Konzepte zur Charakterisierung klinischer CT-Systeme unter Einbeziehung von Bildqualität und Dosis

Characterization of clinical CT systems using a dose efficiency index (DEI)

Aktenzeichen / FKZ : BfS AG-F 3 – 08313 / 3613S20007

Der Bericht gibt die Auffassung und Meinung des Augtragnehmers wieder und muss nicht mit der Meinung der Auftraggeberin übereinstimmen.

Final report

Damien Racine, Anaïs Viry, Sue Edyvean, Francis R. Verdun

Introduction

In Germany, as in most Western countries, the exposure to the population due to CT examinations has increased over the last twenty years, despite major technological progresses which might lead to the expectation that doses would decrease. The last survey organized in Germany showed that in 2014 the average dose per inhabitant due to CT was 1.0 mSv compared to 0.8 mSv in 2007. Thus, the use of CT is a major source of concern for Public Health Authorities such as the "Bundesamts für Stahlenchutz" in Germany. Among the ways to control the population exposure creeping upwards one should ensure that the principle of justification is correctly applied and that the radiological examination and intervention are optimized in terms of the balance of low dose against required image quality. In the context of optimization, the units used should work efficiently to ensure that there is no waste of radiation in the imaging process. The aim of this project is to propose a way to benchmark CT units, allowing the assessment of the efficiency of CT units using clinically relevant protocols.

The project was made of four work-packages. The first work-package was a review of the literature dedicated to image quality assessment in CT imaging. During this work-package the clinically relevant protocols used for benchmarking were also defined. The second work-package was dedicated to the set-up of the methodology to be used when dealing with the benchmarking of CT units using two different concepts. The third work-package was the application of these concepts on eight CT units; including two models from each of four manufacturers. Finally, the fourth work-package consisted in the analysis of the data with the proposal of a strategy to establish a dose efficiency index (DEI) to quantify the CT unit performances.

Outcomes of the project

Arbeitspaket 1 (AP 1) – Work package 1 (WP 1)

Analysis of the state of science and technology – a literature review

The demand for clinically relevant image quality characterization has increased with the development of medical imaging. For example, with CT in particular, the detector efficiency, image reconstruction and processing have all evolved and improved. This has necessarily resulted in the adaptation and evolution of assessment methods. The review performed in the framework of this work-package presents the spectrum of various methods that have been used to characterize image quality in CT. It includes the description of measurements of physical parameters, as well as clinical task-based approaches (such as the model observer (MO) and pure human observer approaches). Advantages and limitations of these methods have also been mentioned. The review performed includes solutions that have been proposed to provide figures of merit (FOM) enabling the benchmarking of CT units. Figure 1 summarizes the various ways to assess image quality in CT, as well as to establish a figure of merit.

The outcome of this work-package confirmed that task-based image quality assessment is the method of choice to be used when dealing with patient dose optimization strategies. One should not aim at producing the highest level of image quality (assessed by traditional metrics such as noise power spectra (NPS) or modulation transfer function (MTF)), but one should provide enough information to enable the diagnosis to be made. Thus, the focus should be on so called "task-based" image quality metrics; metrics that are linked with tasks, such the detection of a low contrast lesion, and which can be termed classification tasks. One could also choose the estimation of the size or shape of a lesion which would be other relevant tasks in the framework of medical diagnosis.

The outcome of the review also confirmed the need to be cautious when using the Fourier domain when dealing with iterative reconstruction methods. Fourier metrics should only be used with linear shift invariant systems; conditions that do not apply anymore with the use of iterative reconstruction. It is of note that, even with the standard filtered back-projection method, CT images are not shift invariant; however the use of Fourier metrics is largely accepted for these images. With the high non linearity nature of iterative reconstruction this is no longer possible. Finally patient dose optimization could be thought of as using two approaches: the first approach aiming to get the maximum information available in the images, and the second approach aiming to get the information that a human observer could extract from the image. We suggest the first approach to be used to benchmark CT units, whereas the second approach should be used to benchmark protocols that are used in the clinical environment. These two approaches will constitute the two concepts requested for this project: concept 1 proposed to benchmark CT unit, and concept 2 proposed to benchmark CT protocols.
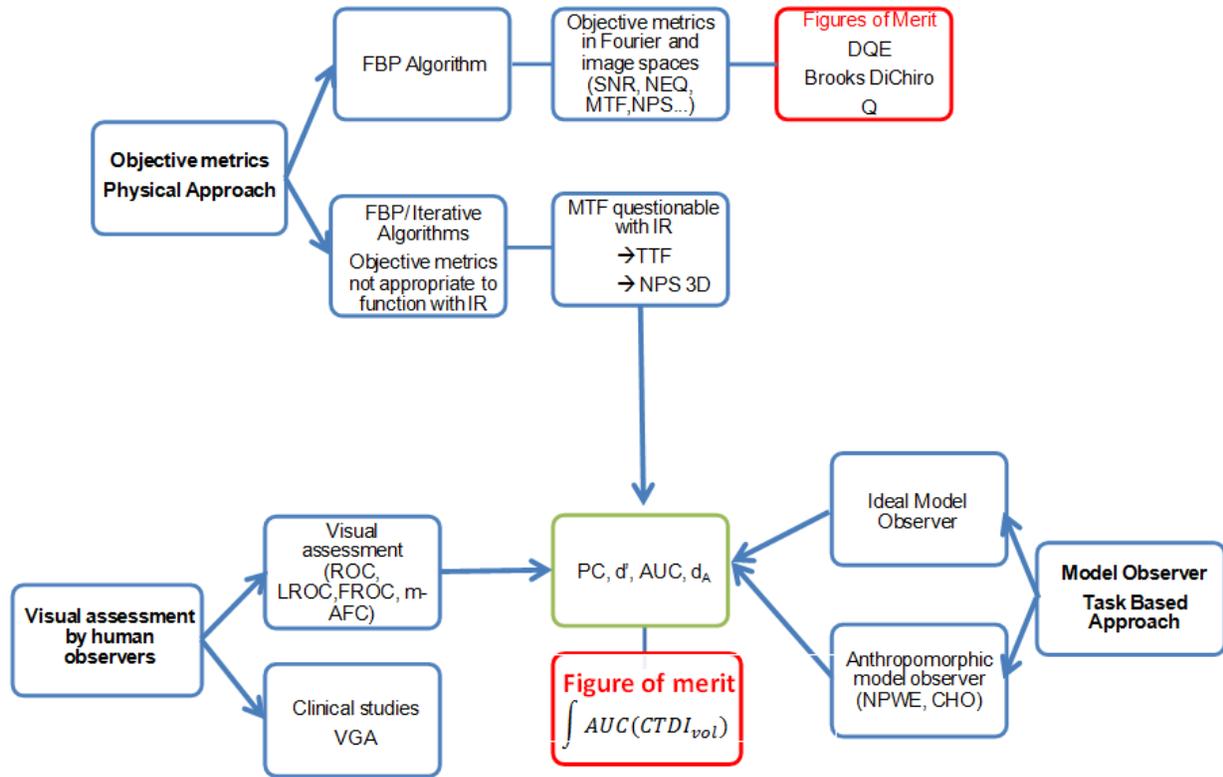
Figure 1: Various way to assess imaging performances of CT units [1]

Diagnostic image quality of CT examinations requires the assessment of a large number of parameters: from the standard low contrast and high contrast resolution, to temporal resolution, energy resolution when dealing with kV optimization or dual energy imaging, etc. In this project, we will constrain the benchmarking of units to the low and high contrast detectability; parameters that are essential in the current use of CT unit. This constitutes a limitation of the project. For example, the outcome of this method will not be able to discriminate between CT units with better temporal resolution (that is an essential property when dealing with cardiac CT), better energy discrimination (that is an essential property when dealing with tissue characterization) or with better structure size assessment.

The review of the spectrum of various methods that have been used to characterize image quality in CT was published in physica medica in 2015 [1].

To finalize this work-package a set of seven clinical protocols to be used with the two concepts has been proposed and adapted after extensive discussions with radiologists from Switzerland, the UK and Germany. These protocols are given in Annexe 1.

Arbeitspaket 2 (AP 2) – Work package 2 (WP 2)

Development of the methodology to characterize a CT unit

The outcome of WP 1 focussed on two particular aspects of task-based image quality criteria: low contrast detectability and spatial resolution of high contrast structures. The necessity to consider those two parameters is because low contrast detectability is highly dependent on the noise level of the image and spatial resolution is highly dependent on the spatial frequency characteristics of the imaging chain. Thus, it was proposed to assess those two parameters for both concepts. As proposed at the beginning of the project for both concepts, a classification task-based image quality assessment methodology (signal present/absent) was used.

Material and methods

Spatial resolution assessment

Methodology

The parameter usually used to assess the spatial resolution when dealing with CT units is the Modulation Transfer Function (MTF) using a standard test object such as a Catphan phantom, as proposed by the Phantom Laboratory (Salem, NY - USA). However, iterative Reconstruction algorithms (IR), are known to be highly non-linear and therefore might introduce an influence of the image contrast and noise on the spatial resolution [2-4], indeed the spatial resolution can decreased or increased with the strength of iterative algorithm in terms of the contrast of the structure. To overcome this problem the target transfer function (TTF) metric was proposed to characterise the spatial resolution incorporating any noise and contrast dependency [2-4]. The MTF and TTF are similar metrics but differ from one another in the sense that MTF only applies to a single given contrast level while the TTF can be applied to different contrasts and dose levels [4].

In a former study the TTF concept was applied to the assessment of spatial resolution when dealing with musculoskeletal CT imaging [4]. For that study, as well as this project, a dedicated custom-made phantom which contained circular edges of different contrast materials (Teflon, Polyethylene and Plexiglas) surrounded by water was used to mimic cortical bone, fat and cartilaginous tissue. The external diameter of the TTF phantom used in the present study is 250 mm, and contains rods of contrast materials of 100 mm diameter (see Figure 2). This phantom allows also the assessment of the Noise Power Spectrum (NPS), (extensive details and explanations on the methodology can be found in the paper of Ott et al. [4]). With this phantom, all acquisitions were performed at a reconstructed FOV of 25 cm (using a 512x512 matrix size).
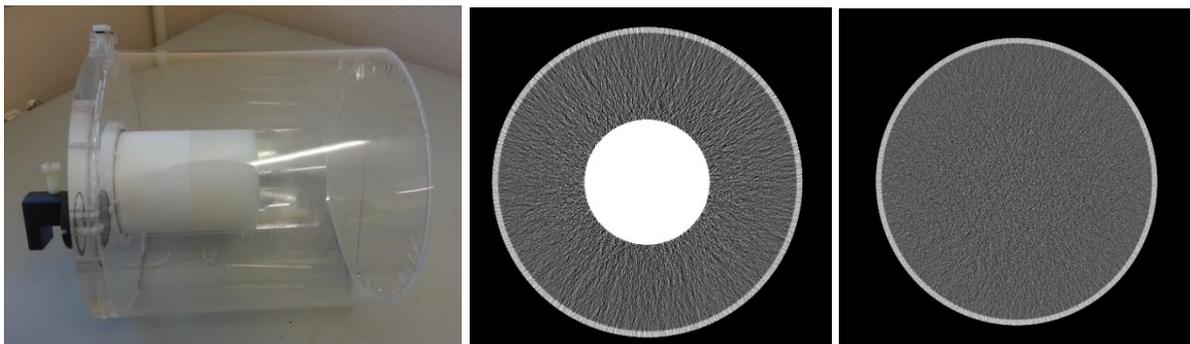

Figure 2: From left to right: Picture of the phantom used to assess the spatial resolution characteristics using the TTF metrics [4]; the Teflon part used to assess the spatial resolution and the homogeneous part used to assess the noise in the image.

