Assessment of Diagnostic Imaging Technologies for Cardiovascular Disease

Majanka H. Heijenbrok-Kal

Assessment of Diagnostic Imaging Technologies for Cardiovascular Disease

Acknowledgements

The work presented in this thesis was supported by the Netherlands Organization for Scientific Research (NWO) (Program Grant no. 904-66-091).

The printing of this thesis was supported by the department of Epidemiology & Biostatistics and the department of Radiology of the Erasmus University Medical Center Rotterdam.

Cover design: Optima Grafische Communicatie
Lay-out: MH Heijenbrok-Kal
Printed by Optima Grafische Communicatie, Rotterdam

ISBN 90-8559-006-X

Assessment of Diagnostic Imaging Technologies for Cardiovascular Disease

Evaluatie van diagnostische beeldvormende technieken voor hart- en vaatziekten

Proefschrift

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

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 24 november 2004 om 15.45 uur

door

Majanka Henriëtte Kal, ev Heijenbrok

geboren te Vlaardingen

Promotiecommissie

Promotor:	Prof.dr. M.G.M. Hunink
Overige leden:	Prof.dr. Th. Stijnen Prof.dr. P.J. de Feyter Prof.dr. F.F.H. Rutten
Co-promotor:	K.M. Kuntz, ScD

Contents

Manuscripts based on studies described in this thesis

Chapter 2
Heijenbrok-Kal MH, Fleischmann KE, Hunink MGM. Non-invasive imaging tests for coronary artery disease: where do we stand? *Submitted.*

Chapter 3
Heijenbrok-Kal MH, Fleischmann KE, Hunink MGM. Non-invasive imaging tests for peripheral, renal, and carotid artery disease: where do we stand? *Submitted.*

Chapter 4
Heijenbrok-Kal MH, Kuntz KM, Hunink MGM, Fleischmann KE. Value of pharmacological stress testing in patients with chest pain: a cost-effectiveness analysis. *JAMA: revision requested.*

Chapter 5
Heijenbrok-Kal MH, Kuntz KM, Fleischmann KE, Hunink MGM. Cost-effectiveness of electron beam computed tomography in patients with chest pain. *Submitted.*

Chapter 6
Heijenbrok-Kal MH, Kuntz KM, Fleischmann KE, Hunink MGM. Estimation of optimal test characteristics based on cost-effectiveness analysis: flexible versus fixed test characteristics. *In preparation.*

Chapter 7
Heijenbrok-Kal MH, Buskens E, Nederkoorn PJ, Van der Graaf Y, Hunink MGM. The optimal peak systolic velocity threshold on duplex ultrasonography for the indication carotid endarterectomy: a decision analytic approach. *Accepted for publication in Radiology.*

Chapter 8
Heijenbrok-Kal MH, Nederkoorn PJ, Buskens E, Van der Graaf Y, Hunink MGM. Diagnostic performance of duplex ultrasonography in patients suspected of carotid artery disease: symptomatic versus asymptomatic artery. *Submitted.*

Chapter 9
Heijenbrok-Kal MH, Hunink MGM. Adjusting for bias in diagnostic reports. *Am J Med. 2002;112:322-324.*

Chapter 10
Groot Koerkamp B, Heijenbrok-Kal MH, Kuntz KM, Fleischmann KE, Hunink MGM. Guiding future clinical research of competing diagnostic tests for coronary artery disease. *In preparation.*

1

Introduction

Introduction

Cardiovascular disease is the leading cause of death in industrialized countries. In the Netherlands approximately 50,000 patients die each year from cardiovascular disease, which is 35% of all deaths [1]. In the United States, this number is over 930,000 patients per year [2]. Cardiovascular disease includes all diseases of the heart or blood vessels. This thesis concentrates on arterial stenosis in the cardiovascular system, specifically in the coronary arteries, the carotid arteries, the peripheral arteries, and the renal arteries. Arterial stenosis in these areas may lead to severe morbidity due to, for example, myocardial infarction, stroke, limb amputation, or renal dysfunction, and thus has an enormous impact on the quality of life of the patient and may lead to death[3-6]. Therefore, it is important to diagnose cardiovascular disease accurately, so that adequate treatment can be initiated.

Imaging technologies for cardiovascular disease studied in this thesis

The first abdominal aortography was performed in 1929 and catheterization and angiography have been developed since that time [7]. In the 1960s conventional angiography became clinically useful to diagnose the presence and extent of arterial stenosis in the arterial system. However, conventional angiography carries a risk of morbidity and mortality due to the use of contrast agents and radiation exposure[8]. In addition, the test is costly and burdensome to the patient, because it requires several hours of hospitalization. Therefore, the focus is currently on the development and improvement of non-invasive imaging technologies with the ultimate goal to replace conventional angiography.

An enormous variety of non-invasive imaging tests exists to determine the presence or extent of cardiovascular disease. Tests that are routinely used in coronary artery disease include stress echocardiography and stress single-photon-emission computed tomography[9]. These tests are relatively inexpensive. In contrast to angiography, however, these tests use indirect measures of arterial stenosis, such as wall motion scores or perfusion scores, which are functional measures of myocardial ischemia. During the last decade, the role of electron-beam computed tomography (EBCT) has been studied for the diagnosis and prognosis of coronary artery disease[10]. Although calcium scoring on EBCT is generally used to determine prognosis in a patient with coronary artery disease, it can also be used to estimate the degree of arterial stenosis. EBCT has a very short scanning time, but requires special equipment. A new technological development involves computed tomographic angiography that has been adapted for diagnosing coronary artery disease[11]. This technology is becoming more widely available and can provide three-dimensional images of the heart. For each of theses non-invasive imaging technologies diagnosing coronary artery disease is a technological challenge due to the continuous beating of the heart, respiratory motion, and the small vessel size.

For the diagnosis of peripheral, carotid, and renal artery disease duplex ultrasonography has since long been used[12-14]. Duplex ultrasonography is inexpensive but depends on the expertise of the operator. New technological developments for these anatomical areas also include computed tomographic angiography and (contrast enhanced) magnetic resonance angiography[15]. These tests are able to provide a roadmap of the vascular tree, like conventional angiography. The

trade-off between spatial and temporal resolution in these techniques, however, may lead to a reduced imaging quality.

During the last decade numerous studies have been published on the diagnostic performance of non-invasive imaging tests for cardiovascular disease. The question remains which studies provide the best-available evidence for the clinician and which test should be chosen in a particular clinical setting.

Methods of diagnostic test evaluation used in this thesis

Several methods have been developed to compare the performance of imaging tests for the diagnosis of cardiovascular disease. Traditionally, the determination of sensitivity and specificity of a new technology have been used to demonstrate good diagnostic performance. However, studies have shown that these measures may be affected by the choice of patient population and the clinical setting in which the test is performed [16]. Furthermore, if sensitivity of test A is higher than test B, but specificity of test A is lower than test B, it is still not clear which test has the best diagnostic performance. Therefore, positive and negative predictive values, likelihood ratios, and diagnostic odds ratios have been evaluated in addition to measures of sensitivity and specificity.

To summarize the currently available evidence on the diagnostic performance of non-invasive imaging tests published in the literature a systematic review can be performed. In a systematic review the available evidence on the subject of interest is systematically collected from the literature using in- and exclusion criteria[17]. Outcome measures of interest, for example measures of sensitivity and specificity, are extracted from the included studies and a summary measure of diagnostic performance is calculated. Traditionally, pooling the measures of sensitivity and specificity independently, weighted for sample size, has been performed.

During the last decade, summary receiver operating characteristic (SROC) analysis has been used to summarize measures of sensitivity and specificity taking into account their inverse relationship[18]. SROC analysis is a regression technique that accounts for changes in the cut-off value of the test and offers the opportunity to adjust for study characteristics. The area under the SROC curve can be used as an overall measure of diagnostic performance.

A large number of diagnostic imaging tests have been published showing a good diagnostic performance for the diagnosis of cardiovascular disease. From a health care policy perspective, however, the value of a diagnostic test should not only reflect the diagnostic performance, but also the cost and risks associated with the test and the treatments that may be induced by the test, the pre-test probability of disease, and the health related quality of life of the patient. Cost-effectiveness analysis can be performed to integrate these variables and to guide the decisions on which test should be used in specific clinical settings[19].

Using the outcomes of a cost-effectiveness analysis in combination with a SROC analysis, the optimal sensitivity and specificity for a specific clinical setting can be calculated[20]. For each potential combination of sensitivity and specificity on the SROC curve the net health benefit, a measure that integrates cost and effectiveness, is calculated for a specific patient population[21]. The operating point on the SROC curve with the maximum net health benefit is defined as the optimal operating point. Furthermore, using the ROC curve from a single study, the optimal cut-off value of a test can be calculated for a specific clinical setting.

Depending on the referral pattern and test protocols used in clinical settings, diagnostic test evaluations may generate biased results[22]. It has been recognized that in particular verification bias may have a large influence on measures of diagnostic performance. Verification bias exists if patients are referred to the reference test based on the results from the test under evaluation. As a result of verification bias sensitivity may be inflated and specificity deflated if positive test results are verified preferentially. If data are collected on all patients undergoing the test that is evaluated, it is possible to correct for verification bias under the assumption that the probability of verification and the disease status are conditionally independent[23].

Despite substantial clinical research, uncertainty remains about which test should be chosen in a specific clinical setting. Future quantitative clinical research can help resolve this uncertainty. To estimate which variables in diagnostic test evaluations have the largest impact on the choice of the optimal test strategy, value of information analysis can be performed [24]. Using value of information analysis the expected benefit of various study parameters can be estimated for future test evaluations.

Aim and outline of this thesis

The aim of this thesis was to assess diagnostic imaging technologies for the diagnosis of cardiovascular disease using various methods of diagnostic test evaluations and to determine the optimal diagnostic test strategy in specific clinical settings. A secondary aim was to develop the methods available for the assessment of diagnostic imaging technology, including meta-analytical and decision analytical techniques.

In this thesis several methods of diagnostic test evaluation are described and applied to the field of diagnostic imaging technologies in cardiovascular disease: from systematic review and comparison of the sensitivity, specificity, and diagnostic odds ratios in chapters 2 and 3, to cost-effectiveness analysis in chapters 4 and 5. In addition, methods of test optimization are described for the determination of the optimal sensitivity and specificity in chapter 6 and optimizing cut-off values of diagnostic tests in chapters 7 and 8. Chapter 9 focuses on potential biases that may affect the results of diagnostic test evaluations, in particular verification bias. In chapter 10 value of information analysis is used to determine which study parameters should be evaluated in future research on imaging technologies for coronary artery disease.

Each method discussed in this thesis has its specific benefits and limitations. In the general discussion these will be put into perspective in order to select a methodology of diagnostic test evaluation that may help solve a specific clinical problem.

References

1. Koek HL, Van Leest LATM, Verschuren WMM, Bots ML. Hart- en vaatziekten in Nederland 2003, cijfers over leefstijl- en risicofactoren, ziekte en sterfte. Den Haag: Nederlandse Hartstichting; 2003.
2. http://www.cdc.gov/nccdphp/aag/pdf/aag_cvh2004.pdf.
3. Halliday A, Mansfield A, Marro J, et al. Prevention of disabling and fatal strokes by successful carotid endarterectomy in patients without recent neurological symptoms: randomised controlled trial. Lancet. 2004;363:1491-1502.
4. Hallin A, Bergqvist D, Fugl-Meyer K, Holmberg L. Areas of concern, quality of life and life satisfaction in patients with peripheral vascular disease. Eur J Vasc Endovasc Surg. 2002;24:255-263.
5. Brorsson B, Bernstein SJ, Brook RH, Werko L. Quality of life of chronic stable angina patients 4 years after coronary angioplasty or coronary artery bypass surgery. J Intern Med. 2001;249:47-57.
6. Fergany AF. Ischemic nephropathy. Epidemiology and clinical implications. Urol Clin North Am. 2001;28:805-813.
7. Pepine CJ, Hill JA, Lambert CR. History of the development and application of cardiac catheterization. In: Diagnostic and therapeutic cardiac catheterization. Baltimore: Williams and Wilkins; 1989.
8. Pohler E, Gunther H, Diekmann M, Eggeling T. Outpatient coronary angiography--safety and feasibility. Cardiology. 1994;84:305-309.
9. 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. Circulation. 2002;106:1883-1892.
10. O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/American Heart Association Expert Consensus Document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. J Am Coll Cardiol. 2000;36:326-340.
11. Schoenhagen P, Halliburton SS, Stillman AE, et al. Noninvasive Imaging of Coronary Arteries: Current and Future Role of Multi-Detector Row CT. Radiology. 2004;232:7-17.
12. Vasbinder GB, Nelemans PJ, Kessels AG, Kroon AA, de Leeuw PW, van Engelshoven JM. Diagnostic tests for renal artery stenosis in patients suspected of having renovascular hypertension: a meta-analysis. Ann Intern Med. 2001;135:401-411.
13. Visser K, Hunink MG. Peripheral arterial disease: gadolinium-enhanced MR angiography versus color-guided duplex US--a meta-analysis. Radiology. 2000;216:67-77.
14. Nederkoorn PJ, van der Graaf Y, Hunink MG. Duplex ultrasound and magnetic resonance angiography compared with digital subtraction angiography in carotid artery stenosis: a systematic review. Stroke. 2003;34:1324-1332.
15. Donnelly R, Hinwood D, London NJ. ABC of arterial and venous disease. Non-invasive methods of arterial and venous assessment. BMJ. 2000;320:698-701.
16. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. Epidemiology. 1997;8:12-17.
17. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Radiology. 2003;226:24-28.
18. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med. 1993;12:1293-1316.
19. Graham JD, Corso PS, Morris JM, Segui-Gomez M, Weinstein MC. Evaluating the cost-effectiveness of clinical and public health measures. Annu Rev Public Health. 1998; 19:125-152.
20. Hunink MGM, Glasziou PP. Multiple test results. In: Decision making in health and medicine. Integrating evidence and values. 1st ed. Cambridge: Cambridge University Press; 2001.
21. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998;18:S68-80.

22. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999;282:1061-1066.
23. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983;39:207-215.
24. 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.

2

Non-invasive imaging tests for coronary artery disease: where do we stand?

Abstract

Objectives: The purpose of this study was to review the literature and to compare the diagnostic performance of stress echocardiography, stress single-photon-emission computed tomography (SPECT), and electron-beam computed tomography (EBCT) for the assessment of coronary artery disease using meta-analytic techniques.
Background: Current guidelines suggest that a non-invasive imaging test should be performed to determine whether patients with suspected coronary artery disease should undergo coronary angiography.
Methods: We searched PubMed from January 1990 through May 2003 for meta-analytic studies on the diagnostic performance of imaging tests for coronary artery disease. Meta-analyses were selected that used angiography as reference standard and that presented the absolute numbers of true-positive, false-positive, true-negative, and false-negative results of the source studies. Data were extracted on the level of the meta-analysis and the level of the source studies included in each meta-analysis. Duplicate source studies were excluded. A summary-receiver-operating-characteristic analysis was performed on the source data.
Results: Seven meta-analyses were selected, including 246 patient series with 24,761 patients. The coronary tests showed little differences in diagnostic performance. The diagnostic odds ratios of exercise, dobutamine, and dipyridamole echocardiography were higher than the corresponding stress SPECT tests and EBCT, but not significantly. The diagnostic performance was significantly better in studies with a high proportion of men and significantly declined over the years.
Conclusions: We conclude that the differences in diagnostic performance between imaging tests for coronary artery disease are small, that the diagnostic performance of imaging tests is significantly better in men than in women, and that the diagnostic performance has declined over the years.

Introduction

Coronary artery disease is the main cause of death in industrialized countries. Coronary events can be minimized if patients suspected of coronary artery disease are correctly and safely diagnosed and treatment is initiated accordingly. The accepted reference standard for the diagnosis of coronary artery disease is coronary angiography. However, coronary angiography is costly and severe complications occur in almost 0.5% of patients undergoing coronary angiography.[1] Therefore, non-invasive tests with optimal sensitivity and specificity should be performed to determine which patients should undergo coronary angiography. Current guidelines recommend the use of non-invasive imaging tests, such as stress echocardiography and stress SPECT, prior to coronary angiography in patients suspected of coronary artery disease.[2] The role of EBCT for the diagnosis of coronary artery disease is controversial. Its use is both discouraged [3] and recommended [4] for patients suspected of coronary artery disease.

A number of studies have been published in which the sensitivity and specificity of non-invasive imaging tests have been evaluated. In the last decade, several meta-analyses have been performed summarizing the published data for each test. However, direct comparison of test accuracy measurements from various studies is complicated, because several factors that are associated with patient selection and study design may have differed between the studies. Furthermore, it is uncertain whether the sensitivity or specificity should take priority in a particular clinical setting. Therefore, it is useful to combine the sensitivity and specificity into one estimate of test performance, such as the diagnostic odds ratio. The potential bias of clinical factors and study characteristics on the diagnostic odds ratio can be accounted for in a linear regression analysis using the methods of summary receiver operating characteristics analysis (SROC).[5-7] We questioned whether significant differences existed in the diagnostic performance of stress echocardiography, stress SPECT, and EBCT, when adjusting for study characteristics.

The purpose of this study was to review the literature on non-invasive imaging tests for the diagnosis of coronary artery disease, to compare the diagnostic performance of the tests, and to identify study characteristics that may influence the diagnostic performance.

Methods

Study selection

A literature search was performed in PubMed from January 1990 through May 2003 to identify meta-analyses on the diagnostic performance of non-invasive imaging tests for the diagnosis of coronary artery disease. We used the following search terms: *meta-analysis* or *systematic review* or *structured review* and *imaging test*. We also checked the references of the included articles for additional meta-analyses. Articles in the English language were included if they met the following inclusion criteria: (a) the diagnostic performance of imaging tests for coronary artery disease was studied, (b) meta-analytic methods were used, (c) coronary angiography was used as reference standard, (d) the absolute numbers of true-positive, false-negative, true-negative and false-positive results of the source studies were available or derivable from the meta-analyses, and (e) it was published after 1990. If inclusion criterion *d* was not met, the authors were requested to provide these data. Because of the rapid advances in

imaging technology, meta-analyses published before 1990 were excluded. If a meta-analysis compared an imaging technology with a technology that is no longer in use or with a non-imaging technology, only the data on the newer imaging technology were included.

Data extraction

The data were extracted on two levels: on the level of the meta-analysis and on the level of the source studies that were included in each meta-analysis. The following data were extracted per meta-analysis: author, journal, year of publication, type of test, number of studies included, total number of patients, summary estimates of sensitivity and specificity, and details of the meta-analytic methods. In addition to the study identification and study size, the mean age of the patients, the proportion of men included, the type of stress used, and the absolute number of true-positive, false-negative, false-positive, and true-negative test results were extracted per source study. If source studies of meta-analyses overlapped and the data presented in the meta-analyses were identical, the duplicate data were excluded. If discrepancies existed between duplicate source studies, the original publication of the source study was retrieved and, based on the original paper, the discrepant source data were excluded. If a meta-analysis was done on the same source studies for several definitions of significant disease, the data were extracted only for one definition: this definition was 50-100% stenosis in at least one coronary artery. However, in some meta-analyses, the definition of significant disease was not used as selection criterion, but as a covariate in the analysis. The non-invasive imaging tests included exercise echocardiography, pharmacological stress echocardiography, exercise SPECT, pharmacological stress SPECT, and EBCT.

Data-synthesis

Pooled weighted analysis

First, the data were evaluated on the level of the meta-analysis. The published summary estimates of sensitivity and specificity were plotted per test type with 1-specificity on the horizontal axis and sensitivity on the vertical axis. If the meta-analysis did not present summary estimates of sensitivity and specificity, the pooled values of sensitivity and specificity were calculated from the absolute numbers of true and false test results of the source studies. Because most of the meta-analyses presented fixed summary estimates, we calculated the pooled values of sensitivity and specificity weighted for sample size only. After exclusion of the duplicate studies, we re-calculated the pooled values of sensitivity and specificity per test type using a random effects meta-analysis to account for heterogeneity between studies.

SROC analysis

We performed a random effects SROC analysis on the data set of source studies to calculate the relationship between sensitivity and specificity, taking into account the differences in positivity criterion and other factors of heterogeneity between settings. In an SROC analysis the true-positive rate (TPR) (sensitivity), and the false-positive rate (FPR) (1-specificity), are transformed into the logit form.[6,8] Then, a linear regression analysis is performed on the

variables D and S, D representing the natural logarithm of the diagnostic odds ratio and S functioning as proxy for the positivity criterion of the diagnostic test. D and S are defined as the difference and the sum of the logit transformations of TPR and FPR respectively, in formula:

$$D = \ln\frac{TPR}{1-TPR} - \ln\frac{FPR}{1-FPR}$$

$$S = \ln\frac{TPR}{1-TPR} + \ln\frac{FPR}{1-FPR}$$

To prevent undefined values for D and S due to 0's in the equations, 0.5 was added to the absolute numbers of true-positive, false-positive, true-negative and false-negative test results for the calculation of D and S. The regression analysis was weighted with the inverse of the variance of D. We used a random effects model, which takes into account the variability between studies.[9] All data were analyzed in a linear regression model in which D, the log of the diagnostic odds ratio, was modeled as a function of the explanatory variable that indicated the type of test, i.e. exercise, dobutamine, adenosine, and dipyridamole echocardiography; exercise, dobutamine, adenosine, and dipyridamole SPECT; and EBCT. The diagnostic performance of each test is represented by the relative odds ratio: a relative odds ratio equal to 1 indicates similar performance, a relative diagnostic odds ratio larger than 1 indicates better performance, and a relative diagnostic odds ratio smaller than 1 indicates inferior performance as compared with the test that was chosen as reference. Additional covariates were added to the model to evaluate the influence of clinical factors and study characteristics on diagnostic performance. The following covariates were evaluated: total number of patients in the study, mean age, proportion of men included, proportion of diseased patients, and the year of publication of the study. The year of publication of the study was chosen as a proxy for potential secular changes. First, we evaluated the individual effect of each variable on the diagnostic performance in the model that always included S. Starting with the variables with the lowest residual between-study variance (determined by tau-squared), the significant variables were added to the multivariable model in a stepwise forward manner. A significance level of 0.05 was used. Tau-squared, the residual between-study variance calculated with the restricted maximum likelihood method, was used as a measure of the model fit. A lower tau-squared indicates less residual between-study variance and therefore a better model fit and a better explanatory power by the model of the heterogeneity across studies. We used STATA 8.0 for the random effects analyses.

Results

Study selection

The search strategy resulted in 120 abstracts of which 12 meta-analyses were found that reported on coronary artery disease.[3,10-20] Of the 12 meta-analyses, one study was excluded because it was published before 1990.[11] One meta-analysis did not report the absolute numbers of true-positive, false-positive, true-negative, and false-negative test results of the source studies.[14] The author of this meta-analysis was requested to provide the missing information and

Table 1: Overview of the included meta-analyses

Meta-analysis	Publ year	Type of Test	No. series	No. patients	Pooled Sens	Pooled Spec	Significant predictors SROC analysis	Discussion / Conclusion
Fleischmann	1998	Exercise Echo	24	2637	85	77	Test, age, setting,	Discriminatory capabilities
		Exercise SPECT	27	3237	87	64	publication year	Exercise Echo > Exercise SPECT
O'Rourke	2000	EBCT	16	3683	91	49	Calcium score >100	High sensitivity, but low specificity
Picano	2000	Dipyridamole Echo	38	2856	73	91	n/a	Comparable diagnostic accuracy
		Dobutamine Echo	59	5082	81	83		
Berry	2001	EBCT	7	1246	94	48	No. of slices	Specificity EBCT is low
De Albuquerque	2001	Dipyridamole Echo	8	533	72	92	n/a	Similar diagnostic accuracy
Fonseca		Exercise Echo	8	533	79	82		
Kim	2001	Adenosine Echo	6	516	72	91	Definition of disease,	Highest combination of sensitivity
		Dipyridamole Echo	20	1835	70	93	prevalence	and specificity: dobutamine Echo
		Dobutamine Echo	40	4097	80	84		
		Adenosine SPECT	9	1207	90	75		
		Dipyridamole SPECT	21	1464	89	65		
		Dobutamine SPECT	14	1066	82	75		
Nallamothu	2001	EBCT	14	1662	92	51	Sample size >100	Reasonable diagnostic performance

Publ year, publication year; Sens, sensitivity; Spec, specificity; SROC, summary receiver operating characteristic; Echo, echocardiography; EBCT, electron-beam computed tomography; SPECT, single-photon emission computed tomography; n/a, not applicable

responded. Two meta-analyses were excluded because they did not use angiography as reference test, because an emergency setting was evaluated.[13,16] Two meta-analyses were excluded because they were duplicates of other meta-analyses: one was a consensus document published twice[3], the second a critique on meta-methodology.[15] Because these meta-analyses included the same source data and provided the same pooled values of sensitivity and specificity as the original meta-analyses,[12,18] the studies were excluded from further analysis.

Ten non-invasive imaging tests were evaluated in the 7 remaining meta-analyses including 310 patient series, of which 64 identical series were excluded. The final data set consisted of 246 patient series, including 24,761 patients.

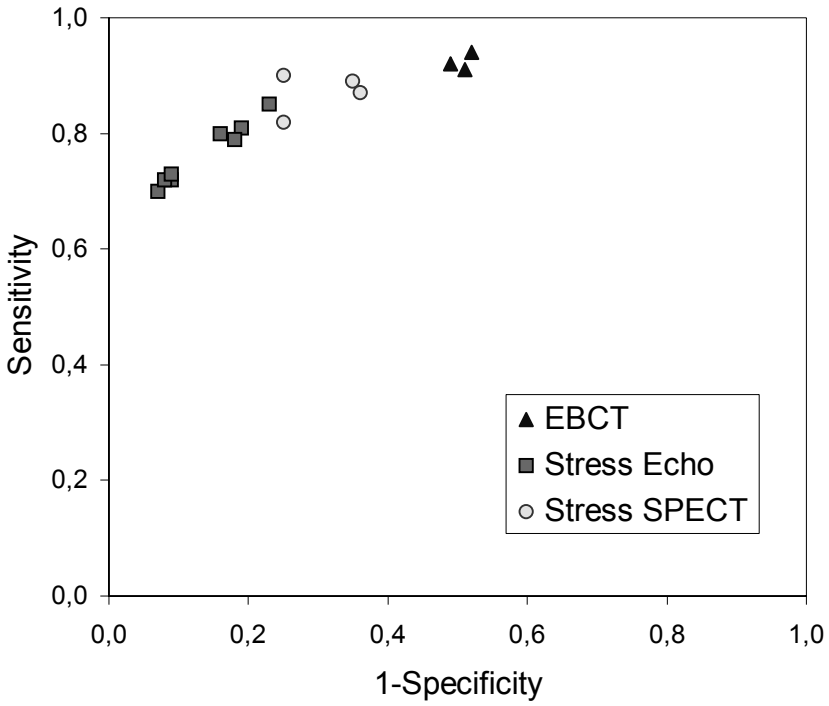


Figure 1: Pooled summary estimates of sensitivity and specificity of stress echocardiography, stress SPECT, and EBCT for the diagnosis of coronary artery disease obtained from meta-analyses. The test characteristics are listed in table 1.

EBCT, electron beam computed tomography; Echo, echocardiography; SPECT, single-photon emission computed tomography

Comparison of test characteristics

Table 1 gives an overview of the meta-analyses that were included in the current study. The table shows the pooled values of sensitivity and specificity per meta-analysis, which were reported or calculated from the source data. In Figure 1 these values are plotted in ROC space per type of test. An interesting result is that the summary estimates from 7 different meta-analyses together have the shape of an SROC curve. Furthermore, the EBCT and stress SPECT tests are clustered at the right tail of this hypothetical SROC curve indicating high sensitivity and reduced specificity, whereas the stress echocardiography tests are clustered at the left tail indicating high specificity and reduced sensitivity. At the upper left of the dots there is room for improvement of both the sensitivity and specificity of the imaging tests for coronary artery disease.

Table 2 shows the pooled sensitivity and specificity per test type calculated using a random effects meta-analysis after exclusion of duplicate studies. The results again indicate a higher specificity for the stress echocardiography tests and a higher sensitivity for the corresponding stress SPECT tests and EBCT.

Figure 2 shows the summary estimates of the diagnostic log odds ratio, which incorporates both sensitivity and specificity, and the corresponding 95% confidence interval for each type of test. Note that all confidence intervals overlap each other indicating no statistical difference between the tests. Adenosine SPECT testing showed relatively high diagnostic performance compared with the other tests, but there was considerable heterogeneity in these studies which is denoted by the wide confidence interval. The diagnostic odds ratios of exercise, dobutamine, and dipyridamole echocardiography were higher than the corresponding stress SPECT tests and EBCT, but not significantly.

In the regression analysis significant predictors of diagnostic performance were the type of test, the year of publication, and the proportion of men. Adenosine SPECT performed significantly better than exercise SPECT. There were no significant differences between the other coronary tests. The relative odds ratios are shown in table 3. Correction for publication year reduced the width of the confidence intervals but did not alter the significance of the differences between the test types. The effect of publication year on diagnostic performance is shown in figure 3. The area under the SROC curve significantly declined when comparing early-published studies and studies published 10 years later. The proportion of men was available in only 161 of the 246 studies. The difference between adenosine SPECT and exercise SPECT remained significant in this subgroup. In addition, there was a trend towards better diagnostic performance ($p<0.10$) for exercise echocardiography versus exercise SPECT (relative odds ratio (ROR) 1.65; 95% CI 0.98-2.77), for adenosine SPECT versus dobutamine SPECT (ROR 2.01; 95% CI 0.89-4.62), and for adenosine SPECT versus EBCT (ROR 2.10; 95% CI 0.91-4.85) in this subgroup analysis.

Table 2: Pooled sensitivity, specificity, and corresponding 95% confidence intervals per type of test using a random effects meta-analysis

Test	N	Sensitivity (%)	95% CI	Specificity (%)	95% CI
Exercise echo	28	83.8	80.3-87.3	80.0	74.5-85.5
Adenosine echo	6	74.2	62.8-85.5	91.2	86.3-96.0
Dipyridamole echo	49	72.1	68.6-75.6	94.5	92.7-96.3*
Dobutamine echo	72	81.2	0.79-83.5	84.1	81.8-86.5*
Exercise SPECT	27	87.6	84.9-90.2	69.4	60.7-78.2
Adenosine SPECT	8	90.9	89.0-92.7†	76.5	64.7-88.2
Dipyridamole SPECT	21	92.0	88.8-95.2†	73.8	64.2-83.5
Dobutamine SPECT	14	85.3	79.9-90.6	76.3	72.0-80.6
EBCT	21	93.1	90.7-95.6‡	54.5	45.3-63.8‡

CI, confidence interval; N, number of studies included; echo, echocardiography; SPECT, single-photon emission computed tomography

* non-overlapping confidence intervals indicating a statistically higher specificity than the corresponding SPECT test

† non-overlapping confidence intervals indicating a statistically higher sensitivity than the corresponding echocardiography test

‡ non-overlapping confidence intervals indicating a statistically higher sensitivity than all other tests, except for dipyridamole SPECT and a statistically lower specificity than all other tests except for exercise SPECT

Table 3: Multivariable model for coronary tests (n=246)

	Regression coefficient	Relative Odds Ratio	95% CI
Constant	2.92		
S	-0.024	0.98	0.88 – 1.08
Type of test			
Exercise Echo	0.24	1.27	0.74 – 2.18
Adenosine Echo	0.33	1.39	0.54 – 3.56
Dipyridamole Echo	0.33	1.39	0.76 – 2.56
Dobutamine Echo	0.20	1.22	0.75 – 1.99
Exercise SPECT	-0.13	0.88	0.52 – 1.49
Adenosine SPECT	0.68	1.97	0.94 – 4.14
Dipyridamole SPECT	0.11	1.12	0.61 – 2.05
Dobutamine SPECT	0.03	1.03	0.55 – 1.95
EBCT (Reference)	0	1	
Study characteristics			
Publication year-1990	-0.051	0.95	0.91 – 1.00

Echo, echocardiography; Ex, exercise; Ad, adenosine; Dip, dipyridamole; Dob, dobutamine; Publication year-1990, publication year after 1990

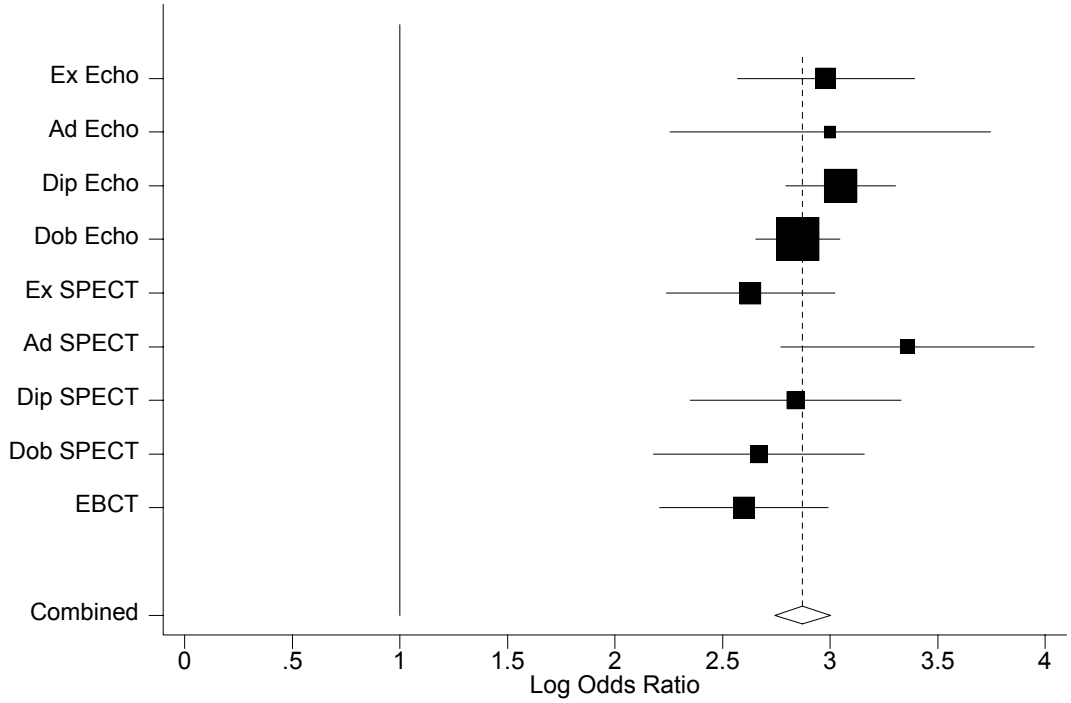


Figure 2: Summary diagnostic log odds ratios (boxes) and 95% confidence intervals (horizontal lines) for each type of test. The area of each box is inversely proportional to the variance in the test group, hence giving more visual prominence to test groups where the effect is more precisely estimated.
Ex, exercise; Ad, adenosine; Dip, dipyridamole; Dob, dobutamine; Echo, echocardiography

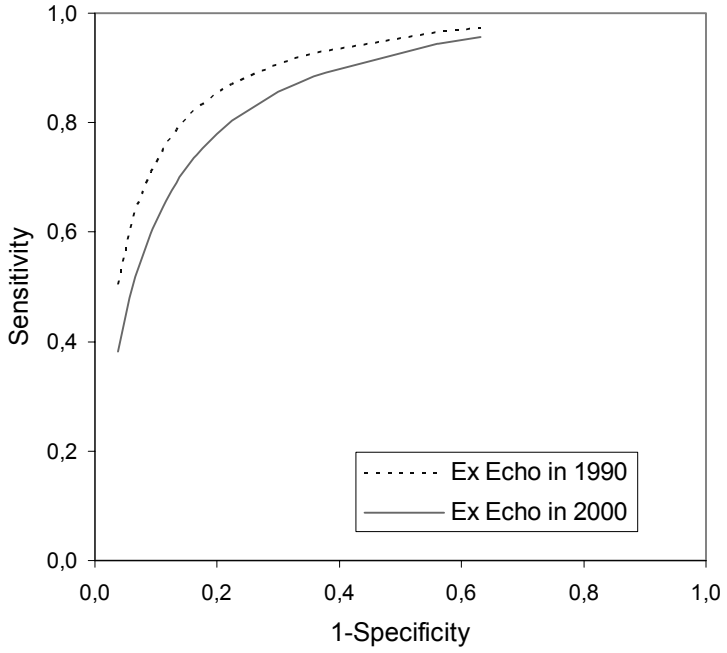


Figure 3: Summary receiver operating characteristic curves for exercise echocardiography (Ex Echo) adjusted for the year of publication. The upper curve represents the studies published in 1990 and the lower curve represents studies published in the year 2000. In 10 years time the area under the curve has significantly declined.

Discussion

In this study we combined data from 7 meta-analyses that included 24,761 patients, who underwent 8 different non-invasive imaging technologies for coronary artery disease. The analysis showed that the differences in diagnostic performance between the coronary tests were small. We showed that stress SPECT and EBCT are more sensitive tests while stress echocardiography is more specific. If a very sensitive test is used, the number of false-negative test results is minimized, resulting in fewer missed diagnoses of truly diseased patients, but it also increases the number of patients that unnecessarily will be referred to invasive coronary angiography. If a very specific test is used, the number of false-positive test results is minimal resulting in less unnecessary angiographies, but it may also delay the diagnosis in patients with false-negative test results. Which of these options should take priority depends on the costs and consequences of false test results in a specific clinical setting, which should be studied in cost-effectiveness analyses.

In general, we found that the diagnostic performance of stress echocardiography tended to be higher than that of stress SPECT and EBCT, but not significantly. The American College of Cardiology/American Heart Association (ACC/AHA) guidelines have discouraged the use of EBCT for screening and diagnosing subjects suspected of coronary artery disease because of the low specificity of this test, which may unnecessarily increase the number of patients referred for further medical testing.[3] In the recently updated guidelines, however, it is acknowledged that in clinical practice asymptomatic patients often present with abnormal EBCT results. For these patients further non-invasive diagnostic work-up is advised.[2] Stress echocardiography and stress SPECT are both recommended as initial test in patients with suspected coronary artery disease.[2] The availability of equipment, practice style, and local expertise usually dictate which test is used in a specific clinical setting. In the literature we found that apart from stress echocardiography and stress SPECT, these patients are also evaluated with EBCT. Therefore we thought it was justified to compare the performance of all these imaging techniques.

Our findings indicate that there is room for improvement of the diagnostic performance of imaging tests in coronary artery disease. The limiting factor in current imaging technology for evaluation of the degree of stenosis in the coronary arteries is that indirect measures of arterial disease are used, such as myocardial wall motion abnormalities, perfusion abnormalities, and the presence of coronary calcium, whereas in coronary angiography, the reference standard, the vessel lumen and blood flow through the vessel are directly visualized. Consequently, stress echocardiography and stress SPECT are more useful tests for functional risk assessment rather than evaluation of the degree of stenosis. Alternatively, it is debated whether angiographic evidence of coronary stenosis is a good reference standard for the diagnosis of coronary artery disease because this reference standard does not reflect the functional importance of the stenosis.[21] Another limitation of imaging the coronary arteries for evaluation of stenosis is the continuous beating of the heart and respiratory motion which both cause motion artifacts. Furthermore, the small vessel size of the coronary arteries requires a very high resolution of imaging technology. More advanced technologies, like contrast-enhanced magnetic resonance angiography (MRA) and computed tomographic angiography (CTA) are still under development for diagnosing coronary artery disease. Nevertheless, preliminary results show that there is a great potential for these techniques to improve non-invasive diagnostic testing in

coronary artery disease.[22,23] These non-invasive angiography techniques directly image the vessel lumen similar to intra-arterial angiography.

We did not exclude studies based on the quality of the data reported. Jüni et al. showed that studies should not be excluded based on composite quality scores, because many quality scales are more closely related to reporting quality rather than to the internal validity of studies.[24] Instead, relevant methodological aspects should be assessed individually and their influence on effect sizes explored. Therefore, we only included studies that used the same reference standard and accounted for differences in methodological quality by adding covariates to our statistical analysis, such as the number of patients included, proportion of diseased patients, gender and publication year.

The year of publication was a significant negative predictor in the coronary tests, indicating a decline in diagnostic accuracy over time despite improvements in imaging technology. This could be the result of selection bias in diagnostic test evaluations. First test results of a new technology, which are primarily optimistic, are usually obtained in highly selected patients using strict in- and exclusion criteria, but when in use longer and applied to a broader patient spectrum the results may become less impressive. Another potential explanation could be the effect of publication bias. Stern and Simes have shown that positive results are not only more likely to be published than negative results, but they also have a significantly shorter time to publication.[25] The decline is probably not due to a change in practice patterns, because such changes were not observed in a study by Miller et al, who evaluated temporal trends in clinical characteristics, stress test results and use of invasive procedures over a 10-year period.[26] Gender also significantly influenced diagnostic performance of coronary testing: diagnostic performance was better in men than in women. The diagnosis of coronary artery stenosis in women has been thought to be more difficult than in men, owing to the overall lower prevalence and severity of coronary disease in women, as well as more subtle clinical presentations.[27]

Our study showed that the differences among the stress tests were small. Only adenosine SPECT showed significantly better diagnostic performance than exercise SPECT, also when correcting for publication year or the proportion of men included. Because the patient population undergoing pharmacological imaging, i.e. patients who cannot perform a physical stress test, is usually different from the population undergoing exercise testing, this finding may have little relevance to clinical practice. However, a few head-to-head comparisons have been performed comparing exercise and adenosine SPECT, showing contradictory results.[28-31] When correcting for the proportion of men included, we found that adenosine SPECT also tended to be better than other types of SPECT testing. Current ACC/AHA practice guidelines do not dictate whether adenosine or dipyridamole should be used in SPECT testing, whereas the use of dobutamine in combination with SPECT is only recommended if vasodilators are contra-indicated.[32] Our findings may suggest that adenosine should be preferred over dipyridamole if pharmacological stress SPECT is considered. However, these results should be interpreted with caution, because the number of adenosine studies was small and the heterogeneity between studies was large.

