



**LUND**  
UNIVERSITY



Master of Science Thesis

**Analysis of the Benefit of  
Quantification in Myocardial  
Perfusion Imaging: A human  
observer ROC evaluation  
using Monte Carlo Simulated  
Data**

**Rouwaida Yassin**

Supervisors

Ulf Nilsson, PhD

Jolanta Bartosik, MD

Department of Clinical Physiology,  
Helsingborg Hospital

Michael Ljungberg, PhD

Department of Medical Radiation Physics,  
Lund University Hospital

Department of Medical Radiation Physics  
Clinical Sciences, Lund University  
2005

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>3</b>
<b>INTRODUCTION</b> .....	<b>4</b>
<b>MATERIALS AND METHODS</b> .....	<b>7</b>
THE COMPUTER PHANTOM .....	7
THE SIMULATED COMPUTER SCINTILLATION CAMERA.....	9
IMAGE RECONSTRUCTION .....	10
QUANTITATIVE PERFUSION SOFTWARE.....	11
ROC- ANALYSIS.....	15
<b>RESULTS</b> .....	<b>17</b>
<b>DISCUSSION</b> .....	<b>23</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>25</b>
<b>REFERENCES</b> .....	<b>26</b>
<b>APPENDIX I</b> .....	<b>28</b>

## Abstract

The purpose of this study was to evaluate the benefit of commercial quantification software in perfusion SPECT imaging by simulating a group of patients and their corresponding SPECT acquisitions using a computer phantom and a Monte Carlo program. Based on decisions from experienced physicians and their visual interpretation of the simulated patient images, without and with addition of the information from the quantitative perfusion software, the benefit of such a quantification program was evaluated from a ROC analysis.

**Methods:** A computer phantom was used to generate twenty male patients of normal size, where six patients were without any perfusion defects in the left ventricular (LV) wall and fourteen patients had perfusion defects at seven different size and locations in the LV wall. The activity uptake in the defects was 75% and 85% of the normal activity in the LV wall, giving a total number of fourteen different defects. The size of the defect varied between 4% and 19 % of the LV wall volume. The activity distribution in the phantom was simulated to correspond to a total administered activity of 600 MBq  $^{99}\text{Tc}^{\text{m}}$  - Sestamibi.

The phantom was connected to a scintillation camera Monte Carlo simulation program (SIMIND) and realistic SPECT projections were simulated. These projections were imported into the clinical reconstruction software (AutoSPECT<sup>+</sup><sup>TM</sup>) and evaluated using the same procedures as for a real patient study. A commercial software, AutoQuant<sup>TM</sup>, was used for the reviewing, analysis and quantification of the myocardial perfusion SPECT images. Four experienced physicians interpreted the twenty computer patient images at two different reading sessions, without and with the use of the quantitative perfusion information. To evaluate the aid of the quantitative perfusion information Receiver Operating Characteristic (ROC) analysis was used. The ROC curve is a plot of sensitivity, or true positive fraction (TPF), of a procedure versus its false positive fraction (FPF = 1- specificity) for all possible cut-points. The area under the ROC curve was used as a measure of accuracy of the diagnostic procedure.

**Results:** The overall ROC curves showed some improvement in the area under the curve when quantitative perfusion information from the software was used. However, this improvement was relatively moderate and not significant. For the individual observers, the result from the ROC analysis varied, probably because of their individual mental thresholds to rank images.

**Conclusion:** The Monte Carlo method together with a realistic computer phantom and ROC analysis can be very useful to obtain information in how observers take advantage of information from quantitative perfusion software in myocardial SPECT.

Key Words: technetium-99m-sestamibi, quantitative perfusion, SPECT, Monte-Carlo, ROC-analysis.

## Introduction

Single Photon Emission Computed Tomography (SPECT) is a non-invasive medical imaging technique that is widely used for the diagnosis of coronary artery disease (CAD). Myocardial perfusion imaging (MPI) with SPECT allows for a direct assessment of regional perfusion changes and therefore the method plays an important role in the diagnosis of CAD in patients presenting symptoms of chest pain. An inducible perfusion abnormality indicates an impaired perfusion reserve, which in turn corresponds to coronary artery obstruction.

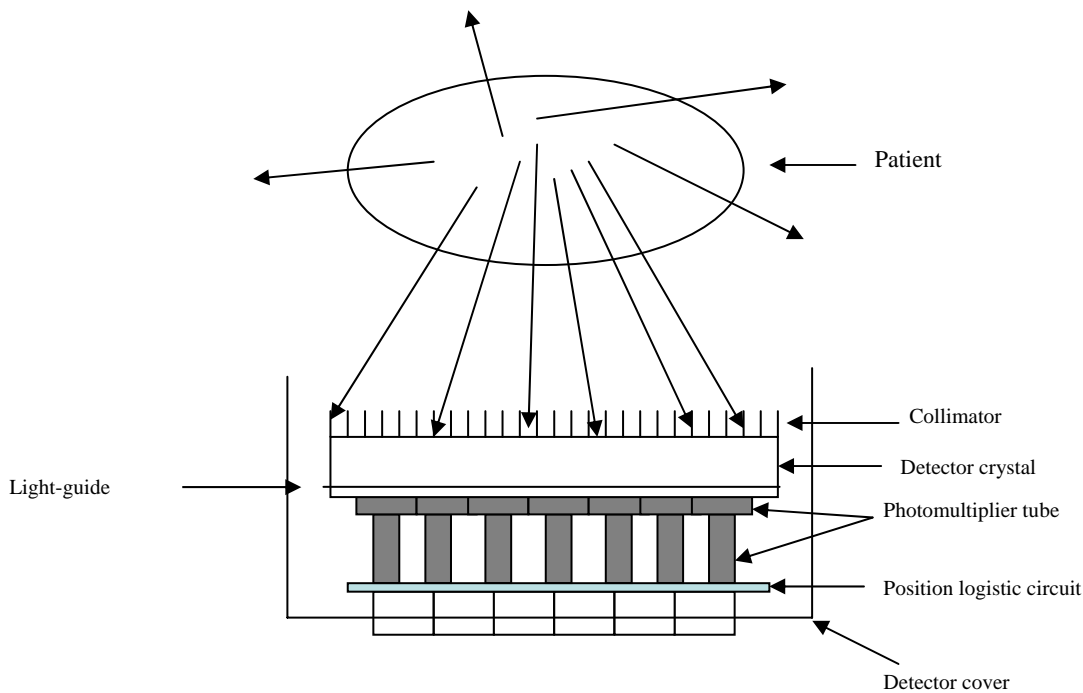
Myocardial perfusion SPECT (MPS) is typically performed as a combination of a stress study and a rest study in order to determine if a defect in the images indicate the presence of reduced blood-flow (ischemia) or cell-death (infarct). In ischemia, atherosclerosis involves the heart arteries and reduces the flow which results in insufficient blood supply to the heart muscle. In the state of rest, there exists sufficient blood supply for the heart, but at stress where the heart may require a factor of 4-6 more oxygen, reduced blood flow can cause pain. As the disease progresses pain may occur with minimal exertion or even at rest.

$^{99}\text{Tc}^{\text{m}}$ -labelled Sestamibi is a tracer useful for measurement of regional myocardial blood flow and regional ventricular function. Sestamibi has the ability to be taken up and be fixed into the myocardial tissue in proportion to regional myocardial blood flow. Once the tracer has entered the myocardium, it remains in the mitochondria as result of a passive diffusion process (1). Thus, the study therefore gives information about the blood perfusion at the time of injection, although the SPECT measurement is performed up to 30-60 minutes afterwards. The myocardial uptake of  $^{99}\text{Tc}^{\text{m}}$  labeled Sestamibi is about 1.2% and 1.5% of administered activity at stress and rest, respectively (2). The  $^{99}\text{Tc}^{\text{m}}$ -Sestamibi is also taken up by the various organs of the body according to the properties of the radiopharmaceutical and the status of the patient.

Myocardial perfusion imaging using  $^{99}\text{Tc}^{\text{m}}$ -Sestamibi can be performed either as one- or two-day protocol. In the two-day protocol, the radiopharmaceutical is administered to the patients when they have their maximum exercise study on the first day (using bicycle ergometer, treadmill or pharmacological stressing). Images obtained from this study are then compared with images obtained the following day when the  $^{99}\text{Tc}^{\text{m}}$ -Sestamibi is injected at rest. If the stress-study is entirely normal the patients usually do not need to proceed with the rest study. In the one-day protocol both rest and stress studies are performed on the same day. Imaging should begin 30-60 min post-injection; longer delays are required for rest and stress studies with vasodilators alone due to higher liver uptake (2).

The location of the decays is determined from the emitted 140 keV photons by a scintillation camera since a fraction of these photons is able to penetrate through the surrounding body tissue. The essential components in the scintillation camera (Fig. 1) are the collimator, the detector crystal, light-guide, several photomultiplier tubes (PMT) and a position logic circuit connected to a data analysis computer. The collimator is necessary to absorb photons in other directions than along the collimator holes, allowing only aligned photons to reach the detectors. Photons passing through the collimator interact in the NaI(Tl) scintillation crystal by photon absorption and/or Compton scattering and the emitted light from indirect scintillations are measured by a grid of PMTs. The emission of scintillation light is guided to the PMTs, where these devices generate an amplified electrical signal proportional to the

imparted energy. These amplified signals are finally converted to a digital position (x, y) signal and an energy signal by an A/D converter.



