years after randomization to angioplasty or bypass surgery for multivessel coronary artery disease. Circulation 2004; 110:1960-1966.
33. Morise AP, Bobbio M, Detrano R, Duval RD. Incremental evaluation of exercise capacity as an independent predictor of coronary artery disease presence and extent. Am Heart J 1994; 127:32-38.
34. Esteves FP, Khan A, Correia LC, et al. Absent coronary artery calcium excludes inducible myocardial ischemia on computed tomography/positron emission tomography. Int J Cardiol 2011; 147:424-427.
35. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 2007; 356:1503-1516.
36. National Institute for Clinical Excellence (NICE). Guide to the methods of technology appraisal. Oxford, England: Radcliff Medical Press, 2004.
37. RIVM. Dutch National Institue for Public Health and the Environment. http://www.rivm.nl/vtv/object_document/o3171n17964.html. Accessed March 2008
38. Groot Koerkamp B, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink MG. Uncertainty and patient heterogeneity in medical decision models. Med Decis Making 2010; 30:194-205.
39. Groot Koerkamp B, Nikken JJ, Oei EH, Stijnen T, Ginai AZ, Hunink MG. Value of information analysis used to determine the necessity of additional research: MR imaging in acute knee trauma as an example. Radiology 2008; 246:420-425.
40. Diamond GA. A clinically relevant classification of chest discomfort. J Am Coll Cardiol 1983; 1:574-575.
41. Min JK, Kang N, Shaw LJ, et al. Costs and clinical outcomes after coronary multidetector CT angiography in patients without known coronary artery disease: comparison to myocardial perfusion SPECT. Radiology 2008; 249:62-70.
42. Min JK, Shaw LJ, Berman DS, Gilmore A, Kang N. Costs and clinical outcomes in individuals without known coronary artery disease undergoing coronary computed tomographic angiography from an analysis of Medicare category III transaction codes. Am J Cardiol 2008; 102:672-678.
43. Nicol ED, Stirrup J, Leatham E, et al. Clinical management and short-term cost - 64-slice MDCT vs. myocardial perfusion scintigraphy. Int J Cardiol 2010; 144:248-250.
44. Shreibati JB, Baker LC, Hlatky MA. Association of coronary CT angiography or stress testing with subsequent utilization and spending among Medicare beneficiaries. JAMA 2011; 306:2128-2136.
45. Min JK, Gilmore A, Budoff MJ, Berman DS, O'Day K. Cost-effectiveness of coronary CT angiography versus myocardial perfusion SPECT for evaluation of patients with chest pain and no known coronary artery disease. Radiology 2010; 254:801-808.
46. Koonce J, Schoepf JU, Nguyen SA, Northam MC, Ravenel JG. Extra-cardiac findings at cardiac CT: experience with 1,764 patients. Eur Radiol 2009; 19:570-576.
47. Bendix K, Jensen JM, Poulsen S, Mygind N, Norgaard BL. Coronary dual source multi detector computed tomography in patients suspected of coronary artery disease: Prevalence of incidental extra-cardiac findings. Eur J Radiol 2010.

Chapter 11

Stress Myocardial Perfusion Cardiac Magnetic Resonance Imaging vs. Coronary CT Angiography in the Diagnostic Work-up of Patients with Stable Chest Pain: Comparative Effectiveness and Costs.

Submitted

Tessa S.S. Genders
Steffen E. Petersen
Francesca Pugliese
Amardeep Dastidar
Kirsten E. Fleischmann
Koen Nieman
M.G. Myriam Hunink

Abstract

Objective

To determine the comparative effectiveness and costs of coronary CT angiography (CCTA) and stress cardiac magnetic resonance imaging (CMR) for diagnosing coronary artery disease (CAD).

Design, setting, and patients

A Markov micro-simulation model for 60-year-old patients with stable chest pain, using the perspective of the United States (US), United Kingdom (UK), and the Netherlands (NL).

Interventions

CCTA, CMR, and CCTA+CMR (CCTA, if positive followed by CMR) were considered and compared to direct catheter-based angiography (CAG) and no testing. The strategies were considered both as conservative strategy (patients with mildly-positive test results are not referred for CAG), and as invasive strategy (all patients with positive test results are referred for CAG).

Main outcome measures

Lifetime costs, quality-adjusted life years (QALY), and radiation exposure.

Results

Differences in effectiveness (QALYs) across strategies were very small. For 60-year old men and women, the CCTA, CMR, and CAG strategies were dominated, because the CCTA+CMR-conservative strategy was slightly more effective, and less expensive. Compared to the CCTA+CMR-conservative strategy, the CCTA+CMR-invasive strategy was slightly more costly and slightly more effective. The CCTA+CMR-invasive strategy was cost-effective for the US and NL, but not for the UK. When patients with false-negative test results were assumed to remain false-negative for 3 years, differences between strategies increased, and the CCTA-invasive strategy became cost-effective for UK and NL.

Conclusions

Quality-adjusted life expectancy was similar across strategies. The CCTA+CMR strategy was cost-effective up to a pre-test probability of 70-90%, depending on the country. Above these thresholds, the CMR-strategy was cost-effective.

Introduction

Although guidelines recommend stress electrocardiography as the first line diagnostic test for patients with stable chest pain (1-3), its diagnostic accuracy is limited (4), The 2010 guidelines from the United Kingdom on recent onset chest pain have eliminated stress electrocardiography from the diagnostic algorithm (5), and recommended coronary CT calcium scoring with or without coronary CT angiography (CCTA) for patients with a pre-test likelihood of 10-29% and non-invasive functional imaging tests for patients with a pre-test likelihood of 30-60%.

Although CCTA is reliable in ruling out the presence of coronary artery disease (CAD) (sensitivity 95-100%) (6-8), it is less useful for identifying patients who would benefit from revascularization, because of the poor correlation between the anatomy and functional significance of a stenosis (9-10). Stress myocardial perfusion cardiac magnetic resonance imaging (CMR) can visualize myocardial perfusion or inducible regional wall motion abnormalities and thus stenoses with hemodynamic significance (11). With a sensitivity and specificity of 83-91% and 81-86%, respectively, for diagnosing obstructive CAD on catheter-based coronary angiography (CAG) (12-14). CMR could be used as an alternative or in addition to CCTA (9). Currently, single-photon emission computed tomography (SPECT) is more widely used for visualizing perfusion and wall motion. However, CMR has a higher spatial resolution, a superior diagnostic performance (12, 15), and has the additional advantage that it does not require radiation exposure.

Our aim was to determine the comparative effectiveness and costs of CCTA and CMR in the diagnostic work-up of patients with stable chest pain, from the perspective of the United States (US), United Kingdom (UK), and the Netherlands (NL).

Methods

Decision model

A micro-simulation model was developed (in DATA Pro 2009 Suite, TreeAge Software Inc, Williamstown, MA, USA) to evaluate the comparative effectiveness and costs of CCTA and CMR. Diagnostic outcomes were modeled with a decision tree, and lifetime prognosis was modeled using a Markov-model. Model parameters were based on the best-available evidence from the literature. The model was analyzed from the perspective of the UK, US, and NL. The supplementary material provides model details and assumptions. Here we summarize the most pertinent points.

Target population

Our target population (Figure 1) consisted of 60-year old patients with stable chest pain presenting to the cardiologist, without a history of coronary artery disease (CAD), percutaneous coronary intervention (PCI), or coronary artery bypass graft surgery (CABG).

Diagnostic pathways

The following diagnostic strategies were modeled: No testing, CCTA, CMR, CCTA if positive followed by CMR, and direct CAG (Figure 2). The CCTA, CMR and CCTA+CMR strategy were analyzed both as a conservative and invasive strategy. The invasive approach

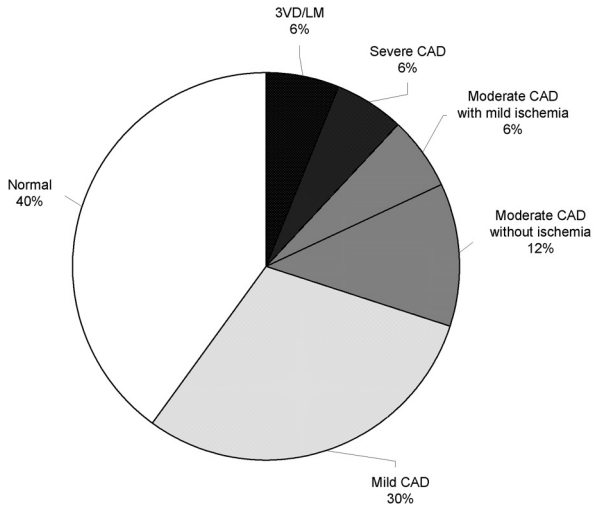


Figure 1. Disease distribution for 60-year old patients with a 30% prevalence of obstructive coronary artery disease (CAD) used in the base case analysis. Mild CAD = <50% stenosis, moderate CAD = ≥50-70% stenosis or ≥70% in small vessels, severe CAD = ≥70% stenosis (see Methods). For patients with severe and moderate CAD, we assumed that 33% had mild ischemia, and 67% had severe ischemia.

implied that patients with obstructive CAD on CCTA (≥50% stenosis in at least 1 vessel but otherwise regardless of severity) and patients with perfusion defects on CMR (regardless of severity), are referred for CAG. In the conservative approach, patients with moderate CAD on CCTA or mild perfusion defects on CMR receive optimal medical treatment (OMT), without referral to CAG (Figure 2).

Severity of disease was categorized into either: 1) normal coronary arteries, 2) mild CAD (<50% obstruction), 3) moderate CAD (50-70% in proximal vessels or ≥70% in small vessels) without perfusion defects, 4) moderate CAD (50-70% in proximal vessels or ≥70% in small vessels) with suspected/mild perfusion defect(s), 5) severe CAD (≥70% stenosis) in 1 or 2 proximal vessels [except left main] with suspected/mild perfusion defect(s), 6) severe CAD (≥70% stenosis) in 1 or 2 proximal vessels [except left main] with severe perfusion defect(s), 7) moderate-severe three-vessel disease (3VD) or left main (LM) disease (≥50%) with suspected/mild perfusion defect(s), or 8) moderate-severe 3VD/LM disease with severe perfusion defect(s). Severity of disease was tracked using a Monte Carlo tracker variable and was used to determine test results, costs, treatments, and prognosis.

Estimates of sensitivity and specificity for CCTA (0.98 and 0.90, respectively) (6-8) and CMR (0.90 and 0.80 respectively) (13-14) for the detection of obstructive CAD on CAG (<50% vs. ≥50%) were derived from published meta-analyses (Table 1). For CCTA and CMR, we assumed that false-positive results show mild CAD and mild perfusion defects, respectively. Incidental findings (i.e. mostly pulmonary nodules) associated with CCTA were modeled by taking into account the probability of occurrence (16), an average cost (16), and a disutility (Table 1). Life-time radiation exposure was calculated based on the effective dose for diagnostic tests and procedures (Table 1).

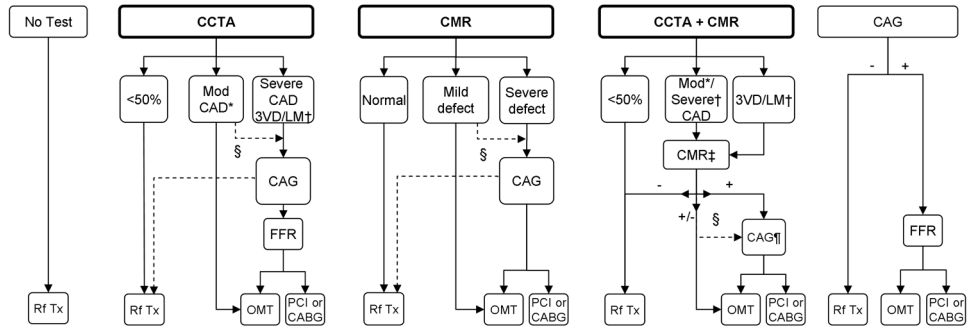
Treatment and prognosis

For simplicity, we assumed that the initial optimal therapy if the true disease severity is known includes risk factor management (Rf Tx) in individuals with normal coronary arteries, mild CAD, and moderate CAD without ischemia; OMT for patients with mild ischemia

and moderate to severe CAD; PCI for patients with severe CAD and severe ischemia, and CABG for patients with 3VD/LM disease. We assumed that PCI and CABG include OMT and that OMT includes Rf Tx. Figure 2 shows the treatment decision based on test results. See Supplementary Material for details on medication use.

The Markov-model for long-term prognosis included the following health states: Alive, Post Myocardial Infarction (MI), and Dead. We modeled the risk of major adverse cardiac events (MACE which consisted of revascularization, non-fatal MI, and cardiac death), depending on disease severity. For patients with 3VD/LM disease, rates of MACE were based on the CABG-arm of the SYNTAX trial (17-18), which compared CABG with drug-eluting stenting for 3VD/LM disease (mean follow-up 3 years). For patients with suspected/mild perfusion defects and moderate to severe CAD (treated with OMT) and patients with severe CAD and severe perfusion defects (treated with PCI), prognosis was based on the OMT- and PCI-arm of the COURAGE trial (19), respectively (mean follow-up 4.6 years). To allow for a higher event rate in the first year after treatment initiation (for OMT, PCI, and CABG), MACE rates were calculated separately for the first year vs. all subsequent years. For patients without CAD and mild CAD, prognosis was based on a recent meta-analysis of the prognostic value of CCTA (20) (Supplementary Table 2). The prognosis of patients with moderate CAD without perfusion defects was assumed to be equal to the prognosis of individuals with mild CAD. The risk of dying from non-cardiac causes was based on the most recent age- and sex-specific mortality rates available for the US (21), UK (22), and NL (23), separately.

Figure 2. Diagnostic strategies



* Moderate CAD is defined as 1- or 2-vessel disease (≥50-70%) or ≥70% stenosis in small vessels (no or mild perfusion defects).

† Severe CAD was defined as 1- or 2-vessel disease with ≥70% (mild or severe perfusion defects). 3VD/LM is defined as 3-vessel disease (≥50%) or left main coronary stenosis (≥50%) (severe perfusion defects)

‡ CMR can show either: no perfusion defects (-), suspected/mild perfusion defect(s) (+/-), or severe perfusion defect(s) (+)

§ The CCTA, CMR and CCTA+CMR strategies were analyzed according to a conservative and invasive approach (§, dashed lines). In the conservative approach, patients with moderate CAD on CCTA or mild perfusion defects on CMR (which includes the false-positives!) are treated medically, without performing CAG. In the invasive strategy (§, dashed lines), patients with moderate CAD on CCTA or mild perfusion defects are referred for CAG (CCTA and CMR false-positives are thus identified as free of obstructive CAD or perfusion defects, respectively).

¶ FFR only if CMR was not performed prior to CAG

CCTA = coronary computed tomography angiography, CMR = cardiac stress perfusion magnetic resonance imaging, CAG = catheter-based coronary angiography, CAD = coronary artery disease, Tx = treatment, Rf = risk factor, OMT = optimal medical treatment, PCI = percutaneous coronary intervention, CABG = coronary artery bypass graft surgery, FFR = fractional flow reserve

Patients with CAD and false-negative test results forgo the benefit of treatment. We assumed that the hazard rate ratio (HRR) of treatment versus no treatment on the outcome MACE was 0.70 (range used in sensitivity analysis: 0.60-0.90). The inverse of the treatment HRR was used in the model to adjust the (known) event rate among treated patients (i.e. those with true positive test results) in order to estimate the (unknown) event rate among untreated patients with CAD (i.e. those with false-negative test results). To reflect clinical practice, we assumed that patients with false-negative test results return to their physicians with persistent complaints, undergo additional testing, and start receiving the appropriate treatment within 1 year (except for patients with moderate CAD without ischemia, in which case we assumed that only 25% would return). After a non-fatal MI, we assumed the rate of MACE increased with an HRR of 1.44 (95%CI: 1.25-1.66) (24)Individuals with false-positive test results are referred for CAG (invasive approach) and subsequently identified as negative. However, in the conservative approach, patients with false-positive test results are not referred for CAG. We assumed such patients receive OMT for their remaining life expectancy, which was assumed to affect costs only (no effect on events).

Costs

Costs were based on best-available evidence and expert opinion (Table 1) and converted to 2011 dollars (US), pounds (UK), and euros (NL) using the country-specific (medical care component for US) consumer price indices.

Quality-of-life

Age- and sex-specific quality-of-life estimates for those without CAD and without chest pain were based on EQ-5D reference values for the general population (25). Published domain-specific SF-36 and RAND-36 scores based on the OMT-arm of the COURAGE trial (26) and the CABG-arm of the SYNTAX trial (27), respectively, were used to calculate a mean EQ-5D utility using a validated algorithm (28) and subsequently used to calculate the quality-of-life decrement relative to the general population of the same age and sex. For the first year, quality-of-life for patients receiving OMT/PCI/CABG was based on the average utility decrement as observed in the trials. For subsequent years, the last observed utility in the trial was carried forward. Baseline utility measurements from the trials were used to model the quality-of-life for patients who forgo the benefit of treatment (false-negatives) (See Supplementary Table 2). The quality-of-life of false-positive individuals was adjusted to reflect the disutility of taking medication (29).

Data analysis

All variables were entered in the model as distributions (Appendix Table 3). Monte Carlo microsimulation (two-level) was used to calculate mean outcomes. Parameter values were randomly drawn from the distributions (10000 samples). For each parameter value set, 1000 random walks (representing individual patients) were simulated and outcomes were averaged across individuals. Parameter uncertainty is reflected in the 95% confidence intervals of costs and QALYs (see Appendix Table 4 and 5). Outcomes included lifetime costs, quality-adjusted life-years, radiation exposure, and incremental cost-effectiveness ratios (ICERs). The ICER of strategy A vs. B is defined as the difference in costs divided by difference in effectiveness (strategy A minus B). An ICER below the willingness-to-pay (WTP) threshold implies that strategy A is a cost-effective alternative to strategy B. Strategies were compared by calculating the ICER compared to the next more costly strategy, eliminating strategies that are dominated (more costly, less effective), and by extended dominance (when the

Table 1. Parameter estimates and country-specific costs/recommendations for cost-effectiveness analysis. See Appendix Table 3 for distributions.

			United States		**United Kingdom**		**The Netherlands**	
CEA recommendations:			(30-32)		(33)		(34-35)	
Perspective:			Societal		Health care		Societal	
Costs:			2011 $		2011 £		2011 €	
Discount rate costs:			3%		3.5%		4%	
Discount rate effectiveness:			3%		3.5%		1.5%	
WTP threshold:			$50 000 / QALY		£25 000 / QALY		€ 80 000 / QALY	
	Mean	**Source**	**Cost***	**Source**	**Cost***	**Source**	**Cost***	**Source**
CCTA			372	CPT 75574	291	Expert opinion	215	EMC fee
Sensitivity	0.98	(6-8)						
Specificity	0.90	(6-8)						
Radiation (mSv)	5							
Mortality (%)	0.0006	(36)						
Disutility	0.0005							
Incidental finding on CCTA †			559	(16)	359	Estimation based on (16)	434	Estimation based on (16)
Prevalence (%)	7	(16)						
Radiation (mSv)	9.4	(16)						
Disutility	0.001							
CMR			689	CPT 75563	548	BSCI	319	EMC fee
Sensitivity (%)	0.90	(13-14)						
Specificity (%)	0.80	(13-14)						
Mortality (%)	0.0006	(36)						
Disutility	0.00075							
CAG			2989	CPT 93454 + APC 80	1052	NHS HRG tariff EA36A	1513	EMC fee
Radiation (mSv)	7	(37)						
Mortality (%)	0.11	(38)						
Peri-procedural MI (%)	0.05	(39)						
Disutility	0.005							
FFR			715	(40)	460	Estimation based on (40)	555	Estimation based on (40)
Outpatient PCI			6529	CPT 92980 + APC 104	3676	NHS HRG tariff EA32Z	4168	DBC 140437, 140439, 140990, 140991 (43)
Radiation (mSv)	15	(41)						
Mortality (%)	1.1	(42)						
Disutility	0.005							
Clopidogrel (1 month)								
CABG			38217	AHRQ CCS 44 (44)	7318	NHS HRG tariff EA14Z	11887	DBC 140923, 140924 (43)
Mortality (%)	1.8	(42)						
Disutility	0.02							
Non-fatal MI			10208	DRG 280-282 (44)	1519 + PCI	NHS HRG tariff EB10Z	3983 + PCI	DBC 141527, 141528 (43)
Disutility	0.04	(41)						
Medication (annual)								
Aspirin (80 mg)			24		10		63	
Simvastatin (40 mg)			1702		14		52	
Atenolol (50 mg)			315	Redbook ‡	9	NHS (45)	82	CVZ (46) §
Isosorbide mononitrate (60 mg)			376		126		145	
Enalapril (20 mg)			587		11		34	

CCTA = coronary computed tomography angiography, CMR = cardiac stress perfusion magnetic resonance imaging, CAG = catheter-based coronary angiography, PCI = percutaneous coronary intervention, CABG = coronary artery bypass graft surgery, FFR = fractional flow reserve, EMC = Erasmus University Medical Center, APC = Ambulatory Procedure Classifications, which reflects the Medicare national average facility fee, CPT = Current Procedural Terminology, which reflects the Medicare national average physician fee, BSCI = British Society for Cardiovascular Imaging, CCS = clinical classification software, DRG = diagnosis-related group, AHRQ = Agency for Healthcare Research & Quality (US), NHS = National Health Service (UK), HRG = health-related group, DBC = Dutch diagnosis-related group, CVZ = Dutch Health Care Insurance Board (NL)

** All costs were assumed to follow a gamma distribution, assuming a standard error of 20% of the mean*

In 2011, $1.00 was equivalent to €0.72 or £0.62

† Incidental findings of indeterminate clinical importance that require follow-up testing.

‡ Mid-range of the average wholesale price (AWP), Redbook accessed on January 9, 2012

§ Total cost per defined daily dosage

ICER exceeds the ICER of another more costly and more effective strategy). See Table 2 for country-specific recommendations for cost-effectiveness analysis.

Using one-way sensitivity analysis, we assessed the effect of varying key parameters across plausible values.

Results

Base case analysis

Estimated quality-adjusted life years (QALYs) were similar across strategies, but varied across countries due to differences in recommended discount rate. The CCTA and combination strategies were less expensive compared to the CMR-strategies in the UK and NL. In the US, the conservative strategies were most expensive (Table 2 and 3, Figure 3A and Figure 4).

Calculations suggest that the invasive CCTA+CMR strategy is cost-effective for both men and women in the US and NL, mainly because it was less expensive compared to the other strategies. For men and women in the UK, the CCTA+CMR-conservative strategy was optimal, because the ICER for the invasive vs. conservative CCTA+CMR strategy exceeded the UK-WTP of £25,000/QALY. The CCTA, CAG, and CMR strategies were dominated, which means that the other strategies were less expensive and more effective (Table 2 and 3).

Sensitivity analysis

Pre-test probability

For the UK, the CCTA+CMR-invasive was optimal below a pre-test probability of 30%, and CCTA+CMR-conservative was optimal for 30-70%. For and above a pre-test probability of 70%, 80%, and 90%, the CMR-invasive strategy was cost-effective for UK, US, and NL women, respectively (Figure 3B, Table 4)

Statin cost in the United States

Because there is a wide variability in pricing of statins, depending on the source used, we reduced the annual statin costs in the US from $1702 to $100 (47), and found that total costs were significantly reduced for all strategies. Although the differences between the strategies decreased, the optimal strategy remained CCTA+CMR-invasive (Figure 4).

Test characteristics CMR

When the sensitivity of CMR was increased to 95% and the specificity to 90%, the CMR-strategies became less costly and more effective, however, the CCTA+CMR strategies remained superior.

Treatment benefit

The HRR of 0.70 reflects the reduction of adverse events through treatment and was varied between 0.60 and 0.90. Although the absolute costs and QALYs changed, the incremental costs and QALYs remained virtually unchanged, and conclusions were unaltered.

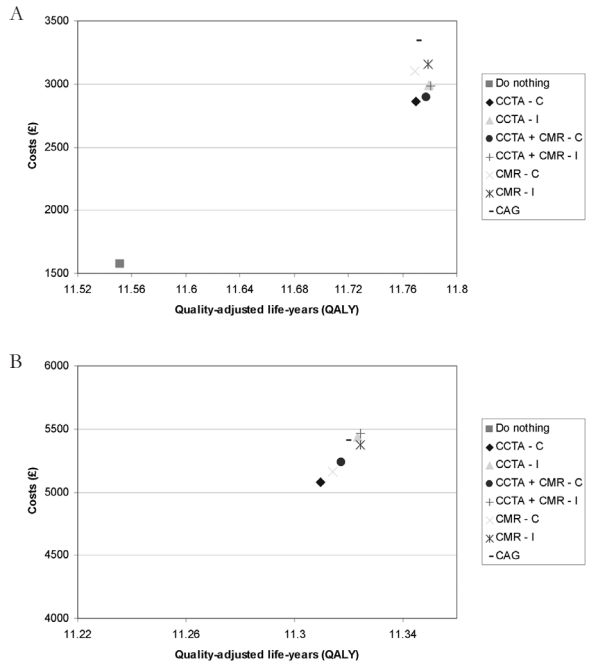


Figure 3. Cost-effectiveness analysis results for 60-year old men (UK analysis) with a 30% (Panel A) and 70%* (Panel B) prevalence of obstructive CAD, respectively. Note the different x-axis and y-axis scale between Panel A and B. See Table 2 for incremental cost-effectiveness ratios. I = invasive approach, C = conservative approach.

* When the pre-test probability was 70%, the "Do Nothing" strategy yielded 10.79 QALYs at a cost of £2690

False negatives

The base case analysis assumed that false-negatives are identified within the first year. Changing this assumption to 3 and 5 years exaggerated the differences between strategies. For the UK, the CCTA-conservative and invasive strategies were most favorable under this alternative assumption (Figure 5, Table 4). For the NL (not shown), the combination strategies

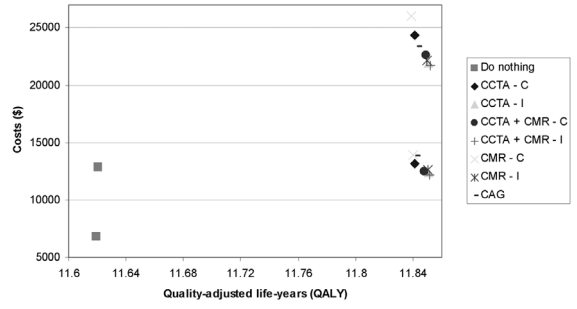


Figure 4. Cost-effectiveness analysis results for 60-year old men (US analysis) (upper dots, base case analysis with annual statin costs of $1702), and sensitivity analysis in which statin costs were reduced (lower dots, annual statin costs $100). I = invasive approach, C = conservative approach

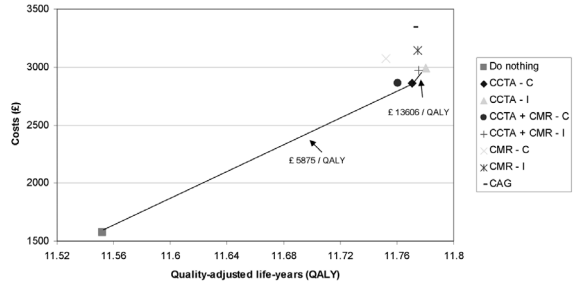


Figure 5. Cost-effectiveness analysis results for 60-year old men (UK analysis). Sensitivity analysis, re-analyzing cost-effectiveness assuming that patients with false-negative test results stay false-negative for 5 years (as opposed to 1 year in the base case analysis, see Figure 3A). I = invasive approach, C = conservative approach.

Table 2. Results of base case cost-effectiveness analysis (60-year old men, pre-test probability of 30%). I = invasive approach, C = conservative approach

	Radiation	US			UK			NL		
	(mSv)	Mean costs ($)	QALYs	ICER*	Mean costs (£)	QALYs	ICER*	Mean costs (€)	QALYs	ICER*
Do nothing	5.1	12863	11.620	-	1582	11.540	-	5983	13.866	-
CCTA - C	11.4	24341	11.841	Dominated	2879	11.761	Ext dominance	7675	14.136	Dominated
CCTA - I	13.2	21912	11.851	Dominated	3011	11.771	Dominated	7736	14.147	Dominated
CCTA + CMR - C	11.1	22648	11.849	Dominated	2911	11.770	**5787**	7479	14.146	**5356**
CCTA + CMR - I	12.5	21912	11.852	**38184**	3006	11.773	28340	7580	14.150	23764
CMR - C	5.5	26054	11.839	Dominated	3120	11.760	Dominated	7819	14.134	Dominated
CMR - I	7.8	22150	11.850	Dominated	3175	11.770	Dominated	7738	14.147	Dominated
CAG	13.2	23385	11.843	Dominated	3359	11.763	Dominated	8463	14.136	Dominated

US = United States, UK = United Kingdom, NL = The Netherlands, QALY = quality-adjusted life-years, ICER = incremental cost-effectiveness ratio, CCTA = coronary computed tomography coronary angiography, CMR = cardiac magnetic resonance imaging, CAG = catheter-based coronary angiography.

Table 3. Results of base case cost-effectiveness analysis (60-year old women, pre-test probability of 30%). I = invasive approach, C = conservative approach

	Radiation		US			UK			NL	
	(mSv)	Mean costs ($)	QALYs	ICER*	Mean costs (£)	QALYs	ICER*	Mean costs (€)	QALYs	ICER*
Do nothing	5.1	14181	12.107	-	1689	11.850	-	4990	14.850	-
CCTA - C	11.4	26403	12.339	Dominated	2995	12.075	Ext dominance	6737	15.140	Dominated
CCTA - I	13.2	23660	12.350	Dominated	3110	12.086	Dominated	6755	15.153	Dominated
CCTA + CMR - C	11.1	24541	12.347	Dominated	3020	12.083	**5697**	6524	15.149	**5118**
CCTA + CMR - I	12.5	23487	12.351	**38170**	3107	12.087	26624	6604	15.154	15943
CMR - C	5.5	28290	12.338	Dominated	3243	12.075	Dominated	6905	15.138	Dominated
CMR - I	7.8	23903	12.350	Dominated	3276	12.086	Dominated	6760	15.153	Dominated
CAG	13.2	25138	12.341	Dominated	3462	12.076	Dominated	7481	15.139	Dominated

US = United States, UK = United Kingdom, NL = The Netherlands, QALY = quality-adjusted life-years, ICER = incremental cost-effectiveness ratio, CCTA = coronary computed tomography coronary angiography, CMR = cardiac magnetic resonance imaging, CAG = catheter-based coronary angiography, c/w = compared with
** Boldface ICERs are calculated compared with the reference strategy "Do nothing". Other ICERs are compared with the next least costly strategy"*
† Strategy eliminated by extended dominance

were least expensive, however, the CCTA-invasive strategy was demonstrated to be cost-effective compared to the CCTA+CMR-invasive strategy (ICER €24514/QALY). For the US (not shown), the ICER for the CCTA-invasive strategy versus the CCTA+CMR-invasive strategy was $64119/QALY, which exceeded the US WTP-threshold of $50000/QALY. Consequently, the CCTA+CMR-invasive strategy remained optimal for the US. However, when statin costs were reduced at the same time, the ICER dropped to $46174/QALY.