A limitation of this study was that by using data published in meta-analyses as the basis for our statistical analyses, the most recent studies on imaging tests are not included. Another

limitation is that we could not evaluate the effect of all significant predictors from each individual meta-analysis, because these predictors were not commonly published. Furthermore, publications from one research group that were excluded from a meta-analysis because of overlapping data may have been included in another meta-analysis. Therefore, even though we took care to exclude duplicate data sets, some residual overlap may have existed in the final data set.

A problem that is often present in diagnostic test comparisons is the possibility of verification bias. Verification bias may occur when patients are referred to the reference test based on the results of the non-invasive test under investigation. Sensitivity may be inflated and specificity deflated if patients with a positive test result are more likely to be verified. Verification bias may have been present in the source data of our study and thus in our summary estimates of sensitivity and specificity of the diagnostic tests. Consequently, tests should not be compared based on point estimates of sensitivity and specificity alone. In our meta-analysis, however, we focused on the diagnostic odds ratio, a measure in which sensitivity and specificity are combined. The diagnostic odds ratio is unbiased if the verification rate is not directly influenced by the unknown disease status of the patient.[33] The same assumption of conditional independence between the verification rate and the disease status is used in studies in which the measures of sensitivity and specificity are corrected for verification bias.[34] Our approach is supported by the results of Lijmer and coworkers, who have shown that the relative diagnostic odds ratio in studies with partial verification was similar to studies with complete verification.[5]

In conclusion, our results show that in coronary artery disease, the differences in diagnostic performance between imaging tests are small. Diagnostic performance of non-invasive imaging tests is significantly higher in men than in women suspected of coronary artery disease. Since the established coronary tests show a decline in diagnostic performance over time, future research should be directed to development and improvement of new technologies, like CTA and contrast-enhanced MRA, for the diagnosis of coronary artery disease.

References

1. Pohler E, Gunther H, Diekmann M, Eggeling T. Outpatient coronary angiography--safety and feasibility. Cardiology. 1994;84:305-309.
2. Gibbons RJ, Abrams J, Chatterjee K, et al. ACC/AHA 2002 guideline update for the management of patients with chronic stable angina--summary article: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on the Management of Patients With Chronic Stable Angina). J Am Coll Cardiol. 2003;41:159-168.
3. O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/ American Heart Association Expert Consensus Document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. J Am Coll Cardiol. 2000;36:326-340.
4. Hecht HS. Practice guidelines for electron beam tomography: a report of the Society of Atherosclerosis Imaging. Am J Cardiol. 2000;86:705-706.
5. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999;282:1061-1066.
6. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making. 1993;13:313-321.
7. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med. 1993;12:1293-1316.
8. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med. 1994;120:667-676.
9. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. Stat Med 1995;14:395-411.
10. de Albuquerque Fonseca L, Picano E. Comparison of dipyridamole and exercise stress echocardiography for detection of coronary artery disease (a meta-analysis). Am J Cardiol. 2001;87:1193-1196.
11. Detrano R, Gianrossi R, Mulvihill D, et al. Exercise-induced ST segment depression in the diagnosis of multivessel coronary disease: a meta-analysis. J Am Coll Cardiol. 1989;14:1501-1508.
12. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA. 1998;280:913-920.
13. Ioannidis JP, Salem D, Chew PW, Lau J. Accuracy of imaging technologies in the diagnosis of acute cardiac ischemia in the emergency department: a meta-analysis. Ann Emerg Med. 2001;37:471-477.
14. Kim C, Kwok YS, Heagerty P, Redberg R. Pharmacologic stress testing for coronary disease diagnosis: A meta-analysis. Am Heart J. 2001;142:934-944.
15. Kymes SM, Bruns DE, Shaw LJ, Gillespie KN, Fletcher JW. Anatomy of a meta-analysis: a critical review of "exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance". J Nucl Cardiol. 2000;7:599-615.
16. Lau J, Ioannidis JP, Balk EM, et al. Diagnosing acute cardiac ischemia in the emergency department: a systematic review of the accuracy and clinical effect of current technologies. Ann Emerg Med. 2001;37:453-460.
17. Nallamothu BK, Saint S, Bielak LF, et al. Electron-beam computed tomography in the diagnosis of coronary artery disease: a meta-analysis. Arch Intern Med. 2001;161:833-838.
18. O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/ American Heart Association Expert Consensus document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. Circulation. 2000;102:126-140.
19. Picano E, Bedetti G, Varga A, Cseh E. The comparable diagnostic accuracies of dobutamine-stress and dipyridamole-stress echocardiographies: a meta-analysis. Coron Artery Dis. 2000;11:151-159.

20. Berry E, Kelly S, Hutton J, et al. A systematic literature review of spiral and electron beam computed tomography: with particular reference to clinical applications in hepatic lesions, pulmonary embolus and coronary artery disease. Health Technol Assess 1999;3:1-118.
21. Ellestad MH. The time has come to reexamine the gold standard when evaluating noninvasive testing. Am J Cardiol. 2001;87:100-101.
22. Kim WY, Danias PG, Stuber M, et al. Coronary magnetic resonance angiography for the detection of coronary stenoses. N Engl J Med. 2001;345:1863-1869.
23. Nieman K, Cademartiri F, Lemos PA, Raaijmakers R, Pattynama PM, de Feyter PJ. Reliable noninvasive coronary angiography with fast submillimeter multislice spiral computed tomography. Circulation. 2002;106:2051-2054.
24. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282:1054-1060.
25. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ. 1997;315:640-645.
26. Miller TD, Roger VL, Hodge DO, Hopfenspirger MR, Bailey KR, Gibbons RJ. Gender differences and temporal trends in clinical characteristics, stress test results and use of invasive procedures in patients undergoing evaluation for coronary artery disease. J Am Coll Cardiol. 2001;38:690-697.
27. Tong AT, Douglas PS. Stress echocardiography in women. Cardiol Clin 1999;17:573-82.
28. Levine MG, Ahlberg AW, Mann A, et al. Comparison of exercise, dipyridamole, adenosine, and dobutamine stress with the use of Tc-99m tetrofosmin tomographic imaging. J Nucl Cardiol. 1999;6:389-396.
29. Cuocolo A, Nicolai E, Soricelli A, et al. Technetium 99m-labeled tetrofosmin myocardial tomography in patients with coronary artery disease: comparison between adenosine and dynamic exercise stress testing. J Nucl Cardiol. 1996;3:194-203.
30. Abe S, Takeishi Y, Chiba J, Ikeda K, Tomoike H. Comparison of adenosine and treadmill exercise thallium-201 stress tests for the detection of coronary artery disease. Jpn Circ J. 1993;57:1111-1119.
31. Muller-Suur R, Eriksson SV, Strandberg LE, Mesko L. Comparison of adenosine and exercise stress test for quantitative perfusion imaging in patients on beta-blocker therapy. Cardiology. 2001;95:112-118.
32. Gibbons RJ, Chatterjee K, Daley J, et al. ACC/AHA/ACP-ASIM 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 (Committee on Management of Patients With Chronic Stable Angina). J Am Coll Cardiol. 1999;33:2092-2197.
33. Knottnerus JA. The effects of disease verification and referral on the relationship between symptoms and diseases. Med Decis Making. 1987;7:139-148.
34. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983;39:207-215.

3

Non-invasive imaging tests for peripheral, renal, and carotid artery disease: where do we stand?

Abstract

Purpose: The purpose of this study was to review the literature and to compare the diagnostic performance of computed tomographic angiography (CTA), magnetic resonance angiography (MRA), and duplex ultrasonography across anatomic regions for the assessment of peripheral, renal, and carotid artery disease, using meta-analytic techniques.

Materials and methods: We searched PubMed from January 1990 through May 2003 for meta-analytic studies. Meta-analyses on the diagnostic performance of imaging tests for arterial disease were selected that used angiography as reference standard and that presented the absolute numbers of true-positive, false-negative, true-negative and false-positive results of the source studies. Data were extracted on the level of the meta-analysis and on the level of the source studies that were included in each meta-analysis. Duplicate source studies were excluded. We combined data from 10 meta-analyses with a total of 260 patient series including 14,640 patients and 52,976 arterial segments. A summary-receiver-operating-characteristic (SROC) analysis was performed on the source data using a weighted random-effects model.

Results: The diagnostic performance was significantly better for contrast-enhanced MRA (relative odds ratio (OR) 4.08; 95% confidence interval (95% CI) 2.38-6.99) and CTA (OR 12.49; 95% CI 3.10-50.35) versus duplex ultrasonography, and significantly worse for renal (OR 0.59; 95% CI 0.36-0.97), and carotid artery disease (OR 0.60; 95% CI 0.39-0.92) versus peripheral disease. Diagnostic performance increased with the number of arterial segments analyzed (OR 1.06 per 100 segments; 95% CI 1.01-1.12).

Conclusions: Our study suggests that CTA and contrast-enhanced MRA have better diagnostic performance than non-enhanced MRA and duplex ultrasonography for the diagnosis of arterial disease.

Introduction

In patients with arterial disease in the peripheral, renal, or carotid arteries the planning of revascularization procedures requires precise localization and determination of severity of arterial stenoses. The accepted reference standard for the diagnosis of arterial disease in each of these anatomic regions is intra-arterial angiography. However, because of the risks and costs associated with this invasive procedure, intra-arterial angiography is not suitable as initial test. For this reason, non-invasive imaging technologies, such as duplex ultrasonography, magnetic resonance angiography (MRA), and computed tomographic angiography (CTA) are increasingly used in the diagnostic work-up in patients with peripheral, renal, and carotid artery disease.

Over the years, a number of studies have been published in which the diagnostic performance of non-invasive imaging tests has been evaluated for peripheral, renal and carotid artery disease. A broad spectrum of values for sensitivity and specificity has been reported in various studies on diagnostic testing. In the last decade, several meta-analyses have been performed summarizing the published data for each test and for each anatomic region. To understand the role of non-invasive tests for arterial disease we were interested in comparing the diagnostic performance of various non-invasive tests across anatomic regions. Furthermore, we questioned whether the diagnostic performance could be predicted by study characteristics. Therefore, we set out to review systematically all published meta-analyses on imaging the peripheral, renal, or carotid arteries by duplex ultrasound, CTA, and/or MRA, in order to quantify the diagnostic performance across anatomic regions. We followed the recently published standards for reporting of diagnostic accuracy and the guidelines for meta-analyses of observational studies when appropriate for a meta-analysis on diagnostic testing.[1,2]

The purpose of this study was to provide an overview of the diagnostic performance of non-invasive imaging tests in peripheral, renal, and carotid artery disease, to compare the diagnostic performance across anatomic regions, and to identify study characteristics that may influence the diagnostic performance.

Materials and Methods

Study selection

A literature search was performed in PubMed from January 1990 through May 2003 to identify meta-analyses on the diagnostic performance of non-invasive imaging tests for the diagnosis of arterial disease. We used the following search terms: *meta-analysis* or *systematic review* or *structured review* and *imaging test*. We also checked the references of the included articles for additional meta-analyses. Articles were included if they met the following inclusion criteria: (a) the diagnostic performance of imaging tests for peripheral, renal, or carotid artery disease was studied, (b) meta-analytic methods were used, (c) angiography was used as reference standard, (d) the absolute numbers of true-positive, false-negative, true-negative and false-positive results of the source studies were available or derivable from the meta-analyses, and (e) it was published after 1990. If inclusion criterion *d* was not met, the authors were requested to provide these data. Because of the rapid advances in imaging technology, meta-analyses published before 1990 were excluded. If a meta-analysis compared an imaging

technology with a technology that is no longer in use or with a non-imaging technology, only the data on the newer imaging technology were included.

Data extraction

The data were extracted on two levels: on the level of the meta-analysis and on the level of the source studies that were included in each meta-analysis. The following data were extracted per meta-analysis: author, journal, year of publication, type of test, anatomic region, number of studies included, total number of patients, total number of arterial segments, summary estimates of test characteristics and meta-analytic methods. In addition to the study identification and study size, the publication year, the number of arterial segments, mean age, percentage of men, and the absolute number of true-positive, false-negative, false-positive, and true-negative test results were extracted per source study. If source studies of meta-analyses overlapped and the data presented in the meta-analyses were identical, the duplicate data were excluded. If discrepancies existed between duplicate source studies, the original publication of the source study was retrieved and, based on the original article, the discrepant source data were excluded. The anatomic regions considered were the carotid, renal, and peripheral regions. If a meta-analysis was done on the same source studies for several definitions of significant stenosis, the data were extracted only for one definition: for peripheral and renal artery disease this definition was 50-100% stenosis and for carotid artery disease it was 70-99% stenosis, in accordance with the indications for further diagnostic work-up or therapeutic intervention.[3-6] The non-invasive imaging tests included duplex ultrasonography, (contrast-enhanced (CE-)) MRA, and CTA.

Data-analysis

Comparison of meta-analyses

First, the data were evaluated on the level of the meta-analysis. The published summary estimates of sensitivity and specificity were plotted per test type and grouped per anatomic region, with 1-specificity on the horizontal axis and sensitivity on the vertical axis. If the meta-analysis did not present summary estimates of sensitivity and specificity, the pooled values of sensitivity and specificity were calculated from the absolute numbers of true and false test results of the source studies. Because most of the meta-analyses presented fixed summary estimates, we calculated the pooled values of sensitivity and specificity weighted for sample size only.

Summary receiver operating characteristic (SROC) analysis

We performed a random effect SROC analysis on the data set of source studies to calculate the relationship between sensitivity and specificity, taking into account the differences in positivity criterion and other factors of heterogeneity between settings.[7-9] In an SROC analysis the true-positive rate (TPR) (sensitivity), and the false-positive rate (FPR) (1-specificity), are transformed into the logit form. Then, a linear regression analysis is performed on the variables D and S, D representing the natural logarithm of the diagnostic odds ratio and S functioning as

proxy for the positivity criterion of the diagnostic test. D and S are defined as the difference and the sum of the logit transformations of TPR and FPR respectively, in formula:

$$D = \ln\frac{TPR}{1-TPR} - \ln\frac{FPR}{1-FPR}$$

$$S = \ln\frac{TPR}{1-TPR} + \ln\frac{FPR}{1-FPR}$$

To prevent undefined values for D and S due to 0's in the equations, 0.5 was added to the absolute numbers of true-positive, false-negative, true-negative and false-positive test results for the calculation of D and S. The regression analysis was weighted with the inverse of the variance of D. We used a random effects model, which takes into account the variability between studies.[10] All data were analyzed in a linear regression model in which D, the log odds ratio, represented the diagnostic performance depending on a variable indicating the type of test. The differences between the tests are represented by the relative odds ratios, with a value equal to 1 indicating no difference in diagnostic performance, a value larger than 1 indicating better performance and values smaller than 1 indicating reduced performance. Additional covariates were added to the model to evaluate the influence of other factors on diagnostic performance and to correct for confounders when comparing tests. The following covariates were considered: the anatomic region (peripheral, renal, or carotid), the number of patients tested, the number of arterial segments included, the number of arterial segments per patient, the mean age of the patients, the proportion of segments diseased, the proportion of men, and the year of publication. First, we evaluated the individual effect of each variable on the diagnostic performance in the model that always included S. Starting with the variables with the lowest residual between-study variance (determined by tau-squared), the significant variables were added to the multivariable model in a stepwise forward manner. A significance level of 0.05 was used. Tau-squared, the residual between-study variance calculated with the restricted maximum likelihood method, was used as a measure of the model fit. A lower tau-squared indicates less residual between-study variance and therefore a better model fit and a better explanatory power by the model of the heterogeneity across studies. We used STATA 8.0 for the statistical analyses.

Results

Study selection

The search strategy resulted in 120 abstracts of which 12 meta-analyses were found that reported on peripheral, renal, or carotid arterial disease.[11-22] Of the 12 meta-analyses, one study was excluded because it was published before 1990.[21] Four meta-analyses did not report the absolute numbers of true-positive, false-positive, true-negative, and false-negative test results of the source studies.[11,12,14,22] The authors of these meta-analyses were requested to provide the missing information, of which 3 responded. Thus, 1 meta-analysis was excluded for missing data of source studies.[22] Of the remaining 10 meta-analyses, 2 studies focused on carotid artery disease, 6 on peripheral arterial disease and 2 on renal artery disease. Six non-invasive tests were evaluated in these 10 meta-analyses.

Table 1: Overview of the included meta-analyses

Meta-analysis	Publ year	Disease	Type of Test	No. series	No. patients	No. units	Units/ Patient	Pooled Sens	Pooled Spec	Significant predictors SROC analysis	Conclusion
Westwood[11]	2002	CAS	CE-MRA	4	148	200	1.35	98	99	None	Diagnostic performance trend CE-MRA>MRA not significant
			MRA	18	695	1438	2.07	93	94		
Nederkoorn[12]	2003	CAS	DUS	65	6670	11377	1.71	83	86	Type MRA scanner, verification bias, difference in cut-off	Discriminatory power (CE-)MRA>DUS
			CE-MRA	4	126	208	1.65	92	80		
			MRA	16	681	1306	1.92	92	87		
De Vries[13]	1996	PAD	DUS	14	613	4906	8.00	86	96	Color flow	Diagnostic performance Color DUS>DUS
Koelemay[14]	1996	PAD	DUS	16	759	1499	1.97	85	97	n/a	DUS is accurate tool
Nelemans[15]	2000	PAD	CE-MRA	10	253	2578	10.19	92	95	Test, post-processing technique	Diagnostic accuracy CE-MRA>MRA
			MRA	13	344	3396	9.87	89	89		
Visser[16]	2000	PAD	CE-MRA	9	216	2051	9.50	98	96	Test, time between tests	Discriminatory power CE-MRA>DUS
			DUS	18	1059	11221	10.60	88	95		
Koelemay[17]	2001	PAD	CE-MRA	26	803	7581	9.44	94	96	Test, no. of segments	Diagnostic performance CE-MRA>MRA
			MRA	27	885	6476	7.32	91	91		
Berry[18]	2002	PAD	CE-MRA	6	210	3182	15.15	96	97	None	Diagnostic performance not significantly different
			MRA	10	301	1738	5.77	92	87		
Vasbinder[19]	2001	RAS	CE-MRA	6	288	468	1.63	98	94	Sample size>50	Diagnostic accuracy CTA and CE-MRA> other tests
			MRA	12	458	809	1.77	91	87		
			CTA	5	317	734	2.32	96	96		
			DUS	24	1592	2437	1.53	84	90		
Tan[20]	2002	RAS	CE-MRA	15	499	990	1.98	94	85	n/a	Diagnostic accuracy CE-MRA>MRA
			MRA	12	499	993	1.99	97	93		

Publ year, publication year; Sens, sensitivity; Spec, specificity; SROC, summary receiver operating characteristic; CAS, carotid artery stenosis; PAD, peripheral arterial disease; RAS, renal artery stenosis; DUS, duplex ultrasonography; (CE-) MRA, (contrast-enhanced) magnetic resonance angiography; CTA, computed tomographic angiography; n/a, not applicable

Two of the tests included in an analysis of renal artery disease were excluded, the captopril test (non-imaging) and captopril renal scintigraphy[19] because these tests are not used for peripheral and carotid arterial disease . The remaining 4 imaging tests, duplex ultrasonography, CTA, MRA, and contrast-enhanced MRA, were studied in 330 patient series, of which 70 duplicate series were excluded (21%). The final data set consisted of 260 patient series, 94 on peripheral, 66 on renal, and 100 on carotid artery disease, including 14,640 patients and 52,976 arterial segments.

Comparison of meta-analyses

Table 1 gives an overview of the meta-analyses that were included in the current study. The table shows the pooled values of sensitivity and specificity per meta-analysis, which were reported or calculated from the source data. The majority of the meta-analyses concluded that the diagnostic performance of contrast-enhanced MRA was significantly higher than non-enhanced MRA and duplex ultrasonography. Comparing the analytical methods we found that two[14,20] out of the 10 meta-analyses performed only a pooled analysis and did not perform a SROC analysis. Furthermore, we observed a very high number of arterial units analyzed per patient in the peripheral arterial disease studies compared with the renal and carotid regions.

In Figure 1 the sensitivity and specificity summary estimates that were obtained from the meta-analyses are plotted in ROC space per anatomic region. They are all scattered over the upper left corner of ROC space, indicating high diagnostic performance.

SROC analysis

Figure 2 presents the summary estimates of the log odds ratio for each test per anatomic region. The peripheral tests showed higher diagnostic log odds ratios than the renal and carotid tests. Furthermore, there was a trend towards higher diagnostic log odds ratios for MRA, contrast-enhanced MRA and CTA compared with duplex ultrasonography. In carotid artery disease, the number of contrast-enhanced MRA studies was small and the heterogeneity was substantial. The CTA-studies, only available for renal artery disease, showed the best diagnostic performance.

These results were in agreement with the results of the multivariable regression analysis, in which the type of test, anatomic region and total number of segments analyzed were significant predictors of diagnostic performance (table 2). CTA and contrast-enhanced MRA were significantly better than non-enhanced MRA and duplex ultrasonography, corrected for the anatomic region and the total number of arterial segments analyzed. Publication year, the number of patients, the number of arterial segments per patient, age, percentage of men, and percentage of diseased segments did not influence diagnostic performance significantly.

In Figure 3 the SROC curves are plotted per test and per anatomic region assuming a study with 100 arterial segments analyzed. The upper curves of MRA and duplex ultrasonography represent testing for peripheral arterial disease and the lower curves for renal artery disease. Each pair of curves together forms an SROC area. The curves for carotid arterial disease lie within these SROC areas. The SROC curves for CTA and contrast-enhanced MRA are closest to the upper left corner of SROC space, indicating excellent diagnostic performance. Diagnostic performance of non-enhanced MRA did not differ significantly from duplex ultrasonography and is therefore not shown in this figure.

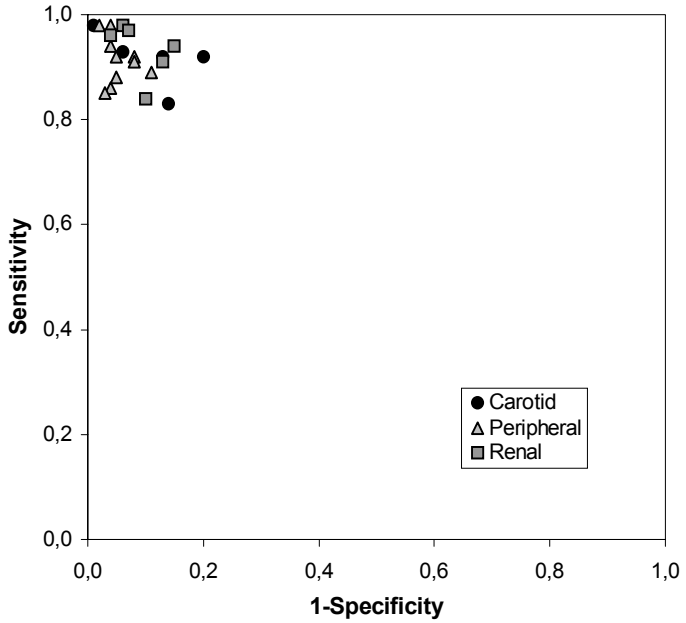


Figure 1: Pooled summary estimates of sensitivity and specificity of several non-invasive imaging tests for peripheral, renal, and carotid artery disease obtained from published meta-analyses. The tests are listed in table 1.

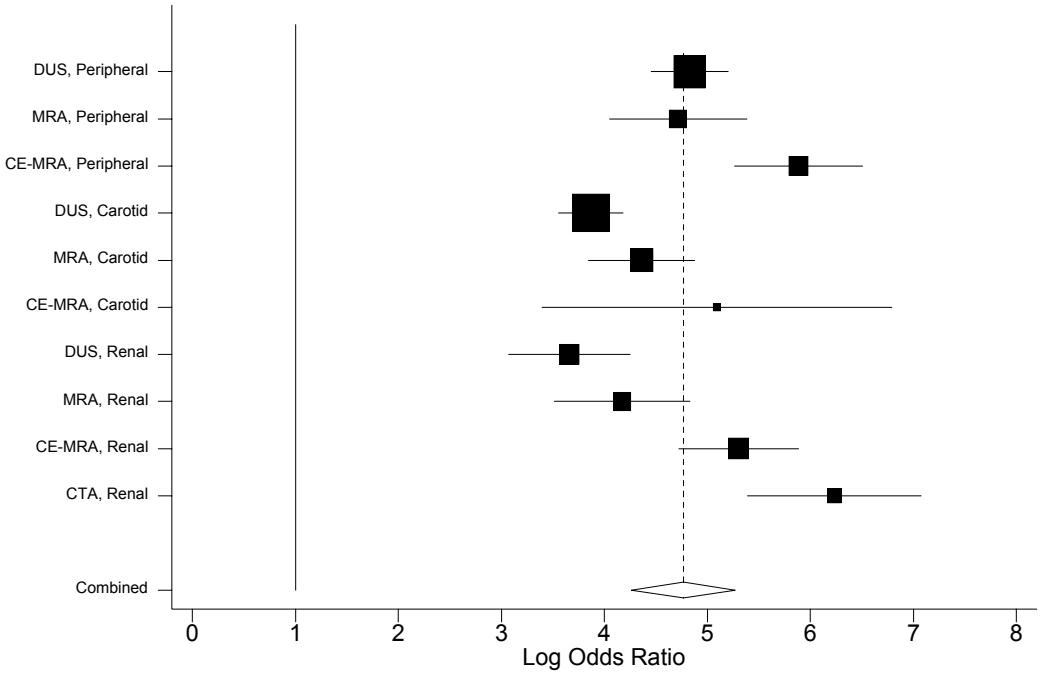


Figure 2: Summary diagnostic log odds ratios (boxes) and 95% confidence intervals (horizontal lines) for each non-invasive test for peripheral, renal, and carotid artery disease. The area of each box is inversely proportional to the variance in the test group, hence giving more visual prominence to test groups where the effect is more precisely estimated.

DUS, duplex ultrasonography; (CE-) MRA, (contrast-enhanced) magnetic resonance angiography; CTA, computed tomographic angiography

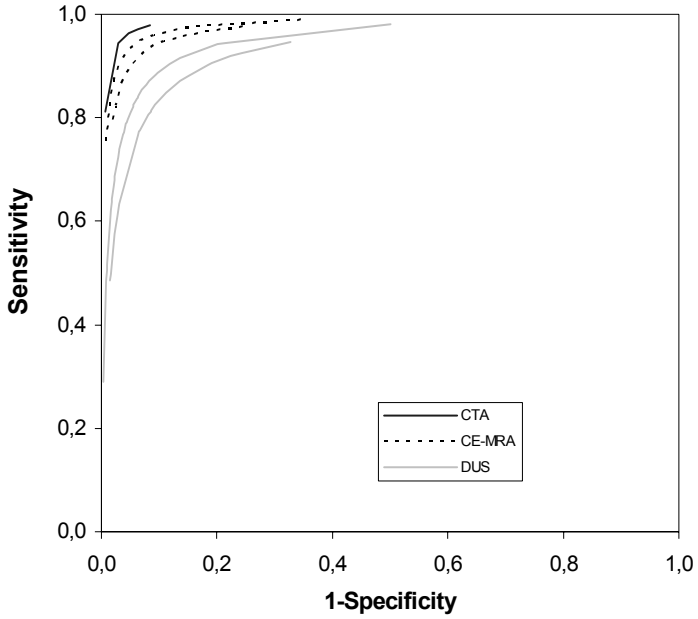


Figure 3: Summary receiver operating characteristic curves for renal CTA, and SROC areas for duplex ultrasonography and contrast-enhanced MRA for a study population of 100 arterial segments. CTA, computed tomographic angiography; CE-MRA, contrast-enhanced magnetic resonance angiography; DUS, duplex ultrasonography

Table 2: Multivariable model for non-invasive imaging tests for arterial disease (n=260 series)

	Regression coefficient	Relative Odds Ratio	95% CI
Constant	4.26		
S	-0.084	0.92	0.80 – 1.05
Type of test			
CTA	2.53	12.49	3.10 – 50.35
CE-MRA	1.40	4.08	2.38 – 6.99
MRA	0.38	1.46	0.95 – 2.23
DUS (Reference)	0	1	
Anatomic region			
Carotid	-0.51	0.60	0.39 – 0.92
Renal	-0.53	0.59	0.36 – 0.97
Peripheral (Reference)	0	1	
Study characteristics			
No of segments/100	0.062	1.06	1.01 – 1.12

95% CI, 95% confidence interval; DUS, duplex ultrasonography; CTA, computed tomographic angiography; (CE-) MRA, (contrast-enhanced) magnetic resonance angiography; No of segments/100, total number of arterial segments per 100 extra segments

Discussion

In this study we combined data from 10 meta-analyses containing 260 patient series, including 14,640 patients and 52,976 arterial segments. We compared the diagnostic performance of 4 non-invasive imaging technologies for peripheral, renal, and carotid artery disease. Our study suggests that contrast-enhanced MRA and CTA have significantly better diagnostic performance than non-enhanced MRA and duplex ultrasonography for the diagnosis of arterial disease, adjusted for the anatomical regions. These findings are consistent with the results of the individual meta-analyses on non-invasive imaging in peripheral and renal artery disease. In both meta-analyses available on carotid artery disease, however, contrast-enhanced MRA did not perform significantly better than non-enhanced MRA, probably due to the low number of source studies available.[11,12] Our analyses showed that despite a substantial heterogeneity in the carotid studies, the same effect was seen in the carotid studies as in the renal and peripheral studies indicating a better performance if contrast enhancement is used in MRA testing. With contrast-enhanced MRA and CTA the vessel lumen is visualized with contrast medium as in intra-arterial angiography, the reference standard. With duplex ultrasonography and non-enhanced MRA physical characteristics of the blood flow are used to image the lumen and to determine stenosis severity. Contrast-enhancement diminishes the possibility of flow-related artifacts, which may explain the better diagnostic performance.[23] However, the diagnostic performance of non-invasive angiography is not equivalent to intra-arterial angiography. There is some debate whether intra-arterial angiography is the best reference standard for arterial disease. Some studies have shown that intra-arterial angiography may underestimate the degree of stenosis compared with non-invasive techniques.[24,25] Furthermore, characterization of plaque morphology is not possible using intra-arterial angiography in contrast to non-invasive imaging techniques.

When comparing the performance of non-invasive imaging technologies across anatomic regions we found that the best diagnostic performance was obtained for the peripheral arteries. There were no differences in diagnostic performance between the renal and carotid arteries. One explanation for this finding could be that in peripheral arterial disease the sensitivity and specificity are inflated due to the subdivision of the arteries into multiple segments. However, the difference still existed after correction for the number of arterial segments studied. Another explanation for the worse performance of duplex ultrasonography and non-enhanced MRA in the renal arteries could be the presence of accessory renal arteries. In the carotid arteries slow flow or turbulent flow near the carotid bifurcation may cause overestimation of stenosis when non-enhanced imaging techniques are used. In contrast, the peripheral arteries are predominantly not tortuous, unidirectional, and have steady nonturbulent flow, and thus provide an ideal environment for non-enhanced imaging techniques.[26]

The inclusion of studies in the current analysis was not based on the quality of the data reported. Jüni et al. showed that studies should not be excluded based on composite quality scores, because many quality scales are more closely related to reporting quality rather than to the internal validity of studies.[27] Instead, relevant methodological aspects should be assessed individually and their influence on effect sizes explored. Therefore, we only included meta-analyses that used the same reference standard and we accounted for differences in methodological quality by adding covariates to our statistical analysis, such as the number of

patients or arterial segments included, percentage diseased, gender and publication year. Only the number of arterial segments studied significantly influenced diagnostic performance: the more segments studied, the higher the diagnostic performance. Dividing the arterial tract into multiple segments may raise estimates of sensitivity and specificity by increasing the sample size and mutual dependence within the sample. For example, including the arteries on the asymptomatic side overestimates test specificity, because the number of true negative results is increased. Similarly, dividing a severely and extensively diseased arterial tree into multiple segments overestimates test sensitivity because the number of true positive results is increased. Therefore, one must be careful when interpreting optimistic results of studies with a small number of patients that include both symptomatic and asymptomatic sides or divide the arteries into a high number of segments. In our opinion, the choice of whether the patient, side, or segment should be used as the unit of measurement in the calculations of test sensitivity and specificity depends on the associated clinical decision. If the decision is patient-based, e.g. whether or not to proceed to diagnostic angiography, the analysis should be performed with patients as the unit of measurement. If the decision is based on the arterial segment, e.g. whether or not to insert a stent, the analysis should be performed at the segmental level. If the asymptomatic side is irrelevant to the clinical decision, as is usually the case in carotid and peripheral arterial disease, then only the symptomatic side should be included.

In the current study we performed an overall analysis of the results from the source studies that were included in various meta-analyses. The overlap of source studies in these meta-analyses was substantial; we had to exclude 70 studies out of 330 (21%) of the source studies because of identical data. However, publications from one research group that were excluded from a meta-analysis because of partially overlapping data may have been included in another meta-analysis. Therefore, even though we took care to exclude duplicate data sets, some residual overlapping data may have existed in the final data set.

A limitation of using data published in meta-analyses as the basis for our statistical analyses is that the most recent studies on imaging tests are not included. Furthermore, CTA studies were analyzed in only one meta-analysis on renal artery disease and therefore these results should be interpreted with caution. Another limitation is that we could not evaluate the effect of all significant predictors from each individual meta-analysis, because these predictors were not commonly published.

A problem that is often present in diagnostic test comparisons is the possibility of verification bias. Verification bias may occur when patients are referred to the reference test based on the results of the non-invasive test under investigation. Sensitivity may be inflated and specificity deflated if patients with a positive test result are more likely to be verified. Verification bias may have been present in the source studies reporting the sensitivity and specificity of a diagnostic test. However, we used from each study the diagnostic odds ratio, in which sensitivity and specificity are combined. The diagnostic odds ratio is unbiased if the verification rate is not directly influenced by the unknown disease status of the patient.[28] Our approach is supported by the study from Lijmer and coworkers who have shown that the relative diagnostic odds ratio in studies with partial verification was comparable with studies with complete verification.[29]

In conclusion, our results suggest that CTA and contrast-enhanced MRA have better diagnostic performance than non-enhanced MRA and duplex ultrasonography for the diagnosis of arterial disease, adjusted for anatomical region. When comparing tests for arterial disease, the number of arterial segments studied should be taken into account: studies using a high number of arterial segments may overestimate diagnostic performance.

References

1. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Radiology. 2003;226:24-28.
2. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283:2008-2012.
3. Rutherford RB, Baker JD, Ernst C, et al. Recommended standards for reports dealing with lower extremity ischemia: revised version. J Vasc Surg. 1997;26:517-538.
4. Martin LG, Rundback JH, Sacks D, et al. Quality improvement guidelines for angiography, angioplasty, and stent placement in the diagnosis and treatment of renal artery stenosis in adults. J Vasc Interv Radiol. 2003;14:S297-310.
5. Barnett HJ, Taylor DW, Eliasziw M, et al. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. N Engl J Med. 1998;339:1415-1425.
6. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). Lancet. 1998;351:1379-1387.
7. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making. 1993;13:313-321.
8. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med. 1994;120:667-676.
9. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med. 1993;12:1293-1316.
10. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. Stat Med. 1995;14:395-411.
11. Westwood ME, Kelly S, Berry E, et al. Use of magnetic resonance angiography to select candidates with recently symptomatic carotid stenosis for surgery: systematic review. BMJ. 2002;324:198.
12. Nederkoorn PJ, van der Graaf Y, Hunink MG. Duplex ultrasound and magnetic resonance angiography compared with digital subtraction angiography in carotid artery stenosis: a systematic review. Stroke. 2003;34:1324-1332.
13. de Vries SO, Hunink MG, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis of the diagnostic performance of duplex ultrasonography in peripheral arterial disease. Acad Radiol. 1996;3:361-369.
14. Koelemay MJ, den Hartog D, Prins MH, Kromhout JG, Legemate DA, Jacobs MJ. Diagnosis of arterial disease of the lower extremities with duplex ultrasonography. Br J Surg. 1996;83:404-409.
15. Nelemans PJ, Leiner T, de Vet HC, van Engelshoven JM. Peripheral arterial disease: meta-analysis of the diagnostic performance of MR angiography. Radiology. 2000;217:105-114.
16. Visser K, Hunink MG. Peripheral arterial disease: gadolinium-enhanced MR angiography versus color-guided duplex US--a meta-analysis. Radiology. 2000;216:67-77.
17. Koelemay MJ, Lijmer JG, Stoker J, Legemate DA, Bossuyt PM. Magnetic resonance angiography for the evaluation of lower extremity arterial disease: a meta-analysis. JAMA. 2001;285:1338-1345.
18. Berry E, Kelly S, Westwood ME, et al. The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review. Health Technol Assess. 2002;6:1-155.
19. Vasbinder GB, Nelemans PJ, Kessels AG, Kroon AA, de Leeuw PW, van Engelshoven JM. Diagnostic tests for renal artery stenosis in patients suspected of having renovascular hypertension: a meta-analysis. Ann Intern Med. 2001;135:401-411.
20. Tan KT, van Beek EJ, Brown PW, van Delden OM, Tijssen J, Ramsay LE. Magnetic resonance angiography for the diagnosis of renal artery stenosis: a meta-analysis. Clin Radiol. 2002;57:617-624.
21. Feussner JR, Matchar DB. When and how to study the carotid arteries. Ann Intern Med. 1988;109:805-818.

22. Blakeley DD, Oddone EZ, Hasselblad V, Simel DL, Matchar DB. Noninvasive carotid artery testing. A meta-analytic review. Ann Intern Med. 1995;122:360-367.
23. Tatli S, Lipton MJ, Davison BD, Skorstad RB, Yucel EK. From the RSNA refresher courses: MR imaging of aortic and peripheral vascular disease. Radiographics. 2003;23:S59-78.
24. Pan XM, Saloner D, Reilly LM, et al. Assessment of carotid artery stenosis by ultrasonography, conventional angiography, and magnetic resonance angiography: correlation with ex vivo measurement of plaque stenosis. J Vasc Surg. 1995;21:82-88; discussion 88-89.
25. Elgersma OE, Wust AF, Buijs PC, van Der Graaf Y, Eikelboom BC, Mali WP. Multidirectional depiction of internal carotid arterial stenosis: three-dimensional time-of-flight MR angiography versus rotational and conventional digital subtraction angiography. Radiology. 2000;216:511-516.
26. Yucel EK, Anderson CM, Edelman RR, et al. AHA scientific statement. Magnetic resonance angiography : update on applications for extracranial arteries. Circulation. 1999;100:2284-2301.
27. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA. 1999;282:1054-1060.
28. Knottnerus JA. The effects of disease verification and referral on the relationship between symptoms and diseases. Med Decis Making. 1987;7:139-148.
29. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999;282:1061-1066.

4

Value of pharmacological stress testing in patients with chest pain: a cost-effectiveness analysis

Abstract

Context: Many studies have been performed to estimate the sensitivity and specificity of pharmacological stress testing, but little is known about the cost-effectiveness of these tests.
Objective: To estimate and compare the cost-effectiveness of adenosine, dipyridamole and dobutamine stress in combination with either echocardiography or single-photon-emission computed tomography (SPECT).
Design, Setting, and Patients: Cost-effectiveness analysis using a Markov decision model to compare six pharmacological stress testing strategies with routine coronary angiography and no imaging in patients presenting with chest pain. Several cohorts of patients were compared, defined by age, sex, and the type of chest pain. The analyses were performed from a societal perspective.
Main Outcome Measures: Lifetime costs, quality-adjusted life-years (QALYs) saved, and incremental cost-effectiveness ratios.
Results: Pharmacological echocardiography cost $31,700 - $33,100 per QALY saved in 55-year-old men and $37,100 - $42,300 per QALY in 65-year-old women with atypical angina. Pharmacological SPECT was only cost-effective in men with atypical angina if its cost was less than $420 or if angiography cost more than $5,000, whereas in women pharmacological SPECT cost more than $60,000 per QALY compared with echocardiography. Routine angiography was the optimal strategy in both men and women with typical chest pain ($25,400 per QALY in men and $31,200 in women). The optimal test strategy depended mostly on the pre-test probability of disease and the willingness-to-pay threshold.
Conclusions: Pharmacological stress echocardiography is cost-effective for the diagnosis of coronary artery disease in middle-aged men and women at intermediate risk of coronary artery disease if society is willing to pay more than $30,000 per QALY saved. Adenosine SPECT may be cost-effective for patients at intermediate risk in specific clinical situations or at higher willingness-to-pay thresholds. Routine coronary angiography is the optimal test strategy for patients with typical angina.

Introduction

Non-invasive stress testing is commonly used for the diagnosis of coronary artery disease in patients presenting with chest pain. Exercise testing is preferable if the patient can exercise to an appropriate level of cardiovascular stress. However, between one-third and one-half of patients in most centers are not able to exercise adequately and are therefore referred for pharmacological stress testing.[1]

The commonly used stress agents are the coronary vasodilators, adenosine and dipyridamole, and the catecholamine dobutamine, which are used in combination with echocardiography or single-photon-emission computed tomography (SPECT). Dobutamine is primarily used in combination with echocardiography, especially in the United States.[2,3] Vasodilators are the agent of choice in SPECT testing,[2,4] but dobutamine can also be used.[5] In Europe, however, the coronary vasodilators are also used in combination with echocardiography.[6]

Little information exists to guide the clinician about which test to order or to inform policy makers about which tests are cost-effective. A recent meta-analysis compared the sensitivity and specificity of pharmacological stress tests and showed that maximum specificity can be attained by vasodilator echocardiography, maximum sensitivity by vasodilator SPECT, and a best compromise between sensitivity and specificity by dobutamine echocardiography.[7] However, to determine which tests are most valuable in clinical practice we need to take into account not only the diagnostic accuracy of each test, but also the risks and costs of the test, and the effects of true- or false-positive, and true- or false-negative test results on subsequent treatment and long-term follow-up. A cost-effectiveness model integrates all relevant effects and costs and has the ability to determine the optimal diagnostic strategy out of various alternatives using a decision-analytic approach.