The TTFs were used to compute the detectability index (*d'*) of a 2-mm-diameter structure having contrast values corresponding to: Teflon in water; Polyethylene in water; and Plexiglas in water, using mathematical model observers.

Methodology used to assess adult head protocols (high resolution)

To assess the spatial resolution properties of adult head protocols when dealing with a task that requires high spatial resolution characteristics, the TTF phantom was scanned using the imaging protocols summarized in Annexe 1.1. In particular, a bone filter was applied, the pitch value was systematically close to 1.0, and the X-ray tube current was limited to ensure the use of the small focal spot, as agreed with our panel of radiologists. One acquisition with fixed mAs providing a $CTDI_{vol}$ value of 15 ± 1 mGy was made, and fifty images per contrast level were used to compute the TTF metric.

Methodology used to assess adult abdomen/pelvis protocols (high resolution)

For this protocol, the images from TTF phantom were also acquired, but with a standard reconstruction filter (soft tissue), a pitch value close to 1.4, and a high tube current value to be more realistic of the standard abdominal acquisitions. The imaging protocols actually used are summarized in Annexe 1.2. One acquisition with fixed mAs providing a $CTDI_{vol}$ value of 15 ± 1 mGy was made, and fifty images per contrast level were used to compute the TTF function.

Methodology used to investigate image quality of chest protocols (high resolution)

When dealing with chest examinations the spatial resolution parameter is of uppermost importance. Low contrast detection requirements do exist for the detection of mediastinal lesions but in that case one can use the outcomes of low contrast detectability assessed with abdominal phantoms. To characterize image quality when dealing with this protocol the raw data acquired for the high resolution abdomen/pelvis protocols were reconstructed using a reconstruction kernel adapted to standard chest imaging - see Annexe 1.3.

For these three protocols the ATCM, automatic tube current modulation, was not used to ensure a constant $CTDI_{vol}$. Due to the circular shape of the phantom the use of the ATCM functionality would anyway have been useless.

Low contrast detectability

Choice of the phantom to be used to investigate image quality of adult head protocols

In order to avoid the use of phantoms provided by only one manufacturer and, additionally, to acquire experience in the approach required by the United States Food and Drug Administration (FDA) for CT image quality assessments, the use of the head MITA phantom (The Phantom Laboratory, Salem USA) was proposed. This phantom has been specifically developed for image quality assessments by means of mathematical model observers. It is made of a cylinder (Ø 20 cm) and contains four rods of various sizes in diameter producing low contrast structures (9 HU/2 mm; 5 HU/3mm; 4 HU/4mm and 3 HU/5 mm). For each dose level, images were acquired using five successive scans of the phantom (without any phantom position change). This enabled 40 ROI's to be obtained with the targets, and 90 ROIs without any target, with each ROI having a size of 80 pixels. The actual protocol used is given in Annexe 1.0 and Figure 3 presents the phantom. Other designs would be necessary to take into account partial volume effects or beam hardening. However, it appears that the diameter is quite large for head protocols; maybe a 16 cm modulus with an annulus in Teflon would have been a better option.
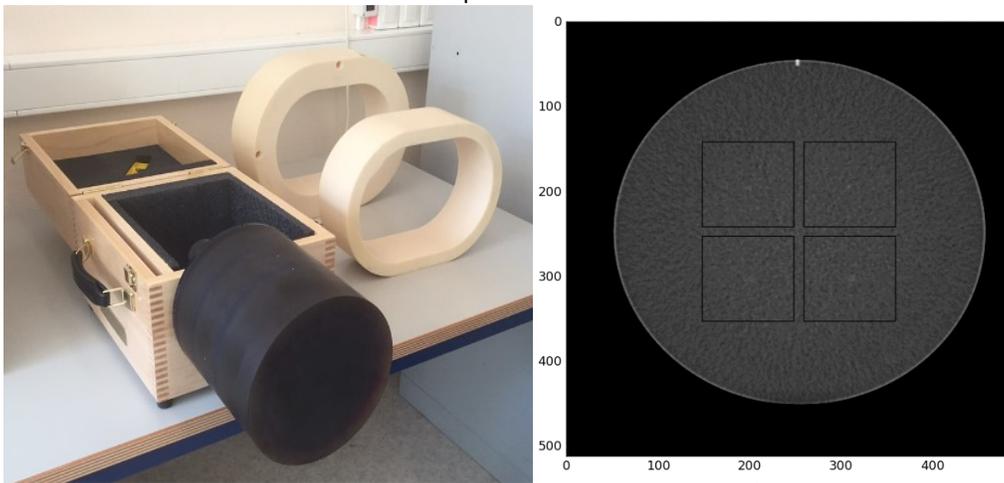

Figure 3: Picture of the head MITA phantom and an image obtained

Choice of the phantom to be used to investigate image quality of paediatric head protocols

Head CT in children are the ones that are the most frequent. It was thus proposed to focus on that protocol using a smaller phantom than the MITA one to assess the CT performances when dealing with very young children. The option taken was to use a modified version of the low contrast detectability module proposed by QRM (Moehrendorf, Germany) shown in Figure 4. The phantom used has a diameter of 100 mm and a length of 200 mm. It is made of tissue-equivalent plastic and contains spherical contrast inserts of -20 HU relative to background (Ø 8, 6, 5, 4 and 3 mm). To use the same size of the region of interest (ROI) considered in all the calculations, whatever the target (lesion) size, image quality assessments were performed using only the spheres of 5 mm, 6 mm and 8 mm in diameter. Images have been acquired using ten successive scans of the phantom (without any phantom position change) allowing 40 ROI to be obtained with the targets, and 90 ROI without any target; an approach that limits the number of acquisitions while providing reliable data sets. The actual protocol used is given in Annexe 1.6. To ensure that imaging characteristics were comparable when dealing with images with or without target scanning, a similar homogeneous modulus (containing no contrast) was scanned in similar conditions (the same position within the phantom was thus considered).
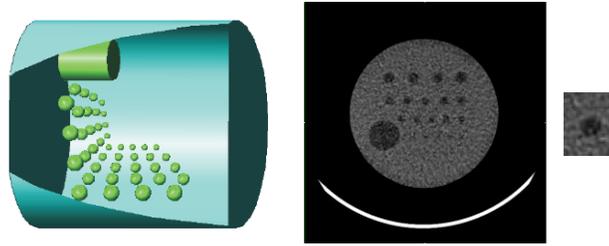
Figure 4: Schematics of the QRM 3D module, an example of image obtained, and an example of the ROI used to assess the detectability

Choice of the phantom to be used to investigate image quality of abdominal protocols

The module shown in Figure 4 was inserted into an abdominal envelope; a slightly modified QRM 401 phantom (Moehrendorf, Germany) (length of 200 mm instead of 100 mm). This phantom has an equivalent diameter of 240 mm. The envelope of the phantom is made of muscle, liver, spleen and bone (vertebrae) tissue equivalents. As for the paediatric protocol, the LCD was assessed using the spheres of 5, 6 and 8mm in diameter. A homogeneous module was also used to produce regions of interest (ROI) containing noise only.

When dealing with abdominal scans, the diagnostic image quality requirements might vary significantly. For example the requirement of some pathologies concerning LCD performance may be higher than others. In addition, IR might allow drastic dose reductions while still providing an acceptable appearance of the image. Thus, in order to benchmark CT units, this phantom was systematically scanned using three dose levels to assess the added value of the IR solutions proposed by manufacturers. Moreover, the use of several dose levels ensure that image quality assessment is not made at a local optimum provided by IR.

In order to benchmark clinical protocols, the set of data was completed with acquisitions performed with additional annuli (also provided by QRM) to investigate the effect of the anatomy on LCD. The acquisition protocols are summarized in Annexe 1.4 and 1.5
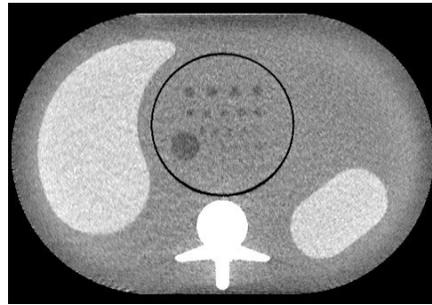
Figure 5: Example of Images provided by the QRM abdominal phantom

Choice of the mathematical model observers for high contrast detectability

Concept 1

As mentioned previously, the aim of concept 1 is to extract as much information as possible, and for a model observer study a so-called "ideal model observer" should be used. In this project it is proposed to use the pre-whitening model observer (PW) to assess the transfer of high contrast structures by the CT unit from the object to the image. This kind of model removes the noise correlation and this operation is called "pre-whitening", aiming at assessing the performance as if the noise was white (constant amplitude at all spatial frequency). The figure of merit used will be d'; signal-to-noise ratio, SNR, of the measurements (cf. Equation 1)

$$d' = \sqrt{2\pi}\,\Delta HU \sqrt{\int_0^{f_{Ny}} \frac{S^2(f)TTF^2(f)}{NPS(f)}\,f\,df}\;(1)$$

Where, f_{Ny} is the Nyquist frequency, ΔHU is the contrast difference between the signal and the background and $S(f)$ is the Fourier transform of the input signal. For a disc of radius R (R=1 mm in our case) we have: $S(f) = \frac{R}{f}\,J_1(2\pi Rf)$, J_1 being a Bessel function of the first kind.

At this stage, it is important to realize that some precautions have to be taken concerning the effect of scatter radiations. The TTF is reduced by scatter as well as the image contrast and this effect should not be taken into effect twice. For this project it was chosen to include the scatter effect by using the measured contrasts rather than the nominal contrast. The TTF's have been fitted in a way to avoid the effect of scatter (spatial resolution drop in the low frequency range) as presented in ref. [4].

Alternative methods for high resolution image quality assessments

In addition to the task-based image quality assessment methods, it was also proposed to investigate the outcomes of the results when using alternative methods such as the TTF integral, Acceptability index or Q2 factors.

TTF integral

When dealing with PW models observers with IR, drastic NPS reductions might have a major impact on the metric without having that much impact of target conspicuity. In such a context, it was proposed to compute the integral of the TTF to characterize the spatial resolution characteristics, ignoring the noise component of the acquisition. This approach is not a task-based one and has limitations mentioned previously in the introduction part.

