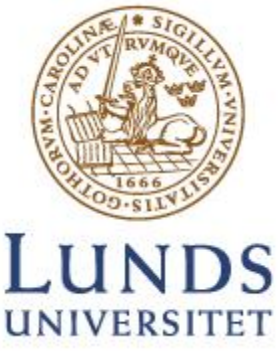

LUNDS UNIVERSITET

Master of Science dissertation:

Optimization of activity level in rCBF SPECT using the observer study Visual Grading Regression

Sofia Kvernby

Supervisors:
Anna Olsson, Agnetha Gustafsson and Michael Ljungberg

This thesis has been performed at the Department of Radiation Physics
Linköping University Hospital, Linköping, Sweden

Department of Medical Radiation Physics, Lund University, Lund, Sweden, 2012

List of abbreviations

3D	Three-dimensional
ANOVA	Analysis of Variance
CDRF	Collimator-Detector Response Function
CT	Computed Tomography
FWHM	Full Width at Half Maximum
HMPAO	Hexamethylpropylene amine oxime
IC	Image Criteria
ICS	Image Criteria Scoring
LEHR	Low Energy High Resolution
NMSE	Normalized Mean Square Error
OR	Odds Ratio
OS-EM	Ordered Subsets Expectation Maximization
PSF	Point Spread Function
ROC	Receiver Operating Characteristics
rCBF	regional Cerebral Blood Floow
SNR	Signal to Noise Ratio
SPECT	Single Photon Emission Computed Tomography
SSM	Strålsäkerhetsmyndigheten (Swedish Radiation Safety Authority)
TNF	True Negative Fraction
TPF	True Positive Fraction
VGA	Visual Grading Analysis
VGC	Visual Grading Characteristics
VGR	Visual Grading Regression

Abstract

The purpose of this work is to assess the activity level needed to achieve satisfactory diagnostic information in regional cerebral blood flow single photon emission computed tomography (rCBF SPECT) by introducing a new evaluation method to be used for and hopefully facilitate optimization studies in nuclear medicine in the future. The purpose is further to perform a visual grading study and investigate the use of this new evaluation method, Visual Grading Regression (VGR). Image quality criteria applicable to rCBF SPECT images will be defined in this work and their relevance for evaluating rCBF SPECT image quality is investigated.

The study comprises a material of 21 consecutive patients with dementia issue that have undergone an rCBF SPECT examination. An administered activity of 1000 MBq ^{99m}Tc labelled HMPAO was injected to all patients in the study. From one single examination, five studies corresponding to different activity levels (500, 625, 750, 875 and 1000 MBq) were generated by using a gated acquisition. Iterative image reconstruction, OS-EM, including corrections for attenuation, scatter and distance dependent resolution was used.

Three experienced observers, i.e. specialists in nuclear medicine, evaluated the images by rating their confidence about the fulfilment of specific image quality criteria. Seven criteria were defined in this study, developed in collaboration with experienced specialists in nuclear medicine with comprehensive knowledge on how to evaluate rCBF SPECT images. The result of the observers assessment were analysed using Visual Grading Regression, a method based on ordinal logistic regression with the aim to analyse data from visual grading experiments.

The result shows that there is a significant difference in perceived image quality between 500 MBq and the reference activity, 1000 MBq, in five of the seven image quality criteria. No statistical significant degradation was found between any other activity level than 500 MBq and the reference activity (1000 MBq). This study doesn't prove that any other activity level provides the same image quality as 1000 MBq, only because no difference was seen, but it gives an indication that the activity level could be reduced without losing too much diagnostic information.

The analysis method used, Visual Grading Regression, has proven to be convenient and easy to use for this kind of optimisation studies in nuclear medicine. The defined criteria cover the areas of the brain that are of interest in blood flow examinations and the results of this study showed that the observers used the whole confidence rating scale for each criterion, which is desirable. Some of the criteria had a very low proportion of rating scores corresponding to a fulfilment of the criterion, meaning that the satisfaction of the observers is low. A reversion or adjustment of these criteria might be needed to investigate whether the low satisfaction level is due to the formulation of the criteria or if only so the particular area is difficult to assess.

Author: Sofia Kvernby
Supervisors: Anna Olsson, Agnetha Gustafsson and Michael Ljungberg
Title: Optimization of activity level in rCBF SPECT using the observer study Visual Grading Regression

Master of Science Thesis in Medical Radiation Physics, 30 credits
Department of Medical Radiation Physics, Lund University

Kan man få lika bra bildkvalitet till lägre stråldos?

Möjligheten att visualisera olika sjukdomstillstånd med hjälp av bilder har inom sjukvården blivit mer och mer populärt. Nuklearmedicinska bilder framställs genom att en liten mängd av ett radioaktivt ämne injiceras i patienten och sedan detekteras med hjälp av en gammakamera. Genom att utnyttja olika upptagsmekanismer kan funktionen av ett område eller ett specifikt organ i kroppen undersökas. Blodflödet i hjärnan kan studeras genom att injicera ett radioaktivt ämne som är kopplat till en fettmolekyl, vilken passerar blod-hjärnbarriären och ackumuleras i hjärnan. Upptaget av det radioaktiva ämnet avspeglar det regionala blodflödet i hjärnan vid tidpunkten för injektionen eller strax därefter. Regionala blodflödesstudier i hjärnan utförs på patienter som stöd vid diagnostisering av demenssjukdomar då flertalet av dessa leder till en minskning av det regionala blodflödet i hjärnan.

Aktivitetsmängden som injiceras påverkar både bildkvaliteten och stråldosen till patienten. För att säkerställa den aktivitet som krävs för att ge tillräckligt god bildkvalitet till lägsta möjliga stråldos krävs ett optimeringsarbete. Detta görs vanligtvis med hjälp av fysikaliska parametrar såsom brus, upplösning och kontrast. Ett annat sätt att bedöma bildkvalitet är med hjälp av erfarna observatörer som gör en visuell bedömning utifrån kliniskt relevanta bildkriterier.

I arbetet introduceras en ny utvärderingsmetod, Visual Grading Regression (VGR), som inte tidigare har använts vid optimeringsstudier inom nuklearmedicin. Syftet med arbetet är att optimera aktivitetsnivån för regionala blodfödesundersökningar med gammakamera med avseende på bildkvalitet och stråldos till patient.

I studien ingick 21 konsekutivt utvalda patienter med demensfrågeställning som alla genomgått en blodflödesundersökning av hjärnan med gammakameratomografi. En aktivitetsmängd på 1000 MBq injicerades till samtliga patienter. Genom att dela upp insamlingstiden i delmängder är det möjligt att generera flera bilder, vilka motsvarar olika mängder administrerad aktivitet, från en och samma undersökning. Från en studie genererades fem bilder, vilka motsvarade aktivitetsnivåer av 500, 625, 750, 875 och 1000 MBq.

Kliniskt relevanta bildkriterier definierades i samarbete med en erfaren radiolog. Bilderna, motsvarande de olika aktivitetsnivåerna, bedömdes med en fyrgradig skala utifrån uppfyllandet av de definierade kriterierna. Tre erfarna radiologer med specialistkompetens inom nuklearmedicin deltog som observatörer i studien och granskade bilderna en och en i en slumpmässigt vald ordning. Resultatet av bedömningen analyserades med hjälp av den valda utvärderingsmetoden, VGR, vilken är speciellt utformad för visuella granskningsstudier och har fördelen att kunna ta hänsyn till variationer mellan patienter och mellan observatörer.

Resultatet visar att en skillnad i bildkvalitet kan ses mellan bilder som motsvarar en injicerad aktivitet på 500 MBq och 1000 MBq i fem av de sju kriterierna. Ingen skillnad kan ses mellan någon av de andra aktivitetsnivåerna och referensaktiviteten, en administrerad aktivitet på 1000 MBq. Studien visar inte att någon av de andra aktivitetsnivåerna ger samma bildkvalitet som 1000 MBq, endast på grund av att ingen skillnad sågs, men det ger en indikation på att aktivitetsnivån kan sänkas utan att förlora alltför mycket diagnostisk information.

Table of content

1. Introduction

Image quality in nuclear medicine is affected by several parameters, for example the characteristics of the scintillation camera, choice of collimator, radionuclide, amount of administered activity and reconstruction algorithm. The size of the patient i.e. the amount of scattered material will largely affect the image quality. To achieve an optimized examination, each and every of these influencing parameters has to be optimized. Optimization of image quality in nuclear medicine is hence an extensive and time consuming work.

The Swedish Radiation Safety Authority (SSM) requires that ionizing radiation are used in an optimized way with respect to radiation dose and medical purpose[4]. Dose optimisation means the process of adapting the radiation dose, i.e. the injected amount of activity, to a patient in a diagnostic examination in such a way that the radiation dose is as small as reasonably achievable, still ensuring that the needed diagnostic information is obtained. The work to optimize the procedures and protocols used in a diagnostic examination is thus both important and essential. The activity level needed to give the definite diagnostic information in cerebral blood flow single photon emission computed tomography (rCBF SPECT) is complex and the available guidelines give different recommendations [5-6].

Contrast, resolution, noise and normalized mean square errors (NMSE) are typically used as image parameters to determine the image quality and are often used in image quality optimization studies. These physical criteria are easily measured in phantoms, unfortunately in clinical relevant images they are more difficult to assess. Phantoms can to this purpose be used in the same manner as for daily, weekly and monthly quality controls of the systems performance, thus according to the frequently used standards of the National Electric Manufacturers Association (NEMA) report [7].

An image with physical optimal image quality, as high contrast and resolution together with low noise level, does not necessarily mean an image with satisfactory diagnostic image quality, i.e. an image quality where the observer can feel confident in the diagnostic decision. How observers interpret the diagnostic information in an image and how the image quality is perceived has a large impact on the decision whether satisfactory diagnostic image quality is achieved or not. The need for a more clinically appropriate method to evaluate image quality has made visual evaluation of image quality more popular. Performing an optimization study in the same manner as diagnostic decision is made, which often is based on visual impression, is thus desirable.

Before clinically implementing the results of an optimization study, based on physical parameters or Monte Carlo simulations, some kind of clinical assessment is usually made where anatomic structures and background activity are included. This clinical assessment can be made by using receiver operating characteristics (ROC), which thus is a commonly used method in clinical studies. A drawback with ROC is that the true diagnosis of the patient has to be known, i.e. if the patient has a disease or not. That can be done by biopsies or long term follow up, hence a time-consuming part of the study.

Another way of performing clinical assessment is to use visual grading experiments. In a visual grading study it's possible to directly use clinical image quality criteria, implying criteria applicable to clinical relevant images containing substructures and anatomic noise. This type of criteria shall correspond to the clinical reality that observers perceive in a nuclear medicine image. Visual grading involves human observers who rate their confidence about the fulfillment of clinical relevant image criteria.

The validity of visual grading studies can be assumed to be high since the image quality criteria are based on clinically relevant structures which reflect a clinical situation where the anatomical background is included. An advantage of using visual grading studies, especially in comparison with ROC studies, is the easiness of conducting such a study. This is an important aspect when optimizing equipment and examination protocols

at the local level on a small nuclear medicine department. The time consumption required of the observers is often moderate and thus an additional reason why to prefer a visual grading study in a small hospital.

Comparative studies have been made [8-11] between methods based on visual grading or receiver operating characteristics and method based on calculations of physical parameters, which in special cases have shown good agreement. This validates the underlying assumption and the basic idea in visual grading, meaning that the possibility to detect pathology correlates to the reproduction of anatomy.

Aim of the project

The aim of this thesis was to assess the activity level needed to achieve satisfactory diagnostic information in regional cerebral blood flow single photon emission computed tomography (rCBF SPECT) by introducing a new evaluation method to be used for optimization studies in nuclear medicine. From one single patient rCBF SPECT examination with an injected activity of 1000 MBq, five studies with different activity levels (500 MBq, 625 MBq, 750 MBq, 875 MBq and 1000 MBq) were generated by using a gated acquisition.

The aim was further to perform a visual grading study and investigate this new evaluation method, Visual Grading Regresion (VGR). VGR is based on ordinal logistic regression and applied to data from image studies with visual grading scores selected on an ordinal scale, i.e. a non-parametric rank-invariant statistical test is required. Visual Grading Regression is thus used to investigate whether different amount of administered activity changes the assessed image quality in rCBF SPECT using ^{99m}Tc labelled HMPAO. Image quality criteria, which are clinically based, were defined and used in the Visual Grading study. Experienced observers rated their confidence about the fulfillment of the criteria and a Visual Grading Regression analysis was made of the study. The aim of this thesis was also to evaluate whether these criteria are suitable for use in nuclear medicine to evaluate rCBF SPECT image quality. Hopefully this method is supposed to be useful in, and facilitate, optimization studies in nuclear medicine in the future.

2. Background

2.1 Scintillation camera

In nuclear medicine imaging, the scintillation camera is used to detect photons emitted from radiopharmaceuticals. The scintillation camera visualizes the distribution of a radioactive isotope in a patient's body. Its construction and principle were published by Hal O Anger in 1957 [12]. It consists of a collimator, a scintillation crystal, light guide and several photomultiplier tubes within a lead shielding (Figure 1).

The main assignment for the scintillation crystal is to absorb the incoming photons and effectively reemit the deposited energy in form of scintillation photons, usually in the range of visible light. The phenomenom is called radio luminescence and arises in some organic molecules or inorganic doped crystals. The produced scintillation quanta are proportional to the deposited energy. The desired properties of the scintillation crystal are several. High density implies a high cross section for photoelectric absorption and a possibility to reduce the crystal size without losing detection efficiency. Short decay time and fast response time are also desirable. Usually the crystal consists of sodium iodide with thallium doping.