To explore the relative merit of pharmacological imaging tests in diagnosing coronary artery disease, we evaluated the cost-effectiveness of adenosine, dipyridamole, and dobutamine stress in combination with echocardiography or SPECT compared with initial testing with coronary angiography and a strategy without diagnostic imaging.

Methods

Target population

We focused on patients presenting with non-specific chest pain, atypical angina, or typical angina, as defined in the Coronary Artery Surgery Study (CASS) study.[8] In CASS, chest pain was defined as typical angina if it was substernal, was associated with physical exertion and was relieved by nitroglycerin. Atypical chest pain was defined as symptoms meeting two out of the three criteria, while nonspecific chest pain met only one of three. The type of chest pain determined the pre-test probability of coronary artery disease, which was 18%, 71%, and 95% in men and 10%, 48%, and 81% in women for patients with nonspecific chest pain, atypical angina, and typical angina respectively.[8] We modeled several cohorts of patients, defined by age, sex, and type of chest pain. As a base case a cohort of 55-year-old men with atypical angina was analyzed.

Markov Model

We modified and updated a previously developed Markov decision model[9] to estimate the lifetime costs and quality-adjusted life expectancy of pharmacological stress testing and the resulting treatment in patients presenting with chest pain. Six pharmacological stress test strategies were compared with direct coronary angiography or no imaging test at all: adenosine, dobutamine, and dipyridamole stress each in combination with echocardiography and SPECT. Each of the stress test strategies had 5 possible outcomes: true-positive, false-positive, true-negative, false-negative, or non-diagnostic.

We assumed that a stress test would be followed by coronary angiography if the test result was positive or non-diagnostic. The angiogram result was categorized into no disease, single, double, or triple vessel disease or left main coronary disease. The prevalence of each of these categories, stratified by age, sex, and type of chest pain was based on the CASS registry data.[8] We assumed that patients diagnosed with single vessel or double vessel coronary disease would undergo a percutaneous coronary intervention (PCI), whereas in patients with triple vessel or left main coronary artery disease coronary artery bypass grafting (CABG) would be performed. A PCI included percutaneous transluminal coronary angioplasty (PTCA) with routine stenting in 70% of the cases and PTCA without stenting in 30%.[10] Patients with a negative stress test result and patients with no coronary artery disease on angiography were treated medically. Figure 1 presents the flow of patients undergoing the diagnostic and therapeutic procedures.

For each treatment strategy, long-term survival was modeled based on the initial patient variables (age, sex, extent of coronary artery disease). Risk reductions by treatment strategies were calculated based on 10-year survival data reported in a systematic review of clinical trials [11]. Risk reductions of stenting versus PTCA were obtained from a meta-analysis on clinical trials comparing stenting and PTCA procedures.[10] Nonfatal myocardial infarction and revascularization procedures during follow-up were modeled to evaluate their effects on costs and quality of life. We allowed differences in revascularization probabilities among the three treatments to persist for 10 years. The risks and risk reductions for the base case, stratified for the extent of coronary artery disease are presented in table 1.[9]

Test characteristics

The sensitivity and specificity estimates for each stress test were obtained from a recent meta-analysis.[7] These are shown in table 2. We assumed that stress SPECT results were non-diagnostic in 1.3%[12] of the cases and stress echocardiography results in 10%,[13] independent of the type of pharmacological stress agent used.

Stress testing involves risks of morbidity and mortality. The rates of non-fatal myocardial infarctions associated with each pharmacological agent were obtained from large safety studies[12,14,15] and were assumed to depend only on the type of stress agent used and not on the type of imaging (Table 3). The mortality rates associated with stress testing are very low and can only be detected in very large studies. In the largest safety study including 73,806 patients a mortality rate of 0.0095% (i.e. less than 1 in 10,000) was found in dipyridamole myocardial perfusion imaging.[14] As no such large studies could be found reporting mortality rates for the other stress agents we assumed that the mortality risk was negligible for each stress agent.

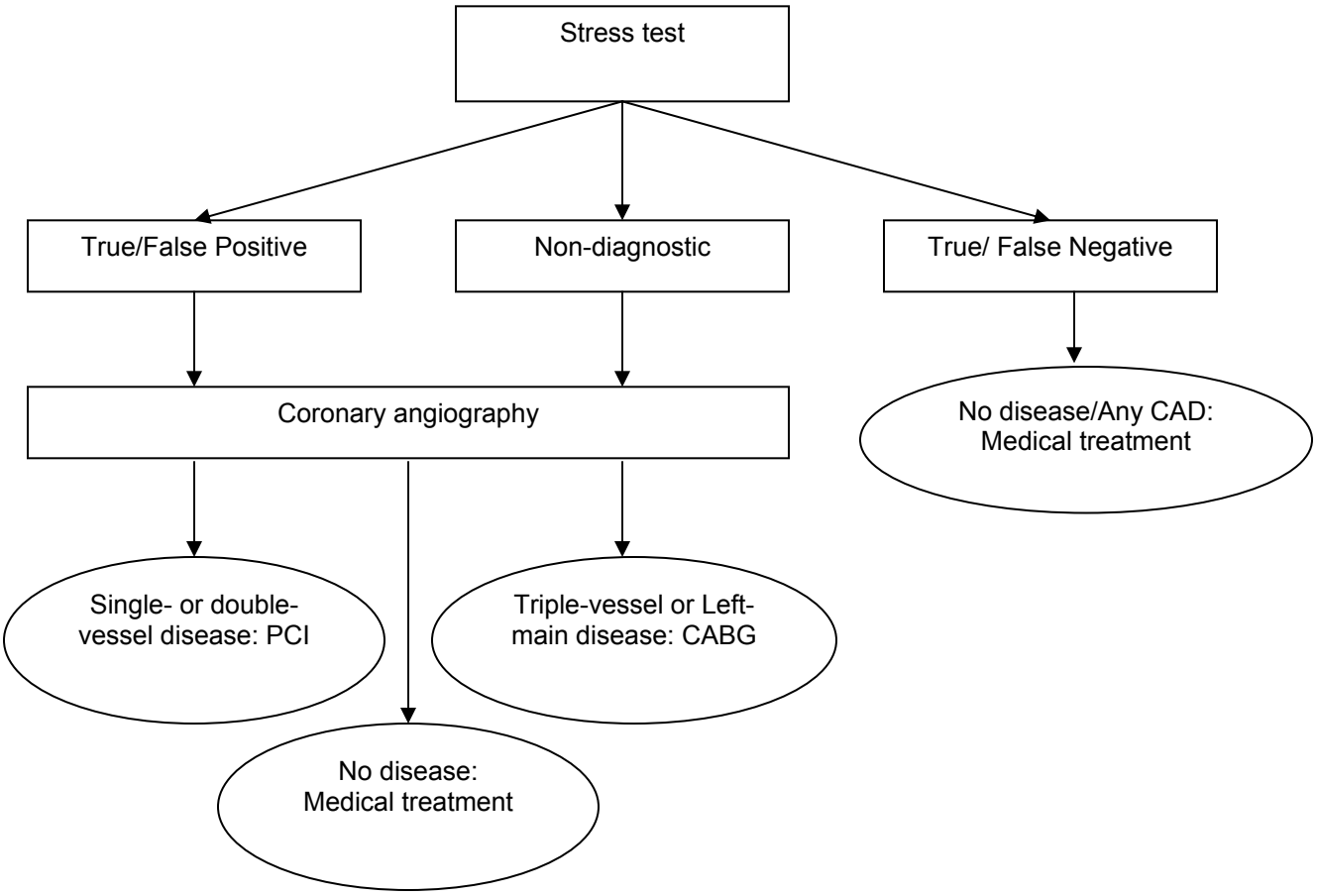


Figure 1: Flow chart of test procedures and subsequent treatment for patients presenting with chest pain. CAD, coronary artery disease; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting

Costs

The cost estimates of the diagnostic tests and treatment procedures were obtained from the Medicare-allowed global reimbursement rates based on the 2002 Current Procedural Terminology code system for outpatient procedures and the 2002 Diagnosis Related Group codes for in-patient procedures, including the technical and professional fees. For stress echocardiography the reimbursement rate was $241 and for stress SPECT it was $599. The Medicare reimbursement for stress testing did not depend on the pharmacological stress agent used. The lifetime costs were calculated by accumulating the costs of diagnostic testing, the costs of possible complications, costs of treatment and possible events, and the annual cost of medical care for chest pain, including follow-up visits, subsequent testing and medical treatment. All costs were adjusted to 2002 U.S. dollars using the medical care component of the Consumer Price Index and were discounted at an annual rate of 3%. The cost estimates are shown in table 4.

Table 1: Selected model variables with ranges for single-/double-, and triple-vessel/left main coronary artery disease for the base case (mainly from reference 9, table 1)

Variable	Single/double vessel disease	Triple vessel/ left main disease
Risk ratio of mortality	2.3 (1.9-2.8)	3.6 (3.1-4.1) / 9.6 (6.1-14.3)
Annual risk of nonfatal MI	0.022 (0.016-0.029)	0.028 (0.021-0.035)
Annual risk of revascularization		
Medical therapy	0.010 (0.001-0.022) /0.042 (0.028-0.056)	0.075 (0.061-0.089)
CABG	n/a	0.018 (0.011-0.025)
PCI	0.025 (0.018-0.033)*	n/a
Risk reduction of mortality (%)		
CABG	n/a	48 (32-64) / 67 (43-87)
PCI	15 (0-49)	n/a
Risk reduction of MI (%)		
CABG	n/a	42 (29-55)
PCI	17 (12-22)	n/a

* Based on an odds ratio of 0.54 for the revascularization rate of PTCA without stent (30%) versus PTCA with stent (70%);[10]

MI, myocardial infarction; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting; PTCA, percutaneous transluminal coronary angioplasty; n/a, not applicable

Table 2: Test characteristics (from reference 7, table 2)

Test	Sensitivity (%)	(95% CI)	Specificity (%)	(95% CI)
Stress echocardiography				
Adenosine	72	(62-79)	91	(88-93)
Dipyridamole	70	(66-74)	93	(90-95)
Dobutamine	80	(77-83)	84	(80-86)
Stress SPECT				
Adenosine	90	(89-92)	75	(70-79)
Dipyridamole	89	(84-93)	65	(54-74)
Dobutamine	82	(77-87)	75	(70-79)

95% CI, 95% confidence interval; SPECT, single-photon emission computed tomography

Table 3: Risk of pharmacological stress testing (from references 11, 13, and 14)

Pharmacological stress agent	Risk of non-fatal MI (%)	(95% CI)
Adenosine	0.011	(0.008-0.027)
Dipyridamole	0.018	(0.000-0.032)
Dobutamine	0.025	(0.000-0.073)

MI, myocardial infarction; 95% CI, 95% confidence interval

Table 4: Cost estimates in the model based on the 2002 Medicare allowed global reimbursement rates

Variable	Cost ($)	Range
Stress echocardiography	241	170 - 315
Stress SPECT	599	420 - 780
Coronary angiography:		
Out-patient	2,395	1,675 – 3,115
In-patient	5,030	3,521 – 6,540
PCI:		
PTCA	9,945	6,960 – 1,295
PTCA with stent	11,600	8,120 – 15,080
CABG	23,052	16,135 – 29,970
Myocardial infarction	6,690	4,683 – 8,697
Annual cost		
No angina	200	140 - 260
Mild angina	2,000	1,400 – 2,600
Severe angina	4,375	3,063 – 5,688

Stress SPECT, stress single-photon-emission computed tomography; PCI, percutaneous coronary intervention; PTCA, percutaneous transluminal coronary angioplasty; CABG, coronary artery bypass grafting

Quality of life

We estimated the quality-adjusted life expectancy by modeling annual transitions between 4 different health states: no chest pain, mild chest pain, severe chest pain, and death. Quality of life without chest pain was valued as 0.87, mild chest pain was valued as 0.82, and severe chest pain as 0.67 compared with perfect health.[9] These health values are based on a survey of patients with chronic stable angina by using the standard gamble method, where 0 represents death and 1 represents perfect health [16]. The transitions from one chest pain state to another as a result of treatment were based on the CASS quality of life study.[17] We assumed that a PCI procedure was 85% as effective as CABG for immediate change in chest pain severity and that the effects were similar during follow-up.[18] Furthermore, we assumed that the differences in effects among the three treatments persisted for 10 subsequent years. To incorporate the inconvenience of treatment or experiencing an adverse event, we subtracted life days from the quality-adjusted life expectancy. One day was subtracted for each angiogram performed, 2.5 days for each PCI, 30 days for each CABG, and 10 days each time a non-fatal MI occurred.

Sensitivity analysis

To evaluate the effect of varying the baseline estimates over plausible ranges on our results we performed one- and two-way sensitivity analyses. Ranges were based on the 95% confidence intervals of the baseline estimates. For the cost-estimates, we used a range of 30% above and below the baseline estimate. Furthermore, we performed a sensitivity analysis including only the pharmacologic stress tests recommended by the American College of Cardiology and the American Heart Association (ACC/AHA), i.e. dobutamine echocardiography and adenosine or dipyridamole SPECT.

Table 5: Results for a cohort of 55-year-old men with non-specific chest pain, atypical, and typical angina

55-year-old men	Cost ($)	Effect (QALYs)	ICER ($/QALY)
Non-specific chest pain			
No imaging	24,500	14.967	.
Dipyridamole Echo	26,600	15.001	60,483
Adenosine Echo	26,700	15.002	ED
Dobutamine Echo	27,000	15.005	99,186
Dobutamine SPECT	27,400	15.005	ED
Adenosine SPECT	27,500	15.009	150,555
Dipyridamole SPECT	27,800	15.008	D
Coronary angiography	29,000	15.100	1,165,159
Atypical angina			
No imaging	26,250	12.670	.
Dipyridamole Echo	32,300	12.860	31,767
Adenosine Echo	32,400	12.865	ED
Dobutamine Echo	33,100	12.884	33,051
Dobutamine SPECT	33,500	12.885	ED
Adenosine SPECT	34,100	12.905	ED
Dipyridamole SPECT	34,100	12.902	D
Coronary angiography	34,900	12.929	40,401
Typical Angina			
No imaging	27,600	11.221	.
Dipyridamole Echo	35,900	11.543	ED
Adenosine Echo	36,100	11.551	ED
Dobutamine Echo	36,900	11.582	ED
Dobutamine SPECT	37,300	11.584	ED
Dipyridamole SPECT	38,100	11.614	ED
Adenosine SPECT	38,200	11.619	ED
Coronary angiography	38,800	11.662	25,431

ED, extended dominated; D, dominated; Echo, echocardiography; SPECT, single-photon emission computed tomography; ICER, incremental cost-effectiveness ratio; QALY(s), quality-adjusted life year(s)

Results

Base case

The results of the base case (55-year-old men with atypical angina) are presented in the middle section of table 5. The incremental costs associated with each test strategy relative to the increment in effectiveness determine which test strategy is optimal – contingent on the societal willingness to pay for a quality-adjusted life year (QALY). Only dipyridamole SPECT was clearly dominated by other test strategies in the base case, meaning that it was more expensive and less effective than other strategies. Adenosine echocardiography and dobutamine or adenosine SPECT were weakly dominated by the other strategies, which means that the incremental cost-effectiveness ratio was higher than for a more effective strategy.

Cost-effective non-invasive strategies for the base case were dipyridamole and dobutamine echocardiography with incremental cost-effectiveness ratios of respectively $31,800 and $33,100 per QALY. Routine coronary angiography was more effective than the non-invasive imaging strategies in the base case, but at a higher cost ($40,400 per QALY).

Other cohorts

In 65-year-old women with atypical angina, the ranking of the tests was the same as in men, but at higher incremental cost-effectiveness ratios (Table 6). Furthermore, adenosine SPECT was cost-effective in this cohort, if society would be willing to pay at least $61,500 per QALY compared with dobutamine echocardiography.

In patients with non-specific chest pain the ranking of the imaging strategies was the same as in the base case, but all imaging strategies cost more than $60,000 per QALY in men (Table 5) and more than $77,000 in women (Table 6).

In patients with typical angina routine coronary angiography was the optimal strategy at a cost of $ 25,400 per QALY in men (Table 5). All non-invasive strategies were ruled out because they had higher incremental cost-effectiveness ratios than angiography. In women with typical angina, however, dipyridamole and dobutamine echocardiography were cost-effective strategies, both costing less than $ 30,000 per QALY, whereas angiography cost $31,200 per QALY (Table 6).

Sensitivity analysis

Prevalence

The incremental cost-effectiveness ratios were sensitive to the prevalence of disease. The optimal test strategies for the base case are plotted in figure 2 as a function of the prevalence for several willingness-to-pay thresholds. At a very low prevalence of coronary artery disease (<0.10) or if society is willing to pay less than $30,000 per QALY saved, managing the patient medically without diagnostic imaging or revascularization was preferred. For a higher willingness-to-pay threshold, for example $50,000 per QALY, dipyridamole echocardiography was the optimal imaging strategy at a low to intermediate pre-test probability of disease, dobutamine echocardiography at an intermediate pre-test probability, and coronary angiography was optimal at a high prevalence. If society is willing-to-pay more than $53,000 per QALY, adenosine SPECT becomes the optimal strategy at an intermediate pre-test probability of disease.

Test characteristics

If the baseline estimates of sensitivity and specificity of the pharmacological SPECT tests were varied over their 95% confidence intervals, there were no changes in the rankings of the incremental cost-effectiveness ratios.

Varying the sensitivity and specificity of the pharmacological echocardiography tests, however, resulted in small changes of the rankings of the echocardiography tests. Dobutamine echocardiography was always cost-effective, except if the sensitivity of adenosine echocardiography exceeded 79%. In addition to dobutamine and dipyridamole echocardiography, adenosine echocardiography was cost-effective if the sensitivity or specificity of dobutamine echocardiography decreased. Furthermore, adenosine echocardiography replaced dipyridamole echocardiography if the sensitivity or specificity of adenosine echocardiography increased or if the sensitivity or specificity of dipyridamole echocardiography decreased.

Changes in the rate of non-diagnostic test results over a plausible range associated with the type of imaging did not influence the ranking of the tests.

Cost

Varying the cost of stress echocardiography over a range of 30% above and below the baseline estimates had no effect on the cost-effectiveness or ranking of the tests.

In the base case, i.e. 55-year-old men with atypical angina, adenosine SPECT was only cost-effective if its cost was lower than $420 or if the cost of angiography was higher than $5,000. For example, if all patients need to be hospitalized for a diagnostic coronary angiogram, which has a cost of $5,030 per hospitalization, adenosine SPECT was a potentially cost-effective alternative with an incremental cost-effectiveness ratio of $52,600 per QALY compared with dobutamine echocardiography.

Selection of tests recommended by the ACC/AHA

In a sensitivity analysis we also compared a selection of tests in the base case, namely dobutamine echocardiography, adenosine SPECT, and dipyridamole SPECT, which are recommended by the ACC/AHA, compared with the no-imaging strategy and with routine coronary angiography. Dobutamine echocardiography had an incremental cost-effectiveness ratio of $31,900 per QALY compared with the no-imaging strategy. Routine coronary angiography cost $40,400 per QALY compared with dobutamine echocardiography. The dipyridamole SPECT strategy was more expensive and less effective than adenosine SPECT. Adenosine SPECT was more costly and more effective than dobutamine echocardiography, but its incremental cost-effectiveness ratio was higher than the more effective coronary angiography strategy.

Table 6: Results for a cohort of 65-year-old women with non-specific chest pain, atypical, and typical angina

65-year-old women	Cost ($)	Effect (QALYs)	ICER ($/QALY)
Non-specific chest pain			
No imaging	20,600	12.780	.
Dipyridamole Echo	22,200	12.800	77,194
Adenosine Echo	22,200	12.801	ED
Dobutamine Echo	22,500	12.802	152,623
Dobutamine SPECT	22,900	12.802	ED
Adenosine SPECT	23,000	12.805	226,809
Dipyridamole SPECT	23,300	12.804	D
Coronary angiography	24,600	12.804	D
Atypical angina			
No imaging	21,000	11.305	.
Dipyridamole Echo	25,500	11.427	37,067
Adenosine Echo	25,600	11.430	ED
Dobutamine Echo	26,100	11.442	42,280
Dobutamine SPECT	26,600	11.442	ED
Adenosine SPECT	27,000	11.455	61,512
Dipyridamole SPECT	27,100	11.453	D
Coronary angiography	28,100	11.470	79,807
Typical Angina			
No imaging	22,800	9.612	.
Dipyridamole Echo	30,400	9.877	28,890
Adenosine Echo	30,600	9.883	ED
Dobutamine Echo	31,400	9.909	29,172
Dobutamine SPECT	31,800	9.910	ED
Dipyridamole SPECT	32,500	9.935	ED
Adenosine SPECT	32,600	9.939	ED
Coronary angiography	33,400	9.973	31,225

ED, extended dominated; D, dominated; Echo, echocardiography; SPECT, single-photon emission computed tomography; ICER, incremental cost-effectiveness ratio; QALY(s), quality-adjusted life year(s)

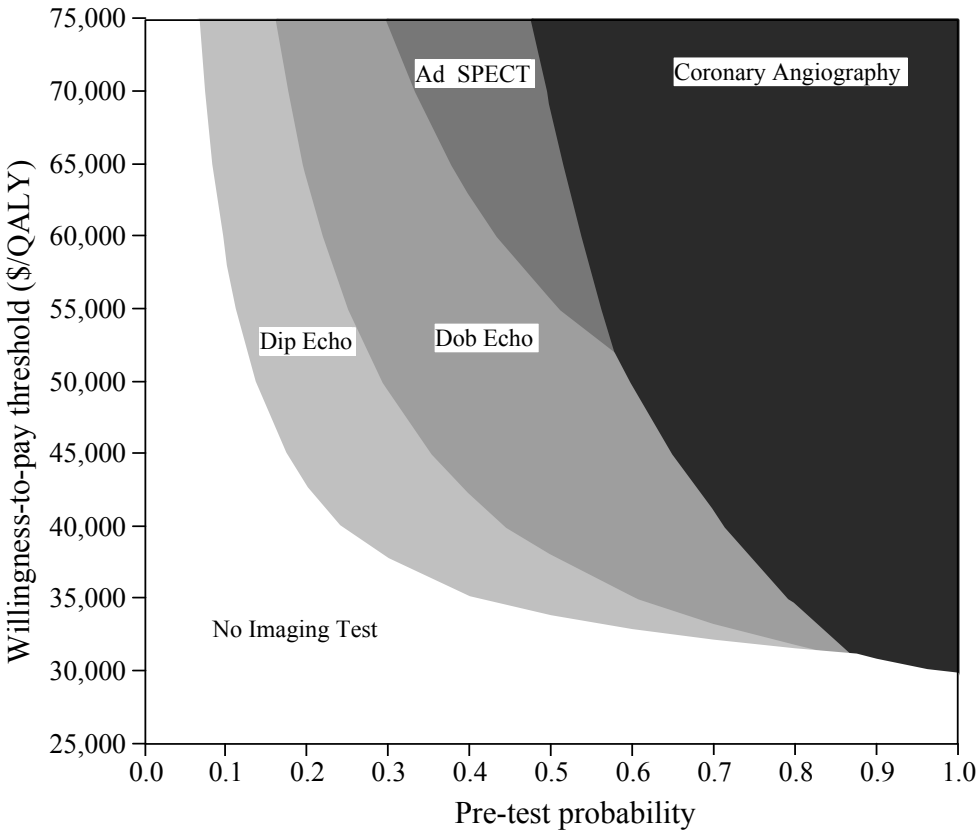


Figure 2: Sensitivity analysis of the pre-test probability of coronary artery disease and the willingness-to-pay threshold. Shaded regions indicate the optimal approach for the base case.
QALY, quality-adjusted life-year; Dip echo, dipyridamole echocardiography; Dob echo, dobutamine echocardiography; Ad SPECT, adenosine single-photon-emission computed tomography

Discussion

We developed a decision-analytic model to integrate the best available evidence on the costs and health outcomes associated with various pharmacological stress tests and their associated follow-up treatments. We found that pharmacological stress echocardiography is cost-effective in both men and women at low to intermediate risk of coronary artery disease, if society is willing to pay at least $30,000 per QALY saved. In particular clinical situations pharmacological SPECT may be cost-effective if society is willing to pay more than $50,000 per QALY saved. Routine coronary angiography was cost-effective in both men and women with typical angina.

In addition, we found that the choice of the most cost-effective imaging strategy was primarily determined by the pre-test probability of disease and the threshold amount that society is willing to pay per quality-adjusted life-year saved. The rankings of the different tests corresponded with

the sensitivity of each test: the higher the sensitivity of the test, the more cost-effective it is to utilize the test at a higher prevalence of disease. Tests that are more specific are associated with more favorable cost-effectiveness ratios at lower pre-test probabilities of disease. In patients with non-specific chest pain, for example, who have a low pre-test probability of disease, stress echocardiography (more specific) resulted in more favorable cost-effectiveness ratios than stress SPECT (more sensitive). In general we found that the lifetime outcomes associated with very sensitive but less specific test strategies are more expensive, but also more effective in a high risk population because more patients will be tested positive, diseased or not, and consequently referred to angiography, which will provide the final diagnosis.

Comparison to previous studies

In a previous cost-effectiveness study comparing exercise echocardiography and exercise SPECT, our group found that exercise SPECT had less favorable incremental cost-effectiveness ratios than exercise echocardiography.[9] Kim et al. found similar results in a cost-effectiveness study on exercise testing in women.[19] Furthermore, Garber and coworkers found that the incremental cost-effectiveness ratio of SPECT ranged between $64,000 to nearly $150,000 per QALY gained compared with stress echocardiography.[20] In the latter study no distinction was made between the types of pharmacological and exercise stress. Our results on pharmacological stress SPECT are consistent with the results of these studies. Specifically, dobutamine has similar physiological effects compared with exercise and, meta-analyses suggest similar sensitivity of exercise in conjunction with echocardiography and SPECT, but lower specificity with SPECT than with echocardiography.[7,21,22] This explains why the dobutamine SPECT strategy was never optimal.

In contrast, a recent cost-effectiveness study concluded that pharmacological stress SPECT was more cost-effective than pharmacological stress echocardiography.[23] However, in that study the authors assumed that the sensitivity and specificity estimates of SPECT and echocardiography were equal. Meta-analysis of the published data[7] suggests that stress echocardiography is a more specific test, whereas stress SPECT is a more sensitive test. Our findings show that pharmacological echocardiography is generally cost-effective for the evaluation of patients with non-specific and atypical angina and that pharmacological SPECT may be cost-effective for particular clinical settings in patients with atypical angina.

Selection of pharmacological stress agent

In the United States, dobutamine is the recommended stress agent that is primarily used in combination with echocardiography, whereas in Europe both dobutamine and dipyridamole echocardiography are commonly used.[2,6,24] Our study suggests that both pharmacological agents can be cost-effective, but dipyridamole is associated with more favorable cost-effectiveness ratios in patients at lower pre-test probability compared with dobutamine, likely related to the higher specificity of dipyridamole and higher sensitivity of dobutamine.

Adenosine echocardiography and dobutamine and dipyridamole SPECT did not emerge as optimal imaging strategies in the baseline analysis compared with the other strategies. Because the mechanism of action of the vasodilators dipyridamole and adenosine is similar, we expected that dipyridamole could be replaced by adenosine or vice versa with little additional cost or change in effectiveness. We did see this effect in echocardiography: sensitivity analyses showed that small

changes in the cost of the test or the test characteristics determined whether the incremental cost-effectiveness ratio of adenosine or dipyridamole echocardiography was most favorable. In SPECT testing, however, we found that the cost-effectiveness of dipyridamole SPECT was never optimal. This can be explained by the lower point estimate for specificity of dipyridamole SPECT as compared with adenosine SPECT.[7] Therefore, our results suggest that adenosine should be used rather than dipyridamole in combination with SPECT testing. However, this result should be considered with caution. The estimates of sensitivity and specificity of the diagnostic tests in our model were obtained from a meta-analysis.[7] In this meta-analysis only 15 studies were available on adenosine testing compared with 54 on dobutamine and 41 on dipyridamole testing. Publication bias and spectrum bias may have influenced the estimates of the test characteristics, especially for adenosine testing. These biases may occur when only positive results of a new test are published using highly selected patients. Subsequent publications of studies that use less selected patients may report lower values of test performance. If the reported sensitivity and specificity of adenosine SPECT decreases in further studies to be more similar to dipyridamole SPECT, then the cost-effectiveness ratios associated with adenosine SPECT may become less favorable.

If we only compared the tests recommended by the ACC/AHA, i.e. dobutamine echocardiography, adenosine SPECT, and dipyridamole SPECT, we found that dobutamine echocardiography was a cost-effective non-invasive imaging strategy. Pharmacological stress SPECT was not preferred as initial imaging strategy because of its high cost.

Potential implications for clinical management

The present study has the following implications with respect to the evaluation of coronary artery disease in patients presenting with chest pain. In patients with non-specific chest pain with a pre-test probability of 10-30%, dipyridamole or adenosine echocardiography was the optimal imaging strategy for the selection of candidates for coronary angiography and subsequent treatment. In patients presenting with atypical angina with a pre-test probability of 30-80%, dobutamine echocardiography was generally the optimal imaging strategy. If society is willing to pay more than $50,000 per QALY saved and in specific clinical situations, adenosine SPECT or routine coronary angiography may be cost-effective in patients with atypical angina. For example, if a diagnostic coronary angiography is considered which requires hospitalization for a specific patient with atypical angina, adenosine SPECT may be a cost-effective alternative. Patients with typical angina, with a pre-test probability of 80-100%, should routinely be referred to coronary angiography.

Limitations

A limitation of our study is that we only compared stress tests that use pharmacological agents to induce stress, which is applicable for patients with limited exercise capacity. However, we assumed that the prevalence of coronary artery disease and the risks and benefits of treatment for our target population was the same as for all patients presenting with chest pain. Because patients with limited exercise capacity may have more comorbidity, we may have overestimated the effect of treatment in these patients. However, this overestimation would be equal for all test strategies and would therefore not impact the relative outcomes.

Another limitation is that our study is a synthesis of the published data on the risks, benefits, and costs of different diagnostic and therapeutic options in patients with coronary artery disease. Due to the rapid changes and improvements in imaging technology and coronary interventions, such as the use of contrast agents and harmonic imaging in echocardiography, new isotopes and attenuation correction in SPECT, and the use of new drug-eluting stents in PCI, some of our data may not reflect current practice in highly specialized care environments. To test the robustness of our conclusions, we examined, where possible, the effect of varying our estimates over plausible ranges on the outcomes of our study. We found that these sensitivity analyses did not change the results substantially, which implies that the conclusions remained the same.

Conclusions

In conclusion, pharmacological stress echocardiography is cost-effective for the diagnosis of coronary artery disease in middle-aged men and women at low and intermediate risk of coronary artery disease if society is willing to pay more than $30,000 per QALY saved. Adenosine SPECT can be cost-effective for patients at intermediate risk of coronary artery disease if society is willing to pay more than $50,000 per QALY saved in particular clinical situations. Routine coronary angiography is the optimal test strategy for patients with typical angina. Careful pretest risk determination is essential to optimal use of cost-effective diagnostic strategies.

References

1. Anagnostopoulos C, Henein MY, Underwood SR. Non-invasive investigations. Br Med Bull. 2001;59:29-44.
2. Ali Raza J, Reeves WC, Movahed A. Pharmacological stress agents for evaluation of ischemic heart disease. Int J Cardiol. 2001;81:157-167.
3. Gibbons RJ, Chatterjee K, Daley J, et al. ACC/AHA/ACP-ASIM 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 (Committee on Management of Patients With Chronic Stable Angina). J Am Coll Cardiol. 1999;33:2092-2197.
4. Matte GG, Barnes DC, Abrams DN. Pharmacologic interventions in nuclear medicine assessment of cardiac perfusion. J Pharm Pharm Sci. 2001;4:255-262.
5. Elhendy A, Bax JJ, Poldermans D. Dobutamine stress myocardial perfusion imaging in coronary artery disease. J Nucl Med. 2002;43:1634-1646.
6. Imran MB, Palinkas A, Picano E. Head-to-head comparison of dipyridamole echocardiography and stress perfusion scintigraphy for the detection of coronary artery disease: a meta-analysis. Comparison between stress echo and scintigraphy. Int J Cardiovasc Imaging. 2003;19:23-28.
7. Kim C, Kwok YS, Heagerty P, Redberg R. Pharmacologic stress testing for coronary disease diagnosis: A meta-analysis. Am Heart J. 2001;142:934-944.
8. Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). Circulation. 1981;64:360-367.
9. 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.
10. Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C. Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review. Health Technol Assess. 2000;4:1-153.
11. Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. Lancet. 1994;344:563-570.
12. Cerqueira MD, Verani MS, Schwaiger M, Heo J, Iskandrian AS. Safety profile of adenosine stress perfusion imaging: results from the Adenoscan Multicenter Trial Registry. J Am Coll Cardiol. 1994;23:384-389.
13. Geleijnse ML, Fioretti PM, Roelandt JR. Methodology, feasibility, safety and diagnostic accuracy of dobutamine stress echocardiography. J Am Coll Cardiol. 1997;30:595-606.
14. Lette J, Tatum JL, Fraser S, et al. Safety of dipyridamole testing in 73,806 patients: the Multicenter Dipyridamole Safety Study. J Nucl Cardiol. 1995;2:3-17.
15. Mathias W, Jr., Arruda A, Santos FC, et al. Safety of dobutamine-atropine stress echocardiography: A prospective experience of 4,033 consecutive studies. J Am Soc Echocardiogr. 1999;12:785-791.
16. Nease RF, Jr., Kneeland T, O'Connor GT, et al. Variation in patient utilities for outcomes of the management of chronic stable angina. Implications for clinical practice guidelines. Ischemic Heart Disease Patient Outcomes Research Team. JAMA. 1995;273:1185-1190.
17. Coronary artery surgery study (CASS): a randomized trial of coronary artery bypass surgery. Quality of life in patients randomly assigned to treatment groups. Circulation. 1983;68:951-960.
18. Pocock SJ, Henderson RA, Rickards AF, et al. Meta-analysis of randomised trials comparing coronary angioplasty with bypass surgery. Lancet. 1995;346:1184-1189.
19. Kim C, Kwok YS, Saha S, Redberg RF. Diagnosis of suspected coronary artery disease in women: a cost-effectiveness analysis. Am Heart J. 1999;137:1019-1027.
20. Garber AM, Solomon NA. Cost-effectiveness of alternative test strategies for the diagnosis of coronary artery disease. Ann Intern Med. 1999;130:719-728.

21. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA. 1998;280:913-920.
22. Kymes SM, Bruns DE, Shaw LJ, Gillespie KN, Fletcher JW. Anatomy of a meta-analysis: a critical review of "exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance". J Nucl Cardiol. 2000;7:599-615.
23. Lee DS, Jang MJ, Cheon GJ, Chung JK, Lee MC. Comparison of the cost-effectiveness of stress myocardial SPECT and stress echocardiography in suspected coronary artery disease considering the prognostic value of false-negative results. J Nucl Cardiol. 2002;9:515-522.
24. Williams SV, Fihn SD, Gibbons RJ. Guidelines for the management of patients with chronic stable angina: diagnosis and risk stratification. Ann Intern Med. 2001;135:530-547.

5

Cost-effectiveness of electron-beam computed tomography in patients with chest pain

Abstract

Background: Many studies have been performed on the diagnostic value of electron-beam computed tomography (EBCT) in patients suspected of coronary artery disease, but little is known about its cost-effectiveness.

Objectives: The aim of this study was to evaluate the cost-effectiveness of EBCT compared with exercise echocardiography and exercise single-photon-emission computed tomography (SPECT) in patients presenting with chest pain to identify those who should undergo coronary angiography for the diagnosis of coronary artery disease.

Methods: We used a Markov decision model to estimate the lifetime costs and effects for different test strategies and clinical follow up from the health care perspective. As a base case we evaluated a cohort of 55-year-old men with atypical chest pain.

Results: In the base-case, the EBCT strategy was slightly more effective than exercise echocardiography and exercise SPECT at a relatively high cost and therefore extended dominated by the more effective coronary angiography. In cohorts of 65-year-old women with atypical chest pain the incremental cost-effectiveness ratio of EBCT was $66,600 per quality-adjusted life-year (QALY) saved compared with exercise echocardiography. Sensitivity analyses showed that EBCT may be a cost-effective strategy for men with a 40%-65% pre-test probability of coronary artery disease at a cost of at least $47,500 per QALY. If society is willing to pay $50,000 per QALY or less, the strategy with optimal cost-effectiveness for patients with non-specific chest pain was the no imaging test strategy, for patients with atypical chest pain it was exercise echocardiography, and for patients with typical chest pain the optimal strategy was routine coronary angiography.

Conclusions: Using EBCT to decide which patients with chest pain should undergo coronary angiography for the diagnosis of coronary artery disease is, in general, not cost-effective. In a selected group of patients with an intermediate pre-test probability of disease EBCT provides a small health benefit at a relatively high cost.

Introduction

Coronary calcification on electron-beam computed tomography (EBCT) is commonly used to identify patients with coronary artery disease who are at high risk of developing a myocardial infarction. The same technology is also used as a diagnostic test in patients with chest pain suspected of having coronary artery disease.[1-9]

The Society for Atherosclerosis Imaging (SAI) stated that there is evidence and/or general agreement that EBCT is useful and effective as an initial diagnostic test in ambulatory adults <= 65 years of age with atypical chest pain symptoms, in the absence of established cardiovascular disease, and as a supplementary diagnostic test in similar patients with indeterminate stress test results.[10] A recent meta-analysis showed that sensitivity and specificity of EBCT in symptomatic patients are comparable with other diagnostic tests.[11] Similar meta-analytic results were found by the American College of Cardiology and the American Heart Association.[12] They concluded that EBCT is not superior to other currently available diagnostic procedures but, given its relatively low cost and ease of implementation, recommended more research into its potential cost-effectiveness as a diagnostic tool. So far, however, only the initial costs and effects of EBCT have been studied compared with other test strategies.[13,14] No information is available on the long-term cost-effectiveness of EBCT.

The purpose of the current study was to evaluate the long-term cost-effectiveness of EBCT for the diagnosis of coronary artery disease compared with other imaging strategies in patients presenting with chest pain.

Methods

Target population

We analyzed several cohorts of patients suspected of coronary artery disease, defined by age, gender, and type of chest pain. Chest pain was categorized into nonspecific, atypical, or typical chest pain as in the Coronary Artery Surgery Study (CASS). Nonspecific, atypical, or typical chest pain was present if respectively one, two, or three of the following criteria were met: the chest pain is located substernally, follows physical exertion, and is relieved by nitroglycerin. The prevalence of coronary artery disease in patients with nonspecific, atypical, and typical chest pain was respectively 18%, 71%, and 95% in 55-year-old men and 10%, 48%, and 81% in 65-year-old women.[15] As a base case, we evaluated a cohort of 55-year-old men with atypical chest pain.

Markov decision model

We revised a previously developed Markov decision model[16] to compare the cost-effectiveness of five diagnostic imaging strategies: (1) EBCT, (2) exercise echocardiography, (3) exercise SPECT, (4) coronary angiography, and (5) no imaging. If the test result of EBCT, exercise echocardiography or exercise SPECT was negative, no additional test was performed and the patient was treated medically. If the non-invasive test result was positive or non-diagnostic, it was followed by coronary angiography to reach the final diagnosis. In accordance with the guidelines of the Society of Atherosclerosis Imaging[10] we also studied the cost-effectiveness of EBCT as a supplementary

test if exercise echocardiography or SPECT were non-diagnostic, followed by angiography if EBCT was positive or non-diagnostic.

Coronary angiography was considered the reference standard with 100% sensitivity and 100% specificity and was assumed to be interpretable in all cases. Significant coronary artery disease was defined as at least 50% diameter stenosis in at least one coronary artery. The result of coronary angiography was quantified as no, single-, double-, or triple-vessel disease, or left main coronary artery disease. The prevalence of the underlying coronary artery disease status was obtained from the Coronary Artery Surgery Study registry.[15] We assumed that patients diagnosed with triple-vessel or left main coronary artery disease underwent coronary artery bypass graft surgery (CABG) and patients with single- or double-vessel disease underwent a percutaneous coronary intervention (PCI), consisting of percutaneous transluminal coronary angioplasty (PTCA) with additional stent implantation in 70% of cases.[17] Patients with no or non-significant disease were treated medically. We also modeled possible events such as nonfatal myocardial infarction and revascularization procedures occurring after the initial treatment decision to evaluate their effects on cost and quality of life.

Test characteristics

Sensitivity and specificity estimates of EBCT, exercise echocardiography and exercise SPECT for diagnosing patients suspected of coronary artery disease were obtained from meta-analyses.[11,18] EBCT was considered positive for significant coronary artery disease if the coronary calcium score was greater than zero.[1] Presence of new or worsening regional wall motion abnormalities after exercise was defined as a positive exercise echocardiogram. Using SPECT, a perfusion defect at rest or after exercise indicated significant coronary artery disease.

Based on the study by Rumberger et al, we assumed that EBCT was non-diagnostic in 2% of cases and has no associated morbidity or mortality.[13] The test characteristics are shown in Table 1.