**Figure 1:** The components of the scintillation camera head.

Scintillation camera imaging is inherently a two-dimensional procedure where photons, that are emitted from a source within the patient along lines (or projections), perpendicular to the detector face, are detected. The number of photons detected by the camera is a measure of the activity along that line if no attenuation in the patient occurs. To create three-dimensional images, several projections need to be acquired at different angles around the patient. These projections are then used to reconstruct transversal images of the radionuclide distribution either by filtered backprojection methods (FBP) or by iterative reconstruction methods.

Direct backprojection essentially reverses the process of image acquisition by smearing the projection data along the geometry from which it was obtained. As the projection data from a number of different angular profiles are smeared back onto the image matrix, areas of high and low intensity reinforce each other building up the image in the computer. Direct backprojection is not useful since it produce blurred images with poor spatial resolution. FBP involves taking the 1D Fourier transform of the acquired projection data and applying a 1D ramp filter to appropriately weight the frequency components of the projection data. The inverse Fourier transform of the projection data is then performed and backprojection is used to obtain the image estimate of the original object.

Iterative algorithms are based on a statistical criterion and estimate the reconstructed image from the measured projection data using iterative estimation techniques designed to meet the specific statistical criterion. Iterative algorithms often start with an initial estimate of the final image, usually an initial FBP reconstructed image or uniform distribution. Projections are then calculated from the image estimate and compared to the measured projection data that was obtained by the SPECT scintillation camera. The image estimate is then updated from the

information of the quotients between the measured and calculated projections. The process is repeated until the image estimate is determined by some statistical criterion to be a close representation of the true image.

Tomographic reconstruction of projection images produces 3D data sets, which often are displayed as separate transaxial images orientated perpendicular to the long axis of the patient. Because the orientation of the heart relative the patient's long axis varies from patient to patient, it is a common practice to reorient the transaxial images into short-axis images, which are perpendicular to the long axis of the left ventricle, vertical and horizontal long-axis images.

By comparing stress and rest images, the physician can decide about the presence of an infarct or an ischemic disease. However, the images are often subjected to photon attenuation and this and other degradation effects make it sometimes difficult to accurately evaluate the severity of the disease. In order to account for this, quantitative perfusion software (QPS) has been developed that from a software-specific database that has been developed from studies of normal patients make comparison to the present study and quantify the regional differences in uptake by a scoring system. The relative activity distribution (i.e. the relative perfusion) of the left ventricular wall is thus compared to the normal relative distribution in the database. The database contains patients, who have made a SPECT examination at similar conditions (male/female, Sestamibi) and have been decided as being normal regarding heart disease.

The aid, provided by these programs, can be difficult to evaluate. The overall problem is that the 'true' activity in a patient (i.e. the actual radiopharmaceutical distribution) is not known. Phantom experiments by using physical phantoms can be helpful but generally these are limited by being not so very patient-realistic and flexible. An alternative to physical phantom measurements is to use a computer phantom where lesions can be defined with a great flexibility. If such phantoms are used in a scintillation camera Monte Carlo simulation program, realistic SPECT projections can be simulated. These projections can then be imported into the clinical software and evaluated using the same procedures as for a real patient study.

In order to fully evaluate the aid from these quantification programs, human observer receiver operating characteristic (ROC) analysis can be used, which is used in many studies; for example King et al (3) used to evaluate the different reconstruction methods in combination with attenuations, scatter and resolution compensation to decide which combination improves best the accuracy of detection of CAD. By subjecting images of known defects in a controlled manner to different observers with the task given to grade the probability of the presence or the absence of a lesion, the fraction of true positive decision (the sensitivity) can be plotted against the fraction of false-positive decision (1-specificity). These curves are often called ROC curves. The ROC curves are a plot of the sensitivity of a particular investigation versus its false-positive fraction for all possible cut-points – i.e. the point used to define positive and negative test results. The area under the ROC curve is also useful as a measure of the accuracy of a diagnostic procedure for CAD detection. The term *Receiver Operating Characteristic* curve comes from the idea that, given a particular curve, we, the receiver of the information, can use (or operate at) any point of the curve by using the appropriate cut-point. The clinical application determines which cut-point to be used (5).

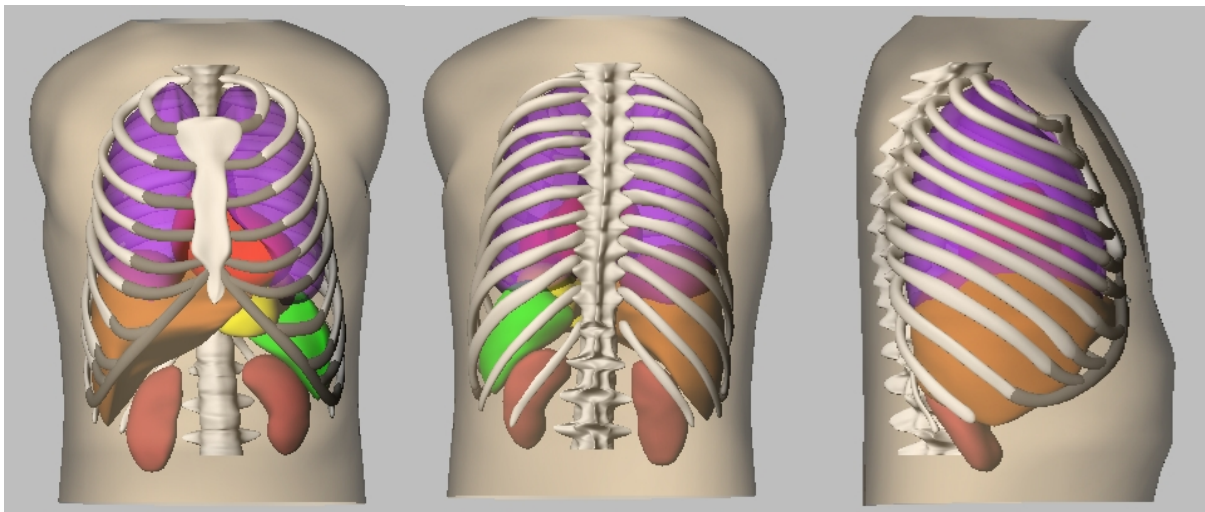
The purpose of this study was to evaluate the benefit of commercial quantification software in perfusion SPECT imaging by simulating a group of patients and their corresponding SPECT

acquisitions using a well-established computer phantom and a Monte Carlo program. Based on decisions from experienced physicians (specialists in nuclear medicine) and their visual interpretation of simulated patient images without and with addition of the information from the quantitative perfusion software, the benefit of such a quantification program was evaluated from a ROC analysis.

## Materials and Methods

### *The computer phantom*

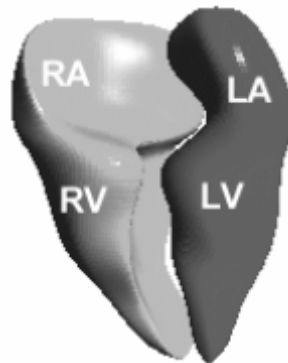
We have used the NCAT computer phantom, developed by Segars et al. (4), because of its flexibility in its way of defining the patient geometry, the shape of the heart and the ability to put defects within the heart. The phantom uses non-uniform rational B-splines (NURBS) fitted to actual patient data. The NURBS primitives define continuous surfaces allowing the phantom to be defined at any resolution. By fitting NURBS to patient data, the phantom is more realistic than those based on solid geometry. The phantom may also be extended to the fourth dimension, time, to model cardiac and respiratory motion (4). Some of the parameters that can be changed are: the size of heart ventricles; the size of the body, heart and ribcage; the breast type (supine or prone) and their dimension; the type of the motions (beating heart only, respiratory motion only or both), the cyclic period of respiration and heart; the extent of diaphragms motion. All of the organ and skeletal models (Fig. 2) are based on CT scans from the Visible Human male data set (7) with the exception of the heart, which is based on a gated MRI cardiac scan of a normal patient (Fig. 3).



**Figure 2:** Surfaces for the various organs in the torso and for the skeletal system used in the NCAT phantom. (Figure taken from: <http://www.bme.unc.edu/~wsegars/>).

In this study, the NCAT phantom was used to generate twenty male patients of normal size, where six patients were without any perfusion defects in the left ventricular (LV) wall and fourteen patients had different perfusion defects in the LV wall. By averaging eight frames of the phantom over the cardiac cycle we included the image blur due to cardiac motion. The activity distribution in the phantom was simulated to correspond to a total administered activity of 600 MBq  $^{99}\text{Tc}^{\text{m}}$ -Sestamibi, which is the activity level used in the clinical situation

for patients with a weight ranging between 80 and 100 kg. The relative organ uptake of  $^{99}\text{Tc}^{\text{m}}$ -Sestamibi per unit of volume is given in Table 1 (8).