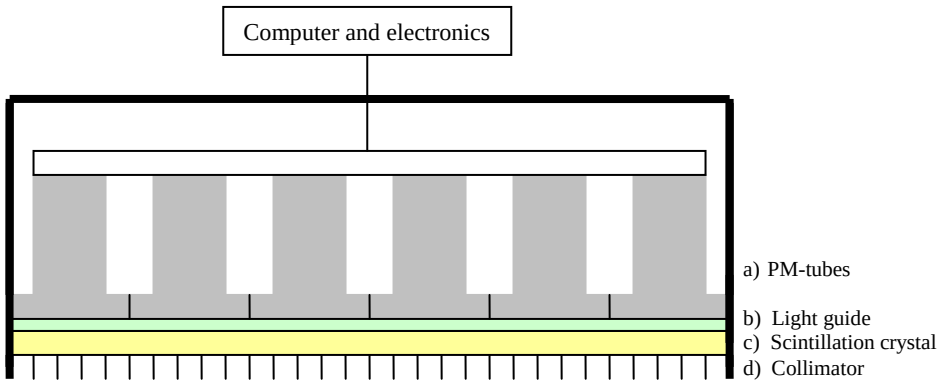

Figure 1: Construction of the scintillation camera

When the produced scintillation photons enter the photomultiplier tube, they strike the photocathode material and free electrons are produced. These electrons are directed by a focusing electrode towards a chain of dynodes. Between each dynode they are accelerated by the electric field and when an electron strikes the first dynode, a bunch of low energy electrons are emitted and accelerated towards the second dynode. This accumulation of electrons implies a sharp pulse when the cascade reaches the anode and indicates the detection of a photon. Several photomultiplier tubes are used and constitute a hexagonal detector array. The use of many photomultiplier tubes enables to determine where in the scintillation crystal the interaction took place. By using the relative signal information from all photomultiplier tubes, it is possible to calculate the centroid position of the signals and thus determine both the position and the energy of the incoming photon. This kind of processing is often referred to as Anger Logic.

A collimator is placed at the most surface part of the scintillation camera. The collimator controls the spatial resolution and the sensitivity of the camera. It performs a selection of the incoming photons with respect to direction. Depending on the diameter and length of the holes together with the thickness of the walls, different acceptance angles can be obtained. The photons will be absorbed in the collimator walls if the incident angle is larger than the acceptance angle. Large collimator holes will, at the expense of spatial resolution, increase the sensitivity by transmitting many photons and in high resolution collimators, with small diameter or long length of the collimator holes, the sensitivity will decrease with an increasing resolution. Lower sensitivity means fewer transmitted photons and thereby a higher noise level.

2.2 Single Photon Emission Computed Tomography (SPECT)

2.2.1 Reconstruction

A two-dimensional planar image which represents the activity distribution in a patient can be obtained with the scintillation camera. A planar image means information from a three dimensional object have been superposed to two dimensions. That implies a loss of information in one direction, in depth, because of overlaying structures and thus difficulties to distinguish uptake from different organs. The need to obtain a three-dimensional visualization of the activity distribution is satisfied by Single Photon Emission Computed Tomography (SPECT). SPECT imaging has several advantages over Planar Imaging, for example better contrast sensitivity and estimation of location. The ability to make quantification more accurate, by using volume instead of area, is important in dose calculations.

Tomographic reconstruction is applied in Single Photon Emission Computed Tomography to obtain cross-sectional images of patients. The photons emitted from the patient are collected in several projections, which are done by rotating the camera around the patient. Each projection of the object at a given angle consists of a set of line integrals through the object. This is the Radon transform which is fundamental in image reconstruction. It represents the line integral of an unknown density distribution:

$$P(\theta, t) = \int_{-\infty}^{\infty} f(x, y) d\theta dt \quad (1)$$

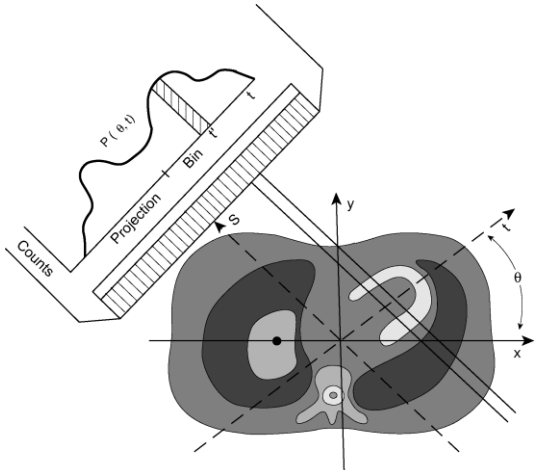

Figure 2: Principle of data acquisition in SPECT[1]

The information from the Radon transform is used to reconstruct cross sectional images of the unknown density distribution f(x,y). The Radon transform has to be inverted, which can be made in different ways, analytically or iteratively [13].

The aim of analytical methods, such as Filtered Back Projection, is to obtain an inversion formula for the continuous Radon transform. The measurement takes place only in a finite number of angels and in a finite number of data points, i.e. pixels, the model must therefore be discretised. The first step in the reconstruction process is to create a sinogram, which serves as a map where the collected image data is compiled in a structured way. A sinogram is created by picking out a certain row of the image matrix collected in the first projection angle and place it on the first row in the sinogram matrix. Then the same row is picked out of the image matrix collected in the second projection and is placed on the second row in the sinogram, and so on.

The resulting sinogram is a matrix where the number of projection angles determines the number of rows in the sinogram matrix. The sinogram contains all information needed to reconstruct one cross section of the object.

The next step, to create a tomographic image, is to backproject each row of the sinogram, which means that each row is smeared over a new matrix. For every row in the sinogram matrix that is backprojected, the matrix over which the values are smeared out is rotated with the same angle as the angular difference between two projection angles. The data from the sinogram is then superimposed in a way that corresponds to a cross sectional image slice of the object. A problem with backprojection is that the actual activity distribution is blurred out into the surrounding region. In filtered back projection, the collected data is filtered before it is back projected. A ramp filter, which is a high pass filter, is used to reduce the blurring of the image. Reducing the low frequencies in the image involves not only an enhancement of the details but also an increase in noise level. Sometimes the ramp filter is used in combination with a low pass filter if the high noise level is bothersome.

Tomographic reconstruction can also be made by using an iterative reconstruction method. There are several types of iterative reconstruction methods available, for example ML-EM (Maximum Likelihood Expectation Maximization) and OS-EM (Ordered Subsets Expectation Maximization). A fundamental difference between iterative and analytic image reconstruction methods is that the image is estimated in several steps through optimization in the iterative method, unlike the analytical method in which the calculation of the image matrix is made in a single step based on an analytical formula. A difference between iterative and analytic reconstruction methods can also be found in the initial image matrix, the one that the initial sinogram is created from. In an analytic reconstruction method, the initial image matrix is created from the collected data belonging to different projections. In an iterative reconstruction method is the sinogram initially created from a guessed image matrix. The pixels in the guessed image are often initially assigned a constant positive non-zero value. The data from the guessed image is forward projected and then in the same way as in filtered back projection a sinogram can be created.

Next step is to compare the measured, i.e. the real, sinogram with the sinogram created from the guessed image. Taking the ratio for each projection between the measured and calculated projection usually makes that, as in ML-EM. The resultant error projection is then back projected and multiplied with the initial guessed image. The product of these two results in a new guessed image and one iteration is performed. This process is repeated until the error projection is as small as possible and good image quality is achieved. The principle of iterative reconstruction is described in figure 3.

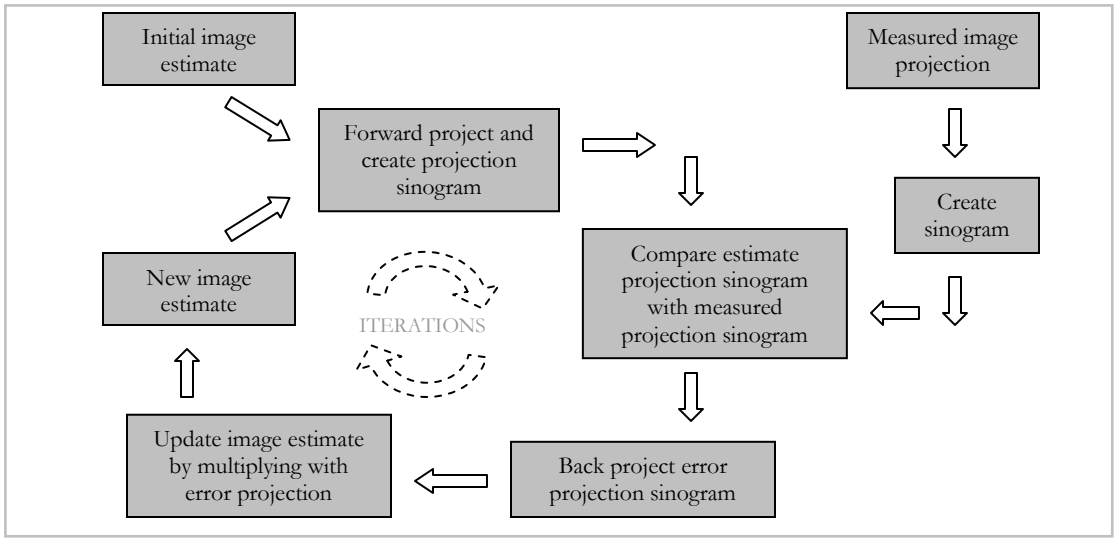

Figure 3: Schematic principle of iterative reconstruction

This method is quite time consuming and can be speeded up. Instead of using data from all projections to estimate the error, only a subset of the data can be used and thus improve the speed of the reconstruction. Only a subset of the projections are used in OS-EM to produce an error image and thus for updating the guessed image. All projection angles are still used in each iteration but the difference is that the estimated error is calculated from only a subset of the projections and several times during each iteration. There is a risk of reducing the accuracy in the error image if using too few angles in each subset. Fewer angels in each subset imply less time needed to reconstruct but also less accuracy, meaning that a balance must be struck.

A general problem with the iterative methods is the increase of noise level with increasing number of iterations. There are different ways of handling this problem. One way is to perform either a post- or pre-filtration with a low pass filter. Another way is to stop the reconstruction early, i.e. reducing the number of iterations. Stopping the reconstruction early gives the same effect on the image noise as low pass filtering because of the low frequencies converge faster. Maximum A Posteriori (MAP) algorithms can also be used to handle this problem, which implies that restrictions about the created error ratio are added to the reconstruction process. Such restrictions penalize large local variations in pixel values in a known homogeneous area, which dampens the noise.

The iterative reconstruction methods have several advantages over the analytical methods. An iterative method can include corrections in the model for example for non homogeneous attenuation, scatter, collimator blurring and septum penetration of high energy photons. Some of these corrections will be shortly described in the next sections.

2.2.2 Attenuation effect and correction

Attenuation occurs when photons interact with matter. Attenuation is the collective name for photo electric effect, Compton scatter, coherent scatter and pair production. The dominant processes of attenuation in nuclear medicine are photo electric effect and Compton scatter, due to interaction in tissue at photon energies in the interval ~70 – 360 keV. Attenuation of photons means that the number of primary photons, i.e. detected photons which have not interacted, decreases along their path through tissue or any sort of matter. In nuclear medicine imaging, attenuation will cause some problems and limits the ability to accurate determines the distribution of the radiopharmaceutical in the object. The tissue densities and the amount of tissue between the uptake volume and the detector vary between the different projection angles. This issue may create areas with false lack of uptake, which may mean that the observer detects false pathologies. Attenuation can be described by an exponential function:

$$N(x) = N_0 \cdot e^{-\frac{\mu}{\rho}\rho \cdot x} \quad (2)$$

Where N(x) is the number of photons counted after a passage through a material of thickness x and density ρ. The mass attenuation coefficient $\frac{\mu}{\rho}$ is dependent on material composition and photon energy. The number of emitted photons is notated as N_0.

Attenuation correction can be made in different ways; common for them all is that they are based in finding out values for the actual mass attenuation coefficient by using transmission data. Today it is common to use hybrid systems where the SPECT camera is combined with a CT (computer tomography)[14]. By incorporating the techniques of CT in the image acquisition and reconstruction process it becomes possible to make a careful attenuation correction by using the detailed 3D density map that CT provides. The attenuation map is estimated by segmentation of anatomical structures and smoothing of voxel values to a size which is adapted for SPECT image voxel size, together with transformation of Hounsfield units to linear attenuation coefficients, μ, for the relevant photon energy. High quality and high spatial resolution images of cross sectional anatomy can also be provided within these hybrid systems, which gives the observer good anatomic

landmarks that can be correlated to the SPECT image. Attenuation correction using CT data can be incorporated in an iterative reconstruction method.

2.2.3 Scatter effect and correction

The primary attenuation mechanism in tissue throughout the diagnostic energy range is Compton scatter, which corresponds to a loss of energy and a change of direction of the photons. Scattered photons means less primary photons and thus reduced contrast together with higher noise level. The change of direction of a photon may result in a missed count because of the change in trajectory may cause that the collimator prevent it from reach the detector. If the scattered photon pass the collimator and is detected, it will lead to a false background and a degradation of the image quality. One way to overcome or reduce the problem with scatter is to use a small energy window, so that only primary photons are detected. This method is not possible when using a NaI-detector because of its poor energy resolution. To avoid a high noise level, a wide energy window is thus needed to obtain an acceptable counting statistics.