Table 1: Test characteristics

	EBCT	Ex Echo	Ex SPECT
Sensitivity (%, range)	92 (91-94)	85 (83-87)	87 (86-88)
Specificity (%, range)	51 (48-55)	77 (74-80)	64 (60-68)
Mortality (%, range)	0 (0-0.005)	0.005 (0.002-0.008)	0.005 (0.002-0.008)
Morbidity (%, range)	0 (0-0.05)	0.05 (0.02-0.08)	0.05 (0.02-0.08)
Non-diagnostic (%, range)	2 (1-3)	10 (5-15)	2 (1-3)

EBCT, electron-beam computed tomography; Ex Echo, exercise echocardiography; Ex SPECT, exercise single-photon-emission computed tomography

Cost-effectiveness calculations

Outcomes of the Markov decision model were defined in terms of lifetime costs and quality adjusted life years (QALYs). The incremental cost-effectiveness ratio (ICER) was calculated for each diagnostic strategy, which was defined as the difference in costs divided by the difference in

QALYs compared with the next less expensive strategy. Strategies were dominated if they were more costly and less effective compared with another strategy or extended dominated if the incremental cost-effectiveness ratio was greater than a more effective strategy. The analysis was performed from the perspective of the health care system.

Table 2: Estimates of cost and utility

Variable	Estimate	Range
Cost ($)		
EBCT	271	190 - 352
Exercise echocardiography	241	169 - 313
Exercise SPECT	599	419 - 779
Coronary angiography:		
Out-patient	2,395	1,675 - 3,115
In-patient	5,030	3,521 - 6,540
PCI:		
PTCA	9,945	6,960 - 1,295
PTCA with stent	11,600	8,120 - 15,080
CABG	23,052	16,135 - 29,970
Myocardial infarction	6,690	4,683 - 8,697
Annual cost		
No angina	200	140 - 260
Mild angina	2,000	1,400 - 2,600
Severe angina	4,375	3,063 - 5,688
Utility		
No chest pain symptoms	0.87	0.77 - 1.00
Mild chest pain symptoms	0.81	0.68 - 1.00
Severe chest pain symptoms	0.67	0.40 - 0.98

EBCT, electron-beam computed tomography; exercise SPECT, exercise single-photon-emission computed tomography; PCI, percutaneous coronary intervention; PTCA, percutaneous transluminal coronary angioplasty; CABG, coronary artery bypass grafting

Cost calculation

The cost estimates of the diagnostic tests and treatment procedures were obtained from the Medicare reimbursement system (Table 2). For outpatient procedures we used the 2002 Current Procedural Terminology (CPT) coding system including the technical and professional fees. For inpatient procedures, we used the 2002 Diagnosis Related Group (DRG) coding system plus the professional fees according to the CPT system. We assumed that the Medicare reimbursement for EBCT would be equal to the reimbursement for a CT of the chest. The lifetime costs were calculated by summing the cost of the tests, the cost of possible test complications, the cost of treatment following the test results, the cost of possible events, and the annual cost of medical care for chest pain, including medications, subsequent tests and outpatient visits. All costs were

adjusted to 2002 U.S. dollars using the medical care component of the Consumer Price Index and were discounted at an annual rate of 3%.

Effect calculation

Quality of life without chest pain was valued at 87% of the quality of life in full health, with mild chest pain at 81%, and with severe chest pain at 67%.[16] These health values are based on a survey of patients with chronic stable angina by using the standard gamble method, where 0 represents death and 1 represents perfect health.[19] The change in severity of chest pain due to treatment was estimated using the CASS quality of life study.[20] This study reported the proportion of patients for each chest pain severity level over time, stratified by underlying coronary artery disease status and initial treatment. We assumed that a PCI procedure was 85% as effective as CABG for immediate change in chest pain severity and that the effects on symptoms were similar thereafter.[21] The differences among the three treatment strategies were modeled to persist for 10 subsequent years.

Mortality risk ratios, mortality risk reductions, and the risk for nonfatal myocardial infarction or revascularization procedures during follow-up depended on the extent of coronary artery disease and type of initial treatment (Table 3).[20,22-25]

Table 3: Selected model variables for single-/double-, and triple-vessel/left main coronary artery disease for the base case[16]

Variable	Single/double vessel disease	Triple vessel/ left main disease
Risk Ratio of mortality	2.3 (1.9-2.8)	3.6 (3.1-4.1) / 9.6 (6.1-14.3)
Annual Risk of non-fatal MI	0.022 (0.016-0.029)	0.028 (0.021-0.035)
Risk Reduction of mortality (%)		
CABG	n/a	48 (32-64) / 67 (43-87)
PCI	15 (0-49)	n/a
Risk Reduction of MI (%)		
CABG	n/a	42 (29-55)
PCI	17 (12-22)	n/a
Annual Risk of revascularization		
Medical therapy	0.010 (0.001-0.022) /0.042 (0.028-0.056)	0.075 (0.061-0.089)
CABG	n/a	0.018 (0.011-0.025)
PCI	0.025 (0.018-0.033)*	n/a

Based on an odds ratio of 0.54 for the revascularization rate of PTCA without stent (30%) versus PTCA with stent (70%)[17];

MI, myocardial infarction; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting; PTCA, percutaneous transluminal coronary angioplasty; N/a, not applicable

Sensitivity analysis

We performed sensitivity analyses on all of the variables in our model to assess the effect of varying baseline estimates within plausible ranges on our results. All the costs were varied using a

range of +/- 30% of the base-case estimates. The ranges of the test characteristics, risk ratios and annual risks were based on 95% confidence intervals. The ranges of the utilities were obtained from the 25th and 75th percentiles of the data.

Results

Cost-effectiveness analysis

Table 4 shows the quality-adjusted life expectancy, lifetime costs, and incremental cost-effectiveness ratios of the test strategies for 55-year-old men presenting with nonspecific, atypical, or typical chest pain. Table 5 shows the results for three cohorts of 65-year-old women. The results of the base case, 55-year-old men with atypical chest pain, are shown in the middle section of table 4.

In the base case, the EBCT strategy provided a slightly higher effectiveness compared with exercise echocardiography and SPECT at a relatively high cost. The incremental cost-effectiveness ratios (ICER) of EBCT, and of exercise SPECT, were higher than the more effective angiography strategy, meaning that the EBCT and exercise SPECT strategies were extended dominated. Exercise echocardiography and routine coronary angiography were both cost-effective strategies in the base case, if society is willing to pay $30,000 to 40,000 per QALY.

In 65-year-old women with atypical chest pain the incremental cost-effectiveness ratio of EBCT was $66,600 per QALY compared with exercise echocardiography, which had an ICER of $37,900 per QALY compared with no imaging. In both men and women with nonspecific chest pain all imaging strategies had ICERs of more than $50,000 per QALY. In patients with typical chest pain, however, routine coronary angiography was the optimal strategy in 55-year-old men, whereas both exercise echocardiography and angiography were cost-effective strategies in 65-year old women.

In the secondary analysis in which EBCT was implemented as a supplementary test if exercise echocardiography or SPECT was non-diagnostic, we found that the combination strategies were less costly than the single test strategies, but also less effective, and therefore ruled out by extended dominance.

Sensitivity analysis

Test characteristics

We performed sensitivity analyses on the test characteristics of EBCT to assess the effect of varying baseline estimates over their 95% confidence intervals on our results. EBCT was cost-effective compared with exercise echocardiography if the sensitivity was equal to or more than 93% (ICER $40,100 per QALY). Varying the specificity of EBCT over its 95% confidence interval had no influence on our results. If the cost of EBCT was less than $240, EBCT was a cost-effective strategy compared with exercise echocardiography, having an ICER of $40,000 per QALY. Varying the percentage of non-diagnostic EBCT tests, the morbidity, or the mortality associated with EBCT-testing did not change our conclusions.

Table 4: Results for a cohort of 55-year-old men with non-specific chest pain, atypical, and typical angina

55-year-old men	Cost ($)	Effect (QALYs)	ICER ($/QALY)
Non-specific chest pain			
No imaging	24,500	14.967	.
Exercise Echo	27,200	15.007	68,108
Exercise SPECT	27,800	15.007	ED
EBCT	27,900	15.010	213,147
Coronary angiography	29,000	15.013	465,551
Atypical angina			
No imaging	26,300	12.670	.
Exercise Echo	32,500	12.897	31,867
Exercise SPECT	33,900	12.899	ED
EBCT	34,100	12.912	ED
Coronary angiography	34,900	12.932	40,525
Typical Angina			
No imaging	27,600	11.221	.
Exercise Echo	37,500	11.604	ED
Exercise SPECT	37,900	11.608	ED
EBCT	38,100	11.630	ED
Coronary angiography	38,800	11.665	25,263

ED, extended dominated strategy, i.e. the incremental cost-effectiveness ratio is higher than the next more effective strategy.
QALY, quality-adjusted life-year; ICER, incremental cost-effectiveness ratio; Echo, echocardiography; SPECT, single-photon emission computed tomography; EBCT, electron-beam computed tomography

Pre-test probability of coronary artery disease

If the pre-test probability of coronary artery disease was 65% or less EBCT was a cost-effective strategy in the base case (ICER $47,500 per QALY) compared with exercise echocardiography. Figure 1 shows the optimal strategy as a function of the pre-test probability of coronary artery disease and the willingness-to-pay threshold. The figure shows that the EBCT strategy was the optimal strategy for a small proportion of patients with atypical chest pain at a relatively high cost per QALY.

Health utilities

The ranking of the incremental cost-effectiveness ratios was insensitive to the values of the health utilities assigned to patients without chest pain, with mild, and with severe chest pain resulting after treatment. However, the incremental cost-effectiveness ratios of the tests changed. If no adjustments were made for quality of life, the incremental cost-effectiveness ratios for exercise echocardiography and angiography were $46,100 and $58,300 per life year saved respectively.

Table 5: Results for a cohort of 65-year-old women with non-specific chest pain, atypical, and typical angina

65-year-old women	Cost ($)	Effect (QALYs)	ICER ($/QALY)
Non-specific chest pain			
No imaging	20,600	12.780	.
Exercise Echo	22,700	12.804	89,726
Exercise SPECT	23,300	12.804	ED
EBCT	23,400	12.806	346,491
Coronary angiography	24,600	12.807	1,037,588
Atypical angina			
No imaging	21,000	11.305	.
Exercise Echo	26,500	11.450	37,887
Exercise SPECT	27,000	11.451	ED
EBCT	27,200	11.460	66,562
Coronary angiography	28,100	11.472	78,162
Typical Angina			
No imaging	22,800	9.612	.
Exercise Echo	31,900	9.927	28,822
Exercise SPECT	32,300	9.930	ED
EBCT	32,600	9.948	ED
Coronary angiography	33,400	9.976	30,874

ED, extended dominated strategy, i.e. the incremental cost-effectiveness ratio is higher than the next more effective strategy.

QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio; Echo, echocardiography; SPECT, single-photon emission computed tomography; EBCT, electron-beam computed tomography

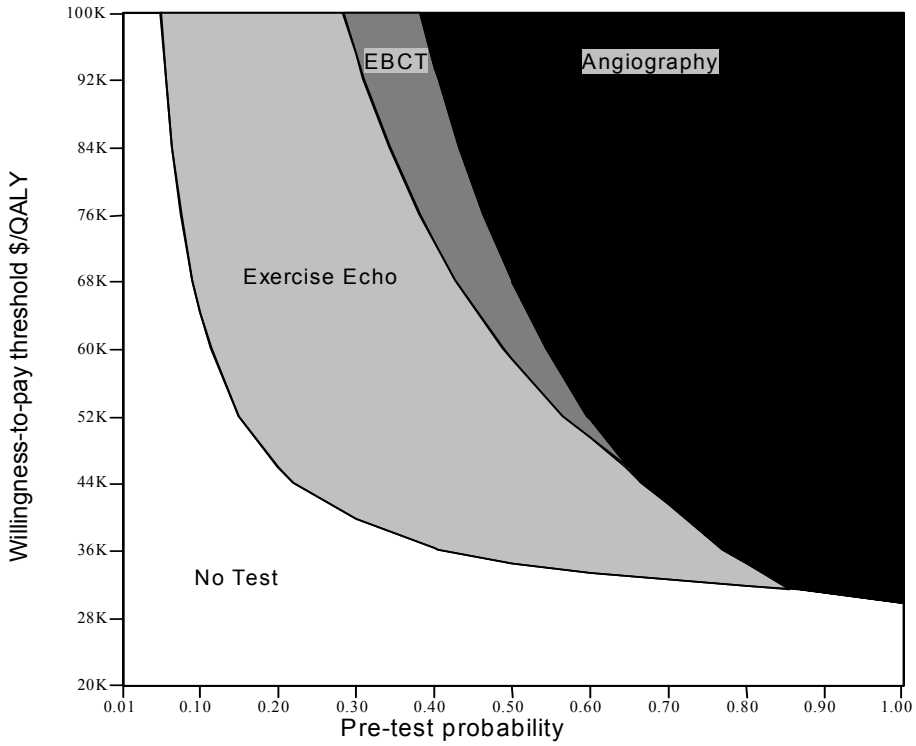


Figure 1: Sensitivity analysis on pre-test probability of disease and the willingness-to-pay threshold. Shaded regions indicate the optimal test strategy for the base case.
Exercise Echo, exercise echocardiography; EBCT, electron-beam computed tomography.

Discussion

This study showed that the cost-effectiveness of EBCT to refer patients to coronary angiography for the diagnosis of coronary artery disease depended mostly on the pre-test probability of disease and the willingness-to-pay threshold. EBCT was only cost-effective for a small range of patients with atypical chest pain at an intermediate pre-test probability of disease (40%-65%) if society is willing to pay more than $47,500 per QALY saved. Exercise echocardiography was optimal at a lower willingness-to-pay threshold for a broader range of patients with atypical chest pain (pre-test probability 20%-80%). The high sensitivity of EBCT and the associated low specificity explain these results: in the EBCT strategy a large proportion of the patients, diseased or not, will be referred to coronary angiography, which is costly but the most effective test in patients at high risk of having coronary artery disease. Therefore, routine coronary angiography dominated the EBCT strategy for patients at high risk of coronary artery disease (70%-100%).

The results of the current study differed from two previous studies that evaluated the short-term cost-effectiveness of EBCT.[13,14] Raggi et al. studied the cost-effectiveness of EBCT compared with exercise electrocardiography in patients with chest pain at low to intermediate pre-test probability of coronary artery disease.[14] The authors concluded that the diagnostic pathway based on EBCT provides a substantial cost benefit over the exercise electrocardiography strategy. However, only initial costs of the diagnostic pathway were evaluated in their study and other imaging tests were not included.

Rumberger et al found that EBCT was cost-effective in patients with a moderate disease prevalence (<70%) compared with exercise electrocardiography, exercise SPECT and exercise echocardiography. This result was obtained by increasing the cut-off value of a positive EBCT-test from a calcium score>0 to a score of 37, 80, or 168.[13] Higher cut-off values are associated with a relatively lower sensitivity and higher specificity. Furthermore, in their analysis cost-effectiveness was defined as the total direct costs of the diagnostic work-up divided by the total number of patients correctly diagnosed with disease. This criterion for cost-effectiveness does not account for long-term effects and also emphasizes optimizing specificity to avoid the expense of angiography.

A limitation of the current study is that tests such as exercise SPECT and exercise echocardiography are physiologic tests that measure myocardial ischemia, whereas EBCT is based on anatomical findings. In the literature, however, tests for the diagnosis of coronary artery disease are compared with the same reference standard: coronary angiography, which is limited to diagnosing the presence and extent of luminal stenosis. Our decision model did not incorporate the functional information provided by exercise testing. Despite the anatomical reference, however, exercise echocardiography was the most efficient imaging test in most cohorts.

Another limitation of our study is that we did not model the prognostic value conditional on the amount of coronary calcium identified with EBCT, which is the more common usage of this technology. Nevertheless, our analysis does consider prognosis by modeling life expectancy conditional on identifying coronary artery disease and the probability of severe disease.

A limitation of every decision model is that input variables are retrieved from various data sources. We derived the test characteristics from meta-analyses, which provided summary estimates from the published literature. Verification bias and publication bias may have affected these estimates. Furthermore, test characteristics, risks, and cost variables may change over time, when the test is applied to a wider patient spectrum. To account for the uncertainty of the variables in our model we used sensitivity analyses and found the results of exercise echocardiography to be robust, whereas the incremental cost-effectiveness ratio of EBCT was sensitive to changes in cost, sensitivity and pre-test probability of disease.

Recent technological developments in scanning techniques for CAD have been spiral and multi-detector computed tomography (MDCT) adapted to acquire cardiac images. The advantages of MDCT over EBCT are that MDCT scanners are less expensive and are becoming more widely available. Currently, the test characteristics of spiral CT and MDCT for coronary imaging are being investigated. A recent study that compared spiral CT and EBCT using a working phantom heart showed that spiral CT had a significantly higher sensitivity and specificity than EBCT.[26] Other studies showed very high correlation coefficients for spiral (MD)CT and EBCT in calcium measurements, varying from 0.90 to 0.99.[27-30] Furthermore, the ability of MDCT angiography

techniques to help identify both calcified and noncalcified atherosclerotic plaque components is emerging.[31, 32] Because MDCT scanners are less expensive than EBCT scanners and because the diagnostic performance of MDCT may be better than EBCT, MDCT may become a cost-effective test for CAD. If more patient data become available the cost-effectiveness of MDCT could be compared with other imaging strategies.

In conclusion, our results suggest that using EBCT to decide which patients with chest pain should undergo coronary angiography is not cost-effective. In a selected group of patients with an intermediate pre-test probability of disease EBCT provides a small health benefit at a relatively high cost.

References

1. 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.
2. Baumgart D, Schmermund A, Goerge G, et al. Comparison of electron beam computed tomography with intracoronary ultrasound and coronary angiography for detection of coronary atherosclerosis. J Am Coll Cardiol. 1997;30:57-64.
3. Bielak LF, Kaufmann RB, Moll PP, McCollough CH, Schwartz RS, Sheedy PF, 2nd. Small lesions in the heart identified at electron beam CT: calcification or noise? Radiology. 1994;192:631-636.
4. Braun J, Oldendorf M, Moshage W, Heidler R, Zeitler E, Luft FC. Electron beam computed tomography in the evaluation of cardiac calcification in chronic dialysis patients. Am J Kidney Dis. 1996;27:394-401.
5. Budoff MJ, Georgiou D, Brody A, et al. Ultrafast computed tomography as a diagnostic modality in the detection of coronary artery disease: a multicenter study. Circulation. 1996;93:898-904.
6. Fallavollita JA, Brody AS, Bunnell IL, Kumar K, Canty JM, Jr. Fast computed tomography detection of coronary calcification in the diagnosis of coronary artery disease. Comparison with angiography in patients < 50 years old. Circulation. 1994;89:285-290.
7. Kajinami K, Seki H, Takekoshi N, Mabuchi H. Noninvasive prediction of coronary atherosclerosis by quantification of coronary artery calcification using electron beam computed tomography: comparison with electrocardiographic and thallium exercise stress test results. J Am Coll Cardiol. 1995;26:1209-1221.
8. Yaghoubi S, Tang W, Wang S, et al. Offline assessment of atherosclerotic coronary calcium from electron beam tomograms. Am J Card Imaging. 1995;9:231-236.
9. Yao Z, Liu XJ, Shi R, et al. A comparison of 99mTc-MIBI myocardial SPET with electron beam computed tomography in the assessment of coronary artery disease. Eur J Nucl Med. 1997;24:1115-1120.
10. Hecht HS. Practice guidelines for electron beam tomography: a report of the Society of Atherosclerosis Imaging. Am J Cardiol. 2000;86:705-706.
11. Nallamothu BK, Saint S, Bielak LF, et al. Electron-beam computed tomography in the diagnosis of coronary artery disease: a meta-analysis. Arch Intern Med. 2001;161:833-838.
12. O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/American Heart Association Expert Consensus document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. Circulation. 2000;102:126-140.
13. Rumberger JA, Behrenbeck T, Breen JF, Sheedy PF, 2nd. Coronary calcification by electron beam computed tomography and obstructive coronary artery disease: a model for costs and effectiveness of diagnosis as compared with conventional cardiac testing methods. J Am Coll Cardiol. 1999;33:453-462.
14. Raggi P, Callister TQ, Cooil B, Russo DJ, Lippolis NJ, Patterson RE. Evaluation of chest pain in patients with low to intermediate pretest probability of coronary artery disease by electron beam computed tomography. Am J Cardiol. 2000;85:283-288.
15. Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). Circulation. 1981;64:360-367.
16. 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.
17. Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C. Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review. Health Technol Assess. 2000;4:1-153.
18. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA. 1998;280:913-920.

19. Nease RF, Jr., Kneeland T, O'Connor GT, et al. Variation in patient utilities for outcomes of the management of chronic stable angina. Implications for clinical practice guidelines. Ischemic Heart Disease Patient Outcomes Research Team. JAMA. 1995;273:1185-1190.
20. Coronary artery surgery study (CASS): a randomized trial of coronary artery bypass surgery. Quality of life in patients randomly assigned to treatment groups. Circulation. 1983;68:951-960.
21. Pocock SJ, Henderson RA, Rickards AF, et al. Meta-analysis of randomised trials comparing coronary angioplasty with bypass surgery. Lancet. 1995;346:1184-1189.
22. Myocardial infarction and mortality in the coronary artery surgery study (CASS) randomized trial. N Engl J Med. 1984;310:750-758.
23. Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. Lancet. 1994;344:563-570.
24. Wong JB, Sonnenberg FA, Salem DN, Pauker SG. Myocardial revascularization for chronic stable angina. Analysis of the role of percutaneous transluminal coronary angioplasty based on data available in 1989. Ann Intern Med. 1990;113:852-871.
25. King SB, 3rd, Lembo NJ, Weintraub WS, et al. A randomized trial comparing coronary angioplasty with coronary bypass surgery. Emory Angioplasty versus Surgery Trial (EAST). N Engl J Med. 1994;331:1044-1050.
26. Hopper KD, Strollo DC, Mauger DT. Comparison of electron-beam and ungated helical CT in detecting coronary arterial calcification by using a working heart phantom and artificial coronary arteries. Radiology. 2002;222:474-482.
27. Becker CR, Jakobs TF, Aydemir S, et al. Helical and single-slice conventional CT versus electron beam CT for the quantification of coronary artery calcification. AJR Am J Roentgenol. 2000;174:543-547.
28. Becker CR, Kleffel T, Crispin A, et al. Coronary artery calcium measurement: agreement of multirow detector and electron beam CT. AJR Am J Roentgenol. 2001;176:1295-1298.
29. Carr JJ, Crouse JR, 3rd, Goff DC, Jr., D'Agostino RB, Jr., Peterson NP, Burke GL. Evaluation of subsecond gated helical CT for quantification of coronary artery calcium and comparison with electron beam CT. AJR Am J Roentgenol. 2000;174:915-921.
30. Horiguchi J, Nakanishi T, Ito K. Quantification of coronary artery calcium using multidetector CT and a retrospective ECG-gating reconstruction algorithm. AJR Am J Roentgenol. 2001;177:1429-1435.
31. Schoenhagen P, Halliburton SS, Stillman AE, et al. Noninvasive Imaging of Coronary Arteries: Current and Future Role of Multi-Detector Row CT. Radiology. 2004;232:7-17.
32. Schoepf UJ, Becker CR, Ohnesorge BM, Yucel EK. CT of Coronary Artery Disease. Radiology. 2004;232:18-37.

6

Estimation of optimal test characteristics based on cost-effectiveness analysis: flexible versus fixed test characteristics

Abstract

Purpose: To estimate the optimal test characteristics of electron-beam computed tomography (EBCT) for the workup of patients suspected of coronary artery disease based on a cost-effectiveness analysis and to compare these with pooled published values of sensitivity and specificity.

Methods: A Markov decision model was developed to model lifetime effectiveness, lifetime costs, and net health benefits of the true- and false-positive and true- and false-negative test outcomes and subsequent treatment for patients suspected of coronary artery disease. The sensitivity of EBCT was modeled as a function of the false positive rate using summary receiver operating characteristic (SROC) analysis. Using sensitivity analysis on the false-positive rate, the net health benefit was calculated for each combination of sensitivity and specificity on the SROC curve. The optimal test characteristics of EBCT were determined at the point on the SROC curve where the net health benefit reached its maximum. Optimal test characteristics were evaluated for 6 cohorts depending on age, sex, and type of chest pain. The analyses were performed from the health-care perspective. We assumed a threshold willingness-to-pay of $50,000 per quality-adjusted life year (QALY) gained. The published data on sensitivity and specificity were also pooled, weighted by sample size, and compared to the optimal test characteristics derived from the cost-effectiveness analysis.

Results: For the workup of a cohort of 55 year-old men with atypical angina (pre-test probability is 0.71), the optimal operating point for EBCT was found at a sensitivity of 93% and specificity of 34%.For a cohort of 65-year-old women with atypical angina (pre-test probability is 0.48) the optimal sensitivity and specificity were 87% and 73% respectively. Pooled published data demonstrated a sensitivity of 92% and a specificity of 51% for EBCT. In patients with non-specific chest pain no imaging was optimal, whereas in patients with typical chest pain routine coronary angiography was the optimal strategy. With an increase in the willingness-to-pay to $75,000 per QALY the optimal sensitivity increased to 99%, whereas specificity decreased to 14% in 55-year-old men with atypical angina.

Conclusions: Optimal test characteristics based on cost-effectiveness analysis differ from pooled values of sensitivity and specificity. They depend on the setting in which the test is used and on the willingness-to-pay threshold. This study suggests that, where possible, the interpretation of test results should be adjusted to optimize the use of such tests.

Introduction

In cost-effectiveness models that evaluate the performance of diagnostic tests normally fixed values of sensitivity and specificity are used. Generally, published pooled values of sensitivity and specificity are obtained from a meta-analysis. However, pooled values of sensitivity and specificity indicate neither the variability of the test characteristics nor the best combination of sensitivity and specificity for a particular clinical setting. To evaluate the uncertainty around these estimates sensitivity analyses are often performed over the 95% confidence intervals of the point estimates. This, however, disregards the inverse relationship between sensitivity and specificity and ignores the possibility of adjusting the interpretation of the test results so as to optimize the test characteristics.

The inverse relationship between sensitivity and specificity can be calculated using (summary) receiver operating characteristic ((S)ROC) analysis. Furthermore, various methods have been reported in the literature to determine the optimal combination of sensitivity and specificity on the SROC curve of a diagnostic test. Some investigators have used the point on the SROC curve that is closest to the upper left corner as 'optimal' operating point[1], the so-called J-point. Others have used the operating point with a likelihood ratio equal to 1.[2] Other criteria for determining the optimal operating point include selection of the Q-point: the point where sensitivity equals specificity[3], maximizing accuracy or the sum of sensitivity plus specificity,[4,5] or accepting a preset level of sensitivity (or specificity) and determining the corresponding specificity (or sensitivity respectively).[6] All these methods minimize the number of false-positive and false-negative test results. However, none of these methods accounts for the fact that the consequences of a false-negative test result in those with the disease may be far worse than a false-positive test result in those without the disease or vice-versa. Therefore, depending on the prevalence of disease and the consequences of the test result, the sensitivity should be maximized in a particular clinical setting, whereas the specificity should be optimized in another setting.

The aim of this study was to determine the optimal combination of sensitivity and specificity of electron-beam computed tomography (EBCT) testing for patients with chest pain based on a cost-effectiveness analysis. A secondary objective was to compare outcomes of the cost-effectiveness model using the optimal test characteristics and the pooled published values of sensitivity and specificity.

Methods

Target population

We focused on patients presenting with non-specific chest pain, atypical angina, or typical angina, as defined in the Coronary Artery Surgery Study (CASS).[7] Chest pain was defined as typical angina for pain that is substernal, follows physical exertion, and is relieved by nitroglycerin. If two of these criteria were met, chest pain was defined as atypical angina and if only one criterion applied it was defined as nonspecific chest pain. We considered 6 cohorts of patients defined by age, sex, and type of chest pain. Cohorts of 55 year-old men and 65-year-old women were evaluated because these are the average ages at which men and women present with chest pain. The pre-test probabilities of coronary artery disease (CAD) in 55-year-old men with

typical, atypical, and nonspecific chest pain, based on the CASS, were respectively 95%, 71% and 18%. For 65-year-old women the pre-test probabilities were somewhat lower: 81%, 48% and 10% respectively.[7] As a base case we evaluated a cohort of 55-year-old men with atypical chest pain.

Markov model

We updated a Markov model[8] to evaluate the cost-effectiveness of EBCT compared with no imaging and routine coronary angiography for the diagnosis of CAD in patients presenting with chest pain. The model estimated the lifetime costs and quality-adjusted life years (QALYs) for cohorts of patients with a true- or false-positive and those with a true- or false-negative test result for CAD from the perspective of the health care system.[8] Patients with a negative EBCT result were treated medically, whereas patients with a positive or non-diagnostic test result were referred to coronary angiography, which provided the final diagnosis. Based on coronary angiography, patients were stratified into categories of no, single-, double-, and triple-vessel or left main disease. The post-catheterization treatment strategy in the model was coronary artery bypass graft surgery (CABG) for triple-vessel or left main coronary artery disease, percutaneous coronary intervention (PCI) for single- or double-vessel stenosis and medical therapy if no significant CAD was found. Significant angiographic disease was defined as the presence of a luminal stenosis of 50% or more in at least one coronary artery. Nonfatal myocardial infarction and revascularization procedures occurring after the initial treatment decision were modeled to evaluate their effects on cost and quality of life. We estimated the quality-adjusted life expectancy by modelling annual transitions between 4 different health states: no chest pain, mild chest pain, severe chest pain, and death. Quality of life with no symptoms was valued at 0.87, with mild chest pain at 0.81, and with severe chest pain at 0.67 on a scale from 0, representing death, to 1, representing full health, based on a standard gamble evaluation in patients with chronic angina.[8,9]

We transformed the two outcomes of the Markov model, costs and QALYs, into one balanced health measure: the net health benefit (NHB).[10] The NHB was defined as the lifetime effectiveness (QALYs) minus the lifetime costs ($), the latter divided by the societal willingness-to-pay (WTP) threshold to save one QALY ($/QALY), in formula:

$$\text{NHB} = \text{QALYs} - \frac{\text{Cost}}{\text{WTP threshold}} \qquad \text{(eq 1)}$$

The net health benefit is expressed in QALY-equivalents. We considered 3 generally accepted thresholds of societal willingness-to-pay: $25,000, $50,000 and $75,000 per QALY gained.[11]

Data sources

Test characteristics

Sensitivity and specificity estimates of EBCT and pooled values of sensitivity and specificity for diagnosing patients suspected of coronary artery disease were obtained from a published meta-analysis.[12] A standard method of EBCT calcium scoring was used that defined significant CAD based on a calcium score >0.[13]

Table 1: Test characteristics of EBCT and coronary angiography. The sensitivity of EBCT can be calculated using equation 2 and the intercept and slope of the SROC curve for each value of the specificity.

	EBCT	CAG
Costs ($)	271	2,395*
Intercept SROC curve, *i*	2.310	n/a
Slope SROC curve, *b*	0.09875	n/a
Mortality (%)	0	0.1
Morbidity (%)	0	1.6
Uninterpretable test result (%)	2	0

* cost for outpatient CAG procedure (70% of CAG procedures); inpatient CAG cost $5,030 (30% of CAG procedures)
EBCT, electron-beam computed tomography; CAG, coronary angiography; SROC, summary receiver operating characteristic; n/a, not applicable

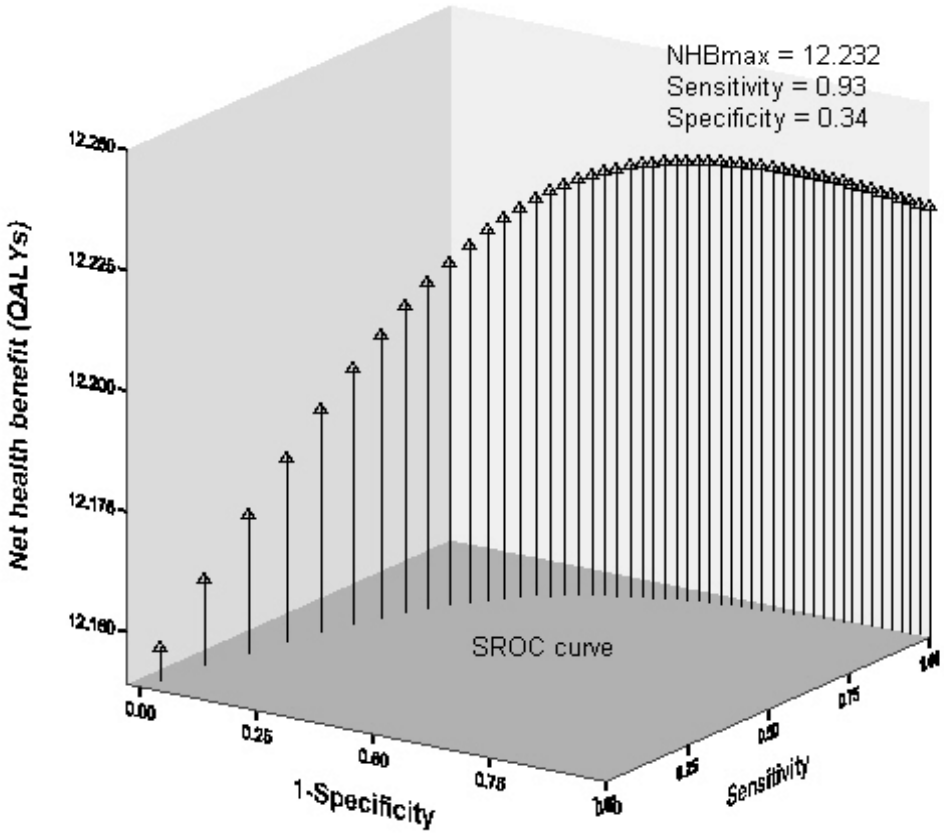


Figure 1: Net health benefit (NHB) plotted for each combination of sensitivity and 1-specificity on the summary receiver operating characteristic curve (NHB-SROC curve) for the base case at a willingness-to-pay threshold of $50,000 per QALY gained.

The cost estimates of the diagnostic tests and treatment procedures were obtained from the Medicare-allowed global reimbursement rates based on the 2002 Current Procedural Terminology (CPT) code system for outpatient procedures and the 2002 Diagnosis Related Group (DRG) codes for in-patient procedures, including the technical and professional fees. We assumed that the Medicare-allowed reimbursement for EBCT would be equal to the reimbursement for a CT of the chest. All costs were adjusted to 2002 U.S. dollars using the medical care component of the Consumer Price Index and were discounted at an annual rate of 3%. Furthermore, we assumed that EBCT was uninterpretable in 2% of cases and has no associated morbidity or mortality.[14] The test characteristics are shown in Table 1.

SROC analysis and calculation of the optimal operating point on the SROC curve

The published sensitivity and specificity estimates were re-analyzed with a summary receiver operating characteristic (SROC) analysis using a random effects model, which accounts for the variation among studies (STATA 7.0).[15,16] Each point on an SROC curve represents a combination of sensitivity (or true-positive rate (TPR)) and 1-specificity (or false-positive rate (FPR)) at a certain threshold value of a diagnostic test.

In a SROC analysis, the TPR and the FPR are transformed into their logit form. Subsequently, the difference (D) of the logit transformations, which is equal to the diagnostic odds ratio, and the sum (S) of the logit transformations, a proxy for the cut-off value of the test, are calculated. Then, a linear regression analysis is performed, in which D is the dependent variable and S is a predictor, weighted for the inverse of the variance of D. The coordinates of the SROC curve can be calculated by back transformation and rearrangement of the regression equation using the following formula:

$$TPR = \frac{1}{1 + \left[e^{i/(1-b)} * \left(\frac{FPR}{1 - FPR} \right)^{(1+b)/(1-b)} \right]^{-1}} \qquad \text{(eq 2)}$$

in which *i* is the intercept and b the slope of the regression model. This formula was entered into the Markov model as variable for the TPR of EBCT. A sensitivity analysis on the FPR was performed. For each possible value of the FPR the model calculated the associated TPR and the resulting net health benefits. The optimal combination of the true- and false-positive rate, or the optimal operating point, was defined as the point on the SROC curve where the net health benefit reaches its maximum. This analysis was repeated for the six cohorts and for the three willingness-to-pay thresholds for the base case.

The optimal combination of sensitivity and specificity or the optimal operating point on the SROC curve is determined by the prior probability of disease, and the long-term risks, benefits and costs associated with the test outcomes. In particular, a trade-off needs to be made between the net loss due to false-positive (FP) test results relative to true-negative (TN) test results and the net loss due to false-negative (FN) test results relative to true-positive (TP) test results[17,18], in formula:

$$\frac{\Delta TPR}{\Delta FPR} > \frac{1-p}{p} * \frac{NHB_{TN} - NHB_{FP}}{NHB_{TP} - NHB_{FN}} \qquad \text{(eq 3)}$$

Results

Optimal operating points

For each combination of sensitivity and 1-specificity on the SROC curve the associated net health benefit was calculated for each of the six cohorts. The results of the base case are plotted in a three-dimensional graph, which shows the SROC curve in the x-y plane and the net health benefit on the z-axis (NHB-SROC curve; Figure 1). The top of the NHB-SROC curve represents the maximum net health benefit that can be attained with EBCT at the optimal combination of sensitivity and specificity for the base case.

Figure 2 shows for each of the six cohorts the net health benefit as a function of the false-positive rate at a willingness-to-pay threshold of $50,000 per QALY. For simplicity the sensitivity is not plotted in these graphs. In patients with non-specific chest pain the net health benefit reached its maximum at a FPR of 0%, or a specificity of 100% with an associated sensitivity of 0%, which implies that the test should not be performed at all. Using these test characteristics all patients would have a negative test result and would be treated medically. As a result, the no imaging strategy was the optimal strategy for these cohorts. In patients with typical chest pain, however, the maximum net health benefit was reached at a FPR of 100% or a specificity of 0% with an associated sensitivity of 100%, which implies that in these patients EBCT should not be performed either, because all patients would have a positive test result and would be referred to coronary angiography for the final diagnosis. The optimal test strategy for patients with typical chest pain was therefore routine coronary angiography. EBCT was only cost-effective in cohorts of patients with atypical chest pain. For a 55-year-old male the maximum net health benefit was 12.232 QALYs, which was attained at a sensitivity of 93% and a specificity of 34% at a willingness-to-pay threshold of $50,000 per QALY. For a 65-year-old female with atypical chest pain the maximum net health benefit was 10.918 QALYs, which was achieved at a sensitivity of 87% and a specificity of 63%. In the cohort of 65-year-old women with atypical angina EBCT showed a higher net health benefit than coronary angiography over a large part of the SROC curve. This implies that EBCT was the optimal strategy for this cohort if the FPR of EBCT ranged between 15% and 80%, which is equal to a specificity of 20%-85% with an associated sensitivity ranging from 99% to 61%. Routine coronary angiography was the optimal strategy for the remainder of the SROC curve, thus at a very low or very high FPR of EBCT.

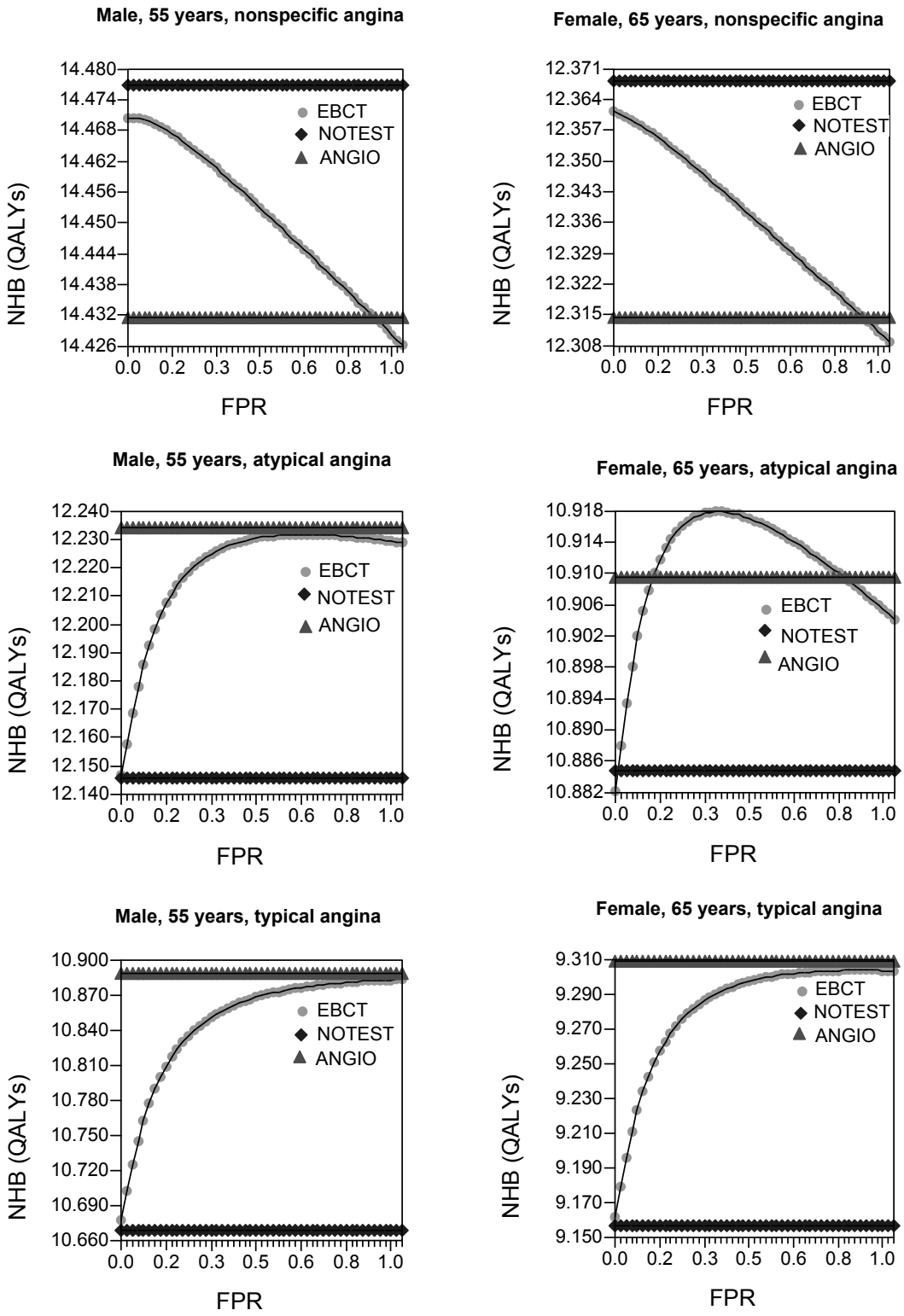


Figure 2: NHB-SROC curves for 3 cohorts of 55-year-old men and 3 cohorts of 65-year-old women with non-specific, atypical, and typical angina at a willingness-to-pay threshold of $50,000 per QALY.