NPS integral

To better understand the outcomes of the model observer approach not only the integral of the TTF was used but the integral of the NPS was also computed. In the process of interpreting the results of model observers it appears important to establish some links with more traditional metrics.

AbnahmeIndikator from DIN 6868-161

From the standard DIN 6868-161 (01-2013) for image quality assurance of diagnostic cone beam X-ray devices, a relationship between contrast-to-noise ratio, dose and spatial resolution is proposed.

$$AI = \frac{CNR}{D\,(\frac{1}{2}\,v_{50\%})^2},$$

where CNR is the contrast-to-noise ratio calculated from equation, D, as indicated above is a measure of the radiation dose and $v_{50\%}$ is a measure of the in-plane spatial resolution (MTF$_{50}$).
The contrast-to-noise ratio (CNR) was calculated using the following equation:

$$CNR = \frac{ROI_{material} - ROI_{background}}{\sqrt{\frac{1}{2}(SD\ material^2 + SD\ background^2)}},$$

where $ROI_{material1}$ and $ROI_{background}$ are the mean HU values in the corresponding material and SD is the standard deviation in the corresponding material or in the background.

Q2 from Impact (UK)

The generalized form of a relationship to combine dose and image quality in terms of a dose efficiency value can be presented as follows:

$$Q2 = \sqrt{\frac{f_{av}^3}{\sigma^2 z_1 CTDI_{vol}}}$$

where σ is the image noise, f_{av} is a measure of the in-plane spatial resolution (MTF$_{50}$ + MTF$_{10}$)/2, z is a measure of the spatial resolution along the z-axis (in image space, and a measure of the z-sensitivity), and D, as indicated above, a is measure of the radiation dose. This is the approach used by the ImPACT CT scanner evaluation group [18].

Concept 2

Using the TTF and NPS functions, SNR values were computed using a more anthropomorphic model observer than the pre-whitening one. According to the current status of knowledge, the non-prewhitening match filter, filtered with human eye response, model observer (NPWE - Equation 2) reasonably matches human performances, and was therefore proposed as concept 2 in this project [4-5]. In opposition to the PW model, the NPWE model does not include the noise decorrelation process. It can thus be considered ideal when it is applied on image that contains only white noise.

$$d' = \frac{\sqrt{2\pi}\,\Delta HU \int_0^{f_{Ny}} S^2(f) TTF^2(f)\, VTF^2(f) f\, df}{\sqrt{\int_0^{f_{Ny}} S^2(f) TTF^2(f)\, NPS(f) VTF^4(f) f\, df}} \quad (2)$$

Where, $VTF(f)$ is the visual transfer function of the human eye $VTF(f) = f^{1.8}\exp(-0.6f^2)$ [4].

Choice of the mathematical model observers for low contrast detectability

Model observers are mathematical models based on the statistical decision theory to estimate the detection performance of ideal or human observers. In this study we chose a linear anthropomorphic Channelized Hotelling (CHO) model observer. With a linear observer model, the decision variable is given by the dot product between the template **w** and the reconstructed image **g**. The Hotelling observer (HO) model takes into account all the statistical knowledge of signal and noise to create a template that allows a noise decorrelation (by dividing by the covariance matrix) to extract the maximum of information from the image. The CHO model is a derivative of the HO which is too computing intensive to be used in practice. To reduce the dimensionality of HO, the image is first passed through a set of J channels; where J is significantly lower than N (N being the number of pixels of the image).

These channelized models take advantage of the spatial selectivity behaviour of the human visual system. It is then much easier to implement the Hotelling strategy by dividing by the JxJ covariance matrix resulting from the output of channels.

With the adopted notation, a channel is a Nx1 column vector that produces a scalar output when multiplied by the image **g**. The ensemble of the J channels can therefore be written as an NxJ matrix where each column is one of channel **u**$_j$.

$$U = \left[\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_J \right]$$

The channel output **v**$_i$ is obtained by the dot product between the channel **u**$_j$ and the image **g**.

$$\mathbf{v}_i = \left(\mathbf{u}_j\right)^T \mathbf{g} \quad \text{(Eq. 1)}$$

The template of the resulting covariance matrix is calculated from the images containing no signal.

$$\mathbf{W}_{CHO} = \left(\mathbf{K}_{v/n}\right)^{-1} \left\langle \mathbf{v}_s - \mathbf{v}_n \right\rangle \quad \text{(Eq. 2)}$$

where $\mathbf{K}_{v/n}$ represents the covariance matrix computed from channelized images containing no signal, $<\mathbf{v}_s>$ represents the mean vector containing the data of the signal images as seen through the channels, and $<\mathbf{v}_n>$ represents the mean vector containing the data of the signal-absent as seen through the channels.

The decision variable of CHO model is given by:

$$\lambda_{CHO} = \mathbf{w}^T_{CHO} \mathbf{v} \quad \text{(Eq. 3)}$$

where $\mathbf{w}$ is the vector of the linear template and $\mathbf{v}$ the channelized image.

Concept 1

This concept should extract as much information as possible. Thus one should use an ideal model observer which can be computed in the Fourier domain (as previously for the characterization of spatial resolution) or in the image domain. To avoid the use of noisy data, such as a low contrast TTF's, it was proposed to remain in the image domain and use Channelized Hotelling model Observers, CHO, already proposed by several authors in the field of medical imaging [5-14]. In this project it is proposed to use the CHO model observer with Laguerre-Gauss channels which is considered to be ideal. In this project 40 images with target and 90 images without target were used to provide a reliable covariance matrix inversion process leading to sound metrics such as the area under the ROC curve (AUC); the task of this model observer is a classification one (target present/absent) that is usually characterized by a receiving operating curve (ROC curve).

Laguerre Gauss channels

In this implementation we used up to the sixth order of the Laguerre-Gauss polynomials and one orientations (rotationally invariant) resulting in a total of 6 channels.

$$\boldsymbol{u}_p(r|a_u) = \frac{\sqrt{2}}{a_u} \exp\left(\frac{-\pi r^2}{a_u^2}\right) L_p\left(\frac{2\pi r^2}{a_u^2}\right)$$

With p=6 and where a_u is the width of the Gaussian function and equal to 10.

$$L_p(x) = \sum_{k=0}^{p} (-1)^k \binom{p}{k} \frac{x^k}{k!}$$

Concept 2

Several channels used with the CHO model observer are considered to be anthropomorphic. Their performances are lower than the ideal model observers and tend to better match human performances. In the framework of this project the difference of dense Gaussian channels, D-DoG has been chosen.

Difference of dense Gaussian channels, D-DoG

In this study the target to be detected has a spherical symmetry. In such a case the channels recognized as a good approximation of human vision is the dense of difference of Gaussian (D-DoG) which allows the use of a limited number of channels. This is particularly important since the more channels to be used the more images need to be produced (as a rule of thumb a total of at least J^2 images; J being the number of channels). In our situation it is admitted than 10 channels are sufficient to produce sound results. The D-DoG has the advantage of using fewer channels in comparison to other anthropomorphic channels (such as the Gabor ones). The radial profile of each frequency of the D-DoG is given by the following formula:

$$C_j(\rho) = e^{-\frac{1}{2}\left(\frac{\rho}{Q\sigma_j}\right)^2} - e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2}$$

where ρ is the spatial frequency, J the channel number channels, Q the bandwidth of the channel and, σ_j the standard deviation of each channel. Each σ_j values are given by $\sigma_j = \sigma_0 \alpha^{j-1}$. Factor Q is the bandwidth of the filter. Generally the parameters used are: $\sigma_0 = 0.005$, $\alpha = 1.4$ and $Q = 1.67$ [20].

In order to improve the readability of this report a summary is proposed in Table 1. Each row specifies the tasks and concepts used. For example; low or high contrast detectability, concepts and parameters (TTF, NPS, d' …), and phantoms.

Table 1: Summary the proposed approaches:

	High contrast detectability (head, chest abdomen)	
	concept 1 (ideal model)	concept 2 (anthropomorphic model)
Model	PW	NPWE
Phantom	TTF phantom (Figure 2)	TTF phantom (Figure 2)
Number of ROIs	50 for each contrast	50 for each contrast
Dose levels	15 mGy	15 mGy
	Low contrast detectability – head protocol	
	concept 1 (ideal model)	concept 2 (anthropomorphic model)
Model	CHO with Laguerre Gauss	CHO with DDoG
Phantom	MITA and QRM	MITA and QRM
Number of ROIs	40 for signal 90 for noise	40 for signal 90 for noise
Dose levels	45, 55 and 65 mGy for adult and 1mGy for paediatric	45, 55 and 65 mGy for adult and 1mGy for paediatric
	Low contrast detectability – abdomen protocols	
	concept 1 (ideal model)	concept 2 (anthropomorphic model)
Model	CHO with Laguerre Gauss	CHO with DDoG
Phantom	QRM	QRM
Number of ROIs	40 for signal 90 for noise	40 for signal 90 for noise
Dose levels	5, 10 and 15mGy for fixed dose and different dose levels for protocols with tube current modulation	5, 10 and 15mGy for fixed dose and different dose levels for protocols with tube current modulation

Area Under the Curve Weighted (AUC$_w$)

For each dose level and phantom size, we used an image quality metric called "AUC$_w$" which combines the area under the ROC curve (AUC) of each target size. In the ROC analysis, if λ (for image with signal present) is above a given threshold, the response is considered as a true positive. If λ (for image with noise only) is above a given threshold, the response is considered as false positive. So, for a given threshold, a true positive fraction (TPF) and a false positive fraction (FPF) can obtained and a ROC curve can constructed from pairs of TPF and FPF while varying the decision threshold. The area under the curve is calculated for each target size by the trapezoidal method using 100 points. Then the synthetic metric is computed according to:

$$AUC_w = \frac{\sum_{i,c} \frac{AUC_{lesion(i,c)}}{i * c}}{\sum_{i \in (8;6;5)} \frac{AUC_{lesion(i,c),max}}{i * c}} \quad (7)$$

where i represents the lesion sizes (8, 6, or 5 mm for the phantom QRM and 2, 3, 4 and 5mm for the MITA phantom), c represents the contrast of the lesion (20 HU for phantom QRM and 9, 5, 4 and 3HU for MITA phantom), and AUC$_{lesion(i)}$ represents the outcome of model observer for each lesion size and contrast. With such a definition, AUC$_{lesion (I, c) max}$ corresponds to the value of this metric when the performance is maximal for each lesion (AUC$_{lesion (I, c) max = 1.0}$).

