Optimization of Paediatric CT Examinations

An Approach to Minimize Absorbed Dose to Patients with Regard to
Image Quality and Observer Variability

Kerstin Ledenius

Department of Radiation Physics, Institute of Clinical Sciences
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden, 2011

To my beloved family

Optimization of Paediatric CT Examinations
An Approach to Minimize Absorbed Dose to Patients with Regard to Image Quality and Observer Variability

Kerstin Ledenius

Department of Radiation Physics, Institute of Clinical Sciences at Sahlgrenska Academy, University of Gothenburg, SE−413 45 Gothenburg, Sweden

Abstract

The absorbed dose to paediatric patients is important bearing in mind the increased risk of radiation-induced cancer due to exposure to X-rays at young ages. Questions have also been raised of whether a CT examination of the paediatric brain might lead to a reduction in cognitive function. Considering the difference in anatomy and thus in X-ray attenuation, children have a special need in CT image quality and require separate scanning protocols and thus separate optimization from adults.

The overall aim of the work described in this thesis was to find an optimization approach to minimize the absorbed dose to paediatric patients undergoing CT examinations, while maintaining the diagnostic image quality and taking into account observer variability. In a first study, the effect of reducing the tube current on the diagnostic image quality was evaluated for paediatric cerebral CT examinations using the non-parametric statistical method of inter-scale concordance. The observer variability was evaluated by means of Svensson's method in a second study. The approaches in these two studies were then combined in a third study to optimize the noise index in abdominal paediatric CT examinations. The aim of the fourth study was to estimate the variability in the results when using inter-scale concordance. A post-processing 2D adaptive filter, claiming to enable reductions in radiation exposure, was investigated in the third study, and in a separate fifth study.

Artificial noise was added to copies of raw data of paediatric CT examinations in order to simulate a reduction in radiation exposure without having to expose paediatric patients to further scans. When the adaptive filter was tested, all images were created in duplicate: one set being post-processed. All images, including the images duplicated for test-retest assessments were evaluated blindly and randomly by three (two in one study) observers using a software viewing station. The radiologists assessed the image quality visually by grading the reproduction of high- and low-contrast structures and overall image quality on a 4-point rating scale.

For the cerebral CT examinations reductions in radiation exposure were possible for patients 1 to 10 years old. It was possible to further reduce the radiation exposure for shunt-treated patients. The original image quality for patients under 6 months of age was found to be inadequate. Noise index 11 was sufficient for a routine abdominal examination for patients aged 6 to 10 years, noise index 12 was considered sufficient for patients aged 11 to 15 years. The variability in results was less than 20 % between two cerebral studies regarding routine CT examinations. The post-processing filter enabled reductions in radiation exposure of approximately 15 % for some age groups.

The approach used in this work enabled the inter-scale relations between radiation exposure and diagnostic image quality to be determined for paediatric cerebral and abdominal CT examinations. Observer variability was also evaluated and a minimum radiation exposure to paediatric patients was suggested. Applying the approach to post-processed images indicated a possible reduction in radiation exposure to paediatric patients.

Optimering av datortomografiundersökningar av barn
Ett tillvägagångssätt att minska stråldosen till barn med hänsyn till bildkvalitet och variation bland granskarna

Populärvetenskaplig sammanfattning

Datortomografen är en röntgenmaskin där röntgenrör och detektor roterar runt patienten, på så sätt får detaljrika snittbilder till skillnad från konventionell röntgen. Datortomografi (DT) har blivit en alltmer populär undersökning trots att den ofta ger högre stråldoser än vid konventionell röntgen. Eftersom nya tekniska förbättringar av datortomografen presenteras i rask takt så har processen kring optimering av maskinen (d.v.s. se till att man använder lägsta möjliga stråldos som fortfarande resulterar i en tillräckligt bra bildkvalitet) fått kontinuerligt nya förutsättningar att ta hänsyn till.

Stråldosen till barn är speciellt intressant eftersom barn är extra känsliga för strålning. Med avseende på de anatomiska skillnader som finns mellan vuxna och barn så behöver barn specifik bildkvalitet och separata undersökningsprotokoll. Forskning kring lämpliga protokoll för barn är viktig och speciellt kring enkla tillvägagångssätt för röntgenkliniker att på egen hand ta tag i optimeringen av protokollen.

Denna avhandling syftar till att presentera ett tillvägagångssätt att minska stråldosen till barn med hänsyn till bildkvalitet och variation bland granskarna. I en första studien undersöktes effekten som en sänkning av stråldosen har på bildkvaliteten genom att simulera sänkningar av den så kallade rörströmmen för DT-undersökningar av hjärnan på barn. På så sätt kunde den lägsta stråldosen som fortfarande gav en acceptabel bildkvalitet identifieras. I en andra studie låg fokus på att utvärdera variationen mellan de granskare som studerat bilderna. Teknikerna från studierna sammanfördes i en tredje studie där bildkvaliteten i DT-undersökningar av magen på barn undersöktes. I en fjärde studie utvärderades tillvägagångssättet att identifiera minsta stråldosen, genom att genomföra en ny studie av DT-undersökningar av hjärnan på barn och studera variationen i resultat mellan denna och första studien. En mjukvara som påstods vara till hjälp med att sänka stråldosen testades i den tredje studien men även i en separat, femte studie.

För att finna relationen mellan stråldos och bildkvalitet tillfördes artificiellt brus till redan genomförda undersökningar för att simulera en sänkning i stråldos. Med denna teknik behövde inga patienter ställa upp på undersökningar i rent forskningssyfte. För tester av mjukvaran skapades dubbletter till alla bilder där ena kopian behandlades med mjukvaran. Vissa bilder dubblerades för att utvärdera hur konsekventa granskarna var i sina bedömningar. Radiologer bedömde bildkvaliteten efter hur väl strukturer i bilden syntes, samt helhetsintrycket av bildkvaliteten. Den bedömda kvaliteten matchades ihop med stråldosen med hjälp av en så kallad rang-baserad statistisk metod och på så sätt kunde man få fram den lägsta stråldosen som representerar en viss bildkvalitet. Granskarvariationen utvärderades med ytterligare en statistisk metod som fokuserar på att analysera olikheter mellan granskare.

För DT-undersökningar av hjärnan på barn så fanns det marginal att minska stråldosen till patienter mellan 1 och 10 år. Ytterligare minskningar i stråldos var möjligt för uppföljningsundersökningar av shunt-behandlade barn. För barn under 6 månader visade sig stråldoserna vara för låga redan från början. Noise index 11 var tillräckligt för barn mellan 6 och 10 år medan noise index 12 var tillräckligt för barn mellan 11 och 15 år. Variationen i resultat mellan två studier angående rutinundersökning av hjärnan på barn var under 20 %. Mjukvaran som påstods hjälpa till vid dosreducering kunde sänka doserna med ca 15 % för vissa åldersgrupper.

Slutsatsen i denna avhandling är att tillvägagångssättet som användes (och undersöktes) i studierna är användbart till att identifiera lägsta rörström som ger en viss bildkvalitet vid DT-undersökningar av barn.

List of Papers

This thesis is based on five papers, which will be referred to in the text by their Roman numerals.

I.   Ledenius K, Gustavsson M, Johansson S, Stålhammar F, Wiklund LM and Thilander-Klang A
     Effect of tube current on diagnostic image quality in paediatric cerebral multidetector CT images
     Br J Radiol. 2009 Apr; 82(976): 313-320

II. Ledenius K, Svensson E, Stålhammar F, Wiklund LM and Thilander-Klang A
     A method to analyse observer disagreement in visual grading studies: example of assessed image quality in paediatric cerebral multidetector CT images
     Br J Radiol. 2010 Jul; 83(991): 604-611

III. Ledenius K, Stålhammar F, Jönsson M, Boström H and Thilander-Klang A
     Optimization of noise index in paediatric abdominal computed tomography images
     Submitted to European Radiology

IV. Ledenius K, Båth M, Stålhammar F, Wiklund LM and Thilander-Klang A
     Estimating the variability in optimization by repeating a study on paediatric cerebral CT examinations
     Submitted to British Journal of Radiology

V.   Ledenius K, Stålhammar F, Wiklund LM, Fredriksson C, Forsberg A and Thilander-Klang A
     Evaluation of image-enhanced paediatric computed tomography brain examinations
     Radiat Prot Dosimetry. 2010 Apr-May; 139(1-3): 287-292

Preliminary results have been presented at the following conferences:

Ledenius K, Gustavsson M, Johansson S, Stålhammar F, Söderberg J, Wiklund L-M and Thilander Klang A
Optimization of Absorbed Dose versus Image Noise in Paediatric Multi-Slice CT Examinations
Oral presentation at the Annual Swedish X-ray Conference (Röntgenveckan), September 15-19, 2003, Norrköping, Sweden

Ledenius K, Gustavsson M, Johansson S, Stålhammar F, Söderberg J, Wiklund L-M and Thilander Klang A
A Method to Predict the Image Noise in Paediatric Multi-Slice CT Examinations
Oral presentation at the Second Malmö Conference on Medical Imaging, April 23-25, 2004, Malmö, Sweden

Ledenius K, Gustavsson M, Johansson S, Stålhammar F, Wiklund L-M and
Thilander Klang A
Paediatric cerebral MDCT: A balance between radiation dose and image quality
Oral presentation at the Annual Swedish X-ray Conference (Röntgenveckan), September 19-23, 2005, Malmö, Sweden

Ledenius K, Gustavsson M, Johansson S, Stålhammar F, Wiklund L-M and
Thilander Klang A
Quality Assessment of Simulated Dose-Reduced Pediatric Cerebral CT Scans
Oral presentation at the Annual Conference of the Radiological Society of North America, November 27 to December 2, 2005, Chicago, USA

Ledenius K, Gustavsson M, Johansson S, Stålhammar F, Wiklund L-M and
Thilander Klang A
Reduction of radiation dose in pediatric CT brain examinations – A pilot study
Presented as a poster at the Annual Swedish X-ray Conference (Röntgenveckan), September 18-22, 2006, Örebro, Sweden

Ledenius K, Stålhammar F, Wiklund L-M, Fredriksson C, Forsberg A and Thilander Klang A
Evaluation of image enhanced paediatric computed tomography brain examinations
Presented as a poster at the Third Malmö Conference on Medical Imaging, June 25-27, 2009, Malmö, Sweden

Ledenius K, Johansson S, Stålhammar F, Wiklund L-M and Thilander-Klang A
Experiences of optimizing paediatric CT examinations
Oral presentation at the Annual Swedish X-ray Conference (Röntgenveckan), September 1-4, 2009, Jönköping, Sweden

Table of Contents

Abbreviations

ALARA	As low as reasonably achievable
AUC	Area under the curve
ATCM	Automatic tube current modulation
CRT	Cathode ray tube
CT	Computed tomography
$CTDI_{vol}$	Computed tomography dose index by volume
$CTDI_w$	Weighted computed tomography dose index
DLP	Dose-length product
DICOM	Digital imaging and communications in medicine
DRL	Diagnostic reference levels
IEC	International electrotechnical commission
LCD	Liquid crystal display
MDCT	Multi-detector computed tomography
MRI	Magnetic resonance imaging
NI	Noise index
PA	Percentage agreement
PMMA	Polymethyl methacrylate
ROC	Receiver operation characteristics
RC	Relative concentration
RP	Relative position
RV	Relative rank variance
RTPA	Rank transformable pattern of agreement
SNR	Signal-to-noise ratio
TCM	Tube current modulation
VGA	Visual grading analysis
VGC	Visual grading characteristics
ViewDEX	Viewer for digital evaluation of X-ray images
VRS	Verbal rating scale

1 Introduction

1.1 Background

Computed tomography (CT) is a medical X-ray imaging system where the X-ray tube and the detectors are rotated around the patient. The data obtained are then used to create cross-sectional images of the examined patient. CT has undergone a tremendous technical development since the invention of the first CT equipment in the 1970s. A milestone in the development of CT was the introduction of multiple rows of small detector elements in 1998, which replaced the one row of large detectors, enabling many images to be reconstructed for each rotation instead of one. The main advantage of the multi-detector technique is the radical reduction in scan time as there is no longer a need to limit the beam width to the nominal slice thickness. For paediatric patients, this meant a reduction in the sedation necessary and fewer motion artefacts. Replacing a row of large detectors with several small ones resulted in a loss of detection efficiency; however, the resolution along the z-axis (along the patient) was dramatically increased.

The first decade of the new millennium saw intense competition between manufacturers to produce CT systems that could collect the most images per rotation, with the thinnest possible nominal slice thickness, using the fastest gantry rotation time. However, physical limitations have started to show. For example, the beam has become wider which results in more secondary radiation and distortion of images, as the beam changes from fan shaped to cone shaped. There are also practical limitations in the transportation of the patient through the gantry as too rapid table movements create motion artefacts. Rapid gantry rotation causes mechanical strain on the equipment and high requirements on X-ray tube and detectors. The focus has since been changed to improving the components of the CT, such as the X-ray tube, detector efficiency and data processing. The new objectives in the development of CT are the introduction of the iterative image reconstruction and dual-energy scanning, which will change CT imaging as we know it today.

CT is an important diagnostic tool in modern healthcare. However, CT has a reputation for high radiation exposure of patients compared with conventional X-ray examinations. Ionizing radiation is associated with health risks to humans at effective doses higher than 100 mSv; cancer being one of the stochastic risks [1]. Opinions on the effects of low doses of ionizing radiation (below 100 mSv) differ as to whether there are any stochastic risks or not. Performing scientific studies on the subject is difficult for practical reasons as it requires enormous samples in order to maintain statistical precision and power. For example, to be able to draw conclusions regarding the effects of an effective dose of 10 mSv, a sample size of approximately 5 million subjects would be required [1].

The effective doses resulting from paediatric CT examinations today is normally in the range of <1 to 30 mSv [2]. However, repeated examinations of patients are very common [3, 4], resulting in larger accumulated effective doses. It is recommended that exposure levels are kept as low as reasonably achievable to reduce the potential risks [5], this recommendation is also known as the ALARA principle. Not only should radiation exposure be kept as low as possible, the use of CT must also be justified, and other diagnostic methods should be considered when possible.

Introduction

The risk of radiation-induced cancer as a result of ionizing radiation is higher in children than in adults [5, 6]. This is partly because the radiation-sensitive organs, such as bone-marrow, represent a higher proportion of the body mass in children than in adults. Cell division is also more active in children, which increases the sensitivity of cell damage. The risk of a paediatric patient developing cancer is up to ten times higher than that for an adult [7]. Hall et al. [8] have also raised the question of whether irradiation from CT of the paediatric brain may lead to a reduction in cognitive function. Not only are children more sensitive to radiation, but if the radiation exposure is not adjusted for children, they will receive higher organ doses than adult patients, as their bodies do not attenuate the same number of photons before the organs of interest are reached, resulting in a higher energy deposition per unit mass.

In an American study published in 2001 [6], it was roughly estimated that 500 out of the paediatric patients undergoing a CT examination during a year would ultimately die as a result of radiation-induced cancer following the CT examination in the United States. This estimate was based on a linear extrapolation of the cancer risk, and that approximately 600 000 paediatric CT examinations are carried out annually. This roughly equals 1 patient in a 1000. It should be noted that this estimate does not include children expected to recover from CT-induced cancer. Berrington de González et al. [9] estimated that a total of 29 000 cases of cancer were related to CT scanning in the United States in 2007; 15 % of which were estimated to be due to the scanning of patients under 18 years of age.

The improvement in image quality and the diagnostic ability of CT has led to an increase in its popularity, and conventional X-ray examinations are increasingly being replaced by corresponding CT examinations. According to estimates in 1997 [10], CT represented about 4 % of all diagnostic X-ray examinations and almost 40 % of the total radiation dose from medical diagnostic examinations in the United Kingdom. Data from 2006 [11] indicated that in the United States, CT represented about 15 % of all diagnostic X-ray examinations (excluding dental examinations), and more than half of the collective dose resulting from medical diagnostic examinations.