Pooled weighted analysis

The pooled sensitivity and specificity of EBCT were 92.3% (95% CI 90.7-94.0) and 51.2% (95% CI 47.2-54.9) respectively.[12] The net health benefit of EBCT at this combination of sensitivity and specificity was 12.230 QALYs for the cohort of 55-year-old men with atypical angina at a willingness-to-pay threshold of $50,000 per QALY. For the cohort of 65-year-old women with atypical angina, the net health benefit was 10.917 QALYs (Table 2). The outcomes in Table 2 show that for each of the willingness-to-pay thresholds, the net health benefit was lower when the pooled values of sensitivity and specificity were analyzed than when optimal test characteristics were used.

Sensitivity analysis

Varying the willingness-to-pay threshold from $25,000 to $75,000 per QALY increased the maximum net health benefit that could be achieved with EBCT and moved the optimal operating point to a higher sensitivity and a lower associated specificity, resulting in a higher net health benefit (Table 2). At a higher sensitivity of EBCT more patients will be selected for coronary angiography, which is more expensive, but also results in higher effectiveness, in spite of the risks associated with angiography.

Table 2: Maximum net health benefits (NHBmax) obtained using optimal operating points for EBCT compared with the net health benefits (NHB) obtained using the pooled published values of sensitivity and specificity (93% and 51% respectively) of EBCT for two cohorts of patients with atypical chest pain using willingness-to-pay thresholds of $25,000, $50,000, and $75,000 per QALY.

	Optimal operating point EBCT			Pooled values EBCT
Willingness-to-pay	Sensitivity	Specificity	NHBmax	NHB
$/QALY	%	%	QALYs	QALYs
Men, 55 years, atypical angina				
25,000	0	100	11.605	11.548
50,000	93	34	12.232	12.230
75,000	99	14	12.464	12.458
Women, 65 years, atypical angina				
25,000	0	100	10.449	10.373
50,000	87	63	10.918	10.917
75,000	95	44	11.098	11.098

NHB, net health benefit; max, maximum; QALY, quality-adjusted life year

Discussion

We demonstrated that the optimal sensitivity and specificity of EBCT accounting for long-term costs and effectiveness differed from the pooled published values of sensitivity and specificity.

Furthermore, the optimal test characteristics depended on the setting in which the test was used and on the willingness-to-pay threshold. Although the differences were small, our results suggest that the net health benefit of a diagnostic test could be improved by adjusting the interpretation of the test result so as to optimize the test characteristics. Moreover, using optimal test characteristics instead of fixed values of sensitivity and specificity in a cost-effectiveness analysis of diagnostic test strategies could potentially change the decision with respect to the optimal test strategy. For example, if NHB-SROC curves are plotted for more than one test the curves may cross, which implies that one test may maximize net health benefit at a low FPR, whereas the other test may provide the same net health benefit at a higher FPR. In cohorts with a high probability of disease, the optimal operating point moved to a higher sensitivity and a lower associated specificity compared to a cohort with low probability of disease. In clinical practice this would result in more patients being referred to coronary angiography, which yielded better outcomes in terms of effectiveness and higher costs than if the pooled sensitivity and specificity values were considered. A higher sensitivity results in fewer patients with CAD that are misdiagnosed (false negatives), but, because of the associated lower specificity, more patients also unnecessarily undergo coronary angiography (false positives). On the contrary, in patients with a lower probability of disease the optimal test characteristics moved to a lower sensitivity and higher specificity. In patients with non-specific chest pain for example, our model suggested that a sensitivity of 0% and specificity of 100% would be optimal implying that EBCT testing would lead to unnecessary coronary angiography.

SROC curves reflect the discriminatory power of a diagnostic test in different settings whereas the pooled values of sensitivity and specificity reflect a fixed setting. Conventional ROC curves take into account the variation in sensitivity and specificity when the cut-off value of the diagnostic test is varied within one patient group. SROC curves, however, are derived from published studies in various patient cohorts in different settings, each with its own methods for performing and interpreting the tests, and chosen cut-off value of the test variable. In the meta-analysis that we used for our SROC analysis, for example, a standardized calcium scoring method was used in each source study. However, variations between the study protocols existed, which included different definitions of the minimum area of tissue attenuation required for a lesion to be considered calcium and not artifact; minimum areas ranged in size from 0.5 to 2 mm.[12] Furthermore, the number of tomograms obtained in each examination (range 20-40) and whether a standard inspiration or expiration breath-hold was used during image acquisition varied across studies. SROC curves reflect this variation in study protocols and populations studied in addition to potential differences in cut-off values.

Due to the variation in patient populations and disparity in methods for performing and interpreting the test results we were, unfortunately, unable to calculate the explicit cut-off value of the calcium score corresponding to the optimal sensitivity and specificity. The ease of altering cut-off values through adjusting minimum size criteria for lesions or calcium scores is an advantage of EBCT compared with traditional tests such as exercise echocardiography or exercise single-photon emission computed tomography, which yield categorical rather than continuous test results. In general, our results suggest that more lenient criteria should be applied for patients with high risk factors and stricter criteria for patients with low risk factors for CAD. For EBCT this would mean that a low calcium score should be used as cut-off value for high-risk patients with severe chest

pain, while a high calcium score should be used as cut-off value for low-risk patients with mild chest pain, to get the maximum net health benefit from the test. This implies that for women, who have a lower pre-test probability of CAD than men, the calcium cut-off should be set at a higher value than for men. Rumberger et al. attempted to define ranges for EBCT calcium scores to predict the severity of CAD.[19] Although their definition of the optimal operating point did not account for the long-term outcome, the authors found the same principle in optimizing sensitivity or specificity for high- and low-risk patients respectively.

We assumed that the societal willingness-to-pay threshold can be defined, although we recognize that the threshold value is difficult to determine because it may fluctuate with time, differs across countries, and may differ across types of interventions. The results of the sensitivity analysis in which the willingness-to-pay threshold was varied suggested that the optimal test characteristics of EBCT for the diagnosis of CAD depended highly on what society is willing to pay for health care. At a willingness-to-pay of $50,000 per QALY, for example, our results suggest that if the test characteristics of EBCT are optimized, EBCT was a cost-effective test for patients with atypical angina. However, if the willingness-to-pay threshold was set at $25,000 per QALY, EBCT was too expensive for the same patient cohorts. On the other hand, the optimal sensitivity increased and specificity decreased if the willingness-to-pay threshold increased to $ 75,000 per QALY. This finding suggests that more coronary angiographies could and should be performed if society is willing to pay more, and this will lead to a higher effectiveness despite increased risks and cost.

The results of our study should be interpreted in the context of the following limitations. The input variables in our Markov model were retrieved from various data sources and assumptions were required to model the diagnostic workup and treatment of patients suspected of CAD. We derived the test characteristics from a meta-analysis, which included studies that varied widely in demographics, prevalence of CAD, and EBCT protocol. Verification bias and publication bias may have affected these estimates. Furthermore, test characteristics, risks and cost variables may change over time, when the test is applied to a wider patient spectrum. We did not explore the uncertainty of all the variables in our model, because we wanted to demonstrate the differences in outcome of the model depending on the flexibility of the test characteristics alone.

In conclusion, this analysis suggest that flexible test characteristics should be analyzed in cost-effectiveness models of diagnostic tests instead of using one set of fixed values of sensitivity and specificity. Optimal test characteristics based on cost-effectiveness analysis differ from pooled values of sensitivity and specificity. They depend on the setting in which the test is used and on the willingness-to-pay threshold. This study suggests that, where possible, the interpretation of test results should be adjusted to optimize the use of such tests.

References

1. Criswell BK, Langsfeld M, Tullis MJ, Marek J. Evaluating institutional variability of duplex scanning in the detection of carotid artery stenosis. Am J Surg. 1998;176:591-597.
2. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000;45:23-41.
3. Westwood ME, Kelly S, Berry E, et al. Use of magnetic resonance angiography to select candidates with recently symptomatic carotid stenosis for surgery: systematic review. BMJ. 2002;324:198.
4. Carsten CG, 3rd, Elmore JR, Franklin DP, Thomas DD, Mordan F, Wood GC. Use of limited color-flow duplex for a carotid screening project. Am J Surg. 1999;178:173-176.
5. AbuRahma AF, Robinson PA, Strickler DL, Alberts S, Young L. Proposed new duplex classification for threshold stenoses used in various symptomatic and asymptomatic carotid endarterectomy trials. Ann Vasc Surg. 1998;12:349-358.
6. Fillinger MF, Baker RJ, Jr., Zwolak RM, et al. Carotid duplex criteria for a 60% or greater angiographic stenosis: variation according to equipment. J Vasc Surg. 1996;24:856-864.
7. Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). Circulation. 1981;64:360-367.
8. 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.
9. Nease RF, Jr., Kneeland T, O'Connor GT, et al. Variation in patient utilities for outcomes of the management of chronic stable angina. Implications for clinical practice guidelines. Ischemic Heart Disease Patient Outcomes Research Team. JAMA. 1995;273:1185-1190.
10. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998;18:S68-80.
11. Graham JD, Corso PS, Morris JM, Segui-Gomez M, Weinstein MC. Evaluating the cost-effectiveness of clinical and public health measures. Annu Rev Public Health. 1998;19:125-152.
12. Nallamothu BK, Saint S, Bielak LF, et al. Electron-beam computed tomography in the diagnosis of coronary artery disease: a meta-analysis. Arch Intern Med. 2001;161:833-838.
13. 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.
14. Rumberger JA, Behrenbeck T, Breen JF, Sheedy PF, 2nd. Coronary calcification by electron beam computed tomography and obstructive coronary artery disease: a model for costs and effectiveness of diagnosis as compared with conventional cardiac testing methods. J Am Coll Cardiol. 1999;33:453-462.
15. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making. 1993;13:313-321.
16. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med. 1994;120:667-676.
17. Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. Med Decis Making. 1988;8:279-289.
18. Hunink MGM, Glasziou PP. Multiple test results. In: Decision making in health and medicine. Integrating evidence and values. 1st ed. Cambridge: Cambridge University Press; 2001.
19. Rumberger JA, Sheedy PF, Breen JF, Schwartz RS. Electron beam computed tomographic coronary calcium score cutpoints and severity of associated angiographic lumen stenosis. J Am Coll Cardiol. 1997;29:1542-1548.

7

The optimal peak systolic velocity threshold on duplex ultrasonography for the indication carotid endarterectomy: a decision analytic approach

Abstract

Purpose: To determine the optimal peak systolic velocity (PSV) threshold on duplex ultrasonography for the indication of carotid endarterectomy based on the long-term cost-effectiveness of diagnostic testing and treatment.

Materials and Methods: From January 1997 to January 2000, a prospective study was conducted including 350 patients with symptoms of amaurosis fugax, TIA, or minor stroke, who underwent duplex ultrasonography and digital subtraction angiography. In total 236 non-occluded carotid arteries were available for evaluation by both imaging modalities. Receiver operating characteristic (ROC) curves were constructed for the diagnosis of either a 70-99% or a 50-99% stenosis. Based on the lifetime costs and quality-adjusted life-years, obtained from a cost-effectiveness analysis, and the prevalence of disease, the optimal likelihood ratio was calculated. Finally, the associated optimal sensitivity, specificity and threshold PSV were derived from the ROC-curves.

Results: For the diagnosis of a 70-99% stenosis the optimal likelihood ratio was 0.21, which was associated with a threshold PSV of 220 cm/s, a sensitivity of 97%, and specificity of 49%. For the diagnosis of a 50-99% stenosis the optimal likelihood ratio was 0.38, which was associated with a threshold PSV of 180 cm/s, a sensitivity of 95%, and specificity of 69%.

Conclusion: Based on the lifetime consequences of diagnostic testing and subsequent treatment, the optimal threshold PSV for the diagnosis of a 70-99% carotid artery stenosis in patients with amaurosis fugax, TIA, or minor stroke is 220 cm/s and for the diagnosis of a 50-99% stenosis the optimal threshold PSV is 180 cm/s.

Introduction

Two large randomized trials, the North American Symptomatic Carotid Endarterectomy Trial (NASCET) and the European Carotid Surgery Trial (ECST), have shown a considerable benefit for carotid endarterectomy in patients with a 70-99% stenosis, but also a small benefit for patients with a 50-69% stenosis [1-6]. In these trials, severe carotid artery stenosis was diagnosed by cerebral angiography, the reference standard. However, cerebral angiography carries a risk of mortality and morbidity and has a financial cost. Its routine use in patients with amaurosis fugax, transient ischemic attack (TIA), or minor stroke who are potential endarterectomy candidates is therefore undesirable. Consequently, many clinicians nowadays use noninvasive studies, such as duplex ultrasonography (DUS), magnetic resonance angiography, and computed tomographic angiography, to select patients for carotid endarterectomy [7]. We recently showed that DUS was the optimal test strategy to select patients for carotid entarterectomy in a cost-effectiveness analysis [8].

Moreover, the Society of Radiologists in Ultrasound published recommendations for the interpretation of DUS results in the diagnosis of internal carotid artery stenosis [7]. That study was based on the test characteristics and diagnostic accuracy of DUS obtained from the literature. Maximization of accuracy as criterion to define the optimal threshold of the peak systolic velocity (PSV) assumes that a false-negative test result has the same importance as a false-positive result. However, because the DUS result determines whether or not carotid endarterectomy is performed, the consequences of missing a significant stenosis may be more harmful with respect to cost and/or effectiveness outcomes than performing endarterectomy for a non-significant stenosis. Therefore, it is more clinically relevant to account for the cost and effectiveness associated with false-positive or false-negative test results when selecting an optimal threshold for referral for endarterectomy.

Accordingly, the aim of this study was to determine the optimal peak systolic velocity threshold on duplex ultrasonography for the indication of carotid endarterectomy taking into account the cost-effectiveness outcomes of diagnostic testing and subsequent treatment.

Methods

Study population

From January 1997 to January 2000 a prospective diagnostic study was performed in two academic hospitals and one non-academic hospital in the Netherlands [9]. After informed consent was obtained, patients with symptoms of amaurosis fugax, TIA, or minor stroke underwent carotid duplex ultrasonography and carotid digital subtraction angiography within a time frame of 4 weeks. Patients were enrolled at the University Medical Center Utrecht, Erasmus Medical Center Rotterdam, and Medical Spectrum Twente. The medical ethical committee approved the study for each hospital.

Carotid duplex ultrasonography

The peak systolic velocity (PSV) was measured on a continuous scale in cm/s in the proximal part of the symptomatic internal carotid artery of each patient. A carotid artery was called

symptomatic if the neurological symptoms, i.e. amaurosis fugax, TIA, or minor stroke corresponded with the stenosed side. If no detectable flow was present patients were classified as having an occlusion (100% stenosis). Slow flow in combination with a visualized severe stenosis was defined as a near occlusion. Arteries with an occlusion or near occlusion on ultrasonography were excluded from the analysis, because peak systolic velocity cannot reliably be measured in these arteries.

Digital subtraction angiography (DSA)

DSA was performed by selective positioning of an intra-arterial catheter in the common carotid artery. From the carotid bifurcation, three projections (lateral, posteroanterior, and oblique) were acquired. Additional projections of occasionally performed rotational DSA examinations were not used in the context of this study.

The DSA test results were read by one observer for each hospital. The observers were blinded for clinical information and for the duplex results. The observers used printed hard copies to read the DSA images. The percentage stenosis was measured according to the NASCET criteria [1]. The degree of stenosis is defined as the remaining lumen at the site of the stenosis as percentage of the normal lumen distal to the stenosis. The maximum degree of stenosis of the three projections was used in the analyses. Angiography was considered the standard of reference.

Markov Model

The lifetime costs and quality-adjusted life years following a true or false, positive or negative DUS test result (TP, FP, TN, FN) were derived from a previously published Markov model [8]. These included the costs and utilities associated with subsequent treatment, complications of treatment and progression of disease. Several health states were modeled for the severity of neurological disease (TIA, minor stroke, major stroke either as presenting symptom or as complication) for patients with initially less than 50% stenosis, 50-69% stenosis, and 70-99% stenosis. Medical therapy including aspirin treatment was assumed to be the optimal treatment for patients with less than 50% stenosis. For patients with more than 50% stenosis two criteria for carotid endarterectomy were considered: 70-99% stenosis and 50-99% stenosis. Progression of disease and death were modeled by allowing patients to transition to more severe health states during follow-up. The Markov model was built in DATA Pro 11.0 (TreeAge).

The lifetime costs and effects were integrated into one measure: the net health benefit (NHB), which is defined as the lifetime effects in quality-adjusted life years minus the lifetime costs, the latter divided by the amount that society is willing to pay (WTP threshold) for saving one quality-adjusted life year (QALY) [10]. In formula:

$$\mathrm{NHB} = \mathrm{QALYs} - \frac{\mathrm{Cost}}{\mathrm{WTP\ threshold}} \qquad \text{(eq 1)}$$

We assumed two threshold amounts that society would be willing to pay for saving one quality-adjusted life year: $25,000 and $50,000 per QALY.

Recommended PSV thresholds

The recommended PSV thresholds of 230 cm/s for the diagnosis of a 70-99% stenosis and 125 cm/s for a 50-99% stenosis were applied to our study data [7]. For both PSV thresholds, we calculated the associated sensitivity and specificity.

Optimal PSV threshold

We calculated the sensitivities and specificities associated with different PSV-thresholds (R_i) using DSA as the reference standard. We used two definitions of carotid disease, in agreement with the two indications for carotid endarterectomy: for a 70-99% and for a 50-99% angiographic stenosis. Using receiver operating characteristic (ROC) analysis, all combinations of sensitivity and specificity were plotted in a graph in which the y-axis represents sensitivity and the x-axis represents 1-specificity [11]. Smooth ROC curves were created using the methods of summary ROC analysis [12-14]. For each combination of sensitivity and specificity on the smooth curves the result-specific likelihood ratio (LR_{Ri}), i.e. the probability of a specific test result in the group with the disease divided by the probability of that specific test result in the group without the disease, was calculated which is equal to the tangent or slope of the curve [15, 16]. In formula:

$$LR_{R_i} = \frac{P(R_i \mid Disease)}{P(R_i \mid No\ Disease)} \qquad \text{(eq 2)}$$

Taking into account the lifetime consequences of diagnostic testing and subsequent treatment the optimal likelihood ratio (LR_{opt}) depends on the prevalence of disease (p) and the ratio of the net loss due to false-positive test results compared with true-negative results ($NHB_{TN} - NHB_{FP}$) and the net loss due to false-negative results compared with true-positive results (NHB_{TP}-NHB_{FN}), in formula [15, 16]:

$$LR_{opt} > \frac{1-p}{p} * \frac{NHB_{TN} - NHB_{FP}}{NHB_{TP} - NHB_{FN}} \qquad \text{(eq 3)}$$

The optimal combination of sensitivity and specificity (the optimal operating point) was derived from the smooth ROC curve at the operating point where the result-specific likelihood ratio (eq 2) equals the optimal likelihood ratio based on the Markov model (eq 3). Subsequently, we selected the observed PSV value closest to the point on the smooth curve, which we define as the optimal threshold PSV.

For the summary ROC curve analysis we used the statistical package SPSS 11.0 and for the construction of the ROC curves we used Excel 2000.

Results

Study population

In the prospective study 350 patients were included having a mean age of 67 (range 39-88) and 76% being male [9]. The symptomatic carotid artery could be imaged in 323 patients with DSA and in 330 with DUS. Both tests were available in 313 patients.

In table 1 the categorized angiographic and duplex results are cross-tabulated. Angiographic stenosis of 0-49% stenosis was found in 14% of the patients, 50-69% stenosis in 19% of the patients, 70-98% stenosis in 41%, near occlusions (99% stenosis) in 5%, and occlusions in 20%. Duplex ultrasonography identified 61 occlusions. The sensitivity and specificity of diagnosing an occlusion by duplex ultrasound were 94% and 99% respectively. Slow flow in combination with a visualized severe stenosis on DUS was found in 16 arteries, indicating a near occlusion. In total 236 arteries without (near) occlusions on DUS were available for the analysis. In figure 1 (A and B) the absolute PSV measurements are plotted against the angiographic degree of stenosis. The figure shows that the relationship between the PSV and the degree of stenosis is non-linear. In figure A the indication for endarterectomy is set at a minimum angiographic stenosis of 70% and in figure B at 50% (horizontal lines). Shifting the PSV threshold (vertical lines) to the left will result in an inversely related increase of the true-positive test results and a decrease of true-negative test results. Thus, lowering the PSV threshold is associated with a higher sensitivity and lower specificity. Conversely, shifting the PSV threshold upwards will result in a lower sensitivity and higher specificity.

Markov Model

The long-term outcomes of the cost-effectiveness analysis (Table 2) showed that referral for carotid endarterectomy instead of medical therapy resulted in a loss in QALYs and increase in costs in patients with a 0-49% stenosis. A small gain in QALYs and slight cost-savings were obtained with endarterectomy in patients with a 50-69% stenosis and a relatively large gain in QALYs and large cost-savings in patients with a 70-99% stenosis (Table 2).

Recommended PSV thresholds

The Society of Radiologists in Ultrasound recommended the use of a PSV of 230 cm/s for the diagnosis of a 70-99% stenosis and a PSV of 125 cm/s for the diagnosis of a 50-99% stenosis [7]. Applying the recommended threshold of 230 cm/s in our population for diagnosing a 70-99% stenosis would result in a sensitivity of 95% and specificity of 51%. If we would apply a threshold of 125 cm/s to our population for detecting a 50-99% stenosis the sensitivity would be 99%, whereas the specificity would be 36%.

Table 1: Categorized measurements of the internal carotid artery: duplex versus digital subtraction angiography.

		DSA stenosis categories					
	PSV cm/s	0-49%	50-69%	70-98%	99%	Occlusion	Total
	0-124	16	1				17
DUS	125-230	20	17	5	1		43
categories	>230	8	42	117	8	1	176
	Near occl	1	1	5	6	3	16
	Occlusion			1	1	59	61
	Total	45	61	128	16	63	313

DSA, digital subtraction angiography; DUS, duplex ultrasonosgraphy; PSV, peak systolic velocity; Near occl, near occlusion

Table 2: Lifetime costs, quality-adjusted life years (QALYs), and net health benefits (NHB) for the base case with a positive or negative duplex ultrasonography result for several stenosis categories.[#]

	Stenosis category		
	0-49%	50-69%	70-99%
Negative test result: medical therapy			
Cost ($)	30,599	36,427	46,444
QALYs	11.36	11.09	10.71
NHB $25k	10.14	9.63	8.85
NHB $50k	10.75	10.36	9.78
Positve test result: carotid endarterectomy*			
Cost ($)	35,638	35,638	35,638
QALYs	11.12	11.12	11.12
NHB $25k	9.70	9.70	9.70
NHB $50k	10.41	10.41	10.41
Endarterectomy vs medical therapy			
Δ Cost ($)	5,139	-789	-10,806
Δ QALYs	-0.24	0.03	0.41
Δ NHB $25k	-0.44	0.06	0.85
Δ NHB $50k	-0.34	0.05	0.63

[#] Source: Buskens et al [8]. Modified data from table 3: data include treatment and follow-up for patients with TIA or stroke.

* Risks and prognosis following carotid endarterectomy were assumed to be independent of the underlying stenosis category.

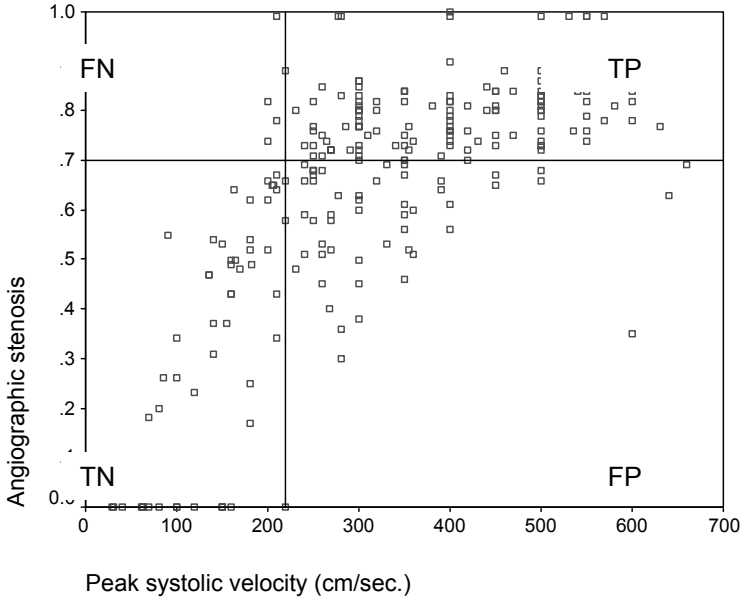
NHB $25k and NHB $50k indicate that the net health benefit was calculated with a willingness-to-pay threshold of $25,000 and $50,000 per QALY respectively.

Optimal PSV threshold

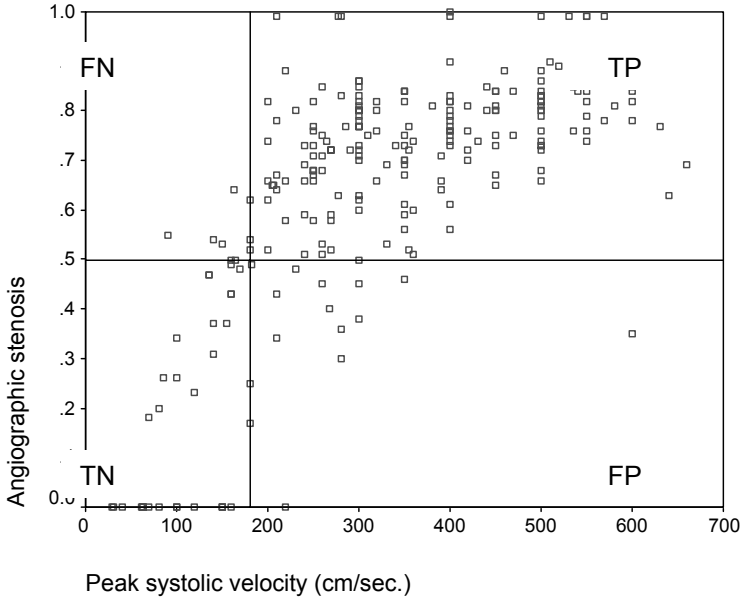
For a 70-99% stenosis indication for endarterectomy the optimal result-specific likelihood ratio based on the Markov model (calculated with equation 3) was 0.21. Figure 2A shows the observed ROC curve and the smooth ROC curve. Each dot on the observed curve represents the sensitivity and 1-specificity for a specific PSV threshold for predicting a 70-99% angiographic stenosis, i.e. eligible for endarterectomy. The optimal likelihood ratio or slope of the ROC curve is indicated in this figure. The associated optimal sensitivity and specificity were 97% and 49% respectively. The optimal threshold PSV that corresponded with this sensitivity and specificity was 220 cm/s (figure 2A). The optimal PSV threshold is also shown in figure 1A. The south-east quadrant of this figure represents the large number of false-positive test results as compared with the north-west quadrant which contains a small number of false-negative test results using the optimal PSV threshold.

If a 50-99% stenosis was used as indication for endarterectomy the optimal result-specific likelihood ratio was 0.38. The associated optimal sensitivity and specificity were 95% and 69% respectively (figure 2B). The optimal threshold PSV that corresponded with this sensitivity and specificity was 180 cm/s (figure 1B and 2B).

Changing the willingness-to-pay threshold from $25,000 to $50,000 per QALY hardly affected the optimal likelihood ratio and thus, had no effect on the optimal threshold PSV. If a higher willingness-to-pay threshold is accepted the net health benefit (eq 1) increases. The ratio of the difference in net health benefits associated with specific test results, however, only changed minimally if a higher willingness-to-pay threshold was used (eq 3).

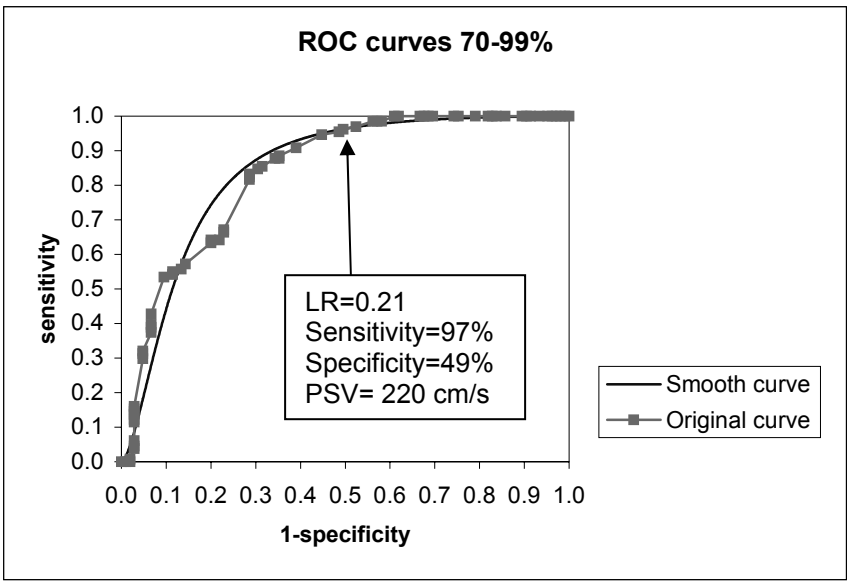


1A

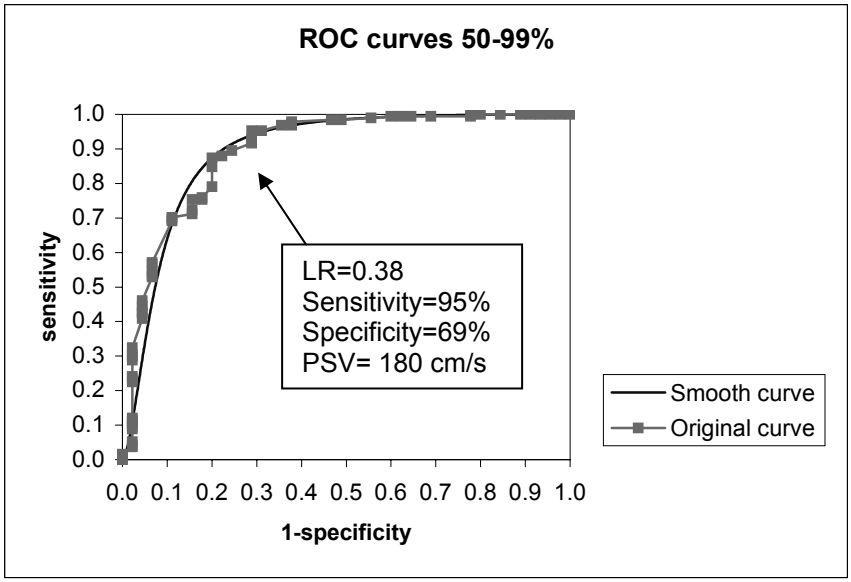


1B

Figure 1: Scatterplot of absolute PSV measurements and angiographic stenosis measurements. In figure 1A the threshold PSV (220 cm/s) for the indication to operate a 70-99% stenosis is shown and in figure 1B the threshold PSV (180cm/s) for the indication to operate a 50-99% stenosis. FN, false negative; TP, true positive; TN, true negative; FP, false positive



1A



1B

Figure 2: Observed and smooth receiver operating characteristic curves for (1A) the indication to operate a 70-99% stenosis and (1B) for the indication to operate a 50-99% stenosis. The optimal likelihood ratio (LR) or slope of the ROC curve is indicated with the associated sensitivity and specificity and the optimal threshold PSV.

Discussion

We found that the optimal threshold PSV to select symptomatic patients for carotid endarterectomy was 220 cm/s if the indication for endarterectomy was a 70-99% stenosis and 180 cm/s if the indication for endarterectomy was a 50-99% stenosis. The first result is close to the threshold recommended by the Society of Radiologists in Ultrasound (230 cm/s for a 70-99% stenosis), but the latter result is higher than recommended (125 cm/s for a 50-99% stenosis) [7].

The Society of Radiologists in Ultrasound based their recommendations on a literature review, focused on optimizing accuracy, and did not take into account the variable impact of false-negative test results as opposed to false-positive test results. In our study we determined the optimal threshold values based on the long-term consequences of false test results in costs and effects using a Markov model. The results of our model showed that referring a non-significant stenosis (<50%) for endarterectomy is more harmful than missing a diagnosis of 50-69% stenosis, which explains the fairly high PSV threshold for discriminating a 50-99% stenosis from a 0-49% stenosis. If we would use the recommended threshold of 125 cm/s in our population the sensitivity for selecting a 50-99% stenosis would increase to 99%, whereas the specificity would decrease dramatically to only 36%.

The natural history of the disease to be identified and the efficacy of treatment play a major role in determining the relative importance of sensitivity and specificity. Especially in high-grade carotid stenosis (>70%), undiagnosed severe carotid disease has a high cost in monetary terms and in life expectancy [8]. To identify such patients, DUS criteria should be highly sensitive, with a minimum number of false-negatives, since these patients carry a high excess morbidity and mortality if left untreated. Therefore, in identifying patients with a 70-99% stenosis a high sensitivity of the diagnostic test is more important than a high specificity.

In identifying patients with a 50-99% stenosis suitable for endarterectomy a relatively high number of patients will benefit from the procedure. However, the overall losses associated with missing a diagnosis are smaller, because most of these patients have a 50-69% stenosis for which the missed benefit is smaller than for a 70-99% stenosis. This led to a somewhat lower optimal sensitivity and a higher optimal specificity as compared with identifying a 70-99% stenosis.

The prior probability in the population being evaluated also plays a critical role in defining the optimal test criterion. Ideally, if the ROC analysis were performed on measurements in unselected carotid endarterectomy candidates, the prior probability would be known. In our study population the prevalence of 70-99% stenoses was 46% and the prevalence of 50-99% stenoses was 66%. If the prior probability would be lower the derived slope would be steeper (eq 3) and the optimum cut point on the ROC curve would shift to the left, implying lower sensitivity, higher specificity, and thus a higher threshold PSV.

Varying the societal willingness-to-pay threshold from $25,000 to $50,000 per QALY did not influence the choice of the optimal PSV threshold in our study. Although the net health benefit for a particular test result changed when varying the WTP threshold, the ratio of the differences in net health benefits in equation 3 hardly changed. Note, however, that this does not necessarily always apply. In fact, the optimal likelihood ratio can either increase or decrease, depending on the proportions of true and false test results in a particular test situation, which can lead to both

a higher or lower threshold of the test under investigation in other studies when increasing the WTP-threshold.

Various methods have been reported in the literature to determine the optimal diagnostic cut point on the receiver operating characteristics curve. Some investigators have used the point on the curve that is closest to the upper left corner of the ROC curve as 'optimal' cut point [17] or the cut point with a likelihood ratio equal to 1 [18]. Other criteria for determining the optimal cut point include selection of the Q-point: the point where sensitivity equals specificity [19], maximizing accuracy or the sum of sensitivity plus specificity [20, 21], or accepting a preset level of sensitivity or specificity and determining the corresponding operating characteristic [22]. All these methods minimize the number of false-positive and false-negative test results. However, none of these methods takes into consideration the prevalence of disease or the consequences of correctly or incorrectly classifying a test result as positive or negative in terms of costs and quality of life.

The decision analytic approach of determining the optimal diagnostic threshold that we applied was described many years ago [15]. Practical applications of this method, however, are scarce. In carotid artery disease publications we found only two studies that based their optimal test criteria on patient outcome rather than test accuracy [23, 24]. Wilterdink and coworkers based their criteria on 2-year mortality and morbidity of a severe stenosis treated medically versus surgically as reported in NASCET. They found a slope or optimal likelihood ratio of 0.09, which is more lenient than the slope that we found, which implies a higher sensitivity and lower specificity [23]. The difference can be ascribed to the fact that they used DUS to select patients for angiography, whereas we used DUS to select patients for endarterectomy. The harm of unnecessary angiography is much smaller than the harm of unnecessary endarterectomy in false-positives, whereas false-negative test results lead in both studies to a missed opportunity of reducing events by performing endarterectomy, which explains the lower slope compared with our results. Furthermore, we used the updated results from the NASCET study in our model, which show a small but significant benefit for patients with a 50-69% stenosis, we integrated both cost and effects on life expectancy, and we modeled lifetime outcomes.

Kuntz et al. chose the PSV cut point that minimized the probability of stroke at 2 years for symptomatic patients. They found an optimal threshold PSV of 229 cm/s for one laboratory and an optimal threshold PSV of 340 cm/s for another laboratory [24]. Unfortunately, they did not report the optimal likelihood ratio or slope of the ROC curve. We did not evaluate the potential differences between hospitals. However, we are aware that the optimal threshold PSV may depend on hospital equipment. If, however, hospitals do have data on angiography and PSV measurements they could construct their own ROC curve and determine their own optimal threshold using the optimal slope from our decision analysis.

In conclusion, based on the lifetime consequences of diagnostic testing and subsequent treatment, the optimal threshold for the diagnosis of a 70-99% carotid artery stenosis is a PSV of 220 cm/s and for a 50-99% stenosis the optimal threshold PSV is 180 cm/s in patients with symptoms of amaurosis fugax, TIA, or minor stroke if duplex ultrasonography is used for referring patients to carotid endarterectomy.

References

1. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. N Engl J Med. 1991;325:445-453.
2. MRC European Carotid Surgery Trial: interim results for symptomatic patients with severe (70-99%) or with mild (0-29%) carotid stenosis. European Carotid Surgery Trialists' Collaborative Group. Lancet. 1991;337:1235-1243.
3. Barnett HJ, Taylor DW, Eliasziw M, et al. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. N Engl J Med. 1998;339:1415-1425.
4. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). Lancet. 1998;351:1379-1387.
5. Rothwell PM, Gutnikov SA, Warlow CP. Reanalysis of the final results of the European Carotid Surgery Trial. Stroke. 2003;34:514-523.
6. Rothwell PM, Eliasziw M, Gutnikov SA, et al. Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. Lancet. 2003;361:107-116.
7. Grant EG, Benson CB, Moneta GL, et al. Carotid artery stenosis: gray-scale and Doppler US diagnosis--Society of Radiologists in Ultrasound Consensus Conference. Radiology. 2003;229:340-346.
8. Buskens E, Nederkoorn PJ, Buijs-Van Der Woude T, et al. Imaging of carotid arteries in symptomatic patients: cost-effectiveness of diagnostic strategies. Radiology. 2004;233:101-112.
9. Nederkoorn PJ, Mali WP, Eikelboom BC, et al. Preoperative diagnosis of carotid artery stenosis: accuracy of noninvasive testing. Stroke. 2002;33:2003-2008.
10. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998;18:S68-80.
11. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology. 2003;229:3-8.
12. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med. 1993;12:1293-1316.
13. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making. 1993;13:313-321.
14. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med. 1994;120:667-676.
15. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8:283-298.
16. Hunink MGM, Glasziou PP. Multiple test results. In: Decision making in health and medicine. Integrating evidence and values. 1st ed. Cambridge: Cambridge University Press; 2001.
17. Criswell BK, Langsfeld M, Tullis MJ, Marek J. Evaluating institutional variability of duplex scanning in the detection of carotid artery stenosis. Am J Surg. 1998;176:591-597.
18. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000;45:23-41.
19. Westwood ME, Kelly S, Berry E, et al. Use of magnetic resonance angiography to select candidates with recently symptomatic carotid stenosis for surgery: systematic review. BMJ. 2002;324:198.
20. Carsten CG, 3rd, Elmore JR, Franklin DP, Thomas DD, Mordan F, Wood GC. Use of limited color-flow duplex for a carotid screening project. Am J Surg. 1999;178:173-176.
21. AbuRahma AF, Robinson PA, Strickler DL, Alberts S, Young L. Proposed new duplex classification for threshold stenoses used in various symptomatic and asymptomatic carotid endarterectomy trials. Ann Vasc Surg. 1998;12:349-358.
22. Fillinger MF, Baker RJ, Jr., Zwolak RM, et al. Carotid duplex criteria for a 60% or greater angiographic stenosis: variation according to equipment. J Vasc Surg. 1996;24:856-864.

23. Wilterdink JL, Feldmann E, Easton JD, Ward R. Performance of carotid ultrasound in evaluating candidates for carotid endarterectomy is optimized by an approach based on clinical outcome rather than accuracy. Stroke. 1996;27:1094-1098.
24. Kuntz KM, Polak JF, Whittemore AD, Skillman JJ, Kent KC. Duplex ultrasound criteria for the identification of carotid stenosis should be laboratory specific. Stroke. 1997;28:597-602.