**Figure 3:** Reconstruction of the heart into NURBS surfaces. Models of the inner heart chambers are shown with right atrium (RA), right ventricle (RV), left atrium (LA), and left ventricle (LV) indicated. Image is taken from reference (6).

Table1.

Organ	Relative activity per unit volume
LV myocardium	100
RV myocardium	60
Liver	40
Lungs	5
Gall bladder	250
Kidneys	140
Spleen	22
Stomach	3
Remaining thorax	2

Myocardial perfusion defects can occur in a variety of locations, shapes and sizes and consequently there are no specific guidelines for modeling myocardial defects (9). We decided to use defect sizes and locations based on the examination of SPECT images of real patients. The defects were both transmural, with an activity reduction through the whole LV wall and subendocardial, where the activity reduction is only through a fraction of the LV wall. The size of the defect (*extent*) varied between 4% and 19 % of the left ventricular wall volume (Table 2). The location of the defect was varied throughout the left ventricular wall (as is shown in Fig. 4) with the exception that no defects were placed at the apex. The reason for this was two-folded. Firstly, it is very difficult to define a defect in the apex in the computer phantom, because there is a risk that the defect will land outside the left ventricular wall. Secondly, observations of clinical images have showed that the detection of defects situated close to the apex is very difficult (9). The activity uptake in the defects for the seven locations, described in Table 2, was 75% and 85% of the normal activity in the left ventricular wall. This leads to total number of 14 different defects.



Table 2  
Location and size of the LV wall defect.

Defect	Lesion	Location	Size rel. the LV wall
A	TM	Apical septal	10.5%
B	TM	Mid anteroseptal	18.3%
C	SE <sub>30%</sub>	Mid anteroseptal	11.2%
D	TM	Mid anterolateral	14.3%
E	SE <sub>50%</sub>	Mid inferoseptal	3.9%
F	TM	Mid inferoseptal	7.6%
G	TM	Apical lateral	9.5%

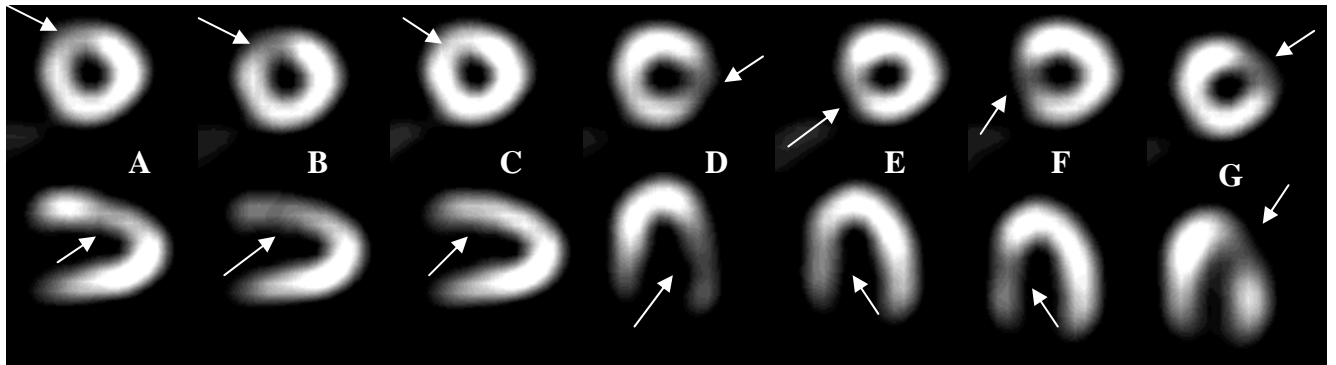
LV = Left ventricular

Mid = Midventricular

TM= Transmural lesion

SE<sub>30%</sub> = Subendocardial lesion through 30% of the LV wall starting from the inside

SE<sub>50%</sub> = Subendocardial lesion through 50% of the LV wall starting from the inside



**Figure 4:** Size and location of the 7 defects used in this study are shown. Top row shows the short-axis images and bottom row shows either vertical long-axis (A-C) or horizontal long-axis (D-G) images. Arrows indicate defect location. The size of the defects relative the LV wall volume was as follows: A (10.5% -TM), B (18.3% -TM), C (11.2% - SE<sub>30%</sub>), D (14.3%-TM), E (3.9%-SE<sub>50%</sub>), F (7.6%-TM), and G (9.6%-TM).

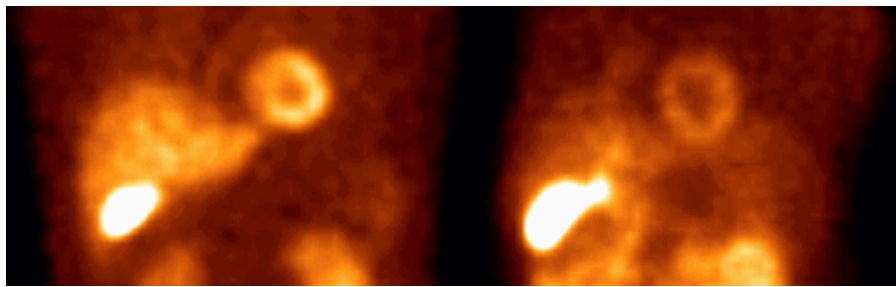
### *The simulated computer scintillation camera*

The NCAT program outputs transversal images of the activity distribution and a corresponding density distribution. These images are then used as input by the Monte Carlo program SIMIND (10) to define the patient. The simulated SPECT system mimicked a Vertex™ SPECT scintillation camera system (Philips Medical Systems) equipped with a low-energy high-resolution parallel-hole collimator and a 9.35 mm thick NaI(Tl) crystal. Projection data were generated in 64x64 matrix with a pixel size of 0.63 cm. The simulation was done for 64 angular views round 180° starting at angle 315°. The energy resolution was 10% at 140 keV and a 20% energy window was centred on the 140 keV <sup>99</sup>Tc<sup>m</sup> photo-peak.

The Monte Carlo method is a general and widely used technique for simulation of stocastical processes. In this case, the photons emitted from a decay is followed in a computer phantom and each direction and energy deposition is calculated based on random numbers and cross-section data. The great advantadge with this method is that each detail is known and the complex scheme that build-up an image can be studied in detail. It is also possible to very

accurate define 'diseases' in the patient since it is based on a model. The program SIMIND that we have used has been developed at Lund University for many years and has been validated by many users. It is written in Fortran-90 and more information can be provided at the web site [www.radfys.lu.se/simind](http://www.radfys.lu.se/simind))

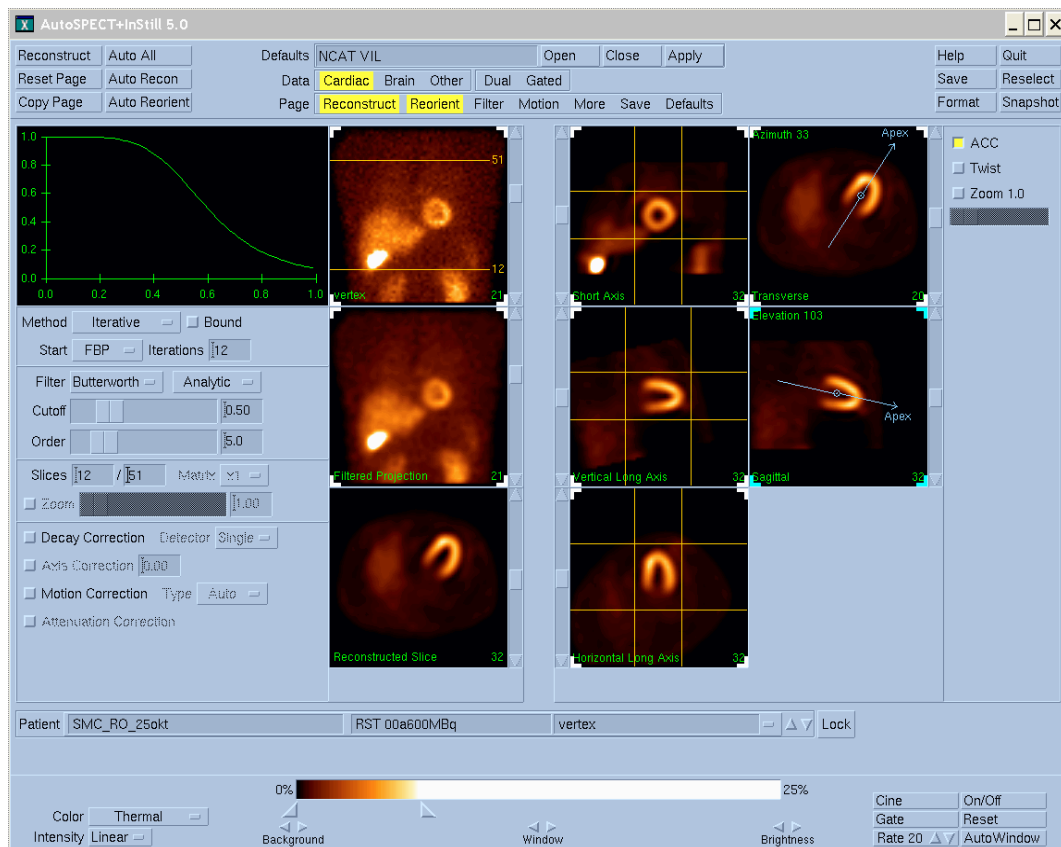
SIMIND provides images (Fig. 5) normalized to an imaging situation where the counts in the images match an activity in the phantom of one MBq and an acquisition time of one second. To achieve a proper noise level, the projection was multiplied by an acquisition time per projection of 20 seconds and an activity within the field-of-view of  $0.4 \cdot 600$  MBq (the factor 0.4 is an estimation of the activity present in the torso per unit administered activity administered to the whole body). After rescaling of the projection data, Poisson distribution noise was added pixel-by-pixel to produce images of clinically realistic noise levels.