Scatter is a more significant problem at low photon energies because the percentage of scatter in the energy window is decreasing with increasing photon energy. For all scattering angles θ, when the energy of the incident photon ($h\nu$) increase, the energy of the scattered photon ($h\nu'$) is moving towards a threshold. That means the electron, which the incident photon collide with, will carry away a larger part of the incident photon energy. The energy gap between the incident and scattered photon will thus be larger for higher photon energies and the amount of scattered photons in the energy window will be smaller. The photon energy before and after the collision with the electron is described as:

$$h\nu' = h\nu \cdot \frac{1}{1+\frac{h\nu}{m_0c^2}(1-cos\theta)} \tag{3}$$

where m_0 is the rest mass of the electron and c is the speed of light.

Different scatter correction methods are available and is often included in the reconstruction algorithm. The basic idea is to estimate the scatter in the data and remove it. The scatter can be estimated by using one or two additional energy widows, which are located next to the full absorption peak, and from them estimate the amount of scattered photons in the primary energy window. A drawback with removing photons from the data is that the noise increase. When using window-based scatter correction methods in an iterative reconstruction is the scatter component added to the guessed image in the forward projection step[15]. Because of adding photons instead of removing, it does not provide the same problem with noise.

2.2.4 Distance dependent resolution correction

The spatial resolution in SPECT, when imaging in air with the absence of scatter and septal penetration, is mainly affected by two components. The intrinsic resolution of the detector and electronics, which often can be modeled as a Gaussian function, together with the geometrical acceptance of photons through the holes in the collimator, which also typically is approximated to a Gaussian function for a parallel hole collimator. The total spatial resolution of a scintillation camera is often described by a point spread function (PSF) which thus for a parallel hole collimator has the approximate shape of a Gaussian function. The PSF describes the response of an imaging system to a point source of activity, placed at a certain distance from the detector. The full width at half maximum (FWHM) of the PSF is often used as a measure of the system spatial resolution.

The intrinsic resolution is due to uncertainty in position estimation in the detector system. It depends on the crystal thickness, the photomultiplier tubes and the accuracy of the electronic components to correctly position the detected photons.

The geometric component, i.e. the geometric response function, is determined by the design of the collimator. The diameter and size of the collimator holes and the septal thickness are parameters that affect the geometric response function. The spatial resolution is also very much a function of the distance between the source and the collimator surface. Therefore, it is important to positioning the camera as close to the object as possible. The standard deviation σ_c of the PSF is a linear function of the distance between the object and the face of the collimator:

$$\sigma_c(d) = \sigma_0 + \sigma_d \cdot d \tag{4}$$

Where d is the distance from the face of the collimator to the object, σ_0 is the standard deviation at the face of the collimator and σ_d is the change in standard deviation with distance. The effect of the geometric component to the spatial resolution is illustrated in figure 4.

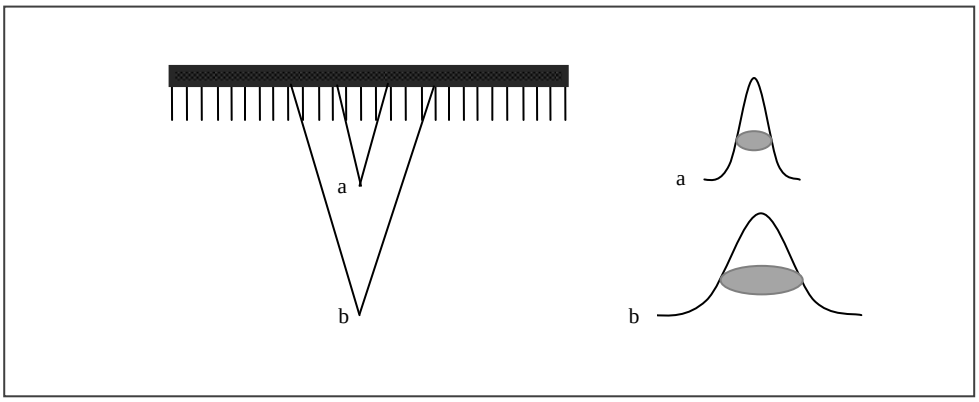

Figure 4: Illustration of the geometric component to the spatial resolution. The appearance of the PSF, i.e. the blurring of the image, at different distances (a and b) from the face of the collimator.

The image can be seen as the convolution of the object with the point spread function. A convolution in the spatial domain corresponds to a multiplication in the frequency domain. By dividing the image with a model of the degradation, as the PSF, in the frequency domain, it is possible to perform a restoration filtering. The goal of a restoration filtering is to recovering the original image of the object. The filter compensate for the characteristics of the system that allows the degradation of the resolution. This image restoration method also affects the image noise, unfortunately in a manner that is not appreciated. The noise contribution, i.e. Poisson noise which can be described by the noise power spectrum, is constant due to the total number of counts in the image, while the object power spectrum decreases rapidly with the frequency. Application of an inverse filter will result in a significant amplification of noise at high frequencies. One way to overcome this problem is to use the inverse filter at low frequencies, where the contributions from the signal dominates the image, and at the higher frequencies, where the noise dominates, switch to a low-pass filter used to reduce the noise. An example of such combination filter is the Wiener filter.

Another method for distance dependent resolution correction is to incorporate the model of blurring degradation, i.e. an accurate model of the collimator-detector response function (CDRF), into the iterative reconstruction process, resulting in improved spatial resolution. This type of correction is based on modeling the spatial resolution with the Point Spread Function, whose standard deviation increases with the distance from the detector. The forward projected plane in the iterative reconstruction process, i.e. the projection of the updated guessed matrix, is compensated for collimator-detector response by convolving each plane, parallel to the camera surface, with corresponding point spread function (figure 5). This procedure is intended to increase the similarities between the measured sinogram and the guessed sinogram and thus reducing the resulting error projection. Correction for collimator in an iterative process is acting a bit like a low-pass filter.

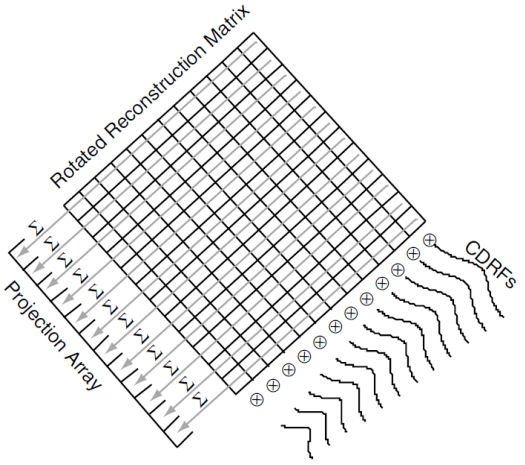

Figure 5: Illustration of incorporating the collimator-detector response function (CDRF) in iterative reconstruction by convolving each plane, parallel to the camera surface, in the matrix with corresponding PSF[3].

2.2.5 Image noise

Statistic noise is a problem in SPECT, which degrading the image by reducing the contrast. The image contrast is often measured as the signal to noise ratio and thus a high noise level implies reduced contrast. Statistical noise arises because of the limited number of counts. The effectiveness of the collimator and the amount of activity injected together with the acquisition time are parameters affecting the number of counts. To improve any of these parameters means a compromise with another parameter affecting the image quality. Higher effectiveness of the collimator implies reduced spatial resolution, higher amount of injected radioactivity means increased absorbed dose to the patient and thus increased risk of cancer and longer acquisition time involves more artifacts due to movement. An optimization process of these parameters is needed to reduce the statistical noise.

The radioactive decay and thus the emission of photons together with the interaction process in the detector and thus the generation of scintillation quanta are stochastic events. These stochastic processes implies an uncertainty in the measurement, meaning that two identical measurements do not result in two identical results. The differences between the two measurements can be referred to as image noise.

The noise is often approximated to a Poisson distribution, which describes the probability of occurrence of rare events and hence the probability of detecting *n* photons can be described according to:

$$P(n) = \frac{e^{-m}m^n}{n!} \tag{5}$$

where *m* is the mean number of events, i.e. mean number of detected photons for large number of repeated measurements, and *n!* is factorial *n*. For the Poisson distribution both the mean and the variance are equal to *m*. The standard deviation, which often is the parameter describing the image noise, thus equals $\sqrt{m}$. The signal to noise ratio (SNR) is thus:

$$SNR = \frac{m}{\sqrt{m}} = \sqrt{m} \tag{6}$$

2.3 Regional cerebral blood flow

The blood flow in the brain is closely related to the regional brain metabolism. Radiopharmaceutical injected in the blood enables functional brain imaging with SPECT and the possibility to assess regionally brain metabolism and diagnose pathological changes. Regional cerebral blood flow (rCBF) SPECT examination has several applications, for example in diagnosing Alzheimer's disease and investigation of epilepsy.

The available guidelines give different recommendations about the amount of injected activity in rCBF SPECT. The Swedish Radiation Safety Authority provides diagnostic reference level for different diagnostic examinations in nuclear medicine. The diagnostic reference level for rCBF SPECT using ^{99m}Tc labelled HMPAO is 1000 MBq [5], which provide an effective dose of 9,3 mSv [16]. The European Association of Nuclear Medicine gives procedure guidelines for brain perfusion SPECT [6]. The recommended amount of injected activity in this guideline is 555-1110 MBq (typically 740 MBq), which provide an effective dose between 5,7 – 10,3 mSv (typically 6,9 mSv) [16].

2.3.1 Radioactive isotope and production

The most commonly used radiopharmaceutical in rCBF SPECT is ^{99m}Tc labelled hexamethylpropylene amine oxime (HMPAO, Ceretec®, General Electric Healthcare). Technetium is the most widely used radioisotope in diagnostic nuclear imaging. The reason is that it fulfills many of the properties desired of a radiopharmaceutical used in diagnostics. It has a half life of six hours, which is long enough to perform a medical examination and short enough to let the patient leave the hospital soon afterwards. The emitted type of radiation, i.e. high yield of photons and low yield of emitted electrons, are beneficial due to radiation dose. The energy of the emitted photons are 140.5 keV, which are high enough to penetrate tissue of a thickness corresponding to a human, but low enough to be absorbed in a thin scintillation crystal.

Technetium-99m is the decay product from Molybdenum-99, which has a half life of 66 h and is hence allowed to be transported over fairly long distances. Molybdenum-99 is in most ^{99m}Tc generators in form of molybdate MoO_4^- which is adsorbed onto alumina acid. When molybdenum decay in this form and in this environment, pertechnetate TcO_4^- is formed. Pertechnetate is less tightly bound to the aluminium and it is hence easy to elute soluble ^{99m}Tc pertechnetate by rinse the generator column with saline solution.

2.3.2 Transport and behaviour in human brain

To obtain an rCBF-SPECT image there is a need for using a radiopharmaceutical that reflects the rCBF. In order to represent the rCBF, the radiopharmaceutical needs to have the ability to pass the Blood Brain Barrier and thus be lipophilic. The lipophilic compound of radiopharmaceutical is following intravenous administration and is rapidly bound to protein. The radiopharmaceutical is transported through the blood vessels and reaches the brain. The extraction fraction, which describes the amount extracted from the blood vessels to the brain, is about 80% for HMPAO and a large proportion is due to first pass extraction, which means the extracted amount during the first passage through the brain [17]. The method of rCBF SPECT is based on the fact that the distribution of the tracers is proportional to the rCBF. The activity ratio of grey to white matter is approximately 2-3 to 1 for HMPAO which enables contrast and hence tissue separation in the brain.

The patient is kept in a calm environment with low light, reduced noise and minimal disturbance before injection to achieve standardized condition. These conditions should be maintained for about 10 minutes before and after the injection. The HMPAO tracer reaches the brain after approximately 15 seconds and the major uptake in the brain takes place during the following 40 seconds. The image acquired is therefore representing the cerebral perfusion during these first minutes after injection, which often is referred to as frozen image.

2.3.3 Evaluation and assessment of rCBF SPECT

The issue of rCBF SPECT is whether there is a functional disruption of brain activity that is consistent with a specific disease. Signal loss is assessed in particular by looking at symmetries and differences between the hemispheres together with differences between grey and white matter. It is primarily the relative differences in uptake that is of interest in an rCBF SPECT examination. Functional areas of the brain that are more robust regardless of pathology, such as thalamus and cerebellum, are used for uptake comparison. Relative differences in uptake may sometimes be quantitative and thus compared to a normal material by using standardized anatomical areas, for example Brodmann areas.

The spatial resolution in an SPECT image is relatively poor, which implies that functional changes in fairly large regions are often required to detect a pathological blood flow reduction.

3. Theory of visual image quality evaluation

3.1 Generally about visual image evaluation

There are many ways to assess image quality[18-19]; one very popular way is to involve human observers which for example are made in a Receiver Operating Characteristics study and in a Visual Grading experiment.

Receiver Operating Characteristics (ROC) analysis is a well-established method for image quality evaluation where, in a typical manner, the observer state whether an image belongs to a healthy patient or weather the patient has a disease. ROC analysis is a binary paradigm meaning that the patient either does have a disease or does not have a disease and the observers' task is to state weather the patient has a disease or not. The method is used to study the human observer's performance or the difference between two imaging modalities.

The observer has to rate his confidence about the decision on typically a four-point or six-point scale. The results of the observers' assessment are compared with the actual true diagnoses. Sensitivity or the True Positive fraction (TPF) is the probability that a patient with an actual disease is determined as having a disease by the observer and specificity or the True Negative Fraction (TNF) is the probability that a healthy patient is determined as being healthy by the observer. By deliberately varying the decision threshold of abnormality, it is possible to establish the balance or trade-off between TPF and TNF. A ROC curve is usually generated with the TPF given as a function of the FPF i.e. sensitivity as a function of 1 – specificity. An illustration of different decision threshold and the corresponding ROC-curve can be seen in figure 6 and 7.