8

Diagnostic performance of duplex ultrasound in patients suspected of carotid artery disease: the symptomatic versus asymptomatic artery

Abstract

Objective: To evaluate duplex ultrasonographic thresholds for the determination of 50-99% and 70-99% stenosis of the symptomatic and asymptomatic internal carotid artery in patients with symptoms of amaurosis fugax, transient ischemic attack (TIA), or minor stroke based on two criteria: maximizing accuracy and optimizing cost-effectiveness and to compare these with current recommendations.

Methods: From January 1997 to January 2000, a prospective multicenter study was conducted including 350 consecutive patients with symptoms of amaurosis fugax, TIA, or minor stroke, who underwent bilateral duplex ultrasonography and digital subtraction angiography. In total 504 non-occluded carotid arteries were available for evaluation by both imaging modalities. A linear regression analysis was performed that estimated the degree of angiographic stenosis as a function of the peak systolic velocity (PSV). Potential differences between the symptomatic and asymptomatic arteries were evaluated. PSV thresholds were calculated for both the symptomatic and asymptomatic carotid arteries based on two criteria: maximizing accuracy and optimizing outcome in terms of cost-effectiveness.

Results: The PSV measurements significantly overestimated the angiographic stenosis in the asymptomatic artery (9.5%; 95% CI 6.3-12.7%) compared with the symptomatic carotid artery. The recommended PSV threshold for the diagnosis of 70-99% stenosis is 230 cm/s. If separate PSV thresholds were used for the symptomatic and asymptomatic arteries the accuracy increased from 75.8% to 78% for symptomatic arteries and from 87.3 to 94% for asymptomatic arteries. Maximizing accuracy the optimal PSV threshold for the symptomatic artery was 280cm/s and for the asymptomatic artery 370 cm/s for diagnosing a 70-99% stenosis. Optimizing cost-effectiveness, the optimal PSV threshold was 220 cm/sec for symptomatic and 290 cm/sec for asymptomatic carotid arteries.

For the diagnosis of 50-99% stenosis the recommended PSV threshold was 125 cm/s. Maximizing accuracy, the optimal PSV threshold was 160 cm/s for symptomatic and 190 cm/s for asymptomatic arteries. Optimizing cost-effectiveness, the optimal PSV threshold was 180 cm/s for symptomatic and 230 for asymptomatic arteries.

Conclusions: PSV measurements overestimate the degree of angiographic stenosis in the asymptomatic carotid artery in patients with symptoms of amaurosis fugax, TIA, or minor stroke. PSV thresholds that optimize cost-effectiveness differ from the recommended thresholds and from thresholds that maximize accuracy.

Introduction

Duplex ultrasonography is currently the principal non-invasive test for evaluating carotid artery disease. Screening for this disease has become particularly important since publication of the results from the North American Symptomatic Carotid Endarterectomy Trial (NASCET) and the European Carotid Surgery Trial (ECST).[1-5] These large randomized trials demonstrated that carotid endarterectomy is not only beneficial in symptomatic patients with severe stenosis (70-99%), but also in moderate stenosis (50-69%).

The reference standard for detection of significant carotid artery stenosis is digital subtraction angiography. Because of the risks and cost associated with this invasive procedure, there is a trend towards relying more on non-invasive imaging techniques for the diagnosis of carotid artery disease, such as duplex ultrasonography. Since the publication of the NASCET and ECST results, many studies aimed at determining optimal ultrasonographic criteria for identifying an angiographic carotid artery stenosis of 70-99% or 50-99%. However, the reported threshold values differ. For example, the optimal threshold value of the peak systolic velocity (PSV) has been reported at 175 cm/s, 200 cm/s, and 230 cm/s[6-8] for the diagnosis of a 70-99% stenosis and at 130 cm/s and 140 cm/s[8,9] for the diagnosis of a 50-99% stenosis. Based on a literature review, a recent consensus study recommended the use of a PSV of 230 cm/s in the internal carotid artery as the optimal threshold for the diagnosis of a 70-99% stenosis and a PSV of 125 cm/s for a 50-99% stenosis.[10] These studies did not, however, evaluate potential differences between thresholds for the asymptomatic and symptomatic artery in symptomatic patients. Furthermore, various criteria were used in the reviewed studies to determine the optimal threshold value. Most studies maximized diagnostic accuracy. The consequences of a false-negative test, however, are probably more harmful and more expensive than the consequences of a false-positive test. Therefore, it is more clinically relevant to use clinical outcome and cost as criteria for the determination of the optimal threshold value for diagnosing a significant stenosis. In a previous study we calculated the optimal PSV threshold for the symptomatic carotid artery based on a cost-effectiveness analysis, which resulted in a PSV threshold of 220 cm/s for a 70-99% stenosis and 180 cm/s for a 50-99% stenosis.[11]

The aim of this study was to evaluate potential differences in duplex ultrasonographic thresholds between the asymptomatic and symptomatic internal carotid artery for the determination of 50-99% or 70-99% stenosis in patients with symptoms of amaurosis fugax, transient ischemic attack (TIA), or minor stroke based on two criteria: maximizing accuracy and optimizing cost-effectiveness and to compare these thresholds with current recommendations.

Methods

Study population

A prospective diagnostic study was performed in two academic hospitals and one non-academic hospital in the Netherlands from January 1997 to January 2000.[12] After informed consent was obtained, patients with symptoms of amaurosis fugax, transient ischemic attack, or minor stroke underwent bilateral carotid duplex ultrasonography and carotid digital subtraction angiography within a time frame of four weeks. Clinical data were collected on symptoms,

neurological examinations and test results. Patients were enrolled in the University Medical Center Utrecht, Erasmus University Medical Center Rotterdam, and Medical Spectrum Twente. The medical ethical committee approved the study for each hospital.

Carotid duplex ultrasonography

The peak systolic velocity was measured bilaterally in the proximal part of the internal carotid artery. A carotid artery was recorded as symptomatic if its perfusion territory was associated with the symptoms of amaurosis fugax, TIA, or minor stroke. If no detectable flow was present patients were classified as having an occlusion (100% stenosis). Slow flow in combination with a visualized severe stenosis was defined as a near occlusion. Arteries with a (near) occlusion on duplex ultrasonography were excluded from the analysis, because the peak systolic velocity is absent or not reliably measurable in these arteries. The hospital in which the patient was tested was recorded, because the results of duplex ultrasonography may depend on the equipment used, local imaging protocols, and the image-processing software used.

Digital subtraction angiography (DSA)

DSA was performed by selective positioning of an intra-arterial catheter sequentially in each common carotid artery. Each carotid bifurcation was imaged in three projections (lateral, posteroanterior, and oblique).

The DSA test results were read by one observer in each hospital. The observers were blinded for clinical information and for the duplex results. Printed hard copies were used by the observers to read the DSA images. The percentage stenosis was determined according to the North American Symptomatic Carotid Endarterectomy Trial (NASCET) criteria.[1] The degree of stenosis was defined as the percentage reduction in lumen diameter comparing the narrowest diameter of the residual lumen at the site of the stenosis to the normal lumen distal to the stenosis. The maximum degree of stenosis on the three projections was used in the analysis. DSA was considered the standard of reference.

Linear regression analysis

The absolute PSV values obtained with duplex ultrasonography were compared with the percentage stenosis measured with DSA, the reference standard, using linear regression analysis. The relationship between the absolute PSV and the percentage angiographic stenosis was non-linear. Therefore, the PSV was transformed to its natural logarithm (ln), for which a linear relationship with the degree of angiographic stenosis was found. Three linear models were assessed. In the first model the degree of angiographic stenosis was modeled as a function of the ln(PSV). In the second model, potential differences between the asymptomatic and symptomatic carotid artery were evaluated by adding a dummy variable to the first model.

In formula:

$$\text{Angiographic stenosis} = \beta_0 + \beta_1 * \ln(\text{PSV}) + \beta_2 * \text{S} \qquad \text{(eq. 1)}$$

in which the angiographic degree of stenosis is the dependent variable, β_0 is the intercept, β_1 and β_2 are regression coefficients, ln(PSV) is the natural logarithm of the PSV, and S is the

dummy variable which takes the value 1 for symptomatic arteries and 0 for asymptomatic arteries.

In the third model, variables were added to adjust for potential confounders, such as hospital, age, and sex. Variables were kept in the model if the model performance improved significantly, which was assessed by testing the change in R squared (F-test), and if the regression coefficient of the added variable had a significance level of $p<0.05$. The statistical package SPSS 11.0 was used for the regression analysis.

Receiver operating characteristic analysis

Receiver operating characteristic (ROC) analysis was performed to calculate sensitivities and specificities associated with each observed PSV result using DSA as the reference standard for the diagnosis of a 70-99% and a 50-99% stenosis. Separate ROC curves were drawn for the asymptomatic arteries and the symptomatic arteries. The statistical package SPSS 11.0 was used for the ROC analysis.

Peak systolic velocity thresholds

The recommended PSV thresholds of 230 cm/s for the diagnosis of a 70-99% stenosis and 125 cm/s for a 50-99% stenosis were applied to our study data. For both PSV thresholds, we calculated the associated sensitivity, specificity, and accuracy for the symptomatic and the asymptomatic carotid arteries separately.

We subsequently calculated our own PSV thresholds using two approaches: 1) maximizing overall accuracy and 2) optimizing outcome in terms of cost-effectiveness. The first approach, in which accuracy is maximized, assumes that a false-positive test result has the same importance as a false-negative test result. In the second approach we used outcomes from a previous study in which we calculated the optimal threshold PSV based on the prevalence of disease, the consequences of false-positive results relative to true-negative results, and the consequences of false-negative results relative to true-positive results in terms of costs and quality-adjusted life-years saved.[11] The consequences of the test results were based on a cost-effectiveness analysis which modeled lifetime outcome as a result of diagnostic testing and subsequent treatment of the symptomatic carotid artery in patients with amaurosis fugax, TIA, or minor stroke.[13] Our previous analysis resulted in an optimal threshold PSV of 220 cm/s for the diagnosis of 70-99% stenosis and 180 cm/s for the diagnosis of 50-99% stenosis for symptomatic arteries.[11] In the current study, we calculated the optimal threshold PSV for asymptomatic arteries by first substituting the optimal PSV of 220 cm/s and 180 cm/s respectively in equation 1. We assumed that the angiographic thresholds for symptomatic and asymptomatic arteries were equivalent. Subsequently, we derived the optimal PSV threshold for the asymptomatic arteries by rearranging equation 1:

$$\text{Threshold PSV}_{\text{asymptomatic artery}} = \exp\left(\frac{\text{Angiographic stenosis} - \beta_0}{\beta_1}\right) \qquad \text{(eq. 2)}$$

The sensitivities and specificities associated with each threshold PSV value were derived from the ROC analysis. In addition, the accuracy was calculated for each PSV threshold. The

outcomes were compared with the recommended threshold PSV values of 230 cm/s and 125 cm/s for 70-99% and 50-99% stenosis respectively.

Table 1: Linear regression models for the relationship between the degree of angiographic stenosis and the peak systolic velocity (PSV) with and without adjustment for asymptomatic and symptomatic arteries, hospital, age, and sex

	Model 1	Model 2	Model 3
Coefficients (95% confidence intervals)			
Intercept	-1.538	-1.410	-1.461
	(-1.638; -1.439)	(-1.515; -1.304)	(-1.567; -1.354)
Ln(PSV)	0.379	0.346	0.355
	(0.361; 0.398)	(0.325; 0.368)	(0.333; 0.376)
Symptomatic artery		0.095	0.091
		(0.063; 0.127)	(0.060; 0.122)
Hospital C vs A, B			0.093
			(0.049; 0.137)
Male sex			0.031
			(0.001; 0.061)
Age			0.0015
			(0.000; 0.003)
Model performance			
R^2	0.760	0.775	0.786
R^2 change	0.760	0.015	0.011
Sig. F change	p=0.000	p=0.000	p=0.000

Ln(PSV), natural logarithm of the PSV; R^2, R squared; Sig. F change, significance of F-test for model change

Results

Study population

In the prospective study 350 patients were included (mean age: 67 years (range 39-88); 76% male; 24% female).[12] The carotid arteries could be imaged in 323 patients with DSA and in 330 with duplex ultrasonography. Both tests were available for 621 arteries.

Duplex ultrasonography identified 99 occlusions. The sensitivity and specificity of diagnosing an occlusion by duplex ultrasound were 95% and 99% respectively. Slow flow in combination with a visualized severe stenosis was found in 18 arteries, indicating a near occlusion. The analysis was performed on all arteries without (near) occlusions on duplex ultrasonography (n=504). Angiographic stenosis of 0-49% stenosis was found in 239 arteries (47%), 50-69% stenosis in 95 arteries (19%), 70-99% in 169 arteries (34%), and 100% stenosis in 1 artery (0.2%).

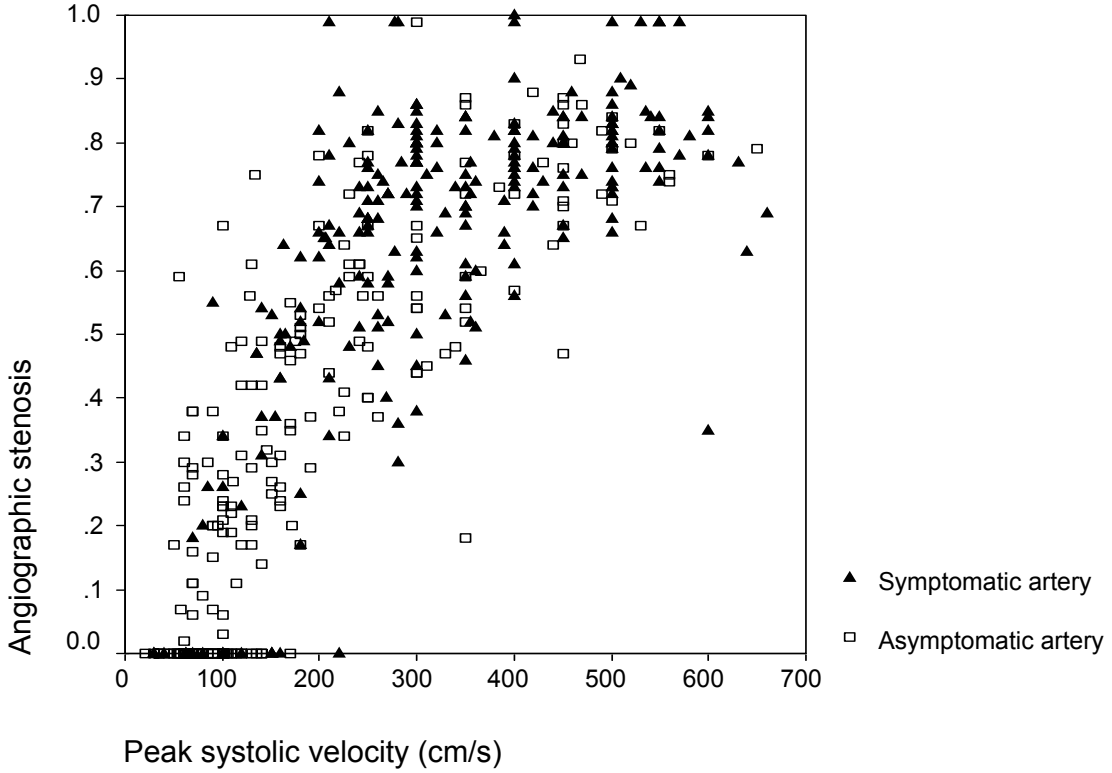


Figure 1: Scatterplot of peak systolic velocity and angiographic stenosis for symptomatic and asymptomatic carotid arteries

Linear regression analysis

Figure 1 demonstrates the distribution of the angiographic stenosis as a function of the PSV. The figure shows that the relationship between angiographic stenosis and the PSV is non-linear. Therefore the natural logarithm of the PSV was used in the regression analysis, which shows a good linear relationship with angiographic stenosis (Table 1, Model 1: R^2=0.76). In model 2 we added a variable to the regression model indicating whether the perfusion territory of the carotid artery was associated with symptoms of amaurosis fugax, TIA, or minor stroke. The angiographic stenosis was significantly more severe in the symptomatic artery than in the asymptomatic artery for a given peak systolic velocity (Table 1, Model 2: regression coefficient 0.095; 95% confidence interval (CI) 0.063-0.127). In other words, the PSV was significantly higher in asymptomatic arteries than in symptomatic arteries for a given degree of angiographic stenosis (Figure 2). In model 3, we added variables to the regression model in order to adjust for potential confounders, such as differences between hospitals, age, and sex. Although these variables significantly influenced the degree of angiographic stenosis, the difference between symptomatic and asymptomatic arteries remained the same (Table 1, Model 3: regression coefficient 0.091; 95%CI: 0.060-0.122). Therefore, we used Model 2 to determine the optimal PSV thresholds for symptomatic and asymptomatic arteries.

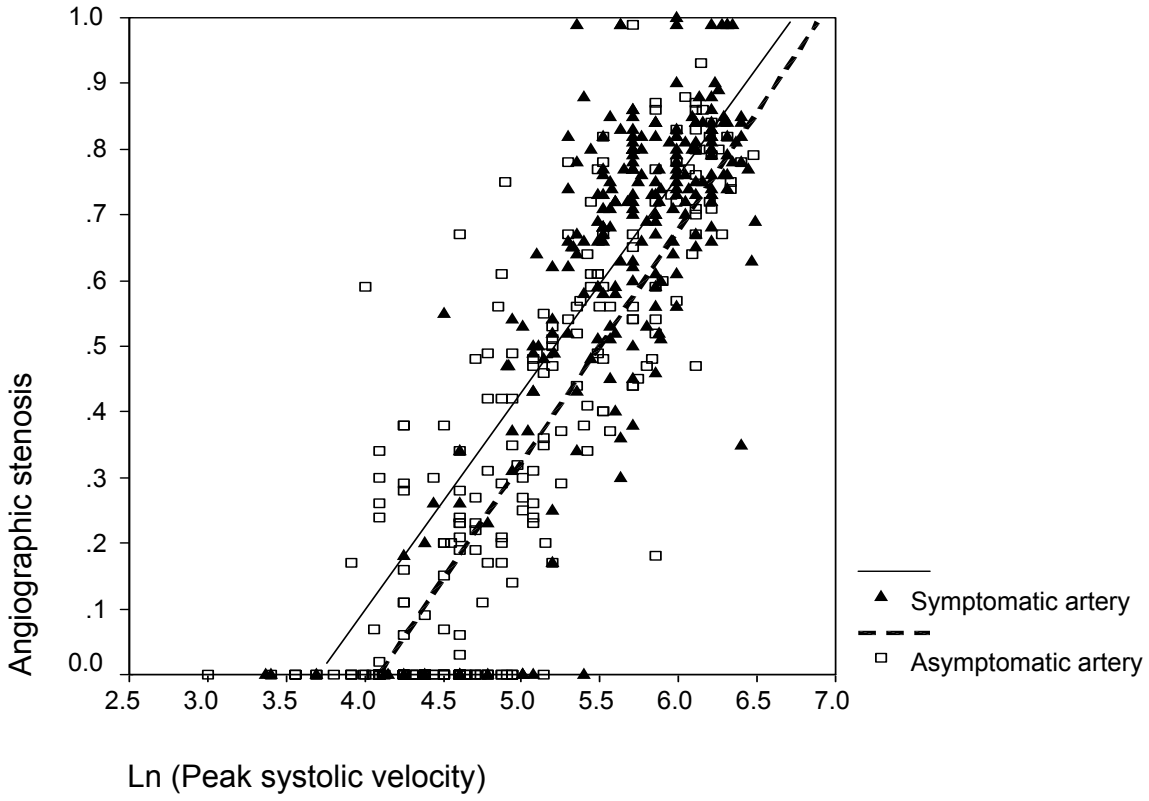
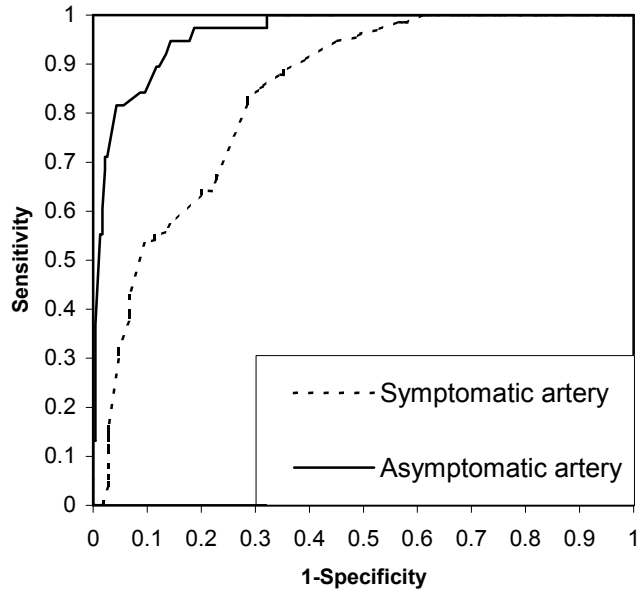


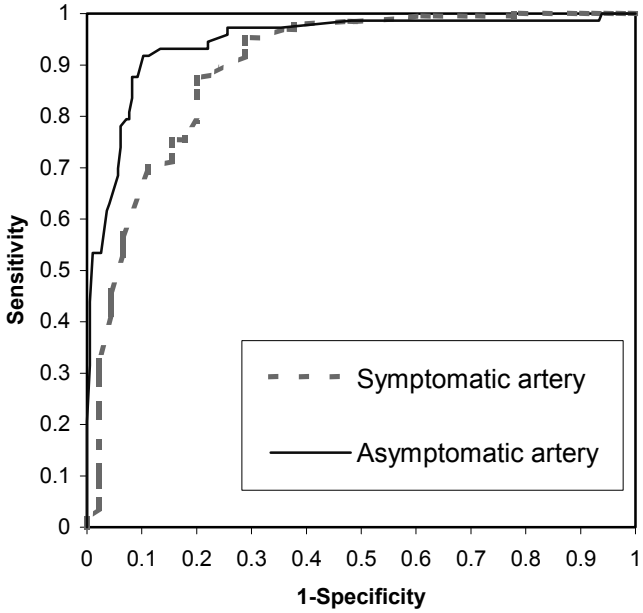
Figure 2: Scatterplot of the natural logarithm of the peak systolic velocity and angiographic stenosis. Regression lines indicate relationships for symptomatic and asymptomatic arteries.

ROC analysis

In Figure 3 the ROC curves are presented for the symptomatic and asymptomatic arteries for A) the diagnosis of a 70-99% stenosis and B) the diagnosis of a 50-99% stenosis. The figures demonstrate that the area under the ROC curves is larger for the asymptomatic arteries than for the symptomatic arteries, indicating a higher diagnostic performance of duplex ultrasonography in asymptomatic arteries than in symptomatic arteries. This is a result of the large error range in PSV measurements in high-grade stenosis, which is primarily found in the symptomatic arteries, as compared with the relatively small error range in low-grade stenosis, which is mainly found in the asymptomatic arteries (Figure 1). For the same reason, lowering the diagnostic threshold from 70-99% to 50-99% stenosis resulted in a larger area under the ROC curve, and thus an increase in diagnostic performance of duplex ultrasonography in the symptomatic arteries (Figure 3B).



3A



3B

Figure 3: Receiver operating characteristic curves for PSV measurements in the symptomatic and asymptomatic arteries for prediction of A) a 70-99% stenosis and B) a 50-99% stenosis.

Table 2: Thresholds of the peak systolic velocity (PSV) and corresponding sensitivity, specificity, and accuracy for the diagnosis of 70-99% stenosis

Threshold definition	PSV threshold (cm/s)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Recommended threshold [10]				
Symptomatic artery	230	95.4	51.4	75.8
Asymptomatic artery	230	92.1	86.5	87.3
Maximum accuracy				
Symptomatic artery	280	84.7	69.5	78.0
Asymptomatic artery	370	71.1	97.8	94.0
Optimal cost-effectiveness				
Symptomatic artery	220	96.2	55.5	75.8
Asymptomatic artery	290	84.2	91.4	90.3

Table 3: Thresholds of the peak systolic velocity (PSV) and corresponding sensitivity, specificity, and accuracy for the diagnosis of 50-99% stenosis

Threshold definition	PSV threshold (cm/s)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Recommended threshold [10]				
Symptomatic artery	125	99.5	35.6	87.3
Asymptomatic artery	125	95.9	74.4	80.2
Maximum accuracy				
Symptomatic artery	160	97.9	62.2	91.1
Asymptomatic artery	190	87.7	91.8	90.7
Optimal cost-effectiveness				
Symptomatic artery	180	95.3	68.9	90.3
Asymptomatic artery	235	74.0	93.9	88.4

Box 1. Optimal peak systolic velocity (PSV) thresholds (cm/s) based on cost-effectiveness criteria.

	Optimal PSV threshold (cm/s)	
	70-99% stenosis	50-99% stenosis
Symptomatic artery	220	180
Asymptomatic artery	290	235

PSV threshold analysis

In Tables 2 and 3 the results of the threshold analyses are presented for the diagnosis of a 70-99% stenosis and a 50-99% stenosis respectively. For both indications for carotid endarterectomy, the recommended PSV thresholds are highly sensitive and less specific for the symptomatic as well as the asymptomatic arteries.

If maximization of accuracy is used as criterion for the determination of the PSV threshold, then specificity becomes more important than sensitivity in asymptomatic arteries. The accuracy increased with 7 percentage points for diagnosing a 70-99% stenosis and with 11 percentage points for diagnosing a 50-99% stenosis in the asymptomatic arteries.

Using cost-effectiveness criteria for the determination of the optimal PSV threshold resulted in a high sensitivity for the evaluation of the symptomatic arteries and a high specificity for the evaluation of the asymptomatic arteries. For the symptomatic arteries, the optimal threshold was a PSV of 220 cm/s for the diagnosis of a 70-99% stenosis and 180 cm/s for a 50-99% stenosis (Box 1). For the asymptomatic side, the optimal PSV thresholds were 290 cm/s and 235 cm/s for diagnosing a 70-99% and 50-99% stenosis respectively.

Discussion

This study shows that the estimated angiographic stenosis in asymptomatic carotid arteries is overestimated if criteria are used based on the PSV measurements of symptomatic arteries. For a given degree of angiographic stenosis the PSV measurements were consistently higher in asymptomatic arteries than in symptomatic arteries. As a result, we found that the optimal PSV threshold for identifying 70-99% or 50-99% angiographic stenosis varied between asymptomatic and symptomatic arteries. Specifically, the optimal PSV threshold in asymptomatic carotid arteries was higher than in symptomatic arteries. Similar observations were made by Fujitani et al. and AbuRahma et al. for the peak systolic frequency in arteries with contralateral severe stenosis or occlusions.[14,15] The physiologic mechanism for this observation has been referred to as the compensatory flow phenomenon, which explains that flow would have to be increased in the artery with contralateral stenosis to maintain the same level of blood volume to the brain as before the stenosis emerged.[16] Since this increased volume of blood must pass through a vessel with a relatively fixed cross sectional area, the velocity of the blood must increase. This has been illustrated in studies that evaluated changes in peak systolic velocity in arteries before and after contralateral carotid endarterectomy[16-19] and carotid stenting.[20] In these studies the PSV measured in the asymptomatic artery significantly decreased after contralateral endarterectomy or stenting, which often led to reclassification of stenosis in the asymptomatic artery. These findings suggested that when duplex ultrasonography is used as the sole imaging modality before carotid endarterectomy, patients with severe bilateral carotid stenosis must have an additional duplex examination before operation on the second side. Our study shows that over the whole range of angiographic stenosis there is a consistent difference in peak systolic velocity for symptomatic versus asymptomatic arteries and therefore suggests that separate PSV

thresholds should be used for the diagnosis of significant stenosis in symptomatic and asymptomatic arteries in patients with symptoms of amaurosis fugax, TIA, or minor stroke.

There are several limitations of our study. First, we focused on the presence or absence of amarousis fugax, TIA, or minor stroke in the perfusion territory of the carotid artery to define whether the artery was symptomatic or not. We did not take into account the severity of disease or the presence of actual stenosis. Several studies have shown that the hemodynamic effect is more pronounced if the contralateral artery is severely stenosed or occluded.[14,15,17] However, since we do not know the actual severity of the contralateral stenosis at the time of duplex ultrasonography we cannot reliably adjust with the actual percentage. Our results show that the presence of contralateral symptoms of amaurosis fugax, TIA, or stroke, has an equivalent effect on stenosis estimation to that described for the presence of contralateral stenosis.

Second, we assumed that the angiographic thresholds for the symptomatic and asymptomatic arteries were the same, implying that the indication for carotid endarterectomy (70-99% stenosis or 50-99% stenosis) was the same for both arteries. Recently, the results of the Asymptomatic Carotid Surgery Trial (ACST) showed that immediate carotid endarterectomy in asymptomatic patients with a more than 70% stenosis on ultrasound, halved the 5-year stroke risk from 12% to about 6%.[21] This study included 24% of patients with previous contralateral symptoms and carotid endarterectomy. The Asymptomatic Carotid Artery Study showed similar results in asymptomatic patients with a more than 60% stenosis, also including 20% of patients with previous symptoms and contralateral endarterectomy.[22] The indication for carotid endarterectomy for 50-99% stenosis, however, has not been evaluated in asymptomatic arteries.

Third, verification bias may have influenced our results. Verification bias may exist if patients are referred to the reference test based on the results of the test under investigation. Sensitivity may be inflated and specificity deflated if selection for the reference standard takes place. This may have been the case in our study for the symptomatic arteries: only patients with a clinical indication for carotid endarterectomy and a high PSV were included in the study. Indeed, we did see a high sensitivity and low specificity in the symptomatic arteries. By including both the symptomatic and the asymptomatic carotid artery of each patient in our analysis, representing the whole spectrum of arterial stenoses, we partially corrected for verification bias. We showed that sensitivity decreased and specificity increased, which is equal to the effect of mathematically correcting for verification bias based on the probability of verification.[23] However, our results may not be applicable for patients without a clinical indication for carotid endarterectomy, because these patients were not included in this study.

Finally, several studies have shown that different laboratories should establish their own thresholds[24-26] and update them regularly.[27] PSV measurements tend to differ due to differences in manufacturers or technologists. In our study we also observed that one hospital measured consistently higher PSV values than the other two hospitals. However, when corrected for type of hospital the difference between asymptomatic and symptomatic arteries remained significant. Therefore, our results show that if laboratories establish their own laboratory specific duplex criteria the side of carotid disease should be taken into account.

The recently published recommendations by the Society of Radiologists in Ultrasound advocate the use of a threshold of 230 cm/sec for determination of a 70-99% stenosis and 125

cm/s for a 50-99% stenosis.[10] Application of these criteria to our data resulted in a high sensitivity, a moderate specificity and good overall diagnostic accuracy. We used two criteria for selecting the optimal threshold for the peak systolic velocity: 1) maximization of diagnostic accuracy and 2) optimization of outcome in terms of cost-effectiveness. In the first case, the goal is to minimize the number of false-positive and false-negative results. This approach assumes that a false-negative result has the same importance as a false-positive result. The second approach takes into account that missing a diagnosis may be more (or less) harmful than further work-up in a patient without disease. Although many of the studies that define duplex thresholds use accuracy maximization as their goal, an approach that balances the clinical consequences of missed diagnoses and unnecessary endarterectomies is clinically more relevant than an approach that maximizes accuracy. Because the majority of asymptomatic arteries have a non-significant stenosis, a highly specific threshold is clinically more relevant for the asymptomatic arteries compared with the symptomatic arteries, for which a highly sensitive threshold is desirable.

This study showed that accuracy and the area under the ROC curve are higher for asymptomatic than for symptomatic arteries. Furthermore, our results also demonstrated that duplex ultrasonography is more accurate in diagnosing 50-99% stenosis than in diagnosing 70-99% stenosis. Both these findings can be explained by the smaller error range in PSV measurements in less severe stenosis.

In conclusion, PSV measurements overestimate the degree of angiographic stenosis in the asymptomatic carotid artery in patients with symptoms of amaurosis fugax, TIA, or minor stroke. Therefore, separate PSV thresholds should be used for the evaluation of the symptomatic and asymptomatic carotid arteries by duplex ultrasound. Optimal PSV thresholds, based on cost-effectiveness criteria, for diagnosing a 70-99% stenosis are 220 cm/s for symptomatic arteries and 290 cm/s for asymptomatic arteries. For the diagnosis of a 50-99% stenosis optimal PSV thresholds are 180 cm/s for symptomatic and 235 cm/s for asymptomatic arteries.

References

1. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. N Engl J Med. 1991;325:445-453.
2. MRC European Carotid Surgery Trial: interim results for symptomatic patients with severe (70-99%) or with mild (0-29%) carotid stenosis. European Carotid Surgery Trialists' Collaborative Group. Lancet. 1991;337:1235-1243.
3. Barnett HJ, Taylor DW, Eliasziw M, et al. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. N Engl J Med. 1998;339:1415-1425.
4. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). Lancet. 1998;351:1379-1387.
5. Rothwell PM, Eliasziw M, Gutnikov SA, Warlow CP, Barnett HJ. Endarterectomy for symptomatic carotid stenosis in relation to clinical subgroups and timing of surgery. Lancet. 2004;363:915-924.
6. Grant EG, Duerinckx AJ, El Saden S, et al. Doppler sonographic parameters for detection of carotid stenosis: is there an optimum method for their selection? AJR Am J Roentgenol. 1999;172:1123-1129.
7. Koga M, Kimura K, Minematsu K, Yamaguchi T. Diagnosis of internal carotid artery stenosis greater than 70% with power Doppler duplex sonography. AJNR Am J Neuroradiol. 2001;22:413-417.
8. Huston J, 3rd, James EM, Brown RD, Jr., et al. Redefined duplex ultrasonographic criteria for diagnosis of carotid artery stenosis. Mayo Clin Proc. 2000;75:1133-1140.
9. AbuRahma AF, Robinson PA, Strickler DL, Alberts S, Young L. Proposed new duplex classification for threshold stenoses used in various symptomatic and asymptomatic carotid endarterectomy trials. Ann Vasc Surg. 1998;12:349-358.
10. Grant EG, Benson CB, Moneta GL, et al. Carotid artery stenosis: gray-scale and Doppler US diagnosis--Society of Radiologists in Ultrasound Consensus Conference. Radiology. 2003;229:340-346.
11. Heijenbrok-Kal MH, Buskens E, Nederkoorn PJ, Van der Graaf Y, Hunink MGM. The optimal peak systolic velocity threshold on duplex ultrasonography for the indication carotid endarterectomy: a decision analytic approach. Radiology. 2004; accepted for publication.
12. Nederkoorn PJ, Mali WP, Eikelboom BC, et al. Preoperative diagnosis of carotid artery stenosis: accuracy of noninvasive testing. Stroke. 2002;33:2003-2008.
13. Buskens E, Nederkoorn PJ, Buijs-Van Der Woude T, et al. Imaging of carotid arteries in symptomatic patients: cost-effectiveness of diagnostic strategies. Radiology. 2004;233:101-112.
14. Fujitani RM, Mills JL, Wang LM, Taylor SM. The effect of unilateral internal carotid arterial occlusion upon contralateral duplex study: criteria for accurate interpretation. J Vasc Surg. 1992;16:459-467; discussion 467-458.
15. AbuRahma AF, Richmond BK, Robinson PA, Khan S, Pollack JA, Alberts S. Effect of contralateral severe stenosis or carotid occlusion on duplex criteria of ipsilateral stenoses: comparative study of various duplex parameters. J Vasc Surg. 1995;22:751-761; discussion 761-752.
16. Busuttil SJ, Franklin DP, Youkey JR, Elmore JR. Carotid duplex overestimation of stenosis due to severe contralateral disease. Am J Surg. 1996;172:144-147; discussion 147-148.
17. Henderson RD, Steinman DA, Eliasziw M, Barnett HJ. Effect of contralateral carotid artery stenosis on carotid ultrasound velocity measurements. Stroke. 2000;31:2636-2640.
18. Ray SA, Lockhart SJ, Dourado R, Irvine AT, Burnand KG. Effect of contralateral disease on duplex measurements of internal carotid artery stenosis. Br J Surg. 2000;87:1057-1062.
19. Abou-Zamzam AM, Jr., Moneta GL, Edwards JM, Yeager RA, Taylor LM, Jr., Porter JM. Is a single preoperative duplex scan sufficient for planning bilateral carotid endarterectomy? J Vasc Surg. 2000;31:282-288.
20. Sachar R, Yadav JS, Roffi M, et al. Severe bilateral carotid stenosis: the impact of ipsilateral stenting on Doppler-defined contralateral stenosis. J Am Coll Cardiol. 2004;43:1358-1362.

21. Halliday A, Mansfield A, Marro J, et al. Prevention of disabling and fatal strokes by successful carotid endarterectomy in patients without recent neurological symptoms: randomised controlled trial. Lancet. 2004;363:1491-1502.
22. Endarterectomy for asymptomatic carotid artery stenosis. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. JAMA. 1995;273:1421-1428.
23. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983;39:207-215.
24. Fillinger MF, Baker RJ, Jr., Zwolak RM, et al. Carotid duplex criteria for a 60% or greater angiographic stenosis: variation according to equipment. J Vasc Surg. 1996;24:856-864.
25. Alexandrov AV, Vital D, Brodie DS, Hamilton P, Grotta JC. Grading carotid stenosis with ultrasound. An interlaboratory comparison. Stroke. 1997;28:1208-1210.
26. Kuntz KM, Polak JF, Whittemore AD, Skillman JJ, Kent KC. Duplex ultrasound criteria for the identification of carotid stenosis should be laboratory specific. Stroke. 1997;28:597-602.
27. Elgersma OE, van Leersum M, Buijs PC, et al. Changes over time in optimal duplex threshold for the identification of patients eligible for carotid endarterectomy. Stroke. 1998;29:2352-2356.

9

Adjusting for bias in diagnostic reports

Bias in diagnostic test evaluations

The performance of a diagnostic test can be evaluated by comparing its results with a reference standard. Measures of sensitivity, specificity and accuracy can then be used as summary estimates of the test's performance. There are several limitations and threats to the internal and external validity of a diagnostic test evaluation.[1,2] Like any other type of research, flaws in patient selection, study conduct and data analysis can lead to biased results. First, selection bias, referral bias or spectrum bias may occur if patients are not included consecutively in the study. In this case, the selected patient group is not representative of the target group of the diagnostic test. Verification or work-up bias, which is in fact a form of referral bias, may exist if the test under evaluation influences the decision to perform the reference test and not all patients tested undergo the reference standard. Disease progression may bias performance parameters if the time lag between the diagnostic test and the reference test is too long. Excluding indeterminate test results and patients lost to follow-up may further cause overestimation of the test characteristics. Intra- and interobserver variation, awareness of the results of the diagnostic test when interpreting the reference standard or vice versa, and awareness of clinical characteristics may also influence test performance.

As long as reporting of the diagnostic test evaluation is complete, one can judge the potential for bias and the generalizability of the test results reported. The quality of reporting diagnostic test evaluations, however, is commonly suboptimal. Comparing publications reporting diagnostic test evaluations requires, at the very least, a complete description of the test, its positivity criterion, and the reference standard. Often this critical information is lacking. Especially for the purpose of meta-analysis, reporting critical information on the design and conduct of diagnostic studies is essential, so that statistical adjustments can be made.

The Standards for Reporting of Diagnostic Accuracy (STARD) working group has developed a checklist and flow diagram that may help accurate and complete reporting of diagnostic test evaluations. Recommendations from the STARD working group can be found on their website: www.consort-statement.org/stardstatement.htm

Adjusting for verification bias

In this issue an article (Miller et.al.) deals with an example of referral and verification bias concerning SPECT testing in patients suspected of coronary artery disease. In 10 years time, a large patient population from the Mayo clinic underwent stress SPECT testing (N=14273). Only 13% of all patients tested with SPECT were referred for coronary angiography. Whereas 26% of patients with positive SPECT results underwent coronary angiography, only 1% of the normal SPECT results were followed by angiography. The implication is that negative test results are less frequently verified by the reference standard test and thus false-negative and true-negative results will be missed yielding high apparent sensitivity and low apparent specificity.

The authors of the paper, however, recognized the potential verification bias in their estimates and they, justifiably, estimated test parameters adjusted for this bias. They used the correction method described by Begg and Greenes and Diamond's correction method.[1] The Begg and Greenes method can correct for both pre- and post-test referral bias, that is, it corrects for both the patient characteristics and the test result that together determine whether a patient will

be referred for the reference test. The Diamond method corrects only for post-test referral bias, that is, only the bias due to the influence the test result has on the decision to refer the patient for the reference test.

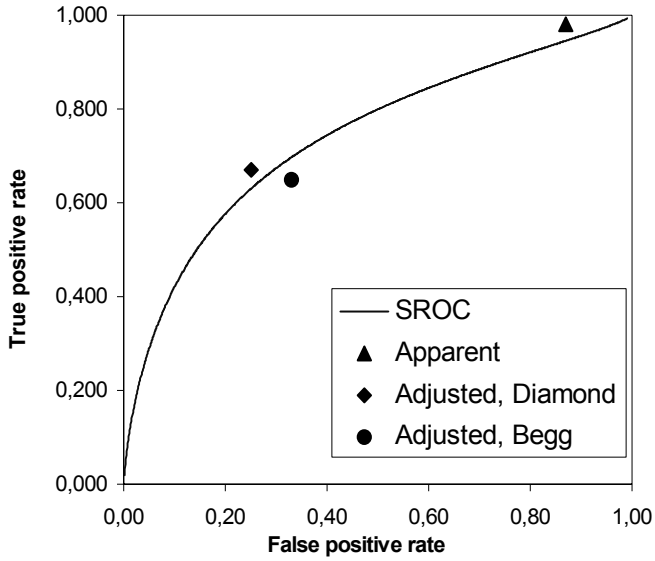


Figure: Apparent and adjusted point estimates (Diamond and Begg methods) of the Miller-report superimposed on SROC curve of SPECT from a published meta-analysis.