Optimizing paediatric CT examinations, i.e. keeping radiation exposure levels to a minimum without jeopardizing the diagnostic image quality, is highly important considering the increased use of CT and the increased risk of radiation induced-cancer in children. In 2001 a series of papers published in the American Journal of Roentgenology [6, 12, 13] highlighted the problem of high radiation exposures in paediatric CT scanning. Paterson et al. [13] published the results of an investigation on paediatric scanning settings, showing that many hospitals still used adult scanning settings for children, resulting in very high effective doses. One of the problems at that time was the lack of tube current modulation (TCM). This technique, which adjusts the radiation exposure in relation to patient attenuation, was only available in conventional X-ray examinations. Without this technique, CT operators had to reduce the radiation exposure by hand for children. As there was a risk of reducing the radiation too much, resulting in images of inadequate quality, there was a tendency not to adjust the tube current at all, leading to high radiation doses to young patients [13, 14]. Some hospitals even increased the radiation exposure for paediatric patients to ensure a high image quality. Scientific efforts were made to estimate the relation between tube current and patient size [15-18]; the patient size providing a rough estimate of the patient attenuation. However, as is common in the development of CT, technology advanced, and TCM was introduced in CT a few years later.

With the introduction of TCM, new issues of how to optimize CT examinations arose. Today, TCM can be used to adjust the tube current to provide a relatively constant image quality throughout the examination, regardless of the patient's morphology. However, the reference image quality set by the CT operator is, by author's experience, often not optimized with regard to the diagnostic purpose of the image but rather based on previous settings or manufacturer's recommendations. Some argue that using a generally high image quality regardless of the diagnostic purpose will result in additional findings in some patients. Whether this can be regarded as a sufficient reason to overexpose the majority of patients and thus increase the number of future cancer cases can definitely be questioned.

CT examinations can be optimized in a number of ways, e.g. better technical solutions and correct use of the equipment. Optimizing a scanning protocol can mean identifying the scanning parameters that result in the highest image quality for a fixed effective dose. In this case there are many statistical methods that are appropriate for comparing two or more parameters with each other. It can, however, mean finding the lowest effective dose without reducing the diagnostic image quality. In this case, it is common to focus on either the tube current, and thus the value of the CT dose index by volume, $CTDI_{vol}$ (see Section 1.2), or the tube voltage. Reducing the tube current results in an increase in image noise and comparing a higher level of tube current to a lower would thus most probably indicate a change in image quality in favour of the higher tube current when using established statistical methods. However, there is often no information of whether the lower image quality is sufficient or not for the diagnostic purpose.

Scanning protocols should be optimized not only with regard to specific indications, but also for specific patient groups. For example, large patients do not require the same low level of image noise as smaller patients [19-21]. One reason for this could be that fat is less attenuating than soft tissue, and appears darker than soft tissues in CT images with abdominal window settings. Soft tissues are therefore better delineated in images of more corpulent patients. Separate optimization is required also for paediatric patients as their anatomy differs from that of adults. Apart from being smaller, the anatomical structures in children have different proportions. Many organs have different CT numbers than for adults [22, 23] thus resulting in e.g. different contrast. For example, the skull bone is much softer and thinner in children, as can be seen in Figure 1. They also lack the fat embedding the organs mentioned earlier (see Figure 2). It is also important to bear in mind that $CTDI_{vol}$ and reference image quality do not represent the same image quality between different CT scanners because of the specific technical solutions used by each manufacturer. Optimization is thus specific for each kind of scanner.

Optimizing CT scanning protocols for children is also more limited than for adults for several reasons. Firstly, the number of patients examined is smaller. Secondly, different protocols are required for different age groups. Thirdly, specialist radiologists trained in interpreting paediatric images are required. There is thus a need to develop a method of optimization that can easily be used in hospitals for paediatric CT scanning protocols. This method should be possible to apply to a limited number of patients, and should be able to differentiate between images of adequate and inadequate quality.

Figure 1. Images from cerebral CT examinations of patients of various ages. Age of the patient to the upper left; 8 months, to the upper right; 10 years, and below; 54 years. The window settings and pixel size are not identical.

Figure 2. Illustration of an abdominal CT examination of a 7 year old patient (to the left) and a 66 year old patient (to the right). For the older patient, fat around the kidney (reproduced with a black colour) enhances the delineation of the organ. The window settings and pixel size are not identical.

1.2 Dosimetry

The computed tomography dose index (CTDI) provides an estimate of the absorbed dose (mGy) within the scanned plane (xy-plane). More mathematically, CTDI can be seen as the area under the dose profile divided by the beam width (i.e., the nominal slice thickness multiplied by the number of slices per rotation), see Figure 3. The length along the z-axis, over which the dose profile is integrated, varies according to the defin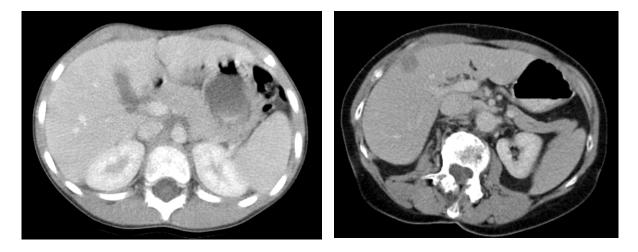ition of CTDI. For example, the American Food and Drug Administration recommends a 14-slice width ($CTDI_{FDA}$) [24], while the International Electrotechnical Commission (IEC) and International Atomic Energy Agency (IAEA) recommends a constant 100-mm width ($CTDI_{100}$ and $C_{a,100}$, respectively) [25, 26].

The general definition of CTDI is:

$$CTDI = \frac{1}{nt} \int_{z-axis} D(z)dz \qquad (1)$$

where n is the number of slices per rotation, t is the nominal slice thickness and D(z) is the value of the dose quantity (air kerma according to IAEA [25] or absorbed dose in air according to IEC [26] and FDA [27]) at different positions along the z-axis. CTDI can be measured either free in air ($CTDI_{air}$) or in a polymethyl methacrylate (PMMA) phantom. The weighted CTDI ($CTDI_w$) is the sum of weighted $CTDI_{100}$ measurements at different positions within the PMMA phantom and is used as an approximation of the average absorbed dose within the xy-plane. It is defined as follows [26, 28]:

$$CTDI_w = \frac{1}{3} CTDI_{100(Central)} + \frac{2}{3} CTDI_{100(Peripheral)} \qquad (2)$$

The positions of the central and peripheral measurements in the PMMA phantom can be seen in Figure 3. $CTDI_w$ is only valid as an estimate of the dose contribution within the scanned plane if the pitch (the relation between beam width and table movement per rotation) equals one.

Today, the CTDI by volume ($CTDI_{vol}$) is the recognized measure, which takes into account the spacing between rotations.

$$CTDI_{vol} = \frac{CTDI_w}{pitch} \qquad (3)$$

The tube current is often incorrectly used as an estimate of the dose to the patient, and in many articles the results are present in terms of the tube current (mA) instead of $CTDI_{vol}$ (mGy). The exact relation between tube current and the resulting dose, $CTDI_{vol}$, is scanner-specific which means that the tube current can only be used when presenting relative changes in dose. To estimate the total exposure of the patient, the dose length product (DLP, Gy cm) takes the irradiated volume into account:

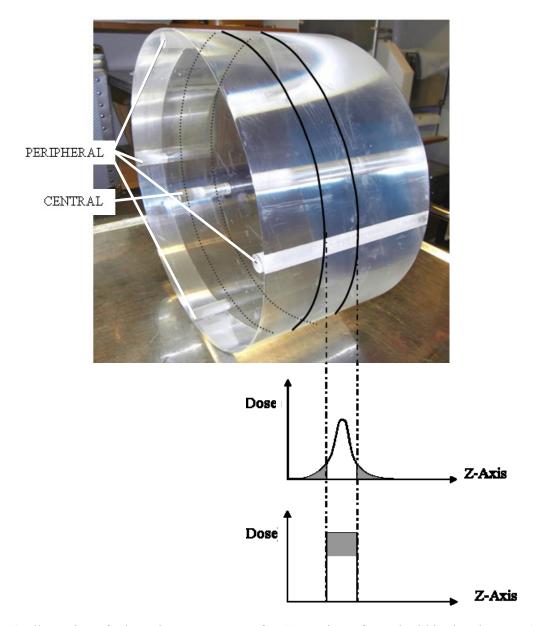$$DLP = CTDI_{vol} \cdot scan\ length \qquad (4)$$

Introduction

Figure 3. Illustration of where the measurement for CTDI$_w$ is performed within the phantom (central and peripheral positions). It also illustrates how the dose profile is summed over the beam width to include the penumbra in the measurements.

DLP can be used to estimate the effective dose, E, when multiplied with a conversion factor, E$_{DLP}$, which depends on the anatomical area scanned [29]. The conversion factors are for a standard adult (70 kg), but conversion factors for children have been suggested [30]. At present, there are several methods of calculating more accurate values of effective dose from CT examinations, depending on the desired level of accuracy: the more accurate the value, the more complicated the calculation. There are also several dose-calculation programs employing program-specific weighting factors for children, resulting in different values of effective dose. Whether CTDI$_{vol}$ or effective dose should be used when comparing doses resulting from CT examinations is currently a topic of heated debate within this area of research [31-34]. The effective dose is perhaps the most correct quantity to use when comparing radiation doses to patients undergoing different CT examinations; however, it is more complicated to calculate, as the mean absorbed dose to each organ is included in the

definition. However, most scientists agree that $CTDI_{vol}$ will soon have to be measured for distances greater than 100 mm along the z-axis, as 100 mm underestimates the absorbed dose to the patient [35, 36] for wide beams due to the penumbra not being sufficiently covered. For some CT models 100 mm is not even sufficient to cover the whole beam width (at the time of writing, beam widths up to 160 mm exists).

Another issue concerning $CTDI_{vol}$ is that it today only is given for two reference phantoms (16 and 32 cm in diameter) after an examination. A newborn child can have cranial and abdominal diameters as small as 10 cm, and a corresponding $CTDI_{vol}$ value in such a phantom can be up to a factor 2.5 higher [37]. Using $CTDI_{vol}$ to estimate effective dose would be more accurate if a size regulating factor were used [30, 38, 39]. Identical output between CT scanners does not necessarily mean identical image quality, but it enables comparisons of the levels of output for different examinations and different patient groups.

1.3 Image noise

Image noise in a CT image can generally be divided in to three kinds, quantum noise, system noise and noise from the reconstruction process and calibration of the data [40]. Quantum noise is the main contributor of noise to the image, and it is inversely proportional to the square root of the absorbed dose to the detector [41]. The absorbed dose in the detector is in turn determined by the output from the X-ray tube, additional filtering (such as the bow-tie filter), the attenuation by the patient and the efficiency of the detector. System noise results from the physical limitations of the different components of the CT scanner, such as electronic noise in the data acquisition and the detector elements, and scattered radiation, among other factors. Noise from reconstruction could for example originate from the enhancement of high signal frequencies in high-resolution kernels. The pre-processing techniques used to calibrate and condition the collected data are also sources of error; small artefacts sometimes influence the standard deviation of the pixel values [40].

Noise can be regarded as heterogeneous pixel values in the image of a homogeneous object. The standard deviation of the pixel values can be used as a measure of the level of noise but it is not a general measure that can define the quality of the image. Noise can appear differently in the image depending on which reconstruction filter has been used. Uncorrelated Poisson noise has equal noise power in all frequencies in the Fourier space. When introducing a reconstruction filter, the noise becomes correlated within the image. When the noise is greater at low frequencies it appears as coarse grains in the image, high frequency noise results in fine grains. The noise in an image can thus have completely different appearances even when the standard deviation of the pixel values is the same [42]. As different manufacturers use different reconstruction filters, direct comparisons between CT scanners are difficult. Observers are often biased as they prefer the image quality they are used to [43]. Trying to define a range of standard deviation values that is suitable for an examination is thus manufacturer and filter specific.

Image noise negatively influences the diagnostic image quality, i.e. the ability to visualize important structures. As image noise increases, the visibility of structures decreases, an effect that depends on the contrast and size of the structure [44]. Too low an image quality will prevent the detection of poorly visible pathology, while too high an image quality implies a higher radiation dose than necessary. It is therefore important to find a balance for optimized image quality.

Introduction

It is often of interest to reduce the noise within an image. Some reconstruction algorithms aim to reduce image noise, but this comes at the expense of resolution. There are several post-processing filters on the market claiming to have the ability to enhance X-ray image quality by reducing image noise, and enhancing anatomical structures; the aim being to enable reductions in the radiation exposure. One of these filters was tested in paediatric abdominal examinations (Paper III) and in paediatric cerebral examinations (Paper V) in order to evaluate the possibility of reducing the radiation exposure.

Artificial noise can be added to CT images through manipulation of the raw data [45, 46] or of the image itself [47]. This is done to simulate a reduction in tube current and thus a reduction in $CTDI_{vol}$ as it is proportional to tube current [48], see Figure 4. Adding noise directly to raw data ensures that it is filtered through the same reconstruction filter as the true image noise. Adding noise directly to the image requires more work as the characteristics of the added noise must match those of the real image noise. Adding noise to the raw data is preferable, although it requires a close cooperation with the manufacturer in order to get access to the raw data.
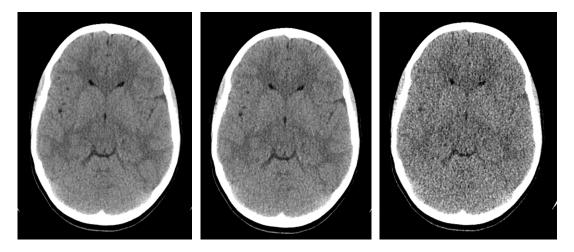

Figure 4. Illustration of images in which simulated noise has been added to the raw data from a paediatric cerebral CT examination of a 9 year old girl. The original examination (left) was performed with a $CTDI_{vol}$ of 42 mGy ($CTDI_{vol}$ is given for a 16-cm CTDI phantom). The simulated images represent $CTDI_{vol}$ values of 31 mGy (centre) and 9 mGy (right).

Adding noise is useful when finding the sufficient image quality required for a specific diagnosis or a general scanning protocol, as the effects of stepwise reductions in the radiation exposure can be visualized without having to scan the patient further. Simulating noise also has the benefit of providing identical images apart from the noise. This excludes the risk of bias from other factors such as patient movements between scans.

A fixed tube current must be used if TCM is not activated or not available for a CT examination. This means that, for example, an image of the shoulders (which are highly attenuating) will have a higher level of image noise (and most probably streak artefacts) compared to an image of the lungs (which are low attenuating). Adjusting the tube current to either the shoulders or the lungs will thus result in either a poor image of the shoulders or a high absorbed dose to the lungs. The use of TCM will not only reduce the dose to the patient [49], but may also provide relatively constant image noise throughout the image (if no limitations in the tube current occurs). This means that efforts can be devoted to finding an

image quality (with regard to image noise) that is sufficient for the diagnostic purpose of the examination, to a minimum CTDI$_{vol}$.

Manufacturers do not only use different technical solutions for TCM, but also different definitions of reference image quality. GE Healthcare and Toshiba Medical Systems use standard-deviation-related measures, while Philips Healthcare and Siemens Healthcare use reference tube current values. This means that CT scanners must be optimized with regard to manufacturer. Paper III describes the investigation of the reference image quality "noise index" sufficient for paediatric abdominal examinations at minimum radiation exposure on a CT from GE Healthcare. According to GE Healthcare, the noise index value will "approximately equal the standard deviation in the central region of the image when a uniform phantom (with the patient's attenuation characteristics) is scanned and reconstructed using the standard reconstruction algorithm" [50].