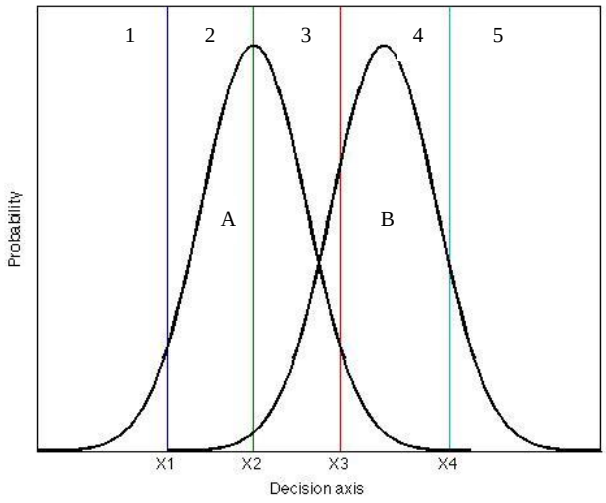

Figure 6: Probability distributions of noise/no disease (A) and signal/disease (B) of a detection task with four decision thresholds (X_1-X_4). A rating of confidence level of the observer that belongs to the first rating category (1) corresponds to X-values < X_1 and the second rating category (2) corresponds to X-values X_1 <X< X_2. The diagnostic decision is thus dependent of decision threshold and the probability distributions. Due to the overlap of these distributions is the diagnostic decision not obvious, at descision threshold X_1 is the sensitivity high but the specificity low and at decision threshold X_4 is the sensitivity low but the specificity high.

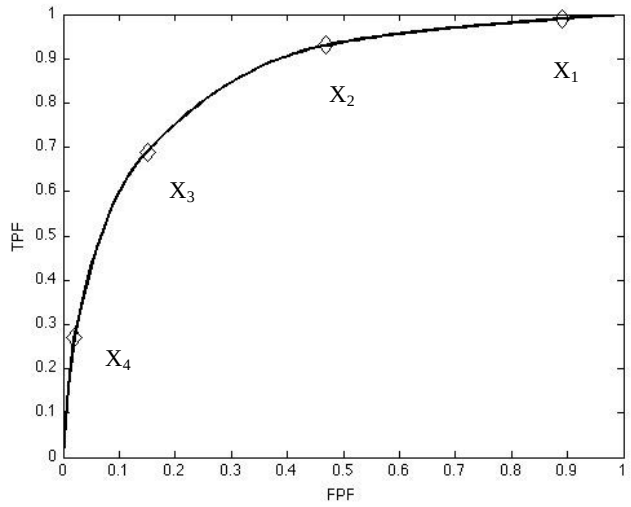

Figure 7: Illustration of ROC-curve where the diamonds correspond to the different decision thresholds in figure 8. The ROC-curve shows the true positive fraction as a function of the false positive function.

The area under the ROC curve is often used as evaluation tool of the study. An area under the ROC curve with the value of 1.0 corresponds to perfect detection while a detection governed by chance is represented with a value of 0.5. The steepness of the curve correlates to how confident the observer is in his answer. A steep curve implies a system where the observers have high confidence about their decision.

A major drawback in Receiver Operating Characteristics is that the true diagnosis of the patients has to be known, i.e. if the patient has a disease or not. That is often the most time-consuming part of the study, which can be done by biopsies or long term follow up. Large number of cases is often required in order to produce statistically significant results. Receiver Operating Characteristics can therefore be difficult to conduct in a small nuclear medicine department.

In an ROC study the focus is on the observers' ability to detect and correctly interpret the visible signals of pathology in an image. In a visual grading experiment the focus is on how well normal anatomy is reproduced. The basic idea behind visual grading is that the visibility of a certain structure or the reproduction of anatomy corresponds to the possibility to detect pathology. A visual grading study is relatively easy to conduct compared to a ROC study and not as time consuming. The workload for the observers are usually fairly low, thus a study is easy to justify in an economical point of view. Comparison of results from a ROC study and a visual grading experiment study has been made and shows good agreement [8].

In a visual grading experiment, a number of observers will rate their confidence about the fulfillment of a specific image criterion. Image quality criteria defined by the European Union [20] are available and are often used in visual grading studies. These criteria include such as the visibility of a certain anatomic structure and are intended to be used in radiographic imaging. There are no established image quality criteria applicable and designed for Visual grading studies in nuclear medicine imaging available.

A visual grading study can be performed by letting the observers view either one image at the time or simultaneous viewing of two images. The method which involves simultaneous viewing of two images is intended to increase the sensitivity to small differences in image quality, where the grading scores are based on comparison between the two images.

The scale steps used in a visual grading experiment are defined on an ordinal scale, meaning that they have a natural ordering. The scale steps are often labeled with numeric values but they still belong to an ordinal scale. That means that there is no guarantee that the difference between 1 and 2 are equal to the difference between 2 and 3. Analyzing the data from a visual grading study, i.e. data defined on an ordinal scale, may introduce some statistical problems. Data derived from an ordinal scale can't be assumed to be normally distributed. The analysis can thus not be made with methods which seek to minimize the sum of squares distances between predicted and observed values such as t-test or ANOVA, which requires that the dependent variable is defined on an interval scale. This problem has resulted in development of some new methods. Visual Grading Characteristics and Visual Grading Regression handle the problem with ordinal scale in a statistically correct way. These two analysis methods will be described in the next section.

A variant of visual grading experiments is Image Criteria Scoring which is a simple visual grading method. It is performed by letting the observers' state whether an image quality criterion are fulfilled or not. The observers are not supposed to rate their confidence about the decision, the assessment of the visibility of a certain structure lies in the criteria and the task of the observers is straightforward scoring fulfilled image criteria. The separating power of Image Criteria Scoring is hence weaker than that of Visual Grading Analysis, which is a method involving several confidence scale step but that uses an insufficient statistical analysis method. The criteria defined in the European Guidelines on Quality Criteria for Diagnostic Radiographic are those deemed necessary to produce an image of standard quality, according to the European Commission. A commonly used evaluation method in Image Criteria Scoring is to calculate the proportion of fulfilled criteria known as image criteria score (ICS):

$$ICS = \frac{Number\ of\ criteria\ fulfilled}{Total\ number\ of\ criteria} \tag{7}$$

$(0 \leq ICS \leq 1)$

Image Criteria Scoring has the advantage that a parametric statistical method can be used in the analysis. Parametric statistics assumes that the underlying data is normally distributed, which the Image Criteria Score is since it is the mean of a variable that can either take the value of zero or unity.
When the visibility of a specific structure is close to the decision threshold for the observer, the observer may have difficulties in decision whether an image criteria is fulfilled or not. This uncertainty of the observer in how far away from the decision threshold the visibility is, are not taken into account when analyzing the results.

3.2 Visual Grading Characteristics (VGC)

Visual Grading Characteristics (VGC) is an analysis method used to compare and evaluate differences in image quality between two modalities. The method has been developed by Båth and Månsson [21]. Its principle is based on elements from Image Criteria Scoring or Visual Grading Analysis (VGA) and used together with a mathematic formalism similar to that of Receiver Operating Characteristics (ROC).

The observer participating in a VGC study will be presented a series of images from each modality. He will state his confidence about the fulfillment of specific criterions in each image. A criterion may be for example "Visually sharp reproduction of the bones" [20] and the rating scale may be for example "Confident that the criterion is not fulfilled", "Somewhat confident that the criterion is not fulfilled", "Indecisive whether the criterion is fulfilled or not", "Somewhat confident that the criterion is fulfilled" and "Confident that the criterion is fulfilled".

The number of images considered belonging to the different confidence levels or rating categories is compiled in a table for each modality. By deliberately varying the decision threshold for which confidence level the criterion is considered to be fulfilled, Image Criteria Scores (ICS) can be obtained for the two modalities. As already described, ICS is the proportion of fulfilled criteria and are in this method used to create a VGC curve. A VGC curve can be generated by plotting the Image Criteria Score for the first modality against the Image Criteria Score for the second modality. The points on the VGC curve corresponds to pairs of the proportion of fulfilled criteria for the two modalities. The first point on the curve is derived from that only the highest confidence level, "Confident that the criterion is fulfilled", is counted as the criterion is fulfilled. The second point on the curve is derived from that the highest and second highest confidence levels, "Confident that the criterion is fulfilled" and "Somewhat confident that the criterion is fulfilled", are counted as the criterion is fulfilled. The last point is derived from including all confidence categories in the concept of fulfilled criterion, meaning that $ICS_1=ICS_2=1.0$.

In the same manner as in Receiver Operating Characteristics, the area under the VGC curve is used as an evaluation tool to determine which modality is the best. By using available software packages it is possible to perform a statistical analysis. The area under the VGC curve can in most cases, exceptions are made for small number of cases or if the area are close to unity, be treated as a normally distributed variable which implies that standard statistical test can be used [21]. If the area under the VGC curve is 0.5 it corresponds to an image quality that are equal for the two modalities. A value that differs in any direction represents an advantage for one of the modalities.

Visual grading characteristics analysis make it possible to perform a visual grading study with multiple scale steps as Visual Grading Analysis and use a valid statistical method instead of being limited to use a two-step rating scale as Image Criteria Scoring because of insufficient statistical methods. The VGC analysis method can be applied directly on the image quality criteria defined by the European Comission [20] if desired, but the designer of the study are not limited to these criteria.
The use of the area under the VGC curve as an evaluation tool can be questioned in the same way as in ROC analysis. When it comes to compare two areas, it may be possible that the size of the areas are the same if the two curves crossing each other and thus implying difficulties in interpreting the result[22] .A major drawback in VGC analysis is the disability to handle a study situation which involves several observers. The lack of procedure to make a correction for inter observer variations and the difficulties in using this method for situations involving several independent variables has led to development of other visual grading analysis methods.

3.3 Visual Grading Regression (VGR)

Visual Grading Regression is a method used to analyse data from visual grading experiments with an ordinal logistic regression model. The method has been developed by Smedby and Fredriksson [2]. Ordinal logistic regression models handle advantageous situations involving several influencing factors, such as different modalities, observers and post processing methods. The method is also able to handle variations that arise due to differences between patients and between observers.

Visual Grading Regression may be applied on both visual grading studies using single-image rating and visual grading studies using image-pair rating. There are slight variations between these two types with regard to organising and analysing the data. The recommended way of organising the data is to create a table and assign one column for each independent variable, see section 3.3.1 "Example of practical implementation of Visual Grading Regression". Patient, observer and image modality are examples of independent variables. The observers are, in the same manner as in other visual grading studies, grading their confidence about the fulfilment of an image quality criterion. The ordinal rating scale, consisting of different confidence levels, is often converted into a numeric scale that, however, still is an ordinal scale. The ordinal logistic regression model is then applied to the organised data and the outcome gives information about the effect of each independent variable on the dependent variable and how well the model fits the data. The basics behind ordinal logistic regression will be shortly described.

The ratio of the probability of an event occurring to the probability of an event not occurring is called the odds for the event:

$$odds = \frac{p}{1-p} \tag{8}$$

Taking the logarithm of the odds creates a logistic function:

$$\text{logit}(p) = \log(odds) = \log\left(\frac{p}{1-p}\right) \tag{9}$$

The logistic function is very useful since it can take as an input any value from minus infinity to plus infinity while the output is confined to a value between zero and one (figure 8). In the simplest case with one independent continuous variable x, the logistic function takes the form:

$$\text{logit}(p) = ax + b \tag{10}$$

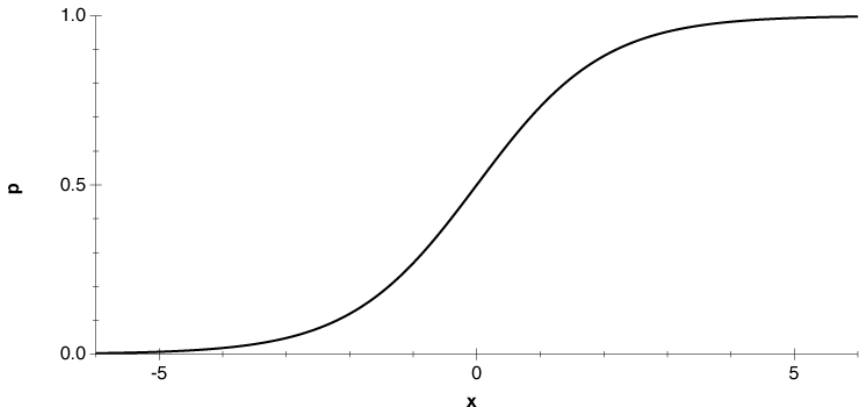

Figure 8: Probability p as a function of a single independent variable x, according to a simple logistic regression model with a= -1 and b= 0 [2].

Implying that the probability p, which represents the probability of a particular outcome, is given by:

$$\log\left(\frac{p}{1-p}\right) = ax + b \tag{11}$$

$$p = \frac{1}{1+\frac{1}{e^{(ax+b)}}} \tag{12}$$

In the case of more than one independent variable, a linear combination of the independent variables is needed. The probability of a particular outcome predicted by the model is given by:

$$p = \frac{1}{1+e^{-z}} \tag{13}$$

Where z is a weighted sum of independent numerical variables, z is hence a measure of the total contribution of all the independent variables used in the model. A value of the probability for a particular outcome is obtained for each combination of the independent variables. Dummy variables are often used for the different independent variables, meaning they are binary response variables. The independent variables can thus take the value of either zero or unity, depending on their contribution to a particular outcome in the probability calculation.