Another difference between the Begg and Diamond methods is that whereas the Begg method can be applied to all types of ROC curves, the Diamond method can only be applied to symmetrical ROC curves.

Technically Miller et al. used a modified version of the Begg method: they calculated the post-test probability or predictive value of CAD conditional on clinical factors and test results among those patients that underwent coronary angiography. Then they assumed that the derived prediction equation would also apply among those who did not undergo coronary angiography. In essence the assumption is that the predictive values of the diagnostic test are the same for the verified and source populations, which is equivalent to assuming that disease status and verification by the reference test are conditionally independent.[1]

Implications

The effect of correcting for verification bias on the point estimates of test performance can best be shown using a published meta-analysis of SPECT studies.[3] Although one can never fully adjust for bias in the source studies, it is possible to evaluate several types of bias in meta-analyses. Presence of verification bias was not a significant predictor of test performance in the cited meta-analysis. We superimposed the apparent and the adjusted operating points that Miller et al. reported on the SROC curve from the meta-analysis of SPECT studies (Figure). The SROC curve presented here was adjusted for age to be representative of the patient population from the Miller report. Whereas the correction for verification bias has a large effect on the

apparent point estimates of test performance, as reported by Miller, the figure shows that the apparent and corrected operating points lie along the same SROC curve. This suggests that the overall diagnostic performance as represented by an (S)ROC curve is not necessarily affected by verification bias but that the operating point corresponding to a particular positivity criterion (threshold value) is. This, in turn, implies that when we choose a positivity criterion for a test we need to take into account verification bias in the estimated test performance.

To determine the optimal positivity criterion (threshold value for calling a test score positive) we first need to determine the optimal operating point on the (S)ROC curve. The optimal operating point on the (S)ROC curve represents the best combination of sensitivity and specificity for a patient population in terms of effectiveness and cost. The optimal operating point is determined by the prior probability of disease, the benefit of correctly diagnosing the disease compared with it going undetected (true-positive compared with false-negative), and the losses associated with incorrectly labelling a subject diseased compared with not doing so (false-positive compared with true-negative). In other words, we need to adjust our operating point so that it is consistent with the trade-offs in risks and benefits.[4] Taking verification bias into account implies another adjustment when translating the derived optimal operating point to the threshold test score to be used in practice. After considering the risks and benefits we may, for example, conclude that we want a true-positive rate of 0.80 with a corresponding false-positive rate of 0.50. Since the reported scoring system corresponded with an adjusted sensitivity of 0.67 we would need to use a more lenient criterion to achieve a sensitivity of 0.80. For example, instead of only considering reversible perfusion defects and moderate fixed defects as a positive test result, one could also take into consideration mild fixed defects in order to diagnose any coronary artery disease. In general, verification bias overestimates apparent sensitivity if positive test results are verified preferentially, which implies that to achieve a particular sensitivity we would need to use a more lenient criterion than it would seem based on the verified sample only. In diagnosing CAD we could achieve this by adjusting the scoring system, or the threshold score, based on the SPECT images.

In summary, given that sensitivity and specificity estimates depend on the conduct of the test, the setting, characteristics of the patients involved, and interpretation of test results obtained, we should strive for accurately reporting how the test was evaluated and how the data were analyzed, so that bias can be recognized, validity can be judged, and appropriate adjustments made.

References

1. Begg CB, Greenes A: Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983; 39: 207-215.
2. Lijmer JG, Mol BW, Heisterkamp S, et al.: Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282: 1061-1066.
3. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS: Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA 1998; 280: 913-920.
4. Phelps CE, Mushlin AI: Focusing technology assessment using medical decision theory. Med Decis Making 1988; 8: 279-289.

10

Guiding future clinical research of competing diagnostic tests for coronary artery disease

Abstract

Background: Several imaging tests are available for the diagnosis of coronary artery disease (CAD) in patients with chest pain. Uncertainty exists about which imaging test is best for a patient given age, gender, and type of chest pain. Future clinical research could reduce this uncertainty and is therefore expected to result in better patient outcomes.
Objective: To guide future clinical research pertaining to the choice between imaging tests for CAD.
Methods: Decision modeling and value of information analysis were applied to estimate the benefit of future clinical research, using currently available evidence from the literature. The outcomes of interest for the imaging tests were quality-adjusted life expectancy and costs from the health care perspective. The benefit of future research was expressed in quality-adjusted life days (QALD) gained per patient and annual quality-adjusted life years gained for the USA.
Results: For patients with a pre-test probability of CAD below 18% it was best not to test; above 83% it is best to perform an angiography; for probabilities between these values, computed tomographic angiography (CTA) was the best test for most patients, although considerable uncertainty remained. The total expected value of perfect information (EVPI) – a ceiling for the harm of uncertainty – ranged from 0 to 13 QALD per patient. The annual population EVPI for all uncertain parameters – a ceiling for the benefit of future research – was 2891 quality-adjusted life years (QALYs). The annual population EVPI for obtaining information on the utilities of chest pain states was 1598 QALYs. Treatment effects, costs, and test characteristics had smaller but substantial associated annual population EVPIs.
Conclusion: The optimal imaging test for patients with chest pain is uncertain for most patients. Our results suggest that more clinical research to reduce this uncertainty is justified. Resolving uncertainty about utilities of chest pain states is most beneficial.

Introduction

Coronary artery disease (CAD) is the leading cause of death in the United States. Each year half a million patients are newly diagnosed with CAD, of whom about half initially present with chest pain.[1] It is important to identify patients with CAD, since they can benefit from a coronary artery bypass graft (CABG) or a percutaneous coronary intervention (PCI). To diagnose CAD in patients with chest pain, various diagnostic imaging tests are available.

Coronary angiography can perfectly distinguish patients with CAD from patients without CAD due to the high spatial resolution. Moreover, angiography makes direct intervention possible. Unfortunately, angiography has considerable drawbacks: it has a risk of mortality and morbidity, and it is expensive. Noninvasive imaging tests are less expensive and have, at most, a minimal risk. Their test characteristics (e.g., sensitivity and specificity), however, are imperfect: they misclassify both patients with CAD and those without CAD. The initial risk and cost of angiography versus the harm and cost of misclassifying patients with chest pain is the main trade-off when choosing between these imaging tests.

Exercise echocardiography and exercise single-photon-emission computed tomography are well established noninvasive tests for asymptomatic or minimally symptomatic stages of CAD. Computed tomographic angiography (CTA) is a more recent but promising test for the detection of CAD.[2] Sensitivity of four-detector row CTA, however, has been considered too low for routine clinical practice.[3] 16-detector row CTA appears to have better test characteristics, although these results are based on small samples.[4,5]

Given a patient's age, gender, and type of chest pain, substantial uncertainty remains about which imaging test is optimal. Future quantitative clinical research will - to some extent - resolve this uncertainty. This could result in better decisions, benefitting patients with chest pain and/or save money. Various uncertain parameters are responsible for the decision uncertainty, such as: test characteristics, treatment effects, and cost of tests and treatments. With value of information analysis we can estimate the benefit of obtaining more information on an uncertain parameter in future clinical research.[6] This benefit varies across parameters; a higher benefit implies that a parameter is more important. The objective of this study was to estimate the benefit of collecting information on subsets of uncertain parameters to guide future clinical research pertaining to the choice between imaging tests for CAD.

Methods

Decision model

We synthesized the available evidence regarding the costs and effects of imaging tests for diagnosing CAD into a decision model from the health care system perspective.[7,8] The model extrapolated the evidence on costs and effects over the entire remaining lifetime of patients. We compared four imaging tests for the diagnosis of CAD: coronary angiography, exercise echocardiography, exercise single-photon-emission computed tomography (SPECT), and 16-detector row computed tomographic angiography (CTA). In a fifth strategy patients with chest pain were not tested and received medical therapy only. We considered patients with chest pain without a history of myocardial infarction.

If a noninvasive test result was positive or uninterpretable, an angiography followed. Patients with a positive angiography received a PCI for one or two vessel disease and a CABG for three vessel disease and left main disease. Both treatments had an associated disutility and a short-term risk of mortality and myocardial infarction. The beneficial effects of treatment were three-fold: reduction in long-term mortality, reduction in long-term myocardial infarction, and reduction of chest pain severity. Age/gender-specific life tables were used to model the subsequent lifetime outcomes with Markov simulation. During this subsequent lifetime three chest pain states were distinguished: no, mild, and severe chest pain. Moreover, each year patients could suffer a myocardial infarction or undergo a CABG or PCI, depending on the extent of the coronary artery disease and treatment history. We modeled the cost of tests and treatments, as well as the annual cost – depending on chest pain severity and left ventricular ejection fraction - for patients with CAD.

The effects and costs of the imaging tests depend on patient characteristics. Patients differ with respect to the pre-test probability of having CAD, the short-term mortality from CABG, and the age/gender-specific annual mortality rates. We distinguished the same 30 subgroups as the CASS-registry, based on: gender, age (30-39, 40-49, 50-59, 60-69, and 70-79), and type of chest pain (typical, atypical, and non-specific).[9] Table 2 of the results section presents subgroup-specific pre-test probabilities of CAD and the proportion of all chest pain patients residing in each subgroup.

The model outcome of each imaging test was quality-adjusted life expectancy (QALE) and expected lifetime costs. We transformed this two-dimensional outcome into one composite dimension: net health benefit.[10] The net health benefit is the QALE minus the QALE equivalent of the costs, which is the costs divided by the societal willingness-to-pay. For example, if the QALE is 0.3 quality-adjusted life years (QALY), the costs are $5000 and the willingness-to-pay is $50000/QALY, then the net health benefit equals 0.3-5000/50000 = 0.2 QALY. Multiplying the net health benefit with the willingness-to-pay yields the net monetary benefit. We used a – generally accepted – willingness-to-pay of $50,000 per QALY gained.

Data sources

From the literature we derived for all uncertain parameters a mean value and a probability distribution representing the uncertainty. If this information was unavailable in the literature, it was estimated by experts. Table 1 lists for all uncertain parameters the mean values, a 95%-confidence interval, and the source of the data.

Probabilistic sensitivity analysis

We performed probabilistic sensitivity analysis (PSA, or 2nd order Monte Carlo simulation) to assess the uncertainty about the optimal test for each subgroup.[11] PSA propagates the uncertainty about the input parameters through the model, resulting in probability distributions for the model outcomes. These distributions reflect the uncertainty about the effects and costs of each test strategy (including not testing). We used these distributions to calculate the probability that given the currently available evidence we would make the wrong decision.

Table 1A: Model parameters – test characteristics and pre-test probability of CAD

nr.	parameter	mean (%)	95% confidence limits		source
	angiography: short-term risks				
1	myocardial infaction if 0-VD	0.02	0.02	0.03	[24]
2	myocardial infaction if 1-VD	0.06	0.05	0.07	[24]
3	myocardial infaction if 2-VD	0.08	0.07	0.10	[24]
4	myocardial infaction if 3-VD	0.08	0.07	0.09	[24]
5	myocardial infaction if LMD	0.17	0.15	0.18	[24]
6	die if 0-VD	0.02	0.01	0.02	[25]
7	die if 1-VD	0.05	0.04	0.06	[25]
8	die if 2-VD	0.07	0.06	0.08	[25]
9	die if 3-VD	0.12	0.11	0.13	[25]
10	die if LMD	0.55	0.52	0.58	[25]
	computed tomographic angiography				
11	sensitivity	99	95	100	[4]
12	specificity	88	59	100	[4]
13	percentage uninterpretable	1.69	0.04	6.16	[4]
14	percentage morbidity	0.05	0.04	0.06	expert opinion
	exercise echocardiography				
15	sensitivity	85	83	87	[26]
16	specificy	77	74	80	[26]
17	percentage uninterpretable	5	0	10	[27]
18	percentage morbidity	0.05	0.04	0.06	expert opinion
19	percentage mortality	0.005	0.004	0.006	expert opinion
	exercise SPECT				
20	sensitivity	87	86	88	[26]
21	specificity	64	60	68	[26]
22	percentage uninterpretable	2	0	4	expert opinion
23	percentage morbidity	0.05	0.04	0.06	expert opinion
24	percentage mortality	0.005	0.004	0.006	expert opinion
25	pre-test probability of CAD given age, gender, and chest pain type				see appendix B

Table 1B: Model parameters – costs

nr.	parameter	mean (US$)	credible interval ‡		source
	cost tests				
26	cost angiography	3186	2230	4141	Medicare (30% inpatient)
27	cost CTA	705	494	917	Medicare
28	cost echocardiography	241	169	313	Medicare
29	cost SPECT	599	419	779	Medicare
	cost treatments				
30	cost CABG	23052	16136	29968	Medicare
31	cost PCI	11795	8257	15334	Medicare (70% stents)
	cost events				
32	cost myocardial infarction	6690	4683	8697	Medicare
	annual costs chest pain patient				
33	no, abnormal LVEF	3870	2709	5031	Medicare
34	mild, abnormal LVEF	5689	3982	7395	Medicare
35	mild, normal LVEF	1818	1272	2363	Medicare
36	severe, abnormal LVEF	9148	6403	11892	Medicare
37	severe, normal LVEF	4105	2874	5337	Medicare

‡ For costs we assumed a ±30% credible interval on Medicare estimates and correlations of 1 between the cost parameters.

Table 1C: Model parameters – utilities

nr.	parameter	mean	95% confidence limits		source
	utilities chest pain states				
38	no	0.87	0.80	0.92	[28]
39	mild	0.81	0.76	0.86	[28]
40	severe	0.67	0.56	0.77	[28]
	disutilities (in days lost)				
41	CABG	30	24	36	expert opinion
42	angiography	1	0.8	1.2	expert opinion
43	myocardial infarction	10	8	12	expert opinion
44	PCI	2.5	2	3	expert opinion

Table 1D: Model parameters – medical treatment

nr.	parameter ‡	mean	95% confidence limits		source
	effects medical treatment only				
45	risk myocardial infarction with 0-VD	1.0	0.7	1.5	[29]
46	risk myocardial infarction with 1/2-VD	2.1	1.5	2.9	[29]
47	risk myocardial infarction with 3-VD / LMD	2.7	2.1	3.5	[29]
48	relative risk die with 1/2-VD *	2.3	1.9	2.8	[30]
49	relative risk die with 3-VD *	3.6	3.1	4.1	[30]
50	relative risk die with LMD *	9.6	6.2	14.3	[30]
51	% for annual transitions between chest pain states after medical treatment				see appendix A
52	risk future procedure 0-VD	0.5	0.2	1.0	[31]
53	risk future procedure 1-VD	1.0	0.5	1.7	[31]
54	risk future procedure 2-VD	4.2	3.0	5.5	[31]
55	risk future procedure 3-VD/LMD	7.5	6.0	9.2	[31]
56	% CABG of future procedures 0/1-VD	16	13.8	18.3	[31]
57	% CABG of future procedures 2-VD	58	55	61	[31]
58	% CABG of future procedures 3-VD/LMD	87	85	89	[31]
	LVEF (treatment independent)				
59	risk abnormal LVEF with 0-VD	0.5	0.4	0.7	[29]
60	risk abnormal LVEF with 1-VD	2.0	1.7	2.3	[29]
61	risk abnormal LVEF with 2-VD	5.0	4.5	5.5	[29]
62	risk abnormal LVEF with 3-VD	6.0	5.5	6.5	[29]
63	risk abnormal LVEF with LMD	6.0	5.5	6.5	[29]

‡ All risks – except relative risks – are annual risks and expressed in %.

* Relative risks of dying are in comparison to life table mortality.

Table 1E: Model parameters – PCI

nr.	parameter ‡	mean	95% confidence limits		source
	short-term risks with PCI				
64	myocardial infarction if 1-VD	3.5	2.7	4.4	[32]
65	myocardial infarction if 2-VD	5.2	4.3	6.2	[32]
66	die if 1-VD	0.2	0.1	0.5	[33, 34]
67	die if 2-VD	0.9	0.5	1.4	[33, 34]
	effects PCI				
68	reduction in myocardial infarction	17	12	22	[35]
69	reduction in mortality if 1/2-VD	15	0	50	[31]
70	% for annual transitions between chest pain states after PCI				see appendix A
71	relative risk chest pain improvement of PCI vs. CABG	0.85	0.80	0.90	[31]
72	relative risk of reprocedure for stent vs. PTCA	0.5	0.2	0.8	[36]
73	% eligible for PCI	76	61	91	[36]
74	annual risk of reprocedure after PCI	4	2	5	[31]
75	% CABG of reprocedures	22	19	25	[31]

‡ All values are expressed in %, except the relative risks.

Table 1F: Model parameters – CABG

nr.	parameter ‡	mean	95% limits	confidence	source
	short-term risks CABG				
76	myocardial infarction if 3-VD	7.0	5.9	8.2	[32]
77	myocardial infarction if LMD	7.0	5.9	8.2	[32]
78	die if 35yr male	2.3	0.2	6.6	[37]
79	die if 35yr female	2.9	0.1	10.9	[37]
80	die if 45yr male	2.7	1.6	4.2	[37]
81	die if 45yr female	3.5	1.6	6.1	[37]
82	die if 55yr male	3.2	2.3	4.2	[37]
83	die if 55yr female	4.1	2.6	5.8	[37]
84	die if 65yr male	4.4	3.5	5.4	[37]
85	die if 65yr female	5.7	4.1	7.5	[37]
86	die if 75yr male	6.3	5.1	7.6	[37]
87	die if 75yr female	8.0	6.0	10.3	[37]
	effects CABG				
88	reduction in myocardial infarction	42	29	55	[29]
89	reduction in mortality if 3-VD	48	29	64	[30]
90	reduction in mortality if LMD	67	40	84	[30]
91	% from mild to no angina	60	55	65	[29]
92	% from severe to mild angina	70	65	75	[29]
93	% from severe to no angina	10	7	13	[29]
94	% for annual transitions between chest pain states after CABG				see appendix A
95	annual risk of reprocedure after CABG	2	1	3	[31]
96	% CABG of reprocedures	7	5	9	[31]

‡ All values are expressed in %.

VD, vessel disease; LMD, left main disease; SPECT, single-photon emission computed tomography; CAD, coronary artery disease; LVEF, left ventricular ejection fraction; CTA, computed tomographic angiography; CABG, coronary artery bypass grafting; PCI, percutaneous coronary intervention; PTCA, percutaneous transluminal coronary angioplasty

Value of information analysis

A substantial probability of making the wrong decision is not necessarily important if the difference between the outcomes of competing strategies is negligible. Estimation of the expected value of perfect information (EVPI) acknowledges this by considering both the probability and the consequences of making the wrong decision.[12, 13] The EVPI is a measure for the importance of uncertainty to individual patients. The higher the EVPI, the more important the uncertainty is. The EVPI – or value of information (VOI) in general – is expressed in the same unit as the model outcome (e.g., QALY, life years, or $). An EVPI of 0.1 QALY means that if we would perform an infinitely large study to eliminate all uncertainty (that is, obtain perfect information), we can expect an increase in QALE of 0.1 QALY per patient. Because an infinitely large study is hypothetical, the EVPI is a ceiling value for the expected benefit of future research.[13]

The population EVPI is the EVPI multiplied with the number of patients that can benefit from future clinical research.[13] It is a measure for the importance of uncertainty to society from a utilitarian perspective. Population EVPI is expressed in the same unit as the model outcome: we used QALY and dollars. Expressing population EVPI in dollars enables comparison with research costs: it is the maximum amount that we should spend on future research to decrease current uncertainty. If the population EVPI does not exceed the estimated cost of future research, more research is not justified. We assumed that the annual population that could benefit from future clinical research on imaging tests for CAD would be 500,000 for the US.[1] Because the value of information differs across the subgroups, we calculated the population EVPI as a weighted sum of these subgroups.

The partial expected value of perfect information (partial EVPI) is the EVPI of one or more uncertain parameters. The parameters with the highest partial EVPI yield the highest benefit if more information is collected on them and are therefore the most important parameters to measure in future studies. For six subsets of parameters we estimated the partial EVPI. Table 1 presents these subsets as sub-tables A to F. Subset A includes the test characteristics of all imaging tests and the pre-test probability of CAD given age, gender, and type of chest pain. The cost parameters are grouped in subset B. The utilities of chest pain states and the disutilities of tests and events are in subset C. Parameters for the effects from medical treatment and the risk of an abnormal left ventricular ejection fraction reside in subset D. Subsets E and F represent parameters for the short-term risks and the effects of PCI and CABG, respectively.

We estimated the (partial) expected value of perfect information of these subsets of parameters using many (1000 or more) "what if"-simulations. For each "what if"-simulation a value for each uncertain parameter of interest was randomly drawn from its distribution. Each resulting set of parameter values had a probability of being the "true" set of parameter values. For these values we then identified the optimal strategy and the loss in cost and/or effect if this strategy was better than the optimal strategy based on current evidence. The expected value of perfect information was determined by calculating the mean (=expected) loss in cost and/or effect. The algorithms to perform VOI analyses were recently described in detail.[14,15]

Results

The best imaging test

The model outcomes for the five strategies were compared for each subgroup, resulting in the subgroup-specific best strategies (Table 2). Exercise echocardiography and exercise SPECT were never the optimal test. For patients with a pre-test probability of CAD below 18% it was best not to test. Above a pre-test probability of CAD of 83%, angiography was always the optimal strategy. For pre-test probabilities between these values, CTA was often, but not always, the best test. Subgroups with similar pre-test probabilities could have different optimal tests.

Uncertainty about the best test

Probabilistic sensitivity analysis showed how certain we are that the test identified as best, really is the best test. In table 2 the uncertainty is presented as the probability that the best test truly is the best test. For most subgroups this probability was well below 95%. For subgroups where CTA was the best test, this probability never exceeded 76%. Figure 1 shows the 95% credible interval of the net health benefit of each strategy for the 6 largest subgroups. Because the outcomes of the different strategies are correlated, the probability that a strategy is optimal can be close to 100% regardless of overlapping intervals. Figure 1 also illustrates that the difference between the best test and the second best test tends to be only a few quality-adjusted life days. The benefit of testing compared with "not testing", however, is for some subgroups as high as 4 quality-adjusted life months.

Importance of uncertainty: total and population EVPI

The total EVPI per patient - a measure of the importance of uncertainty for individual patients - differed considerably across subgroups (Table 2). If the probability of the best test being truly best exceeded 95%, the total EVPI was zero or very small. For lower probabilities, the total EVPI ranged from 2 to 13 quality-adjusted life days. The probability of the best test being truly best was not a good measure of the total EVPI per patient. For example, for a probability of the best test being truly best of 60-65%, the total EVPI per patient ranged from 2 to 13 quality-adjusted life days.

The annual population EVPI – a measure of the importance of uncertainty for society - was especially high for large subgroups with considerable uncertainty about the best test, for example, 40-49 yr men with typical or atypical chest pain. The population EVPI for all patients with chest pain was 2891 QALYs per year, which is equivalent to $145 million per year.

Benefit of assessing subsets of parameters

Figure 2 presents the population partial EVPI of the six subsets of parameters (A to F) corresponding to the subtables of Table 1. Subset C, representing the utilities of chest pain states and the disutilities of tests and events, was associated with the highest population partial EVPI. The population partial EVPI for test characteristics, costs, and treatment effects were similar. Only subset D – parameters related to medical treatment – had a much smaller population partial EVPI.

Table 2: Optimal test for each subgroup

age-group	angina type	% of total ‡	% CAD §	best test	% best *	EVPI/pat (QALD) ¶	popEVPI (QALY)†
MEN							
30-39	nonspec	3	3	NO TEST	100	0	0
40-49	nonspec	5	12	NO TEST	93	0	12
50-59	nonspec	5	18	CTA	50	3	221
60-69	nonspec	2	27	CTA	63	2	52
70-79	nonspec	0	63	CTA	61	2	8
30-39	atypical	2	46	NO TEST	78	2	67
40-49	atypical	7	57	CTA	49	7	725
50-59	atypical	11	71	CTA	56	2	296
60-69	atypical	5	78	ANGIO	89	2	132
70-79	atypical	1	94	ANGIO	99	0	1
30-39	typical	1	83	NO TEST	64	11	124
40-49	typical	5	88	ANGIO	77	5	353
50-59	typical	10	95	ANGIO	97	1	109
60-69	typical	7	95	ANGIO	100	0	0
70-79	typical	1	97	ANGIO	100	0	1
WOMEN							
30-39	nonspec	2	4	NO TEST	99	0	1
40-49	nonspec	5	4	NO TEST	100	0	0
50-59	nonspec	7	6	NO TEST	98	0	4
60-69	nonspec	3	10	NO TEST	61	2	71
70-79	nonspec	0	0	NO TEST	100	0	0
30-39	atypical	0	20	NO TEST	76	2	16
40-49	atypical	3	31	NO TEST	62	4	178
50-59	atypical	5	30	CTA	55	4	249
60-69	atypical	3	48	CTA	76	2	95
70-79	atypical	0	56	CTA	76	2	12
30-39	typical	0	70	NO TEST	64	13	22
40-49	typical	1	56	NO TEST	77	3	39
50-59	typical	2	68	CTA	63	3	78
60-69	typical	2	81	ANGIO	91	1	22
70-79	typical	0	96	ANGIO	94	2	6

‡ % of all chest pain patients in each subgroup.

§ Pre-test probability of CAD given age, gender, and angina type for patients with chest pain.

* Probability that the optimal test given available evidence is the “true” best test.

¶ Total EVPI per patient in quality-adjusted life days.

† Population EVPI in quality-adjusted life years per year for the USA.

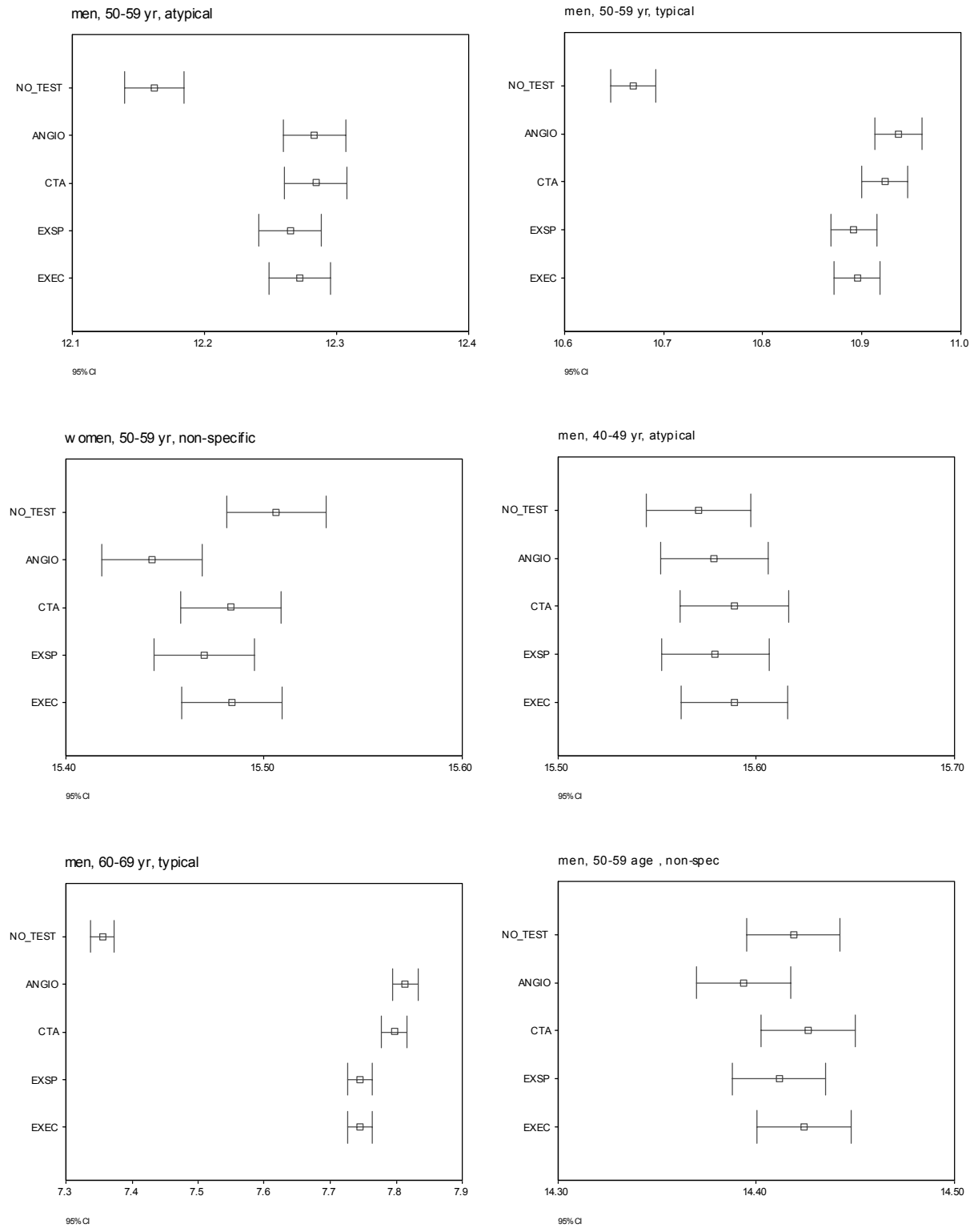


Figure 1: Credible intervals (95%) for the net health benefit of all strategies for the six largest subgroups

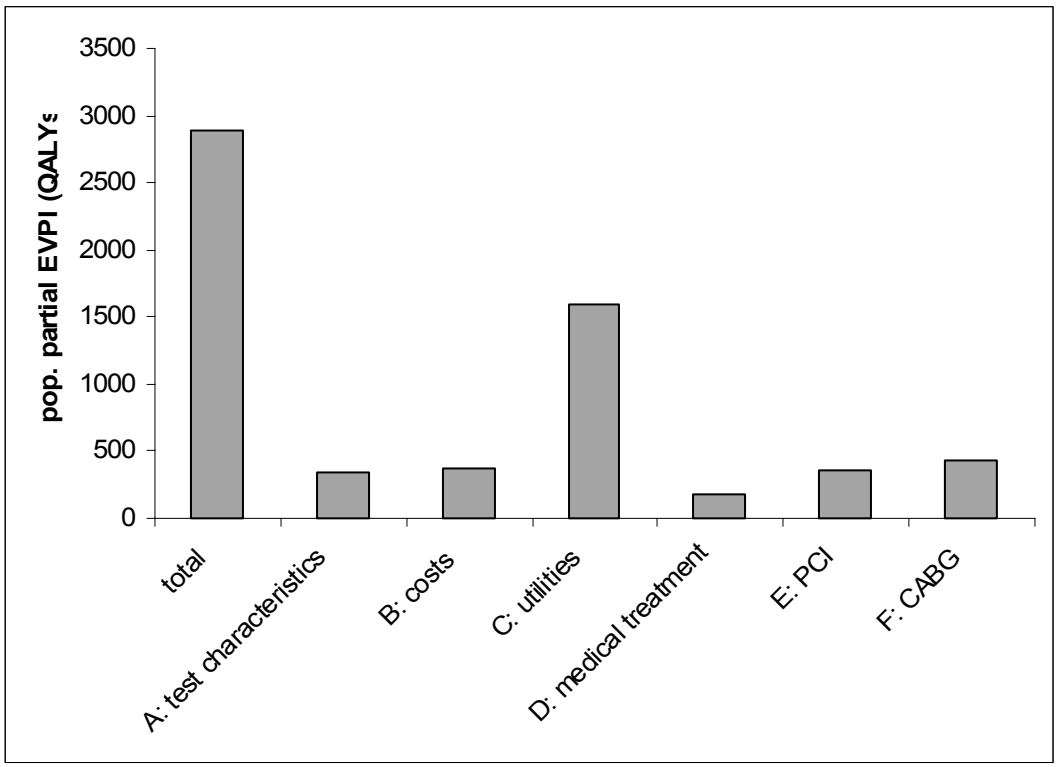


Figure 2: Population EVPI in QALYs per year of the subsets of parameters of Table 1

Discussion

We used a decision model and value of information analysis to assess the uncertainty, and the importance of uncertainty, pertaining to the choice between imaging tests to diagnose coronary artery disease for 30 patient subgroups. In addition, we explicitly estimated the expected benefit of resolving uncertainty of subsets of uncertain parameters.

For most subgroups the decision is uncertain: the probability that with current evidence we would choose the "true" best strategy was below 95%. For the subgroups where CTA was the best test, this probability did not exceed 76%. This uncertainty, however, is in itself not worrisome if the harm of choosing a suboptimal test is negligible. Value of information analysis showed that for many subgroups the expected harm of uncertainty was substantial, ranging from 0 to 13 quality-adjusted life days per patient. These may seem small losses but they are in the same range as, for example, the benefits of breast cancer or colon cancer screening programs.[16] The importance of uncertainty for individual patients was highest for men and women, aged 30-39 years, with typical chest pain and men, aged 40-49 years, with atypical or typical chest pain. The probability of choosing the "true" best test was not a good measure for the importance of uncertainty: for similar probabilities of choosing the "true" best test EVPIs differed up to 6-fold.

The annual population EVPI was 2891 QALYs for the US, which is equivalent to $145 million. Uncertainty of the utilities of the chest pain states (no, moderate, severe) and the disutilities of tests and events was most important. Although a study with a finite sample size

will result in a smaller benefit than estimated for perfect information, obtaining better estimates of the utilities seems the best allocation of research funds. The importance of the utilities can be explained by the considerable uncertainty in the literature about their values and covariance structure, in combination with their impact on false-negative test results.

Obtaining better estimates of test characteristics, costs, and treatment effects was associated with similar population partial EVPIs. The research cost for these subsets of parameters, however, may differ considerably. Most of these parameters could be assessed in observational studies, though some of the treatment effects need to be addressed in randomized controlled trials. In further analyses we will compare the benefit of various study designs.

Our analyses have some limitations that are inherent to value of information analyses. The medical literature generally contains little evidence regarding correlations between input parameters for a decision model. Due to this lack of information, correlations are generally ignored (i.e., implicitly set to zero). Alternatively, prefect correlations can be assumed, as we did for the cost parameters. Unfortunately, the value of information is very sensitive to correlations. In addition, uncertainty about the structure of the decision model is a generally ignored source of uncertainty that could influence the value of information. Uncertainty about the number of future patients that could benefit from future clinical research is another important issue. It is not clear whether we should consider all patients world-wide, limit the analysis to one country, or limit it to a particular local setting. Moreover, the number of years during which patients are expected to benefit from future clinical research is uncertain. These problems, however, are not drawbacks of value of information analysis, but are inherent to the underlying decision problem of how to allocate our limited resources for clinical research wisely.

The well-known alternative to value of information analysis is to apply the arbitrary cut-off for the p-value (0.05) – or its Bayesian analogue: the probability of optimality – to decide whether more research is justified. P-values below 0.05 are often considered as sufficient proof for superiority of a health care intervention and p-values above 0.05 as justification to embark on a future, often larger and more expensive, study. The p-value, however, only tells us how certain we are that two interventions differ in outcome. The medical importance of the uncertainty depends on both the probability and the consequences of not choosing the "true" best imaging test given current available evidence. A substantial probability that the best strategy is not really the best is of little importance if the difference between these strategies is infinitesimal. On the other hand, a small probability could be of great concern, if being wrong has huge consequences.

Value of information analysis is a method from decision theory that was described several decades ago[12,17,18] and is now increasingly applied to decisions in health care and medicine.[19-22] It considers research as an investment and estimates the expected return on investment. Analogous to requiring that health care interventions are cost-effective, value of information analysis requires that clinical research is cost-effective. This idea is not new in the medical literature: Detsky used a method to estimate the benefit of clinical trials, to enable comparison with the cost of these trials.[23] His method, however, was applied after the trials were carried out and used an arbitrary cut-off for the p-value as well as for the minimum clinically important difference.

In conclusion, for most subgroups of patients considerable uncertainty exists about the optimal imaging test to diagnose coronary artery disease. The expected harm of this uncertainty ranged from 0 to 13 quality-adjusted life days per patient, depending on age, gender, and type of chest pain. Future clinical research could reduce this uncertainty to some extent. An observational study yielding more precise estimates of the quality-of-life weights for varying severity of chest pain seems most beneficial. In addition, patients could benefit from more precise estimates of test characteristics, costs, and treatment effects. In further analyses we will compare the benefit of various study designs. Value of information analysis seems a practical method to prioritize and guide future clinical research using available evidence.

References

1. Elveback LR, Connolly DC, Kurland LT. Coronary heart disease in residents of Rochester, Minnesota. II. Mortality, incidence, and survivorship, 1950-1975. Mayo Clin Proc. 1981; 56:665-672.
2. Schoenhagen P, Halliburton SS, Stillman AE, et al. Noninvasive imaging of coronary arteries: current and future role of multi-detector row CT. Radiology. 2004; 232:7-17.
3. Schoepf UJ, Becker CR, Ohnesorge BM, Yucel EK. CT of coronary artery disease. Radiology. 2004; 232:18-37.
4. Nieman K, Cademartiri F, Lemos PA, Raaijmakers R, Pattynama PM, de Feyter PJ. Reliable noninvasive coronary angiography with fast submillimeter multislice spiral computed tomography. Circulation. 2002; 106:2051-2054.
5. Ropers D, Baum U, Pohle K. Detection of coronary artery stenoses with thin-slice multi-detector row spiral computed tomography and multiplanar reconstruction. Ciculation. 2003; 107:664-666.
6. 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.
7. 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.
8. Heijenbrok-Kal MH, Kuntz KM, Fleischmann KE, Hunink MGM. Cost-effectiveness of electron beam computed tomography in patients with chest pain. Submitted.
9. Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). Circulation. 1981; 64:360-367.
10. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998; 18:S68-80.
11. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Med Decis Making. 1985; 5:157-177.
12. Raiffa H. Decision Analysis: Introductory Lectures on Choices under Uncertainty. Reading, MA: Addison-Wesley, 1968.
13. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999; 18:341-364.
14. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. Med Decis Making. 2004; 24:207-227.
15. Groot Koerkamp B, Hunink MGM, Stijnen T, Weinstein MC. Identifying key parameters in cost-effectiveness analysis using value of information: a comparison of methods. Health Econ. Submitted.
16. Wright JC, Weinstein MC. Gains in life expectancy from medical interventions - standardizing data on outcomes. N Eng J Med. 1998; 339:380-386.
17. Raiffa H, Schlaifer R. Applied Statistical Decision Theory: Harvard University, 1961.
18. Howard RA. Information value theory. IEEE Transact Systems Science and Cybernetics 1966; SCC-2:22-26.
19. Coyle D, Buxton MJ, O'Brien BJ. Measures of importance for economic analysis based on decision modeling. J Clin Epidemiol. 2003; 56:989-997.
20. Payne N, Chilcott J, McGoogan E. Liquid-based cytology in cervical screening: a rapid and systematic review. Health Technol Assess. 2000; 4:1-73.
21. Phillips CV. The economics of 'more research is needed'. Int J Epidemiol. 2001; 30:771-776.
22. Claxton K, Neumann PJ, Araki S, Weinstein MC. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. Int J Technol Assess Health Care. 2001; 17:38-55.
23. Detsky AS. Are clinical trials a cost-effective investment? JAMA. 1989; 262:1795-1800.
24. Johnson LW, Lozner EC, Johnson S, et al. Coronary arteriography 1984-1987: a report of the Registry of the Society for Cardiac Angiography and Interventions. I. Results and complications. Cathet Cardiovasc Diagn. 1989; 17:5-10.

25. Lozner EC, Johnson LW, Johnson S, et al. Coronary arteriography 1984-1987: a report of the Registry of the Society for Cardiac Angiography and Interventions. II. An analysis of 218 deaths related to coronary arteriography. Cathet Cardiovasc Diagn. 1989; 17:11-14.
26. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA. 1998; 280:913-920.
27. Rumberger JA, Behrenbeck T, Breen JF, Sheedy PF, 2nd. Coronary calcification by electron beam computed tomography and obstructive coronary artery disease: a model for costs and effectiveness of diagnosis as compared with conventional cardiac testing methods. J Am Coll Cardiol. 1999; 33:453-462.
28. Nease RF, Jr., Kneeland T, O'Connor GT, et al. Variation in patient utilities for outcomes of the management of chronic stable angina. Implications for clinical practice guidelines. Ischemic Heart Disease Patient Outcomes Research Team. JAMA. 1995; 273:1185-1190.
29. Coronary artery surgery study (CASS): a randomized trial of coronary artery bypass surgery. Survival data. Circulation. 1983; 68:939-950.
30. Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. Lancet. 1994; 344:563-570.
31. Pocock SJ, Henderson RA, Rickards AF, et al. Meta-analysis of randomised trials comparing coronary angioplasty with bypass surgery. Lancet. 1995; 346:1184-1189.
32. Holmes DR, Jr., Holubkov R, Vlietstra RE, et al. Comparison of complications during percutaneous transluminal coronary angioplasty from 1977 to 1981 and from 1985 to 1986: the National Heart, Lung, and Blood Institute Percutaneous Transluminal Coronary Angioplasty Registry. J Am Coll Cardiol. 1988; 12:1149-1155.
33. Yusuf S, Zucker D, Chalmers TC. Ten-year results of the randomized control trials of coronary artery bypass graft surgery: tabular data compiled by the collaborative effort of the original trial investigators. Part 2 of 2. Online J Curr Clin Trials. 1994; Doc No 144.
34. Yusuf S, Zucker D, Chalmers TC. Ten-year results of the randomized control trials of coronary artery bypass graft surgery: tabular data compiled by the collaborative effort of the original trial investigators. Part 1 of 2. Online J Curr Clin Trials. 1994; Doc No 145.
35. Wong JB, Sonnenberg FA, Salem DN, Pauker SG. Myocardial revascularization for chronic stable angina. Analysis of the role of percutaneous transluminal coronary angioplasty based on data available in 1989. Ann Intern Med. 1990; 113:852-871.
36. Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C. Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review. Health Technol Assess. 2000;4:1-153.
37. Hannan EL, Kilburn H, Jr., O'Donnell JF, Lukacik G, Shields EP. Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates. JAMA. 1990; 264:2768-2774.