**Figure 5:** Example of projection images. Left figure show a SIMIND simulated projection using the NCAT phantom and right figure show a projection from a real patient examination.

### *Image Reconstruction*

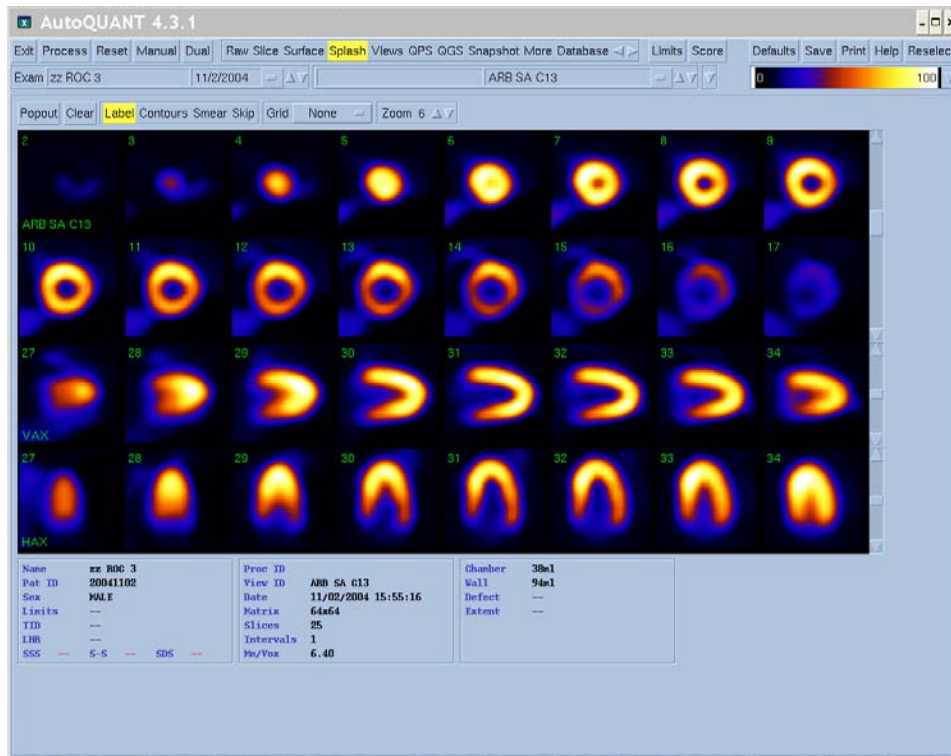
The SPECT projections were transferred to a clinical image processing station and reconstructed to a set of transversal images by the AutoSPECT™ software (Philips Medical Systems) using an iterative ML-EM algorithm. Twelve iterations were used according to the clinical procedure at our department. The reconstruction included a Butterworth post-filtering of the projections with a cut-off frequency of 0.25 cycles/pixel and a filter order of 5. No attenuation or scatter correction was applied, because the normal database (see *Observer study*) is based on data from images reconstructed without attenuation and scatter correction. The transaxial tomographic slices were reoriented into short-axis, horizontal and vertical long-axis images. To maintain the same heart position the azimuth and elevation angles were fixed to  $33^\circ$  and  $103^\circ$ , respectively, for all reconstructed images (Fig.6).



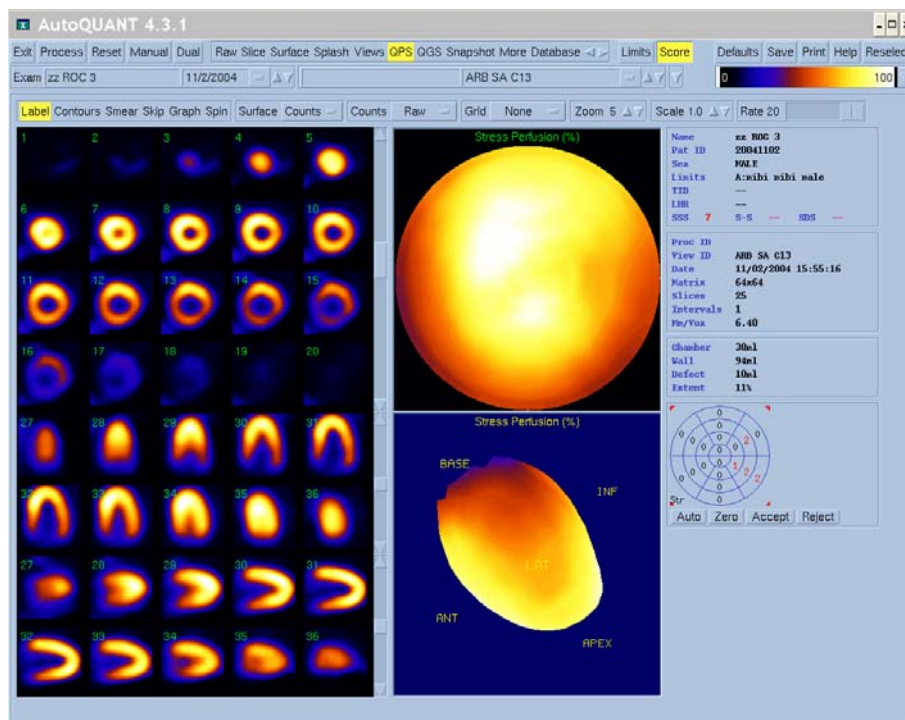
**Figure 6:** The reconstruction window of AutoSPECT<sup>+</sup> where the reconstruction method, the number of iterations, the filter and the azimuth and elevation angles can be seen.

### *Quantitative Perfusion Software*

A commercial software, AutoQuant<sup>TM</sup> (Philips Medical Systems), was used at both reading sessions. This software is developed especially for the reviewing, analyze and quantification of myocardial perfusion SPECT images. Two viewing windows were used: The splash window where the display of the short, horizontal and vertical long-axis images of the left ventricle was utilized (Fig. 7) and 2) the Quantitative Perfusion SPECT (QPS) window where the display of the short, long-axis images, 2D and 3D polar perfusion maps were utilized (Fig. 8). If so desired, the observers were allowed to change the color table, the intensity and brightness, and the magnification (zoom) of the images.

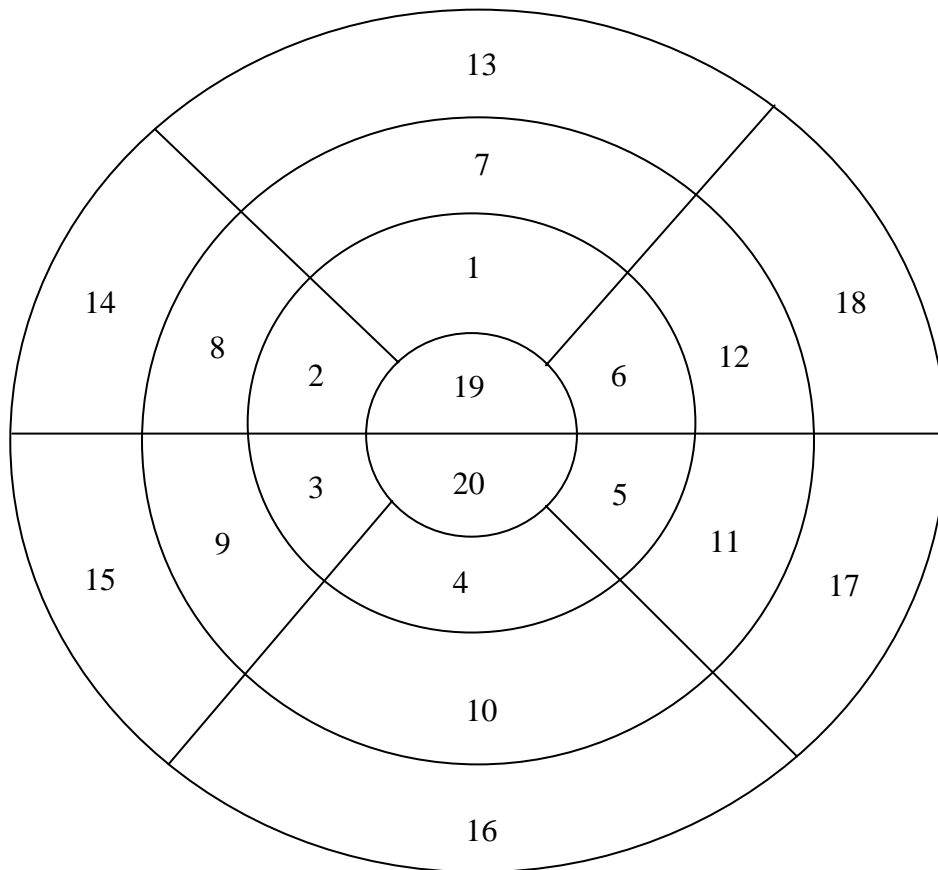


**Figure 7:** Example of information in the splash window at the first reading session. Short-axis images, horizontal and vertical long-axis images, but no quantitative information, are presented.