Applying this analysis method on a visual grading study, the dependent variable y is defined on an ordinal scale and represents the rating score predicted by the model. The value on the ordinal rating scale is notated as n, i.e. n takes the values of 1, 2, 3 or 4 if a four-point scale is used. The cumulative probability $P(y \leq n)$ hence represents the probability of obtaining a value not greater than n for a particular combination of the independent variables.

$$P(y \leq n) = \frac{1}{1+e^{-z}} \tag{14}$$

The weighted sum z is said to be the logistic function of the cumulative probability, which is given by:

$$\text{logit}\big(P(y \leq n)\big) = \log\left(\frac{P(y \leq n)}{1-P(y \leq n)}\right) = z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_i x_i - c_n + D_P + E_0 \tag{15}$$

Where β_0 is the intercept, i.e. the value z takes if all the independent variables are zero, and $\beta_1, \beta_2 \ldots \beta_i$ are coefficients to be estimated. The parameters C_n, D_P and E_O have specific values for each quality level, patient and observer. The independent variables are notated as $x_1, x_2 \ldots x_i$, which for example represent different activity levels or post processing methods.

Statistical software programs in the data evaluation process provide the user with numerical results of the ordinal logistic regression. Information on how well the model fits and explain the studied data together with the effect of each independent variable on the dependent variable are provided by such programs. Significance tests like the likelihood ratio chi-square test are often performed to assess the effect of the independent variables and parameters like R^2, the percentage of variance explained by the model, are used in describing the goodness of fit. The coefficients in equation 15 are estimated and the uncertainty in these estimates is usually indicated by standard errors and confidence limits. A regression coefficient that has a positive value means that the explanatory variable increases the probability of the specific outcome, while a negative regression coefficient means that the variable decreases the probability of that outcome. A large regression coefficient,

either positive or negative, means that the variable has a large influence on the probability of that specific outcome, while a small or near-zero coefficient implies that the variable has a small impact on the probability. For example, a negative coefficient will contribute to a reduction of the value of z and thus a decrease in probability of obtaining a rating value y not greater than n.

The estimated coefficients can also be interpreted as Odds Ratios (OR), which might be easier to understand [23]:

$$OR = e^{"estimate"} \tag{16}$$

For example if the estimated value on the coefficient for imaging modality two is 1.5, the Odds Ratios for modality one versus modality two is thus:

$$OR = \frac{\left(\frac{P(y\leq n)}{P(y>n)}\right)_{mod\,1}}{\left(\frac{P(y\leq n)}{P(y>n)}\right)_{mod2}} = e^{1,5} \approx 4,5 \tag{17}$$

which indicates that modality two corresponds to an increase in value of the rating score scale, i.e. a decrease of $P(y \leq n)$. If rating score 1 corresponds to the best image quality, this outcome can hence be interpreted as modality two implies a degradation of the image quality.

A graphical presentation of the results can be made in some different ways. The cumulative probability can be compiled in a graph and plotted against values of the linear combination of the independent variables, z. The value of z can be thought of as a risk score, which is the sum of the combination of each factor potentially affecting the visual grading score. If a five-point scale is used, four curves are created which corresponds to the values of the parameters C_n in equation 15. If a vertical line is drawn in the graph, representing a specific z value, the length of every vertical line segment between two curves corresponds to the probability that the model predict for that particular score. A particular example of this graphical presentation of the results is illustrated in figure 9.

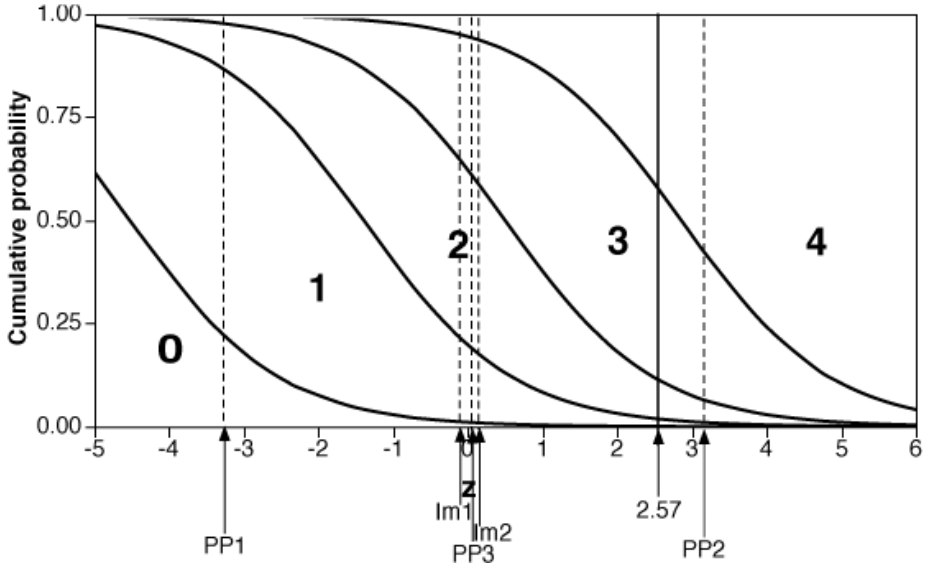

Figure 9: Graph of predicted cumulative probabilities ($P(y \leq n)$) according to an ordinal logistic regression model applied to a single-image grading experiment. The dashed vertical lines represent the predictions for two different types of imaging equipment (Im1 and Im2) and three different post-processing techniques (PP1, PP2 and PP3), and the solid vertical line refers to a numerical example of z corresponding to a particular combination of imaging modality, post-processing techniques and observer. The length of every vertical line segment between two curves corresponds to the predicted probability for a particular score.

The fact that the data in a visual grading study belongs to an ordinal scale are handled correctly in VGR by using ordinal logistic regression, instead of treating the dependent variable as an interval variable and applying t-tests or analysis of variance as in VGA. Visual Grading Regression analysis has the advantage of taking into account the potential bias arising from individual differences between patients or observers. These two parameters should be handled as random effect parameters, thus they are seen as random samples from a larger or infinite population [24]. The method also has the advantage of simultaneously include several factors that possibly affect the image quality. Ordinal logistic regression has not yet been applied to so many visual grading studies of image quality and those found are outside the field of nuclear medicine [25-27].

One should be cautious as the number of independent variables or parameters to be estimated, approaches or exceeding the number of observations. In such situation there is a risk that the model fits the data too well, meaning that the dependence of the number of observations are major and small changes in data can result in large changes to the estimated parameters. Another problem that is general when fitting statistical model to data is that the set of data used to create the model often gives the best value of goodness of fit. A general recommendation is therefore to use one set of data to create the model and another set of data to test the model. This is in practice often difficult to perform.

3.3.1 Example of practical implementation of Visual Grading Regression

Consider a visual grading experiment where a number of patients (P) have undergone an examination with different amount of injected activity (A) and the result is evaluated by a number of observers (O).

Using dummy variables for the independent variable A (Administered activity), according to the data organized in table 1, equation 15 will take the form:

$$\text{logit}\big(P(y \leq n)\big) = z = \beta_1 A_{500} + \beta_2 A_{625} + \beta_3 A_{750} + \beta_4 A_{875} + \beta_5 A_{1000} - c_n + D_P + E_0$$

Implementing the data in table 1 into a statistical software program requires that one category of the independent variables are chosen as the reference category, for example A_{1000}, which the other categories are compared against.

Table 1: Data from a Visual Grading Experiment. Images based on different amount of administered activity are compared with respect to an image quality criterion. Dummy variable coding of the independent variable A (administered activity) is used Combining 5 levels of adminitered activity (A), 3 observers (O) and 21 patients (P) yields a dataset with 315 observations. The score is defined on a scale ranging from 1 (best image quality) to 4 (worst image quality).

Observation number	A_{500}	A_{625}	A_{750}	A_{875}	A_{1000}	O	P	Score
1	1	0	0	0	0	1	1	2
2	0	1	0	0	0	1	1	3
3	0	0	1	0	0	1	1	2
4	0	0	0	1	0	1	1	1
5	0	0	0	0	1	1	1	1
6	1	0	0	0	0	1	2	4
7	0	1	0	0	0	1	2	1
8	0	0	1	0	0	1	2	2
9	0	0	0	1	0	1	2	1
10	0	0	0	0	1	1	2	2
11	1	0	0	0	0	1	3	3
…	…	…	…	…	…	…	…	…
315	0	0	0	0	1	3	21	1

4. Material and Methods

The study comprises a material of 21 consecutive patients with dementia issue. A regional cerebral blood flow single photon emission computed tomography (rCBF SPECT) has been made on these patients at Umeå university hospital. The patients should have undergone this examination regardless of this study, meaning that this study does not affect them further. All images in the study were anonymized and can not be traced to a specific patient.

Image quality criteria applicable to visual grading studies that correspond to rCBF SPECT have been defined. The criteria were developed in collaboration with experienced radiologists in nuclear medicine, who have extensive knowledge on how to evaluate rCBF SPECT images.

Three experienced observers, i.e. specialists in nuclear medicine, took part in the study and evaluated the images according to the evaluation method used, Visual Grading Regression.

4.1 Acquisition

An amount of 1000 MBq ^{99m}Tc labelled HMPAO was injected to all patients in the study. The injected amount of radiopharmaceutical corresponds to the amount in the standard clinical protocol used at Umeå University Hospital, which is where the examinations took place. The examinations were performed using an hybrid SPECT/CT system (Infinia Hawkeye 4, GE Healthcare, Fairfield, Connecticut, USA). The acquisition parameters were chosen due to standard rCBF SPECT examination; LEHR (Low Energy High Resolution) collimator, 30 s acquisition time in each projection, 120 projections, 128*128 matrix size, 1.5 in zoom, 2.95 mm pixel size and approximately 15 cm radius of rotation.

A gated acquisition is performed, meaning that the acquisition time in each projection is divided into several parts. Each part usually corresponds to the length of the cardiac cycle. The cardiac cycle is further subdivided into multiple frames of equal duration, which means that image data for each of the frames are acquired repeatedly over many cardiac cycles and stored separately in the computer [28]. All data from a particular frame is then added together to construct a specific phase of the cardiac cycle. In gated myocardial perfusion SPECT, which is the most common application of gated acquisition, the cardiac cycle is often divided into eight frames.

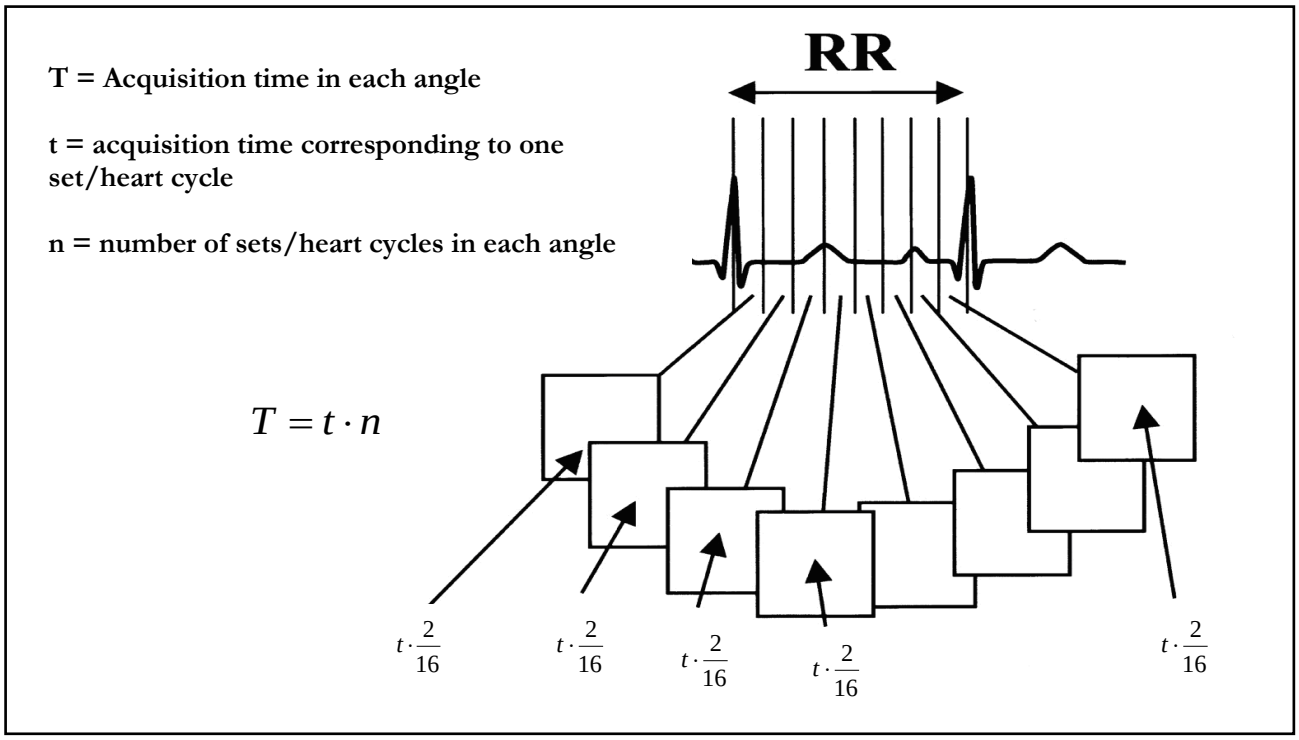

Figure 10: Illustration of how the principle of gated acquisition was used in the study. RR corresponds to the interval between two R-waves of the heart cycle.