Disutility of taking medication

If we assume that the disutility for taking medication is zero, the CCTA+CMR-conservative strategy became more effective compared to the CCTA+CMR-invasive strategy. Because the CCTA+CMR-conservative strategy was already less expensive, it dominated the CCTA+CMR-invasive strategy in the UK and NL analyses. For the US, CCTA+CMR-invasive remained optimal.

Discussion

Summary of key findings

Analysis of our decision model suggests that CCTA, if positive followed by CMR, is optimal for 60-year-old men and women presenting with chest pain and a prevalence of obstructive CAD below 70%. The CMR-invasive strategy was cost-effective for patients with a prevalence of obstructive CAD >70%-90%. As expected, the differences in terms of quality-adjusted life expectancy between the strategies were marginal. This is in part explained by our assumption that patients with false-negative test results have persisting symptoms and return to the hospital for additional work-up. Thus, the optimal strategy is mainly determined by costs. We performed a sensitivity analysis assuming that false-negatives remain false-negative for 3 and 5 years and found that the optimal strategy changed to CCTA-invasive, which is because CCTA has the lowest number of false-negative test results (highest sensitivity).

The CMR-strategies were most expensive, which is explained by the higher cost for CMR and the lower sensitivity and specificity which results in more false-positive and false-negative test results. False-positives (~14% of the population if pre-test probability = 30%) either receive unnecessary medical treatment, or undergo unnecessary CAG, which increases costs. Similarly, patients with false-negative test results temporarily forego the benefit of treatment (quality-of-life □, MACE □) and require additional testing (costs and disutility □) during follow-up. In the US, the conservative strategies were more expensive compared to the invasive strategies, explained by higher medication costs. In the conservative strategies, patients with mildly positive test results (including the false-positives) are not referred for CAG and subsequently treated with OMT (unnecessary for false-positives), which results in lifetime medication use and high costs.

For pre-test probabilities below 30% and 40% for UK men and women, the CCTA+CMR-invasive strategy was cost-effective, whereas the CCTA+CMR-conservative strategy was cost-effective up to a pre-test probability of 70%. This may seem counter-intuitive, but can be explained by the absolute numbers of false-positives, which are highest in numbers when the pre-test probability is low. The conservative approach implies lifelong OMT for the false-positives, whereas the invasive strategy results in referral to CAG and prevents un-

necessary medical therapy. Our analysis suggests that the latter is more cost-effective when the pre-test probability is low.

Previous publications

The 2010 NICE guidelines recommend functional non-invasive imaging in patients with chest pain and a 30-60% pre-test probability of CAD, whereas our model suggests that CCTA, if positive followed by CMR is cost-effective up to a pre-test probability of 70%. Above 70%, our model suggests that the CMR-invasive strategy is optimal, whereas the guidelines recommend CAG if the pre-test probability is 61-90%, and "manage as angina" if >90%.

Although the diagnostic performance of both CCTA and CMR has been studied extensively, literature on direct and long-term comparisons between CCTA and CMR is scarce. Nevertheless, the incremental prognostic value of perfusion imaging is well established, and complementary roles of CCTA and CMR in the diagnostic work-up for chest pain have been proposed (9).

Table 4. Two-way sensitivity analysis on the pre-test probability of obstructive CAD (the model was re-analyzed at discrete values for the pre-test probability, e.g. 10%, 20%, etc. until 90%), and the impact of false-negative test results

Base-case: false-negatives are identified within the 1st year

Pre-test probability		10%	20%	30%	40%	50%	60%	70%	80%	90%
UK	Men	CCTA+CMR-I		CCTA+CMR-C				CMR-I		CMR-C
	Women	CCTA+CMR-I		CCTA+CMR-C				CMR-I		CAG
US	Men	CCTA+ CMR-I							CMR-I	
	Women	CCTA+CMR-I							CMR-I	
NL	Men	CCTA-CMR-I								
	Women	CCTA-CMR-I							CMR-I	

Alternative assumption: false-negatives are identified within 3 years

Pre-test probability		10%	20%	30%	40%	50%	60%	70%	80%	90%
UK	Men	CCTA+ CMR-I		CCTA-I					CAG	
	Women	CCTA+ CMR-I		CCTA-I					CAG	
US	Men	CCTA+CMR-I							CMR-I	
	Women	CCTA+CMR-I							CMR-I	
NL	Men	CCTA-CMR-I				CCTA-I			CMR-I	CAG
	Women	CCTA-CMR-I				CCTA-I				

A previously published cost-effectiveness analysis compared several CCTA-based strategies with myocardial perfusion SPECT and direct CAG, and found that the CCTA-based strategies were optimal up to a prevalence of obstructive CAD of 80% (48). However, other researchers found that US Medicare costs for patients who underwent CCTA were higher compared to patients who underwent stress testing (49).

Limitations

Some limitations of our decision model deserve attention. First, we analyzed five diagnostic strategies from the perspective of the UK, US, and NL. Health-care costs can vary considerably across countries, which is why we used country-specific cost estimates. Although we recognize that diagnostic strategies and treatment decisions may also vary across countries (e.g. due to differences in guidelines and local practices), these model characteristics were held constant to be able to make a fair comparison between the country-specific results.

Second, we assumed that all patients with severe CAD (or moderate CAD with severe ischemia) undergo revascularization. For clarity and simplicity, we did not take into account exceptions to the rule. For example in patients with diabetes, CABG may be preferred over PCI. Furthermore, we assumed that the relative benefit of treatment (adjusted hazard rate ratio) is the same for OMT, PCI, and CABG. However, our simplifications with regard to treament effects and quality-of-life affect all diagnostic strategies to the same extent, approximately. Therefore, these assumptions are unlikely to change the optimal decision.

Third, we assumed that all false-positive CMR results were due to artifacts occuring in patients without obstructive CAD. However, perfusion defects in women without obstructive CAD on CAG may represent "real" ischemia (as opposed to artifacts) caused by small vessel disease and microvascular dysfunction.

Fourth, we focused on the visualization of myocardial perfusion abnormalities, whereas CMR can also be used to assess regional wall motion abnormalities. Including information on wall motion generally results in a higher specificity for diagnosing obstructive CAD (14). However, when we increased the specificity of CMR to 90% in a sensitivity analysis, our conclusion did not change.

Fifth, patients with false-positive test results in conservative strategies are not referred for CAG and are assumed to receive OMT for their remaining life-expectancy. Using a disutility for taking medication, we adjusted the quality-of-life for patients with false-positive test results. On the other hand, potential benefits of OMT in patients without obstructive CAD were not considered. We performed a sensitivity analysis, and found that when the disutility of taking medication was zero, the CCTA+CMR-conservative strategy was preferable over the CCTA+CMR-invasive strategy for the UK and NL. For the US, CCTA+CMR-invasive remained optimal, which is explained by the higher costs for medical therapy.

Generalizability

Our model analyzed diagnostic strategies for patients with stable chest pain without a history of CAD, myocardial infarction, PCI, or CABG, which limits the population to which our results can be generalized. Furthermore, we did not consider the possibility of contra-indications for either CCTA or CMR (e.g. renal insufficiency, contrast allergy, claustrophobia). This implies that our results apply to patients in whom both tests are considered safe and appropriate.

Future research

In the current analysis, we only considered CT coronary angiography and stress CMR for the diagnostic work-up of patients with chest pain. Other strategies that involve stress SPECT, echocardiography, and positron-emission tomography could also be considered. Recent technical developments, such as measurement of myocardial perfusion and fractional flow measurement by CT may be of interest for future cost-effectiveness analysis.

Conclusion

Quality-adjusted life expectancy was very similar across strategies. For the UK, US, and NL, the CCTA+CMR strategy was cost-effective for men and women with stable chest pain and a pre-test probability up to 70%, 80%, and 90%, respectively. Above these thresholds, the CMR-strategy was cost-effective.

Acknowledgement

The authors would like to thank Sotiris Antoniou, Consultant Pharmacist, Cardiovascular Medicine at the Barts and The London NHS Trust, for the useful discussions regarding costs and medication use.

Appendix

Assumptions

- Revascularizations during follow-up: 75% PCI, 25% CABG. Assumption based on Erasmus MC cohort (50), and consistent with observations from COURAGE (19) and SYNTAX (18).

- Travel costs per hospital visit UK: £ 0 (health care perspective) US: $ 25 (guestimate), NL:€ 6 (34).

- We assumed conditional independence with regard to the sensitivity and specificity for CCTA and CMR.

- Proportion of background mortality that was considered non-cardiac was calculated based on Dutch life tables. The following International Classification of Disease (ICD) codes were considered as cardiac deaths: I-11, I-20 – I-25, I-42, I-44 – I-50, R-96.

- Harmful effects of radiation exposure were not modeled. Cumulative (life-time) radiation exposure is reported in Table 2 and 3.

- Patients with Moderate CAD without ischemia and a negative test result did not forgo benefit of treatment.

- Patients with false-negative test results (see Methods) return to their physician within the 1st year and undergo additional testing. Additional testing consisted of CMR in the CCTA strategy, and CCTA in the CMR strategy. Subsequent additional testing was equal to the diagnostic strategies as described (invasive vs. conservative).

- Progression of disease was modeled through the major adverse cardiac event rates (MACE).

- Medication use depends on the CCTA findings and the treatment assigned (see Appendix Table 2). In the CMR strategy, the distinction between normal coronary arteries, non-obstructive disease and mild CAD cannot be made for patients without perfusion defects on CMR. For those patients, we assumed that baseline medication is maintained.

- For all diagnostic strategies, the prognosis for patients without CAD, mild CAD, and moderate CAD without ischemia, was based on a meta-analysis on the prognostic value of CCTA. CCTA findings are likely to alter medical management in the patients included in the meta-analysis, which in turn may benefit prognosis. This issue may have overestimated the event-free survival for patients who undergo CMR only, have a normal result and maintain baseline medication.

- For patients with severe CAD and 3VD/LM, we assumed that 33% has mild ischemia, and 67% has severe ischemia.

Appendix Table 1. Quality-of-life (EQ-5D utilities) and annual rates of major adverse cardiac events (revascularization, non-fatal myocardial infarction, and cardiac death). Age- and sex-specific mortality (non-cardiac) was modeled separately based on country-specific vital statistics. Quality-of-life estimates are shown for 60-year old men and women only, but were modeled using age-specific estimates to account for increasing age over time.

	On treatment	On treatment	Without treatment	Source
	1st year	Subsequent years		
Normal coronary arteries				
MACE	0.0008	0.0008	0.0008	Meta-analysis prognostic value CCTA (20)
Utility in 60-year old men	0.851	0.851	0.851	General population (25)
Utility in 60-year old women	0.824	0.824	0.824	General population (25)
Mild-moderate CAD (no ischemia)				
MACE	0.025	0.025	0.025	Meta-analysis prognostic value CCTA (20)
Utility in 60-year old men	0.851	0.851	0.851	General population (25)
Utility in 60-year old women	0.824	0.824	0.824	General population (25)
Moderate-severe 1- or 2-vessel CAD (mild ischemia) → OMT				
MACE	0.172	0.071	Rate * 1/HRR	COURAGE trial (19)
Utility in 60-year old men	0.734	0.749	0.699	Relative decrement based on COURAGE (26)
Utility in 60-year old women	0.711	0.726	0.677	Relative decrement based on COURAGE (26)
Severe 1- or 2-vessel CAD (severe ischemia) → PCI				
MACE	0.110	0.043	Rate * 1/HRR	COURAGE trial (19)
Utility in 60-year old men	0.740	0.760	0.699	Relative decrement based on COURAGE (26)
Utility in 60-year old women	0.716	0.736	0.677	Relative decrement based on COURAGE (26)
3VD/LM (mild/severe ischemia) → CABG				
MACE	0.096	0.031	Rate * 1/HRR	SYNTAX trial (17-18)
Utility in 60-year old men	0.740	0.820	0.659	Relative decrement based on SYNTAX (27)
Utility in 60-year old women	0.716	0.794	0.638	Relative decrement based on SYNTAX (27)

MACE = major adverse cardiac event, CCTA = coronary computed tomography angiography, CAD = coronary artery disease, OMT = optimal medical treatment, HRR = hazard rate ratio of the treatment benefit, assumed to be 0.70 (range used in sensitivity analysis: 0.6 – 0.9)

Appendix Table 2. Medication use

Treatment	**Baseline**	**Normal coronary arteries**	**Mild CAD**	**Moderate CAD without ischemia**	**OMT**	**PCI**	**CABG**
Platelet inh. (%) Aspirin	48	12	32	73	95	95	83
Statin (%) Simvastatin	22	17	31	72	92	93	86
B-blocker (%) Atenolol	37	17	16	40	86	84	77
Nitrates (%) Isosorbide mononitrate	-	1	5	11	61	47	8*
ACE-inhibitor (%) Enalapril	-	7	11	27	62	64	53
Source	Euro Heart Survey (51)	Cohort Erasmus MC (50, 52)	Cohort Erasmus MC (50, 52)	Cohort Erasmus MC (50, 52)	COURAGE trial (19) OMT-arm at 3 yrs	COURAGE trial (19) PCI-arm at 3 yrs	SYNTAX trial (17) CABG-arm at 3 yrs
Annual costs							
US $	503	391	669	1569	2453	2423	2066
UK £	11	4	16	37	113	96	42
NL €	72	34	61	142	288	268	189

CAD = coronary artery disease, Rf Tx = risk factor treatment, OMT = optimal medical treatment, PCI = percutaneous coronary intervention, CABG = coronary artery bypass graft surgery, NHS = National Health Service, CVZ = Dutch Health Care Insurance Board
** At 1 year (27)*

Appendix Table 3. Model parameters, point estimates, and distributions

Parameter	**Distribution Type**	**Mean**	**Lower CI**	**Upper CI**
CABG costs (NL)	Gamma	11892	7709	16959
CABG costs (UK)	Gamma	7310	4726	10442
CABG costs (US)	Gamma	38147	24698	54446
CABG mortality	Beta	0.018	0.014	0.022
CAG costs (NL)	Gamma	1513	978	2154
CAG costs (UK)	Gamma	1052	679	1505
CAG costs (US)	Gamma	2990	1938	4266
CAG mortality	Beta	0.0011	0.0009	0.0014
CAG radiation	Gamma	7	5	10
CAG rate of MI	Beta	0.0005	0.0003	0.0007
CCTA costs (NL)	Gamma	215	139	307
CCTA costs (UK)	Gamma	291	189	415
CCTA costs (US)	Gamma	372	240	532
CCTA mortality	Beta	0.00001	0.00000	0.00002
CCTA radiation	Gamma	5	3	7
CCTA sensitivity	Beta	0.98	0.95	1.00
CCTA specificity	Beta	0.89	0.83	0.94
Disutility - taking medication	Triangular	0.01	0.00	0.01
Disutility - uncertainty	Gamma	1.00	0.65	1.43
FFR costs (NL)	Gamma	555	359	793
FFR costs (UK)	Gamma	460	298	658
FFR costs (US)	Gamma	715	462	1019
Incidental finding - probability	Beta	0.07	0.06	0.09
Incidental finding costs (NL)	Gamma	433	281	620
Incidental finding costs (UK)	Gamma	359	233	512

Incidental finding costs (US)	Gamma	559	361	798
Medication costs - uncertainty	Gamma	1.00	0.65	1.43
MI costs (NL)	Gamma	8154	5291	11632
MI costs (UK)	Gamma	5197	3366	7414
MI costs (US)	Gamma	10204	6618	14608
MI Hazard ratio	Log-Normal	1.44	1.25	1.66
MRI costs (NL)	Gamma	319	207	455
MRI costs (UK)	Gamma	548	355	785
MRI costs (US)	Gamma	689	446	985
MRI sensitivity	Beta	0.90	0.87	0.93
MRI specificity	Beta	0.80	0.76	0.84
PCI costs (NL)	Gamma	4163	2686	5946
PCI costs (UK)	Gamma	3674	2384	5248
PCI costs (US)	Gamma	6533	4218	9321
PCI mortality	Beta	0.011	0.008	0.015
PCI radiation	Gamma	15	10	21
Quality of life - uncertainty	Gamma	1.00	0.65	1.43
Travel costs (NL)	Gamma	5	3	7
Travel costs (US)	Gamma	25	16	36
Treatment benefit foregone	Log-Normal	1.41	1.09	1.80

Appendix Table 4. Results of base case cost-effectiveness analysis (60-year old men, pre-test probability of 30%), including 95% confidence interval based on 10000 samples from the parameter distributions (1000 trials per sample). I = invasive approach, C = conservative approach

	US		UK		NL	
	Mean costs, $	QALYs	Mean costs, £	QALYs	Mean costs, €	QALYs
Do nothing	12863 (11263-14807)	11.620 (11.308-11.925)	1574 (1186-2041)	11.551 (11.277-11.825)	5983 (5233-6801)	13.866 (13.492-14.248)
CCTA - C	24341 (17752-32440)	11.841 (11.529-12.150)	2862 (2325-3480)	11.770 (11.487-12.050)	7675 (6605-8829)	14.136 (13.756-14.520)
CCTA - I	21912 (16457-28546)	11.851 (11.540-12.160)	2991 (2461-3599)	11.779 (11.502-12.058)	7736 (6717-8811)	14.147 (13.768-14.533)
CCTA + CMR - C	22648 (16683-29976)	11.849 (11.536-12.157)	2893 (2365-3497)	11.777 (11.494-12.055)	7479 (6446-8575)	14.146 (13.774-14.530)
CCTA + CMR - I	21912 (16286-28443)	11.852 (11.540-12.164)	2987 (2459-3598)	11.781 (11.499-12.046)	7580 (6580-8658)	14.150 (13.771-14.525)
CMR - C	26054 (18991-34741)	11.839 (11.533-12.146)	3104 (2531-3758)	11.769 (11.490-12.041)	7819 (6725-9006)	14.134 (13.755-14.522)
CMR - I	22150 (16646-28832)	11.850 (11.542-12.161)	3156 (2604-3797)	11.779 (11.501-12.054)	7738 (6718-8825)	14.147 (13.767-14.536)
CAG	23385 (17788-30123)	11.843 (11.528-12.151)	3340 (2706-4058)	11.771 (11.496-12.050)	8463 (7327-9682)	14.136 (13.757-14.514)

US = United States, UK = United Kingdom, NL = The Netherlands, QALY = quality-adjusted life-years, ICER = incremental cost-effectiveness ratio, CCTA = coronary computed tomography coronary angiography, CMR = cardiac magnetic resonance imaging, CAG = catheter-based coronary angiography.

Appendix Table 5. Results of base case cost-effectiveness analysis (60-year old women, pre-test probability of 30%), including 95% confidence interval based on 10000 samples from the parameter distributions (1000 trials per sample). I = invasive approach, C = conservative approach

	US		UK		NL	
	Mean costs, $	QALYs	Mean costs, £	QALYs	Mean costs, €	QALYs
Do nothing	14181 (12425-16309)	12.107 (11.809-12.401)	1689 (1287-2176)	11.850 (11.588-12.117)	4990 (4321-5764)	14.850 (14.476-15.218)
CCTA - C	26403 (19086-35242)	12.339 (12.048-12.626)	2995 (2431-3634)	12.075 (11.811-12.334)	6737 (5658-7936)	15.140 (14.774-15.508)
CCTA - I	23660 (17602-30921)	12.350 (12.058-12.639)	3110 (2561-3743)	12.086 (11.819-12.346)	6755 (5768-7861)	15.153 (14.787-15.523)
CCTA + CMR - C	24541 (17950-32480)	12.347 (12.059-12.637)	3020 (2480-3645)	12.083 (11.821-12.343)	6524 (5517-7647)	15.149 (14.779-15.515)
CCTA + CMR - I	23487 (17488-30740)	12.351 (12.060-12.643)	3107 (2561-3737)	12.087 (11.825-12.346)	6604 (5635-7697)	15.154 (14.790-15.525)
CMR - C	28290 (20412-37943)	12.338 (12.044-12.624)	3243 (2649-3905)	12.075 (11.818-12.335)	6905 (5779-8144)	15.138 (14.768-15.509)
CMR - I	23903 (17843-31186)	12.350 (12.055-12.642)	3276 (2704-3931)	12.086 (11.827-12.349)	6760 (5753-7869)	15.153 (14.780-15.524)
CAG	25138 (19021-32551)	12.341 (12.045-12.630)	3462 (2815-4187)	12.076 (11.813-12.336)	7481 (6352-8677)	15.139 (14.773-15.501)

US = United States, UK = United Kingdom, NL = The Netherlands, QALY = quality-adjusted life-years, ICER = incremental cost-effectiveness ratio, CCTA = coronary computed tomography coronary angiography, CMR = cardiac magnetic resonance imaging, CAG = catheter-based coronary angiography

References

1. Fox K, Garcia MA, Ardissino D, et al. Guidelines on the management of stable angina pectoris: executive summary - the Task Force on the Management of Stable Angina Pectoris of the European Society of Cardiology. Eur Heart J 2006; 27:1341-1381.
2. Fraker TD, Jr., Fihn SD, Gibbons RJ, et al. 2007 chronic angina focused update of the ACC/AHA 2002 Guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines Writing Group to develop the focused update of the 2002 Guidelines for the management of patients with chronic stable angina. Circulation 2007; 116:2762-2772.
3. Gibbons RJ, Balady GJ, Bricker JT, et al. ACC/AHA 2002 guideline update for exercise testing: summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines). Circulation 2002; 106:1883-1892.
4. Kwok Y, Kim C, Grady D, Segal M, Redberg R. Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol 1999; 83:660-666.
5. Cooper A, Calvert N, Skinner J, et al. Chest pain of recent onset: Assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin. London: National Clinical Guideline Centre for Acute and Chronic Conditions, 2010.
6. Mowatt G, Cook JA, Hillis GS, et al. 64-slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. Heart 2008; 94:1386-1393.
7. Schuetz GM, Zacharopoulou NM, Schlattmann P, Dewey M. Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. Ann Intern Med 2010; 152:167-177.
8. von Ballmoos MW, Haring B, Juillerat P, Alkadhi H. Meta-analysis: Diagnostic Performance of Low-Radiation-Dose Coronary Computed Tomography Angiography. Ann Intern Med 2011; 154:413-420.
9. Groothuis JG, Beek AM, Brinckman SL, et al. Low to intermediate probability of coronary artery disease: comparison of coronary CT angiography with first-pass MR myocardial perfusion imaging. Radiology 2010; 254:384-392.
10. Meijboom WB, Van Mieghem CA, van Pelt N, et al. Comprehensive assessment of coronary artery stenoses: computed tomography coronary angiography versus conventional coronary angiography and correlation with fractional flow reserve in patients with stable angina. J Am Coll Cardiol 2008; 52:636-643.
11. Kurita T, Sakuma H, Onishi K, et al. Regional myocardial perfusion reserve determined using myocardial perfusion magnetic resonance imaging showed a direct correlation with coronary flow velocity reserve by Doppler flow wire. Eur Heart J 2009; 30:444-452.
12. Greenwood JP, Maredia N, Younger JF, et al. Cardiovascular magnetic resonance and single-photon emission computed tomography for diagnosis of coronary heart disease (CE-MARC): a prospective trial. Lancet 2011.
13. Hamon M, Fau G, Nee G, Ehtisham J, Morello R. Meta-analysis of the diagnostic performance of stress perfusion cardiovascular magnetic resonance for detection of coronary artery disease. J Cardiovasc Magn Reson 2010; 12:29.
14. Nandalur KR, Dwamena BA, Choudhri AF, Nandalur MR, Carlos RC. Diagnostic performance of stress cardiac magnetic resonance imaging in the detection of coronary artery disease: a meta-analysis. J Am Coll Cardiol 2007; 50:1343-1353.
15. Schwitter J, Wacker CM, van Rossum AC, et al. MR-IMPACT: comparison of perfusion-cardiac magnetic resonance with single-photon emission computed tomography for the detection of coronary artery disease in a multicentre, multivendor, randomized trial. Eur Heart J 2008; 29:480-489.
16. Machaalany J, Yam Y, Ruddy TD, et al. Potential clinical and economic consequences of noncardiac incidental findings on cardiac computed tomography. J Am Coll Cardiol 2009; 54:1533-1541.
17. Kappetein AP, Feldman TE, Mack MJ, et al. Comparison of coronary bypass surgery with drug-eluting stenting for the treatment of left main and/or three-vessel disease: 3-year follow-up of the SYNTAX trial. Eur Heart J 2011; 32:2125-2134.
18. Serruys PW, Morice MC, Kappetein AP, et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. N Engl J Med 2009; 360:961-972.
19. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 2007; 356:1503-1516.
20. Hulten EA, Carbonaro S, Petrillo SP, Mitchell JD, Villines TC. Prognostic value of cardiac computed tomography angiography: a systematic review and meta-analysis. J Am Coll Cardiol 2011; 57:1237-1247.
21. Arias E. United States Life Tables 2007. Centers for Disease Control and Prevention. National Center for Health Statistics. http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59_09.pdf. Accessed 15 november 2011
22. UK Interim Life Tables 2008-2010. Office for National Statistics. http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Interim+Life+Tables. Accessed 15 november 2011
23. Dutch Central Bureau of Statistics. Life tables 2010. http://statline.cbs.nl/. Accessed 15 november 2011
24. Deckers JW, Goedhart DM, Boersma E, et al. Treatment benefit by perindopril in patients with stable coronary artery disease at different levels of risk. Eur Heart J 2006; 27:796-801.
25. Hanmer J, Lawrence WF, Anderson JP, Kaplan RM, Fryback DG. Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. Med Decis Making 2006; 26:391-400.
26. Weintraub WS, Spertus JA, Kolm P, et al. Effect of PCI on quality of life in patients with stable coronary disease. N Engl J Med 2008; 359:677-687.
27. Cohen DJ, Van Hout B, Serruys PW, et al. Quality of life after PCI with drug-eluting stents or coronary-artery bypass surgery. N Engl J Med 2011; 364:1016-1026.
28. Ara R, Brazier J. Deriving an algorithm to convert the eight mean SF-36 dimension scores into a mean EQ-5D preference-based score from published studies (where patient level data are not available). Value Health 2008; 11:1131-1143.
29. Pletcher MJ, Lazar L, Bibbins-Domingo K, et al. Comparing impact and cost-effectiveness of primary prevention strategies for lipid-lowering. Ann Intern Med 2009; 150:243-254.
30. Gold MR, et al. Cost-effectiveness in Health and Medicine. Journal of Mental Health Policy and Economics 1996; 2:91-92.
31. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996; 276:1339-1341.
32. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. JAMA 1996; 276:1253-1258.
33. National Institute for Clinical Excellence (NICE). Guide to the methods of technology appraisal. Oxford, England: Radcliff Medical Press, 2004.
34. Hakkaart-van Roijen L, Tan SS, Bouwmans C. Dutch Manual for Cost Analysis. Dutch Health Care Insurance Board 2010.
35. Dutch Council for Public Health and Health Care. Sensible and Sustainable Care. http://rvz.net/uploads/docs/Sensible_and_sustainable_care.pdf. Accessed December 2011
36. Katayama H, Yamaguchi K, Kozuka T, Takashima T, Seez P, Matsuura K. Adverse reactions to ionic and nonionic contrast media. A report from the Japanese Committee on the Safety of Contrast Media. Radiology 1990; 175:621-628.
37. Mettler FA, Jr., Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. Radiology 2008; 248:254-263.

38. Noto TJ, Jr., Johnson LW, Krone R, et al. Cardiac catheterization 1990: a report of the Registry of the Society for Cardiac Angiography and Interventions (SCA&I). Cathet Cardiovasc Diagn 1991; 24:75-83.
39. Scanlon PJ, Faxon DP, Audet AM, et al. ACC/AHA guidelines for coronary angiography. A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on Coronary Angiography). Developed in collaboration with the Society for Cardiac Angiography and Interventions. J Am Coll Cardiol 1999; 33:1756-1824.
40. Fearon WF, Bornschein B, Tonino PA, et al. Economic evaluation of fractional flow reserve-guided percutaneous coronary intervention in patients with multivessel disease. Circulation 2010; 122:2545-2550.
41. Genders TS, Meijboom WB, Meijs MF, et al. CT coronary angiography in patients suspected of having coronary artery disease: decision making from various perspectives in the face of uncertainty. Radiology 2009; 253:734-744.
42. Bravata DM, Gienger AL, McDonald KM, et al. Systematic review: the comparative effectiveness of percutaneous coronary interventions and coronary artery bypass graft surgery. Ann Intern Med 2007; 147:703-716.
43. NZA. Dutch National Authority of Health Care. http://dbc-tarieven.nza.nl/Nzatarieven/top.do. Accessed 25 November 2011
44. HCUPnet. Healthcare Cost and Utilization Project (HCUP). http://hcupnet.ahrq.gov/. . Rockville, MD: Agency for Healthcare Research and Quality, 2006.
45. Department of Health. NHS Electronic Drug Tariff. http://www.ppa.org.uk/ppa/edt_intro.htm. Accessed 10 November 2011
46. Dutch Health Care Insurance Board / Medication Information Project [In Dutch]. http://www.gipdatabank.nl/ 2009.
47. Lazar LD, Pletcher MJ, Coxson PG, Bibbins-Domingo K, Goldman L. Cost-effectiveness of statin therapy for primary prevention in a low-cost statin era. Circulation 2011; 124:146-153.
48. Min JK, Gilmore A, Budoff MJ, Berman DS, O'Day K. Cost-effectiveness of coronary CT angiography versus myocardial perfusion SPECT for evaluation of patients with chest pain and no known coronary artery disease. Radiology 2010; 254:801-808.
49. Shreibati JB, Baker LC, Hlatky MA. Association of coronary CT angiography or stress testing with subsequent utilization and spending among Medicare beneficiaries. JAMA 2011; 306:2128-2136.
50. Dedic A, Genders TS, Ferket BS, et al. Stable angina pectoris: head-to-head comparison of prognostic value of cardiac CT and exercise testing. Radiology 2011; 261:428-436.
51. Daly CA, Clemens F, Sendon JL, et al. The initial management of stable angina in Europe, from the Euro Heart Survey: a description of pharmacological management and revascularization strategies initiated within the first month of presentation to a cardiologist in the Euro Heart Survey of Stable Angina. Eur Heart J 2005; 26:1011-1022.
52. Nieman K, Galema T, Weustink A, et al. Computed tomography versus exercise electrocardiography in patients with stable chest complaints: real-world experiences from a fast-track chest pain clinic. Heart 2009; 95:1669-1675.