As mentioned previously, we proposed a dose efficiency index (DEI) to characterize the CT unit performances in terms of low contrast detectability assessment for different fixed dose levels. The DEI is represented by the integral of the area under the curve image quality (AUC$_w$) as a function of dose (figure 6).
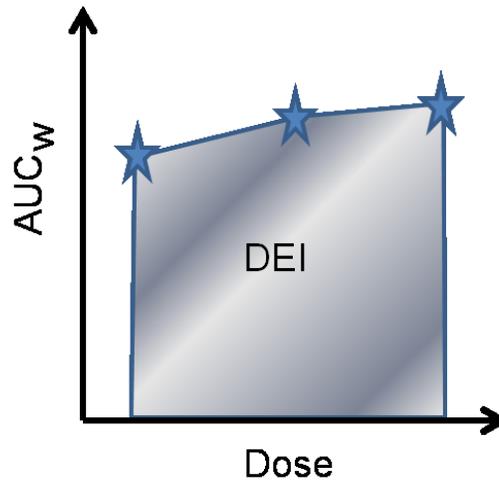

Figure 6: Concept a dose efficiency index (DEI)

CTDI$_{vol}$ assessment

According to Swiss legal requirement, the CTDI$_{vol}$ of the units used in the framework of this study are measured by the manufacturers every three months. Differences between the displayed and measured CTDI$_{vol}$ were within 15%, in conformity with IEC standards (limit of ± 20%).

CT units involved in the study

To ensure impartiality of this work the results will be reported in an anonymous manner; since the goal is to demonstrate the feasibility and potential of the methodology, rather than to evaluate specifically the selected units. The list used is given in the Table 2 where the year of introduction is mentioned. All CT units are least a 16 detector rows CT's. The letter "a" represents the newer CT and "b" the older CT.

Table 2: List of CTs considered in this study

Manufacturer	CT unit	Algorithm	Rows	Year of introduction
GEMS	Revolution	ASIR-V	256	2014
	VCT	ASIR	64	2008
Philips	Ingenuity Core	Idose	128	2011
	Brilliance	FBP	64	2006
Siemens	Force	Admire	192	2012
	Somatom	FBP	64	2003
Toshiba	Aquilion Prime	AIDR 3D	80	2012
	Activion 16	FBP	16	2007

Arbeitspaket 3 & 4 (AP 3 & AP4) – Workpackage 3 & 4 (WP 3 & WP4)

Implementation of the proposed concept

In the framework of this project ninety six conditions had been evaluated (6 protocols, 8 CT scanner models, 2 concepts), an additional protocol was added to include head protocols in neonatology.

WP3.1/WP4 – Application of concept 1 on 48 conditions

As mentioned previously, two model observers have been used to obtain results for concept 1. As the goal of the project is to propose a way to benchmark CT, the PW model and the CHO (using Laguerre Gauss channels) that are considered as ideal model observers were chosen.

Head protocol – HR

Mathematical model observer PW

The high contrast detectability was assessed using the PW model for a structure of 2mm with a nominal contrast of 1080 HU (Teflon).
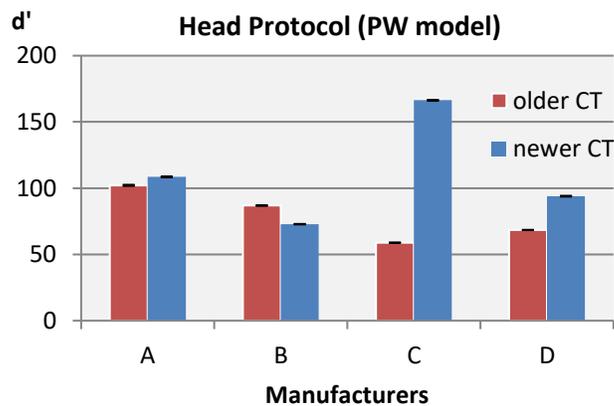
Figure 7: High contrast detectability for head protocol using PW model

The high contrast detectability was assessed using the PW model for a structure of 2mm with a nominal contrast of 1080 HU (Teflon).
As shown in the figure above presented, for three manufacturers A, C and D there is an increase of the detectability index when switching from older to newer CTs. The largest improvement is noted (280 %) for manufacturer C. Nevertheless, for manufacturer B, a slight reduction is observed. For all manufacturers, the detectability index is very high for this task, indicating that the detection of this structure is always trivial. Computing the detectability of structures with different form, size or contrast could be interesting to discriminate CTs. However the change of contrast or diameter change the absolute value, but the ranking remained unchanged [21].
Between newer CTs, the highest score is reached for manufacturer C and manufacturers A and D provide better results than manufacturer B.

Alternative metrics

To assess spatial resolution, it was proposed to use an alternative metric; the TTF integral. The results are given in figure 8a. As with the use of model observer, some differences appeared between CTs, with a major improvement for manufacturer D. However it is not possible to compare spatial resolution performances of CTs without taking into account the noise level. Moreover, TTF is very dependent on the chosen kernel, especially for a high pass kernel. For example, for manufacturer D, the kernel chosen on scanner "Da" is sharper than the kernel chosen on scanner "Db", and this explains why a significant improvement is noted between the older and the newer CT for manufacturer D.

The integral of the NPS (Fig. 8b) was chosen to assess the noise level. For manufacturer C, a major decrease is noted and can explain the major improvement of detectability. Indeed, the iterative reconstruction algorithm used on scanner "Ca" strongly reduces the noise level compared to FBP. These results demonstrate the weakness of the use of the TTF integral concept.
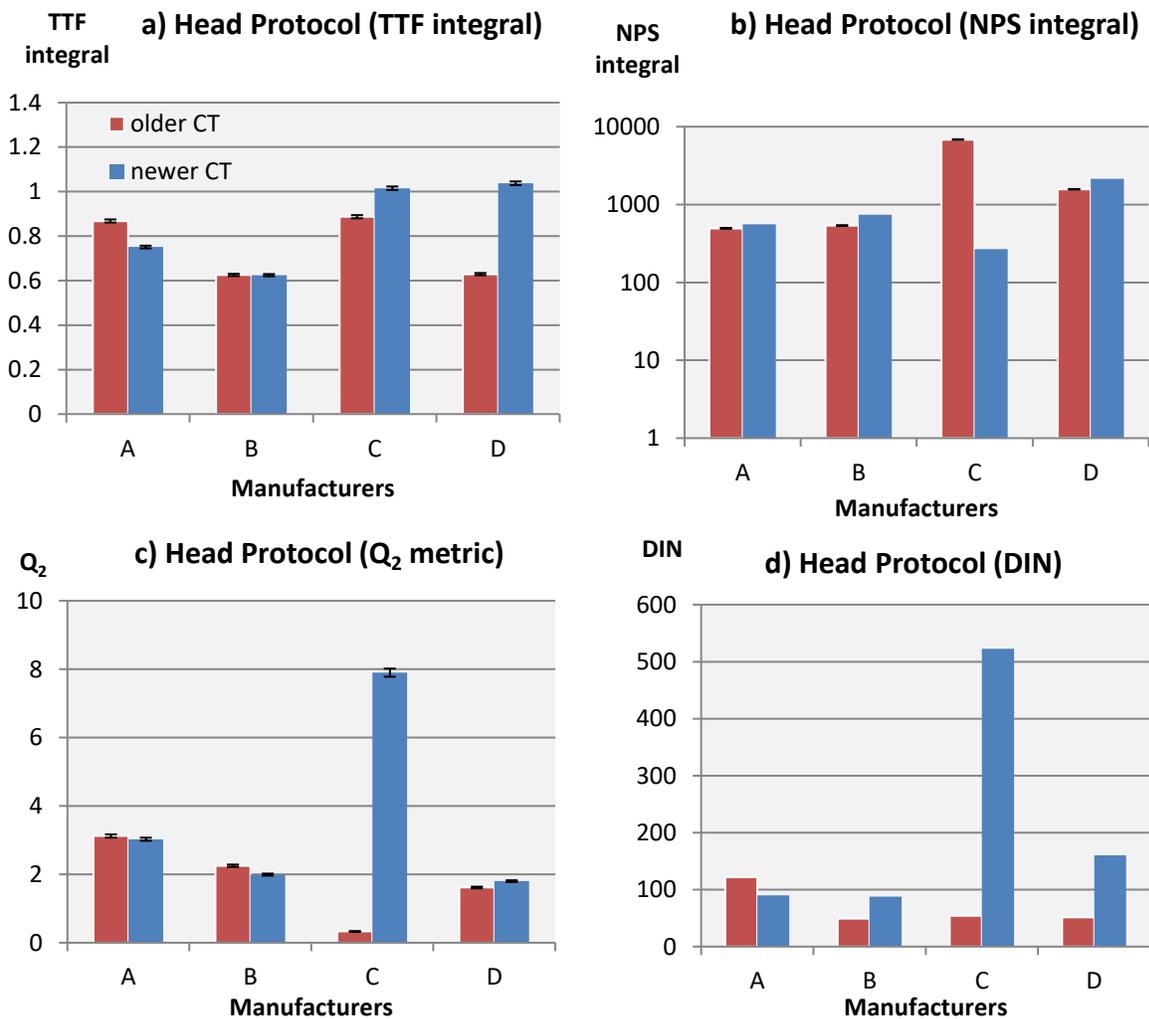
Figure 8: Alternative metrics for HR head protocol: a) TTF's integral; b) NPS's integral; c) Q_2 metric; d) DIN

For the comparison of newer CTs, the analysis of Q2 and DIN metrics provide similar results to the detectability index computed with the PW model. The highest score is reached for manufacturer C. For the comparison between older and newer CTs, Q2 and DIN metrics show a major improvement for manufacturer C. For manufacturer A, B and D with Q2 metric, no major improvement is noted. Nevertheless, even if some correlations could be found between

the Q2 or DIN methods and the model observer approach, those two methods require MTF measurements that are no more adequate when dealing with IR.

Abdomen/Pelvis protocol – HR

Mathematical model observer PW

As shown in figure 9, for the abdominal protocol chosen there is an enhancement of the detectability index (as with the head protocol) for three manufacturers when switching from the older to the newer CTs; for manufacturer A (8 %), for manufacturer C (84 %) and for manufacturer D (39 %). For manufacturer B, the detectability index remains constant within the error bars.

