Optimization of image quality and radiation dose in neuroradiological computed tomography

Löve, Askell

2013

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Optimization of image quality and radiation dose in neuroradiological computed tomography

Áskell Löve
Optimization of Image Quality and Radiation Dose in Neuroradiological Computed Tomography

Abstract

Background: The goal of clinical computed tomography (CT) is to produce images of diagnostic quality using the lowest possible radiation exposure. Degradation of image quality, with increased image noise and reduced spatial resolution, is a major limitation for radiation dose reduction in CT. This can be counteracted with new post-processing image filters and iterative reconstruction (IR) algorithms that improve image quality and allow for reduced radiation doses. Implementation of new methods in clinical routine requires prior validation in phantoms and clinical feasibility studies including comprehensive evaluation of image quality.

Aims: The main objectives of this thesis were to assess new methods for improvement of image quality in CT, explore the associated potential for radiation dose reduction, and to outline a comprehensive approach for evaluation of image quality.

Methods: Extensive phantom testing was performed and a total of 100 human subjects were included in the clinical studies. Image quality and diagnostic acceptability were assessed in brain CT acquired with 30% reduced radiation dose in combination with post-processing filter (Paper I) and IR (Paper II). The potential of IR for image quality improvement, without concomitant radiation dose reduction, was assessed in craniocervical CT angiography (CTA) (Paper III). The performance of six IR algorithms was evaluated in a brain CT phantom model (Paper IV), using different combinations of radiation dose levels and iterative image optimization levels. Throughout the studies, various approaches for subjective and objective evaluation of image quality were used and assessed.

Results: Post-processing image filtering (Paper I) and IR (Paper II) compensated partly or entirely for the loss of image quality caused by 30% reduced radiation dose in brain CT. In both studies, considerable inter-observer variation was seen. In Paper II a discrepancy was seen between results of objective and subjective evaluation of image quality and also between grading and ranking, indicating observer bias. Statistical IR improved image quality in craniocervical CTA (Paper III) with fairly good inter-observer agreement. Despite having different strengths and weaknesses, the six iterative reconstruction algorithms evaluated in Paper IV all improved image quality. Best over-all improvement was seen for one of the model-based IR algorithms, especially at lower radiation doses.

Conclusion: All evaluated methods improved image quality and showed potential for radiation dose reduction while maintaining diagnostic quality. Careful study design and comprehensive evaluation of image quality including objective and subjective evaluation steps may reduce observer bias and improve reliability of study results.

Key words: Computed tomography; image quality; radiation dose; neuroradiology; post-processing image filters; iterative reconstruction; filtered back-projection brain CT; craniocervical CT angiography; phantom.
Optimization of image quality and radiation dose in neuroradiological computed tomography

Áskell Löve

Department of Diagnostic Radiology,
Clinical Sciences Lund, Faculty of Medicine,
Lund University, Skåne University Hospital
2013
In CT, too low or too high image quality is equally inappropriate

– Åskell Löve
Contents

Abstract .................................................................................................................................... 7
Abbreviations used ................................................................................................................... 9
Original papers ..................................................................................................................... 11
Introduction .......................................................................................................................... 13
Background ........................................................................................................................... 15
  Computed tomography ........................................................................................................ 15
    Technique .......................................................................................................................... 15
  Radiation dose ..................................................................................................................... 16
  Image reconstruction ......................................................................................................... 17
  Parameters of image quality .............................................................................................. 17
  Image quality improvement ............................................................................................... 18
Evaluation of image quality .................................................................................................. 19
  Objective evaluation .......................................................................................................... 19
  Subjective evaluation .......................................................................................................... 21
  Observer bias ....................................................................................................................... 22
  Evaluation scales ................................................................................................................ 23
  Research subjects ............................................................................................................... 25
Aims ........................................................................................................................................ 27
  Paper I ................................................................................................................................. 27
  Paper II ................................................................................................................................. 27
  Paper III ............................................................................................................................... 27
  Paper IV ............................................................................................................................... 28
Abstract

**Background:** The goal of clinical computed tomography (CT) is to produce images of diagnostic quality using the lowest possible radiation exposure. Degradation of image quality, with increased image noise and reduced spatial resolution, is a major limitation for radiation dose reduction in CT. This can be counteracted with new post-processing image filters and iterative reconstruction (IR) algorithms that improve image quality and allow for reduced radiation doses. Implementation of new methods in clinical routine requires prior validation in phantoms and clinical feasibility studies including comprehensive evaluation of diagnostic image quality.

**Aims:** The main objectives of this thesis were to assess new methods for improvement of image quality in CT, explore the associated potential for radiation dose reduction, and to outline a comprehensive approach for evaluation of image quality.

**Methods:** Extensive phantom testing was performed and a total of 100 human subjects were included in the clinical studies. Image quality and diagnostic acceptability were assessed in brain CT acquired with 30% reduced radiation dose in combination with post-processing filter (Paper I) and IR (Paper II). The potential of IR for image quality improvement, without concomitant radiation dose reduction, was assessed in craniocervical CT angiography (CTA) (Paper III). The performance of six IR algorithms was evaluated in a brain CT phantom model (Paper IV), using different combinations of radiation dose levels and iterative image optimization levels. Throughout the studies, various approaches for subjective and objective evaluation of image quality were used and assessed.

**Results:** Post-processing image filtering (Paper I) and IR (Paper II) compensated partly or entirely for the loss of image quality caused by 30% reduced radiation dose in brain CT. In both studies, considerable inter-observer variation was seen. In Paper II a discrepancy was seen between results of objective and subjective evaluation of image quality and also between grading and ranking, indicating observer bias. Statistical IR improved image quality in craniocervical CTA (Paper III) with fairly good inter-observer agreement. Despite having different strengths and weaknesses, the six iterative reconstruction algorithms evaluated in Paper IV all improved image
quality. Best over-all improvement was seen for one of the model-based IR algorithms, especially at lower radiation doses.

**Conclusion:** All evaluated methods improved image quality and showed potential for radiation dose reduction while maintaining diagnostic quality. Careful study design and comprehensive evaluation of image quality including objective and subjective evaluation steps may reduce observer bias and improve reliability of study results.
Abbreviations used

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAPM</td>
<td>American association of physicists in medicine</td>
</tr>
<tr>
<td>AIDR</td>
<td>Adaptive iterative dose reduction (Toshiba)</td>
</tr>
<tr>
<td>ALARA</td>
<td>As low as reasonably achievable</td>
</tr>
<tr>
<td>ASIR</td>
<td>Adaptive statistical iterative reconstruction (GE)</td>
</tr>
<tr>
<td>CNR</td>
<td>Contrast-to-noise ratio</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>CTA</td>
<td>Computed tomography angiography</td>
</tr>
<tr>
<td>CTDI</td>
<td>Computed tomography dose index</td>
</tr>
<tr>
<td>DLP</td>
<td>Dose length product</td>
</tr>
<tr>
<td>FBP</td>
<td>Filtered back-projection</td>
</tr>
<tr>
<td>FLD</td>
<td>Filtered low-dose (Paper I)</td>
</tr>
<tr>
<td>HU</td>
<td>Hounsfield unit (attenuation)</td>
</tr>
<tr>
<td>ID2</td>
<td>iDOSE(^4) noise reduction level 2 (Paper II)</td>
</tr>
<tr>
<td>ID4</td>
<td>iDOSE(^4) noise reduction level 4 (Paper II)</td>
</tr>
<tr>
<td>IMR</td>
<td>Iterative model reconstruction (Philips)</td>
</tr>
<tr>
<td>IR</td>
<td>Iterative reconstruction</td>
</tr>
<tr>
<td>IRIS</td>
<td>Iterative reconstruction in image space (Siemens)</td>
</tr>
<tr>
<td>kV</td>
<td>Kilovolt (tube voltage)</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>LD</td>
<td>Low dose (Paper I)</td>
</tr>
<tr>
<td>mA</td>
<td>Milliampere (tube current)</td>
</tr>
<tr>
<td>mAs</td>
<td>Milliampere second (tube charge, radiation quantity)</td>
</tr>
<tr>
<td>mGy</td>
<td>Milligray (absorbed radiation dose)</td>
</tr>
<tr>
<td>mSv</td>
<td>Millisievert (equivalent radiation dose)</td>
</tr>
<tr>
<td>MDCT</td>
<td>Multi-detector computed tomography</td>
</tr>
<tr>
<td>MPR</td>
<td>Multi-planar reconstruction</td>
</tr>
<tr>
<td>MTF</td>
<td>Modulation transfer function</td>
</tr>
<tr>
<td>ND</td>
<td>Normal dose (Papers I–III)</td>
</tr>
<tr>
<td>NPS</td>
<td>Noise-power spectrum</td>
</tr>
<tr>
<td>RD</td>
<td>Reduced dose (Papers II, III)</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristics</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>SAFIRE</td>
<td>Sinogram affirmed iterative reconstruction (Siemens)</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>VGA</td>
<td>Visual grading analysis</td>
</tr>
</tbody>
</table>
Original papers

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals. The complete papers are appended at the end of the theses. Unpublished paper IV is excluded from the e-published version of this thesis. Reprints were made with permission from the respective publishers.