Part V

Other

Chapter 12.1

Stable Angina Pectoris: Head-to-head Comparison of the Prognostic Value of Cardiac CT and Exercise Testing

Radiology. 2011 Nov;261(2):428-36.

Admir Dedic
Tessa S.S. Genders
Bart S. Ferket
Tjebbe W. Galema
Nico R.A. Mollet
Adriaan Moelker
M.G. Myriam Hunink
Pim J. de Feyter
Koen Nieman

Abstract

Purpose

To determine and compare the prognostic value of cardiac computed tomography angiography (CCTA), coronary calcium scoring, and exercise electrocardiography (ECG) in patients with chest pain who are suspected of having coronary artery disease (CAD).

Methods

This study complied with the Declaration of Helsinki, and the local ethics committee approved the study. Patients (n = 471) without known CAD underwent exercise ECG and dual-source CCTA at a 1-day chest pain clinic. Coronary calcification and the presence of 50% or greater coronary stenosis (in one or more vessels) were assessed with CCTA. Exercise ECG results were classified as normal, ischemic, or non-diagnostic. The primary outcome was a major adverse cardiac event (MACE), defined as cardiac death, non-fatal myocardial infarction, or unstable angina requiring hospitalization and revascularization beyond 6 months. Univariable and multivariable Cox regression analysis was used to determine the prognostic values, while clinical impact was assessed with the net reclassification improvement metric.

Results

Follow-up was completed for 424 (90%) patients; the mean duration of follow-up was 2.6 years. A total of 44 MACEs occurred in 30 patients. Four of the MACEs were cardiac deaths and six were non-fatal myocardial infarctions. The presence of coronary calcification (hazard ratio [HR], 8.22 [95% confidence interval (CI): 1.96-34.51]), obstructive CAD (HR, 6.22 [95% CI: 2.77-13.99]), and non-diagnostic stress test results (HR, 3.00 [95% CI: 1.26-7.14]) were univariable predictors of MACEs. In the multivariable model, CCTA findings (HR, 5.0 [95% CI: 1.7-14.5]) and non-diagnostic exercise ECG results (HR, 2.9 [95% CI: 1.2-7.0]) remained independent predictors of MACEs. CCTA findings showed incremental value beyond clinical predictors and stress testing (global χ^2, 37.7 vs.13.7; $P < 0.001$), whereas coronary calcium scores did not have further incremental value (global χ^2, 38.2 vs. 37.7; $P = 0.40$).

Conclusion

CT angiography findings are a strong predictor of future adverse events, showing incremental value over clinical predictors, stress testing, and coronary calcium scores.

Introduction

Over the past decade, coronary computed tomography angiography (CCTA) has emerged as a valuable diagnostic tool to evaluate coronary artery disease (CAD) and has found use as an anatomic alternative to functional testing and a non-invasive alternative to conventional (catheter-based) coronary angiography (1,2). Functional assessment of CAD severity with exercise electrocardiography (ECG) is well established, and its prognostic value has been studied extensively (3,4). Although evidence regarding the prognostic value of coronary CT angiography findings is emerging (5–7), the question remains if this angiographic modality holds incremental prognostic value beyond functional evaluation with exercise ECG. In this study, in a prospectively enrolled population, we sought to determine and compare the prognostic value of coronary CT angiography, coronary calcium scoring (CCS), and exercise ECG in patients with chest pain who were suspected of having CAD.

Materials and Methods

From September 2006 to December 2008, 471 consecutive patients without a history of CAD were evaluated at our 1-day chest pain clinic (mean age, 56 years ± 10 [standard deviation]; 244 male patients) (8). They had been referred by their general practitioner because of chronic complaints of chest pain potentially caused by CAD to undergo additional testing with exercise ECG, coronary CT angiography, and CCS. Information on risk factors was prospectively acquired, and clinical risk estimators— that is, the SCORE (9) and Diamond and Forrester metrics (10)—were calculated from these data. According to the SCORE, 291 patients had a 10-year risk of less than 5% of developing fatal cardiovascular disease, 130 patients had a risk of 5%–10%, and 50 patients had a risk of greater than 10%. The study complied with the Declaration of Helsinki, and the local ethics committee approved the study.

CCS Protocol

Calcium detection was performed with an ECG-triggered step-and-shoot acquisition mode, by using a 120-kV tube voltage, a mean tube current of 78 mAs ± 26, and a section thickness of 3 mm. The coronary calcium score was assessed by using dedicated software (Syngo CaScore; Siemens, Forchheim, Germany) and was quantified with the Agatston method with a standard 130-HU attenuation threshold (11). To account for the skewed distribution, all calcium scores were transformed by taking the natural logarithm of the calcium score +1.

Coronary CT Angiographic Parameters

Contrast material–enhanced dual source multi-section CT (Definition; Siemens) was performed by using the following parameters: collimation, 32 x 0.6 mm; 64-channel acquisition by z-axis focal spot alternation; rotation time, 330 msec; temporal resolution, 83 msec; spiral acquisition mode with prospectively ECG-triggered tube modulation depending on the heart rate regularity; tube voltage, 120 kV; tube current, 380–412 mAs depending on patient size; and variable pitch depending on the heart rate. Iopromide 70–100 mL (Ultravist, 370 milligrams of iodine per milliliter; Schering, Berlin, Germany), followed by a 40-mL saline bolus chaser, was peripherally injected at 5.0–5.5 mL/sec. Patients received sublingual nitroglycerin before the examination but no additional β-blockers. Effective radiation doses for CCS and coronary CT angiography were 0.8 mSv ± 0.2 (range, 0.4–1.6 mSv) and 11.0 mSv ± 3.5 (range, 4.7–17.8 mSv), respectively.

The coronary arteries were evaluated on axial images, multiplanar reconstructions, and maximum intensity projections according to readers' preferences. Readers were blinded to patients' symptoms and exercise ECG results. Vessels were qualitatively scored as significantly stenosed (≥50% diameter narrowing) or less than significantly stenosed (<50%) or free from plaque. To assess the prognostic value of coronary CT angiography, we constructed a model comparing obstructive plaque to non-obstructive plaque. In a second model, we assessed the prognostic value of obstructive and non-obstructive plaque versus no plaque.

Exercise ECG

Patients underwent stress testing on a bicycle ergometer by using a standardized protocol. Use of any □-blockers was ceased 72 hours before stress testing, if possible. During continuous ECG registration and 12-lead printing at 1-minute intervals, the workload was increased from 40 W in 20-W increments at 1-minute intervals. Blood pressure was measured every 2 minutes. In cases of established contraindications, patients did not undergo exercise ECG (12). Criteria for myocardial ischemia included horizontal or down sloping ST-segment depression or elevation of 0.1 mV or greater during or after exercise or typical, increasing angina during exercise. A non-diagnostic test was defined as when there was an inability to perform the test or the test was discontinued without evidence of myocardial ischemia before reaching 85% of the predicted maximum heart rate (12). Additionally, for every patient, we approximated the Duke Treadmill Score by using a standardized formula as follows: DE – (5 STD) – (4 TAI), where DE is duration of exercise in minutes, STD is maximum net ST-segment deviation during or after exercise in millimeters, and TAI is treadmill angina index (where a score of 0 = no complaints, a score of 1 = non-limiting angina, and a score of 2 = angina requiring discontinuation) (3).

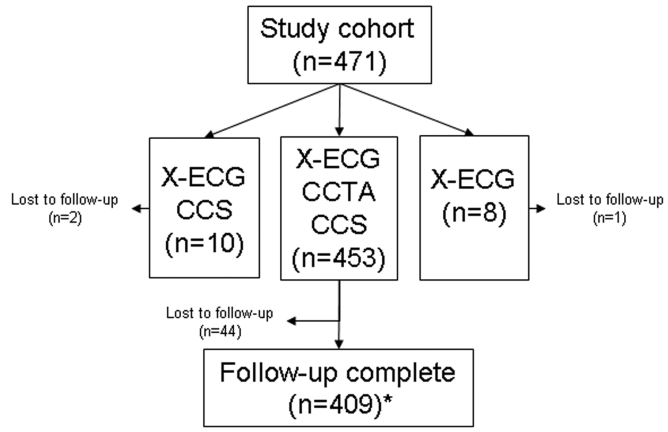
Follow-up

Follow-up data were obtained by consulting the national death registry for the occurrence of mortality and through standardized telephone interviews, questionnaires, or hospital visits. All events were confirmed with death records, hospital records, or correspondence with treating physicians and hospitals. If patients experienced multiple events during follow-up, only the first event counted. Patients lost to follow-up were censored. All events were reviewed by an independent cardiologist blinded to both coronary CT angiography and exercise ECG results.

Outcome Measures

The primary outcome measure was a composite of cardiac death, non-fatal myocardial infarction, unstable angina requiring hospitalization, and coronary revascularization (percutaneous coronary intervention or coronary artery bypass graft surgery). Cardiac death was defined as death caused by acute myocardial infarction, ventricular arrhythmia, refractory heart failure, or cardiogenic shock. Non-fatal myocardial infarction was defined as ischemia resulting in abnormal cardiac biomarkers (>99th percentile of the upper limits of normal). Unstable angina was defined as chest pain with altered frequency or character that was suspicious for acute coronary syndrome; a diagnosis of unstable angina also required ischemic ECG changes or significantly obstructive CAD at invasive conventional angiography (13). Coronary revascularizations initiated as a result of the initial diagnostic work-up and performed within 6 months were not considered as an end point.

Figure 1. Flowchart of patients included in the survival analysis. * Included in multivariable analysis. CCTA = coronary CT angiography, X-ECG = exercise ECG.



Statistical Analysis

Statistical analyses were performed by using SPSS (version 15.0; SPSS, Chicago, Ill) and STATA (version 11.1; Stata, College Station, Tex). All probability values refer to two-tailed tests of significance; $P < 0.05$ was considered to indicate a significant difference. Categorical variables are presented as proportions. Continuous variables are expressed as means ± standard deviations or as medians ± interquartile ranges (IQRs) as appropriate. Clinical patient characteristics are summarized by the SCORE 10-year cardiovascular mortality risk (9).

Univariable Cox regression analyses of each potential clinical predictor, as well as coronary calcification, coronary CT angiography results, and exercise ECG results, were performed to evaluate their effect in predicting major adverse cardiac events (MACEs) during follow-up. Cumulative event rates as a function over time were estimated according to the Kaplan-Meier method. Unadjusted comparison of survival between groups defined by the two coronary CT angiography results, coronary calcium categories, and exercise ECG results was performed by using the log-rank test. Annual event rates were calculated by dividing the cumulative event rates by the median number of follow-up years.

A multivariable Cox proportional hazards model was used to assess the independent prognostic value of exercise ECG, coronary CT angiography, and CCS. Risk adjustment for clinical characteristics was performed by using patients' SCORE results. To assess the incremental value of exercise ECG, coronary CT angiography, and CCS, the global χ^2 values were compared between models with and those without the addition of each incrementally. Using logistic regression, receiver operating characteristic curves with estimates of the area under the receiver operating characteristic curve (AUC, or the C-statistic) were obtained to compare the discriminative performance of the different models.

To quantify the clinical impact of adding exercise ECG, coronary CT angiography, and CCS to the model predicting MACE, we calculated the net reclassification improvement (NRI), which is a measure of correctly reclassified subjects that is penalized for those incorrectly classified. In this context, clinically relevant cutoff values for the risk of a MACE (i.e., the thresholds of risk where reclassification to another category would influence clinical man-

agement) are absent. Thus we used an extension of the traditional NRI that was recently introduced by Pencina et al. (14). The so-called continuous NRI (or category-free NRI) does not depend on the existence of risk categories and allows NRI estimation in the context of survival data. The continuous NRI is the weighted sum of the observed event rate increase among the individuals for whom the predicted risk goes up and observed event rate decrease among those for whom the predicted risk goes down.

Results

Coronary CT angiography could not be performed in 16 (3%) patients because of renal failure (n = 2), known allergy to contrast material (n = 1), patient preferences (n = 8), Parkinson disease (n = 1), patient non-cooperation (n = 1), lack of venous access (n = 1), and severe obesity (n = 2). Two examinations (0.4%) failed because of patient movement or premature scanning initiation. Three patients (0.6%) had mild to moderate allergic reactions.

Exercise ECG could not be performed in 48 (10%) patients because of orthopedic restraints (n = 13), neurologic restraints (n = 8), severe obesity (n = 3), abnormalities at resting ECG (n = 8), pulmonary disease (n = 2), inability to bicycle (n = 3), and a combination of these factors or an unspecified factor (n = 11). Results were considered inconclusive in 140 (30%) patients, mostly because the target heart rate was not reached. These patients together (n = 188) were considered to have had non-diagnostic tests in the survival analysis (Figure 1).

Follow-up

Follow-up data were obtained for 424 (90%) patients. Forty-four adverse cardiac events occurred in 30 patients. These events consisted of four cardiac deaths, six acute myocardial infarctions, 23 coronary revascularizations, and 11 instances of unstable angina pectoris. The occurrence of mortality was obtained for all patients, and, according to the national death registry, another four patients died of a non-cardiac cause—that is, respiratory failure, stroke, or malignancy (n = 2). The median follow-up time was 2.6 years (IQR, 2.1–3.2 years). The overall rate of events was 3.2% per year.

Patients without follow-up were younger (P < 0.001) and more often smokers (P = 0.02). Lower coronary calcium scores and fewer abnormal coronary CT angiography and exercise ECG results were observed for patients without follow-up, although this did not reach statistically significance (Table E1).

There was no significant difference in Duke Treadmill Score between the patients with and those without follow- up (median score, 5 [IQR, 1–7] vs. 5 [IQR, 0.5–7]). Only six patients were classified as having high risk, and they all completed follow-up.

Table 1. Events Sorted according to Test

Test and Result	Total No. of Patients	No. of Patients with MACE	No. of Patients with CD	No. of Patients with AMI	No. of Patients with LR	No. of Patients with UAP	Annual event rate (%)
Coronary calcium score = 0	151	2 (1)	0 (0)	0 (0)	1 (0.7)	3 (2)	0.5
Coronary calcium score > 0	266	28 (11)	4 (2)	6 (2)	22 (8)	8 (3)	4.2
Coronary CT angiography <50%	277	8 (3)	1 (0.4)	2 (0.7)	3 (1)	3 (1)	1.2
Coronary CT angiography ≥50%	132	22 (17)	3 (2)	4 (3)	20 (15)	8 (6)	6.8
Normal exercise ECG result	172	7 (4)	0 (0)	1 (0.6)	6 (4)	3 (2)	1.6
Ischemic exercise ECG result	85	4 (5)	0 (0)	1 (1)	5 (6)	1 (1)	1.9
Non-diagnostic exercise ECG result	167	19 (11)	4 (2)	4 (2)	12 (7)	7 (4)	4.6

Numbers in parentheses are percentages. AMI = acute myocardial infarction, CD = cardiac death, LR = late revascularization, UAP = unstable angina pectoris requiring hospitalization.

Descriptive Analysis

Coronary calcium score

Patients with no coronary calcium had excellent prognoses. No cardiac deaths and no nonfatal myocardial infarctions were observed (Table 1). However, three hospitalizations for unstable angina occurred, and one late revascularization was performed.

Coronary CT angiography

In the group of patients with obstructive CAD, 22 (17%) events occurred, consisting of three cardiac deaths (2%) and four myocardial infarctions (3%). Patients without obstructive CAD experienced significantly fewer total events (eight patients [3%]), one cardiac death (0.4%), and two (0.7%) myocardial infarctions (Table 1). Patients who did not undergo coronary CT angiography or in whom the study was non-diagnostic did not experience a MACE.

Exercise ECG

Patients with a normal exercise ECG result experienced seven (4%) events during follow-up, of which one (0.6%) was a myocardial infarction. An abnormal exercise ECG result was associated with the occurrence of four (5%) events, of which one (1%) was a myocardial infarction. We observed 19 (11%) events in the group of patients who were unable to perform the test or who had an inconclusive test; of these events, four (2%) were cardiac deaths and four (2%) were myocardial infarctions (Table 1).

Univariable Analysis

The clinical risk estimators SCORE and Diamond and Forrester score were both significant predictors of MACEs (Table 2). Of the traditional risk factors, only sex was a statistically significant predictor. A typical angina presentation was associated with an HR of 3.86 (95% CI: 0.88-16.87) compared with non-anginal complaints, while atypical presentation was associated with an HR of 1.91 (95% CI: 0.43-8.47).

Detectable coronary calcium was a significant predictor of MACE (HR, 8.22 [95% CI: 1.96-34.51]; P = 0.004), along with non-diagnostic exercise ECG results (HR, 3.00 [95% CI: 1.26-7.14]; P = 0.01). Regarding coronary CT angiography results, in the first model, obstructive CAD was associated with a significantly higher hazard (HR, 6.22 [95% CI: 2.77-13.99]; P < 0.001). In the second model, the presence of non-obstructive plaque was also associated with a higher hazard (HR, 5.03 [95% CI: 0.62-40.85]), although this did not reach statistical

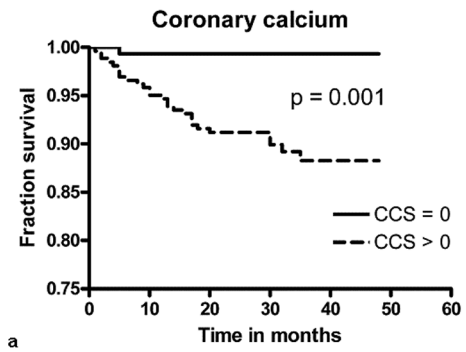
significance (P = 0.13), while obstructive CAD remained a significant predictor (HR, 20.80 [95% CI: 2.80-154.33]; P = 0.003).

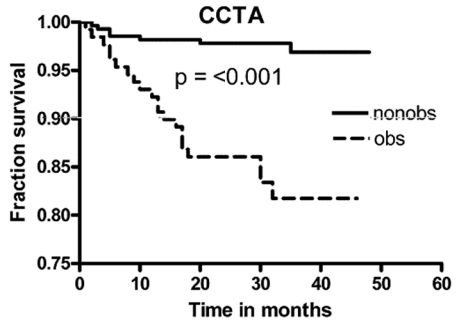
Survival Analysis

Unadjusted comparison between patients with and those without coronary calcium showed significantly higher event-free survival in the group with no calcium (Figure 2a) (P = 0.001). The annual event rate in patients without calcium was 0.5%. Absence of obstructive CAD at coronary CT angiography was associated with a significantly lower event rate than the presence of obstructive CAD at coronary CT angiography (Figure 2b) (P < 0.001). The annual event rates were 6.8% versus 1.2% for obstructive versus non-obstructive CAD (Table 1).

Non-diagnostic exercise ECG results were associated with significantly higher event rates than normal or ischemic exercise ECG results (Figure 2c) (P = 0.016). The observed annual event rate for patients with non-diagnostic results compared with that for patients with normal or ischemic results was substantially higher (4.6% vs. 1.6% and 1.9%, respectively)

Figure 2. Graphs show Kaplan-Meier estimates of survival as compared between (a) patients with and those without visible coronary calcium (CCS), (b) patients with and those without obstructive CAD at coronary CT angiography, and (c) patients with normal, those with ischemic, and those with non-diagnostic exercise ECG results.





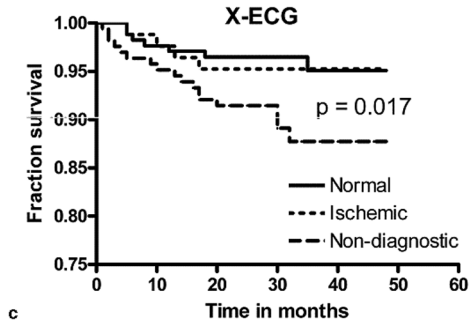


Table 2. Results of Cox Univariable Analysis

Variable	MACE (n=30)	No MACE (n=392)	HR *	*P*-value
Age (y) †	58 ± 8	56 ± 10	1.02 (0.99-1.06)	0.21
Male sex	22 (73)	195 (50)	2.68 (1.19-6.02)	0.02
Smoking	9 (30)	107 (27)	1.15 (0.53-2.52)	0.72
Hypertension	20 (67)	193 (49)	1.99 (0.93-4.25)	0.08
Diabetes	7 (23)	51 (13)	1.92 (0.82-4.47)	0.13
Dyslipidaemia	20 (67)	229 (58)	1.29 (0.60-2.76)	0.52
Family history of CAD	16 (53)	173 (44)	1.39 (0.68-2.85)	0.37
History of vascular disease	2 (7)	28 (7)	0.94 (0.22-3.94)	0.93
Body mass index	26.9	26.8	1.01 (0.94-1.08)	0.86
Chest pain				
Non-anginal	2 (7)	64 (16)	1	
Atypical	13 (43)	207 (53)	1.91 (0.43-8.47)	0.39
Typical	15 (50)	121 (31)	3.86 (0.88-16.87)	0.07
SCORE ‡	6 (2.75-8.25)	3 (1-6)	1.06 (1.01-1.12)	0.03
Diamond & Forrester score ‡	67 (53-92)	54 (28-79)	1.02 (1.01-1.04)	0.003
Low (< 22)			1	
Intermediate (22-78)			2.83 (0.64-12.47)	0.17
High (>78)			5.16 (1.17-22.70)	0.03
Calcium score				
0	2 (7)	149 (39)	1	
0-10	2 (7)	43 (11)	3.19 (0.45-22.65)	0.25
10-100	10 (33)	82 (21)	8.45 (1.85-38.58)	0.006
100-400	7 (23)	59 (15)	8.50 (1.77-40.793)	0.008
>400	9 (30)	54 (14)	11.66 (2.52-54.01)	0.002
>0	28 (93)	238 (62)	8.22 (1.96-34.51)	0.004
CCTA model 1			6.22 (2.77-13.99)	< 0.001
Non-obstructive CAD	8 (27)	269 (71)		
Obstructive CAD	22 (73)	108 (29)		
CCTA model 2				
No plaque	1 (3)	114 (30)	1	
Non-obstructive plaque	7 (23)	155 (41)	5.03 (0.62 – 40.85)	0.13
Obstructive plaque	22 (73)	110 (29)	20.80 (2.80-154.33)	0.003
Exercise ECG				
Normal	7 (23)	164 (42)	1	
Ischemic	4 (13)	80 (20)	1.19 (0.35-4.08)	0.78
Non-diagnostic	19 (63)	148 (38)	3.00 (1.26-7.14)	0.01

Unless otherwise specified, data are numbers of patients, with percentages in parentheses. CCTA = coronary CT angiography
** HR = hazard ratio. Numbers in parentheses are 95% confidence intervals (CIs).*
† Data are means ± standard deviations.
‡ Data are medians, with IQRs in parentheses.

Multivariable Risk-adjusted Analysis

In the multivariable Cox regression analysis, after adjusting for SCORE, the presence of obstructive CAD (HR, 6.61 [95% CI: 2.83-15.43]) and non-diagnostic exercise ECG result (HR, 2.93 [95% CI: 1.23-6.99]) remained independent predictors of MACE (Table 3). The addition of CCS led to a slight decrease in the HRs of coronary CT angiography (HR, 5.00 [95% CI: 1.72-14.52]) and exercise ECG (HR, 2.80 [95% CI: 1.17-6.73]), whereas CCS itself was not an independent predictor (HR, 1.09 [95% CI; 0.89-1.33]).

A statistically significant increase in the global χ^2 value was seen after adding coronary CT angiography results to a risk-adjusted model with exercise ECG (global χ^2 value, 37.7 vs. 13.7; $P < 0.001$) but not with the subsequent addition of CCS (global χ^2 value, 38.2 vs. 37.7; $P = 0.40$).

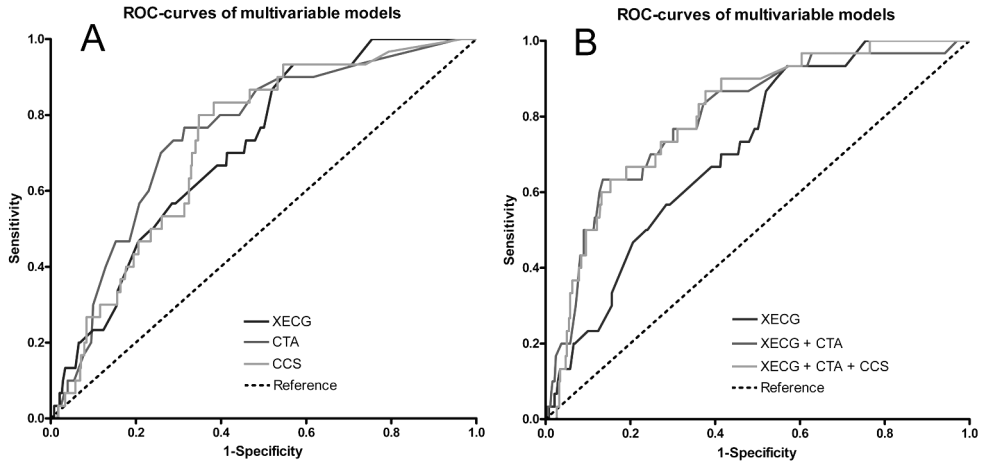
Detectable calcium was an independent predictor in a risk-adjusted model with exercise ECG alone (HR, 1.33 [95% CI: 1.13-1.56]). Subsequent addition of coronary CT angiography improved model performance significantly (global χ^2 value, 38.2 vs. 26.3; $P = 0.002$) (Table E2).

Receiver Operating Characteristics

The obtained curves and their AUC estimates demonstrated that the combination of clinical characteristics and exercise ECG findings resulted in an AUC of 0.71 (95% CI: 0.63-0.79) (Figure 3). A risk-adjusted model with coronary CT angiography resulted in an AUC of 0.75 (95% CI: 0.67-0.83), while clinical characteristics and CCS produced an AUC of 0.73 (95% CI: 0.65-0.81).

When clinical characteristics, exercise ECG findings, and coronary CT angiography findings were combined in a model, we found an AUC of 0.80 (95% CI: 0.72-0.88). Insertion of CCS into this model did not improve its predictive value considerably (AUC, 0.81 [95% CI: 0.73-0.88]).

Figure 3. Graphs show receiver operating characteristics curves of multivariable models of (a) exercise ECG (XECG), CCS, and coronary CT angiography (CCTA) and (b) incremental value of coronary CT angiography and CCS beyond exercise ECG. The models are risk adjusted with the SCORE 10-year cardiovascular disease mortality risk metric (9).



NRI Results

The NRI was 54% (number of patients classified upward: 160; downward: 249) when exercise ECG was added to the clinical Cox proportional hazards model using only the SCORE. The addition of coronary CT angiography to the model with both SCORE and exercise ECG resulted in a NRI of 80% (number of patients classified upward: 132; downward: 277). Finally, the full model that included CCS resulted in an NRI of 47% (number of patients classified upward: 182; downward: 227) (Table 3).

Discussion

Our results show that the presence of obstructive CAD at coronary CT angiography, the degree of exercise tolerance, and the extent of coronary calcification predict future adverse events in patients with stable symptoms of chest pain. After adjustment for clinical characteristics, coronary CT angiography showed incremental value beyond exercise testing, whereas the additive prognostic value of CCS was limited compared with coronary CT angiography.

For the past decades, exercise testing has been the diagnostic cornerstone for the evaluation of ischemic heart disease and can help identify patients at increased risk for adverse events (3,4). Exercise tolerance is a powerful predictor of prognosis (15,16), which is confirmed in our study by the fact that inability to perform and complete exercise ECG predicted unfavorable outcome. Exercise capacity is an important predictor of adverse outcome, representing contractile left ventricular function as well as overall physical health. Most nondiagnostic exercise tests were the result of a low exercise capacity and consequent inability to reach the target heart rate.

Remarkably, ischemic ECG changes were not associated with increased event rates. Presumably, patients with these changes were treated more aggressively, explaining their better event-free survival. Second, patients developing ischemic ECG changes during exercise, but straining themselves outstandingly, apparently tolerate ischemia well and may be in good general shape. The fact that only a minority of the study population was categorized as having high-risk disease according to the Duke Treadmill Score supports this hypothesis.

Non-enhanced coronary CT imaging can be used to help detect and quantify calcified CAD, and its findings have independent prognostic value in both symptomatic and asymptomatic individuals (17–21). Also in our study, the presence of coronary calcium was associated with an adverse outcome. In particular, our results confirm the idea that patients with no detectable calcium have an excellent outcome (annual event rate, 0.5%).

Table 3. Results of Cox Multivariable Analysis

Variable	HR	*P*-value	Global χ^2	Model Comparison P-value	AUC	Continuous NRI (%)
SCORE	1.06 (1.01-1.12)	0.03	5.1		0.66 (0.57-0.75)	reference
SCORE	1.07 (1.01-1.12)	0.02	13.7	0.01	0.71 (0.63-0.79)	54
Normal ECG result	1					
Ischemic ECG result	1.02 (0.29-3.52)	0.98				
Non-diagnostic ECG result	2.96 (1.24-7.04)	0.01				
SCORE	1.02 (0.96-1.09)	0.54	37.7	< 0.001	0.80 (0.72-0.88)	80
Normal ECG result	1					
Ischemic ECG result	0.69 (0.20-2.41)	0.57				
Non-diagnostic ECG result	2.93 (1.23-6.99)	0.02				
CCTA	6.61 (2.83-15.43)	< 0.001				
SCORE	1.02 (0.95-1.08)	0.67	38.2	0.40	0.81 (0.73-0.88)	47
Normal ECG result	1					
Ischemic ECG result	0.69 (0.20-2.39)	0.56				
Non-diagnostic ECG result	2.80 (1.17-6.73)	0.02				
CCTA	5.00 (1.72-14.52)	0.003				
CCS	1.09 (0.89-1.33)	0.41				

Numbers in parentheses are 95% confidence intervals. Here, CCS = natural logarithm of the calcium score +1, ECG = exercise ECG, CCTA = coronary CT angiography.

More recently, coronary CT angiography has emerged as a non-invasive alternative for direct assessment of CAD. Evidence of its prognostic value is emerging (5–7), although results from some studies were affected by work-up bias, selected populations, and short follow-up periods (22–24). In our consecutive population we can confirm the good predictive value of coronary CT angiography for future adverse events.