Comparing the newer CTs of the four manufacturers, CT "Cb" from manufacturer C reached the highest score and manufacturer A and D provide better results than manufacturer B.
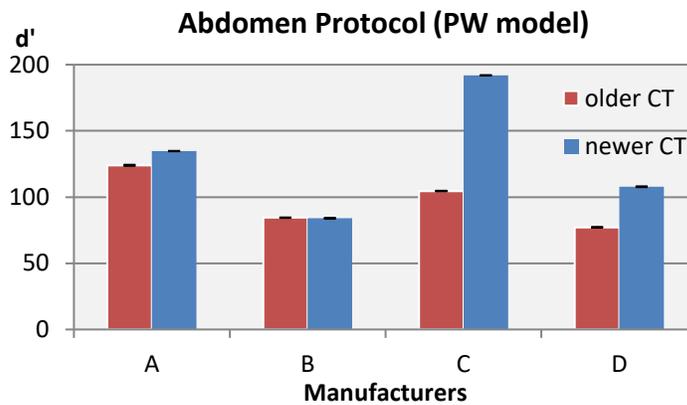
Figure 9: High contrast detectability for abdomen protocol using PW model

Alternative metrics
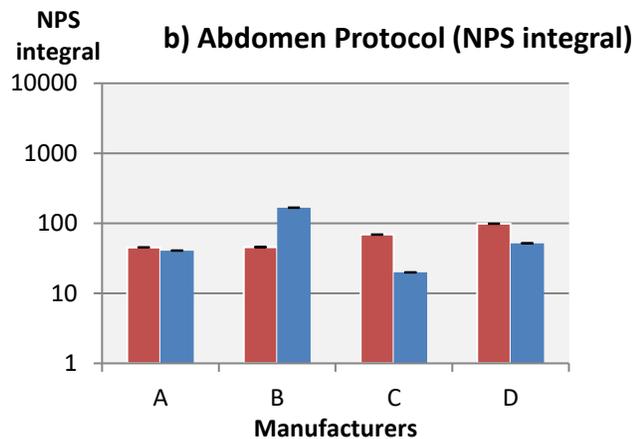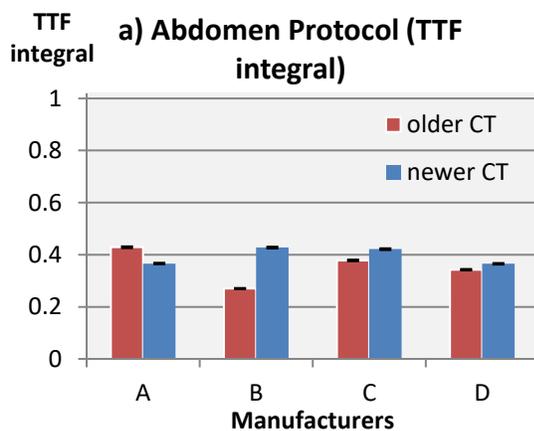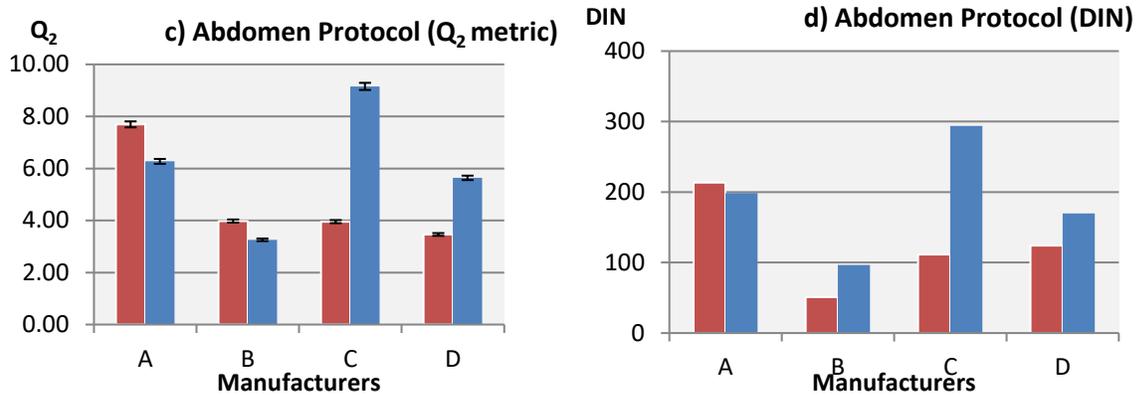
Figure 10: Alternative metrics for HR abdomen protocol: a) TTF's integral; b) NPS's integral; c) Q_2 metric; d) DIN

Concerning the alternative metrics, there is a slight improvement of the TTF integral and a noise reduction (decrease of the NPS integral) for manufacturer C and D. These variations explain also the improvement of the detectability index for these two manufacturers. For the comparison of newer CTs, the analysis of Q2 and DIN metrics provide similar results than the detectability index computed with the PW model.

3 – Chest protocol – HR

3.1 Mathematical model observer PW

For the chest protocol, there is an improvement of the detectability index for manufacturer C and D. As for the abdomen protocol, the detectability remains constant over the time for manufacturer B. However, there is a decrease (36 %) for manufacturer A between the older and newer CTs.
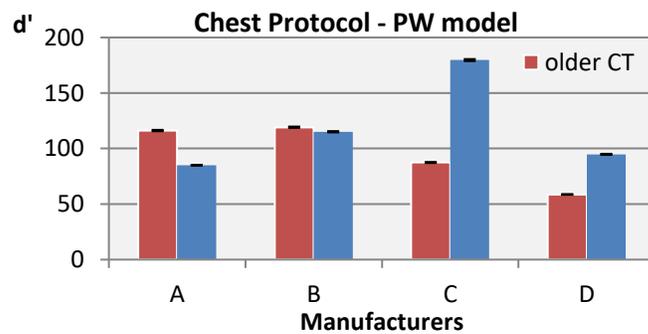
Figure 11: High contrast detectability for chest protocol using PW model

3.2 Alternative metrics

The major increase of the TTF integral for manufacturer C does not reflect only a better spatial resolution between the older and the newer CTs, because significant differences between the kernels exist and can explain the difference between older and newer CTs. An unexpected increase of the NPS integral can be noted for manufacturer A and B compared to manufacturer C and D.
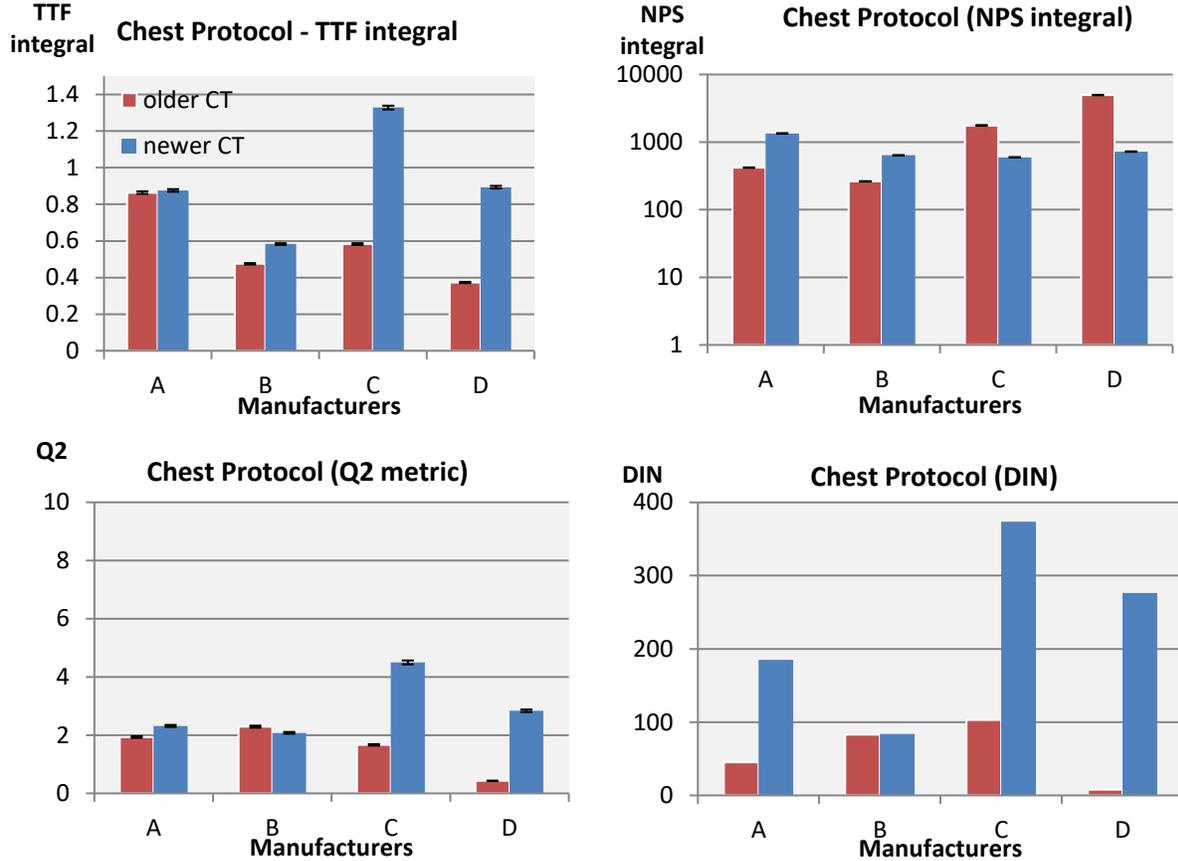
Figure 12: Alternative metrics for HR chest protocol

4- Head protocol – LCD with fixed dose levels
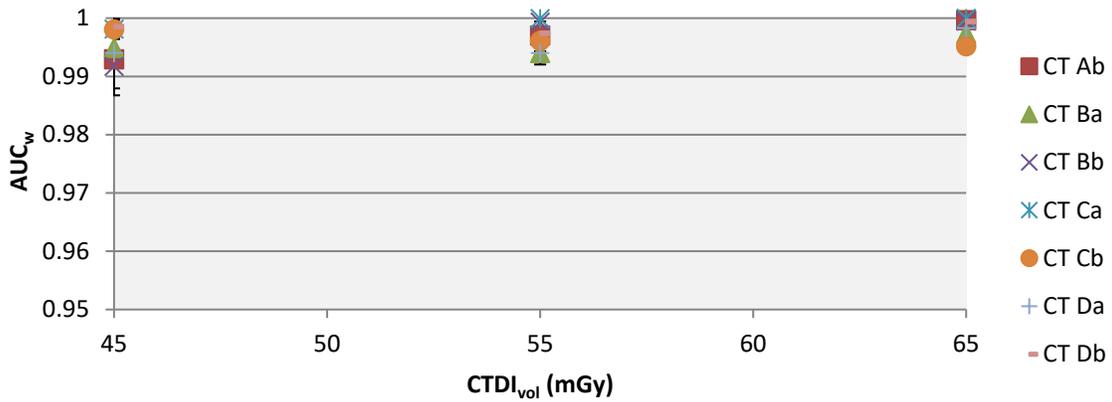
Figure 13: AUC$_w$ as a function of the dose for the assessment of low contrast detectability for Head protocol, with uncertainties (k=2). Note that the AUCw is systematically almost equal to 1.0.

The figure 13, presented above, summarizes the outcomes of the image quality assessment. It appears that "CT Aa" has the lowest performance. It is of note that the most discriminating task is the detection of a rod of 2mm in diameter with a contrast with background of 9HU. For such a task, spatial resolution performances certainly do matter and explain the result obtained. This limitation in spatial resolution has been now improved by the manufacturer. The uncertainty of these measurements, estimated by the bootstrap method (500 iterations), is in the range of ± 0.02. According to the state of the knowledge in the field of model observers it appears that this level of uncertainty might underestimate the actual uncertainty but no alternative method is proposed at the moment. The bootstrap method creates a "new population" by drawing samples with replacement from the initial population (resampling) [19].