1.4 Evaluation of image quality

Physical measures such as detective quantum efficiency and modulation transfer function describe the ability of the equipment to reproduce a given signal to the detectors. This provides however no information regarding the clinical usefulness of the produced image. The perhaps most common physical measure in optimization is the signal-to-noise ratio (SNR) where scan parameters representing the minimum radiation exposure are identified for a fixed level of SNR. Psychophysical measures, in which trained observers determine the amount of detail of different size and contrast visible in a scanned test phantom, can also be employed. This technique is often used for quality assurance in order to detect differences in the performance of equipment over time. These measures are sometimes used when optimizing scanning settings in order to find the minimum radiation exposure with no visible loss of detail. The clinical validity of a study performed on phantoms is; however, always lower than that performed on humans.

Human evaluation of images from real clinical examinations is the most preferable approach when evaluating the diagnostic use of an image. The choice of evaluation tool depends on what is being investigated and the conditions. Receiver operating characteristics (ROC) are preferable when a specific diagnosis is being investigated, and where there is a known distribution of sick and healthy patients in the sample that is being assessed. When optimizing entire scanning protocols, several diagnoses are investigated and thus a more manageable approach to evaluate image quality would be for the observer to assess how well the anatomy is reproduced in the image. This approach is referred to as visual grading [51]. Visual grading does not reflect the ability of the radiologist to make the correct diagnosis, however, it has been shown to agree with methods based on ROC analysis [52, 53], and on calculations of the physical measures in specific cases [54, 55]. This shows that the ability to detect pathology is, to some degree, correlated to the reproduction of anatomical structures, which forms the basis of the visual grading approach.

There are several different approaches based on visual grading, for example, fulfilment of image quality criteria [56], visual grading analysis (VGA) [51], visual grading characteristics (VGC) [57] and visual grading regression [58]. VGA can be divided into relative or absolute VGA. The observers either compare the image quality of two or more images (relative VGA) or they grade the reproduction of anatomical structures in each image using a list of adjectives that describes different levels of visibility (a verbal rating scale) (absolute VGA).

Grading on verbal rating scales (VRS) produces ordered categorical data, also known as ordinal data. It is important that the statistical methods used are appropriate for this type of data, which has rank-invariant properties only. Rank-invariant properties means that the results of data analysis should not depend on the labels of the categories [59, 60]. VGA is unfortunately often used in conjunction with inappropriate statistics, where the numerical labels on the classification scale are treated as if they had a mathematical value.

The improper use of statistics in VGA has resulted in the use of other statistical methods that provide correct and reliable information. These methods are not necessarily new per se, but are not common in the field of radiology. The statistical methods used in this work are examples of such methods. Other examples are VGC, in which the fulfilment of criteria regarding the visualization of anatomical structures is evaluated and analysed with software normally used for ROC-based methods, and visual grading regression, in which logistic regression is used to analyse data. Logistic regression has the advantage of enabling the analysis of several variables simultaneously.

In the present work, rank-based statistical methods were applied to absolute VGA data. Inter-scale concordance [61, 62] was used because of its ability to identify a relation between radiation exposure and diagnostic image quality for an observer. This enables the identification of the minimum radiation exposures corresponding to different levels of diagnostic image quality. For this relation to be representative for the true distribution of assessments, however, it requires reasonably low intra-observer variability and only small variations in original image quality for the different patients included in the study.

1.5 Observer variability

Cohen's Kappa is today the recognized measure of reliability and is a single measure of agreement beyond the chance-expected agreement between and within observers. Despite its popularity it has various unsatisfactory features. The value of Kappa depends on the number of categories; as the number of categories decreases, Kappa increases (the higher the Kappa value, the higher the agreement). It is also assumed that there are unbiased pairs of assessments, which means identical marginal distributions, which is rarely the case in agreement studies [63].

The method of evaluating variability within and between observers used in this work was Svensson's method [64, 65]. Svensson's method has the ability to identify and measure the level of systematic disagreement, when present, separately from additional random variability. Paper II describes the method and demonstrates the kind of information that can be obtained with it. The method was then used to evaluate observer variability (Papers III-V).

1.6 Dose reductions in general

There are several ways of keeping radiation doses at a minimum. Concentrating only on tube current and image quality will not guarantee the lowest possible absorbed dose to the patient. For example, all X-ray examinations must be justified [66]. A radiologist should always be involved in determining whether a patient should undergo CT or not. Other diagnostic techniques not involving ionizing radiation, such as ultrasound and magnetic resonance

imaging (MRI), should be considered first. MRI should especially be considered as an alternative to high radiation exposure examinations.

It is important that the CT operators have sufficient time to prepare the patient for the scan, especially paediatric patients. For example, it is important that the patient is still during the examination in order to reduce the risk of having to repeat a scan. Positioning the patient correctly in the centre of the gantry (the isocentre) has always been important with regard to the bow-tie filter, for both optimal image quality and low patient skin dose. With the introduction of TCM, it has, however, become even more important as incorrect positioning of a few centimetres can result in over- or under-exposure in tube current [67, 68].

For routine cerebral CT examinations the recommendation is to use axial scanning mode (also known as incremental scanning mode) with a tilted gantry or a tilted head position, in order to avoid irradiation of the eye lens [69]. Also, multiple exposures for precontrast imaging should be reduced to a minimum when medically appropriate. A general recommendation regarding scanning modes is to use axial scanning for small scanning lengths, and helical scanning for larger scanning lengths [70], as helical scanning uses an over-scan at the end points of the scanned volume, resulting in greater radiation exposure of the patient than in axial scanning mode. Helical scanning, however, has the benefit of reduced scanning time, and thus reduced risk of motion artefacts. The benefit of the speed of the examination must thus be weighed against the extra dose. The need for multiplanar reconstructions or volume rendering of the scanned volume should also be considered as helical mode is the better choice for this reconstruction technique. Some manufacturers have introduced adaptive collimators to reduce the excess dose from over-scanning at helical scanning [71].

Regarding the collimation of the actual beam width (which is dependent on the detector configuration), as large a beam width as possible (with regard to the minimum nominal slice thickness needed) is often the most dose efficient regarding $CTDI_{vol}$. A broad beam reduces the number of rotations required, and thus the contribution from superimposed penumbras from each rotation. There is however a risk of increased DLP instead, especially for short scan lengths when using helical mode. Adjustments in detector configuration in order to minimize the radiation exposure with a wide beam should thus be done with regard to both $CTDI_{vol}$ and DLP. If the adjustment results in an increase of the scanning volume, it is also important to consider possible effects on radiation sensitive organs.

Pitch can be used to reduce the radiation exposure of the patient for some scanners. Increasing the pitch means increasing the table movement per rotation whilst the beam width remains the same, this leads to a reduction in the total radiation exposure of the patient. The image quality will however be affected. Greater pitch does not necessarily mean lower image quality, as the quality depends partly on the reconstruction algorithm used. However, increasing the pitch means an increase in the distance between the interpolation points which are used to calculate the image. A too high a pitch could reduce the ability to detect small objects. Some manufacturers have implemented automatic adjustment of the radiation output in order to maintain a certain radiation exposure regardless of the pitch.

In conventional X-ray examinations, the tube voltage is adjusted according to the size of the patient although this has not been common practise in CT. However, research indicates that the dose to paediatric patients can be reduced by lowering the tube voltage [72-74]. Using a low tube voltage for the scan projection radiograph (also known as the topogram, scout view,

scanogram, surview or pilot scan, depending on the manufacturer of the CT) is recommended by some [75], however, this could result in an erroneous estimate of the TCM if the actual examination is not performed at the same tube voltage. It is therefore recommended by at least one manufacturer to use the same tube voltage as will be used in the examination [50]. Radiation-sensitive organs should be taken into account when positioning the X-ray tube for the scan projection radiograph. For example, for an anterior-posterior overview, the X-ray tube is better placed facing the back of the patient (posterior-anterior). The absorbed dose to radiation-sensitive organs such as the eye lens, the thyroid and the breasts will then be lower [75].

Shielding radiation-sensitive organs located close to the examined area has been shown to reduce the organ dose resulting from scattered radiation [76, 77]. These organs are commonly the gonads and the thyroid, but other organs that can be shielded are the eye lenses and the breasts. A simple means of reducing the dose is to limit the volume examined, and not to add extra scan length "just in case".

Low image quality despite a high radiation exposure can often be explained by the use of inappropriate settings of the reconstruction parameters. For example, slice thickness affects the resolution and the noise in an image. A thin slice results in a higher image noise compared to a thicker slice if other parameters are fixed, however, increasing the slice thickness results in lower resolution. Finding a balance in slice thickness can be done on already performed examinations by reconstructing new images. One scanning parameter that could have a direct effect on image quality only is the gantry rotation time. It is commonly believed that, the faster the better, however, when a high radiation output is combined with a fast gantry rotation time, the X-ray tube might not be able to deliver the expected tube load. A slight increase in the gantry rotation time with a corresponding decrease in tube current could increase the image quality in such cases.

In order to identify scanning protocols in need of optimization, it is necessary to know which level of radiation exposure is appropriate for a specific type of CT examination. Diagnostic reference levels (DRLs) are values of $CTDI_{vol}$ and DLP for specific CT examinations, based on examination statistics from several hospitals. The third quartile of the distribution of dose values from different hospitals is often used to determine the DRL. This is referred to when establishing whether a hospital is using a high radiation exposure or not. DRLs are thus based on practices at other hospitals, not on the actual optimal level of radiation exposure. International recommendations for DRLs regarding adult scanning have existed for several years [29], although values for paediatric patients only exists on national level for a few countries [78, 79]. More DRLs, especially more recent values, are needed for both adult and paediatric patients.

The combination of high radiation exposure and high image quality is reason to investigate whether the tube current can be reduced without affecting the ability to diagnose. Small reductions in tube current (of the order of 5-10 %) have very little effect on image quality when the original image quality is considered high (especially with regard to a low level of image noise). Introducing small step-wise reductions in tube current clinically and evaluating the image quality retrospectively between reductions has been shown useful [80].

It is important that the scanning protocols defined for paediatric patients do not cover too wide an interval of indications, requiring different levels of image quality. For example,

examining the size of the ventricles in the brain in a follow-up examination of shunt-treated children is considered possible at very low radiation exposure [81, 82]. There should thus be a separate scanning protocol for such examinations.

New recommendations are published constantly, for the interested reader, there are several publications describing general dose awareness more thoroughly [83, 84].

2 Aims

The overall aim of the work described in this thesis was to find an optimization approach with which the absorbed dose to paediatric patients undergoing CT examinations could be minimized with regard to diagnostic requirements on image quality and observer variability.

The aims of the separate studies were:

- To investigate the effect of reduced tube current on the diagnostic image quality in paediatric cerebral multi-detector CT images (**Paper I**).

- To demonstrate a nonparametric statistical method that can identify and explain the components of observer disagreement (**Paper II**).

- To determine the highest acceptable noise index with regard to image quality for routine paediatric abdominal CT examinations (**Paper III**).

- To estimate the variability in results when using an optimization approach based on inter-scale concordance (**Paper IV**).

- To evaluate a 2D post-processing adaptive filter claiming to enable reductions in radiation exposure and thus the absorbed dose to the patient (**Papers III** and **V**).

3.1 The computed tomography scanner

To create CT images, data from thousands of projections around the patient is collected for each rotation, this data is denoted raw data i.e. data that has not been processed yet. All raw data used in this thesis originated from the same multi-detector CT, a Light Speed Ultra (GE Healthcare, Milwaukee, WI, USA) at the Department of Paediatric Radiology & Physiology at the Queen Silvia Children's Hospital in Göteborg, Sweden. This scanner is capable of collecting and producing up to 8 images per rotation. Minimum gantry rotation time is 0.5 s. It is equipped with a HiLight Matrix detector, characterised by 16 detector rows, each representing 1.25 mm at the isocentre, giving the opportunity to scan 20 mm anatomy per rotation. The minimum nominal slice thickness is 0.625 mm when a collimated beam width of 1.25 mm is centred over the lamella between two central rows of detectors. The HiLight Matrix detector system is based on polycrystalline ceramic technology, providing 99 % absorption efficiency. The CT software was upgraded in April 2003, providing the possibility of using TCM. The studies described in this thesis were performed after this upgrade.

3.2 Raw data collection

Raw data was retrospectively collected from clinically performed examinations. Paediatric abdominal CT examinations were the subject of interest in one of the studies (Paper III) and paediatric cerebral CT examinations in the others. The number of patients included in each study (see Table 1) was limited by time and the exclusion criteria. The criteria for exclusion were examinations including pathology that could disturb the evaluation of structures (e.g. covering the organ of interest), the use of non-routine parameters, and interference due to patient movement during the scan. Cerebral CT examinations performed with contrast medium enhancement, and abdominal CT examinations in which the timing of the contrast medium failed were also excluded.

Table 1. The distribution of patients included in each study according to age-based scanning protocols (m=months, y=years).

	0-5 m	6-11 m	1-5 y	6-10 y	11-14 y	>14 y
Paper I	3	1	5	8	5	3
Paper II	3	1	5	8	5	3
Paper III	-	-	-	10	10	-
Paper IV	-	-	10	10	10	-
Paper V	-	-	10	10	-	-

Patients older than 1 year of age that had undergone a cerebral CT examination had been scanned with the axial scanning mode, using a tube voltage of 120 kV, 1 s gantry rotation time, "Head" scan field of view, the soft reconstruction algorithm and 5 mm slice thickness. The same settings had been used for patients under 1 y with the exception of the scan field of view, which was "Ped head", and the gantry rotation time, which was 0.8 s in Papers I-II and 1 s in Papers IV-V. All patients had been scanned with a fixed tube current, see Table 2.

The patients that had undergone an abdominal CT examination had been scanned with a helical mode, using a tube voltage of 120 kV, 0.8 s gantry rotation time, the "Body" scan field of view, the standard reconstruction algorithm, pitch 1.35, 5 mm slice thickness and 2.5 mm increments. TCM had been used in all the abdominal examinations. The noise index values used in the original scans of the 18 patients ranged from 10.0 to 11.0. The mean $CTDI_{vol}$ and the mean DLP were 3.7 mGy and 117 mGy cm, respectively, for patients aged 6 to 10 y, and 4.7 mGy and 190 mGy cm, respectively, for patients aged 11 to 15 y (values were given for the 32-cm CTDI phantom).

3.3 Image noise simulations

A noise simulation program developed by GE Healthcare (Milwaukee, WI, USA) was installed on a separate research CT console. The software creates a copy of the original raw data, identifies the level of noise and then adds a random Gaussian noise distribution, corresponding to the size of the desired reduction in tube current, to the raw data. In this way, the artificial noise is included in the filtering and reconstruction of the new images. By simulating a lower tube current, the effects of a tube current reduction on the image quality can be compared with the original examination.

The software has been validated previously [46, 83], but was tested regarding measurements of mean pixel value, standard deviation and visual assessments of the reproduction of structures using a quality assurance phantom (section CTP 515 in Catphan 600, The Phantom Laboratory, Salem, NY, USA), see Figure 5. Four images were collected at each of the following tube currents: 200, 140, 80 and 40 mA. Noise was added to the four original images at 200 mA to simulate images at 140, 80 and 40 mA. Similarly, noise was added to the 140 mA images to simulate images at 80 and 40 mA, and to the 80 mA images to simulate 40 mA images. The mean pixel value and standard deviation within region-of-interests were determined at the positions illustrated in Figure 5. Evaluation of the numerical data with the paired t-test showed no significant differences (p>0.05). Two radiologists and a physicist visually compared the images side-by-side but could not separate the simulated images from the original.

New images were simulated for the paediatric patients that had undergone a cerebral CT examination, representing image quality at reduced tube currents at decreasing intervals of 20 mA from the tube current used clinically. Table 2 shows the range of tube currents used for cerebral CT examinations in each paper.