Because of the nature of our diagnostic work-up, we could directly compare the prognostic value of the different diagnostic modalities in patients with stable chest pain and a low-to-intermediate probability of CAD. To the best of our knowledge, ours is the first study to evaluate the incremental predictive value of coronary CT angiography over exercise ECG in a consecutive patient population.

In the risk-adjusted multivariable analysis, obstructive atherosclerosis at coronary CT angiography and non-diagnostic exercise ECG results remained independent predictors of late cardiac adverse events, whereas coronary calcium did not remain a significant predictor. This finding is confirmed by a recent study (25) that reported no additive prognostic value of CCS next to coronary CT angiography. It appears that the information obtained with calcium scanning largely overlaps with the information obtained with coronary CT angiography, while the latter also provides such additional characteristics as total plaque burden and luminal obstruction.

The performance of the final model with clinical predictors, exercise ECG, and CCS improved significantly after the addition of coronary CT angiography results.

In addition, we assessed the potential clinical value of the considered predictors by calculating the continuous NRI, an extension of the traditional NRI that is independent of risk categories. Our results suggest that coronary CT angiography is most effective in improving risk prediction, since the NRI was substantial. For CCS, the NRI was 47%, although the addition of CCS did not improve model performance in terms of the χ^2 value or AUC. This finding is explained by the fact that the continuous NRI does not take into account the magnitude of the increase (or decrease) in predicted risk.

In other words, the predictions were not substantially influenced when CCS was added, but among patients with a higher predicted risk, the observed event rate increased compared with the overall mean. Similarly, for patients with a lower predicted risk, the observed event rate decreased compared with the overall mean. Finally, it should be noted that the continuous NRIs reported in the current study cannot be compared directly with the traditional NRIs that have been reported in other studies because of the differences in definition and calculation described earlier.

Although our study population consisted of prospectively enrolled "allcomers" with stable angina complaints, the limitations associated with an observational single-center study were still present. Therefore our results may not necessarily reflect populations or practices elsewhere. Because of the limited number of hard events, we used a composite end point of cardiac death, non-fatal myocardial infarction, unstable angina requiring hospitalization, and coronary revascularization. The use of coronary revascularization as an end point could lead to overestimation of the prognostic value as a result of a potential work-up bias. Revascularizations performed within 6 months from the initial work-up were excluded to

minimize this effect. Even though an experienced cardiologist, blinded to the initial test reports, evaluated all events, unstable angina requiring hospitalization can be a subjective end point.

Given the limited number of events, our analysis could be subject to overfitting. Current guidelines stating a minimum of 10 outcome events per predictor variable have been questioned (26). In this view, our findings regarding the additive value of coronary calcium should be considered explanatory, and conclusions should be made with caution. To limit the number of variables in the multivariable analysis, we dichotomized coronary CT angiography results.

Incomplete follow-up may result in underreporting of adverse events. However, the follow-up rate was substantial (90%), and no deaths occurred in the group without follow-up, as confirmed by the national death registry. Patients lost to follow-up appear to have been at lower risk, with fewer abnormal test results (Table E1). Larger multicenter studies with longer follow-up, or meta-analyses of existing studies, are needed to fully comprehend the prognostic value of these modalities. In conclusion, both functional and anatomic assessment of CAD has prognostic value. Coronary CT angiography findings are strong predictors of future adverse events, with incremental value over clinical predictors, stress testing, and coronary calcification.

Disclosures of Potential Conflicts of Interest: A.D. No potential conflicts of interest to disclose. T.S.S.G. No potential conflicts of interest to disclose. B.S.F. No potential conflicts of interest to disclose. T.W.G. No potential conflicts of interest to disclose. N.R.A.M. No potential conflicts of interest to disclose. A.M. No potential conflicts of interest to disclose. M.G.M.H. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: receives royalties from Cambridge University Press; received travel, accommodations, and/or meeting expenses from the organizers of the Cardiac CT/ MRI course in Cannes, France, in 2010 and 2011. Other relationships: none to disclose. P.J.d.F. No potential conflicts of interest to disclose. K.N. No potential conflicts of interest to disclose.

References

1. Meijboom WB, van Mieghem CA, Mollet NR, et al. 64-slice computed tomography coronary angiography in patients with high, intermediate, or low pretest probability of signifi cant coronary artery disease. J Am Coll Cardiol 2007; 50 (15): 1469 – 1475.
2. Weustink AC, Mollet NR, Neefjes LA, et al. Diagnostic accuracy and clinical utility of noninvasive testing for coronary artery disease. Ann Intern Med 2010; 152 (10): 630 – 639.
3. Mark DB, Shaw L, Harrell FE Jr, et al. Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease. N Engl J Med 1991; 325 (12): 849 – 853.
4. Roger VL, Jacobsen SJ, Pellikka PA, Miller TD, Bailey KR, Gersh BJ. Prognostic value of treadmill exercise testing: a populationbased study in Olmsted County, Minnesota. Circulation 1998; 98 (25): 2836 – 2841.
5. Chow BJ, Wells GA, Chen L, 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 – 1028.
6. Hadamitzky M, Distler R, Meyer T, et al. Prognostic value of coronary computed tomographic angiography in comparison with calcium scoring and clinical risk scores. Circ Cardiovasc Imaging 2011; 4 (1): 16 – 23.
7. Hulten EA, Carbonaro S, Petrillo SP, Mitchell JD, Villines TC. Prognostic value of cardiac computed tomography angiography: a systematic review and meta-analysis. J Am Coll Cardiol 2011; 57 (10): 1237 – 1247.
8. Nieman K, Galema T, Weustink A, et al. Computed tomography versus exercise electrocardiography in patients with stable chest complaints: real-world experiences from a fast-track chest pain clinic. Heart 2009; 95 (20): 1669 – 1675.
9. Conroy RM, Pyörälä K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003; 24 (11): 987 – 1003.
10. Diamond GA. A clinically relevant classifi cation of chest discomfort. J Am Coll Cardiol 1983; 1 (2 Pt 1): 574 – 575.
11. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M Jr, Detrano R. Quantifi cation of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol 1990; 15 (4): 827 – 832.
12. Gibbons RJ, Balady GJ, Beasley JW, et al. ACC/AHA guidelines for exercise testing: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing). Circulation 1997; 96 (1): 345 – 354.
13. Gibler WB, Cannon CP, Blomkalns AL, et al. Practical implementation of the guidelines for unstable angina/non-ST-segment elevation myocardial infarction in the emergency department: a scientifi c statement from the American Heart Association Council on Clinical Cardiology (Subcommittee on Acute Cardiac Care), Council on Cardiovascular Nursing, and Quality of Care and Outcomes Research Interdisciplinary Working Group, in Collaboration With the Society of Chest Pain Centers. Circulation 2005; 111 (20): 2699 – 2710.
14. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassifi cation improvement calculations to measure usefulness of new biomarkers. Stat Med 2011; 30 (1): 11 – 21.
15. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. N Engl J Med 2002; 346 (11): 793 – 801.
16. Prakash M, Myers J, Froelicher VF, et al. Clinical and exercise test predictors of all-cause mortality: results from. 6,000 consecutive referred male patients. Chest 2001; 120 (3): 1003 – 1013.
17. Erbel R, Möhlenkamp S, Moebus S, et al. Coronary risk stratifi cation, discrimination, and reclassifi cation improvement based on quantifi cation of subclinical coronary atherosclerosis: the Heinz Nixdorf Recall study. J Am Coll Cardiol 2010; 56 (17): 1397 – 1406.
18. Vliegenthart R, Oudkerk M, Song B, van der Kuip DA, Hofman A, Witteman JC. Coronary calcifi cation detected by electronbeam computed tomography and myocardial infarction. The Rotterdam Coronary Calcifi cation Study. Eur Heart J 2002; 23 (20): 1596 – 1603.
19. Elias-Smale SE, Proença RV, Koller MT, et al. Coronary calcium score improves classifi cation of coronary heart disease risk in the elderly: the Rotterdam study. J Am Coll Cardiol 2010; 56 (17): 1407 – 1414.
20. Detrano R, Guerci AD, Carr JJ, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. N Engl J Med 2008; 358 (13): 1336 – 1345.
21. Shaw LJ, Raggi P, Schisterman E, Berman DS, Callister TQ. Prognostic value of cardiac risk factors and coronary artery calcium screening for all-cause mortality. Radiology 2003; 228 (3): 826 – 833.
22. Gilard M, Le Gal G, Cornily JC, et al. Midterm prognosis of patients with suspected coronary artery disease and normal multislice computed tomographic fi ndings: a prospective management outcome study. Arch Intern Med 2007; 167 (15): 1686 – 1689.
23. Pundziute G, Schuijf JD, Jukema JW, et al. Prognostic value of multislice computed tomography coronary angiography in patients with known or suspected coronary artery disease. J Am Coll Cardiol 2007; 49 (1): 62 – 70.
24. Gaemperli O, Valenta I, Schepis T, et al. Coronary 64-slice CT angiography predicts outcome in patients with known or suspected coronary artery disease. Eur Radiol 2008; 18 (6): 1162 – 1173.
25. Kwon SW, Kim YJ, Shim J, et al. Coronary artery calcium scoring does not add prognostic value to standard 64-section CT angiography protocol in low-risk patients suspected of having coronary artery disease. Radiology 2011; 259 (1): 92 – 99.
26. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007; 165 (6): 710 – 718.

Appendix

Table E1. Baseline Characteristics

Variable	Follow-up (n=424)	No follow-up (n=47)	*P*-value
Age (y) *	56 ± 10	50 ± 11	< 0.001
Male sex	219 (52)	25(53)	0.88
Smoking	116 (27)	22(47)	0.01
Hypertension	213 (50)	21(45)	0.54
Diabetes	58 (14)	10(21)	0.19
Dyslipidaemia	251 (59)	29(62)	0.88
Family history of CAD	191 (45)	23(49)	0.65
History of vascular disease	30 (7)	1 (2)	0.35
Body mass index †	26.8 (24.2-30.0)	26.9 (24.5-30.5)	0.58
Chest pain			0.31
Non-anginal chest pain	66 (16)	8 (17)	
Atypical chest pain	222 (52)	29 (62)	
Typical chest pain	136 (32)	10 (21)	
SCORE †	3 (1-7)	1 (1-5)	0.02
Diamond & Forrester †	54 (32-79)	46 (13-59)	0.02
Calcium score †	18 (0-158)	0 (0-93)	0.06
0	151 (34)	24 (52)	0.04
0-10	45 (11)	3 (7)	
10-100	92 (22)	9 (20)	
100-400	66 (16)	8 (17)	
>400	63 (15)	2 (4)	
Coronary CT Angiography	453 (96)		0.12
Non-obstructive CAD	277 (68)	35 (80)	
Obstructive CAD	132 (32)	9 (20)	
Exercise ECG	471 (100)		0.76
Normal	172 (41)	18 (38)	
Ischemic	85 (20)	8 (17)	
Non-diagnostic	167 (39)	21(45)	
Duke treadmill score †	5 (1-7)	5 (0.5-7)	0.95
Low risk	190 (51)	24 (55)	
Intermediate risk	181 (48)	20 (45)	
High risk	6 (2)	0	

Unless otherwise specified, data are numbers of patients, with percentages in parentheses.
** Data are means ± standard deviations.*
† Data are medians, with IQRs in parentheses

Table E2. Cox Multivariable Analysis

Variable	HR	*P*-value	Global χ^2	Model comparison *P*-value	AUC (95% CI)	Continuous NRI (%)
SCORE	1.06 (1.01-1.12)	0.03	5.1		0.66 (0.57-0.75)	reference
SCORE	1.07 (1.01-1.12)	0.02	13.7	0.01	0.71 (0.63-0.79)	54
Normal ECG result	1					
Ischemic ECG result	1.02 (0.29-3.52)	0.98				
Non-diagnostic ECG result	2.96 (1.24-7.04)	0.01				
SCORE	1.02 (0.96-1.09)	0.57	26.3	< 0.001	0.77 (0.70-0.85)	72
Normal ECG result	1					
Ischemic ECG result	0.79 (0.23-2.75)	0.71				
Non-diagnostic ECG result	2.40 (1.00-5.79)	0.05				
CCS	1.33 (1.13-1.56)	0.001				
SCORE	1.02 (0.95-1.08)	0.67	38.2	0.002	0.81 (0.73-0.88)	15
Normal ECG result	1					
Ischemic ECG result	0.69 (0.20-2.39)	0.56				
Non-diagnostic ECG result	2.80 (1.17-6.73)	0.02				
CCS	1.09 (0.89-1.33)	0.41				
CCTA	5.00 (1.72-14.5)	0.003				

Numbers in parentheses are 95% confidence intervals. CCS = natural logarithm of the calcium score +1, ECG = exercise ECG, CCTA = coronary CT angiography

Chapter 12.2

Prognostic Value of Cardiac Computed Tomography Angiography

J Am Coll Cardiol. 2011 Jun;57(25):2543-2544.

Tessa S.S. Genders
Admir Dedic
Koen Nieman
M.G. Myriam Hunink

Letter to the Editor

We read with interest the systematic review and meta-analysis reporting on the prognostic value of cardiac computed tomography angiography (CCTA) by Hulten et al. (1). The authors performed an extensive literature search and pooled 18 studies that reported the incidence of cardiovascular events after a CCTA examination. Although we agree with the authors' conclusion that cardiovascular events among patients with normal findings on CCTA are rare, some issues raised our concern.

First, the primary outcome measure—major adverse cardiovascular events (MACE)—was defined as death, myocardial infarction, or revascularization. We noted that 4 large studies (2–5) did not report event rates for revascularization (see Table 2 in [1]), nor did they include revascularization in their definition of MACE. These reports together contribute 67% (6,434 of 9,592) of all patients included in this meta-analysis. Unfortunately, this issue is not mentioned by the authors, and it is unclear to what extent this was accounted for in the analysis. As acknowledged by the authors, the observed increased rate of MACE among patients with coronary artery disease is largely driven by the increased rate of revascularization. Had all studies included revascularization in their definition of MACE, the true absolute rates of MACE are likely to be substantially higher. The effect of this underestimation on sensitivity, specificity, and the likelihood ratio should be explored. Even if MACE was defined consistently across all included studies, the question remains whether MACE including revascularizations is appropriate as the primary endpoint for the evaluation of CCTA. As pointed out by the authors, most revascularization procedures will be initiated based on CCTA findings, resulting in an overestimation of the predictive value of CCTA.

Second, Figure 6 represents a Fagan's nomogram for the prediction of future MACE by CCTA. The figure legend indicates that the left axis represents the pre-test probability of disease, whereas the Results section indicates that it represents the pre-test probability of having future MACE. If using the pre-test probability of disease, it seems inappropriate to perform Bayesian revision combining a pre-test probability of disease with a likelihood ratio that applies to the annual rate of MACE, when calculating the post-test probability of future MACE. If the pre-test probability estimate in Figure 6 refers to the probability of future MACE, it remains unclear how this estimate can be calculated and what time frame is being considered. Assuming a time-frame of 1 year and given the reported annual event rates of patients in this report, a hypothetical patient with a 20% pre-test probability of MACE seems like an unrealistic example.

References

1. Hulten EA, Carbonaro S, Petrillo SP, Mitchell JD, Villines TC. Prognostic value of cardiac computed tomography angiography: a systematic review and meta-analysis. J Am Coll Cardiol 2011;57: 1237– 47.
2. Ostrom MP, Gopal A, Ahmadi N, et al. Mortality incidence and the severity of coronary atherosclerosis assessed by computed tomography angiography. J Am Coll Cardiol 2008;52:1335– 43.
3. Min JK, Shaw LJ, Devereux RB, et al. Prognostic value of multidetector coronary computed tomographic angiography for prediction of all-cause mortality. J Am Coll Cardiol 2007;50:1161–70.
4. Shaw LJ, Berman DS, Hendel RC, Borges Neto S, Min JK, Callister TQ. Prognosis by coronary computed tomographic angiography: matched comparison with myocardial perfusion single-photon emission computed tomography. J Cardiovasc Comput Tomogr 2008;2:93–101.
5. Chow BJ, Wells GA, Chen L, 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;55:1017–28.

Chapter 13

Methods for Calculating Sensitivity and Specificity of Clustered Data: A Tutorial

Accepted for publication in *Radiology*

Tessa S.S. Genders
Sandra Spronk
Theo Stijnen
Ewout W. Steyerberg
Emmanuel M.E.H. Lesaffre
M.G. Myriam Hunink

Abstract

The performance of a diagnostic test is often expressed in terms of sensitivity and specificity compared with the reference standard. Calculations of sensitivity and specificity commonly involve multiple observations per patient, which implies that the data are clustered. Whether analysis of sensitivity and specificity per patient or using multiple observations per patient is preferable, depends on the clinical context and consequences. The purpose of this paper is to discuss and illustrate the most common statistical methods that calculate sensitivity and specificity of clustered data, adjusting for the possible correlation between observations within each patient. This tutorial presents and illustrates the following methods: (1) analysis at different levels ignoring correlation; (2) variance adjustment; (3) logistic random-effects models; and (4) generalized estimating equations. Furthermore, we discuss issues related to missing data, and how to analyze the impact of covariables on the estimate of sensitivity and specificity. We find that results may vary depending on the method used and the strength of the correlation between observations. We conclude that the choice of method and the level of reporting should correspond with the clinical decision problem. If multiple observations per patient are relevant to the clinical decision problem, the potential correlation between observations should be explored and taken into account in the statistical analysis.

Introduction

Researchers commonly encounter clustered data. For example, a study may be carried out in multiple hospitals. Patients from the same hospital tend to be more similar (and thus clustered), because they share the specific characteristics of that hospital (a certain patient selection, physician experience, local protocols, etc). However, apart from clustered individuals, a study may also consider multiple observations per individual, which implies that the observations within a patient are clustered.

Datasets from studies evaluating the diagnostic performance of imaging tests commonly contain multiple observations per patient. For example, we typically have data on 17 segments in 4 vessels per patient in evaluating the coronary arteries, 33 segments in 21 vessels per patient in evaluating the peripheral arteries, 8 segments in evaluating the liver, 5 segments in evaluating the breast, and a large number of possible organs and diagnoses in whole-body imaging. In situations like this, the observations within a patient are clustered.

The choice of level of analysis should be dictated by the clinical decision that follows. For example, if the decision is medical therapy versus no medication, the level of analysis should be the patient. If the next management decision is perform catheter angiography or not, the logical level of analysis is also the patient. However, if the decision is the choice of vessels that may require intervention, then the level of analysis should be the vessel. Furthermore, if the decision is whether to perform balloon angioplasty and stent placement, the logical level of analysis would be the segment.

When the level of interest involves vessels or segments, the observations are clustered within the patient. Observations within patients (clusters) are often positively correlated which may lead to biased estimates of sensitivity (the probability of a positive test result given presence of disease) and specificity (the probability of a negative test result given absence of disease) when the correlated nature of the data within each cluster is ignored. More importantly, an analysis ignoring correlation often yields a misleadingly small estimated standard error (and 95% confidence interval (CI)) for the sensitivity and specificity since it erroneously counts all observations as independent observations (1). Thus, analyses of data that include multiple observations per patient require some form of adjustment to account for the possible correlation between observations.

Notice that the correlation here refers to the presence vs. absence of false-negative results (and false-positive results respectively). The effect of that correlation can only be measured by explicitly taking it into account in the analysis.

The purpose of this tutorial is to present and illustrate statistical methods that calculate sensitivity and specificity from clustered data, adjusting for the possible correlation between observations within each cluster. This tutorial is targeted towards statistical practitioners and therefore the focus is on practical guidance in the use of the methods and the statistical software. We minimize mathematical notation and do not address the theoretical justification for the methods since this is provided by statistical textbooks (1-3) and published papers on this topic (4).

Methods

Example in coronary artery disease

The methods of analysis will be illustrated both theoretically and practically using a dataset consisting of 50 patients with suspected coronary artery disease (CAD) each with 17 coronary segments that have been imaged by both multidetector CT angiography (CT, the index test) and catheter-based angiography (CA, the reference test). In CAD, segments are the level of interest when percutaneous intervention is considered for each segment. Furthermore, most investigators feel that location of disease should have an impact on how an observation is counted. For example, a patient with a left main coronary artery stenosis that is missed by CT and an incorrectly identified distal right coronary artery stenosis would be counted as a true positive observation in a patient-level analysis, whereas in a segment-level analysis it would be counted as one false-negative observation and one false-positive observation. In addition, using the segment as level of analysis has the opportune side-effect of increasing the number of observations (i.e. the sample size). This effect may appear to be beneficial, although the additional observations obtained are not independent observations, and the "effective sample size" is usually smaller than the total number of observations. Therefore, the increase in sample size (more precision, smaller 95% CIs) is counterbalanced by adjusting for the correlated nature of the data (less precision, wider 95% CIs). If there is no correlation, the effective sample size is equal to the number of observations, whereas if the correlation is perfect (no variation in observations within clusters), the effective sample size equals the number of clusters. Sample size calculations for studies with a clustered design should focus on the effective sample size by taking into account the correlation and the cluster size (3).

The vessel-level analysis will be omitted since the analyses are similar. Furthermore, the details of the data collection and image interpretation (5) are irrelevant to the purpose of this paper and we will therefore not elaborate on these. Examples of software code in STATA are provided in the Appendix 1. Background information and mathematical notations are provided in Appendix 2.

Method 1) Analysis at different levels ignoring correlation

The simplest and most commonly used method in the analysis of diagnostic test data is to analyze the data at different levels, i.e. the level of analysis is patient and segment respectively. The difference in results gives some insight into how the multiple observations per patient affect the estimates of sensitivity and specificity and their corresponding CIs.

Several approaches may be taken in the patient-level analysis, where test positivity (*Test+*) and test negativity (*Test-*) are defined at the patient level. A positive/negative test might be based on one observation randomly picked from each patient or, alternatively, may be based on a summary measure of all the observations. In CAD, the test result from a patient is typically counted as true (resp. false) positive if the reference standard CA demonstrated significant (resp. no significant) CAD anywhere in the coronary tree and anywhere in the coronary artery tree a significant stenosis was diagnosed on CT. If the number of patients with a true positive, false-negative, false-positive, and true negative test result are respectively *TP*, *FN*, *FP, and TN,* then we calculate sensitivity and specificity as follows:

$$sens = P(Test+ \mid Disease+) = \frac{TP}{TP+FN} \qquad spec = P(Test- \mid Disease-) = \frac{TN}{FP+TN}$$

The variance and standard error of these estimates can be calculated based on the binomial distribution, as follows:

$$Var(sens) = \frac{sens(1-sens)}{(TP+FN)} \quad \text{and} \quad Var(spec) = \frac{spec(1-spec)}{(FP+TN)}$$

$$SE(sens) = \sqrt{Var(sens)} = \sqrt{\frac{sens(1-sens)}{(TP+FN)}} \quad \text{and} \quad SE(spec) = \sqrt{Var(spec)} = \sqrt{\frac{spec(1-spec)}{(FP+TN)}}$$

The corresponding 95% CIs (for large samples) are:

$$sens \pm 1.96 \times SE(sens) \quad \text{and} \quad spec \pm 1.96 \times SE(spec).$$

The calculation of the 95% CI above uses the approximation to the normal distribution for a binomial proportion which is valid if the dataset is fairly large (≥10 TP, ≥10 FN, ≥10 FP and ≥10 TN observations) (3). Alternatively, one can calculate the exact binomial Cl with statistical software packages (Appendix 1), which will provide a one-sided (97.5%) CI if sensitivity (or specificity) is equal to one.

The remaining part of this tutorial focuses on the segment-level analysis. In the segment-level analysis a segment is counted as a true (resp. false) positive observation if a significant stenosis was diagnosed on CT in that segment when the reference standard CA demonstrated significant (resp. no significant) CAD in the same segment. Although it is recognized that the sensitivity (and specificity) of CT may be different across segments (due to differences in size and location of vessels), this tutorial assumes that the parameter of interest includes the mean estimate of sensitivity (and specificity) across segments. This implies that the sensitivity (and specificity) is assumed to be similar across segments. This assumption is reasonable in most situations, and often used for simplicity. If there is strong evidence that differences in sensitivity (and specificity) across segments exist, one can stratify or use a multivariate approach (not covered by this tutorial).

In the segment-level analysis ignoring correlation, the data from different segments are pooled ignoring the possible dependency and correlation between units within each patient (the cluster). Calculations are as above (Appendix 2). One can examine the correlation between the observations within the clusters to estimate to what extent this may affect the 95% CIs. A low intraclass correlation coefficient (see Method 2b) implies that the analysis with and without correction will yield similar results.

Method 2) Adjusting the variance

If dependency between multiple observations from the same patient (cluster) exists, then we would expect there to be a positive correlation between observations and a corresponding increase in the variance of the parameter estimates. The mean estimate of sensitivity and specificity at the segment-level may be unbiased when the correlation is ignored, but the variance is always underestimated.

The following three approaches assume that the mean estimates of the segment-level sensitivity and specificity obtained using Method 1 are correct and only provide an estimate of the adjusted variance. The adjusted CI is calculated using the adjusted standard error (see above), which is equal to the square root of the variance.

a) Adjusting the variance of sensitivity and specificity with the ratio estimator

An adjusted estimate of the variance can be calculated using the patient- (cluster-) specific estimate of sensitivity and specificity, weights proportional to the ratio of diseased (and non-diseased) observations in each patient and the mean cluster size for the diseased (and non-diseased) observations for sensitivity and specificity, respectively (1, 6). This approach can be applied using a spreadsheet. See Appendix 2 for more details.

b) Adjusting the variance of sensitivity and specificity with the variance inflation factor

The variance of sensitivity and specificity can also be adjusted with the variance inflation factor (VIF) (2). The VIF takes into account the cluster-size-weighted average cluster size (cluster size = number of observations per patient) and the intraclass correlation coefficient, which represents the resemblance between any two observations within a cluster. For every additional observation per patient the variance is inflated by the intraclass correlation. From the formula of the VIF, it can be seen that if the number of observations per patient is low and the intraclass correlation is low, the clustered nature of the data may have a negligible effect on the results. This method can be applied using a spreadsheet if the intraclass correlation coefficient is known. See Appendix 2 for more details.

The intraclass correlation coefficient (ICC) □ is a measure for the amount of clustering, and is defined as the between-cluster variance divided by the total variance (between-cluster + within-cluster variance) (3). If the observations within a cluster are highly correlated (i.e. if the within-cluster variance is small), the ICC will approach one. In the context of sensitivity and specificity (binary data), the ICC can be calculated using the natural estimator (2) for □ in a spreadsheet (see Appendix 2 and Table E2) or by fitting a logistic random-effects model (Method 3) with a random intercept only. The output provides the estimate of □, or provides the standard deviation of the random effect (□ $_u$) which can be used to calculate □ (Appendix 2).

c) Sandwich estimator of the variance

The third approach adjusts the variance of sensitivity and specificity as obtained by fitting a regular logistic regression model that ignores correlation. The formula (not shown) for the calculation of the adjusted variance resembles a sandwich, hence its name "sandwich estimator" (7). This method can be applied by fitting a logistic regression model and specifying the *clustered robust standard error* option (Appendix 1)

Method 3) Logistic random-effects models

Synonyms: Mixed models, hierarchical models, cross-sectional time series

The test characteristics sensitivity and specificity adjusted for correlation can be modeled using a logistic random-effects regression model (8-9). This method is comparable to performing a regular logistic regression, but now it is assumed that the sensitivity and specificity exhibit some variation around a common mean across the patients (clusters). The simplest approach is to consider only a random intercept, i.e. a random effect across clusters, which allows for clustering of observations within patients. The random intercept is

assumed to follow a normal distribution, which implies that the patients specific sensitivities or specificities are assumed to follow a certain distribution (in fact a logit normal distribution). This approach yields the median sensitivity and specificity (corresponding to random intercept = 0) across all patient-specific estimates of sensitivity (and specificity) assuming of course that the assumed model is correct. To obtain an estimate of the average sensitivity and specificity across the patient population (also called 'population-averaged') one needs to average the sensitivity and specificity over the (normal) distribution of the random intercept and this involves a numerical integration. How this needs to be done is, however beyond the scope of this tutorial.

The strengths of the random-effects approach include the efficiency for analyzing small datasets, and the opportunity to draw correct inferences for clustered data in both unpaired and paired study designs (10).

Method 4) Generalized estimating equations (GEE)

Synonyms: Population-averaged model, marginal model

Sensitivity and specificity estimates adjusted for correlation within patients can also be modeled using the generalized estimating equations (GEE) approach (2, 11-13). The GEE approach requires only a rough guess of the correlation structure among the clustered observations and is therefore called a 'working correlation matrix'. When this correlation matrix is specified as 'independent' (which corresponds to no correlation), the GEE estimates are identical to the estimates obtained by using logistic regression and a sandwich estimator for the variance (Method 2c). For a relatively large sample size (it is difficult to define "large"), the confidence intervals obtained with GEE are valid, even when the correlation matrix is not correct. However, the better the correlation matrix is specified, the greater the precision of the results, i.e. smaller CIs. Thus, misspecification will lead to somewhat conservative estimates of the 95% CIs.

Using the robust sandwich estimator of the variance allows us to draw correct inferences even when the correlation structure is misspecified (2). However, as seen above it is wise to make a good guess of the correlation structure (14). For instance, one could choose for an exchangeable correlation structure when it is reasonable to assume that the correlation among observations in the same cluster are equal. When there is a measure of distance between observations (in time or in space) one may let the correlation decrease with an increasing distance. An example of a correlation structure of that type is the autoregressive or m-dependent correlation matrix.

In contrast to random-effects models, the correlation itself is not of primary interest in the GEE approach. Furthermore, in the GEE approach the estimated sensitivity (and specificity) is interpreted as the weighted average across the study population ("population-averaged"), in other words, the GEE approach delivers the mean sensitivity and specificity across all patient-specific estimates of sensitivity (and specificity). This is in contrast to the sensitivity (and specificity) in a random-effects model which has a conditional (on the random effects) interpretation. Therefore, the point estimates obtained with a GEE are in general different from the point estimates obtained with a logistic random-effects. Finally, the GEE approach is less efficient compared to the random-effects approach because it is based on less assumptions (is not based on a fully specified probability model as the random effects approach (3)). On the positive side we can say that, for the same reason, the GEE approach is more robust against model mis-specification.