**Figure 8:** Example of information presented in the QPS window in connection with the second reading session. Relative information in the form of tomographic slices, 2D, 3D polar perfusion maps and quantitative information (defect, extent, and Summed Stress Score, SSS) is presented.

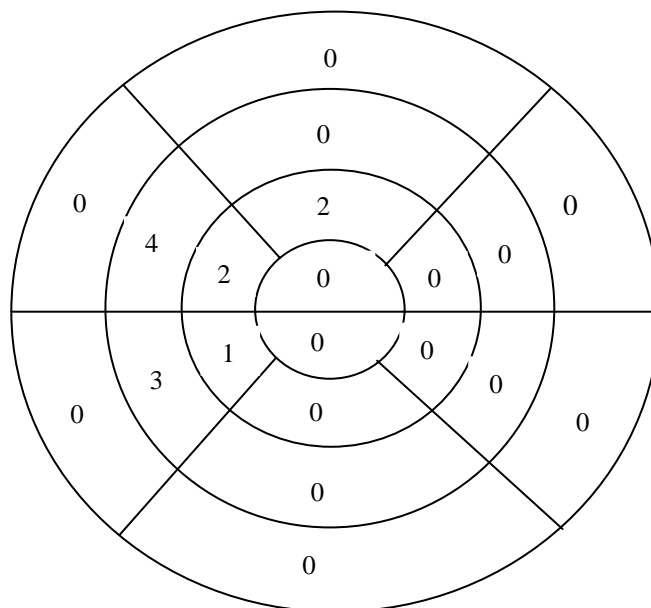
The software produces polar perfusion maps by dividing the left ventricular wall into 20 segments. The mean activity uptake of each segment is calculated and expressed as percentage of maximum activity. The 20-segment scoring system is based on three short-axis parts (distal [apical], mid, and basal) to represent the entire left ventricle, with the apex represented by two segments visualized in a midvertical long axis image. Each of the 20 segments has a distinct number (11), as indicated in Fig 9.



	DISTAL	MID	BASAL
ANTERIOR	1	7	13
ANTEROSEPTAL	2	8	14
INFEROSEPTAL	3	9	15
INFERIOR	4	10	16
INFEROLATERAL	5	11	17
ANTEROLATERAL	6	12	18
ANTEROAPICAL	19		
INFEROPAPICAL	20		

**Figure 9:** Diagrammatic representation of segmental division of the SPECT slices and the assignment of individual segment to individual coronary arteries.

To provide quantitative information, the software compares the SPECT examination with the database of patients, who have made a SPECT examination at similar conditions male/female, Sestamibi) and have been found to be normal using other methods. The relative activity distribution in the polar perfusion map is compared to the normal relative distribution in the database. Activity of each segment is compared to the database and differences in each segment are scored as follows: 0 = normal, 1 = slight reduction of uptake (equivocal), 2 = moderate reduction of uptake (which usually implies a significant abnormality), 3 = severe reduction of uptake, 4 = absence of active uptake. Perfusion defects with scores of 3 or 4 can be reported as consistent with a critical ( $\geq 90\%$ ) coronary stenosis (11). The result of the comparison is presented in the form of a “visual scoring diagram”, Fig. 10. In addition to the segmental score, the *Summed Stress Score* (SSS) is defined as the sum of the stress scores for the 20 segments. Consequently, the *Summed Rest Score* (SRS) is defined as the sum of the rest score and the *Summed Difference Score* (SDS), is defined as the difference between SSS and SRS and measures the degree of the reversability (the difference of activity uptake for each segment in the polar perfusion maps between the rest and stress studies). SSS < 4 are considered normal or nearly normal, SSS of 4-8 are considered mildly abnormal, SSS of 9-13 moderately abnormal, and SSS > 13 severely abnormal (11). The summed scores can thus be considered as global indices of perfusion.



**Figure 10:** The numerical value for a segment in the visual scoring diagram indicates the amount of perfusion deviation in this region relative to the normal database. Values vary from a normal of 0 (normal perfusion), to a maximum abnormal of 4 (no perfusion).

Apart from the quantitative information mentioned (i.e. SSS), information about the *Defect* and the *Extent* is presented (see Fig. 8). The *Defect* is defined as the LV defect volume computed using normal database, and *Extent* is defined as the LV defect volume, as a percentage of LV wall volume (12). The same quantitative information was also presented in the splash window at the second reading session.

### *ROC- analysis*

To evaluate the aid of the quantitative perfusion information Receiver Operating Characteristic (ROC) analysis was used. The observer study and ROC analysis followed the basis methodology established by Metz (13, 14).

At our department the myocardial SPECT examination is performed according to a two-day protocol where the stress study on the first day is followed by a rest study one or two days later. If the stress study is entirely normal, the rest study is mostly omitted. The observer study was set up to mimic the task of reviewing a stress study in order to decide if a rest study is necessary.

Four experienced physicians interpreted the images from the 20 computer patients. The observers were asked to interpret the stress study for each of the 20 computer patient and decide the presence (pathological) or absence (normal) of a defect based on a four-point scoring scheme where 1=normal, 2=probably normal, 3=probably pathological, 4=pathological.

To analyze the benefit of the additional quantitative perfusion information, the viewing of all the 20 computer patients was performed at two different reading sessions. In the first reading session, further on denoted as "V", only visual viewing of the tomographic slices was possible. An example of such a display is given in Fig.7. In the second reading session, denoted "V+Q", the quantitative perfusion information was used in connection with the visual viewing analysis (Fig.8). The time between the two reading sessions was eight weeks or longer for all observers.

The *sensitivity* and *specificity* describe the abilities of a diagnostic procedure to enable one to correctly diagnose disease when disease is actually present and to correctly rule out when the disease is truly absent. The ROC curve is a plot of sensitivity, or true positive fraction (TPF), of a test versus its false positive fraction (FPF = 1- specificity) for all possible cut-points, i.e. the point or threshold used to define "positive" and "negative" test results. Each point on the graph corresponds to a certain cut-point. Straight lines can be used to connect the data points from all possible cut-points. The resulting empirical ROC curve illustrates the dependence between the sensitivity (TPF) and the FPF. A fitted or smoothed ROC curve can be drawn on the basis of an assumed distribution of the test results. By assuming a bi-normal distribution (i.e. two normal distributions where one is for the test results of patients without CAD and the other is for the test for the patients with CAD) we have generated ROC curves by curve fitting using the ROCKIT program, developed by Metz (15). The area under the ROC curve was used as a measure of accuracy of the diagnostic procedure. The area under empirical ROC curves was calculated according to a method described below, and the area under fitted ROC curves was provided by the ROCKIT program.

The ROCKIT program is designed to fit binormal ROC curves to both continuously-distributed and ordinal category (e.g., "confidence-rating") diagnostic test results. The purposes of ROCKIT are; i) to calculate maximum-likelihood estimates of the parameters of a conventional "binormal" model for the input data, ii) to calculate maximum-likelihood estimates of the parameters of a "bivariate binormal" model for data from two potentially correlated diagnostic tests and, thus, to estimate the binormal ROC curves implied by those data and their correlation and iii) to calculate the statistical significance of the difference between two ROC curve estimates using any one of three distinct statistical tests: Bivariate test, Area test and TPF test (15).