In this study the cardiac cycle is not of interest but the ability to divide the acquisition time into subparts in a nice way is of interest, due to the possibility to generate images corresponding to different activity levels. The gating system is in this study not correlated and trigged to the cardiac cycle of a specific patient but correlated to an automatic cycle, i.e. an ECG simulator. This automatic cycle is subdivided into sixteen equal frames, implying the acquisition time in every projection is divided into a number of sets, each consisting of sixteen frames (se figure 10). The acquired counts in a particular frame are summed together for every cycle and projection. An image reconstructed by using the data from just one frame can be interpreted as an examination with the same acquisition time as the whole examination, but with one sixteenth of the injected activity. When all sixteen frames are summed together, this is equivalent to the real ungated examination and thus an activity level of 1000 MBq. Similar examination has been made with the same aim, i.e. generate images corresponding to different activity levels by using a gated acquisition, in myocardial perfusion SPECT [29].

A diagnostic computed tomography (CT) scan were made in combination with the SPECT examination, to be used in the diagnostic evaluation and for the attenuation correction.

4.2 Reconstruction

From one single patient examination, five studies that correspond to different activity levels were generated. A total number of 105 images were reconstructed. The studies have been generated by reconstructing data from 8, 10, 12, 14 and all 16 frames, thus corresponding to images generated by an injected activity amount of 500 MBq, 625 MBq, 750 MBq, 875 MBq and 1000 MBq. An example of the resulting five activity levels are shown in figure 11.

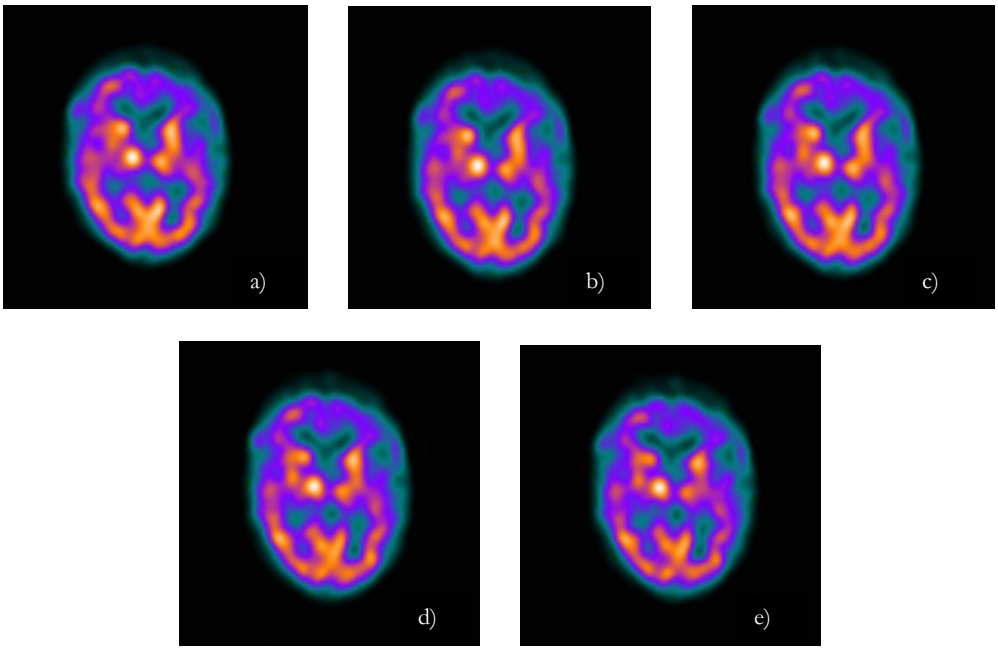

Figure 11: Transaxial section of the brain in a rCBF SPECT image corresponding to an activity level of a) 1000 MBq, b) 875 MBq, c) 750 MBq, d) 625 MBq and e) 500 MBq.

An iterative reconstruction method, OS-EM, were used and the parameters were chosen according to an unpublished paper [Olsson A.], in which the parameters, i.e. number of iterations, number of subsets and filtration properties, were optimised due to image noise and contrast. A clone of General Electric's Xeleris Evolution Bone SPECT were created and adjusted to the chosen properties of the iterative reconstruction; 8 iterations, 10 subsets with each 12 angles and butterworth filter with power 8 and cut-off frequency 0,45. Corrections were included in the iterative reconstruction for scatter, attenuation and distance dependent resolution.

Reorientations and slices were made in the reconstructed images according to clinical standards. Transversal views were used in the study.

4.3 Determination of image criteria

Well established and comprising image quality criteria are available for visual grading studies in the field of diagnostic radiology [20]. In the field of nuclear medicine imaging there is a lack of defined image quality criteria used in visual grading studies. Desirable is to define comprehensive and exhaustive image criteria regarding image quality that covers the whole brain.

Together with experienced specialist in nuclear medicine with extensive knowledge on how to evaluate rCBF studies, image quality criteria applicable to rCBF SPECT images have been defined. Seven image criteria were defined, of which five considered specific anatomic structures and the other two considered image quality in general. The image quality criteria defined are:

1. "Grey and white matter are discernible in the cerebellum"
2. "Grey and white matter are discernible in the medial and lateral parts of the temporal lobe"
3. "Grey and white matter are discernible laterally in the frontal lobes"
4. "Thalamus is bilaterally discernible"
5. "White matter is discernible from the ventricles in the parietal lobes"
6. "The noise level does not have a disturbing effect on the assessment"
7. "Overall image quality is good enough to provide clinical diagnosis"

4.4 Analysis

4.4.1 Image analysis

The reconstructed images were transferred as DICOM-files to a visual grading software package, ViewDEX; University of Gothenburg [30]. The software package was configured with our defined image criteria and the design and performance of the image viewer were edited to a suitable appearance. The images were shown randomly, one by one, and the seven image criteria were shown simultaneously together with a four-point confident scale. Three experienced observers, Umeå University Hospital, in nuclear medicine were visually assessing the images, thus rating their confident of the fulfilment of each criterion. The confident scale used was as follows:

1. "Completely certain that the criterion is fulfilled"
2. "Almost certain that the criterion is fulfilled"
3. "Almost certain that the criterion is not fulfilled"
4. "Completely certain that the criterion is not fulfilled"

Low numerical value thus corresponds to high image quality and vice versa.

The study resulted in 315 observations (3*5*21, i.e. # observers * # activity levels * # patients) and 2205 parameter ratings (3*5*21*7, i.e. # observers * # activity levels * # patients * # criteria). The observers had the possibility to use a demo version and assess five images before assessing the main version of the study. They were also able to freely adjust and use the window settings they are comfortable with. The result of the visual grading were compiled and stored in a text-file.

4.4.2 Statistical analysis

The result file contains information about how the observers have rate their confidence about the fulfilment of each image criterion for every image. The ratings of the observers are shown in a way that they easily can be correlated to the corresponding activity level, patient and observer. The results of the study were analysed using Visual Grading Regression, a statistical analysis method based on ordinal logistic regression. The five different images types, i.e. images based on different activity levels, were compared. The corresponding variables of observers and patients were taken into account in the model, in order to handle the variation that arises due to differences between patients and observers. These variables should be introduced in the model although the researcher is not interested in differences between specific observers or patients.

Applying the model to the whole set of data implies that the different activity levels are compared to the highest activity level, in this case images corresponding to 1000 MBq, with respect to differences in observers' assessment. In this way it is possible to investigate whether there is a significant difference between for example 500 MBq and 1000 MBq. To compare if there is a significant difference between 500 MBq and 750 MBq the activity level of 1000 MBq must be excluded of the study. For comparing all activity levels relative each other mean four statistical analyses have to be made.

The Visual Grading Regression analysis was performed using the statistical analysis software STATISTICA Version 10 (StatSoft, Tulsa, Oklahoma, USA).

5. Results

The observed frequencies of rating scores 1-4 for each of the five different activity levels and for each criterion are summarized in figure 12-18.

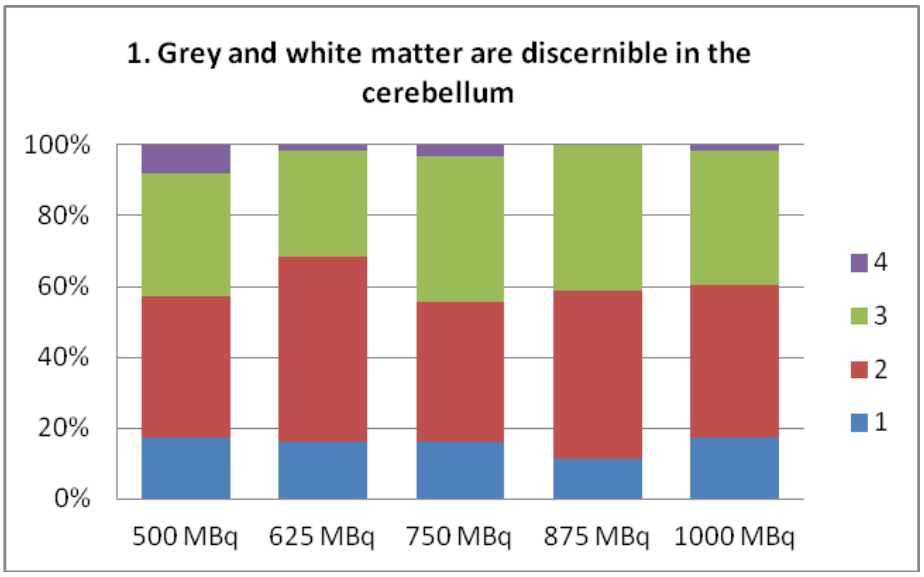

Figure 12: Result of criterion 1, observed frequencies of scores 1-4 for each of the five different activity levels.

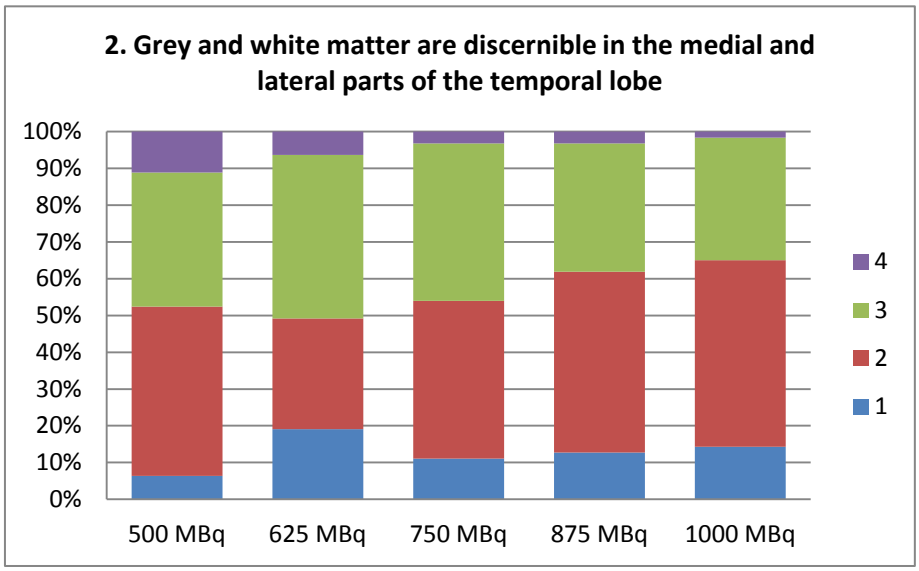

Figure 13: Result of criterion 2, observed frequencies of scores 1-4 for each of the five different activity levels.

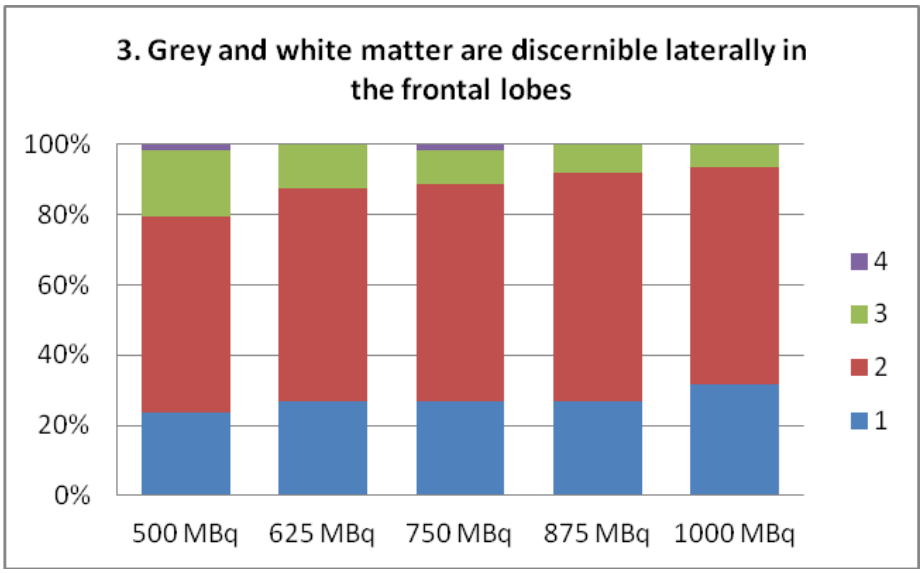

Figure 14: Result of criterion 3, observed frequencies of scores 1-4 for each of the five different activity levels.