---


*Radiation dose reduction in CT of the brain: Can advanced noise filtering compensate for loss of image quality?*


---


*Hybrid iterative reconstruction algorithm in brain CT – A radiation dose reduction and image quality assessment study*

Acta radiologica (accepted 2013 May).

---


*Hybrid iterative reconstruction algorithm improves image quality in craniocervical CT angiography*

American journal of roentgenology (accepted 2013 March).

---


*Six iterative reconstruction algorithms in brain CT – A phantom study on image quality at different radiation doses*

In manuscript.

---
Introduction

Computed tomography (CT) is an important diagnostic tool in neuroradiology. It uses ionizing radiation to produce images of the subject for diagnostic purposes. Continuously increasing utilization of clinical CT imaging has raised concerns over the risks associated with increased population radiation exposure.1-4

The main limitation for radiation dose reduction in CT is the resulting degradation of image quality due to increased noise and reduced spatial resolution.5, 6 Therefore, maintenance of diagnostic image quality while reducing radiation exposure is a major technical challenge. In recent years, various technological advances have been introduced to address the problem. In the beginning, emphasis was mainly on hardware improvement, but in the last few years focus has shifted towards advanced software solutions.

Post-processing image filters (Paper I) and iterative reconstruction (IR) algorithms (Papers II–IV) are two software innovations directly targeting CT image quality. These operate at different stages in the image generation process, with the common objective of reducing noise while maintaining structural information. Both solutions have been shown to improve image quality and allow for radiation dose reduction in CT.7-13

In diagnostic radiology, implementation of new methods in clinical routine is usually preceded by validation in phantoms and clinical feasibility studies. Typically, such evaluation involves both objective and subjective assessment of diagnostic image quality. Subjective assessment is particularly challenging as it introduces various biases – systematic errors from the truth – that can affect study results.14, 15 Knowledge about these biases is important as they can be counteracted with robust study design. Furthermore, the choice of evaluation scales and study groups deserves careful consideration.

This thesis explores the potential of various software innovations to improve image quality in brain CT and craniocervical CT angiography (CTA), and the potential for
radiation dose reduction, in both clinical data and in phantoms. In addition, various aspects of image quality assessment and study design are explored.
Background

Computed tomography

Computed tomography is widely used in diagnostic imaging. Brain CT and craniocervical CT angiography are fundamental examinations in diagnostic neuroradiology and important first-line diagnostic tools in many clinical situations, e.g. in acute stroke.\textsuperscript{16, 17}

Technique

In CT, a rotating source of X-rays (tube) is used to acquire volumetric images of the subject. During the scan the tube continually rotates around the subject while a detector on the opposite side records the remaining beam intensity – how much the subject attenuates the X-rays – at different angles (projections). Attenuation is expressed in Hounsfield units (HU).

\textit{Multi-detector CT} (MDCT) uses an array of detector rows, permitting simultaneous acquisition of data from multiple parallel slices, thus reducing examination time. Current mainstream CT systems allow simultaneous acquisition of 64 to 128 slices, with up to 512-slice systems being available. \textit{Beam collimation} is a product of the number of detector rows and the detector row width. In a 64-slice system with a detector row width of 0.625 mm the collimation is 64 x 0.625 mm, resulting in a total coverage of 4 cm. A disadvantage of MDCT is slightly reduced image quality with increasing number of parallel slices due to cone beam artifacts and scattered radiation.\textsuperscript{18}

CT scans can be run in different modes. In \textit{axial (sequential) mode} the subject is fixed while the X-ray source and detector rotate around it in the x-y plane. In \textit{spiral (helical) mode} the subject simultaneously moves along the z-axis, perpendicular to the plane of rotation of the X-ray source and detector. In axial mode, image quality is marginally better, but acquisition time is increased in scans longer than the detector width.\textsuperscript{19} Speed and continuity of image data acquisition are the main benefits of spiral mode, a
benefit generally considered to outweigh the slight associated reduction in image quality.\textsuperscript{19} To allow enough attenuation data to be gathered for image reconstruction, helical CT scans need to complete at least an extra half rotation before the start and beyond the end of the imaged volume.\textsuperscript{20} Such over-ranging (or over-scanning) is an important source of increased radiation dose in spiral MDCT compared with axial CT, particularly at shorter scan lengths, and increases with collimation and pitch.\textsuperscript{21, 22} In some recent wide detector scanner models this problem has been addressed with dynamic z-axis collimation.\textsuperscript{23}

Radiation dose

The main disadvantage of CT is the use of ionizing radiation, with the associated risk of tissue damage and cancer induction. In CT, radiation dose depends on tube current (mA), slice scan time (s) and peak tube voltage (kV).\textsuperscript{24} Image noise is inversely proportional to radiation dose. According to the Poisson distribution, increase in noise associated with radiation dose reduction equals $\sqrt{\text{mAs}_{\text{original}}/\text{mAs}_{\text{reduced}}}$.\textsuperscript{6} Therefore, 30\% radiation dose reduction is expected to increase image noise by 20\% ($\sqrt{100/70} = 1.20$), thus reducing image quality.

Image noise primarily affects the ability of the CT system to reproduce two adjacent objects with similar CT numbers as separate structures (i.e. low-contrast resolution).\textsuperscript{5, 25} In brain CT, the attenuation difference between gray and white matter is small (~7 HU), requiring low noise levels and high radiation doses to adequately reproduce internal brain anatomy. In chest CT, however, noise is less of a problem since the attenuation difference between vessels or bronchial walls and the surrounding parenchyma is large (~850 HU), allowing for much higher noise levels (and thus lower radiation doses).\textsuperscript{5}

In clinical CT imaging, radiation dose levels should be adjusted according to the anatomical region being examined and the diagnostic task. Basic understanding of CT dosimetry is a prerequisite for successful radiation dose optimization.

The fundamental dosimetric quantity in spiral CT is the computed tomography dose index by volume (CTDI\textsubscript{vol}), expressed in milligrays (mGy). It represents the scanner radiation output, and is directly proportional to tube current in milliamperes (mA).\textsuperscript{26, 27} The CTDI\textsubscript{vol} is calibrated with standardized 16 cm (head) or 32 cm (body) phantoms, and is not a direct measure of dose to the particular subject.\textsuperscript{24, 28} Patient dose is the dose (mGy) absorbed by the particular subject and is dependent on CTDI\textsubscript{vol} and both subject volume and composition.\textsuperscript{28} For subjects smaller than the
calibration phantom, patient dose is higher than indicated by the CTDI<sub>vol</sub>, and vice versa. Total radiation exposure of a CT examination is defined by the dose-length product (DLP = CTDI<sub>vol</sub> x scan length in centimeters).<sup>6</sup> Radiosensitivity indicates the relative susceptibility of tissues to the harmful effects of ionizing radiation. Effective dose equivalent, the overall estimated risk associated with radiation exposure, can be estimated using conversion factors (E<sub>DLP</sub>) reflecting radiosensitivity of the exposed tissues or body regions, and is expressed in millisievert (mSv = DLP x E<sub>DLP</sub>).<sup>29</sup>

**Image reconstruction**

Image contrast is determined by differences in attenuation within the examined volume.<sup>24</sup> During the CT image reconstruction process, attenuation data from a large number of projections (projection domain) is mathematically processed to create an image of the examined volume (image domain). Conventionally this is achieved through a fast mathematical procedure called filtered back-projection (FBP).<sup>6</sup>

Prior to reconstruction, the projection data is filtered to counteract blurring, and to achieve appropriate balance between detail resolution and noise for the diagnostic task.<sup>6</sup> Back-projection involves reversal of the attenuation measurement process where data from thousands of projections is used to reconstruct an image of the subject.