For paediatric patients undergoing an abdominal CT examination, TCM had been used instead of a fixed tube current. The TCM program used in this work modulates tube current between rotations. Based on the last scout view, and a chosen value of the noise index, the TCM program creates a list of tube currents for each rotation of the scan. A maximum and minimum tube current is set to avoid under- or over-irradiation of the patient due to, for example, the incorrect positioning of the patient, or extreme attenuation due to metal implants. As the noise simulation program first identifies the level of noise and then adds noise to simulate a reduction in tube current, an increase in noise index is accomplished by the corresponding reduction in tube current.
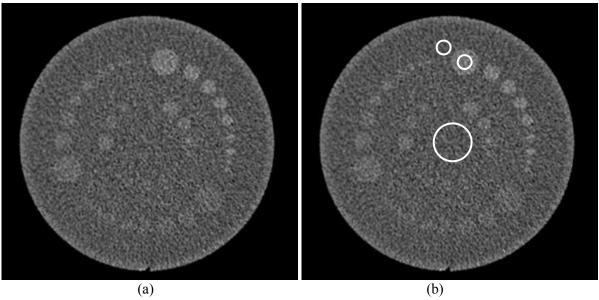
<center>(a)   (b)</center>

Figure 5. (a) Image of the quality assurance phantom (section CTP 515 in Catphan 600) used to test the noise-simulation software. (b) The white rings illustrate the region-of-interest where the mean pixel value and standard deviation were measured within all images.

Table 2. Intervals of tube current (mA), at discrete steps of 20 mA assessed in each study for each patient age group and for each level of the brain. The maximum tube current is the clinically used tube current for routine cerebral CT examinations, regarding age group and paper. Corresponding values of CTDI$_{vol}$ (mGy) are also given corresponding to a 16-cm CTDI phantom.

Age group	Level	Papers I and II		Paper IV		Paper V	
		(mA)	(mGy)	(mA)	(mGy)	(mA)	(mGy)
0-5 m	Upper	30-110	4-15	-	-	-	-
	Lower	30-110	4-15	-	-	-	-
6-11 m	Upper	30-130	4-17	-	-	-	-
	Lower	30-130	4-17	-	-	-	-
1-5 y	Upper	40-180	7-30	50-150	8-25	90-150	15-25
	Lower	60-200	10-33	60-160	10-26	100-160	17-26
6-10 y	Upper	40-200	7-33	60-160	10-26	100-160	17-26
	Lower	60-220	10-39	90-190	15-32	130-190	21-32
11-14 y	Upper	50-230	8-41	130-230	21-41	-	-
	Lower	70-250	12-44	150-250	25-44	-	-
>14 y	Upper	40-240	7-43	-	-	-	-
	Lower	60-260	10-46	-	-	-	-

The relation between the noise index and tube current for a patient can be described by the following equation, proposed by Kanal et al. [85]:

$$D(\varepsilon) = \frac{D}{(1+\varepsilon)^2} \qquad (5)$$

where ε denotes the change in noise index in percent (e.g. an increase of 20 % results in ε = 0.2), and D denotes the dose to the detector. The relation between detector dose and tube current is linear, and thus D can be replaced by the tube current or $CTDI_{vol}$. This equation can thus be rewritten as:

$$mA_2 = \frac{mA_1}{(\frac{NI_2}{NI_1})^2} \qquad (6)$$

where mA represents the tube current value and NI represents the noise index value. Equation 6, which converts a change in noise index to a change in tube current, was tested on water phantoms of different diameters (10 cm, 16 cm and 21.4 cm). The phantoms were scanned with abdominal scanning settings using TCM. Measurements of the mean pixel value and standard deviation were performed, and the original images were visually compared with the simulated images. Evaluation of the numerical data with the paired t-test showed no significant differences ($p > 0.05$) and the images could not be visually separated. Images with noise index values of 11, 12, 13, 14 and 15 were simulated for each patient using the noise simulation program.

3.4 Post-processing filter

A post-processing 2D adaptive filter, SharpView® CT (SharpView AB, Linköping, Sweden), was tested in Paper III and in Paper V to evaluate its ability to enable a reduction in radiation exposure. SharpView® CT analyses an image pixel by pixel to differentiate specific features. It identifies which pixels are part of the same structure and how the structure is oriented. This information is used to enhance the identified structures. It also filters out image noise using an adaptive (2D) filter in the spatial domain. The characteristics of this filter (smoothness/ sharpness) were evaluated by the radiologists prior to the studies in order to find the filter characteristics that were subjectively considered the most suitable for routine paediatric abdominal and cerebral examinations respectively. Examples of post-processed images are shown in Figures 6 and 7. All images (original and simulated) in each study (Papers III and V) were created in duplicates: one set of images was processed using the post-processing filter and the other was not.
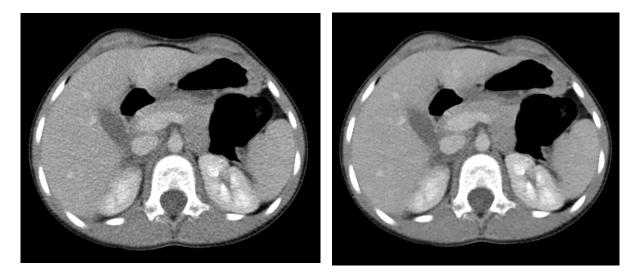
Figure 6. Images from a paediatric abdominal CT examination of a 7 year old girl showing the central level of the assessed stack. The image to the left shows a simulated image representing NI 15. The image to the right represents the post-processed (with SharpView CT®) copy of the image to the left.
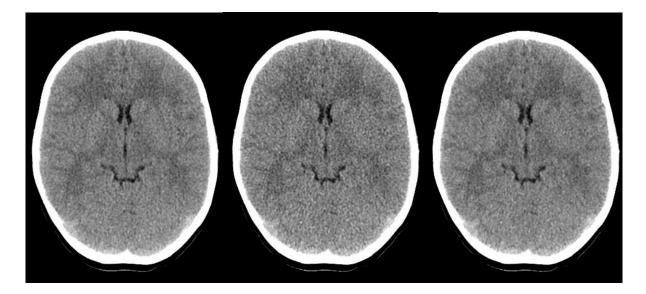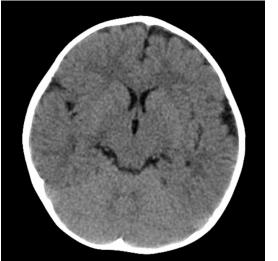

Figure 7. Images from a cerebral CT examination of a 9 year old boy where the image to the left represent the original image acquired at 26 mGy. In the centre: a dose-reduction of the image to the left representing 17 mGy and to the right: a post-processed (with SharpView CT®) copy of the image in the middle.

3.5 Image quality assessments

Images from two different levels in the brain were used for the image quality assessment of the cerebral CT examinations. The upper level contains the lateral ventricles and the basal ganglia, and the lower level contains the posterior fossa at the level of the 4th ventricle, see Figure 8. Each of these levels represents important areas for diagnosis, and contains both high- and low-contrast structures.
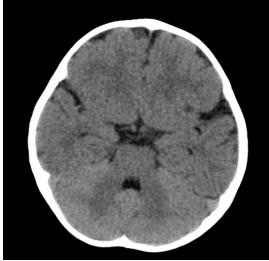
Figure 8. Images from a cerebral CT examination of a 1 year old girl representing the two levels of the brain used for the image quality assessments. The image on the left shows the upper level of the brain, showing the lateral ventricles and the basal ganglia, and the image on the right represents the lower level of the brain, showing the posterior fossa at the level of the 4th ventricle.

Eight images centred at the level of the extrahepatic portal vein were used for the assessment of image quality of the abdominal CT examinations. Figure 6 shows the central level of the assessed stack.

Images from three patients were duplicated in each study for test-retest evaluation to determine intra-observer reliability. The radiologists assessing the images were unaware of the duplicated images. Three observers were used in each study with the exception of Study IV, which only included two radiologists. All observers were either paediatric radiologists or had extensive experience of paediatric patients. All images were evaluated digitally. A cathode ray tube monitor was used in Studies I and II, and a liquid crystal display monitor was used in Studies III to V. Both monitors were medical monitors and calibrated according to DICOM part 14 [86]. The evaluation of the images took place in a quiet, secluded area, where the background light and sound level could be kept constant. The evaluation for Studies IV and V was conjoined and parts of the image material were shared.

All images and image stacks were viewed and evaluated using the computer software ViewDEX (Viewer for Digital Evaluation of X-ray images) [87]. ViewDEX is a Java program developed to present images in random order, without patient or scanning information, and with the possibility to answer the related questions on-screen, see Figure 9. The program allows the observers to scroll, cine, zoom, pan and change window settings. Each radiologist had a personal login ID, and the images were presented in a random order to each radiologist to avoid bias and to ensure that they did not discuss their findings with each other.
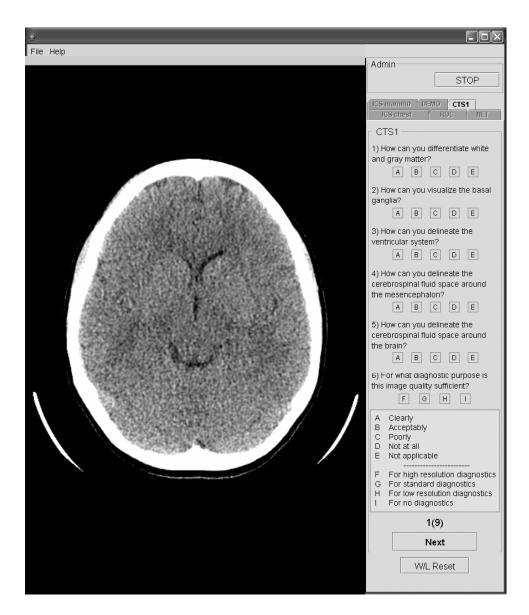
Figure 9. Illustration of how ViewDEX presents the images with the questions and possible responses on the monitor.

The images were assessed using a verbal rating scale in which the first six questions refer to the reproduction of anatomically important structures, and the last question refers to the overall image quality. In the cerebral studies, question 6 was only included in Studies IV and V, thus question 7 is denoted 'question 6' in Papers I and II. The following 7 questions and responses (A to I) were used.

1. How well can you differentiate white and grey matter?
2. How well can you visualize the basal ganglia?
3. How well is the ventricular system delineated?
4. How well is the cerebrospinal fluid space around the mesencephalon delineated?
5. How well is the cerebrospinal fluid space around the brain delineated?
6. How well can you visualize the vessels in the pentagon cistern?

A. Clearly
B. Acceptably
C. Poorly
D. Not at all
E. Not applicable

7. For what diagnostic situation is this image quality sufficient?

F. High-resolution diagnostics
G. Standard diagnostics
H. Low-resolution diagnostics
I. Not diagnostically useful

The reproduction of structures was rated as: "Clearly" = a completely distinct shape of the structure, "Acceptably" = a moderately but still acceptably reproduced structure, "Poorly" = a vaguely reproduced structure, and "Not at all" = the structure could not be discerned. The images were classified as being suitable for: "High-resolution diagnostics" = image quality suitable for indications such as cancer or small haemorrhages, "Standard diagnostics" = image quality suitable for a common trauma case, "Low-resolution diagnostics" = image quality suitable for a follow-up study of a shunt-treated patient, and "Not diagnostically useful" = the image quality was deemed to be of no diagnostic value.

The following 7 questions and responses (A to I) were used regarding the abdominal CT examinations (Paper III).

1. How well is the aorta delineated?
2. How well are the hepatic veins delineated?
3. How well is the liver delineated against the abdominal wall?
4. How well is the extra hepatic part of the portal vein delineated?
5. How well is the pancreas delineated?
6. How well is cruz diafragmatica delineated?

A. Clearly
B. Acceptably
C. Poorly
D. Not at all
E. Not applicable

7. Grade the diagnostic use of this image quality for a routine abdominal exam

F. Excellent
G. Sufficient
H. Insufficient
I. Not applicable

The structures used for evaluation were chosen to represent both high- and low-contrast structures. The responses describing how well the structures could be seen (A-E) were defined as previously described for the cerebral studies. In question 7, the radiologists were asked to

rate the diagnostic use of the overall image quality for a routine abdominal CT examination, using the terms: "Excellent", "Sufficient", "Insufficient" and "Not applicable". A routine abdominal CT examination refers to patients with indications involving abdominal trauma, abdominal pain, indistinct abdominal symptoms and follow-up examinations, when CT is considered the appropriate method.

The minimum tube current-time products (Papers I and II), noise indexes (Paper III) and tube currents (Papers IV and V) required to produce the different levels of image quality were identified with the use of inter-scale concordance (for a detailed description see Section 3.7.2). The value considered sufficient for high-resolution diagnostics of a cerebral CT examination was based on the highest of the minimum values identified by the three observers (two in Paper IV) for response A ("Clearly") and response F ("High-resolution diagnostics"). The highest of the minimum values for response B ("Acceptably") and response G ("Standard diagnostics" for the cerebral examinations and "Sufficient" for the abdominal examinations), were used to define the image quality sufficient for a routine examination (also referred to as standard diagnostics in the text). For an image quality sufficient for low-resolution diagnostics in the studies regarding the cerebral examinations, the highest minimum value identified by the observers for response B ("Acceptably") for question 3 (the reproduction of the ventricles) and response H ("Low-resolution diagnostics") was used.

3.6 The observers

Three observers (one paediatric radiologist, one paediatric neuroradiologist and a neuroradiologist with extensive experience of paediatric patients) assessed the images in the studies presented in Papers I and II. In the study presented in Paper V, the neuroradiologist was replaces with another paediatric radiologist. In Study IV (which aimed to repeat Study I in order to evaluate variability in the results) only the two paediatric radiologists from Study I were able to assess the new images. In the study described in Paper III, two experienced paediatric radiologists new to the approach, and the paediatric radiologist from the previous studies, assessed the images.

3.7 Statistical analysis

3.7.1 Svensson's method
To evaluate observer variability the non-parametric statistical method of Svensson was used [64, 65]. This approach takes into account the properties of ordinal data. The method makes it possible to identify and measure a systematic variability separately from random variability in the assessments.

Cross-classification tables, also known as contingency tables, were created containing the frequency distribution of the paired assessments between two observers X and Y (inter-observer evaluation), or two assessments made by the same radiologist of the images at different occasions X_1 and X_2 (intra-observer evaluation). The main diagonal, representing all agreeing pairs, is oriented from the lower-left to the upper-right corner, see Table 3. Agreement between the assessments was expressed as percentage agreement, PA. PA is simply the percent of agreeing pairs out of all pairs in the table. The presence of systematic disagreement between or within the observers is indicated by different marginal frequency

distributions. For example, in Table 3 a) the marginal distributions are 19, 128, 76, 25 and 0, 85, 88, 75, indicating a difference in assessment.

Two measures of systematic variability apart from the level of random variability were calculated: the relative position, RP, and the relative concentration, RC, both with possible values ranging from -1 to 1. The RP expresses the difference between the probability of the assessments of Observer X being distributed over lower categories (meaning higher image quality in this thesis) than the assessments of Observer Y and vice versa. This difference between the proportions of pairs over and under the main diagonal can be described as $P(X<Y) - P(Y<X)$, or $P(X_1< X_2) - P(X_2<X_1)$ for intra-observer evaluation. Zero values of RP indicate a lack of systematic disagreement in position. Table 3 a shows the paired assessments made by Observers X and Y. The positive RP (0.33) means that 33 % more images were classified as belonging to the higher categories than to the lower ones by Observer Y, when compared with the classifications made by Observer X. This means that Observer Y was more likely to assess an image as being of poorer quality than Observer X.

The RC expresses the difference between the probability that the assessments made by Observer Y (or occasion X_2) are concentrated on the classification levels more than does observer X (or occasion X_1), and vice versa. It can be seen on the marginal distributions in Table 3 b that Observer X tends to have a higher proportion of assessments in the central classification levels than Observer Y, hence a negative value of RC. Zero values of RC indicate a lack of systematic disagreement in concentration.