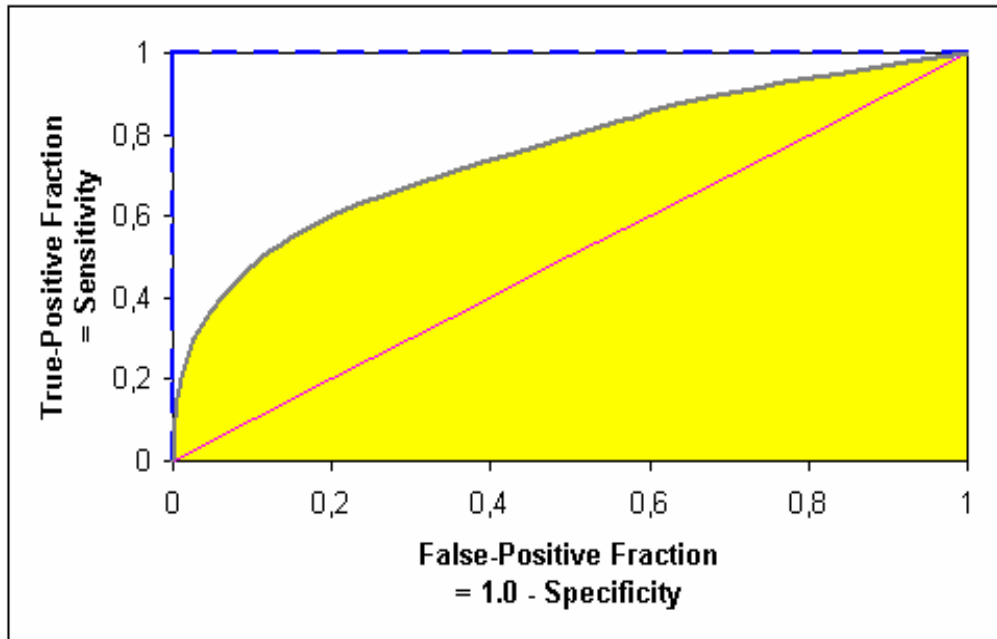
The area under the empirical curve, denoted below as  $A$ , can be estimated as follows (5): Firstly, consider every possible pairing of patients with disease and patients without disease. Secondly give each pair of patients a score of 1.0 if the test result for the patients with disease is higher (i.e., more suspicious for disease), a score of 0.5 if the test results are the same, and a score of 0.0 if the test result for the patients with disease is lower (i.e., less suspicious for disease). Thirdly, take the sum of these assigned scores. If there are  $N$  patients without disease and  $M$  patients with disease in the sample group, then there are  $M \times N$  scores. Finally, divide the sum of the scores by number of scores ( $= M \times N$ ). The result provides an estimate of the area under the empirical ROC curve. The process described can be written mathematically as follows:

$$A = \frac{1}{(M \times N)} \sum_{(j=1)}^M \sum_{(K=1)}^N score(X_j, Y_K) \quad (\text{Equation 1})$$

Where  $X_j$  denote the test score of the  $j_{th}$  patient with disease and  $Y_k$  denote the test score of the  $k_{th}$  patient without disease.  $A$  is the estimate of the area under the empirical ROC curve, and  $score(X_j, Y_k)$  is the score assigned to the pair composed of the  $j_{th}$  patient with disease and the  $k_{th}$  patient without disease. The score will equal 1 if  $X_j$  is greater than  $Y_k$ , 0.5 if  $X_j$  is equal to  $Y_k$ , and 0 if  $X_j$  is less than  $Y_k$ . See Appendix I for an example.

The area under the ROC curve varies between 0.0 and 1.0. A diagnostic procedure for CAD detection with an area under the ROC curve of 1.0 is perfectly accurate because the TPF (sensitivity) is 1.0 when the FPF is 0.0. In contrast, a diagnostic procedure for CAD detection with an area of 0.0 is perfectly inaccurate. The closer the ROC curve area is to 1.0, the more accurate is the diagnostic procedure. The line segment ranging from  $[0, 0]$  to  $[1, 1]$  has an area of 0.5 and is called the *chance diagonal* (Fig. 11). The ROC curve would be expected to match the line of identity (red line) if we relied purely on random guessing to distinguish a patient with disease from a patient without disease. The practical lower bound for the ROC curve area is therefore equal to 0.5 (5).





**Figure 11:** Examples of ROC curves. A perfect test has an area under the ROC curve of 1.0 (blue line). The chance diagonal has an ROC curve area of 0.5 (red line). Tests with some discriminating ability have areas between these two extremes. The closer the ROC curve area is to 1.0, the better the accuracy of the diagnostic procedure.

The area under the ROC curve has several interpretations (5). It can be regarded as 1) the average sensitivity for all possible values of specificity (or FPF's), i.e. the average ROC curve height, 2) the average specificity for all possible values of the sensitivity or 3) the probability of distinguishing correctly between a randomly selected actually positive case and a randomly selected actually negative case. As an example regarding point 3); if the area under the ROC curve is 0.87 this means that if we select two patients at random, one with CAD and one without, then the probability will be 0.87 that the patient with CAD will have more suspicious myocardial perfusion SPECT (5).

## Results

In Table 3 is given the results from the observer's interpretation of the 20 computer patients at the two readings sessions. The results from the first reading session (V) shows that one of four observers interpreted all of the six healthy computer patients as normal, which was a perfect result. Two observers interpreted four and five patients as normal and the remainder observer interpreted three patients as normal. Observer O2 interpreted one of the healthy computer patients as probably pathological. For the fourteen pathological computer patients, observer O2 interpreted two of them as pathological and seven patients as normal. The observer O3 also interpreted two patients as pathological and nine of them as normal.

Observer O1 and O4 interpreted none as pathological and seven and six patients, respectively, as normal.

In the second reading session (V+Q), the observers O2 and O4 interpreted four of the healthy patients as normal and two as probably normal. The first observer interpreted three patients as normal and three as probably normal. The third observer interpreted only two patients as normal and three as probably pathological. For the fourteen pathological patients, observer O3 interpreted seven as pathological and five as normal. Observer O1 and O2 interpreted one patient as pathological and six patient as normal. The fourth observer did not interpret any of the patients as pathological and scored four as normal.

The quantitative information does provide aid to some of the observers to get a more accurate result, but it was not general for all of the observers. It seems like there is a factor that depends on their trust to the quantitative information calculated by the software. As can be seen from Table 3 observer O1, O3 and O4 scored a better result (i.e. more patients interpreted as normal) when evaluating the six healthy computer patients at the first reading session where no quantitative information was provided. On the other hand, for observer O2, the result was more accurate at the second session when the quantitative information was available. From the evaluation of the fourteen pathological computer patients, the accuracy increased (i.e. more patients interpreted as pathological) for observer O1 and O3 at the second reading session when the quantitative information was available. For observer O2, two pathological computer patients, compared to one, were interpreted as pathological when no quantitative information was available. Regarding observer O4, none of the pathological computer patients were interpreted as pathological at the two reading sessions.

Table 3  
Results from the observer study of the 20 computer patients

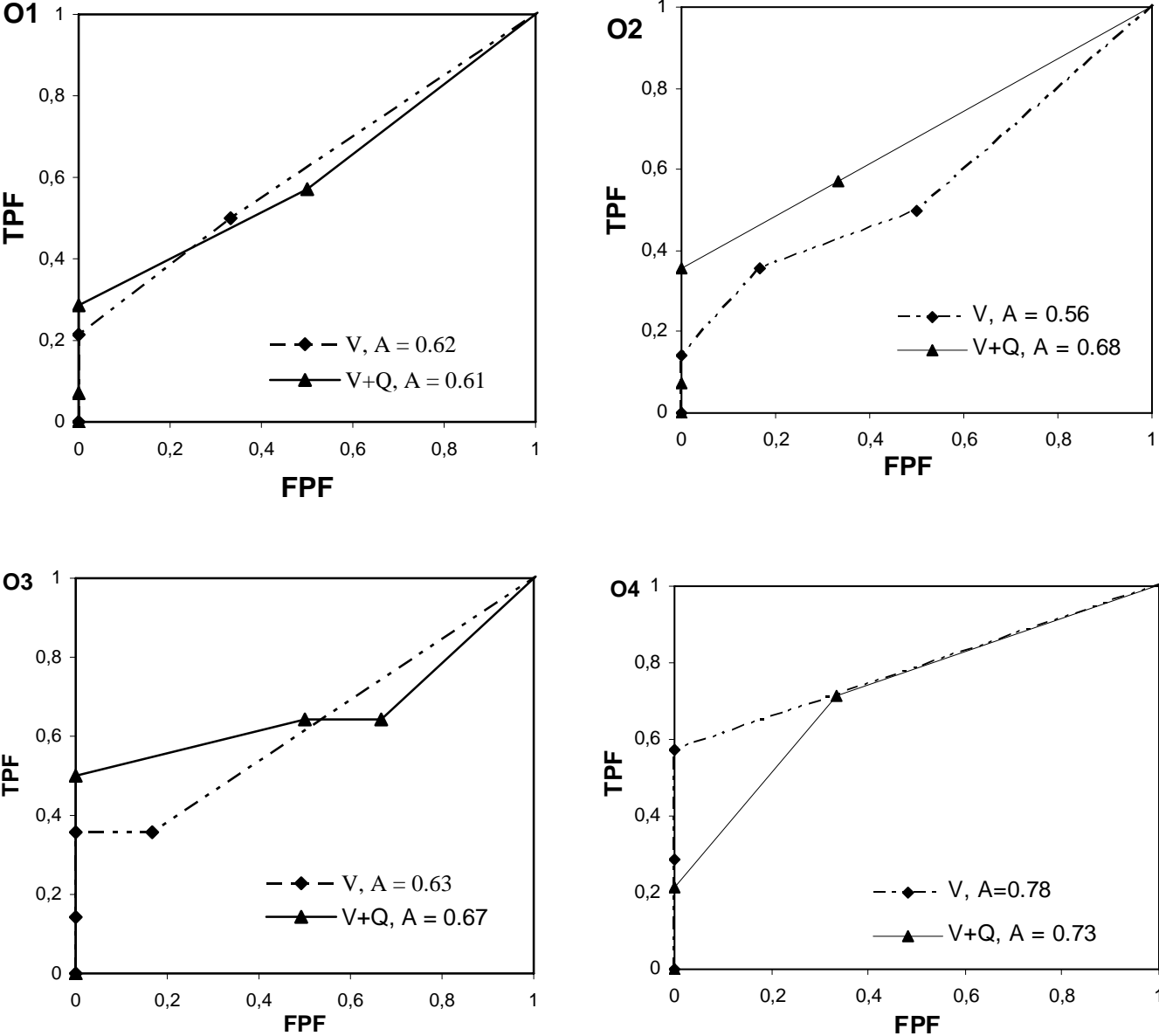
	Clinical Interpretation				Total
	Normal	Probably Normal	Probably Pathological	Pathological	
<u>O1</u>					
V					
Normal	4	2	0	0	6
Pathological	7	4	3	0	14
V+Q					
Normal	3	3	0	0	6
Pathological	6	4	3	1	14
<u>O2</u>					
V					
Normal	3	2	1	0	6
Pathological	7	2	3	2	14
V+Q					
Normal	4	2	0	0	6
Pathological	6	3	4	1	14
<u>O3</u>					
V					
Normal	5	1	0	0	6
Pathological	9	0	3	2	14
V+Q					
Normal	2	1	3	0	6
Pathological	5	0	2	7	14
<u>O4</u>					
V					
Normal	6	0	0	0	6
Pathological	6	4	4	0	14
V+Q					
Normal	4	2	0	0	6
Pathological	4	7	3	0	14