To ensure fast reconstruction, the FBP algorithm assumes an ideal system.<sup>30</sup> All measurements are treated equally and only processed once.<sup>6</sup> This is usually not an issue in CT examinations with standard radiation dose levels, where the signal is much higher than the noise. The limitations of FBP are revealed in low radiation dose acquisitions where image quality is compromised by disproportionally high levels of noise and image artifacts. In recent years, these limitations have been addressed with the introduction of iterative reconstruction (IR) algorithms – discussed further on.

**Parameters of image quality**

The goal of clinical CT imaging is to produce images of diagnostic quality using the lowest possible radiation exposure – often referred to as the ALARA-principle (as low as reasonably achievable).<sup>5, 31, 32</sup> This implies adequate reproduction of clinically important anatomical structures and pathological processes.
There are four fundamental determinants of image quality in CT systems:24

- **Low-contrast resolution** is the ability of the system to reproduce two adjacent objects with similar CT numbers as separate structures. Low contrast resolution is primarily increased with decreased image noise.5, 24

- **Spatial resolution** (high contrast resolution) is the ability of the system to resolve image detail. Spatial resolution is determined by object-to-detector distance, focal spot size, detector size, reconstruction matrix resolution, and slice thickness.5

- **Noise** represents the local statistical fluctuation in CT numbers of individual picture elements.24 Noise decreases with increased radiation exposure (mAs, kV) and is also affected by slice thickness and reconstruction algorithm.5 Noise power spectrum (NPS) is a measure of noise power as a function of spatial frequency (i.e. noise distribution),33 and characterizes noise texture.

- **Image artifact** is defined as a systematic discrepancy between CT numbers in the reconstructed image and true attenuation coefficients of the examined volume. Artifacts can be based on physics (e.g. beam hardening, photon starvation, partial volume), the patient (e.g. metal and motion artifacts), or the CT system (e.g. ring and distortion artifacts).34

**Image quality improvement**

Image quality in CT is a result of interaction between many factors. The optimal balance between image quality and radiation dose is achieved by adjusting scan parameters and the image reconstruction process to fit diagnostic requirements, with respect to clinical indication and the anatomical region in question.

Radiation dose is a major determinant of image quality. Steadily increasing utilization of CT in clinical practice has raised concerns regarding the risks associated with radiation exposure.1-4 In recent years various techniques have been developed that allow radiation dose to be reduced while maintaining diagnostic image quality. Two of these innovations directly target image quality:

- **Post-processing image filters** 35, 36 optimize images that have already been reconstructed with FBP. They use adaptive noise suppression and edge enhancement to reduce noise while maintaining structural information,
resulting in improved image quality. As they operate in the *image domain* only, post-processing filters are unable to retrieve image information already lost during the image reconstruction process. This may be a limitation of the method, especially in combination with aggressive dose reduction, where photon starvation artifacts and high level of noise predominate. Post-processing filters have both been available as proprietary solutions incorporated into a CT systems (e.g. IRIS, Siemens Healthcare, Forchheim, Germany), and as separate vendor independent modules (e.g. SharpView CT, ContextVision AB, Linköping, Sweden).

- **Iterative reconstruction algorithms** are the latest advance in CT technology and come in two basic designs: 
  a) *Statistical iterative optimization*, based on photon statistics, assuming ideal system, and 
  b) *Model-based iterative optimization*, that additionally attempts to model the system and the acquisition process, including system optics. These algorithms perform iterative image optimization, not only in the *image domain*, but also in the *projection domain*. Optimization of projection data prevents noise and artifacts in the projection domain from propagating into the image domain where they might be more difficult to remove. As a result, the IR algorithms can better deal with low signal levels in modern low dose acquisitions, and will ultimately replace conventional FBP.

Both post-processing image filters and IR algorithms have been shown to improve image quality in CT, and allow for considerable radiation dose reduction while maintaining diagnostic image quality.

**Evaluation of image quality**

Optimization of CT image quality requires reliable methods for evaluation of the resulting images to ensure adequate diagnostic quality. Although many key image quality parameters can be directly measured, appreciation of human readers is also important. Ideally, the evaluation process should therefore consist of objective and subjective evaluation steps.

**Objective evaluation**

A variety of objective image quality parameters can be measured in clinical CT images.
- **CT numbers** (HU) are ideally measured in a homogeneous area in a phantom or an organ, to prevent structural information from contributing to the measurements. Although exact absolute CT numbers are infrequently used for diagnostic purposes, they can be important, such as for characterization of incidental adrenal masses.

- **Noise** in CT is typically expressed as standard deviation (SD) of the CT numbers in a ROI, assuming normal distribution. Using SD for direct comparison of noise levels between FBP and extensively manipulated IR may be inappropriate if noise distribution differs (Paper IV). NPS provides information about spatial characteristics (texture) of noise and is assessed in homogeneous phantoms using advanced mathematics. It is ideal for comparison of noise distributions in different reconstructions.

- **Signal-to-noise ratio** can be expressed as signal (HU) divided by noise (SD) in the same homogenous ROI (SNR = HU / SD). High SNR indicates that true information (signal) overpowers noise.

- **Low-contrast resolution** represents the ability of the system to reproduce two adjacent objects with similar CT numbers as separate structures. It can be subjectively evaluated as the smallest discernible object in an image, with specific difference in contrast relative to the adjacent background. Low contrast resolution is highly dependent on image noise, and can be objectively assessed by the contrast-to-noise ratio (CNR), where the contrast between two structures is expressed as a function of noise. Examples of formulas for calculation of CNR between two structures (A and B) in CT images are:

  a) \( \text{CNR} = \frac{(\text{HU}_A - \text{HU}_B)}{[(\text{SD}_A^2 + \text{SD}_B^2)/2]^{0.5}} \), and

  b) \( \text{CNR} = \frac{(\text{HU}_A - \text{HU}_B)}{\text{SD}_B} \).

  The key difference between these is that the first formula takes into account noise levels (SD) in both structures, while the other only includes noise in one of the structures. If noise levels in both structures are identical – as is common in evaluation of low-contrast resolution – the formulas will return identical results. If there are considerable differences in noise levels between the structures (e.g. large attenuation differences), the first formula might produce more accurate absolute CNR values. However, in studies on CT
image quality the result of interest is typically the relative difference in CNR between two methods being compared, and not absolute CNR values.

- **Spatial resolution** (high contrast resolution) represents the ability of the system to resolve image detail. To eliminate interference from image noise, spatial resolution is usually determined by using test objects with large differences in CT numbers in a phantom. Spatial resolution can be assessed subjectively by visualization of line pairs of increasing density. Objectively it can be assessed by the *modulation transfer function* (MTF) which involves complex calculations based on the degree of sharpness observed in the image of a specific test object, typically a high density bead or wire.

**Subjective evaluation**

The ultimate goal of clinical CT imaging is to reproduce anatomical and pathological information for diagnostic purposes – not to produce technically flawless or aesthetically appealing images. The diagnostic process culminates in human interpretation of images and to simulate this, a subjective evaluation step is important in studies on diagnostic image quality.

There are two principal approaches to subjective evaluation of diagnostic image quality:

- **Diagnostic acceptability** is an indirect measure of diagnostic performance. In *visual grading analysis* (VGA), image criteria consisting of clinically important anatomical structures are chosen and visually graded using an ordinal scale. The underlying theory is that the ability to detect pathology correlates well with accurate reproduction of anatomy. Strictly speaking, ordinal grading data from VGA cannot be converted to numerical data, limiting the statistical evaluation to non-parametric rank-invariant methods that can be difficult to comprehend. It may, however, be argued that ordinal data can be treated as normally distributed interval variables, especially in large data sets, and this is widely accepted in medical literature. The main benefit of the VGA approach is that it is not dependent on specific pathological processes or diagnoses, thereby simplifying the study design.

- **Diagnostic accuracy** is a direct measurement of diagnostic performance. It involves detection of abnormal cases against a background of normal cases. Typically, observers are asked to state if the examination is normal or
abnormal and to define their confidence level. The results are presented with receiver operating characteristics (ROC) curves where the true-positive fraction is plotted against the false-positive fraction. Different methods can be compared by measuring the area under the curve. However, assigning interval values to the area under the curve is equally debatable here as in VGA since the original variables are ordinal data.57, 59

Although it can be argued that diagnostic accuracy is a better measure of diagnostic image quality, such studies are more difficult to carry out as they require a large number of subjects with a balanced mix of abnormal and normal cases where the true status of each case is known. Furthermore, patient recruitment with informed consent may be problematic in important applications, such as in acute ischemic stroke, due to the risk of treatment delay. Because of this and due to the rapid technical advances in CT imaging a fast and effective method for subjective evaluation of image quality is favorable, explaining the preference for diagnostic acceptability above diagnostic accuracy in many recent studies.