5- Abdomen protocol with fixed dose levels (small phantom size)

Figure 14 presented below summarizes the results obtained with the small abdomen phantom using fixed CTDI$_{vol}$ values. With a CTDI$_{vol}$ of 15 mGy no major difference among the various scanners appears. Reducing the CTDI$_{vol}$ to 10 mGy, the image quality metrics slightly decreased for all scanners (AUC$_w$ going from 1.0 to 0.985), with a larger reduction observed for scanner "Db" (AUC$_w$ going from 1.0 to 0.945). These variations are statistically significant since the uncertainty of this evaluation is 0.003 (p < 0.05). At the lowest CTDI$_{vol}$ (5 mGy), all newer scanners provided better results than the older ones except for "CT Ab".
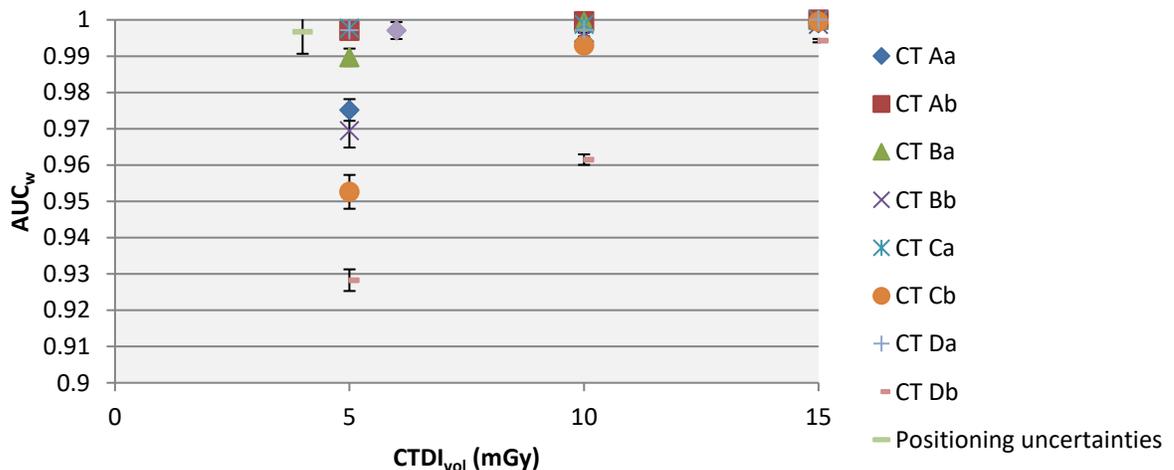
Figure 14: AUC$_w$ as a function of the dose for the assessment of low contrast detectability for abdomen protocol with the small phantom size (three *CTDI*$_{vol}$ levels 5, 10, 15 mGy), with uncertainties (k=2)

As a DEI one could consider the display of the integral of the curve representing the image quality metrics with a fixed range of dose as shown in Fig. 15. The unit would be in mGy. It should be mentioned that the results shown in Fig. 15 is an outcome of Fig. 14 (integral).
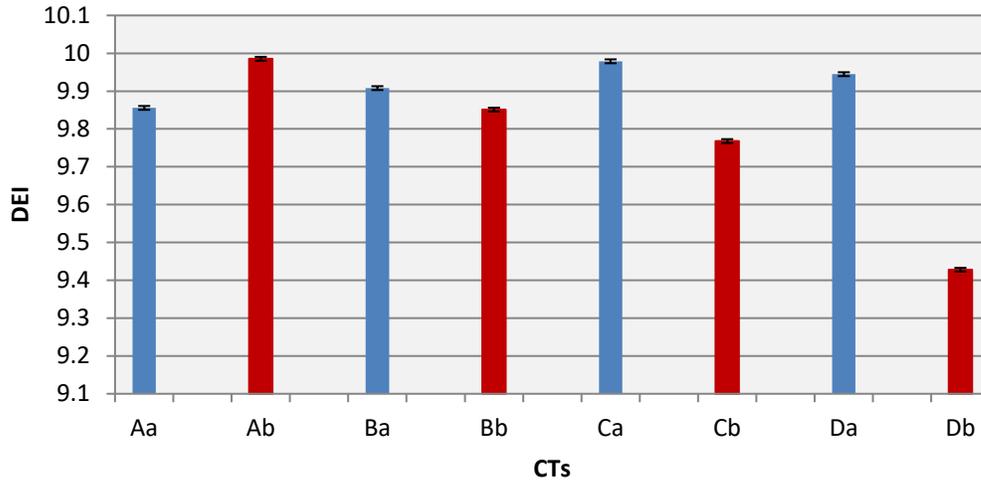
Figure 15: The integral of image quality as a function of the dose could be taken as a Dose Efficient Index (DEI)

Figure 15 shows that it is possible to separate a less efficient system (Db: one of the oldest systems) from average systems. It is of note that this feasibility test has been performed using the smallest abdominal phantom. A better discrimination between newer and older CTs would appear if one would use larger phantom sizes as shown later in this report.

5 - Abdomen protocol with automatic tube current modulation

As shown below in Fig. 16 and 17, using automatic tube current modulation on the small and medium abdominal phantoms, it is possible to reach a similar level of image quality for all scanners (differences within 5%). However, this high level of image quality is obtained at different $CTDI_{vol}$ values. Finally, when using the largest anthropomorphic abdominal phantom (see Fig. 18), large differences in behaviors are observed in terms of dose and image quality.
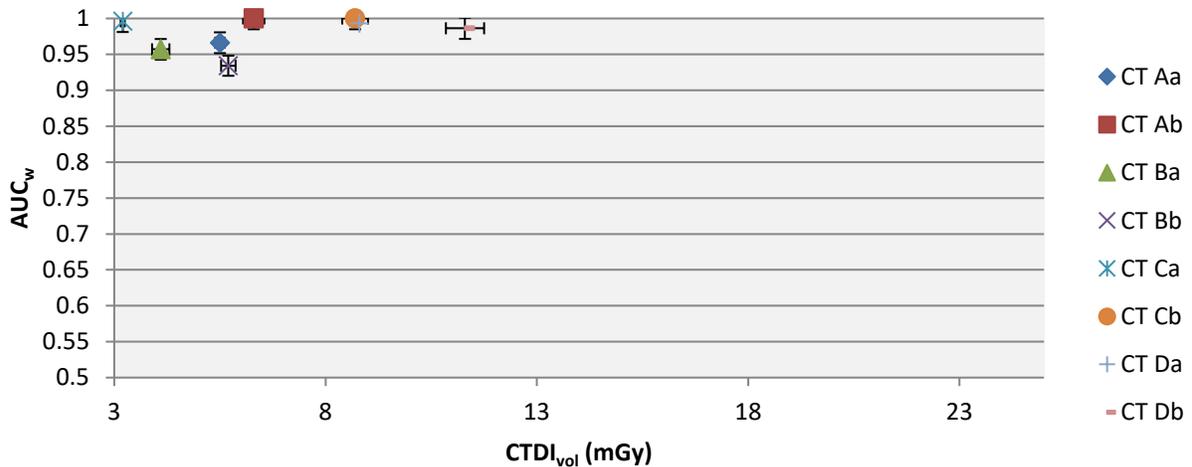
Figure 16: AUC_w for each CT using abdominal protocol with automatic tube current modulation for the small phantom
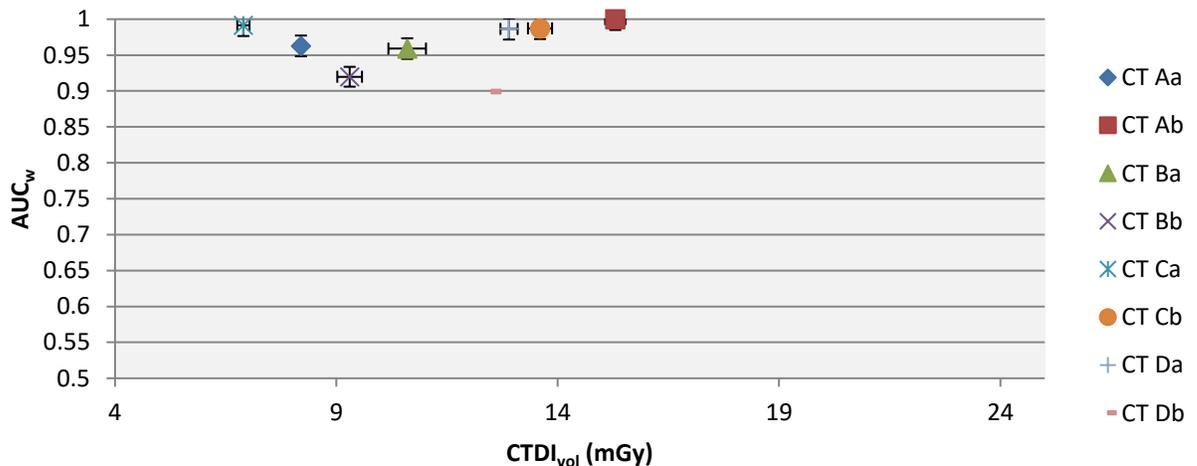
Figure 17: AUCw for each CT using abdominal protocol with automatic tube current modulation for the medium phantom
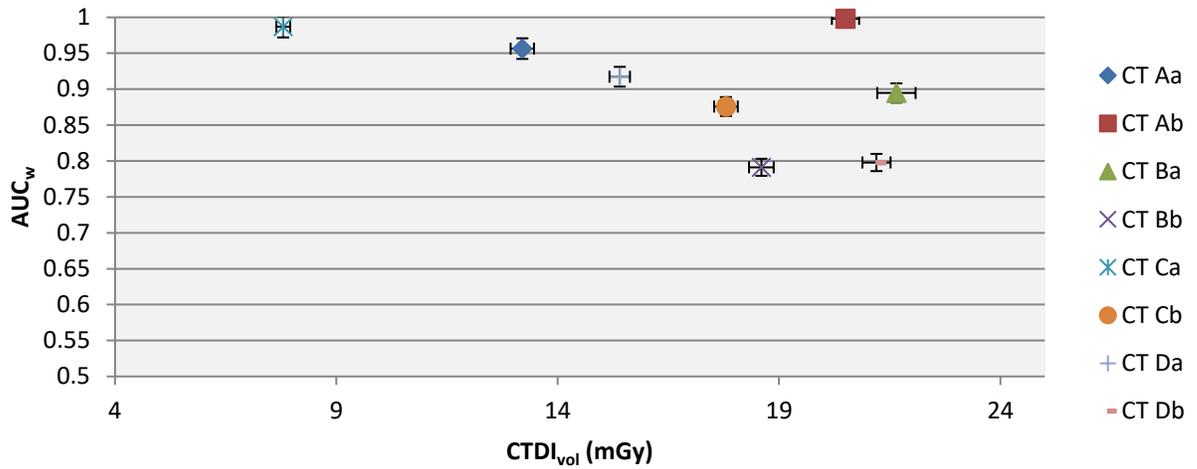
Figure 18: AUCw for each CT using abdominal protocol with automatic tube current modulation for the large phantom