Table 3. Inter-observer assessments for Observer X compared to Observer Y regarding a question with responses F, G, H and I, where F represents the highest image quality and I the lowest. (a) shows a systematic shift in position (RP) on the rating scale as the data pairs are in general positioned over the main diagonal, meaning that Observer X tended to assess the images as being of a higher quality than Observer Y. (b) shows the systematic difference in use of the grading scale (RC) as the marginal distribution reveals that the grades used by Observer X are more concentrated to the central grades than by Observer Y.

(a)

Observer X

		F	G	H	I	Tot.
	I		6	45	24	75
Observer Y	H	1	57	29	1	88
	G	18	65	2		85
	F					0
	Tot.	19	128	76	25	248

PA = 48 %

RP = 0.33

RC = 0.00

RV = 0.01

(b)

Observer X

		F	G	H	I	Tot.
	I		1	28	50	79
Observer Y	H		12	52	7	71
	G	3	50	29		82
	F	4	8	1		13
	Tot.	7	71	110	57	245

PA = 64 %

RP = 0.00

RC = -0.17

RV = 0.01

In intra-observer disagreement, the value of RC is negative when a higher proportion of the images have central classifications at the first review than at the second (provided that X_1 and X_2 are on the x and y-axes, respectively). The cumulative marginal distributions can be illustrated in so-called Q-Q plots, see Figure 10. The dotted line in the figure represents the case of identical marginal distributions. RP is recognized here as the convex shaped curve under the dotted line. The curve diverges from the dotted line as RP diverges from zero. A negative RP results in a curve above the dotted line instead. RC is recognized as an S-shaped curve crossing the dotted line. As RC diverges from zero, the S-shape of the curve enhances. RP and RC often co-exist, which results in many different shapes of the Q-Q plot.

Random variability is estimated by calculating the variance between the ranks of the original data pairs in relation to the ranks of a so-called rank-transformable pattern of agreement (RTPA) between X and Y (or X_1 and X_2). This measure is called relative rank variance, RV. RTPA reflects the relation between X and Y (or X_1 and X_2) when no random variability exists. RTPA can be established if the pairs are ordered by ranks based on the marginal distributions alone. Table 4 a and b illustrate the RTPAs for Table 3 a and b, respectively. Possible values of RV range from 0 to 1, where non-zero RV indicates the presence of random variability, and the higher the value of RV, the more heterogeneous are the pairs in the original data.

The standard errors and 95 % confidence intervals for RP, RC and RV are estimated by means of the jackknife technique. For more details regarding the calculations of RP, RV and RC, the reader is referred to Svensson [64, 65].
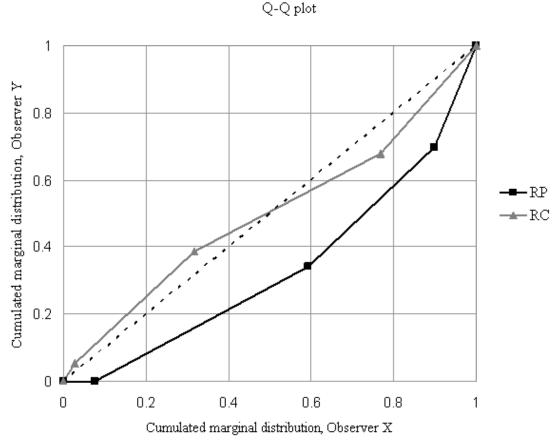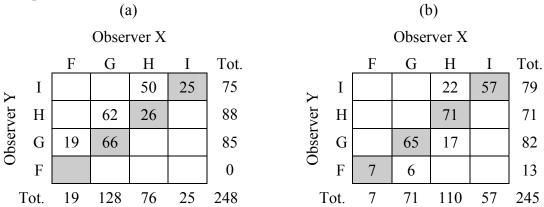

Figure 10. An illustration of the cumulated marginal distributions of two observers relative to each other in a so-called Q-Q plot.

Table 4. (a) and (b) illustrates the rank-transformable pattern of the cross-classification tables in Tables 3 a and b, respectively. The distribution of the pairs is based only on the marginal distributions, and is positioned in order from one corner to another (a so-called rank order). The new pattern matches the marginal distributions, but has no overlapping data pairs.

(a)

Observer X

Observer Y	F	G	H	I	Tot.
I			50	25	75
H		62	26		88
G	19	66			85
F					0
Tot.	19	128	76	25	248

(b)

Observer X

Observer Y	F	G	H	I	Tot.
I			22	57	79
H			71		71
G		65	17		82
F	7	6			13
Tot.	7	71	110	57	245

3.7.2 Inter-scale concordance

Inter-scale concordance focuses on the relationship between different scales. It is also appropriate for paired data, as in this work. Data pairs were created by combining the grade of the assessed image with the tube current (that actually used or simulated) for each structure and observer. The marginal distributions were then calculated by summing all pairs for each tube current and for each image quality. If the left graph in Figure 11 (a, "original data") is used as an example, the marginal distribution would be 5 for each tube current-time product, and 5, 24, 12 and 5 for the image qualities F, G, H, I, where F represents the highest image quality and I the lowest.

Figure 11. Illustration of (a) the original data and (b) the rank-transformable pattern of graph (a). Each ring in the graph represents an assessment, thus there are five rings (one for each patient) for each tube current-time product (clinically used tube current = 200 mA and simulated tube currents = 180 to 40 mA, gantry rotation time was 1s). The pattern in (b) facilitates the identification of the minimum tube current representing the different image qualities (represented by the labels F, G, H and I).

The pairs were then ordered by rank in accordance with the marginal distribution to form the RTPA (as described in the previous section). This results in intervals of tube current (or corresponding noise index) for each level of diagnostic image quality. The original data are a combination of true and disordered pairs due to observer variability and true variations in image quality, and hence the purpose of RTPA is to facilitate the identification of the minimum tube current that results in a certain diagnostic image quality without these effects. The definition of minimum tube current used in this work was the lowest level of tube current in the RTPA where all images were assessed with at least the regarded level of quality. If Figure 11b is used as an example, the intervals in tube current corresponding to the image qualities F, G, H, I are 200, 180-120, 100-60 and 40 mA, respectively. The higher the intra-observer variability in the original data, the less representative the RTPA is.

Intra-observer analysis was used in this work to estimate the observer variability. To keep variations in image quality at the evaluated exposures to a minimum, patient groups of similar size and development were used in all studies. Additionally, tube current modulation was used in the abdominal study.

4 Results

4.1 The effect of tube current in paediatric cerebral CT (Paper I)

The highest values of the minimum tube current-time products for each patient group and diagnostic image quality, according to the three observers, were identified. The corresponding $CTDI_{vol}$ values are given in Table 5. It can be seen that none of the low-contrast structures could be differentiated "Clearly" regardless of patient age. Neither were the criteria for "High-resolution diagnostics" considered fulfilled.

Table 5. The table presents the values of $CTDI_{vol}$ (mGy) corresponding to the highest minimum tube current-time product, among the three observers, required to produce a particular diagnostic image quality in the upper and lower levels of the brain. The values are given for each question and response level (A to C for questions regarding the visibility of specific structures, and F to H for the assessment of overall image quality).

Question	Grade	0-5 months Upper	0-5 months Lower	6-11 months Upper	6-11 months Lower	1-5 years Upper	1-5 years Lower	6-10 years Upper	6-10 years Lower	11-14 years Upper	11-14 years Lower	>14 years Upper	>14 years Lower
1	Clearly	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]
	Acceptably	15	-[a]	-[a]	-[a]	20	17	20	26	25	37	39	46
	Poorly	12	12	9	9	10	10	13	13	15	18	20	20
2	Clearly	-[a]	-[b]	-[a]	-[b]	-[a]	-[b]	-[a]	-[b]	-[a]	-[b]	-[a]	-[b]
	Acceptably	-[a]	-[b]	-[a]	-[b]	23	-[b]	26	-[b]	31	-[b]	39	-[b]
	Poorly	-[a]	-[b]	17	-[b]	13	-[b]	17	-[b]	18	-[b]	23	-[b]
3	Clearly	-[a]	-[a]	-[a]	-[a]	23	30	35	39	41	44	43	43
	Acceptably	15	12	12	9	10	13	17	17	15	21	20	20
	Poorly	7	7	7	7	7	10	10	10	8	12	7	10
4	Clearly	-[a]	-[a]	-[a]	-[a]	30	30	35	39	41	-[a]	-[a]	-[a]
	Acceptably	12	15	12	9	10	13	17	23	18	31	23	26
	Poorly	7	7	7	7	7	10	7	13	12	15	10	13
5	Clearly	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]
	Acceptably	-[a]	-[a]	15	12	17	17	23	35	31	44	35	43
	Poorly	9	12	7	7	10	10	17	26	15	31	17	30
6	High	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]
	Standard	-[a]	-[a]	17	17	23	20	26	30	31	31	39	35
	Low	9	9	12	12	10	10	17	17	21	21	26	20

[a] This image quality was not represented due to a low original image quality

[b] Not applicable in this level of the brain

The final results for a routine cerebral CT examination were based on the highest value regarding the grades "Acceptably" reproduced and "Standard diagnostics" for each age group. For patients between 1 and 5 years of age, the highest value of the minimum tube current-time product was 140 mAs (23 mGy) for the upper level of the brain, and 120 mAs (20 mGy) for the lower level of the brain. This means a possible reduction in dose of 22 % (from 180 mAs to 140 mAs) for the upper level, and 40 % (from 200 mAs to 120 mAs) for the lower level. For patients 6 to 10 years old, the highest value of the minimum tube current-time product was 160 mAs (26 mGy) for the upper level and 200 mAs (35 mGy) for the lower level, which

means possible reductions of 20 % (from 200 mAs to 160 mAs) and 9 % (from 220 mAs to 200 mAs). For patients aged 11 to 14 years, the possible reduction in radiation dose in the upper level of the brain was 17 % (from 230 mAs to 190 mAs (31 mGy)), while no reduction in radiation dose was possible in the lower level of the brain. For patients over 14 years old, the possible reduction in radiation dose was 8 % in the upper level of the brain (from 240 mAs to 220 mAs (39 mGy)). No reduction was considered possible in the lower level of the brain. For patients under 1 year of age, the original image quality was found to be insufficient for a routine examination.

When considering adjustments of radiation dose for high-contrast structures, the highest value was limiting regarding the criteria "Acceptably" reproduced ventricles (question 3) and "Low-resolution diagnostics" (question 6) for each age group. For patients older than 1 year, reductions between 33 % and 67 % were considered possible (33 % in the upper level of the brain for patients >14 years, i.e. from 240 mAs to 160 mAs (26 mGy), and 67 % in the upper level of the brain for patients 1 to 5 years old, i.e. from 180 mAs to 60 mAs (10 mGy)). For newborns, no reduction was considered possible.

4.2 Observer variability analysed with Svensson's method (Paper II)

The percentage agreement in the intra-observer evaluation ranged from 50 % to 83 % for Observer 1, from 63 % to 83 % for Observer 2 and from 60 % to 87 % for Observer 3. None of the RP or RC values were significant for Observer 1 with the exception of a negative value of RP for question 4 in the lower level of the brain, i.e. the observer was more likely to assess an image as being of a higher image quality on the second occasion than on the first. Regarding Observers 2 and 3, the majority of the RP values indicated that both radiologists were more likely to assess an image as being of poorer quality on the second occasion.

Regarding the evaluation of inter-observer variability, the comparison between Observer 1 and 2 showed that the disagreement consisted of Observer 2 systematically grading the images as being of poorer quality than to Observer 1 for all questions except question 2 in the upper level of the brain. Regarding RC, Observer 2 systematically concentrated the assessments more than Observer 1 for all questions except question 6. For the lower level of the brain, Observer 2 continued to concentrate the assessments, whereas the opposite relation was found for RP, meaning that Observer 1 systematically graded the images as being of poorer quality than to Observer 2.

Regarding the disagreement between Observers 1 and 3, Observer 1 systematically tended to concentrate the assessments more than Observer 3 in the upper level of the brain. RP showed no distinct pattern of disagreement regarding all the questions together, but was significant for individual questions. There was less systematic variability in concentration in the lower level of the brain than in the upper level of the brain. The distribution of RP in the lower level of the brain was similar to that in the upper level.

Analysis of the disagreement between Observer 2 and 3 showed that Observer 2 systematically concentrated the assessments more than Observer 3 for both the upper and lower levels of the brain. There was also systematic variability regarding position for questions 3, 4 and 6, where Observer 2 was more critical. For questions 2 and 5, Observer 3 was significantly more critical. The RP values were similar in both the upper and lower levels of the brain, with the exception of question 4.

4.3 The effect of noise index in paediatric abdominal CT (Paper III)

For patients aged 6 to 10 years, the highest possible value of noise index (NI) was considered to be 11 for a routine abdominal examination (the value was limited by the hepatic veins by Observer 2 and by the pancreas by Observer 3). For patients aged 11 to 15 years, NI=12 was considered to be the highest possible value for a routine abdominal examination (limited by the hepatic veins by Observer 2).

In the post-processed images NI=12 was considered to give the lowest acceptable image quality for a routine abdominal scan for patients aged 6 to 10 (limited by the visibility of the pancreas by Observer 3). For patients aged 11 to 15 years, NI=13 was considered sufficient (limited by the visibility of the hepatic veins by Observer 2).

Regarding the analysis of systematic disagreement, Observer 1 assessed 21 % (3 out of 14) of the image stacks to be of a lower image quality at the second assessment than the first assessment regarding question 2 (RP=0.21). Observer 2 assessed 41 % (7 out of 17) of the image stacks to be of a higher image quality at the second assessment than the first for question 5 (RP=-0.41). Observer 3 had significant RP for questions 1, 4 and 7 indicating a trend to assess the image as being of a lower quality the second time. The RV values were negligibly small with the exception of question 4 for Observer 3 (RV=0.1) and question 5 for Observer 1 (RV=0.16).

In the comparison of the assessments between observers, Observer 1 and 2 had a percentage agreement of 39 to 63 %. The disagreement between the observers could mainly be explained by Observer 2 assessing the images as being of a lower quality compared to Observer 1. However, Observer 1 concentrated the responses on the classification scale more than Observer 2. The percentage agreement between Observer 1 and Observer 3 was slightly higher (46 to 82 %). Regarding question 1, Observer 1 significantly assessed the images as being of a lower quality than Observer 3, while the opposite was seen regarding questions 4, 5 and 7. Observer 2 and 3 had a percentage agreement between 39 and 61 %. Observer 2 noticeably assessed the images to be of a lower quality than compared to Observer 3.

4.4 Variability in the results (Paper IV)

The variability in the results when using the optimization approach based in inter-scale concordance was investigated by evaluating image quality in new paediatric cerebral CT images and comparing the results to that in Study I. Table 6 shows the resulting percentage differences in final tube current between Study I and Study IV for standard diagnostics and low-resolution diagnostics.

Table 6. The difference in final tube current between Study I and Study IV in the upper and lower level of the brain for each age group

Age group (years)	Standard diagnostics		Low-resolution diagnostics	
	Upper level	*Lower level*	*Upper level*	*Lower level*
1 to 5	7 %	17 %	50 %	33 %
6 to 10	0 %	-15 %	-20 %	10 %
11 to 14	11 %	-8 %	15 %	31 %

The intra-observer analysis showed ranges of percentage agreement from 50 to 89 % and 59 to 100 % for Observer 1 in the upper and lower levels of the brain, respectively, and 56 to 78 % and 72 to 89 % for Observer 2, respectively. Calculations of RP showed a systematic shift for questions 3 and 4 in the lower level of the brain for Observer 2. Calculations of RC showed no significant systematic difference in concentration on the scale for the observers. RV ranged from 0 to 0.08.