Observer bias

Subjective evaluation of image quality is dependent on human observers. It can be argued that subjective evaluation of image quality is less reliable than objective measurements due to the introduction of various sources of observer bias.14, 15, 60, 61 Nevertheless, subjective assessment is difficult to omit in studies on diagnostic image quality. Knowledge about the most important sources of observer biases is therefore important, as many of these can be counteracted with careful study design, resulting in more reliable study results.

Different subtypes of observer bias are important in the evaluation of image quality:

- **Adaptation bias** occurs when observers accustomed to a certain image appearance and noise texture prefer these images over clinically equivalent images with slightly different appearance. This bias can cause reduced acceptability during the introduction of new methods or equipment that affects image appearance.

- **Recognition bias** becomes evident when observers who are intended to be blinded for an evaluation variable (e.g. a new method or equipment) nevertheless can, or believe that they can, identify which reconstruction they are looking at. Trained radiologists may be able to recognize even the
slightest deviations from usual image characteristics, making complete blinding of observers very difficult to achieve.

- Confirmation bias can arise when observers themselves are responsible for the study design, allowing them to choose study variables to suit their personal preference. Individually tailored study variables serve to maximize the chance for positive results and subsequent publication, but are not necessarily representative for radiologists in general.

Substantial observer variation highlights the fact that in order to produce representative results, studies on subjective image quality not only require a sufficient number of cases, but also an adequate number of observers.

Many recent studies on image quality and radiation dose include only two observers, although it can be assumed that observer diversity, caused by disparate visual and cognitive abilities, will lead to individual differences in interpretation. There is no simple answer to the question how many observers are appropriate, but in general the more observers the better. An article by Obuchowski provides some background information and suggestions based on different study phases. Furthermore, to avoid confirmation bias, it is strongly suggested to keep the number of observers directly involved in the study design at a minimum.

**Evaluation scales**

Subjective evaluation of image quality involves grading using *ordinal scales* where the order of grades is defined, but the degree of difference between them is not. The type of ordinal scale used for subjective evaluation of image quality deserves some consideration as it can affect the study results (Paper II).

**Ranking of image quality**

Ranking of image quality involves a simple form of ordinal scale. Observers are asked to do side-by-side ranking of image quality from best to worst according to image quality criteria (Papers I, II). The ranking steps are only relatively defined and do not include a description of the expected image quality. To compensate for this, one of the ranked images can be used as an internal reference. The internal reference usually represents the image quality produced by the current method, allowing rank comparison with the new method.
The main disadvantage of ranking scales is that observers are usually forced to apply different ranks to images, even in cases where there are no differences in image quality. Furthermore, the size of each ranking step is variable, both between steps and between individual assessments, making it impossible to quantify the difference in image quality between the assigned ranks. Yet another disadvantage of the ranking procedure is that side-by-side presentation unavoidably facilitates recognition bias that may affect the results (Paper II).

**Grading of image quality**

Grading of image quality involves a more sophisticated type of ordinal scales. Observers are typically asked to grade individual images according to image quality criteria (Papers II, III). Each grading step is defined using an absolute description (instead of only relative as in ranking), preferably complemented with an example image to improve consensus between observers. Grading steps can be defined to reflect an approximately linear improvement in image quality, allowing some degree of quantification of differences. Compared with ranking, the main benefits of grading are that the scale is described in absolute terms, allowing images with similar image quality to be assigned the same grade, and that images are reviewed separately (not side-by-side), reducing the risk for recognition and adaptation bias.

**Grading scales**

Grading scales should be tailored for the study purpose. Ideally, in clinical studies on subjective image quality the number and size of the steps should reflect clinically relevant differences. The number of steps in the grading scale is a balance between resolution (small steps) and reproducibility (ease of use). In smaller studies, detection of small differences in image quality might require scales with high resolution. However, scales with many small steps can be difficult to apply in a consistent manner.

It is common for studies on image quality to employ five-step scales. However, through clinical correlation such grading scales can be reduced to only four steps, making them easier to use without significant loss in resolution:

- **Grade 1**: *Non-diagnostic* image quality.
- **Grade 2**: *Sufficient* image quality, some diagnostic limitations.
- **Grade 3**: *Standard* image quality, no significant diagnostic limitations.
Grade 4: Excellent image quality, no diagnostic limitations.

The goal of CT optimization is to achieve image quality without significant diagnostic limitations using the smallest possible radiation dose (ALARA-principle), corresponding to grade 3 in the scale example above. Due to patient diversity and technical limitations this is impossible to achieve for all patients, and therefore more scale steps are required.

To reflect basic diagnostic requirements it is important that the scale allows clear discrimination between diagnostic (grades 2–4) and non-diagnostic (grade 1) examinations. For clinical purposes it is usually not important to know how poor the non-diagnostic examinations are, eliminating the need for further grades at the lower end of the scale. Likewise, grade 4 represents image quality better than what is clinically required, implying usage of unnecessarily high radiation dose that should be lowered. Therefore, further grades at the higher end of the scale are unnecessary.

Results based on such four-grade scales in CT are easy to interpret. First it should be ensured that no examinations are graded 1 (non-diagnostic). Then, as many of the examinations as possible should be graded 3 (standard). Finally, the ratio between grade 2 (sufficient) and grade 4 (excellent) can be used to decide upon an adjustment of the radiation dose. Overrepresentation of examinations with grade 4 image quality would for example indicate a too high radiation dose that should be lowered.

In CT, too low or too high image quality is equally inappropriate.

Research subjects

Evaluation of new algorithms begins in phantoms where repetitive adjustments can be performed under standardized conditions allowing for exact comparison of results. In phantom studies, objective image quality can be automatically evaluated using special software, eliminating human influence with the accompanying variability and bias. Phantom studies are usually followed by clinical studies before the techniques are introduced in clinical routine. Such studies usually involve evaluation of diagnostic acceptability or diagnostic accuracy in a rather small population (<100).

Studies on diagnostic image quality should preferably be dimensioned to be able to detect clinically significant differences. Detection of small intervention effects, such as the effect of different image reconstruction methods on image quality, against a background of relatively large inter-individual variation (i.e. anatomical differences)
requires an adequate number of cases and good matching between cases and controls. Better matching allows for fewer cases to be used. Ideally this involves having identical case and control populations with each case serving as their own control. Using the same case and control population when evaluating different radiation doses can involve increased radiation exposure to the research subjects. The associated risk is dependent on radiation dose, radiosensitivity of the exposed tissues, and age of the subjects. These risk factors can be adjusted accordingly to minimize the resulting risk, for example by only including elderly research subjects (Papers I, II).
Aims

The main objectives of this thesis were to assess new methods for improving image quality in CT, explore the associated potential for radiation dose reduction, and to outline a comprehensive approach for evaluation of image quality.

Specific objectives of the individual papers were as follows:

Paper I

- To evaluate the effect of post-processing image filtering (SharpView CT) on image quality in brain CT acquired with 30% reduced radiation dose.

Paper II

- To evaluate the effect of statistical iterative reconstruction (iDOSE⁴) on image quality in brain CT acquired with 30% reduced radiation dose.
- To illustrate and discuss methodological pitfalls and strengths associated with the methods used for subjective evaluation of image quality.
- To improve on methodological limitations of paper I.

Paper III

- To evaluate the potential of statistical iterative reconstruction (iDOSE⁴) for improvement of image quality in craniocervical CT angiography.
- To assess inter-observer and intra-observer performance.
- To further improve on methodological limitations of papers I and II.
Paper IV

- To evaluate image quality produced by six iterative reconstruction algorithms (AIDR, ASIR, iDOSE\textsuperscript{4}, IMR, SAFIRE, Veo) in a brain CT phantom model, using different radiation dose levels and iterative image optimization levels.
Subjects and methods

Study subjects

The thesis includes a total of 100 human subjects in two human studies (Papers I, III) and one combined human and phantom study (Paper II). The fourth study was a phantom study. Human subjects were recruited amongst patients referred to the examinations on clinical indications not related to the studies. As the two first studies (Papers I, II) involved an additional CT examination leading to increased radiation exposure, informed consent was required. This requirement was waived for the third study by the ethical committee. An overview of study subjects is presented below (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Paper I</th>
<th>Paper II</th>
<th>Paper III</th>
<th>Paper IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examined organ</td>
<td>Brain</td>
<td>Brain</td>
<td>Head/neck</td>
<td>–</td>
</tr>
<tr>
<td>Number of cases/controls(^a)</td>
<td>30</td>
<td>40</td>
<td>30</td>
<td>–</td>
</tr>
<tr>
<td>Subject age range (years)</td>
<td>65–92</td>
<td>65–93</td>
<td>16–80</td>
<td>–</td>
</tr>
<tr>
<td>Median/Mean age (years)</td>
<td>79 / 78</td>
<td>76 / 77</td>
<td>63 / 58</td>
<td>–</td>
</tr>
<tr>
<td>Examinations per subject</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Radiation protection approval</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Research ethics approval</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Informed consent</td>
<td>Yes</td>
<td>Yes</td>
<td>Waived</td>
<td>–</td>
</tr>
<tr>
<td>Phantom examinations</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 1 – Overview of study subjects.