6 - Pediatric head protocol (1 – 4 year old) – High level of LCD

As shown below in Fig.19, all CT units offer similar performances for this criterion; the differences being within the uncertainties of the evaluation. The use of small phantoms requires a drastic dose reduction if one is willing to show differences between units.
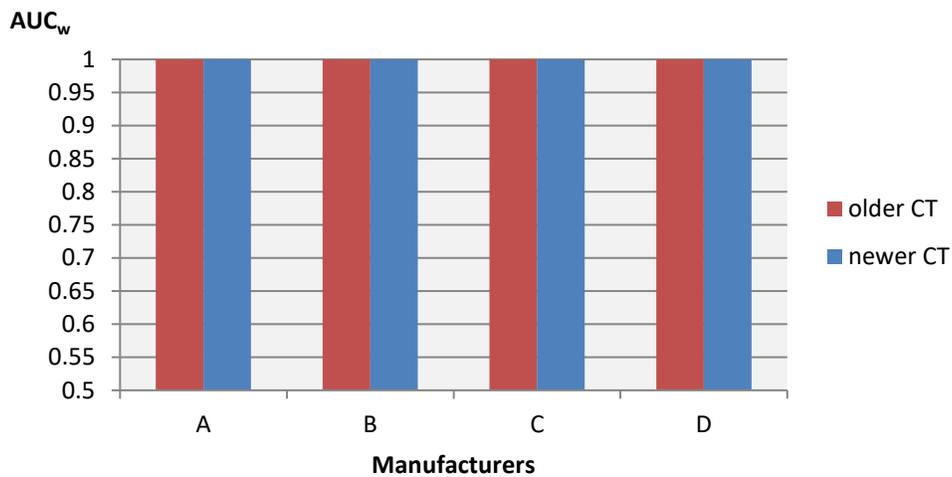
Figure 19: AUCw for each CT for pediatric head protocol at 1mGy

WP3.2/WP4 – Application of concept 2 on 48 conditions

As mentioned previously, two model observers have been used to get results for concept 2. The goal of the project being to propose a way to benchmark clinical protocol, we will focus on the NPWE model observer and the CHO (using 10 D-DoG channels) that are considered as an anthropomorphic model observer.

As shown below in Fig. 20, 21 and 22, for high contrast protocols (head, chest and abdomen protocols), the use of an anthropomorphic model observer (NPWE) leads to lower performances compared to ideal model observer. Nevertheless, similar trends between older and newer CTs can be observed for each manufacturer.

1 – Head protocol – HR
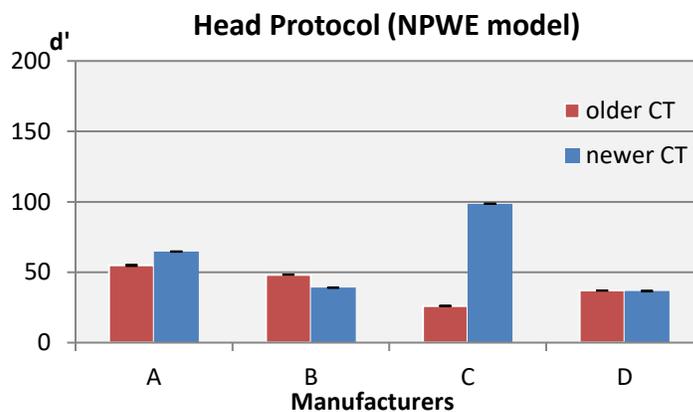
Figure 20: High contrast detectability for head protocol using NPWE model

In comparison to Figure 7, the introduction of the eye filter changed the performance; the noise component is largely affected by the eye filtering.

2. Abdomen/Pelvis protocol – HR
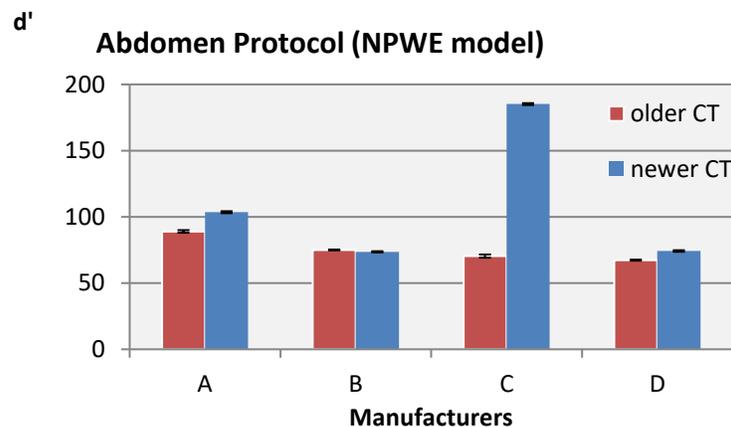
Figure 21: High contrast detectability for abdomen protocol using NPWE model

3 - Chest protocol - HR

Chest Protocol (NPWE model)
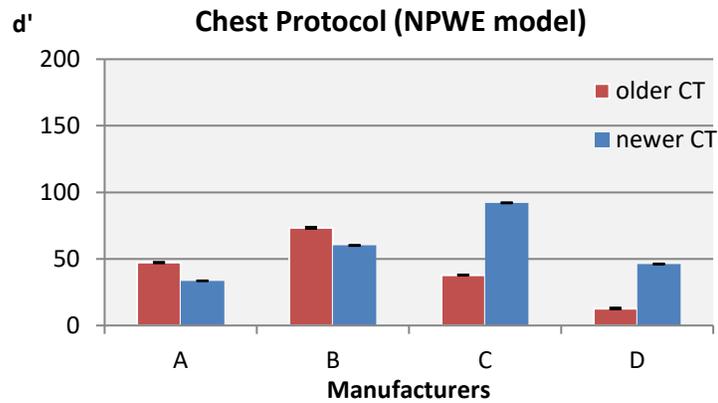
Figure 22: High contrast detectability for chest protocol using NPWE model

Similar statements can be made concerning the results presented below from figure 23 to figure 27 for low contrast protocols (head and abdomen protocol), using an anthropomorphic model observer (CHO with DDoG channels).

4 - Head protocol – LCD with fixed dose levels
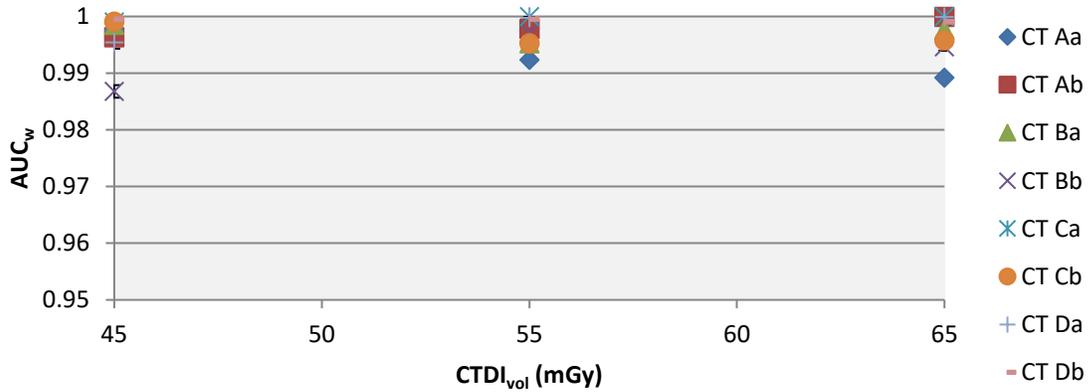
Figure 23: AUC$_w$ as a function of the dose for the assessment of low contrast detectability for Head protocol (CHO model with anthropomorphic DDOG channels)

5- Abdomen/Pelvis protocol with fixed dose levels
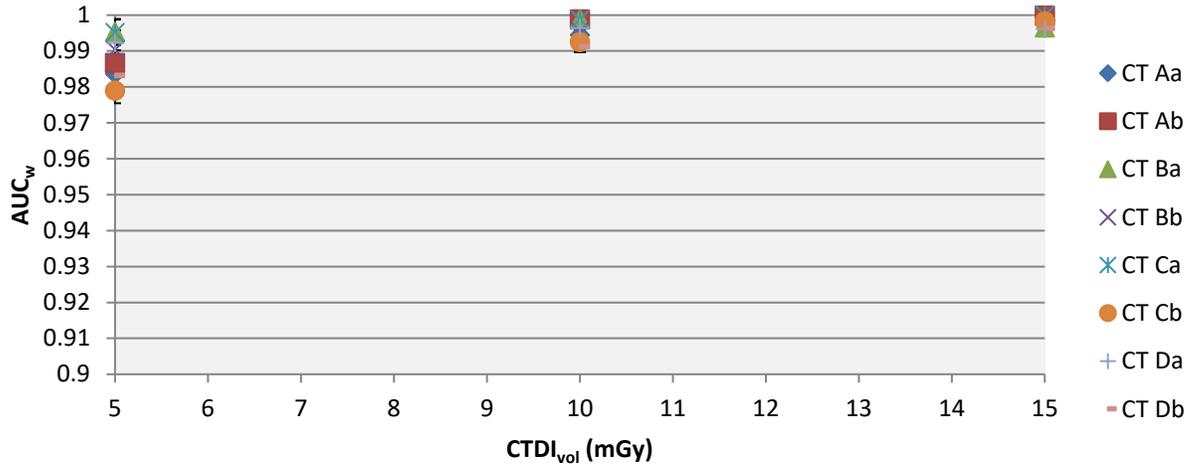
Figure 24 : AUC$_w$ as a function of the dose for the assessment of low contrast detectability for abdomen protocol with the small phantom size (three CTDI$_{vol}$ levels 5, 10, 15 mGy). (CHO model with anthropomorphic DDOG channels)