The impact of covariables on sensitivity and specificity

Although the overall estimate of sensitivity and specificity (adjusted for clustering within patients) is informative and a logical first step in the analysis of diagnostic data, certain factors (i.e. covariables) that may influence sensitivity or specificity are often of interest. For example, patient characteristics may impact the diagnostic performance of a test. For example, the sensitivity of coronary CT angiography can be different in male vs. female patients or in patients with vs. without a history of CAD. In addition, a newer scan protocol, a more experienced reader, or a more advanced imaging technique may result in better diagnostic performance.

The simplest approach to analyzing the impact of covariables is to stratify and repeat the analysis across strata of the covariable, although this may not be possible when considering continuous variables or small datasets. Regression models (Method 2c, 3, and 4) are easily extended with covariables in order to see which factors determine sensitivity and specificity. The next step is then to compare and test the difference between sensitivity and specificity for different levels/values of the covariable.

Since diagnostic studies often aim to compare the diagnostic performance of two different tests (e.g. test A vs. test B), the methods for analyzing the impact of a covariable will be illustrated for this situation. Moreover, when two tests are performed in each patient (i.e. a paired study design), the observations of test A and B within each patient are correlated as well (Figure 1). This correlation (in addition to the correlation between segments in patients) should be taken into account when testing the hypothesis of equal diagnostic performance. If this correlation is ignored, the variance of the difference will be too large making it less likely that the null hypothesis of no difference is rejected (1). On the other hand, when the two tests are performed in different patient populations or when considering patient-specific characteristics (e.g. sex, history of CAD), the study design is unpaired (Figure 1), which requires no further adjustment beyond the adjustment for clustering within patients.

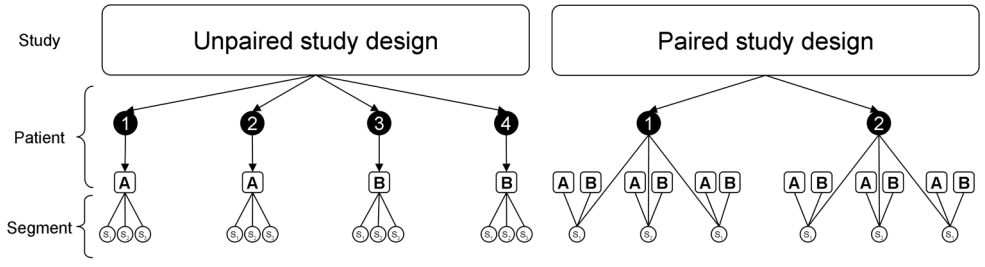


Figure 1. Schematic representation of the unpaired and paired study design. Black circles indicate individual patients, A refers to test A, and B refers to test B, $s_{1,2,3}$ represent multiple segments that are observed per patient. In a paired design, each patient undergoes test A and B. Therefore, not only the correlation between segments in patients, but also the correlation of two test results for each segment should be accounted for. The same concept applies to a design with two readers, two scan protocols etc.

Example: comparing the diagnostic performance of two tests

When comparing sensitivity (or specificity) between two different diagnostic tests, we usually test the null hypothesis that the sensitivity of test A and B are equal, i.e. that the difference equals zero. Hereto, we need to calculate the CI of the *difference*. The appropriate method for this depends on whether the study design was paired or unpaired (Table 1, Appendix 1).

Patient-level and segment-level ignoring correlation

Stratification for type of test and standard tests for independent proportions (and Fisher's exact test for small samples) can be used for unpaired study designs. For a paired study design *McNemar's test* for dependent proportions should be used (which follows a χ^2 distribution for large samples and binomial distribution for small samples). The paired analysis can be extended with covariables using conditional logistic regression.

Variance adjustment: Ratio Estimator and VIF

To calculate the CI of the difference between tests A and B, the variance of the difference needs to be estimated based on the adjusted variances as obtained with the Ratio Estimator or VIF stratified for type of test, for unpaired study designs. For paired study designs, the correlation between sensitivity A and sensitivity B (or specificity A and specificity B) – the covariance – is also required (Appendix 2).

Logistic regression with sandwich estimators, logistic random-effects regression, and GEE

The comparison of tests A and B with regression models is easily carried out by including an indicator variable (e.g. *test_type*) that indicates the test being considered. For unpaired study designs, the Wald-test provided in the statistical output corresponds to the null-hypothesis that the estimated β-coefficient of *test_type* equals zero (i.e. that there is no difference between test A and B) (Table 1). For a paired study design, either a logistic random-effects model with random effects across patients (correlation between segments) as well as random effects across segments within patients (correlation between test A and B), or a GEE model can be used.

Missing data

Missing values occur in almost all clinical studies. An analysis restricted to observations without missing values (complete case analysis) is inefficient because missingness is common. Furthermore, a complete case analysis often results in biased estimates of sensitivity and specificity when systematic differences are present between patients (or segments) with and without complete data with regard to the test results (15). When considering missing test results (missing outcomes), the missing data *mechanism* is important, because the different methods to analyze sensitivity and specificity differ in robustness regarding the missing data mechanism.

The reason for having missing observations (i.e. test results) can be independent of all measured and unmeasured observations. In that case, it is said that the missing data process is missing completely at random, abbreviated as MCAR (Table 2). If missingness of observations is related to measured observations but independent of unmeasured observations, for example if a segment was not analyzed (and therefore "missing") solely because of its location distally to an occlusion, then it is referred to a missing at random process (MAR) (Table 2). Missing not at random (MNAR) occurs when observations are missing because of the nature of the missing observation itself, or when the missingness is related to unobserved or unmeasured data (16). For example, a segment may be uninterpretable and excluded because of severe calcification. In this case, the severity of disease is related to the missingness (Table 2).

The presence of missing data always results in an overestimation of the true variance of the parameter estimates, since less data available results in less precision. Apart from the

Table 1. Overview of methods to analyze sensitivity and specificity of clustered data

Level of analysis	Correction for correlation using:	Advantages (+), and disadvantages (-)	Model for sensitivity (or specificity)*	Method to obtain 95% CI	Method to test the impact of covariables (unpaired design)	Method to test the impact of covariables (paired design)
Patient	Not applicable		Simple proportion (Fig. A1)	Normal approximation of CI Binomial exact CI (small sample)	Stratify and test independent proportions Fisher's exact test (small sample)	McNemar's test for dependent proportions (Fig. A3)
			Logistic regression (Fig. A1)	Statistical output provides CI †	Wald test $\beta_{covariable}$=0 in statistical output	Conditional logistic regression
Segment	No correction	(-) Correlation within patients is ignored	Simple proportion (Fig. A2)	Normal approximation of CI Binomial exact CI (small sample)	Stratify and test independent proportions Fisher's exact test (small sample)	McNemar's test for dependent proportions (Fig. A3)
			Logistic regression (Fig. A2)	Statistical output provides CI †	Wald test $\beta_{covariable}$=0 in statistical output	Conditional logistic regression
	Variance adjustment	(-) Mean estimates are assumed to be unbiased	Ratio estimator (Fig. A1) Variance Inflation factor (Fig. A1)	Use adjusted variance (Appendix 2)	Stratify and use adjusted variance to calculate the variance of the difference	Stratify and use adjusted variance + covariance
			Sandwich estimator (Fig. A2)	Statistical output provides CI †	Wald test $\beta_{covariable}$=0 in statistical output	N/A
	Logistic random-effects modeling	(+) Robust for MAR (-) Sensitivity and specificity estimates conditional on random effects (-) Assumes that model is correctly specified	Logistic regression with random effect across patients (Fig. A2) and segments (if paired design)	Statistical output provides CI †	Wald test $\beta_{covariable}$=0 in statistical output	Wald test $\beta_{covariable}$=0 in statistical output
	GEE	(+) 'Population-averaged' estimates (-) Requires large sample size	As for standard logistic regression, but specifying a correlation structure (Fig. A2)	Statistical output provides CI †	Wald test $\beta_{covariable}$=0 in statistical output	Wald test $\beta_{covariable}$=0 in statistical output

CI = confidence interval; GEE = generalized estimating equations, MAR = missing at random, MCAR = missing completely at random, N/A = not available

** Figure numbers refer to the data structure required to perform the analysis.*

† Statistical output provides the mean and 95% CI on the logit (natural logarithm of the odds) scale, which can be back transformed to sensitivity/specificity scale. When the indicator variable 'test_type' is used (coded 0 for test A and 1 for test B), the intercept of the model reflects Test A, and the sum of the intercept and β_{test_type} refers to Test B. See code for obtaining 95%CI for Test A and B, separately.

bias that occurs because of omitting relevant covariables, missing data can also affect the parameter estimates, however this is dependent on the missing data mechanism and the model being considered.

Applying a GEE model to a dataset with missing test results may cause problems (2), because the GEE approach assumes the MCAR missingness process. In that case the parameter estimates obtained with GEE are consistent (close to the true value). But in many situations missingness occurs at random, for which case weighted GEE methods should be used (17). In contrast, random-effects models allow for MAR missingness, which means that the parameter estimates are consistent (close to the true value) even when the model is applied to data with test results that are missing at random.

For missing data in covariables as opposed to missing data in test results (outcomes), other procedures such as (multiple) imputation methods should be used (16, 18). For missing data in the reference standard that resulted in verification bias, other adjustments are necessary (19-20).

Table 2. Missing data mechanisms.

Mechanism	Description	Missing/ uninterpretable segment caused by:	Results	Solution
Missing completely at random (MCAR)	Missingness independent of other test results and unobserved test results	Motion artifacts, technical problems, high heart rate	Unbiased	-
Missing at random (MAR)	Missingness related to observed data	Variations in coronary anatomy, location distal to an occlusion, segments with stents	Overestimation* of sensitivity and specificity	Use random-effects models
Missing not at random (MNAR)	Missingness related to unobserved test results or the value of the missing observation itself	Calcification	Underestimation of sensitivity	Score segments as positive when uninterpretable due to calcification

** Missing segments caused by variations in coronary anatomy and occlusions are most often distal segments, which are generally smaller and more difficult to interpret. Omitting segments that are more difficult to assess is likely to result in an overestimation of diagnostic performance.*

Results

Table 1 provides an overview of the methods described in this tutorial. See Appendix 1 for statistical code.

As an illustration, Table 3 presents an overview of the results of the dataset on coronary CT with catheter-based angiography as reference test, using the methods described. The number of segments per patient ranged from 9 to 17 (mean 14). The number of diseased segments per patient ranged from 0 to 7 (mean 2.5). The number of non-diseased segments per patient ranged from 4 to 16 (mean 12.2). The intraclass correlation coefficient □ was 0.18 for the diseased segments (38 clusters) and 0.02 for the non-diseased segments (50 clusters) as calculated with the natural estimator for □ . The VIF for sensitivity was 1.46 and for specificity it was 1.22.

Table 3. Estimates of sensitivity and specificity and their 95% confidence intervals calculated with various methods.

Level	Method	Sensitivity	95%CI	Specificity	95%CI
Patient	Binomial proportion	92.1	83.1-101.1	58.3	25.6-91.1
	Binomial proportion (exact CI)	92.1	78.6-98.3	58.3	27.7-84.8
	Logistic regression	92.1	78.2-97.4	58.3	30.8-81.5
Segment	**(1) Analysis ignoring correlation**				
	Binomial proportion	53.8	43.4-64.1	88.3	85.8-90.9
	Binomial proportion (exact CI)	53.8	43.1-64.2	88.3	85.5-90.8
	Logistic regression	53.8	43.6-63.6	88.3	85.5-90.7
	(2) Variance adjustment				
	Ratio estimator	53.8	41.4-66.2	88.3	85.5-91.2
	Variance inflation factor	53.8	41.5-66.0	88.3	85.5-91.2
	Logistic regression with sandwich estimator	53.8	41.4-65.7	88.3	85.2-90.9
	(3) Logistic random-effects model				
	Random effect across clusters *	53.9	37.7-70.1	89.0	85.9-92.1
	(4) Generalized estimating equations				
	Exchangeable correlation matrix	53.1	40.7-65.2	88.3	85.1-90.8

** Estimates conditional on the random effects*

Notice that the patient-level analysis yields a higher sensitivity and lower specificity than the analysis at the segment level. Sensitivity decreased because at the level of segments every stenosis needs to be identified separately to count as a true positive. In other words, by increasing the number of observations (segments) per patient we increase the opportunity to miss a lesion in an individual segment whereas the diagnosis "any" CAD in the patient is usually fairly straightforward. Analogously, specificity increased because at the level of segments multiple segments without disease are added to the dataset increasing the opportunity to call a true negative observation.

The segment-level analyses using variance adjustment yielded slightly wider CIs compared to the analyses that ignored correlation. The three different approaches to perform variance adjustment were very similar. The random-effects model and the GEE approach yielded comparable point estimates and wider CIs compared to the variance adjustment. Please note that a judgement on which method is superior should not be based on the width of the confidence intervals, but rather on whether the estimated confidence intervals reflect the true uncertainty in the data. This cannot be judged from one study but must follow from the above mentioned theoretical considerations and whether these apply to the current data.

The adjusted CIs of the specificity were more similar to the unadjusted CIs, than was the case for the sensitivity. This is explained by the magnitude of the correlation, which was low for the non-diseased segments (= 0.02), and relatively high for the diseased segments (= 0.18).

Discussion

In this paper we summarized commonly used statistical methods for calculating sensitivity and specificity of clustered data, adjusting for the possible correlation between observations within each patient (cluster). In our example, the methods yielded comparable results. However, for other applications results may vary more depending on the method of analysis used, the magnitude of the correlation, and the number of observations per cluster.

In this tutorial we focused on the analysis of data with correlation between observations from the same patient where all observations have been obtained with one or two tests interpreted by one reader. This is the situation that we are most often confronted with. Another similar problem with correlated data has received far more attention in the literature, that is, analysis of multiple observations from the same patient from the performance of multiple tests and/or tests interpreted by multiple readers (8, 21-29). From an analytical point of view the difference between the two situations is irrelevant: both situations involve clustered correlated observations and the methods described can be applied to both. From a practical point of view the adjustment is probably more important in situations with multiple readers. Interpretations from multiple readers are highly likely to be correlated, making adjustment for the clustered nature of the data in those situations really important.

Several previous papers discussed other methods for the analysis of clustered data in the context of diagnostic test evaluation, which were not considered in this tutorial. Our aim was to create a practical guide for radiologists involved in research, discussing the most common methods for adjusting sensitivity and specificity for clustered data. However, adjustment of other statistics and other adjustment methods deserve mentioning. For example, Gonen *et al.* discussed the adjustment of the χ^2 statistic in the case of binary response data and the adjustment of the *t*-statistic in the case of continuous data (30). Both the χ^2 and *t*-statistic can be adjusted by dividing by a correction factor which is equivalent to the VIF described in the methods section (2). In another article, Beam *et al.* described the analysis of clustered data, in particular random-effects models, in the context of receiver-operating-characteristic (ROC) analysis (21-22). Obuchowski *et al.* discussed data analysis of clustered ROC data focusing on the calculation of the area under the ROC curve and its variance (24). Toledano *et al.* presented an ordinal regression methodology for the analysis of correlated ROC curve data (26-27). They applied the method to adjust for multiple interpretations of the same diagnostic study and examination of the same patients with multiple diagnostic modalities. Although an elegant method, it requires dedicated statistical expertise and special software. Rutter *et al.* described a bootstrap approach for estimating sensitivity, specificity, and the ROC curve for patient-clustered data (31).

In this tutorial, we consistently used separate models for sensitivity and specificity. An alternative modeling approach is to use a single model that includes disease status as an independent variable (13). A limitation of using a one-model approach is that the correlations for false-negative and false-positive test results among diseased and non-diseased observations respectively, are combined and assumed to be equal. In our example, the correlation between diseased observations was clearly higher compared to the correlation between non-diseased observations. Furthermore, the two-model approach is easier to understand, calculate, and interpret.

The calculation of sensitivity and specificity relies on the availability of a well-established reference standard. If a reference standard is lacking, kappa statistics could be used to study agreement (32-33).

In some clinical problems involving continuous tissues, the data structure is not as clear as in our example in CAD. The segmentation for continuous tissues can be arbitrary and there can be numerous ways to define it. Subsequently, the estimates of sensitivity and specificity may vary depending on the choice of segmentation. Most importantly, when the number

of segments is increased, the number of non-diseased segments is artificially increased. Lesions are usually small compared to the size of a segment, resulting in relatively more non-diseased segments if a smaller segmentation is chosen. This results in an overestimation of the specificity. The number of diseased segments may also increase, but only if the size of the segments becomes small relative to the usual size of a lesion (one lesion may then cover multiple segments, which increases the number of diseased segments). Zwinderman *et al.* (34) recently proposed a method to calculate sensitivity and specificity taking into account the clustering of the data without the need to define segments. Instead of defining segments, they use a so called "per-lesion" approach and use a Poisson random-effects model to calculate the *number* of false-positive lesions in a patient. Sensitivity defined as the proportion of lesions that is detected was modeled with either GEE or a random-effects model.

In conclusion, this tutorial provides an overview of methods for calculating sensitivity and specificity for clustered data. The choice of method and the level of reporting should always correspond with the clinical context and consequences. If multiple observations per patient are relevant to the clinical decision problem, the potential correlation between observations should be explored and taken into account in the statistical analysis. Statistical code for the various methods is provided in Appendix 1.

Acknowledgements

The authors would like to thank Francesca Pugliese, for providing the patient-level and segment-level data, and Marek Molas for his assistance in writing STATA code.

References

1. Zhou XH, Obuchowski NA, McClish DK. Statistical Methods in Diagnostic Medicine. 2002.
2. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. Hoboken, New Jersey: John Wiley & Sons, Inc., 2003.
3. Kirkwood BR, Sterne JAC. Essential Medical Statistics. Oxford: Blackwell Science, 2003.
4. Ahn C. An Evaluation of Methods for the Estimation of Sensitivity and Specificity of Site-Specific Diagnostic Tests. Biometrical Journal 1997; 39:793-807.
5. Pugliese F, Hunink MG, Gruszczynska K, et al. Learning curve for coronary CT angiography: what constitutes sufficient training? Radiology 2009; 251:359-368.
6. Rao JN, Scott AJ. A simple method for the analysis of clustered binary data. Biometrics 1992; 48:577-585.
7. Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics 2000; 56:645-646.
8. Gatsonis CA. Random-effects models for diagnostic accuracy data. Acad Radiol 1995; 2 Suppl 1:S14-21; discussion S57-67, S61-14 pa.
9. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer, 2000.
10. Mutsvari T, Lesaffre E, Garcia-Zattera MJ, Diya L, Declerck D. Factors that influence data quality in caries experience detection: a multilevel modeling approach. Caries Res 2010; 44:438-444.
11. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986; 42:121-130.
12. Smith PJ, Hadgu A. Sensitivity and specificity for correlated observations. Stat Med 1992; 11:1503-1509.
13. Leisenring W, Pepe MS, Longton G. A marginal regression modelling framework for evaluating medical diagnostic tests. Stat Med 1997; 16:1263-1281.
14. Ronco G, Biggeri A. Estimating sensitivity and specificity when repeated tests are performed on the same subject. J Epidemiol Biostat 1999; 4:329-336.
15. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation and Updating. New York: Springer, 2008.
16. Rubin DB. Inference and missing data. Biometrika 1976; 63:581-592.
17. Jansen I, Beunckens C, Molenberghs G, Verbeke G, Mallinckrodt C. Analyzing Incomplete Discrete Longitudinal Clinical Trial Data. Statistical Science 2006; 21:52-69.
18. Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Hoboken, NJ: Wiley, 2002.
19. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983; 39:207-215.
20. de Groot JA, Bossuyt PM, Reitsma JB, et al. Verification problems in diagnostic accuracy studies: consequences and solutions. BMJ 2011; 343:d4770.
21. Beam CA. Random-effects models in the receiver operating characteristic curve-based assessment of the effectiveness of diagnostic imaging technology: concepts, approaches, and issues. Acad Radiol 1995; 2 Suppl 1:S4-13; discussion S57-60, S68-19 pas.
22. Beam CA. Analysis of clustered data in receiver operating characteristic studies. Stat Methods Med Res 1998; 7:324-336.
23. Obuchowski NA. Multireader, multimodality receiver operating characteristic curve studies: hypothesis testing and sample size estimation using an analysis of variance approach with dependent observations. Acad Radiol 1995; 2 Suppl 1:S22-29; discussion S57-64, S70-21 pas.
24. Obuchowski NA. Nonparametric analysis of clustered ROC curve data. Biometrics 1997; 53:567-578.
25. Obuchowski NA, Lieber ML, Powell KA. Data analysis for detection and localization of multiple abnormalities with application to mammography. Acad Radiol 2000; 7:516-525.
26. Toledano A, Gatsonis CA. Regression analysis of correlated receiver operating characteristic data. Acad Radiol 1995; 2 Suppl 1:S30-36; discussion S61-34, S70-31 pass.

27. Toledano AY, Gatsonis C. Ordinal regression methodology for ROC curves derived from correlated data. Stat Med 1996; 15:1807-1826.
28. Wang F, Gatsonis CA. Hierarchical models for ROC curve summary measures: design and analysis of multi-reader, multi-modality studies of medical tests. Stat Med 2008; 27:243-256.
29. Zhou XH, Li SM, Gatsonis CA. Wilcoxon-based group sequential designs for comparison of areas under two correlated ROC curves. Stat Med 2008; 27:213-223.
30. Gonen M, Panageas KS, Larson SM. Statistical issues in analysis of diagnostic imaging experiments with multiple observations per patient. Radiology 2001; 221:763-767.
31. Rutter CM. Bootstrap estimation of diagnostic accuracy with patient-clustered data. Acad Radiol 2000; 7:413-419.
32. Shoukir MM. Measures of Interobserver Agreement and Reliability. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2011.
33. Vanbelle S, Mutsvari T, Declerck D, Lesaffre E. Hierarchical modelling of agreement. Submitted.
34. Zwinderman AH, Glas AS, Bossuyt PM, Florie J, Bipat S, Stoker J. Statistical models for quantifying diagnostic accuracy with multiple lesions per patient. Biostatistics 2008; 9:513-522.

Appendix 1 – Data structure and statistical code (STATA)

The examples below show the different data structures that are required for patient- and segment-level analysis. A positive test result is coded as 1 and a negative test result is coded as 0. The inverse coding is indicated with "inv_test". When two index tests are of interest, the indicator variable "test_type" can be used.

Patient = patient identification number (identifies the clusters $i = 1 \ldots I$)
j = number of available observations per cluster
TP = number of true positive observations
FN = number of false-negative observations
FP = number of false-positive observations
TN = number of true negative observations
D = number of diseased observations
N = number of non-diseased observations
test = index test result (1 if positive, 0 if negative)
inv_test = inverse of test
disease = presence of disease according to reference test
test_type = indicator variable if multiple index tests are considered, e.g. 0 = test A, 1 = test B

Figure A1. Data structure and coding 1: for **unpaired** and paired* designs

Patient (i)	j	TP	FN	FP	TN	D	N	test	inv_test	disease	test_type
1	14	3	2	1	8	5	9	1	0	1	0
2	17	0	1	0	16	1	16	0	1	1	0
3	12	2	0	2	8	2	10	1	0	1	0
4	15	0	0	2	13	0	15	1	0	0	1
5	16	0	0	0	16	0	16	0	1	0	1
6	17	2	1	5	9	3	14	1	0	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	17	17	17	17	17	17	17	1	1	1	1

** In a paired study design, the variable test_type is omitted and columns n, TP, FN, FP, TN, D, N, test, and inv_test are duplicated. n_A, TP_A…etc. for index test A, and n_B, TP_B… etc. for index test B*

Figure A2. Data structure and coding 2: for **unpaired** and paired* study designs

observation	patient (i)	segment	test	inv_test	disease	test_type
1	1	1	1	0	1	0
2	1	2	0	1	0	0
3	1	3	0	1	0	0
4	1	4	0	1	0	0
5	1	5	0	1	0	0
6	1	6	0	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
17I	I	17	1	1	1	1

** In a paired study design, each observation is entered twice, once with test_type==0 and once with test_type==1*

Figure A3. Data structure and coding 3: for **paired** study designs

observation	Patient (i)	segment	testA	inv_testA	testB	inv_testB	disease
1	1	1	1	0	1	0	1
2	1	2	0	1	0	1	0
3	1	3	0	1	0	1	0
4	1	4	0	1	1	0	0
5	1	5	0	1	0	1	0
6	1	6	0	1	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
17I	I	17	1	1	1	1	1

*(1) Analysis at different levels (ignoring correlation)

** Sensitivity, specificity, and 95% confidence intervals. Patient-level: Fig. A1, segment-level: Fig. A2.*

```
***** Sensitivity
ci test if disease==1
ci test if disease==1, binomial exact
***** Logit of sensitivity
logit test if disease==1

***** Specificity
ci inv_test if disease==0
ci inv_test if disease==0, binomial exact
***** Logit of specificity
logit inv_test if disease==0
```

** Test difference between test A and B for unpaired study design. Patient level: Fig. A1, segment-level Fig A2.*

```
***** Sensitivity
prtest test if disease==1, by(test_type)
tabulate test test_type if disease==1, exact

***** Specificity
prtest inv_test if disease==0, by(test_type)
tabulate inv_test test_type if disease==0, exact
```

** Test difference between test A and B for paired study design. Patient level: Fig. A1, segment-level: Fig. A3.*

- The first line of code tabulates the true positive results for test A vs. B.
  Use the numbers in the table to perform McNemar's test for dependent proportions.

	test_B		Total
test_A	0	1	
0	37	6	43
1	16	34	50
	53	40	93

= mcci 37 6 16 34

```
***** Sensitivity
tabulate test_A test_B if disease==1
mcci … … … …

***** Specificity
tabulate inv_test_A inv_test_B if disease==0
mcci … … … …
```

*(2) Segment-level analysis – a) Ratio estimator

Example excel-sheet can be downloaded via: https://sites.google.com/site/artgroupsite/download

*(2) Segment-level analysis – b) Variance Inflation Factor

Example excel-sheet can be downloaded via: https://sites.google.com/site/artgroupsite/download

*(2) Segment-level analysis – c) Sandwich estimator

** Sensitivity, specificity, and 95% confidence intervals: Fig. A2*

```
***** Logit of sensitivity
logit test if disease==1, vce(cluster patient)
***** Logit of specificity
logit inv_test if disease==0, vce(cluster patient)
```

** Test difference between test A and B for unpaired study designs: Fig. A2*

```
***** Logit of sensitivity
logit test test_type if disease==1, vce(cluster patient)
* Logit sensitivity for Test A
lincom _cons
* Logit sensitivity for Test B
lincom _cons+ test_type

***** Logit of specificity
logit inv_test test_type if disease==0, vce(cluster patient)
* Logit specificity for Test A
lincom _cons
* Logit specificity for Test B
lincom _cons+ test_type
```

*(3) Segment-level analysis – Logistic random-effects model

** Sensitivity, specificity, and 95% confidence intervals: Fig. A2*

```
***** Logit of sensitivity
xtmelogit test if disease==1 || patient:
***** Logit of specificity
xtmelogit inv_test if disease==0 || patient:
```

** Test difference between test A and B for unpaired study designs: Fig. A2*


```
***** Logit of sensitivity
xtmelogit test test_type if disease==1 || patient:
***** Logit of specificity
xtmelogit inv_test test_type if disease==0 || patient:
```

** Test difference between test A and B for paired study designs: Fig.A2*

```
***** Logit of sensitivity
xtmelogit test test_type if disease==1 || patient: || segment:
***** Logit of specificity
xtmelogit inv_test test_type if disease==0 || patient: || segment:
```

*(4) Segment-level analysis – Generalized estimating equations

** Sensitivity, specificity, and 95% confidence intervals: Fig. A2*

```
***** Logit of sensitivity
xtgee test if disease==1, family(binomial) link(logit) i(patient) corr(exchangeable)
vce(robust)
***** Logit of specificity
xtgee inv_test if disease==0, family(binomial) link(logit) i(patient) corr(exchangeable)
vce(robust)
```

** Test difference between test A and B for paired and unpaired study designs: Fig. A2*

```
***** Logit of sensitivity
xtgee test test_type if disease==1, family(binomial) link(logit) i(patient)
corr(exchangeable) vce(robust)
* Logit sensitivity for Test A
lincom _cons
* Logit sensitivity for Test B
lincom _cons+ test_type

***** Logit of specificity
xtgee inv_test test_type if disease==0, family(binomial) link(logit) i(patient)
corr(exchangeable) vce(robust)
* Logit sensitivity for Test A
lincom _cons
* Logit sensitivity for Test B
lincom _cons+ test_type
```

Appendix 2. Background material

(1) Segment-level analysis ignoring correlation

Assume we have $i=1\ldots I$ clusters (i.e. patients), with TP_i true positive, TN_i true negative, FN_i false-negative, and FP_i false-positive observations in the i^{th} cluster. The analysis at the segment-level can then be formulated as follows:

$$sens = \sum_{i=1}^{I} TP_i \Big/ \sum_{i=1}^{I} (TP_i + FN_i) \qquad spec = \sum_{i=1}^{I} TN_i \Big/ \sum_{i=1}^{I} (FP_i + TN_i)$$

From hereon, the summation sign will be used to imply summation across the $i=1\ldots I$ clusters.

Ignoring the clustered nature of the data, the standard errors (and corresponding 95% confidence intervals) would be calculated similar to the above:

$$SE(sens) = \sqrt{\frac{sens(1-sens)}{\sum (TP_i + FN_i)}} \qquad SE(spec) = \sqrt{\frac{spec(1-spec)}{\sum (FP_i + TN_i)}}$$

Comparing the diagnostic performance of test A versus test B

The comparison of sensitivity (or specificity) can be tested using a test of proportions. For **unpaired** study designs (i.e. test A and B are performed in different patients), the test statistics for sensitivity and specificity are:

$$z = \frac{sens_A - sens_B}{\sqrt{Var(sens_A - sens_B)}} \qquad z = \frac{spec_A - spec_B}{\sqrt{Var(spec_A - spec_B)}}$$

which are distributed normally for large samples, with

$$Var(sens_A - sens_B) = Var(sens_A) + Var(sens_B) = \frac{sens_A(1-sens_A)}{\sum (TP_{A,i} + FN_{A,i})} + \frac{sens_B(1-sens_B)}{\sum (TP_{B,i} + FN_{B,i})}$$

$$Var(spec_A - spec_B) = Var(spec_A) + Var(spec_B) = \frac{spec_A(1-spec_A)}{\sum (FP_{A,i} + TN_{A,i})} + \frac{spec_B(1-spec_B)}{\sum (FP_{B,i} + TN_{B,i})}$$

.

For paired study designs, *McNemar's* test for dependent proportions should be used.