\(^a\) In each of the clinical studies (Papers I–III), each patient served as his/her own control.
One of the studies (Paper III) was a sheer image quality evaluation study, while the remaining three (Papers I, II, IV) combined evaluation of image quality with radiation dose reduction. The baseline CT protocols (Table 2) used for the clinical studies (Papers I–III) were based on the clinical routine protocols in use at our clinic at that time. For the phantom study (Paper IV), the reference radiation dose was adjusted to match the European guidelines on quality criteria for CT, while all other examination parameters were chosen according to recommendations for adult brain CT published by the American Association of Physicists in Medicine (AAPM). In studies evaluating the effect of different radiation dose levels (Papers I, II, and IV), 30–90% dose reduction was accomplished by adjusting tube charge (mAs).

<table>
<thead>
<tr>
<th>CT system</th>
<th>Paper I</th>
<th>Paper II</th>
<th>Paper III</th>
<th>Paper IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT system</td>
<td>Philips</td>
<td>Philips</td>
<td>Philips</td>
<td>GE, Philips, Siemens, Toshiba</td>
</tr>
<tr>
<td>Tube voltage (kV)</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>Tube current (mAs)</td>
<td>355 248</td>
<td>498 349</td>
<td>121 b</td>
<td>Variable c</td>
</tr>
<tr>
<td>CTDIvol (mGy) (% of baseline dose)</td>
<td>57 (100%) 40 (70%)</td>
<td>57 (100%) 40 (70%)</td>
<td>6.8 b (–)</td>
<td>120 4 (100%) 84 (70%) 48 (40%) 12 (10%)</td>
</tr>
<tr>
<td>Slice collimation (mm)</td>
<td>16 x 0.625</td>
<td>64 x 0.625</td>
<td>64 x 0.625</td>
<td>Variable c</td>
</tr>
<tr>
<td>Pitch</td>
<td>0.683</td>
<td>0.578</td>
<td>0.98 b</td>
<td>Variable c</td>
</tr>
<tr>
<td>Rotation time (s)</td>
<td>0.75</td>
<td>0.75</td>
<td>0.50 b</td>
<td>Variable c</td>
</tr>
<tr>
<td>Automatic dose modulation</td>
<td>Disabled</td>
<td>Disabled</td>
<td>Enabled</td>
<td>Disabled</td>
</tr>
<tr>
<td>Intravenous contrast medium</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2 – Overview of CT examination parameters used in the individual papers.

a Provided parameters are valid for the clinical study (not the phantom experiments).
b Values represent means since automatic dose modulation was enabled.
Values adapted from the recommendations for adult brain CT published by the *American Association of Physicists in Medicine (AAPM).*

The baseline dose of 120 mGy for the 20 cm image quality phantom was chosen to roughly equal 60 mGy in a standard 16 cm phantom.¹⁹
Image reconstruction

FBP reconstruction was used as baseline for all four studies (Table 3). In one of the studies (Paper I) post-processing filtering was used for improving image quality. IR algorithms were utilized for the three remaining studies (Papers II–IV), using up to three different IR levels. In all other aspects, preparation of images for subjective and objective evaluation followed clinical routine for brain CT or craniocervical CTA, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Paper I</th>
<th>Paper II</th>
<th>Paper III</th>
<th>Paper IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBP reconstruction</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Post-processing filter</td>
<td>SharpView CT</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>IR algorithms</td>
<td>–</td>
<td>iDOSE⁴</td>
<td>iDOSE⁴</td>
<td>AIDR, ASIR, iDOSE⁴, IMR SAFIRE, Veo</td>
</tr>
<tr>
<td>IR levels (n)</td>
<td>–</td>
<td>2</td>
<td>1</td>
<td>1–3</td>
</tr>
<tr>
<td>Slice thickness</td>
<td>5 mm</td>
<td>5 mm</td>
<td>3 mm</td>
<td>5 mm</td>
</tr>
<tr>
<td>MPR-mode</td>
<td>Average</td>
<td>Average</td>
<td>MIP</td>
<td>Average</td>
</tr>
</tbody>
</table>

*The IR algorithms IMR and Veo were evaluated using one IR level while AIDR, ASIR, iDOSE⁴ and SAFIRE were assessed using three different IR levels.

AIDR = Adaptive Iterative Dose Reduction 3D.
ASIR = Adaptive statistical iterative reconstruction
iDOSE⁴ = (product name, not acronym)
IMR = Iterative model reconstruction
SAFIRE = Sinogram affirmed iterative reconstruction
Veo = (product name, not acronym)
Image quality assessment

A variety of evaluation methods were employed in the studies (Table 4). The phantom studies (Papers II, IV) emphasized on objective evaluation of image quality, while clinical studies (Papers I–III) focused more on subjective evaluations.

<table>
<thead>
<tr>
<th>Objective methods</th>
<th>Paper I</th>
<th>Paper II</th>
<th>Paper III</th>
<th>Paper IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT numbers (HU)</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Noise (SD)</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Signal-to-noise ratio (SNR)</td>
<td></td>
<td>•</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contrast-to-noise ratio (CNR)</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>Noise-power spectrum (NPS)</td>
<td>•</td>
<td></td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>Spatial resolution (MTF)</td>
<td></td>
<td>•</td>
<td></td>
<td>•</td>
</tr>
</tbody>
</table>

| Subjective methods                |         |          |           |          |
| Number of observers               | 5       | 6        | 5         | 5        |
| Number of quality criteria        | 3       | 3        | 8         | 2        |
| Low-contrast resolution           | •       | •        |           |          |
| Spatial resolution                |         |          |           | •        |
| Ranking of quality criteria       | •       | •        |           |          |
| Grading of quality criteria       | •       | •        |           |          |
| Inter-observer agreement          | •       | •        |           |          |
| Intra-observer agreement          |         |          |           | •        |

Table 4 – Overview of subjective and objective image quality assessment methods used in the different studies.

In all four studies, multiple independent observers were employed for subjective evaluation of carefully chosen clinically important image quality criteria. Images were presented in randomized order under standardized viewing conditions using identical workstations and medical grade display monitors and standardized window level/width. In the clinical studies (Papers I–III), the radiologists were free to adjust window level/width and zoom level, simulating standard clinical reading process.
Clinical images were anonymized and information on dose level and reconstruction type was masked.

In the brain CT studies (Papers I, II) subjective evaluation focused on low-contrast resolution, employing quality criteria sensitive for this: gray-white matter discrimination and delineation of basal ganglia structures. The subjective evaluation in the contrast-enhanced craniocervical CTA study (Paper III) focused on spatial resolution through evaluation of delineation of vessels in multiple locations. In addition, the studies all included subjective evaluation of general image quality, reflecting the over-all image impression.

Two types of subjective evaluation scales – ranking and grading – were used in the clinical studies (Papers I–III) and in one of these the two methods were compared (Paper II). The number of ranking steps was three (Paper I) or four (Paper II), dependent on the number of images being compared. The grading scales used in Papers II and III both included four steps, adhering to the following basic design:

- Grade 1: Poor (non-diagnostic).
- Grade 2: Fair (sufficient).
- Grade 3: Good (standard).
- Grade 4: Excellent (better than required).

In the phantom study (Paper IV) the subjective evaluation of low-contrast resolution was done in consensus between five observers, using two criteria: smallest discernible cylinder, and smallest sharply defined cylinder.