The range of inter-observer percentage agreement was 53 to 61 % and 39 to 68 % in the upper and lower levels of the brain, respectively. Observer 2 was more critical of the quality of the images of the upper level of the brain than Observer 1 regarding questions 1-3 and 7. There were, however, no significant differences in use of the scale. Observer 2 was more critical of the images of the lower level of the brain regarding questions 1 and 7, while the opposite was found for questions 4 to 6. Observer 1 concentrated the assessments more on the scale than Observer 2 for questions 1, 3 and 7. The RV values were negligible, with the exception of those for questions 4 (RV 0.11) and 5 (RV 0.09) in images of the lower level of the brain.

4.5 Evaluation of the post-processing filter (Papers III and V)

The post-processing filter was tested in both the abdominal CT examinations (Paper III) and in the cerebral CT examinations (Paper V). For the abdominal CT examinations the responses "Excellent" and "Sufficient" were more common for the filtered images than in the case of the unfiltered images. For patients aged 6 to 10, NI=12 was considered to give the lowest acceptable image quality for a routine abdominal scan compared to NI=11 for the unenhanced images. This represents a difference in $CTDI_{vol}$ of 16 %. For patients aged 11 to 15 years, NI=13 was considered sufficient compared to NI=12. This represents a difference in $CTDI_{vol}$ of 15 %.

For the cerebral CT examinations two radiologists assessed the basal ganglia as poorly reproduced in the unenhanced images of patients aged 1 to 5 y in the upper level of the brain but acceptably reproduced in the corresponding enhanced images. For patients aged 6 to 10 y, the requirements for image quality sufficient for routine cerebral CT examinations of the enhanced images were fulfilled at $CTDI_{vol}$ =23 mGy and 28 mGy in the upper and lower levels of the brain, respectively, compared with $CTDI_{vol}$ =27 mGy and 32 mGy for the unenhanced images, respectively, thus indicating possible reductions in radiation exposure of 15 % and 13 %.

For patients aged 1 to 5 y, both enhanced and unenhanced images were sufficient for low-resolution diagnostics at the lowest simulated dose level in the lower level of the brain (17 mGy). In the upper level of the brain, no possible dose reduction was found. For patients aged 6 to 10 y, low-resolution diagnostics of the unenhanced images was considered possible at $CTDI_{vol}$ =20 mGy and 25 mGy in the upper and lower levels of the brain, respectively, and at 17 mGy and 22 mGy in the enhanced images, respectively. This indicates possible reductions in radiation exposure of at least 15 % and 14 % since 17 mGy and 22 mGy represent the lowest levels assessed in the enhanced images.

5 Discussion

The optimization approach based on inter-scale concordance was tested for paediatric cerebral and abdominal CT examinations (Papers I and III). It was also used in a repetitive study regarding paediatric cerebral examinations in order to estimate the variability in results when using the approach (Paper IV). The approach was also used to evaluate a post-processing filter (Paper V). The evaluation of observer variability (Paper II) is an important part of the approach as it provides a measure of both the disagreement between the observers and the individual variability. A large observer variability means that the relation between the scales (diagnostic image quality and radiation exposure), represented by the RTPA, is less representative for the true distribution.

5. 1 Analysis of the results

5.1.1 Paper I

The results showed that the tube current for a routine cerebral CT examination for patients between 1 and 10 years of age could be lowered by approximately 20 %. The tube current for patients over 10 years of age could however not be reduced. For patients under 1 year old, the results showed that the original image quality was insufficient for a routine examination. Discussions regarding these findings of insufficient image quality in clinical use revealed that many of the radiologists had had the impression that the images were of poor quality for this particular age group. The results were thus useful in allowing this situation to be remedied, by increasing the gantry rotation time from 0.8 to 1 s in order to improve the diagnostic image quality. The poor image quality was probably due to previous attempts to restrict the dose to these patients. The approach can thus be of help in identifying poor image quality.

As there are no DRLs for paediatric patients in Sweden yet, the results obtained in this study were compared to those in a national survey of doses in the UK from 2003 [78]. The resulting $CTDI_{vol}$ values found in this study for a routine cerebral CT were close to the values of the first quartile (25th percentile) in the UK survey. The first quartile values were 17 mGy (cerebrum) and 21 mGy (posterior fossa) for patients 0-1 years of age, while the results in Paper I indicated that 15 mGy was considered insufficient and an increase to 18 mGy for both the upper (representing the cerebrum) and lower (representing the posterior fossa) levels of the brain was necessary. For a 5-year-old patient, the values at the first quartile in the UK survey were 21 mGy (cerebrum) and 30 mGy (posterior fossa), while the corresponding values in this study for patients 1-5 years of age were 23 mGy and 20 mGy. For a 10-year-old patient, the first quartile values were 29 mGy (cerebrum) and 40 mGy (posterior fossa) in the UK survey, compared with 26 mGy and 35 mGy in the present study. As the DRLs represent the third quartile, the values found in this study are well below these reference levels.

Scanning protocols especially designed for follow-up examinations of shunt-treated patients were already in use at the department prior to this study; the radiation dose levels being up to 50 % lower than those for routine cerebral examinations. These radiation dose levels agreed relatively well with the results of this study for patients older than 1 year. The results in this study are regarded as a verification of the existing shunt protocols.

5.1.2 Paper II

The evaluation of inter-observer variability showed that the paediatric radiologist was more critical of the overall image quality, while the neuroradiologists were more critical of the reproduction of the basal ganglia. As only three radiologists participated in the study, it is not possible to conclude whether the results are related to their experience.

The statistical method used was able to identify and separately measure a presence of bias apart from additional individual variability within and between the radiologists which is, at the time of writing, not possible with any other statistical approach suitable for paired, ordinal data.

5.1.3 Paper III

The highest possible value of noise index that can be used while ensuring an image of sufficient quality for a routine abdominal examination of patients aged 6 to 10 years was found to be 11. This represents a mean value of $CTDI_{vol}$ of 3.7 mGy (based on the 8 patients in this age group). For patients aged 11 to 15 years, NI=12 was considered the highest possible, representing a mean $CTDI_{vol}$ of 3.9 mGy (based on the 10 patients in this age group). $CTDI_{vol}$ values were given for the 32-cm CTDI phantom.

Comparisons with other studies show similar results. Singh et al. [80] showed that it was possible to increase noise index from 8 to 10 for patients weighing 27 to 45 kg, and from 10 to 12 for patients weighing 46-100 kg. When comparing the patient age group 6 to 10 y in this study with the lower weight group, and the patient age group 11 to 14 y with the higher, the results are quite similar. Honnef et al. concluded that the optimal value of $CTDI_{vol}$ for patients more than 30 kg were 5.9 mGy (n=6) [88]. This value was given for a 16-cm CTDI phantom. Dividing values given for a 16-cm CTDI phantom by two will approximately equal the corresponding $CTDI_{vol}$ value for a 32-cm CTDI phantom. This would result in approximately 3 mGy which can be compared with the values in the present study, which were slightly higher. Different results were reported by Verdun et al. [89], who proposed $CTDI_{vol}$ values of 6.7, 9.4, 15.9 and 24.5 mGy for the weight classes: 2.5-5, 5-15, 15-30 and 30-50 kg, respectively (given for a 16-cm CTDI phantom). Recalculating the values for the two heavier weight classes to a 32-cm CTDI phantom in order to compare to the values in this study would give approximately 8 and 12 mGy, which are noticeably higher.

The results of post-processing the image stacks with an adaptive filter indicated that a higher noise index could be used (corresponding to a reduction in absorbed dose of ~15 %). Because of the approach used, it was not possible to determine whether there was any loss of anatomical information, however, the filter was tested on a quality assurance phantom (section CTP 515 in the Catphan 600 phantom) for images with noise values ranging from 11 to 15, without any visible loss of contrast or details between the post-processed images and the original image.

5.1.4 Paper IV

The comparison of the outcome from Study I and Study IV showed mixed results. For standard diagnostics, the final tube currents for each age group varied by 0 to 11 % compared with the previous results for the upper level of the brain. For the lower level of the brain, the final tube currents for each age group varied by -15 to 17 %. This indicates a variability of less than 20 %. Some difference was expected considering the subjectivity of perceived image

quality, and that four years had elapsed between the studies. This indicates a consistency of the results, at least at this level of image quality.

The results for the rating "Low-resolution diagnostics" differed more. The tube current was reduced in steps of 20 mA, but the reductions should have been made in terms of percent, as a 20 mA reduction results in a much higher percentage difference at the lower tube currents. For patients aged 1 to 5 y, the 20 mA difference between the results from Study I and Study IV represented a 33 % difference at the lower level of the brain. In the upper level of the brain, 90 mA was considered the minimum tube current in Study IV, compared with 60 mA in Study I, resulting in a difference of 50 %. For the patients aged 11 to 14 y, the difference between the two studies was 31 % in the lower level of the brain. This can partly be explained by the fact that the range of tube current assessed in Study IV was 150-250 mA, compared to 70-250 mA in Study I. Acceptable image quality for "Low-resolution diagnostics" in the lower level of the brain was obtained at 130 mA in Study I, while in Study IV, the lowest simulated tube current was 150 mA.

Assessing a too short range of tube currents can affect the results as the ending tube currents (both the highest and lowest represented) cannot be adjusted for random variability. For example, if a single one out of all images is assessed as being of a lower image quality than the others, this assessment has to be represented at the last tube current in the RTPA. In this case, the minimum tube current for the higher image quality cannot be the ending tube current since all images do not have the regarded level of quality. In the example given in Table 7 b, the minimum tube current for the image quality "H" is 80 mA because one of the images was graded as "I" at 60 mA. If tube current steps below 60 mA would have been used, this variability could perhaps have been adjusted for, and the minimum tube current could have been lower.

Table 7. (a) An example of the original distribution of data pairs of tube current (mA) and assessed image quality (F to I), and (b) rank ordering of the data pairs in Table 7 a.

(a) Original data						(b) Rank-ordered data					
mA	F	G	H	I	Total:	mA	F	G	H	I	Total:
60		1	3	1	5	60			4	1	5
80		1	4		5	80			5		5
100		2	3		5	100		4	1		5
120		5			5	120		5			5
140	2	3			5	140		5			5
160	1	4			5	160	3	2			5
Total:	3	16	10	1		Total:	3	16	10	1	

5.1.5 Paper V

The purpose of a post-processing filter is to improve the delineation of the structures and to suppress noise. The approach used in this study did not evaluate the ability of the radiologist to establish the correct diagnose. However, studies on phantoms indicate that more details are detected, using the SharpView® filter [90].

The post-processing filter is limited by the information in the original image; too low radiation exposure, given a signal that cannot be identified above the noise, cannot be

compensated for by the filter. It is thus important to use the filter wisely. This study shows that the filter enables a possible reduction in radiation exposure of ~15 %. This is a relatively low value because of the comparison to optimized values. The benefit of the filter can be assumed as higher at departments without optimized scanning protocols. The credit of the reduction can however not all be given the filter in that case, as the radiation exposure probably would have been reducible even without a filter.

5.2 Sources of errors

The number of paediatric patients used in this work has been small. The number has been limited by both the exclusion criteria and the need to have different age groups. Regarding the number of patients, the statistical methods used have no assumptions or approximations for being valid other than for handling paired ordinal data; the number is thus only limited by being representative for the group. For conclusions regarding all age groups, findings were discussed relative clinical impressions. There were also a small number of radiologists assessing the image quality in the studies. Three radiologists were used in all studies except for Study IV, where only two radiologists participated. Three radiologists were considered sufficient to represent the radiologists at the paediatric radiology department where the results were indented to be implemented. The results from Study IV were only intended to be compared to previous results in order to evaluate variability. The medically responsible radiologist for the CT at the department participated in all studies.

The exclusion criterion of not including images with pathology was based on the possibility that pathology could conceal or overlap structures of interest, making assessments impossible. As the original images were the origin of the simulated images, it was appropriate to use original images that could guarantee a high level of useful data. Since paediatric patients are not examined with MDCT unless there are strong indications, the number of patients without pathology was clearly limited. It can however be argued that excluding more complicated cases facilitates the distinction of structures, and may thus affect the outcome and validity of the study. In a visual grading study, however, different structures of various sizes and contrast are used to allow the results to be generalised to the range of contrast and resolution found in clinical images for a group of individuals. Allowing pathology to change the conditions of these structures randomly among the individuals would undermine the possibility of comparing changes in image quality between and within different age groups.

The analysis of image quality has been based on two images in the cerebral examinations (representing 2×0.5 cm) and on 8 images in the abdominal examinations (representing 4 cm). It is thus not the entire structures of interest that have been evaluated. This could be a source of error if the visibility of a structure was different in the other images. As the results were based on the evaluation of several structures, the risk of this affecting the final results were reduced.

Another source of error could be mispositioning of the patients in the gantry isocentre. Upon viewing the images prior to each study, no image contained a miscentring of such a size that it could have had a large effect on the final results.

Discussion

5.3 Sources of errors due to the approach

It is very important to identify sources of error that origins from the approach, or pitfalls in the use of the approach that could affect the results. Using an optimization approach based on inter-scale concordance is a balance as it means that the categories on the ordinal scale are given a meaning more than just ordering the image quality. The benefit of using statistical methods that evaluate data based on change between two groups is that the observers are allowed to have different interpretations of the categories, and it does not affect the final result. Using inter-scale concordance, as in this work, requires that the radiologists interpret the categories similarly. It does however not require that they are of the same opinion regarding the image quality as the approach takes this into account. Giving more meaning to the categories than just being ordering, simplifies the finding of a specific level of image quality that is considered sufficient for the purpose of the examination. Much effort was put on explaining and discussing the criteria with the radiologists. Prior to all studies in this thesis, there were discussions regarding the structures of interest and how to interpret the categories, there were also sessions of viewing and assessing images together.

One source of error was the variability in the assessments of the observers. This was especially important as the results in this thesis were based on the RTPA. The observer variability is an indicator of how well the RTPA represents the true relation between the scales. Significant RP and RC values in the intra-observer analysis indicated that the radiologists shifted in the way they assessed the images. Whether this was due to true changes in opinion or changes in interpretation of the category was difficult to evaluate. Perhaps these shifts were due to stress and fatigue, which can affect the way in which the radiologists assess the images [91]. As significant RP and RC values mostly were seen in the first study, it was perhaps most likely due to insufficient training of assessing the visibility of structures prior to the study. These shifts were mostly small and only for a few images thus indicating a small effect on the final results. Furthermore, there have been no corrections regarding multiple testing in the intra and inter-observer evaluations. RV (the measure of random variability) was mostly well below 0.1 but exceeded this value on a few occasions. What lacks today is experience on what specific value of RV that is considered too high for the RTPA to be representative.

Figure 2 in Paper I shows the original data and corresponding RTPA ('ordered data') of all three observers. The observer variability for each of the observers regarding this question was small, meaning that RTPA was representative. If assuming that the original data represents the images from a real clinical situation and that the assessments represent actual opinions on whether the image quality is suitable or not for a routine cerebral examination, then the application of the minimum tube current means that some of the images would be considered as being of a lower quality than intended. This brings up the question if the resulting radiation exposure levels in this thesis are too low. It should be kept in mind that obtaining an image quality directly below acceptable reproduction does not automatically mean the need for a rescan. A rescan of a child will most likely require more than that. Comparing the results in Paper I with the DRLs from the UK in 2003 [78] showed that the results approximately were at the level of the first quartile. This means that 25 % of the users in the UK use even lower radiation exposure levels. When comparing the results in Paper III with the results from other optimization studies, there was a relatively good agreement with two of the studies. This indicates that the results gained with this optimization approach are relevant.

There is a risk of bias in assessing the overall impression of image quality as radiologists tend to recognize and favour their old settings [43]. However, this is believed to be of minor concern as the radiologists in this thesis already were accustomed to a wide range of image noise.

The evaluation of the inter-observer variability showed significant difference in how the radiologists assessed the images. The resulting minimum tube currents for the observers were also different. There are evidently differences between the assessments of the observers, presumably in opinion. The strategy of limiting the reduction in dose according to the most critical observer was used so that no radiologist at the department would perceive the general image quality as poor or inadequate. However, the approach of using the most critical observer should be used with caution as it is not acceptable for one radiologist to differ too much regarding requirements on image quality and thus radiation exposure. The evaluation of inter-observer disagreement should serve as the basis for discussions regarding the perception of image quality after the study.