(2) Segment-level analysis – Adjusting the variance

Ratio estimator

(Example excel-sheet can be downloaded via: https://sites.google.com/site/artgroupsite/download)

If $sens_i$ and $spec_i$ are the sensitivity and specificity calculated using the data from the i^{th} cluster only, i.e. $sens_i = TP_i / (TP_i + FN_i)$ and $spec_i = TN_i / (FP_i + TN_i)$, and we have D_i and N_i diseased and non-diseased observations, respectively, in the i^{th} cluster (with $D_i = TP_i + FN_i$ and $N_i = FP_i + TN_i$) then overall sensitivity and specificity can be calculated as:

$$sens = \sum D_i sens_i / \sum D_i \qquad spec = \sum N_i spec_i / \sum N_i$$

This yields the same result as before but it is expressed slightly differently. The corresponding standard errors adjusted for clustering are calculated as:

$$SE(sens) = \sqrt{\frac{1}{I_D(I_D-1)} \sum \left(\left(\frac{D_i}{\overline{D}} \right)^2 (sens_i - sens)^2 \right)}$$

$$SE(spec) = \sqrt{\frac{1}{I_N(I_N-1)} \sum \left(\left(\frac{N_i}{\overline{N}} \right)^2 (spec_i - spec)^2 \right)}$$

where $\overline{D} = \sum D_i / I_D$ and $\overline{N} = \sum N_i / I_N$ are the mean cluster sizes for the diseased and non-diseased observations, respectively, and I_D and I_N are the number of clusters (patients) with diseased and non-diseased segments respectively (1).

Ratio Estimator: Comparing the diagnostic performance of two tests (A vs. B)

In an unpaired study design, the comparison of sensitivity or specificity can be carried out using the z-test as described above. However, the variance of the difference between test A and B should be calculated based on the variance of sensitivity (and specificity) which was adjusted for the correlation between segments.

In a paired study design, the covariance between test A and B should be included when calculating the variance of the difference:

$$Var(sens_A - sens_B) = Var(sens_A)_{adjusted} + Var(sens_B)_{adjusted} - 2 \cdot Cov(sens_A, sens_B)$$

$$Var(spec_A - spec_B) = Var(spec_A)_{adjusted} + Var(spec_B)_{adjusted} - 2 \cdot Cov(spec_A, spec_B)$$

where $Var_{adjusted}$ refers to the variance adjusted with the Ratio Estimator, calculated separately for test A and B.

The covariance (under the null hypothesis that both tests are equal) can be calculated using:

$$Cov(sens_A, sens_B) = \frac{1}{I_D(I_D-1)}\Sigma\left(\left(\frac{D_i}{\overline{D}}\right)^2(sens_{A,i}-sens)(sens_{B,i}-sens)\right)$$

$$Cov(spec_A, spec_B) = \frac{1}{I_N(I_N-1)}\Sigma\left(\left(\frac{N_i}{\overline{N}}\right)^2(spec_{A,i}-spec)(spec_{B,i}-spec)\right)$$

where *sens* (and *spec*) denote the overall sensitivity (and specificity) across test A and B (1).

Variance inflation factor

(Example excel-sheet can be downloaded via: https://sites.google.com/site/artgroupsite/download)

The standard errors adjusted using the Variance Inflation Factor (VIF) are calculated as follows:

$$SE(sens) = \sqrt{VIF_{sens}\frac{sens(1-sens)}{\sum(TP_i+FN_i)}} \quad \text{with} \quad VIF_{sens} = \left(1+(\sum\left(\left(\frac{D_i}{D}\right)D_i\right)-1)\rho_D\right)$$

$$SE(spec) = \sqrt{VIF_{spec}\frac{spec(1-spec)}{\sum(FP_i+TN_i)}} \quad \text{with} \quad VIF_{spec} = \left(1+(\sum\left(\left(\frac{N_i}{N}\right)N_i\right)-1)\rho_N\right)$$

where $D=\sum D_i$ and $N=\sum N_i$ and ρ_D and ρ_N are the intraclass correlation coefficients among the diseased and non-diseased segments respectively. The intraclass correlation coefficient (ICC) □ can be estimated using the natural estimator (2) in a spreadsheet (see Excel file). Furthermore, a random-effects model with a random effect across patients can provide an estimate of □ , but sometimes only provides an estimate of the standard deviation (σ_u) of the random effect. The ICC is then calculated as follows:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2+\frac{\pi^2}{3}}$$

Variance Inflation Factor: Comparing the diagnostic performance of test A versus test B

The following formula can be used to calculate the variance of the difference in sensitivity (or specificity), when the variance inflation factor is used to adjust for correlation (2):

$$Var(sens_A-sens_B) = Var(sens_A)_{adjusted}+Var(sens_B)_{adjusted}-2\cdot\frac{Cov(\sum TP_{A,i},\sum TP_{B,i})}{\sum D_{A,i}\cdot\sum D_{B,i}}$$

$$Var(spec_A-spec_B) = Var(spec_A)_{adjusted}+Var(spec_B)_{adjusted}-2\cdot\frac{Cov(\sum TN_{A,i},\sum TN_{B,i})}{\sum N_{A,i}\cdot\sum N_{B,i}}$$

where $Var_{adjusted}$ refers to the variance adjusted with the Variance Inflation Factor calculated *separately* for test A and B, and where the covariance ("Cov") equals zero when the study design is unpaired (i.e. test A and B were performed in different patients). In a paired study design, the covariance is non-zero, and is calculated as follows:

$$Cov(\sum TP_{A,i},\sum TP_{B,i}) = \rho_{AB,\text{diseased}}\cdot\sqrt{sens_A(1-sens_A)\cdot sens_B(1-sens_B)}\cdot\sum(D_{A,i})(D_{B,i})$$

$$Cov(\sum TN_{A,i},\sum TN_{B,i}) = \rho_{AB,\text{non diseased}}\cdot\sqrt{spec_A(1-spec_A)\cdot spec_B(1-spec_B)}\cdot\sum(N_{A,i})(N_{B,i})$$

with:

$$\rho_{AB,\text{diseased}} = \frac{\sum(TP_{A,i}-(D_{A,i}\cdot sens_A))(TP_{B,i}-(D_{B,i}\cdot sens_B))}{\sum(D_{A,i}\cdot D_{B,i}\sqrt{sens_A(1-sens_A)\cdot sens_B(1-sens_B)})}$$

$$\rho_{AB,\text{non diseased}} = \frac{\sum(TN_{A,i}-(N_{A,i}\cdot spec_A))(TN_{B,i}-(N_{B,i}\cdot spec_B))}{\sum(N_{A,i}\cdot N_{B,i}\sqrt{spec_A(1-spec_A)\cdot spec_B(1-spec_B)})}$$

Chapter 14

Summary and Discussion

Part I. Preface

In this thesis, our aim was to determine the optimal diagnostic strategy for patients who are suspected of having coronary artery disease (CAD). We evaluated several imaging tests (CT angiography, perfusion MRI, SPECT, and echocardiography) in terms of **diagnostic performance**. Study populations consisted of patients presenting with stable chest pain complaints at the cardiology outpatient clinic, presenting with acute chest pain at the emergency department, and asymptomatic individuals at high risk for developing cardiovascular disease. Furthermore, we used decision-analytic models to evaluate the long-term **cost-effectiveness** of several diagnostic strategies (CT angiography, stress perfusion MRI, and exercise electrocardiography) for patients with stable chest pain. To improve decisions regarding the optimal diagnostic test for a patient, we validated, updated, and extended **prediction models** for the estimation of the pre-test probability of obstructive CAD. An online step-wise risk calculator was developed that allows risk estimation as clinical information and test results become available.

Part II. Systematic Reviews and Meta-Analyses

Guidelines on imaging of asymptomatic coronary artery disease

Office-based primary prevention programs aim to identify asymptomatic individuals at high risk for developing cardiovascular disease based on established risk factors such as: age, sex, smoking, lipid levels, and blood pressure. Coronary artery disease (CAD), however, has a pre-clinical detectable phase (i.e. atherosclerosis). Imaging of pre-clinical atherosclerosis could potentially help to identify high-risk individuals. Since randomized controlled trials (RCTs) investigating the impact of such imaging tests are lacking, guidelines for the use of imaging for asymptomatic CAD may vary. We systematically searched the literature for published guidelines and summarized the available recommendations regarding (imaging) test strategies for asymptomatic CAD. Rigor of development was assessed using the AGREE instrument. We found 14 guidelines with 26 recommendations on various tests for CAD: CT calcium scoring, CT angiography, MR angiography, single-photon emission CT, positron emission tomography, stress echocardiography, resting electrocardiography, and exercise tolerance testing. The AGREE scores varied from 21%-93%. Eight of the 14 guidelines recommended against or concluded that there is insufficient evidence for testing of asymptomatic CAD. In the remaining 6 guidelines, testing was only advocated for patients with an a priori elevated risk level. The majority of these guidelines supported consideration of CT calcium scoring in case of intermediate CAD risk. Our systematic review demonstrates that existing guidelines contain conflicting recommendations regarding risk assessment by imaging of asymptomatic CAD. More research, including RCTs, to determine the impact of imaging on clinical outcomes and costs is needed.

Diagnostic performance of myocardial perfusion imaging

Stress myocardial perfusion imaging (MPI) can be used in the diagnostic work-up of patients who present with stable chest pain, by assessing the presence of inducible perfusion defects which in turn is indicative for obstructive CAD. Different techniques are available to quantify myocardial perfusion, including magnetic resonance imaging (MRI), contrast-enhanced echocardiography (ECHO), single-photon emission computed tomography (SPECT), and positron emission tomography (PET). It is difficult to compare the diagnostic performance

of these imaging techniques, because published meta-analyses have used different methodologies regarding publication period, searching the literature, selection of the evidence, and analysis of the data. To overcome this problem we performed a systematic review of different MPI techniques using the same selection criteria and statistical analysis for all techniques in order to make a fair comparison between these imaging tests. We also excluded studies with (potential) verification bias. Our results suggested that stress perfusion MRI is superior to stress perfusion ECHO and SPECT for diagnosing obstructive CAD. We found a similar diagnostic performance for ECHO and SPECT.

Imaging for acute chest pain in the emergency department

Acute chest pain represents a common diagnostic dilemma in the emergency department. Emergency physicians and cardiologists are commonly required to make a decision whether or not to admit a patient, based on clinical judgment and a rough estimation of risk, but without conclusive evidence whether or not an acute coronary syndrome (ACS) is developing. Patients with ACS frequently present with atypical chest pain complaints, a normal physical examination and an ECG that is either poorly interpretable or has normalized at presentation. ACS cannot be ruled out based on the initial assessment alone, which generally requires clinical observation and sequential testing. Optimal triage requires a quick non-invasive test to identify patients with ACS and to accurately identify patients in whom CAD can be ruled out. We systematically reviewed the literature on imaging modalities used for patients with acute chest pain. We found that the diagnostic performance of rest ECHO, rest SPECT, and coronary CT angiography for the detection of ACS (diagnosis group 1) were similar. Furthermore, the diagnostic performance of stress ECHO, stress SPECT and coronary CT angiography for the detection of CAD (diagnosis group 2) was also similar. There was a small, non-significant difference favoring coronary CT angiography, which was the most sensitive test for both outcomes. We concluded with a recommendation for coronary CT angiography and noted that the optimal imaging strategy, given the absence of large differences in diagnostic performance, should also depend on local expertise, availability of equipment, and individual patient characteristics.

Discussion

We performed systematic reviews and meta-analyses to summarize and pool available evidence on guidelines and diagnostic test performance. Some limitations to these methods deserve mentioning. First of all, although the search strategies were carefully designed, we still may have missed relevant studies. Relevant studies may also be excluded during the review process because of poor study design or poor reporting (1). In addition, a 'positive' study (i.e. a study showing a significant effect or a significant difference) may be more likely to be published (2). This phenomenon (i.e. publication bias) leads to selectively including studies showing a (significant) difference, which in turn may lead to overoptimistic results. In Chapter 4, the presence of publication bias was assessed by creating funnel plots. The funnel plots showed evidence for publication bias regarding studies on the diagnostic performance of MRI and SPECT, which means that our summary estimates of sensitivity and specificity for MRI and SPECT may be overestimated.

Other challenges in reviewing the literature include quality-assessment and heterogeneity across studies. We used well-established and validated tools to assess study quality, namely the AGREE instrument (3) for guidelines and the QUADAS tool (1) for studies of diagnostic performance. Apart from the subjectivity involved in scoring study quality, which to

some extent is accounted for by working with two independent reviewers, the quality score may only reflect quality of *reporting*. Thus, a well-performed study may receive a low quality score if the methods are poorly described. Although the reverse is also possible, i.e. a poorly-performed study receives a high quality score , this is less likely to occur.

Even though we systematically searched for studies with a specific PICO-question (Patient, Intervention, Comparison, and Outcome), our search results included studies that were very heterogeneous regarding study population, in- and exclusion criteria, methods, protocols, outcome definition, and follow-up duration. Often, the question is asked whether the studies are comparable and whether they should be pooled at all. In the meta-analyses of diagnostic performance, we used bivariate random-effects models that allow for heterogeneity across studies and take into account the inverse relationship between sensitivity and specificity (4). Furthermore, we use meta-regression analysis to explore the effects of differences between studies. In Chapter 5, we reviewed the evidence on the diagnostic accuracy of different imaging techniques in patients with acute chest pain. The studies obtained were heterogeneous mainly because of differences in the definition of the outcome (i.e. the reference standard used). Some studies used the elevation of cardiac biomarkers as the reference standard, whereas others used obstructive CAD on CAG, or the presence or absence of adverse cardiac events during follow-up. This heterogeneity is likely due to the fact that there is limited consensus among experts regarding the best reference standard. Importantly, different imaging techniques focus on different aspects of the ischemic cascade (see Introduction). For example, a rest perfusion imaging test assesses the presence of myocardial ischemia at rest and is therefore specific in diagnosing a myocardial infarction. Anatomical tests and stress perfusion imaging tests, on the other hand, focus on identifying the presence of obstructive CAD and inducible ischemia, respectively. Obstructive CAD does not necessarily result in a myocardial infarction, which is why these tests may perform better when the outcome is defined as CAD, as opposed to 'myocardial infarction'. To address the issue of heterogeneity with respect to outcome, we stratified our analysis in Chapter 5 for two categories of the outcome. When planning a study on the diagnostic performance of a test for the assessment of acute chest pain, one should consider whether the aim is to exclude the presence of a myocardial infarction or to exclude the presence of obstructive CAD.

Future research

To resolve disagreements between guidelines of early detection of CAD, more evidence regarding the benefits and harms of screening is needed. Although randomized clinical trials can provide such evidence, these studies are expensive, time-consuming and often not feasible. Future research should therefore focus on the development of decision models that evaluate screening strategies by quantitatively weighing the benefits and harms and integrating the best-available evidence from multiple sources.

An increased comparability between outcomes used in studies that evaluate diagnostic tests for acute chest pain would improve the quality of future systematic reviews and meta-analyses. To accomplish this, experts in the field and guideline writing committees could, for example, try to achieve consensus on one or more preferred outcome definitions. If researchers have an incentive to adopt the recommended outcome definitions (e.g. if a paper is more likely to be accepted for publication), then future studies of diagnostic test performance are likely to be more homogeneous regarding the outcome definition.

Part III. Prediction Models

The incremental value of the CT coronary calcium score

The CT coronary calcium score has been demonstrated to predict the occurrence of future cardiovascular events in both asymptomatic individuals and symptomatic individuals suspected of having CAD. In 254 patients with chest pain who were referred for CAG, we validated five existing prediction models for the presence of obstructive CAD and we assessed the incremental value of the CT coronary calcium score as a continuous predictor beyond the clinical predictors included in each model. Age, sex, and symptoms were the most important clinical predictors for the presence of obstructive CAD. The CT coronary calcium score was a strong predictor as well, independent of other predictors in the model. Smoking, hypertension, dyslipidaemia, and diabetes were positively associated with the presence of CAD. Addition of the CT coronary calcium score increased the area under the receiver-operating-characteristic curve (c-statistic) by 0.05-0.09 across the five models, depending on which other predictors were included. Furthermore, adding the CT coronary calcium score resulted in reclassification of 38-47% of the patients, of which 64-84% was correct. The CT coronary calcium score carries incremental diagnostic value for the presence of obstructive CAD, independent of other predictors, and may be considered in the diagnostic work-up of patients with chest pain.

Validation and updating of Diamond & Forrester

Although many risk factors for CAD have been identified, e.g. diabetes, hypertension, lipid levels, smoking, and the CT coronary calcium score, the estimation of the pre-test probability of CAD is often based solely on age, sex, and symptoms, according to the prediction model described by Diamond & Forrester in 1979. The pre-test probability is crucial both for choosing the appropriate diagnostic test for a particular patient and for interpreting the test results. We combined existing data from 14 different hospitals to assess the validity of the Diamond & Forrester prediction model. Furthermore, we updated and extended the model using modern statistical methods. Results suggested that the Diamond & Forrester model overestimates the probability of obstructive CAD, especially in women. The predictive effects of age, sex, symptoms, and hospital were updated and the model was extended to include patients of 70 years and older. The updated model predicted less high probabilities, resulting in down-classification for most patients. Finally, the model was re-calibrated specifically for a low-risk population of patients presenting with chest pain. Online risk calculators were developed for estimating the pre-test probability of CAD in both high- and low-risk populations. These calculators have potential value in clinical practice since an updated estimate of the pre-test probability can help clinicians make better decisions as to whether and which diagnostic strategy is indicated in a particular patient and to decide on further management based on the results of such tests.

A prediction model for the presence of CAD

Current guidelines recommend either the Diamond & Forrester model or the Duke Clinical Score to estimate the pre-test probability of CAD. In contrast to the Diamond & Forrester model, the Duke Clinical Score does take into account cardiovascular risk factors, but has not been validated in populations outside the United States. We combined individual patient data from 18 different hospitals to assess the validity of the Duke Clinical Score. Some hospitals provided a dataset with patients who were referred to coronary CT angiography (low-prevalence setting), whereas other hospitals provided a dataset with patients who were

referred to CAG (high-prevalence setting). We found that the Duke Clinical Score overestimated the probability of CAD. We updated the model adjusting for setting and assessed the incremental value of adding the CT coronary calcium score. Our results confirmed the importance of age, sex, symptoms, and setting in the prediction of CAD. Furthermore, the presence of diabetes, hypertension, dyslipidemia, and smoking were predictive. The coronary CT calcium was a strong predictor which significantly increased the c-statistic. The validity of the prediction model was assessed in a cross-validation procedure. We found satisfactory calibration in most analyses, suggesting that the model is valid for different hospitals. Finally, we created an online step-wise risk calculator that allows estimation of the probability of CAD as clinical information and test results become available.

Discussion

To validate and update prediction models, we combined existing individual patient data from multiple hospitals. An important limitation of this approach concerns heterogeneity. Because our analyses were not the main purpose of the data collection, differences in various aspect of the data collection, local protocols, level of physician experience, and guideline adherence across hospitals could have influenced our results. In Chapter 8 we assessed the heterogeneity in predictor effects across hospitals by performing cross-validation. The five largest datasets with sufficient numbers for reliable validation and the remaining databases combined were each in turn omitted for model development, and used for validation. Some hospitals demonstrated significantly higher or lower predictor effects, but in general they were quite similar. Calibration was satisfactory in most cross-validation procedures.

In Chapter 4 we studied the incremental value of the CT coronary calcium score to predict the presence of CAD in a population referred for CAG. By selecting patients based on the availability of CAG, we selected a high-risk population, which may not represent the population in which the CT coronary calcium score would be most useful. Unfortunately, this limitation is inherent to our study design, which aims to validate and update models that predict the presence of obstructive CAD, for which the reference standard is CAG. In Chapter 7 we also selected patients based on referral to CAG, however, we applied a correction for verification bias in a sensitivity analysis and found similar estimates of the predictor effects. In Chapter 8, our aim was to develop a prediction rule that would be directly applicable to low-risk populations, exactly the population in which such a model would be clinically useful. In addition to patients referred for CAG, we also included patients who underwent coronary CT angiography without subsequent CAG. Since coronary CT angiography is highly sensitive and specific (5-7), CAG results were multiply imputed based on all other covariables as well as the coronary CT angiography result (8). This method too, has limitations, because coronary CT angiography is known to overestimate the severity of CAD, which in turn could have overestimated the severity of disease for patients with imputed CAG results. If present, the bias in our models would tend to advocate further diagnostic workup rather than missing the diagnosis.

Although CAG is commonly considered to be the gold standard for diagnosing CAD, its limitation deserves mentioning. CAG may be the reference standard for determining the degree of stenosis (anatomical test), it is well recognized that the degree of stenosis does not correlate well with the 'hemodynamic' significance of a stenosis. A patients' prognosis and the optimal treatment depend also on the presence of inducible ischemia, which is not captured by CAG. The fractional flow reserve (FFR) of a coronary stenosis can be measured

during a CAG procedure, quantifying its hemodynamic significance. Although FFR measurements increase costs, a combination of a functional and anatomical testing for CAD may represent a better gold standard compared to anatomical testing alone.

Several methods to quantify the performance of a prediction model are available. A prediction model should not only accurately distinguish patients with and without disease (discrimination) and provide a prediction that is close to the observed (calibration), but a new model or the addition of a new predictor should also move patients across clinically meaningful thresholds that would alter clinical management (reclassification) (9). For CAD however, such well-established thresholds are not available. Three categories (low, intermediate, and high) are often used, with an intermediate probability defined as 10-90% or 30-70%, for example. However, there is no consensus on how many categories we should use and what the optimal thresholds would be. Ideally, the thresholds are determined based on cost-effectiveness analyses that take into account long-term benefits and harms of diagnostic tests and treatment for CAD.

Future research

We demonstrated that our updated models outperform existing models. However, we did not study the clinical impact of implementation of the model. Future studies validating our prediction model in independent patient cohorts would support its implementation in clinical practice. A continuous updating strategy could be considered if the model is applied in a new setting. As new patient information becomes available over time, the model could be validated continuously and updated if necessary.

An extension to our work regarding prediction models for the pre-test probability of CAD includes the prediction of prognosis in terms of future cardiovascular events. After a diagnosis of CAD has been established, the optimal treatment depends on the individual patient's prognosis and symptom status (quality-of-life). The prognosis of a patient with CAD depends on the severity of CAD, presence and extent of myocardial ischemia, left ventricular function, and other patient characteristics such as age, sex, and risk factors. Patients with a favorable prognosis are usually treated with medical therapy, whereas patients with an unfavorable prognosis may benefit from revascularization. The variables that determine prognosis and the possible treatment options could be combined into a prediction model and allow for patient-tailored treatment decisions.

Part IV. Cost-effectiveness Analysis

Cost-effectiveness of coronary CT angiography as triage test

Catheter-based coronary angiography (CAG) is the reference standard for diagnosing CAD. Since CAG is expensive, and carries a small risk of complications and death, non-invasive testing is recommended to select patients who would benefit from CAG. Coronary CT angiography has emerged as a non-invasive diagnostic modality for CAD. The cost-effectiveness of coronary CT angiography as a triage test prior to performing CAG was studied using a decision tree and Markov-model. Using best-available evidence from the literature and data from a patient cohort where possible, we analyzed the decision from multiple perspectives using several available recommendations for cost-effectiveness analyses. The threshold probability below which coronary CT angiography would be cost-effective was 44% in men

and 37% in women. The optimal diagnostic work-up depended on the optimization criterion, prior probability of CAD, and the diagnostic performance of coronary CT angiography.

Cost-effectiveness of coronary CT angiography vs. exercise testing

Over the past years, several systematic reviews and meta-analyses have demonstrated the diagnostic and prognostic value of coronary CT angiography. Because of its high negative predictive value (>95%), coronary CT angiography is reliable in ruling out the presence of CAD. Consequently, coronary CT angiography could potentially be used as a first line diagnostic test in patients who present with stable chest pain. We used a Dutch clinical cohort with follow-up in combination with the best-available evidence to construct a decision tree and Markov-model. We designed a hypothetical CT-based strategy which consisted of CT coronary calcium scoring with or without subsequent coronary CT angiography. The reference strategy reflected standard-of-care, which consisted of an exercise test (electrocardiography) and if non-diagnostic a pharmacological stress single-photon emission CT scan. The main difference with the model described in Chapter 9 is the fact that patients with an intermediate stenosis are not necessarily referred for catheter-based angiography. The COURAGE trial (10) has demonstrated that a percutaneous coronary intervention (PCI) does not necessarily reduce the risk of cardiovascular events in patients with CAD (defined as ≥70% stenosis in ≥1 epicardial artery) and documented myocardial ischemia. This is why an initial conservative strategy using optimal medical treatment is currently advocated for patients with mild CAD, with the possibility to intervene at a later stage if symptoms persist despite optimal medical treatment. Therefore, this analysis did not assume that every patient with CAD on coronary CT angiography is referred for CAG. Partly explained by the latter, the CT-based strategy was found to be cost-saving and equally effective as compared to standard-of-care.

Cost-effectiveness of coronary CT angiography vs. stress myocardial perfusion cardiac MRI

Our systematic review and meta-analysis on the diagnostic performance of myocardial perfusion imaging demonstrated that stress perfusion imaging with cardiac MRI (CMR) was superior to stress echocardiography and stress SPECT. However, the question remained whether a diagnostic work-up using CMR (a functional test), or using coronary CT angiography (an anatomical test), or a combination of coronary CT angiography and stress perfusion CMR would be optimal in terms of cost-effectiveness. We developed a decision model that compared these strategies, using both an invasive and conservative approach. The invasive approach implied that all patients with positive test results are referred for CAG, whereas the conservative approach implied that patients with mildly abnormal test results were treated medically without referral to CAG. The decision was analyzed for the United States (US), United Kingdom (UK), and the Netherlands (NL), using country-specific cost-estimates and country-specific recommendations for cost-effectiveness analysis. Our base case analysis (men and women with a 30% pre-test probability of obstructive CAD) demonstrated that a strategy using coronary CT angiography, if positive followed by CMR (invasive approach), was cost-effective for the US and NL. The strategy using CCTA, if positive followed by CMR (conservative approach), was cost-effective for the UK.

Discussion

Determining the optimal diagnostic work-up for a patient presenting with chest pain remains a challenge. Numerous diagnostic tests are available and even more combinations of tests may be considered. In general, the optimal work-up depends on the pre-test probability of coronary artery disease (CAD), the benefit and harm of treating CAD, and the performance (i.e. sensitivity and specificity) of the available diagnostic tests (11). However in practice, the choice of work-up is also driven by the availability of imaging technology, local experience, the tendency to stick with existing protocols, personal physician preferences and biases, and financial incentives.

Ideally, large randomized controlled trials (RCT) comparing multiple diagnostic strategies with standard-of-care should provide the evidence regarding the optimal diagnostic work-up for patients suspected of having CAD. Unfortunately, RCTs are often lacking because performing such trials is often not feasible or too costly. Second, within a RCT only a limited number of strategies can be compared, whereas numerous strategies should be considered. Third, follow-up is usually short whereas the benefits and harms of treating CAD are likely to persist beyond the completion of the trial. Fourth, results of RCTs may only be generalized to patients that meet the in- and exclusion criteria for the trial. Finally, the fact that clinical outcomes are often similar across diagnostic strategies implies that such trials require large sample sizes and long follow-up duration to demonstrate a clinically relevant difference. When RCTs are lacking or results of RCTs are not yet published, decision models that integrate the best-available evidence can guide decision-making. For example, we showed that an initial strategy based on CT coronary calcium scoring with or without subsequent coronary CT angiography is cost-saving and equally effective when compared to standard-of-care (exercise ECG with or without stress SPECT). Similar strategies are now being evaluated in multi-center randomized clinical trials (e.g. the PROMISE and RESCUE trial in the United States and the CRESCENT trial in the Netherlands, ClinicalTrials.gov identifiers: NCT01174550, NCT01262625 and NCT01393028, respectively). As long as trial results are unavailable, clinical decisions have to be made and should be made based on the best-available evidence, which we integrated in our decision models.

Some limitations of cost-effectiveness analyses using decision models deserve mentioning. Firstly, decision models usually constitute a simplification of clinical practice, and assumptions are unavoidable. Modeling of all possible clinical pathways is unpractical and often not feasible. To be able to answer the research question, a decision model should incorporate the most important clinical scenarios and capture the key trade-offs of the clinical problem at hand. Developing a model that is both sufficiently sophisticated to reflect clinical practice and simple enough for readers to understand, will always remain a challenge. Secondly, a cost-effectiveness analysis is often framed from a certain perspective (patient, health care, or payer) for a specific country. Health care costs, recommendations for cost-effectiveness analysis (12-16), and standard care may vary considerably across countries, which limits the generalizability. Lastly but possibly most importantly, the validity of a decision model is always limited by the quality of the input data. We performed extensive one- and two-way sensitivity analysis by varying the parameters values across plausible ranges to see whether such changes would alter the results. Furthermore, we used distributions that reflect the uncertainty around the parameter values, and performed probabilistic sensitivity analysis (17).

The benefits of coronary CT angiography in terms of its diagnostic and prognostic value have been studied extensively. However, the question whether coronary CT angiography actually improves long-term patient outcomes remains unresolved. While large multi-center trials are ongoing, the discussion regarding potential harms of coronary CT angiography continues.

After coronary CT angiography is completed, the physician who interprets the scan has the option to also evaluate non-cardiac structures, which may lead to incidental findings. Whether the evaluation of non-cardiac structures should be a routine procedure, or whether the scan should exclusively be used to assess cardiac structures, is currently under debate (18-19). There is concern that the high prevalence of incidental findings will lead to unnecessary additional diagnostic testing, increasing health care costs, and patient discomfort, especially since there is no evidence that follow-up of incidental findings improves health outcomes. Numerous studies have reported a high prevalence of incidental findings (20). The prevalence of incidental findings of unknown clinical significance (mostly pulmonary nodules) that would require subsequent testing is approximately 7% (21). In Chapter 11, we included the consequences of incidental findings (in terms of extra costs, radiation, and disutility). We conclude that the negative consequences of incidental findings do not outweigh the benefits of coronary CT angiography as compared with CMR or CAG, since the combination strategy (CCTA, if positive followed by CMR) demonstrated favorable cost-effectiveness.