Finally, observer agreement was evaluated in two clinical studies (Papers II, III).
Statistical evaluation

In all four studies descriptive statistics were used, both for evaluation of subjective and objective data (Table 5). The subjective ranking and grading procedures in Papers I–III produce ordinal data that preferably should be statistically evaluated with non-parametric rank-invariant methods such as two-tailed sign test. This is especially true for forced ranking procedures (Papers I, II), where the sizes of the ranking steps can vary widely. In grading procedures (Papers II, III), the combination of scales reflecting approximately linear improvement in image quality and a large number of observations, allows the resulting ordinal data to be treated as normally distributed interval variables. This permits the use of traditional statistical methods in the form of descriptive statistics and linear mixed-models. In addition, grading data were presented using cumulative distribution graphs where the ordinal nature of the data was preserved. In Paper IV each automated measurement was performed only once and the subjective evaluation done in consensus. Therefore, data could be adequately presented using descriptive statistics only.

<table>
<thead>
<tr>
<th></th>
<th>Paper I</th>
<th>Paper II</th>
<th>Paper III</th>
<th>Paper IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descriptive statistics</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Linear mixed-model</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><strong>Subjective methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descriptive statistics</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Two-tailed sign test</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Percent agreement</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Cumulative distribution</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Linear mixed-model</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

**Table 5** – Overview of statistical methods used in the different papers.

- a Used in evaluation of differences in mean attenuation and noise.
- b Used for statistical evaluation or ranking data (non-parametric rank-invariant).
- c Here only used for supplementing evaluation of ranking data to illustrate potential consequences of observer variability in combination with only two observers.
- d Used for assessment of complete observer agreement.
- e Used in evaluation of grade distribution between algorithms within observer.
- f Used for differences in mean grade, including $p$-values and confidence intervals.
Results

Paper I

The study evaluated the combination of 30% radiation dose reduction and post-processing image filtering (SharpView CT) in brain CT (Figure 1).

In pooled data (Table 6), image quality was significantly improved in filtered low dose (FLD) compared with low dose (LD), for all three subjective quality criteria. Objective noise was reduced with about 15%, and thereby in parity with normal dose (ND). Subjectively, ND was still considered superior to FLD, although this difference was not statistically significant.

Considerable inter-observer variation was noted between the 5 observers.

<table>
<thead>
<tr>
<th>Image quality assessment</th>
<th>Normal dose (ND)</th>
<th>Filtered low dose (FLD)</th>
<th>Low dose (LD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean CT numbers (HU)</td>
<td>25.9</td>
<td>25.8</td>
<td>25.7</td>
</tr>
<tr>
<td>Mean noise, SD (HU)</td>
<td>3.0</td>
<td>3.0</td>
<td>3.4</td>
</tr>
<tr>
<td><strong>Subjective methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean rank, gray-white matter discrimination</td>
<td>1.78</td>
<td>1.82^b</td>
<td>2.40</td>
</tr>
<tr>
<td>Mean rank, delineation of basal ganglia</td>
<td>1.65</td>
<td>1.85^b</td>
<td>2.50</td>
</tr>
<tr>
<td>Mean rank, general image quality</td>
<td>1.65</td>
<td>1.98^b</td>
<td>2.37</td>
</tr>
</tbody>
</table>

Table 6 – Summary of results from Paper I. Compared with low dose (LD), image quality was improved in filtered low dose (FLD), although subjectively normal dose (ND) was still considered superior.

^a Lower rank indicates better image quality.

^b Significant difference compared with LD (P < 0.05).
Figure 1 – Typical reconstruction results from a single patient at the level of the centrum semiovale showing: (a) normal dose (ND), (b) filtered low dose (FLD), and (c) low dose images. FD and FLD have very similar appearances, while slight increase in noise can be seen in LD. CT images may not be adequately reproduced in print.
Paper II

The study evaluated the combination of 30% radiation dose reduction and iterative reconstruction (iDOSE<sup>4</sup>) in brain CT (Figure 2) and in a phantom.

Of the two iterative reconstruction levels, iDOSE<sup>4</sup> level 4 (ID4) was generally superior to iDOSE<sup>4</sup> level 2 (ID2) (Table 7). Objectively, image quality was significantly better in ID4 than in ND. Subjectively, this was reversed, with ND being both graded and ranked significantly better than ID4 (and ID2).

Inter-observer percent agreement for the 6 observers ranged from 50–70%.

<table>
<thead>
<tr>
<th>Image quality assessment</th>
<th>Normal dose (ND)</th>
<th>iDOSE&lt;sup&gt;4&lt;/sup&gt; level 2 (ID2)</th>
<th>iDOSE&lt;sup&gt;4&lt;/sup&gt; level 4 (ID4)</th>
<th>Reduced dose (RD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT numbers (HU)</td>
<td>25.6</td>
<td>25.5</td>
<td>25.5</td>
<td>25.5</td>
</tr>
<tr>
<td>Noise, SD (HU)</td>
<td>2.4</td>
<td>2.5&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.8&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>CNR precentral gyrus</td>
<td>3.5</td>
<td>3.6</td>
<td>3.9&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.2&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>CNR caudate nucleus</td>
<td>2.1</td>
<td>2.1</td>
<td>2.4&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.9&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>CNR lentiform nucleus</td>
<td>2.2</td>
<td>2.4</td>
<td>2.6&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.0&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Subjective methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean grade, pooled criteria&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.22</td>
<td>2.87&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.91&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.72&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mean rank, pooled criteria&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.6</td>
<td>2.4&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.7&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.3&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

**Table 7** – Summary of results of the clinical brain CT evaluation in Paper II. According to objective measures, image quality was best in iDOSE<sup>4</sup> level 4 (ID4) and iDOSE<sup>4</sup> level 2 (ID2). Subjectively, normal dose (ND) was considered superior to all others.

<sup>a</sup> Significant difference compared with ND ($P < 0.05$).
<sup>b</sup> Higher grade indicates better image quality.
<sup>c</sup> Lower rank indicates better image quality.

In the phantom evaluation, the IR algorithm reduced noise (SD) and improved SNR in reduced dose acquisitions, while maintaining spatial resolution (MTF) unchanged. Noise-power spectra were practically identical in IR and FBP.
Figure 2 – Typical reconstruction results from a single patient at the level of centrum semiovale showing: (a) normal dose (ND), (b) iDOSE^4 level 2 (ID2), (c) iDOSE^4 level 4, and (d) reduced dose (RD) images. *CT images may not be adequately reproduced in print.*
Paper III

The study evaluated the potential of iterative reconstruction (iDOSE\textsuperscript{4}) to improve image quality in craniocervical CT angiography (Figure 3).

As shown in Table 8, image quality was significantly better in IR than FBP for all criteria, both objective and subjective. Greatest subjective and objective improvement was seen at the level of the vertebral arteries. After excluding one patient that swallowed during the examination, the number of arterial segments receiving “poor” rating was reduced from 19 with FBP to 0 with IR.

Overall the intra-observer agreement was higher (mean 79%, range 69–88%) than the inter-observer agreement (mean 66%, range 59–75%).

<table>
<thead>
<tr>
<th>Image quality assessment</th>
<th>FBP</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective methods</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNR, vertebral arteries</td>
<td>2.8</td>
<td>4.8  \textsuperscript{a}</td>
</tr>
<tr>
<td>CNR, carotid bulbs</td>
<td>6.3</td>
<td>6.9  \textsuperscript{a}</td>
</tr>
<tr>
<td>CNR, basilar artery</td>
<td>5.7</td>
<td>6.5  \textsuperscript{a}</td>
</tr>
<tr>
<td>CNR, middle cerebral artery</td>
<td>5.7</td>
<td>6.5  \textsuperscript{a}</td>
</tr>
<tr>
<td><strong>Subjective methods</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean grade, vertebral arteries\textsuperscript{b}</td>
<td>2.45</td>
<td>3.46 \textsuperscript{a}</td>
</tr>
<tr>
<td>Mean grade, carotid bulbs\textsuperscript{b}</td>
<td>3.40</td>
<td>3.56 \textsuperscript{a}</td>
</tr>
<tr>
<td>Mean grade, basilar artery\textsuperscript{b}</td>
<td>3.41</td>
<td>3.56 \textsuperscript{a}</td>
</tr>
<tr>
<td>Mean grade, middle cerebral artery\textsuperscript{b}</td>
<td>3.29</td>
<td>3.45 \textsuperscript{a}</td>
</tr>
<tr>
<td>Mean grade, general image quality\textsuperscript{b}</td>
<td>3.17</td>
<td>3.58 \textsuperscript{a}</td>
</tr>
</tbody>
</table>

**Table 8** – Summary of key results from Paper III. Image quality was significantly better in iterative reconstruction (IR) than filtered back-projection (FBP) for all criteria, both objective and subjective. Greatest subjective and objective improvement was seen at the level of the vertebral arteries.

\textsuperscript{a} Significant difference compared with FBP (P < 0.05).