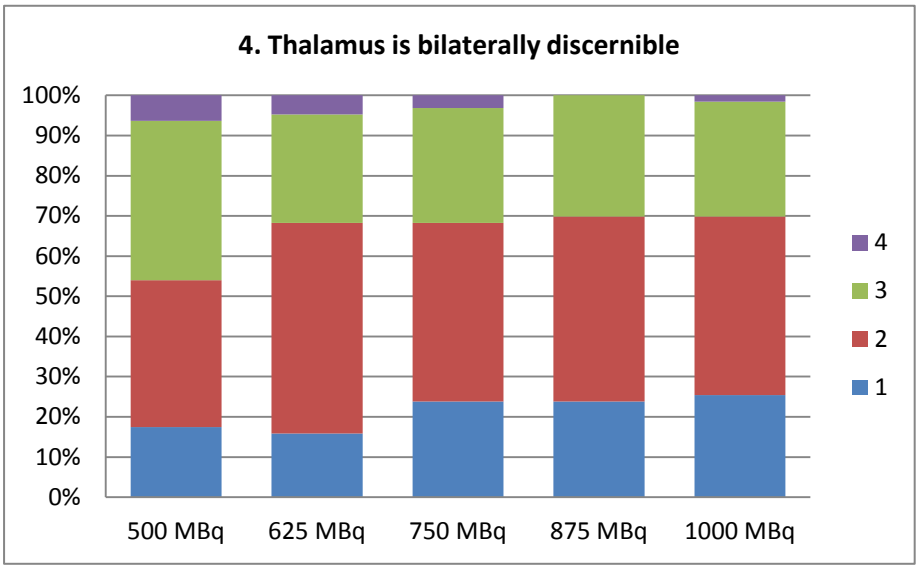

Figure 15: Result of criterion 4, observed frequencies of scores 1-4 for each of the five different activity levels.

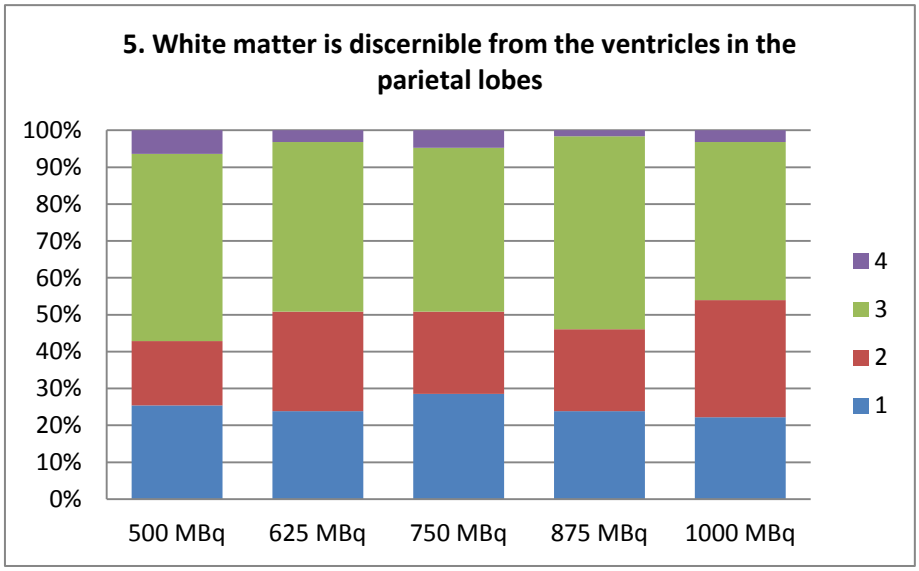

Figure 16: Result of criterion 5, observed frequencies of scores 1-4 for each of the five different activity levels.

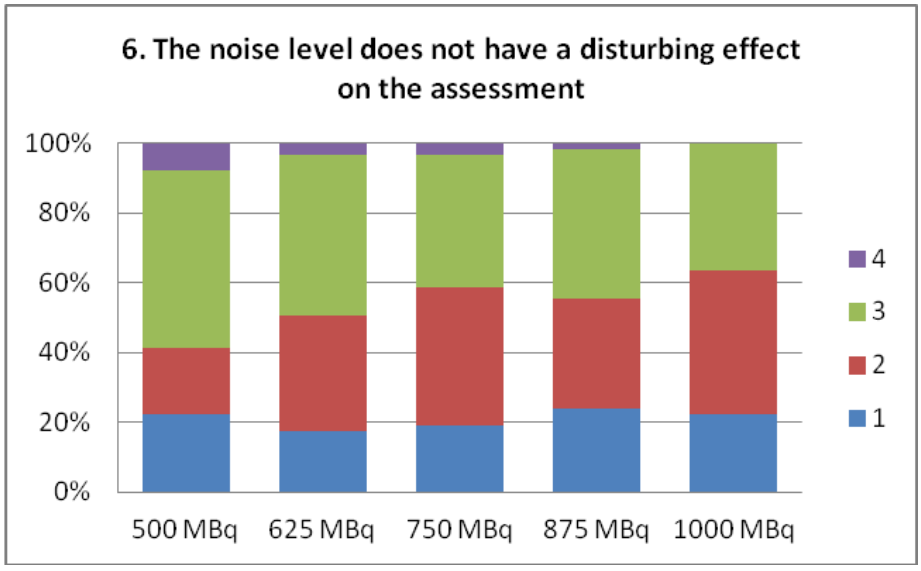

Figure 17: Result of criterion 6, observed frequencies of scores 1-4 for each of the five different activity levels.

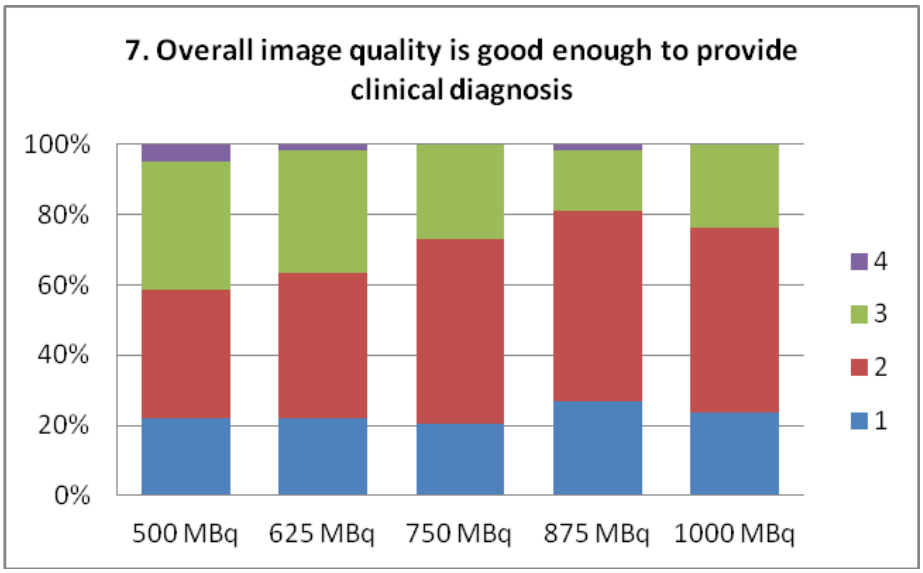

Figure 18: Result of criterion 7, observed frequencies of scores 1-4 for each of the five different activity levels.

Results of the VGR analysis are presented in table 2-4. In those cases where there is significance difference, p-value and parameter estimate can be seen. A negative value of parameter estimate indicates a decrease in $P(y \leq n)$ and thus less probable to obtain a low grading score which corresponds to good image quality, meaning that a negative value implies a degradation of image quality.

Table 2: Results of the VGR analysis using 1000 MBq as reference level of activity.

Criterion	p-value /estimate	p-value /estimate	p-value /estimate	p-value /estimate
	1000 MBq vs. 875 MBq	1000 MBq vs. 750 MBq	1000 MBq vs. 625 MBq	1000 MBq vs. 500 MBq
1	Not significant	Not significant	Not significant	Not significant
2	Not significant	Not significant	Not significant	0,009670 / -0,62476
3	Not significant	Not significant	Not significant	0,002054 / -0,91580
4	Not significant	Not significant	Not significant	0,001575 / -0,75629
5	Not significant	Not significant	Not significant	Not significant
6	Not significant	Not significant	Not significant	0,019504 / -0,55771
7	0,015475 / 0,59	Not significant	Not significant	0,029642 / -0,50412

Table 3: Results of the VGR analysis using 875 MBq as reference level of activity.

Criterion	p-value /estimate	p-value /estimate	p-value /estimate
	875 MBq vs. 750 MBq	875 MBq vs. 625 MBq	875 MBq vs. 500 MBq
1	Not significant	0,038981 / 0,50770	Not significant
2	Not significant	Not significant	0,030237 / -0,50483
3	Not significant	Not significant	0,012728 / -0,70699
4	Not significant	Not significant	0,003714 / -0,67482
5	Not significant	Not significant	Not significant
6	Not significant	Not significant	Not significant
7	Not significant	Not significant	0,049296 / -0,44087

Table 4: Results of the VGR analysis using 750 MBq and 625 MBq as reference level of activity.

Criterion	p-value /estimate	p-value /estimate	p-value /estimate
	750 MBq vs. 625 MBq	750 MBq vs. 500 MBq	625 MBq vs. 500 MBq
1	0,043244 / 0,46618	Not significant	Not significant
2	Not significant	Not significant	Not significant
3	Not significant	Not significant	Not significant
4	Not significant	0,017697 / -0,51875	Not significant
5	Not significant	Not significant	Not significant
6	Not significant	Not significant	Not significant
7	Not significant	Not significant	Not significant

6. Discussion

In this study, the visually assessed image quality of images corresponding to five different activity levels has been compared. The result shows that there is a significant difference in the perceived image quality between 500 MBq and 1000 MBq. This difference can be observed in five of the seven image quality criteria.

The results of the study are represented by the bar graphs, showing the observed frequencies of the rating scores 1-4 for each activity level and each criterion. The results of the VGR analysis are shown in table 1-3. The visual impact of the bar graphs in figure 12-18 is in general that increased activity level provides more observations with lower scores. Since the scoring scale is defined in such a way that the lowest score corresponds to the highest image quality, an increase in activity means thus an improvement in image quality. A typical example of this result can be seen in criterion number 2 (figure 13).

By using these kinds of image quality criteria, which mainly are based on the discernibility and visibility of different anatomical structures, a threshold may exist at a point where an increase of activity no longer implies a higher visibility of the specific anatomical structure. In the bar graphs of criterion number 3, 4, 6 and 7, a plateau can be seen where an increased activity level no longer seems to provide higher satisfaction of the fulfillment of the criterion. In the two remaining criteria, number 1 and 5, it can not be shown that the satisfaction is influenced by the level of administered activity. The assessment of these anatomical areas seems to be regardless of activity level.

Different levels of satisfaction or the frequency of criteria rated as fulfilled, i.e. proportion of the observations belonging to confidence level one and two, can be viewed between the seven criteria. The level of satisfaction is approximately 45 % in criteria number 2, while in criteria number 3 this level is approximately 90 %. These differences might be due to how difficult the particular anatomical area is to assess in an rCBF SPECT image.

Criterion number 2 is defined as "Grey and white matter are discernible in the medial and lateral parts of the temporal lobe". The temporal lobe is of major interest in rCBF SPECT images and especially in the diagnosis of different types of dementia. In the temporal lobe is hippocampus located, which is a part of the brain that plays an important role in the episodic memory. Several factors make this area difficult to assess. The area around the hippocampus has a large anatomical variation and in the area around the Sylvian fissure there can be partial substance defects that might affect the visual assessment. By viewing the observed frequencies of the rating score in figure 13, the assessment of the temporal lobe seems to be easier when using a higher amount of administered activity. The improvement of image quality appears to rise all the way up to 1000 MBq in this criterion. This trend is not shown by a statistically significant result, only between 500 MBq and 1000 MBq a significant difference was found.

Criterion number 3, 4, 6 and 7 has approximately the same appearance where a plateau seems to exist. Criteria number 3 is defined as "Grey and white matter are discernible laterally in the frontal lobes" and the most obvious of this criterion is the high level of observations rated as fulfilled, i.e. the proportion of confidence scale step one and two. The frontal lobe is an area that in general is easy to assess, which largely is due to the area's anatomical regularity of modest variations. The fissures in the surface of the brain cortex, the sulci, has constant depth in the frontal lobe that makes it easy to separate white matter from grey matter, which thus implies a high confidence about the fulfillment of this criterion.

Criterion number 4 is defined as "Thalamus is bilaterally discernible". This criterion has a relatively high level of satisfaction, approximately 70% of the observations. Thalamus is an area that is easy to image with the gamma camera and often has a high uptake of radioactivity which enable it to be used as a reference region. The observations of this criterion shows an increase in the number of satisfied observations between activity level 500 MBq and 625 MBq, but between the higher activity levels no increase is present and the satisfaction

level is stabilized. In criterion 3 and 4 have the satisfaction reached a constant level at 625 MBq, meaning the observer can't se any difference in image quality corresponding to these two criteria between 1000 MBq and 625 MBq.

Criterion 6 and 7 correspond to the image quality in general and are defined as "The noise level does not have a disturbing effect on the assessment" and "Overall image quality is good enough to provide clinical diagnosis". In both criteria the satisfaction level regarding the fulfillment of the criteria seems to reach a plateau at the activity level of 750 MBq. In criterion number 6, concerning the noise level, it can also be seen that the proportion of confidence scale step 4 where the observer is completely certain that the criterion is not fulfilled, decreases with increasing activity level. At 1000 MBq no images are assessed as certainly affected by noise.

Viewing criterion number 7 in figure 18 corresponding to the image quality in general, the confidence scale step 4 has zero observations at the activity level of 750 MBq, implying that the observers are never completely unsatisfied with the image quality at this activity level. The level of satisfaction reaches a proportion of approximately 75% of the total number of observations, which seems to remain constant from 750 MBq to 1000 MBq. Increasing this proportion of satisfaction might be difficult since individual differences of patients with regard to disease states and individual anatomical uptake always will affect the assessment.

The satisfaction level in criterion number 1 and 5 doesn't seem to be affected by the amount of administered activity. It is not possible to deduce a trend due to different activity levels of this criterion. Criterion number 1 is defined as "Grey and white matter are discernible in the cerebellum" and it has a satisfaction level of about 60% of the observations. Cerebellum is a part of the brain that is easy to assess and always has a high uptake of radioactivity. In quantitative measurements of the regional cerebral blood flow, cerebellum is often used as a reference uptake area. This is hence due to the cerebellum's relatively high uptake of the radiopharmaceutical and the assumption that it is often spared any major pathological involvement [31].