Lastly, the exposure to radiation and its potential risk of radiation-induced cancer remains a topic of discussion. In 2009, a multicenter observational study reported an average effective dose for coronary CT angiography of 12 mSv (interquartile range: 8-18), with large variation depending on study site and type of scanner used (22). Epidemiologic models based on data from atomic bomb survivors as well as medical and occupational radiation studies, were used to determine the potential cancer risk of coronary CT angiography. A single cardiac scan with an effective dose of 9 and 14 mSv (for men and women, respectively) was estimated to result in a lifetime attributable risk of radiation-induced cancer of 1 in 1911 and 1 in 715 for 60-year old men and women, respectively (23). In accordance with the as-low-as-reasonably-achievable (ALARA) principle, new-generation scanners and improved scanning protocols have been and are being developed in order to minimize radiation exposure. Nowadays, the radiation exposure in non-obese patients with a low and stable heart rate can be reduced to less than 1 mSv (24), although the average patient may still need ~5 mSv. Incorporating the potential cancer-inducing effects of radiation exposure into a cost-effectiveness analysis is a major challenge, but is feasible when simplifying assumptions are used (25-26). The benefit of a correct diagnosis in a symptomatic patient, however, is likely to outweigh the risks of the radiation exposure. Moreover, alternative diagnostic strategies are usually associated with radiation exposure as well. Consequently, the difference in total radiation exposure across different diagnostic strategies may be small, which in turn makes it less likely to have an effect on the decision.

Future research

Advances in technology are rapidly evolving, and new diagnostic imaging techniques are continuously being developed. Performing randomized trials for all new diagnostic imaging techniques or improvements to existing modalities is not feasible. Therefore, cost-effectiveness analyses that compare existing diagnostic strategies with newly developed diagnostic techniques remain relevant in the future. Such analyses can inform researchers as to whether it is a reasonable alternative and whether it is worth performing randomized trials. If so, the results of such analyses can then be used to guide decision-making for as long as trial results are unavailable.

Furthermore, future decision models should focus on analysis of cost-effectiveness at the individual level as opposed to the population level. Multivariable prediction models that calculate the pre-test probability of CAD, diagnostic test performance, long-term prognosis, and treatment benefits and harms for individuals could be integrated into decision models. Ultimately, such a model could be implemented in electronic patient record systems, and automatically recommend a cost-effective diagnostic work-up and treatment based on patient characteristics from the electronic patient record.

Part V. Other

Prognostic value of coronary CT vs. exercise testing

We studied 471 outpatients with stable chest pain and no history of CAD who were scheduled for exercise testing (exercise electrocardiography) and coronary CT angiography. After a mean follow-up time of 2.6 years, 44 major adverse cardiac events (MACE) (cardiac death, nonfatal myocardial infarction, unstable angina requiring hospitalization, and revascularization) occurred in 30 patients. Patients with a non-diagnostic exercise test and those with obstructive CAD on coronary CT angiography were more likely to experience adverse events. The coronary CT angiography findings demonstrated incremental prognostic value beyond the CT coronary calcium score as well as beyond exercise testing. The CT coronary calcium score provided no incremental prognostic value beyond exercise testing and CT angiography.

Discussion

We studied a population representative of patients who would be eligible for coronary CT angiography in clinical practice. All patients underwent both exercise testing (exercise electrocardiography) and coronary CT angiography. The fact that all patients in our study underwent both tests is a strength of our study, but it is also a weakness. We were able to directly compare the prognostic value of exercise testing and coronary CT angiography, and determine the incremental value of one over the other. However, the decision to perform a revascularization was based on the coronary CT angiography findings as well as the exercise test results, which may have influenced our results because revascularizations were included in our composite outcome. To limit this possible bias, we excluded revascularizations performed within the first 6 months of follow-up.

The clinical impact of adding exercise testing, coronary CT angiography, and the coronary calcium score to models that predict the risk of MACE was determined using reclassification statistics. In this context however, no clinically relevant cutoffs for the risk of MACE are available (i.e. thresholds of risk where reclassification to another category would influence clinical management). We therefore used an extension of the traditional net reclassification improvement (NRI), the so-called continuous NRI (27). This measure does not depend on the existence of risk categories and also allows for application to survival data. The continuous NRI is the weighted sum of the observed event rate increase among the individuals for whom the predicted risk goes up and observed event rate decrease among those for whom the predicted risk goes down. Consequently, it does not take into account the magnitude of change in predicted risk.

Methods for analyzing sensitivity and specificity of clustered data: a tutorial

The performance of a diagnostic imaging test is usually expressed in terms of sensitivity and specificity compared with the reference standard. The calculation of sensitivity and specificity commonly involves multiple observations per patient, which implies that the data are clustered. Whether analysis of sensitivity and specificity per patient or using multiple observations per patient is preferable, depends on the clinical context and consequences. Observations within patients (clusters) are often positively correlated which may lead to biased estimates if the correlated nature of the data within each cluster is ignored. We discussed and illustrated the most common statistical methods that calculate sensitivity and specificity of clustered data: 1) analysis at different levels ignoring correlation; 2) variance adjustment; 3) logistic random-effects models; and 4) generalized estimating equations.

Discussion

In our example analysis based on 50 patients with ±17 coronary segments each, we compared the different methods that adjust for the clustered nature of the data and found similar results. This finding is explained by the relatively low correlation between observations (low intraclass correlation coefficient). However, for other applications, such as a study in which multiple clinicians read the same images, the correlation between observations is likely to be higher and correcting for clustering is probably more important. Furthermore, when two different tests are performed in each patient (paired study design); observations are also likely to be highly correlated.

A limitation of the methods covered in our tutorial article is related to the fact that the models assume that sensitivity and specificity are equal across the correlated observations (in our example: coronary segments). Consequently, the methods do not allow for differences in diagnostic performance across observations. However, as in our example, it may well be the case that the sensitivity for detecting a stenosis in a proximal segment is higher compared to the sensitivity for detecting a stenosis in a distal vessel. In other words, sensitivity and specificity of CT in the detection of obstructive CAD may vary depending on the size and location of the segment. Solutions to this limitation exist. Firstly, one could stratify the analysis by defining groups of segments (e.g. proximal vs. middle vs. distal segments) for which the assumption of equal sensitivity seems more appropriate. Secondly, multivariate models can be used to allow for multiple correlated outcomes (e.g. multiple sensitivities). For simplicity reasons, we did not cover these methods in our tutorial. Furthermore, the assumption of equal sensitivity/specificity across observations may be more reasonable for other clinical areas or when analyzing correlated observations in terms of multiple readers. Lastly, we are generally interested in the mean sensitivity and specificity across segments, for which the methods described are considered appropriate.

Conclusions and implications for clinical practice

In this thesis, our aim was to determine and optimize the diagnostic work-up for patients who are suspected of having CAD. We showed that the diagnostic performance of stress perfusion MRI compares favorably to the diagnostic performance of SPECT. If both tests can be performed in a patient, then stress perfusion MRI should be the preferred test, even more so because it does not involve exposure to radiation.

We demonstrated that the coronary CT calcium score has predictive value beyond existing cardiovascular risk factors for diagnosing obstructive CAD in patients with chest pain. The CT coronary calcium score can be considered as an initial triage test in patients with a low pre-test probability of CAD, preventing (unnecessary) further work-up if the score is zero, and justifying further testing when coronary calcium is present. This strategy is recommended for patients with a pre-test probability between 10-29% by the 2010 NICE guidelines for the management of patients with chest pain (12).

The optimal diagnostic strategy depends on the pre-test probability of CAD, which is traditionally estimated based on the Diamond & Forrester method (28) or the Duke Clinical Score (29). We demonstrated that these prediction rules systematically overestimate the probability of CAD and we updated the models based on contemporary data. The online probability calculator can be accessed via the internet and provides systematically lower (but more accurate) estimates of the pre-test probability. Although we did not study the clinical impact of implementing our prediction model, a more accurate estimate of the pre-test probability leads to better decisions regarding further testing and it could potentially reduce costs since less high probabilities are predicted which in turn may prevent unnecessary diagnostic work-up.

Lastly, we evaluated the long term effectiveness and costs of coronary CT angiography in various different settings and for various countries. In the Dutch setting, coronary CT angiography was found to be cost-effective as triage test prior to CAG if the pre-test probability was below 44% in men and below 37% in women. CT coronary calcium scoring with or without subsequent coronary CT angiography as initial strategy for patients presenting with stable chest pain was less expensive and equally effective compared to standard-of-care. Finally, we showed that a strategy using coronary CT angiography, if positive followed by CMR was cost-effective compared to strategies with coronary CT angiography and CMR alone, for the US, the UK, as well as the Netherlands.

All-in-all, our updated prediction models combined with the results from our decision models and cost-effectiveness analyses provide a practical framework for efficient implementation of diagnostic imaging tests, in particular for the CT coronary calcium score and coronary CT angiography.

References

1. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3:25.
2. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010; 14:iii, ix-xi, 1-193.
3. Agree Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003; 12:18-23.
4. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MG, Stijnen T. Bivariate random effects meta-analysis of ROC curves. Med Decis Making 2008; 28:621-638.
5. Schuetz GM, Zacharopoulou NM, Schlattmann P, Dewey M. Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. Ann Intern Med 2010; 152:167-177.
6. Mowatt G, Cook JA, Hillis GS, et al. 64-slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. Heart 2008.
7. von Ballmoos MW, Haring B, Juillerat P, Alkadhi H. Meta-analysis: Diagnostic Performance of Low-Radiation-Dose Coronary Computed Tomography Angiography. Ann Intern Med 2011; 154:413-420.
8. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011; 30:377-399.
9. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. Ann Intern Med 2009; 150:795-802.
10. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 2007; 356:1503-1516.
11. Hunink MGM, Glasziou PP, Siegel JE, et al. Decision making in health and medicine: Integrating evidence and values. Cambridge: Cambridge University Press, 2001.
12. National Institute for Clinical Excellence (NICE). Guide to the methods of technology appraisal. Oxford, England: Radcliff Medical Press, 2004.
13. Gold MR, et al. Cost-effectiveness in Health and Medicine. Journal of Mental Health Policy and Economics 1996; 2:91-92.
14. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. Jama 1996; 276:1339-1341.
15. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. Jama 1996; 276:1253-1258.
16. Hakkaart-van Roijen L, Tan SS, Bouwmans C. Dutch Manual for Cost Analysis. Dutch Health Care Insurance Board 2010.
17. Groot Koerkamp B, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink MG. Uncertainty and patient heterogeneity in medical decision models. Med Decis Making 2010; 30:194-205.
18. Earls JP. The pros and cons of searching for extracardiac findings at cardiac CT: studies should be reconstructed in the maximum field of view and adequately reviewed to detect pathologic findings. Radiology 2011; 261:342-346.
19. White CS. The pros and cons of searching for extracardiac findings at cardiac CT: use of a restricted field of view is acceptable. Radiology 2011; 261:338-341.
20. Jacobs PC, Mali WP, Grobbee DE, van der Graaf Y. Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review. J Comput Assist Tomogr 2008; 32:214-221.
21. Machaalany J, Yam Y, Ruddy TD, et al. Potential clinical and economic consequences of noncardiac incidental findings on cardiac computed tomography. J Am Coll Cardiol 2009; 54:1533-1541.
22. Hausleiter J, Meyer T, Hermann F, et al. Estimated radiation dose associated with cardiac CT angiography. JAMA 2009; 301:500-507.
23. Einstein AJ, Henzlova MJ, Rajagopalan S. Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. JAMA 2007; 298:317-323.
24. Achenbach S, Marwan M, Ropers D, et al. Coronary computed tomography angiography with a consistent dose below 1 mSv using prospectively electrocardiogram-triggered high-pitch spiral acquisition. Eur Heart J 2010; 31:340-346.
25. Genders TS, Meijboom WB, Meijs MF, et al. CT coronary angiography in patients suspected of having coronary artery disease: decision making from various perspectives in the face of uncertainty. Radiology 2009; 253:734-744.
26. Ladapo JA, Jaffer FA, Hoffmann U, et al. Clinical outcomes and cost-effectiveness of coronary computed tomography angiography in the evaluation of patients with chest pain. J Am Coll Cardiol 2009; 54:2409-2422.
27. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011; 30:11-21.
28. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med 1979; 300:1350-1358.
29. Pryor DB, Shaw L, McCants CB, et al. Value of the history and physical in identifying patients at increased risk for coronary artery disease. Ann Intern Med 1993; 118:81-90.

Samenvatting

Deel I. Introductie

Het doel van het onderzoek beschreven in dit proefschrift was het bepalen van de optimale keuze van diagnostische testen voor patiënten met een verdenking op coronair lijden. De **diagnostische waarde** van verschillende beeldvormende technieken (CT angiografie, perfusie MRI, SPECT en echografie) werden met elkaar vergeleken. De onderzoekspopulaties bestonden uit patiënten met stabiele klachten van pijn op de borst, patiënten die zich op de spoedeisende hulp presenteren met acute klachten van pijn op de borst, en individuen zonder klachten maar met een verhoogd risico op cardiovasculaire ziekten. Daarnaast hebben wij besliskundige modellen ontwikkeld om de lange termijn **kosteneffectiviteit** van verschillende diagnostische strategieën (coronaire CT, stress perfusie MRI en het inspannings-ECG) te kunnen onderzoeken voor patiënten met stabiele pijn op de borst. Ook hebben wij **predictiemodellen** ontwikkeld die de voorafkans op coronair lijden kunnen schatten aan de hand van de klinische presentatie en risicofactoren.

Deel II. Systematisch Literatuuronderzoek en Meta-Analyse

Richtlijnen voor beeldvorming van asymptomatisch coronair lijden

Primaire preventieprogramma's hebben als doel individuen met een verhoogd risico op het ontwikkelen van cardiovasculaire ziektes op te sporen. Dit risico kan worden geschat aan de hand van leeftijd, geslacht, roken, cholesterol en de bloeddruk. Echter, coronair lijden kent een meetbare preklinische fase, namelijk atherosclerose. Beeldvormende technieken die atherosclerose in de coronair vaten afbeelden kunnen daarom van waarde zijn bij het schatten van het risico op cardiovasculaire ziekte en het opsporen van individuen met een verhoogd risico. Omdat er geen gerandomiseerde en gecontroleerde experimentele studies zijn uitgevoerd die deze waarde hebben onderzocht, kan het zo zijn dat richtlijnen hieromtrent verschillend zijn. We hebben systematisch gezocht in de bestaande literatuur naar gepubliceerde richtlijnen en vervolgens de aanbevelingen met betrekking tot (beeldvormende) testen voor asymptomatisch coronair lijden samengevat. De methoden van het tot stand komen van de richtlijnen werden beoordeeld aan de hand van het AGREE instrument. Veertien richtlijnen werden gevonden, met daarin 26 aanbevelingen met betrekking tot verschillende diagnostische testen (CT coronaire kalk score, coronaire CT angiografie, MR angiografie, SPECT, PET, stress echografie, rust echografie, en het inspannings-ECG) voor coronair lijden. De AGREE score varieerde van 21-93%. Acht van de 14 richtlijnen raden diagnostiek af of concluderen dat er te weinig wetenschappelijk bewijs voor is. De overige 6 richtlijnen raden diagnostiek alléén aan bij individuen met een á priori verhoogd risico. De meerderheid van deze richtlijnen beveelt de CT coronaire kalk score aan voor individuen met een "intermediate" risico op coronair lijden. Ons literatuuronderzoek heeft aangetoond dat de bestaande richtlijnen tegenstrijdige aanbevelingen doen met betrekking tot beeldvormende diagnostiek ten behoeve van het opsporen van individuen met een verhoogd risico op cardiovasculaire ziektes. Verder onderzoek, en in het bijzonder gerandomiseerde en gecontroleerde experimentele studies, is nodig om de kosten en effecten van (beeldvormende) diagnostiek in deze setting te kwantificeren.

De diagnostische waarde van myocard perfusie scans

Een perfusiescan van het myocard kan worden gebruikt als diagnostische test voor patiënten met stabiele pijn op de borst. Induceerbare myocardiale perfusie defecten zijn een aanwijzing voor het bestaan van obstructief coronair lijden. Er bestaan verschillende technieken voor het kwantificeren van perfusiedefecten, namelijk magnetische resonantie beeldvorming (MRI), contrast echografie, SPECT, en PET scans. Het is lastig om de diagnostische waarde van deze testen met elkaar te vergelijken op basis van de gepubliceerde meta-analyses, omdat deze allen verschillend zijn wat betreft methoden van zoeken, het selecteren van wetenschappelijk bewijs en het analyseren van de gegevens. Om deze reden hebben wij een systematisch literatuuronderzoek verricht naar verschillende diagnostische testen tegelijkertijd, steeds gebruikmakend van dezelfde methoden, selectiecriteria en statistische analyses. Hierdoor werd een directe vergelijking mogelijk. Daarnaast hebben wij, als extra aandachtspunt, studies met (mogelijke) verificatie bias, geëxcludeerd. Onze resultaten suggereren dat perfusie MRI superieur is ten opzichte van contrast echografie en SPECT voor het diagnosticeren van obstructief coronair lijden. De diagnostische waarde van contrast echografie en SPECT waren vergelijkbaar.

Beeldvorming bij patiënten met acute pijn op de borst

Acute pijn op de borst is een veelvoorkomend ziektebeeld op de spoedeisende hulp (SEH). SEH-artsen en cardiologen moeten beslissen of een patiënt met acute pijn op de borst moet worden opgenomen. Deze beslissing wordt meestal genomen op basis van de kliniek en een grove schatting van het risico, maar niet op sluitend bewijs of er wel of geen sprake is van een acuut coronair syndroom (ACS). Patiënten met een ACS presenteren zich vaak met atypische pijn op de borst klachten, zonder afwijkingen bij het lichamelijk onderzoek, en het ECG is vaak normaal of moeilijk te beoordelen. De aanwezigheid van een ACS kan niet worden uitgesloten op basis van de initiële klinische presentatie, maar vergt vaak klinische observatie en herhaaldelijk testen van markers in het bloed. In de ideale situatie zou een snelle en niet-invasieve test ingezet kunnen worden om een ACS vast te stellen of uit te sluiten. We hebben systematisch gezocht naar bestaande literatuur over de diagnostische waarde van beeldvormende technieken voor patiënten met acute pijn op de borst. De diagnostische waarde van rust ECHO, rust SPECT, en coronaire CT angiografie voor het detecteren van een ACS (diagnose groep 1) was vergelijkbaar. Daarnaast waren ook de diagnostische waarde van stress ECHO, stress SPECT en coronaire CT angiografie vergelijkbaar voor het aantonen van coronair lijden (diagnose groep 2). Een klein, statistisch niet-significant voordeel werd gevonden voor coronaire CT angiografie, die het meest sensitief was voor zowel diagnose groep 1 als 2. We concludeerden met een aanbeveling voor coronaire CT angiografie, waarbij opgemerkt moet worden dat, mede gezien de kleine verschillen in diagnostische waarde, de optimale test strategie in de praktijk ook zou moeten afhangen van de lokale expertise, aanwezigheid van apparatuur en patiënt karakteristieken.

Deel III. Predictiemodellen

De toegevoegde waarde van de CT coronaire kalk score

De voorspellende waarde van de CT coronaire kalk score voor het optreden van cardiovasculaire events is goed beschreven voor zowel asymptomatische als symptomatische individuen met een verdenking op coronair lijden. Op basis van gegevens van 254 patiënten met pijn op de borst die een invasieve hartcatheterisatie ondergingen, hebben wij 5 gepubliceerde predictiemodellen voor de aanwezigheid van obstructief coronair lijden gevalideerd. Vervolgens hebben wij telkens de toegevoegde waarde van de CT coronaire kalk score, als continue voorspeller bovenop de al in het predictiemodel aanwezige voorspellers, onder-

zocht. Leeftijd, geslacht en symptomen waren de meest belangrijke voorspellers voor het aanwezig zijn van obstructief coronair lijden. De CT coronaire kalk score was een sterke voorspeller onafhankelijk van de andere voorspellers. Roken, hypertensie, dyslipidemie en diabetes waren ook geassocieerd met de aanwezigheid van obstructief coronair lijden. Door het toevoegen van de CT coronaire kalk score, steeg de oppervlakte onder de receiver-operating-characteristic curve (c-statistic) met 0.05-0.09 voor de 5 verschillende modellen, die allemaal een andere combinatie van bestaande voorspellers hadden. Daarnaast resulteerde het toevoegen van de CT coronaire kalk score in reclassificatie van 38-47% van de patiënten, bestaande uit 64-84% correcte reclassificatie. We concluderen dat de CT coronaire kalk score toegevoegde waarde heeft bij het voorspellen van de aanwezigheid van obstructief coronair lijden en stellen dat de CT coronaire kalk score gebruikt kan worden in de diagnostische "work-up" van patiënten met pijn op de borst.

Validatie en updating van Diamond & Forrester

Hoewel verschillende risicofactoren voor coronair lijden bekend zijn, zoals diabetes, hypertensie, verhoogd cholesterol, roken en de CT coronaire kalk score, wordt het bepalen van de voorafkans op coronair lijden meestal gebaseerd op leeftijd, geslacht en symptomen naar het model van Diamond & Forrester uit 1979. De voorafkans op coronair lijden is cruciaal voor het bepalen van de optimale diagnostische test en voor het interpreteren van de testresultaten. In dit onderzoek combineerden wij patiëntengegevens afkomstig uit 14 verschillende ziekenhuizen om de validiteit van het Diamond & Forrester model te onderzoeken. Daarnaast hebben wij het model geüpdatet en uitgebreid met behulp van geavanceerde statistische methoden. Onze resultaten suggereren dat het Diamond & Forrester model de voorafkans op obstructief coronair lijden overschat, voornamelijk bij vrouwen. De voorspellende waarden van leeftijd, geslacht, symptomen en ziekenhuis werden geüpdatet en het model werd uitgebreid zodat ook het risico voor patiënten ouder dan 70 jaar kon worden geschat. Het geüpdatete model voorspelde minder hoge voorafkansen, wat erin resulteerde dat veel patiënten omlaag werden geclassificeerd. Het model is vervolgens gerecalibreerd voor laag-risico populaties. Een online risicocalculator werd ontwikkeld voor het voorspellen van de voorafkans op coronair lijden bij patiënten met pijn op de borst, voor zowel hoog- als laag risico populaties. Deze risicocalculator heeft potentiële waarde voor de klinische praktijk, aangezien een betere schatting van de voorafkans kan leiden tot betere beslissingen met betrekking tot het kiezen van de optimale diagnostische test en het interpreteren van testresultaten.

Een predictiemodel voor coronair lijden

De huidige richtlijnen bevelen het Diamond & Forrester model en de Duke Clinical Score aan, voor het schatten van de voorafkans op coronair lijden. In tegenstelling tot het Diamond & Forrester model, neemt de Duke Clinical Score wel cardiovasculaire risicofactoren mee in de berekening van de voorafkans. Echter, de Duke Clinical Score is nooit gevalideerd in externe populaties. In dit onderzoek combineerden wij patiëntengegevens afkomstig uit 18 verschillende ziekenhuizen om de validiteit van de Duke Clinical Score te onderzoeken. Sommige ziekenhuizen leverden gegevens over patiënten die verwezen waren voor een coronaire CT angiografie (laag-prevalente setting), en andere ziekenhuizen leverden gegevens over patiënten die verwezen waren voor een hartcatheterisatie (hoog-prevalente setting). Onze resultaten laten zien dat de Duke Clinical Score de voorafkans op coronair lijden overschat. Het model werd vervolgens geüpdatet, rekening houdend met clustering van patiënten binnen ziekenhuizen, en de toegevoegde waarde van de CT coronaire kalk

score werd berekend. De resultaten bevestigen dat leeftijd, geslacht en symptomen de belangrijkste voorspellers zijn voor het aanwezig zijn van obstructief coronair lijden. Verder zijn ook de aanwezigheid van diabetes, hypertensie, dyslipidemie en roken voorspellend. De CT coronaire kalk score was een sterke voorspeller en resulteerde in een significant betere c-statistic. De validiteit van het nieuwe model werd getest door middel van cross-validatie, waarbij de calibratie voldoende goed werd geacht. Dit suggereert dat het model valide is voor verschillende ziekenhuizen. Tot slot hebben wij een online risicocalculator ontwikkeld, zodat de voorafkans op coronair lijden stapsgewijs bepaald kan worden zodra klinische informatie en testresultaten beschikbaar komen.

Deel IV. Kosteneffectiviteits-analyses

Kosteneffectiviteit van coronaire CT angiografie als triage test

De hartkatheterisatie wordt beschouwd als de gouden standaard voor het vaststellen van coronair lijden. Omdat de kosten voor een hartkatheterisatie hoog zijn en het een klein risico op ernstige complicaties met zich meebrengt, wordt niet-invasieve diagnostiek aanbevolen om patiënten te selecteren die baat hebben bij het ondergaan van een hartkatheterisatie. Coronaire CT angiografie is een veelbelovende niet-invasieve diagnostische test voor het vaststellen van coronair lijden. In dit hoofdstuk hebben wij de kosteneffectiviteit van coronaire CT angiografie als triage test voor de hartkatheterisatie bestudeerd door middel van een beslisboom en Markov-model. Dit model werd ontwikkeld op basis van de best beschikbare gegevens uit de literatuur en ook op basis van een Nederlands patiëntencohort. De beslissing werd geanalyseerd vanuit verschillende perspectieven, en gebruikmakend van verschillende richtlijnen voor het doen van kosteneffectiviteitanalyses. Coronaire CT angiografie was kosteneffectief wanneer de voorafkans op coronair lijden lager was dan 44% in mannen en 37% in vrouwen. De optimale diagnostische "work-up" hangt af van welke uitkomsten belangrijk worden geacht, de voorafkans op coronair lijden, en de diagnostische waarde van de coronaire CT angiografie.

Kosteneffectiviteit van coronaire CT vergeleken met het inspannings-ECG

In de afgelopen jaren zijn verschillende systematische literatuuronderzoeken gepubliceerd die de diagnostische en prognostische waarden van coronaire CT angiografie hebben aangetoond. Omdat coronaire CT angiografie een zeer hoge negatief voorspellende waarde heeft (>95%), sluit een negatief resultaat de aanwezigheid van coronair lijden vrijwel zeker uit. Om deze reden kan coronaire CT angiografie mogelijk een rol spelen bij de diagnostische work-up van patiënten met pijn op de borst. Een Nederlands patiëntencohort met follow-up gegevens en de beste beschikbare gegevens uit de literatuur werden gebruikt om een beslisboom en Markov-model te construeren. Een hypothetische strategie op basis van de CT coronaire kalk score en coronaire CT angiografie werd vergeleken met standaardzorg op basis van een inspannings-ECG en eventueel een nucleaire stress test (SPECT). Het verschil met de analyse uit hoofdstuk 9 is vooral dat wij in deze analyse aannemen dat patiënten met minder ernstig coronair lijden niet allemaal verwezen worden voor een hartkatheterisatie. De COURAGE trial heeft namelijk laten zien dat met een percutane coronaire interventie de overleving en het aantal cardiovasculaire complicaties over het algemeen niet gereduceerd wordt voor patiënten met coronair lijden (≥70% stenose in ≥1 vat) en gedocumenteerde myocardischemie. Daarom wordt in de huidige richtlijnen een initieel conservatieve strategie aanbevolen voor patiënten met mild coronair lijden, op basis van optimale behan-

deling met medicamenten. Wanneer de symptomen niet verbeteren door de medicamenten alleen, kan in een later stadium nog besloten worden om te interveniëren door middel van een percutane coronaire interventie of bypass-chirurgie. De hypothetische CT strategie was goedkoper dan, maar even effectief als standaard zorg.

Kosteneffectiviteit van coronaire CT angiografie vergeleken met myocard perfusie MRI

Het systematische literatuuronderzoek en de meta-analyse met betrekking tot de diagnostische waarde van myocardiale perfusie scans, toonde aan dat perfusie MRI superieur was ten opzichte van contrast echografie en SPECT (Hoofdstuk 4). Echter, de vraag blijft bestaan of een diagnostische "work-up" op basis van perfusie MRI (een functionele test) of een diagnostische "work-up" op basis van coronaire CT angiografie (een anatomische test) optimaal is wat betreft kosteneffectiviteit. In dit hoofdstuk hebben wij een beslismodel ontwikkeld om deze strategieën met elkaar te vergelijken, zowel voor een invasieve als een niet-invasieve aanpak. Dat wil zeggen: bij de invasieve aanpak worden alle patiënten met een positieve testuitslag verwezen voor een hartkatheterisatie, en bij de niet-invasieve strategie worden alleen de patiënten met een ernstig positieve testuitslag verwezen voor een hartkatheterisatie en worden de patiënten met een mild positieve testuitslag direct behandeld met medicamenten. De beslissing werd geanalyseerd vanuit het perspectief van de Verenigde Staten van Amerika (VS), het Verenigd Koninkrijk (UK), en Nederland (NL), gebruikmakend van landspecifieke kostenschattingen en richtlijnen voor kosteneffectiviteitanalyses. Voor mannen en vrouwen met een voorafkans op coronair lijden van 30%, was coronaire CT angiografie gevolgd door perfusie MRI (invasieve aanpak), kosteneffectief voor de VS en NL. Voor de UK was coronaire CT angiografie gevolgd door perfusie MRI (conservatieve aanpak), kosteneffectief.

Deel V. Overigen

Prognostische waarde van coronaire CT vergeleken met het inspannings-ECG

In dit hoofdstuk bestudeerden wij gegevens van 471 poliklinische patiënten met stabiele klachten van pijn op de borst, zonder voorgeschiedenis van coronair lijden, en wie allen een inspannings-ECG en coronaire CT angiografie ondergingen. Na een gemiddelde follow-up duur van 2.6 jaar, vonden 44 cardiovasculaire events (cardiale dood, hartinfarct, of onstabiele angina pectoris met ziekenhuisopname) plaats in 30 verschillende patiënten. Patiënten met een niet-diagnostisch inspannings-ECG hadden een grotere kans op het krijgen van een event. Het coronaire CT angiografie resultaat had prognostische waarde onafhankelijk van de prognostische waarde van het inspannings-ECG. De CT coronaire kalk score had geen toegevoegde prognostische waarde bovenop de coronaire CT angiografie resultaten.

Methoden voor het analyseren van sensitiviteit en specificiteit op basis van geclusterde data

De diagnostische waarde van een beeldvormende test wordt meestal uitgedrukt in sensitiviteit en specificiteit vergeleken met de gouden standaard. Sensitiviteit en specificiteit worden nogal eens berekend op basis van meerdere metingen per patiënt, wat betekent dat de metingen geclusterd zijn. Of de sensitiviteit per patiënt moet worden berekend of gebruikmakend van meerdere metingen per patiënt, hangt af van de klinische context en gevolgen van een positieve/negatieve testuitslag. Meerdere metingen in één patiënt (cluster) zijn vaak positief gecorreleerd, wat kan leiden tot bias in de schattingen en variantie van sensitiviteit en specificiteit als het clustereffect niet wordt meegenomen in de analyse. De meest gangbare methoden om sensitiviteit en specificiteit uit te rekenen op basis van geclusterde data worden besproken en geïllustreerd: 1) Analyse op verschillende niveaus waarbij het clustereffect *niet* wordt meegenomen, 2) aanpassen van de variantie, 3) "logistic random-effects" modellen, en 4) "generalized estimating equations".