A limitation in the evaluation of the minimum radiation exposure was identified in Paper IV at the ending tube current. The lack of a sufficient range of tube currents made it difficult to evaluate the true variability in results in the lower levels of the tube current. This could have been avoided by using a sufficient range of the radiation exposure scale. The reductions should also have been made in terms of percent instead of fixed steps of 20 mA. This resulted in higher percentage difference for disagreeing results at the lower tube current levels than at the higher. According to Mayo et al. [92], the intra observer variability increases at low levels of radiation exposure, perhaps this also could contribute to the larger variability in results for the lower regions of tube current.

Another source of error is the estimation of observer variability based on the assessments of the test-retest images. It would be preferable to evaluate the variability directly on the original data. A suggestion is to divide the tube current scale for the original data into intervals according to the RTPA. If the data from this work were used, this means creating 4×4 cross-classification tables, which Svensson's method could be applied directly upon. The intervals should be based on where the majority of the assessments in the RTPA are. If using Figure 11 b in section 3.7.2 as an example, the tube current scale would be divided into 200 mA = F, 180-100 mA = G, 80-60 mA = H and 40 mA = I. Applying these intervals on the original data (Figure 11 a) would give Table 8. If Svensson's method were used, the information in Table 9 would be obtained.

The information in Table 9 is that the disagreeing pairs (33 %) are explained by random variability and a small systematic difference in position. The variability in this data is considered sufficiently low to conclude that the RTPA is possible to use. What also could be added to the evaluation is the *measure of disorder* (denoted "D"), which also is defined by Svensson [62]. This is a measure of discordance, which equals the proportion of disordered pairs relative the RTPA. Future investigations regarding the most accurate and efficient way of estimating the variability in the original data are necessary.

Discussion

Table 8. A cross-classification table where the tube current scale has been divided into 4 intervals and applied to the data in Figure 11a.

The distribution of assessments

The tube current scale	F	G	H	I	Tot.
I			1	4	5
H		2	7	1	10
G	4	18	3		25
F	1	4			5
Tot.	5	24	11	5	

Table 9. Information obtained when using Svensson's method on the data in Table 8. PA is the percentage agreement, RP is the relative position, RC is the relative concentration, RV is the relative rank variance and D is a measure of disorder.

		Confidence intervals
PA	67 %	-
RP	-0.02	-0.13 to 0.09
RC	0.00	-0.15 to 0.15
RV	0.01	0.00 to 0.03
D	0.03	-

The approach used in this thesis is suggested as a simple tool for optimization of CT scanning protocols. Claiming that the approach is simple requires that the department has access to a noise simulation software. In the moment of writing, this technique, whether it is simulations onto raw data or directly onto the images, is not easily accessible. The technique is however published [47] and there are manufacturers that have started to integrate or to separately sell this technique. The interest for this technique is large and will most likely be more easily attainable in the future.

5.4 Future research

More research and experience regarding the optimization approach based on inter-scale concordance is required. For example, more appropriate tests of the consistency of the results for the lower tube current levels should be conducted. A study testing the findings from Mayo et al. [92] regarding higher intra-observer disagreement at lower radiation exposures would also be interesting and informative. Further investigations regarding the appropriate estimate of the variability in the original data relative RTPA is necessary. It is also necessary to evaluate which observer variability is considered too high.

In the moment of writing, the CT used in all studies has been replaced with a newer CT equipped with iterative reconstruction and dual energy scanning. The efforts of this thesis have not only been of use for the old CT but it has also enabled a chance to compare the new technique, and thus new radiation exposure levels, with the old CT and settings. New research regarding optimal settings of the iterative reconstruction for paediatric patients is already in process. The results regarding the radiation exposure levels found in this thesis have perhaps short validity; the approach however, can hopefully be useful for a longer period.

6 Conclusions

The effect of tube current on diagnostic image quality in paediatric cerebral CT examinations was established in Paper I. The use of inter-scale concordance enabled the identification of minimum $CTDI_{vol}$ values to reproduce different levels of diagnostic image. The study showed that reductions in radiation exposure were possible for patients 1 to 10 years old. No reductions were concluded possible for patients above 10 years old. The original image quality for patients under 6 months of age was found to be inadequate for acceptable reproduction of low-contrast structures. The study also showed that it was possible to further reduce the radiation exposure for shunt-treated patients.

The analysis of intra- and inter-observer variability using Svensson's method was in focus in Paper II. Svensson's method enabled the identification of systematic disagreement and the level of additional individual variability in data from Study I. The obtained information is, at the time of writing, not attainable by any other statistical approach suitable for paired, ordinal data.

In Paper III the approach based on inter-scale concordance was used to identify the highest possible values of noise index (representing the lowest possible absorbed doses) to obtain images of sufficient diagnostic quality for a routine paediatric abdominal CT examination. The results indicate that for patients aged 6 to 10 years (n=8), NI=11 (representing a mean value of $CTDI_{vol}$ of 3.7 mGy) was sufficient for acceptable reproduction of abdominal structures and thus a sufficient image quality for a routine abdominal CT examination. For patients aged 11 to 15 years (n=10), NI=12 (representing a mean value of $CTDI_{vol}$ of 3.9 mGy) was considered sufficient. The $CTDI_{vol}$ values were given for the 32-cm CTDI phantom. The possibility of reducing the radiation exposure using a 2D post-processing adaptive filter was also investigated. For patients aged 6 to 10 years, NI=12 was sufficient for acceptable reproduction of abdominal structures when using the filtered images (representing a mean value of $CTDI_{vol}$ of 3.1 mGy). This suggests a possible reduction in mean absorbed dose of 16 %. For patients aged 11 to 15 years, NI=13 (representing a mean value of $CTDI_{vol}$ of 3.3 mGy) was sufficient, which suggests a possible reduction in mean absorbed dose of 15 %.

In Paper IV the variability in results when using the optimisation approach based on inter-scale concordance was estimated. A less than 20 % difference in the results for standard diagnostics was found between the studies (Study I and Study IV). This indicates a consistency of the results, at least at this level of image quality. Variability in the results for acceptable reproduction of high-contrast structures was difficult to estimate due to insufficient representation of tube currents at the lower levels of the tube current scale. Conclusions regarding how to make the best of the approach were to reduce tube current in steps of percent and to assure that the tube current scale used reached sufficiently low tube currents in order to let observer variability to be adjusted on the scale. The steps of the approach and experiences regarding the methodology are summoned in the Appendix.

Paper V further investigated the 2D post-processing adaptive filter. For patients aged 6 to 10 y it was found that dose reductions from $CTDI_{vol}$ = 27 mGy to 23 mGy (15 %) in the upper level of the brain, and from 32 mGy to 28 mGy (13 %) in the lower level of the brain, were possible for routine cerebral CT examinations. For low-resolution diagnostics, it was found

that reductions from $CTDI_{vol}$ = 20 mGy to 17 mGy (15 %) in the upper level of the brain, and from 25 mGy to 22 mGy (12 %) in the lower level of the brain were possible. Further reductions may have been possible as 17 and 22 mGy were the lowest tube current levels assessed. For patients 1 to 5 y, the results for standard diagnostics in the upper level of the brain and for low-resolution diagnostics in the lower level of the brain were inconclusive. For standard diagnostics in the lower level of the brain and for low-resolution diagnostics in the upper level of the brain, it was found that no reductions were possible.

The overall aim of the work described in this thesis was to find an optimization approach with which the absorbed dose to paediatric patients undergoing CT examinations could be minimized with regard to diagnostic requirements on image quality and observer variability. The approach based on inter-scale concordance has been shown useful for identifying excessive radiation exposure and insufficient image quality, but also for identifying and estimating observer variability. The low variability in results for standard diagnostics indicates that the approach is consistent at least at that level of image quality. Further investigations and applications of the approach are however necessary, especially of how to estimate the observer variability in the original data.

7 Acknowledgements

8 Appendix

The following text is a short step-by-step explanation of the optimization approach presented in this thesis.

1. Collect the image material. The images should represent standard patients for the scanning protocol of concern. The radiologists assessing the images should be representative for the department.

2. Discuss the criteria and possible responses (freeware for Svensson's method is given for 4, 6 and 11 categories) with which the image quality will be evaluated. Structures of interest should be of different size and contrast and they should either be diagnostically interesting in themselves or situated at a diagnostically important level of the head/body. Be aware of different phases of the contrast media that could effect the evaluation. Discuss whether it is practical/necessary to assess the whole stack or just a few images per patient.

3. Arrange with a noise simulation software that either adds artificial noise directly to the raw data or directly to the images. Make sure that the noise simulation tool is accurate for the actual settings used in the scanning protocol of concern. Add noise that represents reductions in radiation exposure in percent. Steps of 10 or 15 % are recommended based on experience.

4. Include a test session with the radiologists with discussions regarding how to interpret the criteria, the appropriateness of the structures and the range of radiation exposure. Make adjustments if required.

5. Perform a training session allowing the radiologists to have some experience of assessing. If using ViewDEX for the evaluation, training sessions can be included within the personal login of the radiologists.

6. Create doublets of at least 10 % of the image material for intra-observer evaluation and include in the study. Make sure assessments from both high and low radiation exposures are represented.

7. Perform the study; be sure that the medical monitor used is appropriate according to clinical practise and calibrated according to standards. The evaluation should take place under conditions representative for the clinical situation.

8. Sort the results for an easy overview in e.g. Microsoft Office Excel®. Create tabulations on the distribution of responses for each patient and radiation exposure level. As an example see the following:

Appendix

| | | Question | | | | | |
Patient	mA	1	2	3	4	5	6
4859	30	C	C	C	C	D	H
4859	50	B	C	C	B	C	H
4859	70	C	C	B	B	B	G
4859	90	B	C	B	B	B	G
4859	110	B	B	B	A	B	G
4975	30	D	D	D	C	C	I
4975	50	C	D	C	C	C	H
4975	70	C	D	C	B	B	H
4975	90	B	F	C	B	B	H
4975	110	C	D	C	B	B	G
4784	30	D	D	C	D	C	I
4784	50	D	D	C	C	C	H
4784	70	C	D	C	D	B	H
4784	90	C	D	B	C	B	G
4784	110	C	C	A	C	A	G

Question 3

	D	C	B	A	
30	1	2			3
50		3			3
70		2	1		3
90		1	2		3
110		1	1	1	3
	1	9	4	1	

Question 3

	D	C	B	A	
30	1	2			3
50		3			3
70		3			3
90		1	2		3
110			2	1	3
	1	9	4	1	

The table to the left shows an overview of the assessments (response A to E for questions 1 to 5 and response F to J for question 6. A and F represents the highest image quality and E and J the lowest) for each level of tube current (30 to 110 mA) for three patients (4859, 4975 and 4784). The assessments are from one observer. The table to the upper right shows the distribution of responses for each tube current level. For example, there is one response A, one response B and one response C for 110 mA regarding question 3. The summarized distribution for each response and tube current is noted at the marginal (marginal distribution). Create an empty table with only the marginal distributions left. Create in this table, new pairs of data by pairing off the tube current with the response according to the marginal distributions by starting in the upper left corner for example (see lower right table). The pairs are now ordered into the RTPA.

9. Create RTPA for each observer, question and patient age group. If using Microsoft Excel, it is recommended by experience to use 'If-rules' in order to automatically sort the original data into RTPA. This speeds up the evaluation.

10. Identify the minimum radiation exposure in the RTPA. In the example above the minimum tube currents are 110 and 50 mA for the responses B and C. The minimum radiation exposure is at the lowest exposure where all images in the RTPA have at least that level of image quality for that exposure.

11. The final level of radiation exposure for a specific age group is set by the highest of the minimum radiation exposures among the observers for the question(s) regarded as relevant for the type of examinations. This provided that the variability between observers seems reasonable.

12. Create cross-classification tables of the distribution of the assessments from the intra- and inter-observer evaluation, for example see the following:

Observer X (X_1)

		A	B	C	D	
Observer Y (X_2)	D	0	0	1	1	2
	C	0	0	10	0	10
	B	2	13	2	0	17
	A	1	0	0	0	1
		3	13	13	1	

The cross-classification table represents the assessments of the test-retest images of one observer at different occasions (X_1 and X_2) (intra-observer evaluation) or for two observers (X and Y) assessing the same images (inter-observer evaluation).

13. Discussions regarding the uncertainty of the resulting minimum radiation exposure should be based on findings of systematic bias (RP and RC) and high RV.

(14.) Estimate the variability in original data relative the RTPA by dividing the original data into intervals according to the RTPA. The intervals should be based on where the majority of the assessments in the RTPA are. An example of this is given in section 5.3 where Table 8 is an evaluation of the RTPA in Figure 11. Apply Svensson's method on the cross-classification table with the radiation exposure scale on one axis and the image quality categories on the other axis. Svensson's method is available for 4, 6 and 11 categories at: http://www.oru.se/Akademier/Handelshogskolan/Kontakt-och-presentation/Personliga-sidor/Statistik/Elisabeth-Svensson/Svensoons-metod/Svenssons-metod---fri-programvara-och-dokumentation/

References

1.   Brenner DJ, Doll R, Goodhead DT, Hall EJ, Land CE, Little JB, et al. (2003). Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proc Natl Acad Sci USA 100, 13761-13766.

2.   National Cancer Institute (2008). Radiation Risks and Pediatric Computed Tomography (CT): A Guide for Health Care Providers. Available via http://www.cancer.gov/cancertopics/causes/radiation/radiation-risks-pediatric-CT. Accessed 2011 January 4.

3.   Mettler FA, Jr., Wiest PW, Locken JA, and Kelsey CA (2000). CT scanning: patterns of use and dose. J Radiol Prot 20, 353-359.

4.   Sodickson A, Baeyens PF, Andriole KP, Prevedello LM, Nawfel RD, Hanson R, et al. (2009). Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. Radiology 251, 175-184.

5.   ICRP Publication 103 (2007). Annals of the ICRP: The 2007 Recommendations of the International Comission on Radiological Protection. 37/2-4.

6.   Brenner D, Elliston C, Hall E, and Berdon W (2001). Estimated risks of radiation-induced fatal cancer from pediatric CT. AJR Am J Roentgenol 176, 289-296.

7.   Brenner DJ (2002). Estimating cancer risks from pediatric CT: going from the qualitative to the quantitative. Pediatr Radiol 32, 228-223; discussion 242-224.

8.   Hall P, Adami HO, Trichopoulos D, Pedersen NL, Lagiou P, Ekbom A, et al. (2004). Effect of low doses of ionising radiation in infancy on cognitive function in adulthood: Swedish population based cohort study. BMJ 328, 19.

9.   Berrington de Gonzalez A, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, et al. (2009). Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med 169, 2071-2077.

10. Shrimpton PC, and Edyvean S (1998). CT scanner dosimetry. Br J Radiol 71, 1-3.

11. Mettler FA, Jr., Thomadsen BR, Bhargavan M, Gilley DB, Gray JE, Lipoti JA, et al. (2008). Medical radiation exposure in the U.S. in 2006: preliminary results. Health Phys 95, 502-507.

12. Donnelly LF, Emery KH, Brody AS, Laor T, Gylys-Morin VM, Anton CG, et al. (2001). Minimizing radiation dose for pediatric body applications of single-detector helical CT: strategies at a large Children's Hospital. AJR Am J Roentgenol 176, 303-306.

13. Paterson A, Frush DP, and Donnelly LF (2001). Helical CT of the body: are settings adjusted for pediatric patients? AJR Am J Roentgenol 176, 297-301.

14. Pages J, Buls N, and Osteaux M (2003). CT doses in children: a multicentre study. Br J Radiol 76, 803-811.

15. Boone JM, Geraghty EM, Seibert JA, and Wootton-Gorges SL (2003). Dose reduction in pediatric CT: a rational approach. Radiology 228, 352-360.