Appendix A

Dirichlet distributions were used to represent uncertainty about the annual transition probabilities between chest pain severity states. These transitions depend on the extent of vessel disease and the treatment. The parameters of the dirichlet distributions are the number of patients that transit to each health state. For example, for patients with 1 vessel disease, medical treatment, and no chest pain, 26 out of 164 patients had mild chest pain at the end of the year. These numbers were derived and extrapolated from the CASS-registry.[9,29]

	sample size	none	mild	severe
Medical treatment, 1-VD				
none	164	136	26	2
mild	309	46	244	19
severe	20	5	7	8
CABG/PCI, 1-VD				
none	263	226	34	3
mild	188	43	134	11
severe	18	5	6	7
Medical treatment, 2-VD				
none	193	154	37	2
mild	449	49	359	40
severe	42	8	10	23
CABG/PCI, 2-VD				
none	428	376	47	4
mild	262	34	225	3
severe	14	3	4	8
Medical treatment, 3-VD/LMD				
none	202	135	65	2
mild	360	86	259	14
severe	31	3	3	25
CABG/PCI, 3-VD/LMD				
none	316	262	51	3
mild	208	31	164	12
severe	15	4	5	6

Appendix B

Dirichlet distributions representing uncertainty about pre-test probability of CAD were based on observational study results.[9] The probabilities depend on gender, age, and type of chest pain. For example, 249 men with non-specific chest pain were observed of whom 5 had one vessel disease.

subgroup	sample size	0-VD	1-VD	2-VD	3-VD	LMD
NON-SPECIFIC						
Men						
30-39	249	242	5	0	2	0
40-49	391	344	35	8	4	0
50-59	440	361	48	18	4	9
60-69	129	94	15	12	6	1
>70	22	8	2	7	5	0
Women						
30-39	135	130	3	1	0	1
40-49	425	408	13	4	0	0
50-59	585	550	23	12	0	0
60-69	215	194	15	2	2	2
>70	23	23	0	0	0	0
ATYPICAL						
Men						
30-39	171	92	32	31	12	3
40-49	568	244	131	80	85	28
50-59	919	267	221	221	147	64
60-69	434	95	91	104	100	43
>70	46	3	10	11	15	7
Women						
30-39	39	31	4	2	0	2
40-49	257	177	46	18	5	10
50-59	414	290	54	37	25	8
60-69	264	137	58	34	24	11
>70	34	15	5	5	8	1
TYPICAL						
Men						
30-39	66	11	25	15	10	5
40-49	402	48	109	105	96	44
50-59	840	42	168	244	269	118
60-69	539	27	86	129	199	97
>70	67	2	6	10	33	16
Women						
30-39	10	3	1	3	2	1
40-49	70	31	15	12	12	0
50-59	161	52	35	31	31	13
60-69	137	26	19	34	44	14
>70	23	1	5	2	13	2

11

Discussion

Discussion

In this thesis various methods of diagnostic test evaluations have been described and applied to specific clinical problems. Each of these methods has its advantages and disadvantages that should be understood in order to correctly interpret the available evidence and to select the best methodology for a specific research question. Furthermore, the diagnostic performance of various tests has been evaluated for the diagnosis of coronary, carotid, peripheral, and renal artery disease. Each of these tests also has its merits and limitations, which should be taken into account in order to choose the best test for a specific clinical problem. In this chapter the methodology and the results of the diagnostic test evaluations will be discussed.

Methods of diagnostic test evaluations

The methods applied in this thesis for the assessment of diagnostic imaging technology include systematic reviews, cost-effectiveness analyses, optimization of test characteristics, optimization of cut-off values of a test, correction for verification bias, and value of information analysis.

Systematic reviews were used to summarize the available evidence. The aim of a systematic review is to collect all available evidence and by integrating the results of independent studies that on their own may have contradictory or non-significant results to increase power so that an overall conclusion can be made[1]. In- and exclusion criteria are used to select relevant articles from the published literature. An advantage of a systematic review is that in principal all available evidence from multiple studies on a specific subject is taken into account. A disadvantage is that studies may be excluded from a systematic review because data are missing or the quality of reporting of the study was insufficient[2]. Studies may also have been missed, because the search criteria may not have detected them. Furthermore, in systematic reviews publication bias may play a role, which implies that studies with positive results are more likely to be published than studies with negative or non-significant results, which could result in overestimation of the diagnostic performance of a test[3]. However, statistical methods exist to correct for potential publication bias.

We evaluated the potential influence of publication bias by including the year of publication in the regression analyses. This method gives information about potential trends over time in published estimates of diagnostic performance of a test, but cannot detect whether studies with negative results are missing in the literature. Other recently developed methods, which we did not perform in this thesis, include the trim and fill method [4]. We did evaluate a correction method for verification bias, which is often present in the evaluation of point estimates of diagnostic performance[5]. In our systematic reviews, however, we used a summary receiver operating characteristic (SROC) analysis to integrate the estimates of sensitivity and specificity of individual studies. We showed that the point estimates of sensitivity and specificity may shift along the SROC curve as a result of verification bias. In our SROC analyses we also accounted for the heterogeneity among study settings by using a random effects model, in contrast to a fixed effects model[6]. The inverse of the variance of the diagnostic odds ratio was used as weighting factor giving more weight to studies with more precise estimates in the regression analysis. The random effects meta-analytical model differs from the mainly used fixed effects

model in that apart from the individual study variances it also takes into account the between-study variance[6].

An advantage of SROC analysis over independent pooling of sensitivity and specificity estimates is that the inverse relationship between sensitivity and specificity is taken into account associated with differences in cut-off values of the test and the potential influence of multiple study characteristics on diagnostic performance can be determined. A limitation of SROC analysis is that it can only use one pair of estimates of sensitivity and specificity from each study. Furthermore, if SROC curves of two or more tests cross, it is still unclear which test has the best diagnostic performance and which test should be used in a specific clinical setting. In particular, it does not provide information on the consequences of the test in terms of cost and effectiveness and the applicability of the test in clinical practice. Therefore, a cost-effectiveness analysis is essential to define the clinical role of a diagnostic test.

The cost-effectiveness analyses in this thesis were based on decision-analytical models. The advantage of modelling studies is that many alternative test strategies can be compared, which cannot be realized in a clinical study for ethical and logistical reasons. A disadvantage of a cost-effectiveness analysis is that input variables come from multiple sources and assumptions have to be made to keep the model tractable. Furthermore, the differences in cost and consequences associated with diagnostic test results may be too small to detect when studied using a lifetime time horizon and a societal perspective. Therefore, uncertainty about the best test strategy for a clinical setting may remain. However, a cost-effectiveness analysis provides more opportunities to explore the diagnostic performance of a test.

As we have shown in our analyses the outcomes of a cost-effectiveness analysis in combination with a SROC analysis can be used for the optimization of test characteristics and cut-off values of a test in different clinical settings. By maximizing the net health benefit of a test, a measure that integrates both costs and effectiveness, for each combination of sensitivity and specificity along the SROC curve the optimal test characteristics can be determined[7]. For this optimal operating point on the SROC curve there is an optimal balance between the trade-offs of cost and effectiveness associated with true-positive and false-negative test results in diseased patients and with true-negative and false-positive test results in non-diseased subjects. This method is therefore more clinically relevant than the various methods used in the literature in which the accuracy of a test is maximized. In the latter case the total number of false test results is minimized, which implies that the consequences of a false-positive test result are assumed equal to the consequences of a false-negative test. In clinical practice, however, the consequences of missing a diagnosis may be far more harmful than referring a patient for treatment of a non-significant stenosis or vice versa.

In addition, we have shown how the optimal operating point on the ROC curve of a single study can be translated into the optimal cut-off value of a test, using the optimal likelihood ratio of the test. The likelihood ratio of a test, which is related to a certain cut-off value, is equal to the slope of the ROC curve. Using this method based on cost and effectiveness outcomes, we found different optimal cut-off values of the peak systolic velocity on duplex ultrasonography for the indication carotid endarterectomy than those recommended in the literature, which were again based on optimizing accuracy[8].

Methods from cost-effectiveness analysis can also be applied to explore the uncertainty of the test variables, namely using probabilistic sensitivity analysis and value of information

analysis[9]. These methods help guide future research by identifying those variables that should be evaluated in a future study. In probabilistic sensitivity analysis a cost-effectiveness analysis is repeated at least 1000 times, each time drawing the parameter values randomly from probability distributions representing the uncertainty that surrounds each point estimate in the model. In contrast to point estimates of cost and effectiveness, this analysis will result in probability distributions for the model outcomes, which reflect the uncertainty about the costs and effects of a test strategy. We then used these distributions in a value of information analysis in which we calculated the probability that, given the currently available evidence, we would make the wrong decision. The larger this probability, the higher the uncertainty and the more we would be prepared to pay to collect more information in a future study. A future study is justified as long as its expected value does not exceed the research cost. Again, basing decisions on the trade-offs between the benefits and harms associated with diagnostic testing is more clinically relevant than calculating an arbitrary p-value for decision making in health care.

In general, all methods used in this thesis provide complementary information on the performance of diagnostic tests. The choice of which methodology to use depends on the research question. If the question is to find out which test is best in terms of test accuracy, i.e. prevention of missing diagnoses and unnecessary treatments, a systematic review including SROC analysis should be performed. If the results are inconclusive or if a clinician wants to know which test should be used in a specific clinical setting a cost-effectiveness analysis should follow the SROC analysis. If the question is how to make optimal use of a new test than the test characteristics and cut-off values of the test should be optimized for the patient population that will undergo the new test. Finally, if a new clinical study is planned the results of a cost-effectiveness analysis should be evaluated using value of information analysis to determine the optimal study design.

Comparison of diagnostic tests for cardiovascular disease and future research

Coronary artery disease

In this thesis various diagnostic imaging tests were evaluated for the diagnosis of coronary artery disease. In a systematic review of meta-analyses, we found little differences between stress echocardiography, stress SPECT, and EBCT, in the relative diagnostic odds ratios. If differences in diagnostic performance are inconclusive a cost-effectiveness analysis is crucial to find out the clinical role of each of the tests. Therefore, we evaluated the same tests using a cost-effectiveness analysis. We found that exercise SPECT was dominated by other test strategies and that the role of EBCT in clinical practice was very small in comparison with exercise echocardiography.

We also evaluated the clinical role of pharmacological stress in combination with either echocardiography or SPECT. The results suggested that dipyridamole and dobutamine echocardiography were cost-effective tests in patients with atypical chest pain at intermediate risk of coronary artery disease. Adenosine SPECT was also cost-effective, but at a higher pre-test risk and at a higher willingness-to-pay threshold. In both cost-effectiveness analyses, the no imaging strategy was optimal for patients with non-specific chest pain, who are at low risk of

coronary artery disease. Both analyses also showed that coronary angiography, the reference test, was the optimal test strategy in patients with typical angina, who are at high risk of coronary artery disease.

Several studies have shown that EBCT is a very sensitive test, but not very specific, which may explain why EBCT was not very cost-effective for diagnostic purposes[10, 11]. Therefore, we evaluated whether the test characteristics of EBCT could be optimized for several cohorts of patients defined by age, sex, and type of chest pain. The results of this study suggested that using optimal test characteristics for specific patient populations did improve the net health benefit that can be obtained with EBCT. In our study, however, this did not lead to a greater role for EBCT in patients with chest pain for the diagnosis of coronary artery disease.

Another CT technology that is currently being investigated is multi-detector (MD-) computed tomographic angiography (CTA) adapted for imaging the coronary arteries[12]. In addition, contrast-enhanced magnetic resonance angiography (MRA) is currently being investigated for the diagnosis of coronary artery disease[13]. Preliminary results show that both MD-CTA and contrast-enhanced MRA are promising techniques[13, 14]. Using value of information analysis, we studied which test variables of MD-CTA are worth the research cost for evaluation in a future study and what study design should be used. We found that we should learn more about the test characteristics of MD-CTA and about the quality of life of patients with several types of chest pain.

Overall, our results suggest that, although differences in test accuracy were non-significant, exercise and pharmacological stress echocardiography showed a better diagnostic performance than exercise and pharmacological SPECT and EBCT for the diagnosis of coronary artery disease in patients with atypical chest pain, if evaluated from a health care perspective. Furthermore, future research should concentrate on new techniques like MD-CTA and contrast-enhanced MRA for the diagnosis of coronary artery disease. Finally, value of information analysis is a useful tool to guide the design of future studies and should be used to allocate research funding efficiently.

Peripheral, Renal, and Carotid artery disease

In a systematic review using SROC analysis we compared duplex ultasonography, computed tomographic angiography (CTA) and (contrast-enhanced) magnetic resonance angiography (MRA). This study showed a significantly better diagnostic performance with CTA and contrast-enhanced MRA compared with non-enhanced MRA and duplex ultrasonography. Furthermore, we found that imaging tests perform significantly better in the diagnosis of peripheral arterial disease than in renal or carotid artery disease. The number of subdivisions of the arterial tree into arterial segments significantly influenced the diagnostic performance. Therefore, we advocate standardization of the subdivision of the arterial tree for the purpose of diagnostic test evaluations in cardiovascular disease. A limitation of this study was that the number of CTA-studies was small and only available for renal artery disease. However, if test accuracy is used as criterion to select the best available test this study suggests that CTA and contrast-enhanced MRA should be used in clinical practice. Furthermore, the technological developments of imaging tests should concentrate on renal and carotid artery disease, because these areas could benefit the most from further improvements.

Because imaging time on CT and MR equipment is often limited and especially in the case of MR expensive, it may still be relevant to study the diagnostic performance of duplex ultrasonography for selected indications. In a recently published cost-effectiveness analysis, for example, a test strategy using duplex ultrasonography as sole imaging test to select patients for carotid endarterectomy was the most cost-effective strategy in patients with symptoms of amaurosis fugax, transient ischemic attack, or minor stroke[15]. We showed that the outcomes of the cost-effectiveness analysis in combination with a ROC curve from a clinical trial comparing duplex ultrasonography with coronary angiography can be used to define optimal test characteristics and optimal cut-off values of duplex ultrasonography. The point on the ROC curve where the net health benefit reached its maximum was defined as the optimal operating point on the ROC curve. This method accounts for the trade-offs in benefits and harms associated with diagnostic testing as opposed to methods that use maximization of accuracy[7]. Using this method, the optimal cut-off values that we found for the indication carotid endarterectomy for a more than 70% or 50% stenosis were higher than the ones published by the Society of Radiologists in Ultrasound [8].

This analysis was followed by a regression analysis, in which we showed that the optimal test characteristics and cut-off values of carotid duplex ultrasonography depended on whether the side of the carotid artery was symptomatic or not in patients with symptoms of minor stroke, transient ischemic attack, or amaurosis fugax. This implies that using different cut-off values for the symptomatic and asymptomatic carotid artery in these patients, the clinical use of duplex ultrasonography can be optimized. However, the optimal cut-off values should be validated in clinical practice.

In summary, the diagnostic performance of CTA and contrast-enhanced MRA is better than that of non-enhanced MRA and duplex ultrasonography for the diagnosis of peripheral, renal, and carotid artery disease in terms of test accuracy. The subdivision of the arterial tree into arterial segments should be standardized in future studies. SROC-analysis in combination with cost-effectiveness analysis can be used for the optimization of a diagnostic test in a specific clinical setting. Value of information analysis should be performed to design future studies on the diagnostic performance of imaging tests.

References

1. Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. J Clin Epidemiol. 1995;48(1):9-18.
2. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Radiology. 2003;226:24-28.
3. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ. 1997;315:640-645.
4. Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J Am Stat Ass. 2000;95:89-98.
5. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983;39:207-215.
6. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. Stat Med. 1995;14:395-411.
7. Hunink MGM, Glasziou PP. Multiple test results. In: Decision making in health and medicine. Integrating evidence and values. 1st ed. Cambridge: Cambridge University Press; 2001.
8. Grant EG, Benson CB, Moneta GL, et al. Carotid artery stenosis: gray-scale and Doppler US diagnosis--Society of Radiologists in Ultrasound Consensus Conference. Radiology. 2003;229:340-346.
9. 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.
10. Nallamothu BK, Saint S, Bielak LF, et al. Electron-beam computed tomography in the diagnosis of coronary artery disease: a meta-analysis. Arch Intern Med. 2001;161:833-838.
11. O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/American Heart Association Expert Consensus Document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. J Am Coll Cardiol. 2000;36:326-340.
12. Schoepf UJ, Becker CR, Ohnesorge BM, Yucel EK. CT of Coronary Artery Disease. Radiology. 2004;232:18-37.
13. Kim WY, Danias PG, Stuber M, et al. Coronary magnetic resonance angiography for the detection of coronary stenoses. N Engl J Med. 2001;345:1863-1869.
14. Nieman K, Cademartiri F, Lemos PA, Raaijmakers R, Pattynama PM, de Feyter PJ. Reliable noninvasive coronary angiography with fast submillimeter multislice spiral computed tomography. Circulation. 2002;106:2051-2054.
15. Buskens Eea. Cost-effectiveness in diagnosing carotid artery stenosis. Radiology. 2004;accepted for publication.

Summary

Summary

This thesis describes studies on the assessment of diagnostic imaging technologies for cardiovascular disease. In chapter 1, the rationale for this research project is presented. Because cardiovascular disease is still the leading cause of death in industrialized countries, it is important to diagnose this disease accurately, so that adequate treatment can be initiated. The accepted reference standard for the diagnosis of arterial stenosis is conventional angiography, which carries a risk of morbidity and mortality. Therefore, research currently focuses on the development and improvement of non-invasive imaging technologies. An enormous variety of non-invasive imaging technologies is available and numerous diagnostic studies have been published. The question remains which studies provide the best-available evidence for the clinician and which diagnostic test should be chosen in a particular clinical setting. Therefore, the aim of this thesis was to assess diagnostic imaging technologies for cardiovascular disease using various methods of diagnostic test evaluations and to determine the optimal diagnostic test strategy in specific clinical settings. A secondary aim was to develop the methods available for the assessment of diagnostic imaging technology, including meta-analytical and decision analytical techniques.

To summarize the currently available evidence on the diagnostic performance of non-invasive imaging tests published in the literature a systematic review can be performed. Using summary receiver operating characteristic (SROC) analysis the diagnostic performance of various tests can be compared and study characteristics can be determined that may influence the diagnostic performance.

In chapters 2 and 3 a systematic review was performed of meta-analyses evaluating diagnostic imaging tests for coronary, peripheral, renal, and carotid artery disease. Each meta-analysis provided sensitivity and specificity estimates of the tests reported in the source studies. A SROC analysis was performed on the data from the source studies. Relative diagnostic odds ratios were compared in which measures of sensitivity and specificity were integrated into one measure of diagnostic performance. The influence of potential confounders was evaluated by adding variables to the regression models. Assessing the coronary tests (chapter 2) we found that differences between tests were small and that the diagnostic performance of the coronary tests diminished over the years.

Comparing tests for peripheral, renal, and carotid artery disease (chapter 3), we found that diagnostic performance of computed tomographic angiography (CTA) and contrast-enhanced magnetic resonance angiography (MRA) were significantly better than non-enhanced MRA and duplex ultrasonography. Furthermore, our results suggested that studies in which the vascular tree was subdivided into a high number of arterial segments the diagnostic performance was inflated compared with studies with no, or less, subdivisions into arterial segments.

A large number of diagnostic imaging tests exist that have shown a good diagnostic performance for the diagnosis of cardiovascular disease. From a health care policy perspective, however, the value of a diagnostic test should not only reflect the diagnostic performance, but also the cost and risks associated with the test and the treatments that may be induced by the

test, the pre-test probability of disease, and the health related quality of life of the patient. Cost-effectiveness analysis can be performed to integrate these variables and to guide decisions on which test should be used in specific clinical settings.

In chapters 4 and 5 cost-effectiveness analyses were performed for comparison of various coronary imaging tests. In chapter 4 the cost-effectiveness of electron beam computed tomography (EBCT) was compared with exercise echocardiography and exercise single-photon emission computed tomography (SPECT) in patients suspected of coronary artery disease. Our results suggested that, in general, EBCT was not a cost-effective test for the diagnosis of coronary artery disease. Only if society is willing to pay a relatively high dollar amount per quality-adjusted life-year (QALY) gained, can a small subset of patients at intermediate probability of coronary artery disease benefit from this test. In chapter 4 three types of pharmacological stress, i.e. adenosine, dipyridamole, and dobutamine, in combination with either echocardiography or SPECT were compared with a no imaging strategy and routine coronary angiography. The cost-effectiveness analysis showed that dipyridamole and dobutamine echocardiography were cost-effective in patient subsets at intermediate pre-test probability of coronary artery disease, whereas adenosine SPECT was cost-effective for patients with at intermediate risk in specific clinical situations or at higher willingness-to-pay thresholds. Routine coronary angiography was the optimal test strategy for patients with typical angina.

The diagnostic performance of a test cannot be expressed in a single fixed estimate of test performance. The diagnostic performance of a test depends on the setting in which a test is used and on the perspective from which the test is evaluated. Several methods have been described to optimize the diagnostic performance of a test, such as the determination of the optimal sensitivity and specificity and the optimal cut-off value of a test to discriminate between the diseased patients and the non-diseased patients. Chapters 6, 7, and 8 describe analyses performed to determine the optimal sensitivity and specificity and cut-off value of tests for coronary and carotid artery disease.

In chapter 6 the sensitivity and specificity of EBCT were optimized for six cohorts of patients defined by age, sex, and type of chest pain from the health care perspective. First a SROC analysis was performed based on a published meta-analysis. The regression model obtained from this analysis was integrated into a cost-effectiveness analysis comparing the cost-effectiveness of EBCT, no imaging, and coronary angiography for patients with chest pain. The results of this study show that if optimal estimates of sensitivity and specificity are used for specific cohorts in a cost-effectiveness analysis instead of pooled fixed values, the net health benefit of a test may be improved.

Chapters 7 and 8 focus on the optimal cut-off value of duplex ultrasonography for the diagnosis of carotid artery disease based on a cost-effectiveness analysis. We found that optimal cut-off values based on cost-effectiveness differ from cut-off values based on maximum accuracy of a test and from recently published recommended cut-off values (chapter 7). Furthermore, we found that a significant difference existed in the optimal cut-off value between the asymptomatic and the symptomatic carotid artery in patients with symptoms of amaurosis fugax, transient ischemic attack, or minor stroke (chapter 8). The cut-off value of the asymptomatic carotid artery was optimal at a significantly higher peak systolic velocity than the optimal cut-off value of the symptomatic carotid artery.

The results of comparing diagnostic imaging tests may be affected by biases, in particular verification bias. Verification bias exists if patients are referred to the reference test based on the results from the test under evaluation. As a result of verification bias sensitivity may be inflated and specificity deflated if positive test results are verified preferentially. Methods have been developed to correct for verification bias.

In chapter 9 we discuss the potential influence of verification bias on point estimates of sensitivity and specificity and on the SROC curve. We found that if correction for verification bias was performed, point estimates of the test characteristics may shift along the SROC curve, implying that the cut-off value of a test may change. We stated that we should strive for accurate reporting of how the test was evaluated and how the data were analyzed so that bias can be recognized, validity can be judged, and appropriate adjustments can be made.

Despite substantial clinical research, uncertainty remains about which test is best in a specific clinical setting. Future quantitative clinical research can help resolve this uncertainty. Using value of information analysis the expected benefit of research on various subsets of study parameters can be estimated.

In chapter 10 six subsets of uncertain parameters in a cost-effectiveness analysis comparing imaging technology for coronary artery disease, namely exercise echocardiography, exercise SPECT, computed tomographic angiography, and coronary angiography, were assessed using value of information analysis. Our results suggested that an observational study yielding more precise estimates of the quality-of-life weights for the severity of chest pain seems most beneficial. In addition, patients could benefit from more precise estimates of test characteristics, costs, and treatment effects. Value of information analysis seems a practical method to prioritize and guide future clinical research using available evidence.

This thesis concludes with a general discussion on the methodology of the presented studies and the clinical implications (chapter 11). This final chapter also provides recommendations on further diagnostic test evaluations.

Samenvatting

Samenvatting

Dit proefschrift beschrijft studies naar de evaluatie van diagnostische beeldvormende technieken voor hart- en vaatziekten. In hoofdstuk 1 wordt de aanleiding voor dit proefschrift gegeven. Omdat hart- en vaatziekten nog altijd de belangrijkste doodsoorzaak vormen in de geïndustrialiseerde landen, is het belangrijk om deze ziekte nauwkeurig te diagnosticeren, zodat een adequate behandeling kan worden ingesteld.

De geaccepteerde referentie standaard voor het diagnosticeren van stenoses in de slagaders is de conventionele angiografie, welke een morbiditeits- en mortaliteitsrisico met zich meebrengt. Daarom richt het onderzoek zich vandaag de dag op het ontwikkelen en verbeteren van niet-invasieve beeldvormende technieken. Er is een enorme variëteit aan niet-invasieve beeldvormende technieken beschikbaar en talrijke diagnostische studies zijn gepubliceerd. De vraag blijft welke studies het best beschikbare bewijs leveren voor de clinicus en welke test uitgevoerd moet worden in een specifieke klinische setting. Daarom was het doel van dit proefschrift om diagnostische beeldvormende technieken voor hart- en vaatziekten, waarbij verschillende methoden van diagnostische testevaluaties worden gebruikt, te evalueren en om vast te stellen wat de optimale diagnostische teststrategie is in een specifieke klinische setting. Een tweede doel was om de methoden die beschikbaar zijn voor de evaluatie van diagnostische beeldvormende technieken te verbeteren, waaronder meta-analytische en besliskundige technieken.

Om de nu beschikbare bewijsvoering over de diagnostische accuratesse van niet-invasieve beeldonderzoeken die gepubliceerd is in de literatuur samen te vatten, kan een systematisch literatuuronderzoek worden uitgevoerd. Met behulp van de zogenaamde summary receiver operating characteristic (SROC) analyse kan de diagnostische accuratesse van meerdere tests worden vergeleken en kunnen determinanten van de studie worden gevonden die van invloed kunnen zijn op de diagnostische accuratesse.

In de hoofdstukken 2 en 3 werd een systematisch literatuuronderzoek naar meta-analyses uitgevoerd die diagnostische beeldonderzoeken bestudeerden voor ziekten van de kransslagaders, de perifere slagaders, de nierslagaders, en de halsslagaders. Iedere meta-analyse presenteerde de schattingen van de sensitiviteit en specificiteit van de tests die waren gerapporteerd in de bronstudies. Relatieve diagnostische odds ratios werden vergeleken waarin de maten van sensitiviteit en specificiteit werden geïntegreerd in één maat van diagnostische accuratesse. De invloed van potentieel verstorende variabelen werd geëvalueerd door variabelen aan het regressiemodel toe te voegen. Bij het bestuderen van de tests voor de kransslagaders (hoofdstuk 2) vonden wij dat de verschillen tussen de tests klein waren en dat de diagnostische accuratesse van deze tests over de jaren verminderde.

Bij het vergelijken van de tests voor perifere slagaders, nierslagaders en halsslagaders (hoofdstuk 3), vonden wij dat de diagnostische accuratesse van CTA en MRA met contrast-toediening significant beter waren dan die van MRA zonder contrast-toediening en duplex echo-onderzoek. Verder suggereerden onze resultaten dat studies waarin de vaatboom was onderverdeeld in een groot aantal slagadersegmenten de diagnostische accuratesse hoger was vergeleken met studies waarin geen of minder onderverdelingen werden gemaakt.

Er bestaat een groot aantal diagnostische beeldonderzoeken die een goede diagnostische accuratesse hebben gedemonstreerd voor de diagnostiek van hart- en vaatziekten. Vanuit het maatschappelijk perspectief, echter, zou de waarde van een diagnostische test niet alleen de diagnostische accuratesse moeten weergeven, maar ook de kosten en risico's verbonden aan de test en de behandelingen die worden ingesteld op basis van de test, de kans op ziekte voorafgaand aan de test, en de aan gezondheid gerelateerde kwaliteit van leven van de patiënt. Een kosten-effectiviteitanalyse kan worden uitgevoerd om deze variabelen te integreren en om beslissingen te helpen maken over welke test gebruikt zou moeten worden in een specifieke klinische setting.

In de hoofdstukken 4 en 5 werden kosten-effectiviteitanalyses uitgevoerd voor het vergelijken van verscheidene beeldonderzoeken voor de kransslagaders. In hoofdstuk 4 werd de kosten-effectiviteit van elektronenbundel tomografie (EBT) vergeleken met inspannings-echocardiografie en inspannings-single-photon-emission computed tomografie (inspannings-SPECT test) bij patiënten bij wie ziekte van de kransslagaders werd verondersteld. Onze resultaten suggereerden dat de EBT test in het algemeen niet een kosten-effectieve test was voor de diagnose van ziekte van de kransslagaders. Alleen als de maatschappij bereid is een relatief hoge prijs te betalen per levensjaar aangepast voor kwaliteit van leven, kan een kleine subgroep van de patiënten met een middelmatige kans op ziekte van de kransslagaders profiteren van deze test. In hoofdstuk 4 werden drie soorten van medicamenteuze stress, te weten adenosine, dipyridamol, en dobutamine in combinatie met ofwel echocardiografie of de SPECT test vergeleken met een strategie zonder beeldonderzoek en een strategie met routinematige angiografie van de kransslagaders. De kosten-effectiviteitanalyse liet zien dat dipyridamol en dobutamine echocardiografie kosten-effectief waren in subgroepen van patiënten met een middelmatige kans op ziekte van de kransslagaders voorafgaand aan de test, terwijl de adenosine SPECT test kosten-effectief was voor patiënten met een middelmatig risico in specifieke klinische situaties of wanneer de gezondheidszorg bereid is tot het betalen van een hogere prijs. Routinematige angiografie van de kransslagaders was de optimale teststrategie voor patiënten met typische pijn op de borst.

De diagnostische accuratesse van een test kan niet worden uitgedrukt in een enkele vaste schatting van de testaccuratesse. De diagnostische accuratesse van een test is afhankelijk van de setting waarin een test wordt toegepast en vanuit welk perspectief een test wordt onderzocht. Verscheidene methoden zijn beschreven om de diagnostische accuratesse van een test te optimaliseren, zoals het bepalen van de optimale sensitiviteit en specificiteit en het optimale afkappunt van een test om onderscheid te maken tussen zieke patiënten en niet-zieke patiënten. Hoofdstukken 6, 7, en 8 beschrijven analyses die werden uitgevoerd om de optimale sensitiviteit en specificiteit en het afkappunt van tests voor ziekten van de kransslagaders en halsslagaders te bepalen.

In hoofdstuk 6 werden de sensitiviteit en specificiteit van de EBT test geoptimaliseerd voor zes cohorten van patiënten die werden gedefinieerd door leeftijd, geslacht, en type pijn op de borst, vanuit het perspectief van de gezondheidszorg. Eerst werd een SROC analyse uitgevoerd die was gebaseerd op een gepubliceerde meta-analyse. Het regressiemodel dat door deze analyse werd verkregen werd geïntegreerd in een kosten-effectiviteitanalyse die de kosten-effectiviteit van de EBT test vergeleek met de strategie zonder beeldonderzoek en met de

strategie met angiografie van de kransslagaders voor patiënten met pijn op de borst. De resultaten van deze studie laten zien dat de netto gezondheidswinst van een test verbeterd zou kunnen worden, wanneer optimale schattingen van de sensitiviteit en specificiteit worden gebruikt voor specifieke cohorten in een kosten-effectiviteitanalyse in plaats van vaste samengevoegde waarden.

De hoofdstukken 7 en 8 focussen op het optimale afkappunt van het duplex echo-onderzoek voor de diagnostiek van ziekte van de halsslagaders die gebaseerd werd op een kosten-effectiviteitanalyse. We vonden dat optimale afkappunten die gebaseerd zijn op kosten-effectiviteitanalyse verschilden van afkappunten die waren gebaseerd op de maximale nauwkeurigheid van een test en van recent gepubliceerde aanbevolen afkappunten (hoofdstuk 7). Verder vonden wij dat er een significant verschil bestond in het optimale afkappunt voor de asymptomatische en de symptomatische halsslagaders bij patiënten met symptomen van amaurosis fugax, een TIA, of een kleine beroerte (hoofdstuk 8). Het afkappunt voor de asymptomatische halsslagader was optimaal bij een significant hogere maximale systolische bloedstroomsnelheid dan het optimale afkappunt voor de symptomatisch halsslagader.

De resultaten van het vergelijken van diagnostische beeldonderzoeken kunnen worden beïnvloed door bias, in het bijzonder door verificatie bias. Er is sprake van verificatie bias als patiënten worden verwezen naar de referentietest op basis van de resultaten van de test die wordt onderzocht. Het gevolg van verificatie bias is dat de sensitiviteit hoger wordt en de specificiteit lager als positieve testuitslagen met voorkeur worden geverifieerd. Er zijn methoden ontwikkeld om voor verificatie bias te corrigeren.

In hoofdstuk 9 wordt de potentiële invloed van verificatie bias op de puntschattingen van sensitiviteit en specificiteit en op de SROC curve bediscussieerd. Wij vonden dat, wanneer de correctie voor verificatie bias was uitgevoerd, de puntschattingen van de testkarakteristieken langs de SROC curve kunnen verschuiven, wat impliceert dat het afkappunt van de test kan veranderen. We stelden dat we moeten streven naar een nauwkeurige rapportage van hoe de test werd onderzocht en hoe de data werden geanalyseerd, zodat bias kan worden herkend, de validiteit kan worden beoordeeld, en geschikte aanpassingen kunnen worden gemaakt.

Ondanks de aanzienlijke hoeveelheid klinisch onderzoek, blijft er onzekerheid bestaan over welke test de beste is in een specifieke klinische setting. Toekomstig kwantitatief klinisch onderzoek kan helpen deze onzekerheid te verminderen. Met gebruikmaking van het zogenaamde value-of-information onderzoek kan de verwachte opbrengst van onderzoek naar verschillende subgroepen van studieparameters worden ingeschat.

In hoofdstuk 10 werden 6 subgroepen van de onzekere parameters onderzocht met behulp van value-of-information analyse afkomstig uit een kosten-effectiviteitanalyse van beeldvormende technieken voor ziekte van de kransslagaders, namelijk inspannings-echocardiografie, inspannings-SPECT, CT-angiografie, en conventionele angiografie. De resultaten suggereren dat een observationele studie die preciezere schattingen bevat van de kwaliteit-van-leven voor de ernst van pijn op de borst het meest gunstig lijkt. Daarnaast zouden patiënten kunnen profiteren van preciezere schattingen van de testkarakteristieken, kosten, en

behandelingseffecten. Value-of-information analyse lijkt een praktische methode te zijn om prioriteiten te stellen in toekomstig klinisch onderzoek op basis van de beschikbare kennis.

Dit proefschrift eindigt met een algemene discussie over de methodologie van de besproken studies en de klinische implicaties (hoofdstuk 11). Dit laatste hoofdstuk verschaft ook aanbevelingen voor verder diagnostisch testonderzoek.

Dankwoord

Dankwoord

In de afgelopen vier jaar zijn er vele personen geweest die hebben bijgedragen aan mijn promotieonderzoek. Een aantal van hen wil ik graag met name noemen.

Prof. Hunink, beste Myriam, ik dank je voor de mogelijkheden die je mij geboden hebt om mij verder te verdiepen in het wetenschappelijk onderzoek. Ik heb niet alleen van het onderzoeksproject zelf veel geleerd, maar ook van de wekelijkse werkbesprekingen van de ART-groep, van mijn onderzoekstijd in Boston, en vooral ook van jouw persoonlijke begeleiding, met name in de laatste fase van het project. Ik bewonder je enorme kennis van zaken en je enthousiasme. Daarom ben ik blij dat je mij een nieuwe positie hebt aangeboden bij de ART-groep. Ik hoop nog lang met je te kunnen samenwerken.

I performed part of my PhD project in Boston under supervision of Dr. Kuntz. Dear Karen, I am honored that you are my co-supervisor. Thank you for giving me the opportunity to work as a teaching assistant in your modeling course. I have had a wonderful time in Boston.

Prof. Stijnen, Prof. De Feyter, en Prof. Rutten dank ik voor hun bereidheid om deel te nemen aan de promotiecommissie en voor het beoordelen van mijn manuscript. Ook de overige leden, Prof. Hofman, Prof. Krestin, Dr. Dippel, en Dr. Van der Lugt wil ik hartelijk danken voor het zitting nemen in de grote commissie.

Bijzonder blij ben ik met mijn twee paranimfen, Ylian en Karen. Beste Ylian, jij hebt veel meer voor mij gedaan dan je van een paranimf kunt verwachten; bedankt voor je geweldige inzet en vriendschap. Beste Karen, jij gaf mij het goede voorbeeld voor mijn promotietraject; bedankt voor al je hulp en vriendschap.

I would like to thank all co-authors for reviewing my manuscripts, especially Dr. Fleischmann and Dr. Kuntz. Dear Kirsten and Karen, I enjoyed our collaboration very much and I hope we will continue to work together in the future. Karen, special thanks for providing your cost-effectiveness model for the coronary tests. Kirsten, special thanks for your help in updating the cost data for the coronary procedures.

Voor een deel van mijn promotieonderzoek heb ik met plezier samengewerkt met enkele medewerkers van het Juliuscentrum in Utrecht. Prof. Van der Graaf en Dr. Nederkoorn (nu AMC) wil ik graag bedanken voor het beschikbaar stellen en uitleggen van hun onderzoeksgegevens. Tevens heb ik gebruik mogen maken van een kosteneffectiviteitmodel van Dr. Buskens. Beste Yolanda, Paul, en Erik, ik ben jullie erg dankbaar voor de prettige samenwerking en voor jullie bijdrage aan de artikelen die hieruit zijn voortgekomen.

Een belangrijke bijdrage in dit proefschrift werd geleverd door mijn collega Bas. Beste Bas, jouw stuk geeft een extra dimensie aan mijn proefschrift.

Alle (ex)medewerkers van de ART-groep, in het bijzonder Ylian, Karen, Rody, Bas, Joke, Ineke, Rogier, Galied en Ankie, wil ik bedanken voor hun belangstelling en suggesties voor mijn onderzoek tijdens de ART-bijeenkomsten, en voor de gezellige koffie-, thee, en lunchpauzes. Ook de andere collega's van de afdeling Epidemiologie & Biostatistiek wil ik danken voor hun collegialiteit en gezelligheid.

Voor vragen over de statistiek kon ik altijd terecht bij Prof. Stijnen. Beste Theo, bedankt voor je heldere uitleg en adviezen. Lidia, bedankt voor je advies over meta-analyses. Eerste hulp bij computerproblemen kreeg ik van Nano en Marcel: bedankt voor jullie vele reddingsacties. Verder bedank ik Marjolijn, Petra, en Marti voor de secretariele ondersteuning. Andries dank ik

voor het maken van een mooie poster voor het jaarlijkse congres van de Society of Medical Decision Making.

Lieve vrienden en vriendinnen, in het bijzonder Helmie en Monique, bedankt voor jullie vriendschap in voor- en tegenspoed.

Lieve familie, bedankt voor jullie liefde en steun en belangstelling voor mijn werk, maar vooral ook daarbuiten. Dit geldt speciaal voor mijn lieve ouders, zus, en broer. Lieve pap, extra dank voor je inhoudelijke interesse en het zetten van de puntjes op de i in mijn manuscript. Lieve mam, extra dank voor het met liefde oppassen op Jens tijdens het voltooien van dit boekje.

De laatste woorden van dit proefschrift wil ik wijden aan de twee belangrijkste mannen in mijn leven: Mark en Jens. Mijn liefste Mark, bedankt voor je eindeloze liefde en ondersteuning in alles wat ik doe; ik hou van je! Onze liefste Jens, bedankt voor je perfecte timing: nu dit proefschrift af is, begin jij te lachen!

About the author

About the author

Majanka H. Heijenbrok-Kal was born on October 18th, 1970 in Vlaardingen, the Netherlands. She graduated in 1989 at the Stedelijk Gymnasium in Schiedam. In the same year she started her study in Human Movement Sciences at the Free University in Amsterdam, where she obtained her degree in 1994. During her studies, she went for eight months to Finland, to take part in a Sport and Health Sciences program at the University of Jyväskyla. After her graduation, she worked as a Clinical Research Coordinator for the Europa study, a multicenter trial on the secondary prevention of coronary artery disease, at the Julius Center for Patient Oriented Research of the University Medical Center Utrecht, in collaboration with Sticares (Cardiovascular Research Foundation). In August 2000, she started the work described in this thesis in the Assessment of Radiological Technology (ART) group, a cooperation of the Department of Epidemiology & Biostatistics and the Department of Radiology of the Erasmus Medical Center Rotterdam. During this work, she obtained in 2002 a Master of Science degree in Clinical Epidemiology at the Netherlands Institute of Health Sciences. In 2003 she worked for four months as a Teaching Assistant at the Harvard School of Public Health in Boston, U.S.A, where part of this work was undertaken. Recently, she continued her research on diagnostic imaging tests for cardiovascular disease as a post-doc in the ART group.

Majanka H. Heijenbrok-Kal