\textsuperscript{b} Higher grade indicates better image quality.
Figure 3 – Typical coronal reconstruction results from a single patient. The improvement in image quality in (a) iterative reconstruction, compared with (b) filtered back-projection, is best seen at the level of the shoulders. CT images may not be adequately reproduced in print.
This study evaluated image quality produced by six different IR algorithms, and the effects of radiation dose reduction in a brain CT phantom model.

As seen in Table 9, the greatest noise reduction was achieved with Philips IMR and GE IR3. All GE and Philips algorithms manage to reduce noise while maintaining spatial resolution (MTF). The model-based algorithms (Philips IMR, GE Veo) reduce noise progressively with reduced radiation dose.

<table>
<thead>
<tr>
<th></th>
<th>120 mGy</th>
<th>84 mGy</th>
<th>48 mGy</th>
<th>12 mGy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SD %</td>
<td>SNR %</td>
<td>MTF %</td>
<td>SD %</td>
</tr>
<tr>
<td>GE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>IR1</td>
<td>86</td>
<td>87</td>
<td>100</td>
<td>87</td>
</tr>
<tr>
<td>IR2</td>
<td>62</td>
<td>117</td>
<td>103</td>
<td>63</td>
</tr>
<tr>
<td>IR3</td>
<td>42</td>
<td>170</td>
<td>106</td>
<td>43</td>
</tr>
<tr>
<td>Veo</td>
<td>97</td>
<td>69</td>
<td>136</td>
<td>98</td>
</tr>
<tr>
<td>Philips</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>IR1</td>
<td>97</td>
<td>104</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>IR2</td>
<td>79</td>
<td>127</td>
<td>101</td>
<td>77</td>
</tr>
<tr>
<td>IR3</td>
<td>66</td>
<td>152</td>
<td>97</td>
<td>63</td>
</tr>
<tr>
<td>IMR</td>
<td>69</td>
<td>141</td>
<td>119</td>
<td>59</td>
</tr>
<tr>
<td>Siemens</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>IR1</td>
<td>85</td>
<td>117</td>
<td>88</td>
<td>85</td>
</tr>
<tr>
<td>IR2</td>
<td>71</td>
<td>140</td>
<td>87</td>
<td>71</td>
</tr>
<tr>
<td>IR3</td>
<td>59</td>
<td>170</td>
<td>88</td>
<td>58</td>
</tr>
<tr>
<td>Toshiba</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>IR1</td>
<td>77</td>
<td>124</td>
<td>93</td>
<td>77</td>
</tr>
<tr>
<td>IR2</td>
<td>66</td>
<td>139</td>
<td>95</td>
<td>66</td>
</tr>
<tr>
<td>IR3</td>
<td>67</td>
<td>158</td>
<td>94</td>
<td>61</td>
</tr>
</tbody>
</table>

Table 9 – Results from objective measurements of noise (SD), signal-to-noise ratio (SNR) and MFT_{10\%} (MTF) presented as percentages relative to FBP. Values in blue indicate improvement; values in red indicate deterioration; and bold style the best result for the respective CT system/dose.
CNR was generally improved with increasing radiation dose and higher IR-level (Figure 4). Same trend was seen with subjective visual evaluation of cylinders in the low-contrast module (Figure 5).

Visual evaluation of NPS curves revealed almost identical curve shapes (noise texture) between Philips FBP and IR, while they were very dissimilar for both model-based IR algorithms (Philips IMR, GE Veo) indicating different noise textures (Figure 6).

**Figure 4** – Cumulative contrast-to-noise (CNR) ratios for all combinations of radiation doses and reconstruction algorithms from objective measurements at three nominal contrast levels: 1.0% (dark blue), 0.5% (blue), and 0.3% (light blue).
**Figure 5** – All 72 combinations of radiation dose and reconstruction algorithm in the low-contrast phantom at 1.0% nominal contrast level. *CT images may not be adequately reproduced in print.*

**Figure 6** – Noise power spectra for all reconstruction algorithms at 84 mGy.
The studies included in this thesis show that software solutions can be used to improve image quality and reduce radiation doses in neuroradiological CT examinations. Results of the clinical studies (Papers I–III) and the phantom study (Paper IV) all revealed significant improvement of image quality compared with conventional FBP reconstruction. With all other examination parameters kept constant, image quality using the new methods surpassed the reference method (Papers III, IV). Even with considerable radiation dose reduction, certain image quality parameters were improved by the new method compared with the standard dose reference method (Papers II, IV).

Image quality and radiation dose

The post-processing image filter SharpView CT (Paper I), is a stand-alone solution that performs image optimization in the image domain only. The main benefit of the post-processing approach is the ability to serve multiple CT systems from different vendors. A theoretical limitation of the post-processing approach is the inability to optimize projection data.

Optimization of projection data is a key feature of IR algorithms, such as iDOSE, with the main purpose being to prevent noise and artifacts from propagating to the image domain where they might be more difficult to remove. As a result, image quality in IR can be expected to be superior to post-processing filtering.

Unfortunately, no publications have directly compared SharpView CT and the IR algorithms. In a white paper published by Siemens, their statistical IR algorithm (SAFIRE) is stated to improve image quality compared with a post-processing image filter from the same vendor (IRIS). However, this has not been confirmed in the peer reviewed literature. The only publication including both SAFIRE and IRIS does not explicitly compare the two, but the results do not indicate a clear benefit of the IR algorithm compared with the post-processing image filter.
Post-hoc comparison of results from Papers I and II is difficult to perform because of different evaluation approaches used in the studies. Despite both studies using the same internal reference (normal dose FBP), the results of the subjective ranking procedure cannot be directly compared as a different number of ranks was used in the studies: three in Paper I and four in Paper II. However, noise was measured in a similar way in both studies allowing comparison of relative noise reduction. With FBP as baseline, SharpView CT reduced noise by 11%, while iDOSE4 reduced noise by 10.8% (ID2) and 21.4% (ID4). Thus, more noise reduction was achieved with the IR algorithm than the post-processing image filter.

Review of the literature shows that both post-processing image filters and IR algorithms improve image quality and show potential for radiation dose reduction. This is in line with the results presented in this thesis. SharpView CT has been shown to improve image quality, with a potential for 13–30% radiation dose reduction in brain CT\textsuperscript{13, 44}, 30–50% dose reduction in abdominal CT,\textsuperscript{10, 70, 71} and 58% in CT of the paranasal sinuses.\textsuperscript{9}

Likewise, iDOSE4 has been shown to improve image quality in various applications,\textsuperscript{72–74} (Papers II–IV), but so far the potential for radiation dose reduction in clinical examinations has only been explored in cardiac CTA (55–63% reduction)\textsuperscript{8, 75} and temporal bone CT (50% reduction).\textsuperscript{76} One study on lesion detection rates in a liver phantom model reports no benefit of iDOSE4 compared with FBP.\textsuperscript{77}

Similar results have been published for IR algorithms from other vendors. However, studies comparing IR algorithms from different vendors are lacking. Only two studies have compared IR algorithms from two vendors.\textsuperscript{7, 11} All other published studies (and Papers II, III) have compared the IR algorithms with FBP in the same CT-system. Therefore the comparison of six iterative IR algorithms in four CT systems in Paper IV gives a unique insight into the strengths and weaknesses of the algorithms.

Although all IR algorithms evaluated in Paper IV improved image quality compared with FBP reconstruction, they showed different strengths and weaknesses. The overall best performing algorithm in the study was the prototype model-based IR algorithm IMR (Philips), which managed to greatly reduce noise and at the same time improve spatial resolution. Both evaluated model-based IR algorithms (GE Veo and Philips IMR) were most effective at lower radiation dose levels – a finding indicating that this new generation of IR might allow further radiation dose reduction in clinical protocols. This has already been confirmed for GE Veo in various applications.\textsuperscript{11, 78–81}
Comprehensive evaluation of image quality

Diagnostic radiology is constantly evolving through introduction of new technologies and continuous optimizations of current methods. Every innovation or major adjustment necessitates validation before implementation in clinical routine. Ideally, this entails multicenter randomized case-control clinical trials that measure diagnostic clinical performance compared with the clinical reference standard – the best available method for establishing the true status of a subjects’ condition. However, such clinical trials are complicated and time consuming. Considering the fast rate of development in diagnostic radiology, simpler evaluation methods are needed.