16.   Ledenius K, Gustavsson M, Johansson S, Stalhammar F, Soderberg J, Wiklund LM, et al. (2005). A method of predicting the image noise in paediatric multi-slice computed tomography images. Radiat Prot Dosimetry 114, 313-316.

17.   Starck G, Lonn L, Cederblad A, Forssell-Aronsson E, Sjostrom L, and Alpsten M (2002). A method to obtain the same levels of CT image noise for patients of various sizes, to minimize radiation dose. Br J Radiol 75, 140-150.

18.   Nyman U, Ahl TL, Kristiansson M, Nilsson L, and Wettemark S (2005). Patient-circumference-adapted dose regulation in body computed tomography. A practical and flexible formula. Acta Radiol 46, 396-406.

19.   Kalra MK, Maher MM, Kamath RS, Horiuchi T, Toth TL, Halpern EF, et al. (2004). Sixteen-detector row CT of abdomen and pelvis: study for optimization of Z-axis modulation technique performed in 153 patients. Radiology 233, 241-249.

20.   McCollough CH, Bruesewitz MR, and Kofler JM, Jr. (2006). CT dose reduction and dose management tools: overview of available options. Radiographics 26, 503-512.

21.   Wilting JE, Zwartkruis A, van Leeuwen MS, Timmer J, Kamphuis AG, and Feldberg M (2001). A rational approach to dose reduction in CT: individualized scan protocols. Eur Radiol 11, 2627-2632.

22.   Huda W, Lieberman KA, Chang J, and Roskopf ML (2004). Patient size and x-ray technique factors in head computed tomography examinations. II. Image quality. Med Phys 31, 595-601.

23.   Ogden K, Huda W, Scalzetti EM, and Roskopf ML (2004). Patient size and x-ray transmission in body CT. Health Phys 86, 397-405.

24.   Department of Health and Human Services, FDA (2010). CFR 21 Part 1020: Performance standards for ionizing radiation emitting products. Federal register. In 49, 170.

25.   International Atomic Energy Agency (IAEA) (2007). Dosimetry in Diagnostic Radiology: An International Code of practice. Technical reports series, no. 457. ISSN 0074–1914; Printed by the IAEA in Austria, STI/PUB/1294.

26.   International Electrotechnical Commission (2009). International Standard IEC 60601-2-44 Third Edition, Medical electrical equipment – Part 2-44: Particular requirements for the basic safety and essential performance of X-ray equipment for computed tomography. Geneva, IEC.

27.   US Food and Drug Administration (2006). Provision for alternate measure of the computed tomography dose index (CTDI) to assure compliance with the dose information requirements of the Federal Performance Standard for Computed Tomography. Washington, DC, FDA.

28.   Leitz W, Axelsson B, and Szendrö G (1995). Computed tomography dose assessment – a practical approach. Radiation Protection Dosimetry 57, 377-380.

29.   Bongartz G, Golding SJ, Jurik AG, Leonardi M, van Persijn van Meerten E, Rodríguez R, et al. (2004). European Guidelines for Multislice Computed Tomography. Publication EUR 16262 EN. Contract number FIGM-CT2000-20078-CT-TIP. Luxembourg: Office for Official Publications of the European Communities.

30.   Chapple CL, Willis S, and Frame J (2002). Effective dose in paediatric computed tomography. Phys Med Biol 47, 107-115.

References

31. McCollough CH (2006). It is time to retire the computed tomography dose index (CTDI) for CT quality assurance and dose optimization. Against the proposition. Med Phys 33, 1190-1191.

32. Brenner DJ (2006). It is time to retire the computed tomography dose index (CTDI) for CT quality assurance and dose optimization. For the proposition. Med Phys 33, 1189-1190.

33. Brenner DJ (2005). Is it time to retire the CTDI for CT quality assurance and dose optimization? Med Phys 32, 3225-3226.

34. Borras C, Huda W, and Orton CG (2010). Point/counterpoint. The use of effective dose for medical procedures is inappropriate. Med Phys 37, 3497-3500.

35. Dixon RL (2003). A new look at CT dose measurement: beyond CTDI. Med Phys 30, 1272-1280.

36. Mori S, Endo M, Nishizawa K, Tsunoo T, Aoyama T, Fujiwara H, et al. (2005). Enlarged longitudinal dose profiles in cone-beam CT and the need for modified dosimetry. Med Phys 32, 1061-1069.

37. Morgan HT (2002). Dose reduction for CT pediatric imaging. Pediatr Radiol 32, 724-728; discussion 751-724.

38. Khursheed A, Hillier MC, Shrimpton PC, and Wall BF (2002). Influence of patient age on normalized effective doses calculated for CT examinations. Br J Radiol 75, 819-830.

39. Theocharopoulos N, Damilakis J, Perisinakis K, Tzedakis A, Karantanas A, and Gourtsoyiannis N (2006). Estimation of effective doses to adult and pediatric patients from multislice computed tomography: A method based on energy imparted. Med Phys 33, 3846-3856.

40. Hsieh J (2003). Computed Tomography: Principles, Design, Artifacts, and Recent Advances, (SPIE Press Monograph Vol. PM114).

41. Brooks RA, and Di Chiro G (1976). Statistical limitations in x-ray reconstructive tomography. Med Phys 3, 237-240.

42. Boedeker KL, Cooper VN, and McNitt-Gray MF (2007). Application of the noise power spectrum in modern diagnostic MDCT: part I. Measurement of noise power spectra and noise equivalent quanta. Phys Med Biol 52, 4027-4046.

43. Vucich J, Goodenough D, Lewicki A, Briefel E, and Weaver K (1980). Use of anatomical criteria in screen/film selection for portable chest x-ray procedures. In: Cameron J, editor. Optimization of chest radiography. Rockville, MD: HHS Publication (FDA) 80-8124, 237-248.

44. Judy PF, Swensson RG, and Szulc M (1981). Lesion detection and signal-to-noise ratio in CT images. Med Phys 8, 13-23.

45. Frush DP, Slack CC, Hollingsworth CL, Bisset GS, Donnelly LF, Hsieh J, et al. (2002). Computer-simulated radiation dose reduction for abdominal multidetector CT of pediatric patients. AJR Am J Roentgenol 179, 1107-1113.

46. Mayo JR, Whittall KP, Leung AN, Hartman TE, Park CS, Primack SL, et al. (1997). Simulated dose reduction in conventional chest CT: validation study. Radiology 202, 453-457.

47. Li X, Samei E, DeLong DM, Jones RP, Gaca AM, Hollingsworth CL, et al. (2009). Pediatric MDCT: towards assessing the diagnostic influence of dose reduction on the detection of small lung nodules. Acad Radiol 16, 872-880.

48. Lee C, Kim H, Su Jeon S, Cho H, Ra Nam S, and Jung J (2008). Radiation Dose Measurement for Various Parameters in MDCT. Proc. of SPIE 6913.

49. Kalra MK, Maher MM, Toth TL, Kamath RS, Halpern EF, and Saini S (2004). Comparison of Z-axis automatic tube current modulation technique with fixed tube current CT scanning of abdomen and pelvis. Radiology 232, 347-353.

50. GE Healthcare (2008). AutomA/SmartmA Theory. Available via http://www.gehealthcare.com/usen/education/tip_app/docs/AutomA-SmartmA%20Theory.pdf. Accessed 2010 November 13.

51. Månsson LG (2000). Methods for the evaluation of image quality. Radiation Protection Dosimtery 90, 89-99.

52. Sund P, Herrmann C, Tingberg A, Kheddache S, Månsson LG, Almén A, et al. (2000). Comparison of two methods for evaluating image quality of chest radiographs. Proc SPIE 3981, 251–257.

53. Tingberg A, Herrmann C, Lanhede B, Almén A, Besjakov J, Mattsson S, et al. (2000). Comparison of two methods for evaluation of the image quality of lumbar spine radiographs. Radiat Prot Dosim 90, 165–168.

54. Sandborg M, McVey G, Dance D, and Alm Carlsson G (2000). Comparison of model predictions of image quality with results of clinical trials in chest and lumbar spine screenfilm imaging. Radiat Prot Dosim 90, 173–176.

55. Sandborg M, Tingberg A, Dance D, Lanhede B, Almén A, McVey G, et al. (2001). Demonstration of correlations between clinical and physical image quality measures in chest and lumbar spine screen-film radiography. Br J Radiol 74, 520–528.

56. Tingberg A (2000). Quantifying the quality of medical X-ray images. An evaluation based on normal anatomy for lumbar spine and chest images. (Thesis). Lund University.

57. Bath M, and Mansson LG (2007). Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. Br J Radiol 80, 169-176.

58. Smedby O, and Fredrikson M (2010). Visual grading regression: analysing data from visual grading experiments with regression models. Br J Radiol 83, 767-775.

59. Dybkaer R, and Jorgensen K (1989). Measurement, value, and scale. Scand J Clin Lab Invest Suppl 194, 69-76.

60. Stevens SS (1946). On the Theory of Scales of Measurement. Science 103, 677-680.

61. Svensson E (2000). Comparison of the quality of assessments using continuous and discrete ordinal rating scales. Biometrical J 42, 417-434.

62. Svensson E (2000). Concordance between ratings using different scales for the same variable Stat Med 19, 3483-3496.

63. Maclure M, and Willett WC (1987). Misinterpretation and misuse of the kappa statistic. Am J Epidemiol 126, 161-169.

References

64.   Svensson E (1997). A coefficient of agreement adjusted for bias in paired ordered categorical data Biometric J 39, 643–657.

65.   Svensson E (1998). Application of a rank-invariant method to evaluate reliability of ordered categorical assessments. J Epidemiol Biostat 3, 403–409.

66.   Swedish Radiation Safety Authority (2008). Strålsäkerhetsmyndighetens föreskrifter om allmänna skyldigheter vid medicinsk och odontologisk verksamhet med joniserande strålning. [Text in Swedish]. SSMFS 2008:35.

67.   Li J, Udayasankar UK, Toth TL, Seamans J, Small WC, and Kalra MK (2007). Automatic patient centering for MDCT: effect on radiation dose. AJR Am J Roentgenol 188, 547-552.

68.   Matsubara K, Koshida K, Ichikawa K, Suzuki M, Takata T, Yamamoto T, et al. (2009). Misoperation of CT automatic tube current modulation systems with inappropriate patient centering: phantom studies. AJR Am J Roentgenol 192, 862-865.

69.   Tzedakis A, Perisinakis K, Raissaki M, and Damilakis J (2006). The effect of z overscanning on radiation burden of pediatric patients undergoing head CT with multidetector scanners: a Monte Carlo study. Med Phys 33, 2472-2478.

70.   Nicholson R, and Fetherston S (2002). Primary radiation outside the imaged volume of a multislice helical CT scan. Br J Radiol 75, 518-522.

71.   Deak PD, Langner O, Lell M, and Kalender WA (2009). Effects of adaptive section collimation on patient radiation dose in multisection spiral CT. Radiology 252, 140-147.

72.   Huda W (2004). Dose and image quality. How does the choice of kilovoltage affect patient dose, scattered radiation, and image quality in CT examinations? Submitted by Glen Thomson, Middlemore Hospital, New Zealand. Pediatr Radiol 34, 185.

73.   Kim MJ, Park CH, Choi SJ, Hwang KH, and Kim HS (2009). Multidetector computed tomography chest examinations with low-kilovoltage protocols in adults: effect on image quality and radiation dose. J Comput Assist Tomogr 33, 416-421.

74.   Siegel MJ, Schmidt B, Bradley D, Suess C, and Hildebolt C (2004). Radiation dose and image quality in pediatric CT: effect of technical factors and phantom size and shape. Radiology 233, 515-522.

75.   Nauer CB, Kellner-Weldon F, Von Allmen G, Schaller D, and Gralla J (2009). Effective doses from scan projection radiographs of the head: impact of different scanning practices and comparison with conventional radiography. AJNR Am J Neuroradiol 30, 155-159.

76.   Beaconsfield T, Nicholson R, Thornton A, and Al-Kutoubi A (1998). Would thyroid and breast shielding be beneficial in CT of the head? Eur Radiol 8, 664-667.

77.   Chang KH, Lee W, Choo DM, Lee CS, and Kim Y (2010). Dose reduction in CT using bismuth shielding: measurements and Monte Carlo simulations. Radiat Prot Dosimetry 138, 382-388.

78.   Shrimpton PC, Hillier MC, Lewis MA, and Dunn M (2006). National survey of doses from CT in the UK: 2003. Br J Radiol 79, 968-980.

79.   Galanski M, Nagel HD, and Stamm G (2006). Paediatric CT exposure practice in the Federal Republic of Germany. Results of a nation-wide survey in 2005/06. Medizinische Hochschule Hannover. Available via

http://www.mh-hannover.de/fileadmin/kliniken/diagnostische_radiologie/download/Report_German_Paed-CT-Survey_2005_06.pdf. Accessed 2011 January 29.

80. Singh S, Kalra MK, Moore MA, Shailam R, Liu B, Toth TL, et al. (2009). Dose reduction and compliance with pediatric CT protocols adapted to patient size, clinical indication, and number of prior studies. Radiology 252, 200-208.

81. Rybka K, Staniszewska AM, and Bieganski T (2007). Low-dose protocol for head CT in monitoring hydrocephalus in children. Med Sci Monit 13 Suppl 1, 147-151.

82. Udayasankar UK, Braithwaite K, Arvaniti M, Tudorascu D, Small WC, Little S, et al. (2008). Low-dose nonenhanced head CT protocol for follow-up evaluation of children with ventriculoperitoneal shunt: reduction of radiation and effect on image quality. AJNR Am J Neuroradiol 29, 802-806.

83. Frush DP, Donnelly LF, and Rosen NS (2003). Computed tomography and radiation risks: what pediatric health care providers should know. Pediatrics 112, 951-957.

84. Kalra MK, Maher MM, Toth TL, Hamberg LM, Blake MA, Shepard JA, et al. (2004). Strategies for CT radiation dose optimization. Radiology 230, 619-628.

85. Kanal KM, Stewart BK, Kolokythas O, and Shuman WP (2007). Impact of operator-selected image noise index and reconstruction slice thickness on patient radiation dose in 64-MDCT. AJR Am J Roentgenol 189, 219-225.

86. National Electrical Manufacturers Association (2004). Digital Imaging and Communications in Medicine (DICOM) part 14: Grayscale Standard Display Function. NEMA PS 3.14-2004 (NEMA, Rosslyn, VA).

87. Hakansson M, Svensson S, Zachrisson S, Svalkvist A, Bath M, and Mansson LG (2010). VIEWDEX: an efficient and easy-to-use software for observer performance studies. Radiat Prot Dosimetry 139, 42-51.

88. Honnef D, Wildberger JE, Haras G, Hohl C, Staatz G, Gunther RW, et al. (2008). Prospective evaluation of image quality with use of a patient image gallery for dose reduction in pediatric 16-MDCT. AJR Am J Roentgenol 190, 467-473.

89. Verdun FR, Lepori D, Monnin P, Valley JF, Schnyder P, and Gudinchet F (2004). Management of patient dose and image noise in routine pediatric CT abdominal examinations. Eur Radiol 14, 835-841.

90. Martinsen AC, Saether HK, Olsen DR, Skaane P, and Olerud HM (2008). Reduction in dose from CT examinations of liver lesions with a new postprocessing filter: a ROC phantom study. Acta Radiol 49, 303-309.

91. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, and Kim J (2010). Long radiology workdays reduce detection and accommodation accuracy. J Am Coll Radiol 7, 698-704.

92. Mayo JR, Kim KI, MacDonald SL, Johkoh T, Kavanagh P, Coxson HO, et al. (2004). Reduced radiation dose helical chest CT: effect on reader evaluation of structures and lung findings. Radiology 232, 749-756.

No bird soars too high if he soars with his own wings

-William Blake