O1, O2, O3, and O4 are the four observers

The design of the ROCKIT program is such that it does not generate fitted ROC curves for each observer from the empirical data. Therefore, no area under the curve was calculated. Instead, the areas under the empirical ROC curves for each observer were calculated using the method, previously described. The corresponding areas under the empirical ROC curves are included in Fig. 12 and also shown in Table 4.

Fig. 12 shows the empirical ROC curves for the four observers and for the two reading sessions. The result indicates that for observer O2 and O3, the evaluation with quantitative information provided higher detection accuracy, whereas for observer O4 the evaluation

provided higher detection accuracy without the quantitative information. For the first observer, there was no difference in the detection accuracy with and without the quantitative information.



**Figure 12:** The empirical ROC curves for each observer (O1, O2, O3 and O4) at the two reading sessions (V and V+Q) are shown.

Table 4 compares the differences in the calculated area under the empirical ROC curves between the two reading sessions for each observer. The first observer obtained almost the same result at the two reading sessions; the difference in the area was equal to -0.01. The second observer obtained an improved accuracy in the second reading session as compared to the first session; which can be seen in the area difference of 0.12. For observers O3 and O4, the area differences were found to be 0.04 and - 0.06, respectively.

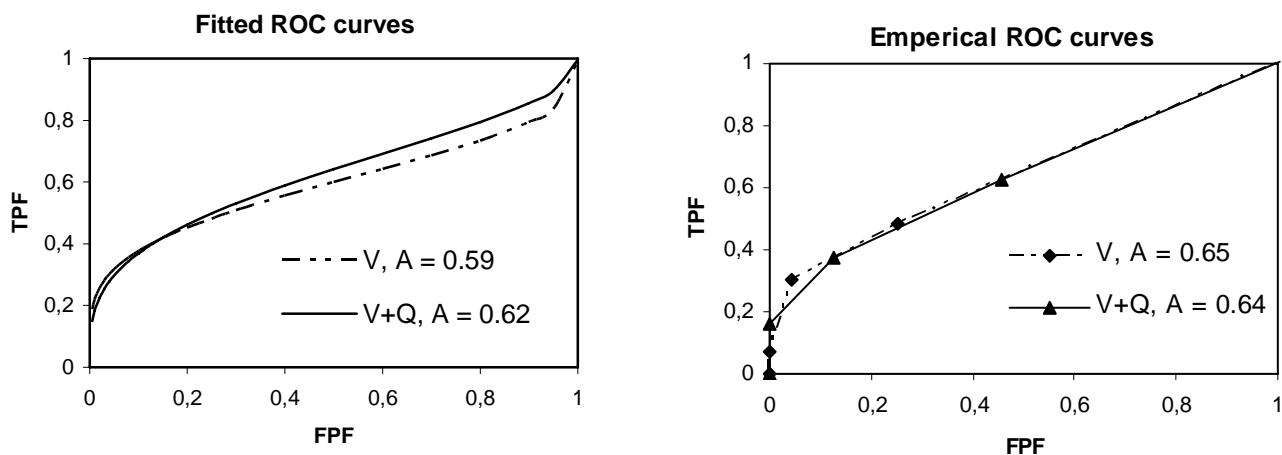
Table 4  
Comparison of the area under the empirical ROC curves at the two reading session (V, V+Q)  
for individual observers

Observer	V	V+Q	V vs. V+Q Comparison of A
	A	A	$A_{V+Q} - A_V$
O1	0.62	0.61	-0.01
O2	0.56	0.68	0.12
O3	0.63	0.67	0.04
O4	0.79	0.73	-0.06
<b>Average</b>	<b>0.65</b>	<b>0.67</b>	<b>0.02</b>

In an attempt to summarize the results over all observers for the two evaluation methods used in this study, we can assume that we have only one observer, who evaluates the 20 computer patients at four different occasions (Table 5) with the two methods V and V+Q. Fig. 15 compares such a “summarize” by displaying fitted and empirical ROC curves for each of the evaluation methods. The area under the fitted ROC curves were calculated by the ROCKIT program for session V and V+Q and were found to be  $0.59 \pm 0.09$  and  $0.62 \pm 0.07$ , respectively (mean $\pm$ SD). The difference in area under the fitted ROC curves between session V+Q and session V was found to be 0.03, which is not a significant improvement. The corresponding area under the empirical ROC curves for the two reading sessions V and V+Q was 0.65 and 0.64, respectively, yielding an area difference of -0.01. The difference in area between the fitted and the empirical ROC curves depends on; the few points used to plot the empirical curves, the resulting empirical ROC curve illustrates the dependence between the sensitivity (TPF) and the FPF and the fitted or smoothed ROC curve can be drawn on the basis of an assumed bi-normal distribution of the test results (see *ROC-analysis*). If we compare the average area (Table 4) for the four observers with the area under the overall fitted and empirical ROC curves this shows that this 'summarize' method was probably not the best method to generalize the results. This is one of several reasons why Metz discourages to use this method to generalize the results for this kind of study. To do that, other programs and other parameters than the area under the ROC curves should be used (16).

Table 5  
Results from this study with 20 computer patients accumulated as if evaluated four times by  
the same observer

	Clinical Interpretation				Total
	Normal	Probably Normal	Probably Pathological	Pathological	
<b>V</b>					
Normal	18	5	1	0	24
Pathological	29	10	13	4	56
<b>V+Q</b>					
Normal	13	8	3	0	24
Pathological	21	14	12	9	56



**Figure 13:** Show the summarized fitted ROC curves at the left and the empirical ROC curves at the right for one assumed observer evaluating the 20 computer patients at four occasions.

Table 7 compares the observer's interpretation and the AutoQuant™ evaluation of the 20 computer patients. According to AutoQuant™ only four computer patients was pathological resulting in a SSS value between 5 and 7. Two of the computer patients showed a SSS equal to 4 where one was pathological and the other was normal. According to Table 7, the SSS values for four of the six healthy computer patients varied between 1 and 4, and the SSS values for five of the fourteen pathological patients became equal to 0. These unwanted scoring results will most likely smooth the differences between V and V+Q.

We found that the observers within the group interpreted the computer patients differently from each other at the two reading sessions V and V+Q. This can be explained by the fact that they have their own mental 'threshold' to rank images and how they shall use the quantitative information, as provided by the software. For example at the second reading session; the patient "G<sup>2</sup>" was interpreted as normal by observer O1 but pathological by the observers O2 and O3 and probably normal by observer O4. The difference in interpretation of the images appears also at the two reading sessions for the same observer; e.g. the first observer interpreted the patient "D<sup>1</sup>" as probably pathological at reading session V and as normal at reading session V+Q.

The calculated SSS was also depending on the defects location. For example, the patients A<sup>1</sup>, C<sup>1</sup>, and G<sup>1</sup> had defects that were almost the same regarding the volume (10.5%, 11.2% and 9.5% relative the left ventricular wall volume) but since they were placed at different locations, the SSS was calculated to 0, 0 and 5, respectively. Here, 0 denotes that the heart is normal and 5 means that the heart is mildly abnormal.