Conclusies en implicaties voor de klinische praktijk

Het doel van dit proefschrift was het bepalen en optimaliseren van de diagnostische "work-up" voor patiënten met mogelijk coronair lijden. Ons onderzoek heeft laten zien dat de diagnostische waarde van de perfusie MRI beter is dan de diagnostische waarde van SPECT. Als beide testen reële alternatieven zijn, verdient de perfusie MRI de voorkeur, temeer omdat er geen stralenbelasting bij komt kijken.

Ons onderzoek laat verder zien dat de CT coronaire kalk score voorspellende waarde heeft onafhankelijk van andere cardiovasculaire risicofactoren. De CT coronaire kalk score kan worden gebruikt als initiële triage test in patiënten met een lage voorafkans op coronair lijden. Hiermee kan verdere (onnodige) diagnostiek worden bespaard wanneer de kalkscore nul is, en kan verdere diagnostiek plaatsvinden als er wel coronaire kalk is.

De optimale diagnostische "work-up" wordt voornamelijk bepaald door de voorafkans op coronair lijden, die traditioneel wordt berekend op basis van het Diamond & Forrester model of de Duke Clinical Score. Ons onderzoek op basis van recente patiëntgegevens toont aan dat beide methoden de voorafkans op coronair lijden systematisch overschatten. Een nieuw predictiemodel werd ontwikkeld, dat als een online calculator gebruikt kan worden door artsen. Hoewel wij niet direct de klinische impact van het nieuwe predictiemodel hebben kunnen bestuderen, is het waarschijnlijk dat een betere schatting van de voorafkans tot betere beslissingen ten aanzien van diagnostiek zal leiden. Daarnaast kan onnodige diagnostiek worden voorkomen omdat de voorafkans op coronair lijden niet langer wordt overschat.

Verder hebben wij de lange termijn kosten en effecten van coronaire CT angiografie in kaart gebracht vanuit verschillende perspectieven en voor verschillende landen. In de Nederlandse situatie bleek dat coronaire CT angiografie als triage test voor een hartkatheterisatie kosteneffectief was voor mannen met een voorafkans lager dan 44% en voor vrouwen met een voorafkans lager dan 37%. Een "work-up" op basis van CT coronaire kalk score ± coronaire CT angiografie was goedkoper dan en even effectief als standaard zorg. Ook hebben wij gevonden dat een strategie op basis van coronaire CT angiografie gevolgd door perfusie MRI kosteneffectief is vergeleken met strategieën op basis van CT of MRI alleen.

Ten slotte, het geüpdatete predictiemodel voor het voorspellen van de voorafkans op coronair lijden en onze resultaten van de kosteneffectiviteitanalyses vormen een samenhangend geheel van wetenschappelijk bewijs voor het implementeren van beeldvormende diagnostisch testen, in het bijzonder voor de CT coronaire kalk score en coronaire CT angiografie.

CAD Consortium Collaborators

Koen Nieman
Department of Cardiology, Erasmus MC, Rotterdam, the Netherlands

Tjebbe W. Galema
Department of Cardiology, Erasmus MC, Rotterdam, the Netherlands

W. Bob Meijboom
Department of Radiology and Cardiology, Erasmus MC, Rotterdam, the Netherlands

Nico R. Mollet
Department of Radiology and Cardiology, Erasmus MC, Rotterdam, the Netherlands

Pim J. de Feyter
Department of Radiology and Cardiology, Erasmus MC, Rotterdam, the Netherlands

Gabriel P. Krestin
Department of Radiology, Erasmus MC, Rotterdam, the Netherlands

Hatem Alkadhi
Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Switzerland

Sebastian Leschka
Institute of Radiology, Kantonsspital St. Gallen, Switzerland

Lotus Desbiolles
Institute of Radiology, Kantonsspital St. Gallen, Switzerland

Matthijs F.L. Meijs
Department of Cardiology, University Medical Centre Utrecht, Utrecht, The Netherlands
Department of Radiology, University Medical Centre Utrecht, Utrecht, The Netherlands

Maarten J. Cramer
Department of Cardiology, University Medical Centre Utrecht, Utrecht, The Netherlands

Filippo Cademartiri
Department of Radiology, Erasmus University Medical Centre, Rotterdam, the Netherlands
Department of Radiology and Cardiology, Azienda Ospedaliero-Universitaria, Parma, Italy
Department of Radiology, Giovanni XXIII Clinic, Monastier, Treviso, Italy

Erica Maffei
Department of Radiology and Cardiology, Azienda Ospedaliero-Universitaria, Parma, Italy
Department of Radiology, Giovanni XXIII Clinic, Monastier, Treviso, Italy

Chiara Martini
Department of Radiology and Cardiology, Azienda Ospedaliero-Universitaria, Parma, Italy
Department of Radiology, Giovanni XXIII Clinic, Monastier, Treviso, Italy

Sara Seitun
Department of Radiology and Cardiology, Azienda Ospedaliero-Universitaria, Parma, Italy

Annachiara Aldrovandi
Department of Radiology and Cardiology, Azienda Ospedaliero-Universitaria, Parma, Italy

Marc Dewey
Department of Radiology, Charité, Medical School, Humboldt University, Berlin, Germany

Elke Zimmermann
Department of Radiology, Charité, Medical School, Humboldt University, Berlin, Germany

Michael Laule
Department of Cardiology, Charité, Medical School, Humboldt University, Berlin, Germany

Francesca Pugliese
Centre for Advanced Cardiovascular Imaging, Barts and The London NIHR Cardiovascular Biomedical Research Unit, Barts and The London School of Medicine and Dentistry & Barts and The London NHS Trust

Steffen E. Petersen
Centre for Advanced Cardiovascular Imaging, Barts and The London NIHR Cardiovascular Biomedical Research Unit, Barts and The London School of Medicine and Dentistry & Barts and The London NHS Trust

L. Ceri Davies
Centre for Advanced Cardiovascular Imaging, Barts and The London NIHR Cardiovascular Biomedical Research Unit, Barts and The London School of Medicine and Dentistry & Barts and The London NHS Trust
Rossella Barbagallo

Valentin Sinitsyn
Department of Radiology, Federal Centre of Medicine and Rehabilitation, Moscow, Russia

Deepa Gopalan
Department of Radiology. Papworth Hospital NHS Trust, Cambridge, United Kingdom

Jan Bogaert
Department of Cardiovascular Diseases, University Hospital Leuven, Belgium

Kaatje Goetschalckx
Department of Cardiovascular Diseases, University Hospital Leuven, Belgium

U. Joseph Schoepf
Department of Radiology, Medical University of South Carolina, Charleston, USA

Garrett W. Rowe
Department of Radiology, Medical University of South Carolina, Charleston, USA

Joanne D. Schuijf
Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands

Jeroen J. Bax
Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands

Fleur R. de Graaf
Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands

Juhani Knuuti
Turku PET Centre, Turku University Hospital, Turku, Finland

Sami Kajander
Turku PET Centre, Turku University Hospital, Turku, Finland

Carlos A.G. van Mieghem
Department of Cardiology, OLV Hospital Aalst, Belgium

Luc van Driessche
Department of Cardiology, St. Blasius Hospital Dendermonde, Belgium

Gudrun Feuchtner
Department of Radiology, Innsbruck Medical University, Austria

Tobias De Zordo
Department of Radiology, Innsbruck Medical University, Austria

Thomas Auer
Department of Radiology, Innsbruck Medical University, Austria

Fabian Plank
Department of Radiology, Innsbruck Medical University, Austria

Guy Friedrich
Department of Cardiology, Innsbruck Medical University, Austria

Simon Wildermuth
Institute of Radiology, Kantonsspital St. Gallen, Switzerland

Björn Stinn
Institute of Radiology, Kantonsspital St. Gallen, Switzerland

Jürgen Fornaro
Institute of Radiology, Kantonsspital St. Gallen, Switzerland

Konstantin Nikolaou
Department of Clinical Radiology, University Hospitals Munich, Grosshadern Campus, Germany

Fabian Bamberg
Department of Clinical Radiology, University Hospitals Munich, Grosshadern Campus, Germany

Ricardo C. Cury
Department of Radiology, Baptist Hospital of Miami and Baptist Cardiac and Vascular Institute, Miami, Florida, United States of America

Juan Battle
Department of Radiology, Baptist Hospital of Miami and Baptist Cardiac and Vascular Institute, Miami, Florida, United States of America

Pál Maurovich-Horvat
Heart Centre, Semmelweis University, Budapest, Hungary

Andrea Bartykowszki
Heart Centre, Semmelweis University, Budapest, Hungary

Bela Merkely
Heart Centre, Semmelweis University, Budapest, Hungary

Dávid Becker
Heart Centre, Semmelweis University, Budapest, Hungary

Martin Hadamitzky
Department of Cardiology, German Heart Centre, Munich, Germany

Jörg Hausleiter
Department of Cardiology, German Heart Centre, Munich, Germany

List of Publications

Genders TS, Meijboom WB, Meijs MF, Schuijf JD, Mollet NR, Weustink AC, Pugliese F, Bax JJ, Cramer MJ, Krestin GP, de Feyter PJ, Hunink MG. CT coronary angiography in patients suspected of having coronary artery disease: decision making from various perspectives in the face of uncertainty. *Radiology*. 2009 Dec;253(3):734-44.

Genders TS, Pugliese F, Mollet NR, Meijboom WB, Weustink AC, van Mieghem CA, de Feyter PJ, Hunink MG. Incremental value of the CT coronary calcium score for the prediction of coronary artery disease. *Eur Radiol*. 2010 Oct;20(10):2331-40.

Tholen AT, de Monye C, **Genders TS**, Buskens E, Dippel DW, van der Lugt A, Hunink MG. Suspected carotid artery stenosis: cost-effectiveness of CT angiography in work-up of patients with recent TIA or minor ischemic stroke. *Radiology*. 2010 Aug;256(2):585-97.

Ferket BS, **Genders TS**, Colkesen EB, Visser JJ, Spronk S, Steyerberg EW, Hunink MG. Systematic review of guidelines on imaging of asymptomatic coronary artery disease. *J Am Coll Cardiol*. 2011 Apr 12;57(15):1591-600.

Genders TS, Steyerberg EW, Alkadhi H, Leschka S, Desbiolles L, Nieman K, Galema TW, Meijboom WB, Mollet NR, de Feyter PJ, Cademartiri F, Maffei E, Dewey M, Zimmermann E, Laule M, Pugliese F, Barbagallo R, Sinitsyn V, Bogaert J, Goetschalckx K, Schoepf UJ, Rowe GW, Schuijf JD, Bax JJ, de Graaf FR, Knuuti J, Kajander S, van Mieghem CA, Meijs MF, Cramer MJ, Gopalan D, Feuchtner G, Friedrich G, Krestin GP, Hunink MG, The CAD Consortium. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating, and extension. *Eur Heart J*. 2011 Jun;32(11):1316-30.

Dedic A, **Genders TS**, Ferket BS, Galema TW, Mollet NR, Moelker A, Hunink MG, de Feyter PJ, Nieman K, Stable Angina Pectoris: Head-to-Head Comparison of Prognostic Value of Cardiac CT and Exercise Testing. *Radiology*. 2011 Nov;261(2):428-36.

de Jong M, **Genders TS**, van Geuns RJ, Moelker A, Hunink MG. Diagnostic performance of stress myocardial perfusion imaging for coronary artery disease: A systematic review and meta-analysis. *European Radiology* (In press)

Dedic A, **Genders TS**, Hunink MG, Nieman K. Imaging strategies for acute chest pain in the emergency department.*American Journal of Roentgenology* (In press)

Genders TS, Ferket BS, Dedic A, Galema TW, Mollet NR, de Feyter PJ, Fleischmann KE, Nieman K, Hunink MG. Coronary computed tomography versus exercise testing in patients with stable chest pain: comparative effectiveness and costs. *International Journal of Cardiology* (In press)

Genders TS, Steyerberg EW, Nieman K, Galema TW, Mollet NR, de Feyter PJ, Krestin GP, Alkadhi H, Leschka S, Desbiolles S, Meijs MFL, Cramer MJ, Knuuti J, Kajander S, Bogaert J, Goetschalckx K, Cademartiri F, Maffei E, Martini C, Seitun S, Aldrovandi A, Wildermuth S, Stinn B, Fornaro J, Feuchtner G, De Zordo T, Auer T, Plank F, Friedrich G, Pugliese F, Petersen SE, Davies LC, Schoepf UJ, Rowe GW, van Mieghem CAG, van Driessche L, Sinitsyn V, Gopalan D, Nikolaou K, Bamberg F, Cury RC, Battle J, Maurovich-Horvat P, Bartykowszki A, Merkely B, Becker D, Hadamitzky M, Hausleiter J, Dewey M, Zimmermann E, Laule M, Hunink MG, The CAD Consortium. A Prediction Model for the Presence of Coronary Artery Disease: a pooled analysis of existing cohorts. *British Medical Journal* (In press)

Genders TS, Spronk S, Stijnen T, Steyerberg WE, Lesaffre EM, Hunink MG. Methods for analyzing sensitivity and specificity of clustered data: a tutorial. Accepted for publication in *Radiology*

Neefjes LA, Rossi A, **Genders TS**, Nieman K, Papadopoulou SL, Dharampal AS, Schultz CJ, Weustink AC, Dijkshoorn ML, ten Kate GR, Dedic A, van Straten M, Cademartiri F, Hunink MG, Krestin GP, de Feyter PF, Mollet NR. Diagnostic Accuracy of 128-Slice Dual Source CT Coronary Angiography Using Different Low Dose Scan Protocols in Patients with Various Heart Rates: A Randomized Study. **Submitted**

Genders TS, Petersen SE, Pugliese F, Dastidar A, Vokó Z, de Brouwer S, Fleischmann KE, Nieman K, Hunink MG. Stress Myocardial Perfusion Cardiac Magnetic Resonance Imaging vs. Coronary CT Angiography in the Diagnostic Work-up of Patients with Stable Chest Pain: comparative effectiveness and costs. **Submitted**

Osnabrugge RL, Kappetein AP, Head SJ, **Genders TS**, Bogers AJ, van Mieghem NM, de Jaegere PT, Jüni P, Kalesan B, Hunink MG. Cost analysis of transcatheter versus surgical aortic valve replacement in intermediate risk patients with symptomatic aortic stenosis: a real-life experience. **Submitted.**

Correspondence

Genders TS, Dedic A, Nieman K, Hunink MG. Prognostic value of cardiac computed tomography angiography. *J Am Coll Cardiol*. 2011 Jun 21;57(25):2543-4.

Book chapters

Genders TS, Hunink MG. The Epidemiology of Coronary Artery Disease. In: *Clinical Applications of Cardiac CT* – Second edition. Springer-Verlag Italia 2012. Editors: Cademartiri F, Casolo M, Midiri, M.

PhD Portfolio

Name PhD student: Tessa Sarah Suzanne Genders
Erasmus MC Department: Epidemiology and Radiology
Research School: Netherlands Institute for Health Sciences (NIHES)
PhD period: November 2008 – February 2012
Promotor: Prof. dr. M.G. Myriam Hunink
Co-promotor: Dr. Koen Nieman

PhD training	Year	Workload (Hours/ECTS)
Research skills		
Master of Science in Clinical Epidemiology, NIHES, Rotterdam. Various courses in research methodology: Principles of Research in Medicine and Epidemiology, Introduction to Data-analysis, Regression Analysis, Topics in Evidence-based Medicine, Meta-analysis, Advanced Topics in Decision-making in Medicine and Diagnostic Research	2006 -2009	120 ECTS
Systematic Research and Meta-analysis, Summer Session II of the Harvard School of Public Health, Boston, USA	2008	20 hours
In-depth courses		
Workshop "Perspective and Uncertainty in Health Technology Assessment" by Karl Claxton, Nederlandse Vereniging voor Technology Assessment in de Gezondheidszorg	2010	10 hours
Book discussions		
Piantadosi, Clinical Trials: A Methodologic Perspective, 2005	2011	20 hours
Kirkwood and Sterne, Essential Medical Statistics, 2003	2010	20 hours
Steyerberg, Clinical Prediction Models, 2009	2010	20 hours

International conferences		
Radiological Society of North America (RSNA)	2008	1 ECTS
European Society of Cardiac Radiology (ESCR)	2009	1 ECTS
Society for Medical Decision Making (SMDM)	2009	1 ECTS
Society for Medical Decision Making (SMDM)	2010	1 ECTS
Society for Medical Decision Making (SMDM)	2011	1 ECTS
Presentations		
CT Coronary Angiography in Patients with Suspected Coronary Artery Disease: decision making from various perspectives in the face of uncertainty (RSNA)	2008	1 ECTS
Validation and updating of a simple clinical prediction rule for the diagnosis of coronary artery disease: The CAD consortium (ESCR)	2009	1 ECTS
Cost-effectiveness of CT Calcium Scoring and CT Coronary Angiography in the Diagnosis of Coronary Artery Disease: The influence of reduced radiation exposure (SMDM)	2009	1 ECTS
Modeling the joint distribution of sensitivity and specificity (SMDM)	2009	1 ECTS
Diagnostic strategies in patients with chest pain (Cleveland Clinic Heart & Vascular Institute)	2010	1 ECTS
Coronary computed tomography versus exercise testing in patients with stable chest pain: comparative effectiveness and costs (SMDM)	2011	1 ECTS
Risk-stratification for patients with stable angina (Invited lecture, University of Antwerp, Belgium)	2012	1 ECTS
Awards		
NIHES Award for best article written under the guidance of a NIHES tutor, Rotterdam	2009	-

Teaching activities	Year	Workload (Hours/ECTS)
Lecturing		
CQM Working Group Methods for the Evaluation of Diagnostic Tests: Diagnostic analysis when there are multiple observations per patient	2011	4 hours
CQM Working Group Methods for the Evaluation of Diagnostic Tests: Evaluation of diagnostic tests without a (good) reference standard	2011	4 hours
Teaching 5th year medical students about "Evidence Based Radiology" during their Radiology clerkship at the Erasmus University Medical Center, Rotterdam, The Netherlands	2009-2011	24 hours
Teaching a 4-day course in "Health Economic Modeling" at Department of Health Economics, University of Oslo, Norway	2010	40 hours
Co-teaching the Short Course "Presentation skills" at the annual meeting of the Society for Medical Decision Making	2010	6 hours

Supervising practical sessions

Teaching assistant at the Harvard School of Public Health for the summer course RDS-288: Methods for Decision Making in Medicine	2010	40 hours
Teaching assistant at Erasmus Winter Programme, Course: Advanced Topics in Decision Making in Medicine	2009-2012	80 hours
Teaching 4th year medical students during; "Sensitivity, specificity, predictive values, Bayesian revision, likelihood ratios and ROC curves", "Medical decision-making + no treat-test/ test-treat thresholds", and "How to read a scientific paper?"	2008-2011	32 hours
Teaching assistant at the Harvard School of Public Health for the summer course RDS-288: Methods for Decision Making in Medicine	2008	40 hours

Dankwoord

Dit boekje en de afgelopen jaren van mijn leven hadden er heel anders uitgezien zonder de hulp van heel veel mensen. Mijn dank is groot!

Professor Hunink, beste Myriam, na wat getwijfel aan het begin (decision... decisions...decisions...!), wist jij me toch over te halen om te gaan promoveren. En zie hier... het resultaat! Sinds dag één ben jij een groot voorbeeld voor mij en een bron van inspiratie. Daarnaast was het geweldig en gezellig om jouw teaching assistant te mogen zijn in Boston! Ik ben heel erg blij met alle kansen die je voor me hebt gecreëerd en alle kennis die ik dankzij jou heb kunnen opdoen. Niet alleen dankzij onze fantastische samenwerking, maar zeker ook door jouw betrokkenheid op het persoonlijke vlak, kijk ik met veel plezier terug op mijn promotie-tijd. Al met al, een succes-formule voor een proefschrift waar ik trots op kan zijn! Duizendmaal dank voor alles!

Dr. Nieman, beste Koen, als cardioloog en co-promotor ben jij onmisbaar geweest voor dit onderzoek en zorgde je ervoor dat ik me niet blind staarde op modellen en ingewikkelde analyses, en gaf me advies over de gang van zaken in de klinische praktijk. Onze besprekingen waren altijd nuttig, motiverend en daarnaast ook nog eens gezellig. Dankzij jou kon ik de Fast Track patiënten database analyseren en mee-werken aan het verzamelen van de follow-up gegevens. Heel erg bedankt voor je toewijding en de prettige samenwerking!

Professor Steyerberg, Ewout, jij was een belangrijke schakel in het tot stand komen van dit proefschrift. Dankzij jouw adviezen en zéker ook dankzij jouw meesterlijke boek “Clinical Prediction Models” (nog even reclame maken!) heb ik veel methodologische problemen kunnen oplossen. Je beantwoordde razendsnel mijn e-mails vol “moeilijke vragen”. En als de vragen ook voor jou moeilijk bleken (of... niet “moeilijk”, maar “te veel werk om per e-mail te beantwoorden”), mocht ik langskomen voor live advies tussen alle andere bedrijven door. Bedankt!

Professor Fleischmann, dear Kirsten, thank you so much for your (early morning) skype-calls, e-mails, and discussions on the clinical context of the cost-effectiveness analyses. And thank you even more for traveling all the way to The Netherlands and being part of the committee.

Professor Krestin, als hoofd van de groeiende en bloeiende afdeling Radiologie heeft u een druk bestaan. Maar u bent altijd aanwezig bij feesten en partijen en toont belangstelling voor iedereen. Bedankt dat u deel wilde uitmaken van mijn promotiecommissie!

Fawakka!! Bart Ferket, je was een super collega, helemaal mee eens! Met je hilarische imitaties van een persoon die ik niet eens ken maakte jij me altijd aan het lachen. Het was altijd fijn om met je te kunnen brainstormen over “rrrrisk modeling”. Een half uurtje gedachten, ideeën en theorieën sparren was erg nuttig om cirkelredeneringen te doorbreken en weer te kunnen begrijpen waar ik nou eigenlijk mee bezig was! Key!!!! Bedankt voor alles en natuurlijk dat je mijn paranimf wilde zijn. Sabbi tog.

Nathalie Grootenboer, lieve Nath... in het begin vond je me maar raar, maar na al die jaren zijn we niet alleen oud-collega’s maar ook vriendinnen. We hebben ontzettend veel leuke dingen beleefd, zoals koffie-breaks, werk-etentjes, shopdagen (sommige moeizaam, andere héél snel), citroen-momenten, stapavonden (de ene nóg fouter dan de andere), Erasmus

Summer Programmes, kooksessies, work-outs met Jimmy (die van Hollander!), bloemschikles en zo kan ik nog wel even doorgaan. Ik wil je heel erg bedanken dat je zowel aan mijn onderzoeksperiode als aan mijn kledingkast zoveel kleur hebt gegeven! Maar ik ben je natuurlijk nog het meest dankbaar voor die avond met eh... in Boudewijn!!! En dat je mijn paranimf was! Ik hoop dat we nog veel meer avonturen zullen beleven (waar dan ook ter wereld) en hopelijk mag ik je binnenkort lastig vallen als bijdehante co-assistent!

I would like to thank all consortium collaborators (see page 339) for sharing their data and contributing to this thesis. Without your contributions, it would not have been possible! Thank you so much!!

Francesca Pugliese and Steffen Petersen, thank you so much for everything! I enjoyed our collaboration and appreciate your ongoing support. Daarnaast wil ik Pim de Feyter, Bob Meijboom, Nico Mollet, Annick Weustink, Carlos van Mieghem, Tjebbe Galema en iedereen die ik vergeten ben, bedanken voor al het harde werk dat nodig was voor het verzamelen van patiëntengegevens, het beoordelen van CT-scans en angio’s, en jullie bijdragen aan de manuscripten.

Admir, heel erg bedankt voor de gezellige en goede samenwerking. We hebben samen aan leuke projecten gewerkt en succesvol gepubliceerd. Met mijn idealistische en perfectionistiche oog voor ieniemienie details was ik natuurlijk geen makkelijke om mee samen te werken, maar je bleef ondanks dat (bijna) altijd rustig :) Ook al werd er soms gekibbeld, we hadden voornamelijk veel (onderbroeken) lol!

Thomas Marwick and Rory Hachamovitch, thank you so much for inviting me to the Cleveland Clinic Heart & Vascular Institute. It was a wonderful experience (also a shocking one when I found out about the McDonalds restaurant within the hospital ;)) and an honor to present my research findings to you! Thanks so much.

Ton, bedankt dat je tussen het surfen en motorrijden door tijd had om de lay-out zo mooi te maken!

Professor Lesaffre, uw uitleg over en hulp bij het schrijven over statistische modellen was onmisbaar! Marek, enorm bedankt voor je hulp met de syntaxen! Nano, bedankt voor je eeuwige geduld met computer- en printer-perikelen.

Verder wil ik mijn (oud-) collega’s van de ART-groep bedanken: Sandra, Ineke, Edwin, Bas, Bob, Farzin, Erica, Cécile, Raluca, Suman, Rachel, Aletta, Marcus, Ruben, Minghui, Loes, Felisia en John. Jullie waren er altijd voor gezelligheid tussendoor, “quick” question rounds, vieze senseo-koffie, snoepjes/kauwgompjes, lunch, pennen & enveloppen, taart, verlopen thee, originele verjaardagscadeautjes, AH bonus-aanbiedingen, of ongezoute kritieken op mijn onderzoek en presentaties. Bedankt!!

Michiel, Lea & Bas, Maarten & Karin, Esther, Joop, en volleybalteam G1, bedankt voor jullie vriendschap en gezelligheid. Esther, ik kijk met plezier terug op onze samenwoon/samenfiets periode, en je hulp met dat PB-tje zal ik niet vergeten ;) Joop, dankzij jouw “To do list” boekje kon ik continu lijstjes maken en taken afvinken. En zó lief dat je die heerlijke cupcakes hebt gemaakt!

Mam en Pap, bedankt dat jullie altijd voor me klaar staan en dat jullie mij intelligentie en doorzettingsvermogen hebben gegeven. Van wie ik nu precies wat heb geërfd, mogen jullie zelf uitmaken ;) Anna, Corrie, Robert, Carlos, Lonneke, Job en de rest van de familie die ik niet heb genoemd: bedankt dat jullie er voor mij zijn. Summer, enorm bedankt dat jij de creatieve kant(en) van dit boekje wilde verzorgen, het is gewoonweg WOW!

Liefste Aaron, alweer meer dan 4 jaar geleden werden wij aan elkaar voorgesteld en sindsdien hebben we zóveel coole dingen meegemaakt! De komende jaren worden wederom bijzonder en ik ben er van overtuigd dat we ons prima zullen redden als “arme” studenten. Bedankt dat je er altijd voor me bent... Baby, I can’t wait for the rest of our life.

About the Author

Tessa Sarah Suzanne Genders was born on 18 February 1986 in Maastricht, The Netherlands. In 2004, she graduated from the Emmauscollege, Rotterdam and started medical school at the Erasmus University in Rotterdam.

In 2006, Tessa was among the top 10% of medical students and was selected to participate in a special program for medical students organized by the Netherlands' Institute of Health Sciences (NIHES). This program enabled her to enroll in a Master of Science in Clinical Epidemiology alongside her medical degree, which consisted of summer- and winter schools during summer recess and winter elective periods. For her research assignment, Tessa worked at the department of Epidemiology and Radiology under the supervision of M.G. Myriam Hunink, which resulted in her first publication. In 2009, she completed her Master of Science in Clinical Epidemiology, and her MSc-thesis was awarded the NIHES prize for best paper written under the guidance of a NIHES tutor during the academic year of 2008-2009.

From November 2008 until February 2012, Tessa worked on her PhD-thesis at the department of Epidemiology and Radiology at the Erasmus University Medical Center, Rotterdam, under the supervision of M.G. Myriam Hunink. Her work focused on the evaluation of diagnostic test strategies for patients suspected of having coronary artery disease. She collaborated with the department of Cardiology and with various Radiology research groups across Europe and the United States (CAD consortium collaborators). Next to her research, Tessa was involved in teaching medical students and supervising practical sessions at the NIHES summer- and winter classes. In 2008 and 2010, Tessa was employed as a teaching assistant to M.G. Myriam Hunink, teaching summer courses in decision-making at the Harvard School of Public Health, Harvard University, Boston, USA.

Currently, Tessa is doing her clinical rotations (co-schappen) at various hospitals across the Rotterdam area, and expects to graduate from medical school in 2014.

ISBN: 978-90-9026566-7

Stellingen behorende bij het proefschrift:

1. Diagnostic Imaging Strategies for Patients with Suspected Coronary Artery Disease

2. CT coronaire angiografie als triage test voor conventionele angiografie is kosten-effectief voor patiënten met pijn op de borst en een voorafkans op obstructief coronairlijden lager dan 40% (dit proefschrift).

3. De CT coronaire kalkscore is een sterke en onafhankelijke voorspeller voor de aanwezigheid van obstructief coronairlijden (dit proefschrift).

4. De thans gangbare methodes voor risicostratificatie overschatten de voorafkans op obstructief coronairlijden (dit proefschrift).

5. Diagnostiek op basis van een CT coronaire kalkscore, eventueel gevolgd door CT coronair angiografie, voor patienten met pijn op de borst is goedkoper dan, en minstens zo effectief als standaard zorg (dit proefschrift).

6. Leeftijd, geslacht, type pijn op de borst, en de CT coronaire kalkscore zijn sterke voorspellers voor de aanwezigheid van obstructief coronairlijden, en kunnen worden gebruikt om de voorafkans te bepalen (dit proefschrift).

7. Viele kleine Leute die in vielen kleinen Orten viele kleine Dinge tun, können das Gesicht der Welt verändern. (Berlijnse muur, Afrikaanse wijsheid)

8. Balance in everything you do will bring you health and happiness.

9. "Peer review is a flawed process, full of easily identified defects with little evidence that it works." (Richard Smith, J R Soc Med 2006;99:178–182)

10. "Essentially, all models are wrong, but some are useful." (George E.P. Box, Empirical Model-Building and Response Surfaces, 1987)

11. Integrity is doing the right thing, even when you think nobody is watching. (Unknown)

12. "Do one thing every day that scares you." (Eleanor Roosevelt)

Tessa Genders
Rotterdam, Juni 2012