In Papers I–III, major simplification was achieved by using diagnostic acceptability testing as it involves evaluation of important anatomical structures instead of specific disease processes. In this approach, diagnostic performance is indirectly measured through evaluation of carefully selected clinically important image quality criteria. The lack of diagnostic accuracy testing can be considered a weakness of the studies. However, diagnostic accuracy is not the ultimate evaluation method and studies exploring diagnostic/therapeutic impact and patient outcome are considered superior.

Most studies on subjective image quality and IR only include multiple research subjects, but only two observers. This is the case for all published studies on IR in brain CT. Employing two observers is sufficient only if inter-observer agreement is high and if the observers can be assumed to be representative for observers in general. Results from Papers I–III show that inter-observer variability is a problem and that it is unreasonable to assume that any two observers are representative for observers in general. Therefore, as recommended in the literature, more observers are needed. However, a weakness of Papers I–IV is that all observers were recruited from the same center. This might for instance have contributed to overestimation of observer agreement.

Increasing the number of observers is not the only way to improve reliability of study results. Observer agreement can be improved by careful study design. In Paper I, considerable observer variation was seen in the subjective ranking procedure. To further explore this, Paper II included both side-by-side ranking and individual grading procedures supported with a more extensive objective evaluation. From this we concluded that grading was more reliable for the assessment of image quality, especially when complemented with a thorough objective assessment. However, observer variation was still considerable in Paper II, indicating that further methodological improvement was possible. An improved version of the grading scale
was introduced in Paper III, including image examples for each grade. This may have contributed to the observed improvement in subjective grading consistency, resulting in reduced observer variation.

As explained above, the main problems associated with subjective evaluation of image quality are related to consistency and reproducibility. Since these are the main strengths of objective evaluation methods, the two methods complement each other well. Various methods for assessment of image quality have been employed throughout this thesis in an attempt to improve on the methodology. Based on the findings presented in the papers, the following outlines for comprehensive evaluation of image quality have been constructed:

- **Evaluation:** Assessment of image quality benefits from using both subjective and objective methods (Papers I–IV) as these have different strengths and weaknesses. An important weakness of subjective evaluation is that it is prone to various observer biases that might affect study results. A disadvantage of objective evaluation is that it does not evaluate human perception of images and observer variation. Therefore, a combined approach should result in more reliable study results.

- **Observers:** In subjective assessment of image quality, multiple observers are required to compensate for observer diversity as illustrated in Papers I–III. This is especially important when the observers themselves are responsible for the study design and therefore also subject to **confirmation bias**.

- **Evaluation scales:** Image quality is preferably evaluated with grading of clinically important quality criteria. Grading scales should be clinically correlated and the grading steps should reflect clinically important differences. Four-grade scales may provide all the necessary information in the context of optimization of image quality and radiation dose in CT (Papers I–III).

- **Research subjects:** The use of identical case and control groups eliminates confounding factors related to individual anatomical and pathological diversity. Therefore, differences in results between the two groups can with greater certainty be assumed to be caused by differences between the methods being evaluated. Due to the perfectly paired situation, significant results can be achieved using small study samples (Papers I–III).
Clinical impact

Positive study results should preferably lead to clinical implementation of the observed advances. Although the results of Paper I (SharpView CT) were positive, they were never implemented in clinical routine at our clinic since the iterative reconstruction algorithm evaluated in Paper II (iDOSE\textsuperscript{4}) was launched close in time and also proved to be effective. However, the results of both Papers II and III have had direct clinical impact on the current routine brain CT and craniocervical CTA protocols at our clinic, leading to reduced radiation doses and improved image quality.

In the near future, continuing advances in CT technology can be expected to further improve image quality. If possible this should lead to further reduction in radiation dose, since in CT, too low or too high image quality is equally inappropriate.
The main objectives of this thesis were to assess new methods for improving image quality in CT, explore the associated potential for radiation dose reduction, and to outline a comprehensive approach for evaluation of image quality.

The following conclusions can be drawn from the individual papers:

Paper I

- Post-processing image filtering (SharpView CT) improves image quality in brain CT acquired with 30% reduced radiation dose.
- Noise levels were reduced and equivalent with normal dose. Subjective image quality was improved although normal dose still was considered non-significantly superior.
- Ranking of image quality is associated with considerable observer variation.

Paper II

- Statistical iterative reconstruction (iDOSE®) improves image quality in both phantom and clinical brain CT acquired with 30% reduced radiation dose.
- Noise levels and contrast-to-noise ratios were significantly improved beyond normal dose examinations. Subjective image quality was also improved although normal dose still was considered significantly superior.
- Observer bias is a problem associated with subjective evaluation of image quality, although less so in grading than in side-by-side ranking.
- To produce representative results studies on subjective image quality require an adequate number of both research subjects and observers.
• Modest inter-observer agreement was seen, possibly due to the use of using written grading scale only, lacking accompanying example images.

Paper III

• Statistical iterative reconstruction (iDOSE4) improves image quality in craniocervical CT, especially at the thoracic inlet.
• Noise levels, signal-to-noise ratios, and contrast-to-noise ratios were all significantly improved compared with filtered back-projection. Subjective image quality was also improved at all levels.
• Perfect matching of cases and controls eliminated many confounding factors possibly affecting image quality in CTA.
• Improved inter-observer agreement was seen, probably reflecting a well-defined written grading scale in combination with example images.

Paper IV

• The iterative reconstruction algorithms have different strengths and weaknesses, but all improve image quality compared with filtered back-projection from the same vendor.
• Noise levels, signal-to-noise ratios, and contrast-to-noise ratios all improved, although to a variable degree. Three of the algorithms also improved spatial resolution. Subjective image quality was generally improved for all algorithms.
• The model-based algorithms improved image quality progressively with decreasing radiation doses, while the effect of statistical algorithms was more constant irrespective of radiation dose.
• Iterative reconstruction affected noise-power spectra, explaining differences in image texture, especially for the model-based algorithms.
Acknowledgements

Working on my dissertation has been an inspiring and fulfilling experience. It has given me the opportunity to work with many wonderful people that have contributed to the success of this project.

First, I want to thank my senior supervisor, Isabella Björkman-Burtscher, for believing in me and encouraging me when I needed it the most. Also I want to thank my assistance supervisor, Roger Siemund, who first introduced me to radiology research. Thank you for your relentless effort in getting this project done in a timely manner!

Next, I want to thank Peter Hochbergs, head of radiology, and professors Pia Maly-Sundgren and Stig Holtås for their interest in supporting my academic efforts and for giving me the opportunity to focus on this project.

Then, I want to thank Peter Höglund for his invaluable statistical guidance and advice, and for always being there for me when I got stuck.

Furthermore, I would like to thank my other co-authors: Danielle van Westen, Lars Stenberg, Cecilia Petersén, Per Undrén, Birgitta Ramgren, Marcus Söderberg, Marie-Louise Olsson and Fredrik Stålhammar – without your help I would never have succeeded.

Besides my co-authors, I also want to acknowledge my other colleagues at the department of neuroradiology: Åke Forssell, Mats Cronqvist, Olof Rudling, Martin Larmark, Håkan Sjunnesson, Jan Anders Ahnlide, Tomas Nilsson, Kasim Abul-Kasim, and Angelica Overgaard – thank you for covering for me at work when I was busy doing research!

My colleague and roommate, Mats Geijer, deserves a special mention. Working in parallel with him on his pursuit of becoming an associate professor, provided the competitive atmosphere that sometimes is necessary to get things done!
Also, I want to thank many radiographers, medical physicists, secretaries and administrators for much appreciated assistance: Mette Forseth, Eva Borglin, Karolina Larsson, Siv Tjärnberg, Helena Szaboova, Amir Vahabi, Alexandra Trudel, Marina Jamtheim, Sonny La, Lars Weber, Mats Nilsson, Åsa Hansson, Gunilla Angelin, Annika Samuelsson, Karin Jönsson, Anita Nilsson, Annika Törling-Ring, and Inger Bjarke ...with apologies if I forgot anyone – well done all of you!

Last, and above all, I want to thank my wife Elva Rut and my two wonderful daughters Sif (6) and Saga (4). Without their patience and support I would never have made it – I love you!
References


38. Hsieh J. Removing the Barriers (ASIR). *CT clarity - The magazine of CT (GE healthcare)*, Buckinghamshire, UK; 2009.


58. Knapp TR. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing research* 1990;39:121-123.


61


82. Draft guidance for industry, clinical investigators, and Food and drug administration staff - Design considerations for pivotal clinical investigations for medical devices. *U.S. Department of health and human services; Food and drug administration (FDA); Center for biologic evaluation and research; 2011. (See: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm265553.htm)
Appendix