Table 7

This table show the results from the evaluation with 4- point scoring for both methods according to each observer, and the SSS values calculated by AutoQuant™

Phantom nr	O1		O2		O3		O4		SSS
	V	V+Q	V	V+Q	V	V+Q	V	V+Q	
1	N	N	N	N	N	N	N	N	0
2	PN	PN	N	PN	N	PP	N	PN	2
3	PN	N	N	N	N	PN	N	N	1
4	N	N	PN	N	N	N	N	N	0
5	N	PN	PP	PN	PN	PP	N	PN	4
6	N	PN	PN	N	N	PP	N	N	1
A <sup>1</sup>	N	N	N	N	N	N	N	PN	0
B <sup>1</sup>	N	N	N	N	N	N	PN	PN	0
C <sup>1</sup>	N	N	N	PN	N	N	N	N	0
D <sup>1</sup>	PP	N	P	PN	PP	P	PP	PP	7
E <sup>1</sup>	PN	N	PN	N	N	P	PN	PN	3
F <sup>1</sup>	PP	PN	PP	PP	P	P	PP	PP	7
G <sup>1</sup>	PP	P	P	PP	P	P	PP	PP	5
A <sup>2</sup>	N	PN	N	N	N	N	N	N	0
B <sup>2</sup>	N	PN	N	N	N	N	N	N	0
C <sup>2</sup>	N	N	N	N	N	PP	PN	PN	2
D <sup>2</sup>	N	PP	PN	PP	N	P	N	PN	5
E <sup>2</sup>	PN	PP	N	PN	N	PP	N	N	2
F <sup>2</sup>	PN	PP	PP	PP	PP	P	PP	PN	4
G <sup>2</sup>	PN	N	PP	P	PP	P	PN	PN	3

1-6: Healthy computer patients. <sup>1</sup> = 75% activity uptake. <sup>2</sup> = 85% activity uptake.

A-G are the defects location shown in Fig.4

SSS = Summed Stress Score as calculated by AutoQuant™

N = Normal, PN = Probably Normal, PP = Probably Pathological, P = Pathological

## Discussion

In this work we have presented a ROC study with the aim of evaluating the benefit of using quantitative perfusion software in addition to the standard three-view heart display. The overall ROC curves show some improvement in the area under the curve for the diagnostic procedure for CAD detection at the second session, V+Q, when comparing to the reading session V. However, this improvement was relatively moderate. For each of the observers, the result from the ROC analysis varied, probably because of their individual mentally thresholds to rank images. The quantitative information calculated by the software (AutoQuant™) may aid the interpretation of the patient examination.

The computer patient cases, chosen for this study, were based on clinical patient cases regarding size and location of the defects and the activity uptake. However, the activity uptake reduction in the defects was relatively small in the pathological patients. The reason for this was because we were interested in the cases where we thought the additional information would really make a difference.

We found that when a defect were situated in the lateral wall, it was easier for the observer to detect it than if the defects were situated in the septal wall. This holds true for both evaluation methods (compare the case A<sup>1</sup> and C<sup>1</sup> situated on the septal wall and G<sup>1</sup> situated in the lateral wall and where A<sup>1</sup> and C<sup>1</sup> were interpreted as normal by G<sup>1</sup> as pathological by the AutoQuant program). The magnitude of the activity uptake reduction in the defect also affect the accuracy in the defect detection, that is, as the activity uptake reduction becomes larger in the defect the easier it will be for the observer to detect the defect.

This study also reveals differences between each of the four observers in how they interpret the information from only viewing tomographic slices compared to also have access to the quantitative information (volume, extent, regional scoring and SSS). An example of this is the results from observer O4 who obtained a reduced accuracy when considering the quantitative perfusion information as compared the case without the information. We think that the result in this study, however, generally shows that quantitative perfusion information in connection with visual viewing may help to get a better diagnostic procedure for CAD detection.

As a conclusion, the Monte Carlo method together with a realistic computer phantom and ROC analysis can be very useful to obtain information in how observers take advantage of information from quantitative perfusion software in myocardial SPECT. Monte Carlo simulated data and ROC analysis can also be useful to evaluate the reconstruction methods (iterative reconstruction algorithm, filtered backprojection) in combination with attenuation correction, scatter correction and resolution compensation to decide which combination improves best the accuracy of detection of CAD.



## **Acknowledgements**

I would like to thank my supervisors Ulf Nilsson, Jolanta Bartosik and Michael Ljungberg for their guidance and active engagement in the thesis. Thank you for an educational and pleasant time.

I also would like to thank all the personals in the Department of Clinical Physiology at Helsingborg Hospital, and a special thanks to the physicians Morten Scheike, Germano De Pedis, Jan Dahlström, and Tomaz Tekavec, for their engagement and help.

A special thank goes to Henrik Hussein El-Ali, Department of Medical Radiation, Lund University for his help with all the computer simulations in this study.

Rouwaida

## References

1. Larock MP, Braat SH, Sochor H, Maisey M, Rigo P. *New development in myocardial imaging technetium  $^{99m}\text{Tc}$  sestamibi*. London, Great Britain: Martin Dunitz; 1993: 11.
2. Martin J, Taylor A. Imaging and processing. In: The European Association of Nuclear Medicine Technologist Committee Education Sub-Committee. *Myocardial Perfusion Imaging. A technologist's Guide*. Bristol-Myers Squibb Medical imaging; 2004: 26-31.
3. Narayanan MV, King MA. Human-Observer Receiver-Operating-Characteristic Evaluation of attenuation, scatter, and Resolution Compensation Strategies for  $^{99m}\text{Tc}$  myocardial Perfusion Imaging. *J Nucl Med*. 2003;44: 1725-1734.
4. Segars WP. *Development and application of the new dynamic NURBS-based cardiac-torso (NCAT) phantom*. Chapel Hill, NC, USA: PhD Dissertation, The University of North Carolina: 2001.
5. Obuchowski NA. Receiver Operating Characteristic curves and their use in Radiology. *Radiology* 2003; 229:3-8.
6. Segars WP, Lalush DS, Tsui BMW. A Realistic Spline-Based Dynamic Heart Phantom. *IEEE Trans Nucl Sci*. 1999; 46: 503-506.
7. The Visible Human project. Bethesda, MD, USA: United States National Library of Medicine, National Institutes of Health.
8. Wackers FJ, Berman DS. Technetium-99m hexakis 2-methoxyisobutyl isonitrile: Human biodistribution, dosimetry, safety, and preliminary comparison to thallium-201 for myocardial perfusion imaging. *J Nucl Med*. 1989 ; 30: 301-311.
9. LaCroix KJ, Tsui BM, Frey EC, Jaszczak RJ. Receiver operating characteristic evaluation of iterative reconstruction with attenuation correction in  $^{99m}\text{Tc}$ -sestamibi myocardial SPECT images. *J Nucl Med*. 2000; 41: 502-513.
10. Ljungberg M, Strand S-E. A Monte Carlo program simulating scintillation camera imaging. *Comp Meth Progr Biomed*. 1989; 29: 257-272.
11. Berman DS, Germano G. An approach to the interpretation and reporting of gated myocardial perfusion SPECT. In: Germano G, Berman DS, eds. *Clinical gated Cardiac SPECT*. Armonk, NY: Futura Publishing Company; 1999: 147-182.
12. AutoQuant™ user's manual. Milpitas, CA, USA: ADAC Laboratories; 1999.
13. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol*; 1989; 24: 234-245.
14. Metz CE. Basic Principles of ROC analysis. *Semin Nucl Med*. 1978; 8: 283-298.

15. Metz CE. *The ROCKIT 0.9 0.1 B program*. Chicago, IL, USA: University of Chicago; 1998.

16. King M, University of Massachusetts Medical Center, Worcester, USA, *personal communication*, 2004.

## Appendix I

Example from our study:

*Visual analysis*

Standard of Reference Result	N	PN	PP	P	Total
Disease Present	7	4	3	0	14
Disease Absent	4	2	0	0	6

N = Normal, PN = Probably Normal.

PP = Probably Pathological

P = Pathological.

Step 1: Identify pairs and assign scores to each pair.

### Possible Pairings

Patient with disease	Patients without disease	No. Of such pairings	Score
Normal	Normal	7*4=28	0.5
Normal	Probably Normal	7*2=14	0
Normal	Probably Pathological	7*0	0
Normal	Pathological	7*0	0
Probably Normal	Normal	4*4=16	1
Probably Normal	Probably Normal	4*2=8	0.5
Probably Normal	Probably Pathological	4*0	0
Probably Normal	Pathological	4*0	0
Probably Pathological	Normal	3*4=12	1
Probably Pathological	Probably Normal	3*2=6	1
Probably Pathological	Probably Pathological	3*0	0.5
Probably Pathological	Pathological	3*0	0
Pathological	Normal	0*4	1
Pathological	Probably Normal	0*2	1
Pathological	Probably Pathological	0*0	1
Pathological	Pathological	0*0	0.5

Step 2: Sum the (MxN) scores. This is written below as (no. of pairings times score)

$$(28*0.5)+(14*0)+(0*0)+(0*0)+(16*1)+(8*0.5)+(0*0)+(0*0)+(12*0)+(6*1)+(0*0.5)+(0*0)+(0*1)+(0*1)+(0*1)+(0*1)+(0*0.5) = 52$$

Step 3: Divide the sum by the total number of pair (MxN). This gives the estimated area under the empirical ROC curve:  $52 / (14*6) = 0.62$