Criterion number 5 is defined as "White matter is discernible from the ventricles in the parietal lobes". This criterion has a satisfaction level of only 45% - 50%, regardless of the amount of administered activity. The formulation of this criterion might be the reason for the low satisfaction level, meaning that white matter is difficult to separate from the ventricles. Looking at the relationship of intensity between different types of tissue in an rCBF SPECT image it can be seen that grey matter is represented by the highest intensity, white matter by slightly lower and the ventricles by the lowest intensity. In an SPECT image representing the area around the ventricles it is sometimes difficult to distinguish the intensity plateau representing white matter. This shortage of the gamma camera implies that the intensity scale goes from grey matter to the ventricles, without any disruption representing white matter intensity. To achieve a better outcome of this criterion regarding the parietal region, a revision of the existing criteria or using an additional criterion could possibly be needed where white matter is replaced by grey matter in the criterion formulation. Including the parietal lobes is important when assessing image quality in rCBF SPECT images, due to the fact that it is essential in dementia investigations as it contains association center.

In this study, image quality criteria corresponding to the visibility, reproduction and discernibility of different anatomical structures in the brain were defined. The criteria were chosen with respect to provide information on image quality of the whole brain, thus including parts of the brain that is most often of interest in rCBF SPECT. Five of the seven criteria used correspond to the visibility and discernibility of certain anatomical structures, while the other two criteria correspond to the overall image quality. The choice of criteria is important in a visual grading study. Desirable is that the entire rating scale is used for every criterion. Such an outcome can be interpreted as the assessing radiologists thus have had a freedom and mobility in their decisions regarding their confidence about the fulfillment of the specific criterion.

Several types of visual evaluation methods are available for example Image Criteria Scoring (IC), Receiver Operating Characteristics (ROC), Visual Grading Analysis (VGA), Visual Grading Characteristics (VGC) and Visual Grading Regression (VGR). In addition to different design of the studies, some of these methods suffer from insufficient statistical analysis methods. This study has several requirements when selecting analysis method according to the chosen design of the study. The method needs to handle data originating from an ordinal scale in a correct way, needs to handle situations where the observers grade their confidence about the fulfillment of a specific criterion instead of just "fulfilled" or "not fulfilled" and the method needs to handle situations involving several observers and patients.

According to these requirements, the choice of visual grading method fell on Visual Grading Regression (VGR). A Visual Grading Regression study gives the observer a freedom in that he can access multiple confidence levels while rating the fulfillment of a criteria, the statistical analysis handle the situation with data on an ordinal scale in a correct way and the analysis method takes the variations between observers and between patients into account as fixed effects. The most correct would be to handle observers and patients as random effects, since they represent a random sample of a larger population. According to a preliminary study[24], handle these parameters as fixed effect instead of random effects don't causes any major differences in the results.

In a visual grading study is not the specific disease state of interest, but only the visibility of a certain structure or the overall quality of the image. Due to the lack of interest in disease status there is no need of knowing the true diagnosis of the patient, unlike in ROC analysis, which hence is a huge advantage for visual grading. If the case is such that the disease state or diagnose is known and there is an interest, for example, to find out how different diagnoses affect the outcome of the observations, there is a possibility to include diagnosis as an independent variable in the VGR analysis. The basic assumption in a visual grading study is that the visibility of a certain structure or the reproduction of anatomy corresponds to the possibility to detect pathology. Criterion number 7 regarding the possibility to provide clinical diagnosis implies an assessment of the ability to detect pathology, thus correlating to the basic assumption of a visual grading experiment.

A number of 21 patients were included in the study; each patient is represented with five reconstructed images which correspond to five different activity levels. Every observer thus evaluates a number of 105 studies which represents a total number of 315 observations when using three observers. As already described, one should be cautious using VGR as the number of independent variables or parameters to be estimated, approaches or exceeding the number of observations. This is not a problem in this study as the number of observations is major in comparison to the number of independent parameters. However, a larger number of observations give better statistics and more reliable results.

The small number of participating observers should be taking into account as a limitation of the study. The evaluation of the images was divided into several sessions which may have contributed to a drift of the observers' assessment between the different studies. In Umeå, where the images were evaluated, slightly different reconstruction parameters and post processing are normally used in clinical practice. This might affect the observers' assessment and has to be taken into account.

Another limitation is that the reconstruction parameters were fixed for all studies and were not optimized for the different activity levels. The reconstruction and post processing parameters are optimized and chosen according to the highest activity level, 1000 MBq. It is conceivable that different number of iterations and subsets together with different post reconstruction filtering, adapted to the level of activity, would lead to an assessment in which the image quality is considered to be better. Such an individual optimization of dose levels would probably lead to a decreased difference in assessment score between the highest and lowest activity level.

The study comprises a material of 21 consecutive patients with dementia issue. The disease state of the patients varies together with the progression of the possible disease. The various disease states of the patients are a factor that may have influenced the assessment of image quality. This has to be taken into account when interpreting the result of the study. By taking the patients consecutively gives a representative sample of the normal patient flow for the department. The situation of interpreting images originating from patients with different diagnosis is in itself nothing strange, rather an analogy of the true clinical assessment situation.

7. Conclusion and future prospects

According to the results in table 1, there is a significant difference in perceived image quality between 500 MBq and the reference activity, 1000 MBq, in five of the seven image quality criteria. No statistical significant degradation was found between any other activity level than 500 MBq and the reference activity (1000 MBq). This study doesn't prove that any other activity level provides the same image quality as 1000 MBq, only because no difference was seen, but it gives an indication that the activity level could be reduced without losing too much diagnostic information.

The analysis method used to evaluate the data from this visual grading experiment, Visual Grading Regression, has proven to be convenient and easy to use for this kind of optimisation studies in nuclear medicine. By using this analysis method, it became possible to use several observers and to make a correction for inter observer variations.

Seven image quality criteria applicable to rCBF SPECT images were defined in the study. The defined criteria cover the areas of the brain that are of interest in blood flow examinations and the results of this study showed that the observers used the whole confidence rating scale for each criterion, which is desirable. Some of the criteria had a very low proportion of rating scores corresponding to a fulfilment of the criterion, meaning that the satisfaction of the observers is low. A reversion or adjustment of these criteria might be needed to investigate whether the low satisfaction level is due to the formulation of the criteria or if only so the particular area is difficult to assess.

Some future prospects are to further investigate parameters that might affect the outcome of the image assessment in this study. The next step is to take more account of individual differences that may affect the uptake of radiopharmaceutical, for example the weight of the patients and the different diagnosis. In some hospitals the amount of administered activity in rCBF SPECT examinations is dependent on weight and in some a fixed amount of activity is used. It would also be of interest to evaluate the result of the study with respect to different disease categories.

Making a comparison of the result from this study with physical parameters, for example the dependence of count statistics in the images, would also be of interest.

The defined image quality criteria will be further developed and adapted to be more suitable for use in nuclear medicine to evaluate rCBF SPECT image quality. Hopefully this method is supposed to be useful in, and facilitate, optimization studies in nuclear medicine in the future.

Acknowledgement

I would like to express my most sincere thanks to my supervisors Anna Olsson and Agnetha Gustafsson for making this project possible. For all the encouraging support, inspiring discussions and indispensable help I've received during this project. Thanks to you, my expectations about this project have been exceeded!

I would also like to extend my warmest thanks to Torbjörn Sundström for your valuable comments and input of medical knowledge. In addition I'd like to thank Michael Ljungberg for your constructive and helpful comments on my report.

Thanks to everyone at the Radiation Physics Department in Linköping for making my time so pleasant!

Linköping, June 2012

Sofia Kvernby

8. References

[1] King AeA. Emission Tomography, Attenuation Scatter and Spatial Resolution Compensation in SPECT,2004.

[2] Smedby Ö and Fredrikson M. Visual grading regression: analysing data from visual grading experiments with regression models. British Journal of Radiology 2010;83:767-75.

[3] Frey EC and Tsui BMW. Collimator-Detector Response Compensation in SPECT Quantitative Analysis in Nuclear Medicine Imaging. In: H Zaidi editor: Springer US; 2006, p. 141-66.

[4] SSM. Strålsäkerhetsmyndighetens föreskrifter om nukleärmedicin SSMFS 2008:34. Strålsäkerhetsmyndighetens författningssamling.

[5] SSM. Strålsäkerhetsmyndighetens föreskrifter och allmänna råd av diagnostiska referensnivåer inom nukleärmedicin SSMFS 2008:4. Strålsäkerhetsmyndighetens författningssamling.

[6] Özlem LKeA. EANM procedure guideline for brain perfusion SPECT using 99mTc-labelled radiopharmaceuticals, version 2. In: EAoN Medicine editor; 2009.

[7] NEMA. Performance measurements of scintillation cameras. Standard publication no.Nu1. Washington, DC: National Electrical manufactures Association; 1994.

[8] Tingberg A, Herrmann C, Lanhede B, Alm_ A, JB, Mattsson S, et al. Comparison of Two Methods for Evaluation of the Image Quality of Lumbar Spine Radiographs. Radiation Protection Dosimetry 2000;90:165-168.

[9] Sandborg M, Tingberg A, Dance DR, Lanhede B, Almén A, McVey G, et al. Demonstration of correlations between clinical and physical image quality measures in chest and lumbar spine screen–film radiography. British Journal of Radiology 2001;74:520-8.

[10] Herrmann C, Sund P, Tingberg A, Keddache S, Mansson LG, Almen A, et al. Comparison of two methods for evaluating image quality of chest radiographs. In: AK Elizabeth editor: SPIE; 2000, p. 251-8.

[11] Sandborg M, Tingberg A, Ullman G, Dance DR, and Carlsson GA. Comparison of clinical and physical measures of image quality in chest and pelvis computed radiography at different tube voltages. Medical Physics 2006;33:4169-75.

[12] Anger HO. Scintillation Camera. Review of Scientific Instruments 1958;29:27-33.

[13] Bruyant PP. Analytic and Iterative Reconstruction Algorithms in SPECT. Journal of Nuclear Medicine 2002;43:1343-58.

[14] Patton JA and Turkington TG. SPECT/CT physical principles and attenuation correction. J Nucl Med Technol 2008;36:1-10.

[15] Hutton BF, Buvat I, and Beekman FJ. Review and current status of SPECT scatter correction. Phys Med Biol 2011;56:R85-112.

[16] ICRP. Radiation Dose to Patients from Radiopharmaceuticals ICRP Publication 53. Ann. ICRP 18 (1-4); 1988.

[17] Sundström T. Human brain function evaluated with rCBF SPECT. Department of Radiation Sciences, Diagnostic Radiology. Umeå: Umeå University; 2006.

[18] Båth M. Evaluating imaging systems: practical applications. Radiation Protection Dosimetry 2010;139:26-36.

[19] Månsson LG. Methods for the Evaluation of Image Quality: A Review. Radiation Protection Dosimetry 2000;90:89-99.

[20] Carmichael JHEeA. European Guidelines on quality criteria for diagnostic radiographic images, EUR 16260 EN, Luxembourg: Office for Official Publications of the European Communities,1996.

[21] Bath M and Mansson LG. Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. Br J Radiol 2007;80:169-76.

[22] McNeil BJ and Hanley JA. Statistical Approaches to the Analysis of Receiver Operating Characteristic (ROC) Curves. Medical Decision Making 1984;4:137-50.

[23] Petrie AaS, C. Binary outcomes and logistic regression. Medical Statistics at a Glance. West Sussex: Wiley-Blackwell; 2009, p. 88-91.

[24] Smedby O, Fredrikson M, Geer JD, and Sandborg M. Visual grading regression with random effects. In: KA Craig and RM-T Claudia editors: SPIE; 2012, p. 831805.

[25] De Geer J, Sandborg M, Smedby Ö, and Persson A. The efficacy of 2D, non-linear noise reduction filtering in cardiac imaging: a pilot study. Acta Radiologica 2011;52:716-22.

[26] Hamer OW, Sirlin CB, Strotzer M, Borisch I, Zorger N, Feuerbach S, et al. Chest Radiography with a Flat-Panel Detector: Image Quality with Dose Reduction after Copper Filtration1. Radiology 2005;237:691-700.

[27] Smedby Ö, Öberg R, Åsberg B, Stenström H, and Eriksson P. Standardized Volume‐Rendering of Contrast‐Enhanced Renal Magnetic Resonance Angiography. Acta Radiologica 2005;46:497-504.

[28] Nabi AKPaHA. Gated Myocardial Perfusion SPECT: Basic Principles, Technical Aspects, and Clinical Applications. Journal of Nuclear Medicine Technology 2004;32:179-87.

[29] Robinson C, van Aswegen A, Julious S, Nunan T, Thomson W, Tindale W, et al. The relationship between administered radiopharmaceutical activity in myocardial perfusion scintigraphy and imaging outcome. European Journal of Nuclear Medicine and Molecular Imaging 2008;35:329-35.

[30] Håkansson M, Svensson S, Zachrisson S, Svalkvist A, Båth M, and Månsson LG. ViewDEX: an efficient and easy-to-use software for observer performance studies. Radiation Protection Dosimetry 2010;139:42-51.

[31] Soonawala D, Amin T, Ebmeier KP, Steele JD, Dougall NJ, Best J, et al. Statistical Parametric Mapping of 99mTc-HMPAO-SPECT Images for the Diagnosis of Alzheimer's Disease: Normalizing to Cerebellar Tracer Uptake. NeuroImage 2002;17:1193-202.