## 5- Abdomen/Pelvis protocol with automatic tube current modulation
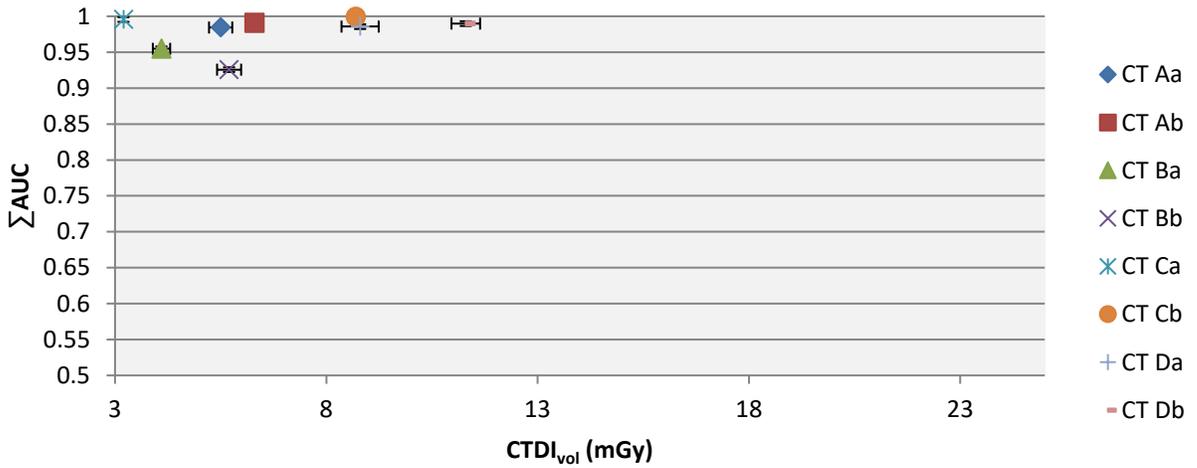
Figure 25 : AUC$_w$ for each CT using abdominal protocol with automatic tube current modulation for the small phantom
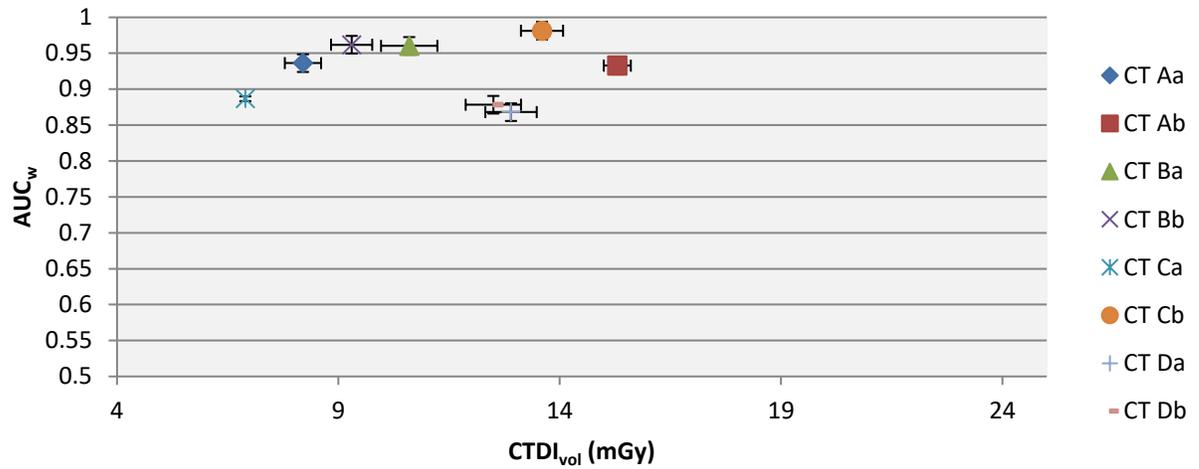
Figure 26: AUC_w for each CT using abdominal protocol with automatic tube current modulation for the medium phantom
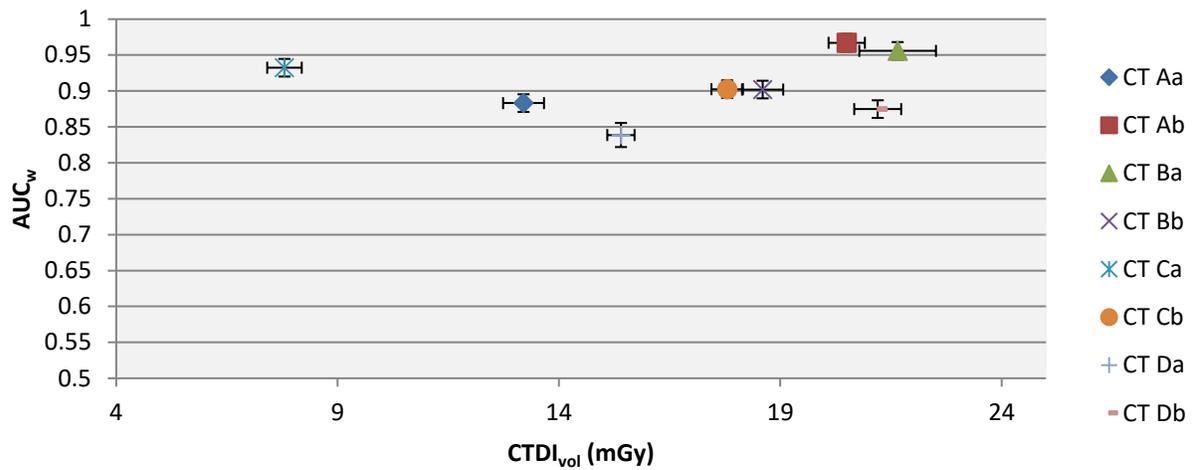

Figure 27: AUC_w for each CT using abdominal protocol with automatic tube current modulation for the large phantom

6 - Paediatric head protocol (1 – 4 year old) – High level of LCD

For this protocol the same conclusion can be drawn as for the use of the ideal model observer.
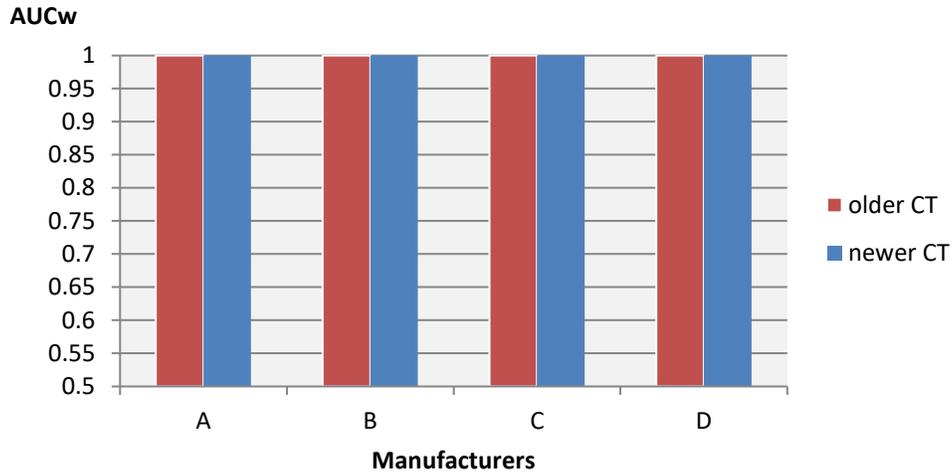
Figure 28: AUC_w for each CT for pediatric head protocol at 1 mGy

WP3.3 – Application of the concept on five different CT's of "CT Da"

In order to test the variability of the outcome of the concept proposed when changing a CT unit of the same kind, one protocol (protocol 4) has been applied on five different "CT Da" units. The results are summarized in the table 2 which demonstrates the robustness of the methodology investigated.

Table 3: AUC values for each lesion sizes and three dose levels for 5 different « CT Da » units

	AUC from CHO DDoG					
	5mGy		10mGy		15mGy	
	5mm/20HU	8mm/20HU	5mm/20HU	8mm/20HU	5mm/20HU	8mm/20HU
CT 1	0.91	1.00	0.98	1.00	0.99	1.00
CT 2	0.95	1.00	0.99	1.00	1.00	1.00
CT 3	0.96	1.00	0.95	1.00	1.00	1.00
CT 4	0.92	1.00	0.99	1.00	0.99	1.00
CT 5	0.94	1.00	0.98	1.00	0.98	1.00

Summary of WP4 – Analysis of the methodology

We proposed a methodology to benchmark CT units and benchmark clinical protocols using mathematical model observers. Advantages and limitations of the methodology chosen (phantom, clinical task and protocol settings) are summarized in table 3.

Table 4: Summary of advantages and limitations of the methodology chosen (phantom, clinical task and protocol settings) to benchmark CT and to benchmark protocols

		Concept 1: CT benchmarking	Concept 2: protocols benchmarking
Head HR (I)	**TTF's phantom**	Not representative of the attenuation and size of adults' heads	The phantom is not adapted to assess head protocols, no beam hardening is produced
	Detection task	Detection of structures with different sizes and contrasts	An estimation task (shape discrimination, size estimation…) rather the simple detection of a structure could be more interesting for a HR protocol
	Protocol settings	To find comparable kernels between manufacturers is very challenging for high-pass kernels	Use of the kernel locally implemented
Abdomen HR (II)	**TTF's phantom**	Methodology proposed adequate	The phantom is adequate but has only one size
	Detection task	Detection of structures with different sizes and contrasts	An estimation task (shape discrimination, size estimation…) rather the simple detection of a structure could be more interesting for a HR protocol
	Protocol settings	Find comparable kernels for abdominal kernels between CTs	Using local kernels, using local settings including tube current modulation
Chest HR (III)	**TTF's phantom**	Methodology proposed adequate	The phantom is not adequate (no air compartment) and has only one size
	Detection task	Detection of structures with different sizes and contrasts	An estimation task (shape discrimination, size estimation…) rather the simple detection of a structure could be more interesting for a HR protocol. A 3D model observer including the scrolling through images could be interesting when dealing with the search of nodules

		Concept 1: CT benchmarking	Concept 2: protocols benchmarking
Head LCD (IV)	**Protocol settings**	Find comparable kernels	Using local kernels, using local settings including tube current modulation
	MITA Phantom	Adequate but not representative of the attenuation and size of adults' heads. It is difficult to assess the constancy of the contrast values when changing CT unit since the sizes of the targets are quite small.	No beam hardening is produced
	Detection task	The detection task is so easy that a localization task could be added for this phantom when dealing with clinically relevant dose levels	Detection task or localization task
	Protocol settings	Use of lower dose levels for a detection tasks	Use of local kernels
Abdomen LCD (V)	**QRM medium size**	Methodology proposed adequate	The use of only one phantom size is limited
	Detection task	Methodology proposed adequate	Methodology proposed adequate
	Protocol with fixed dose levels	As an example: find comparable kernels, using dose levels from 2 to 10 mGy (for medium phantom). As presented, one should use several dose levels	Not adapted with fixed dose levels
	QRM: 3 sizes	One medium or large size might be enough	Methodology proposed adequate
	Detection task	Methodology proposed adequate	Methodology proposed adequate (an estimation of the lesion size could be added)
	Protocol with ATMC	Not adapted with ATMC	Use of local settings of the ATMC
Pediatric head protocol LCD (VI)	**Modulus of QRM phantom**	Not representative for the clinical tasks The size of the anatomy imaged is so small that the discrimination of CT unit seems difficult in a realistic dose range	Not representative for the clinical tasks
	Detection task	Methodology proposed adequate	Methodology proposed adequate (an estimation of the lesion size could be added)
	Protocol settings	Using lower dose levels (< 1 mGy; very different from current DRL)	Methodology proposed adequate

Conclusion

The goal of the project was to propose a way to benchmark CT using a metric that could take into account both the image quality assessment and the dose aspect, to give a DEI (Dose Efficiency Index). We used an index of image quality (like CTDIvol for the dose exposition index); in this context a comparison between model observers' performances and human performances is not mandatory even if human performances stay the gold standard. Such a process has been performed in the past using image quality metrics defined in signal theory that were fully adequate when dealing with a linear shift invariant system. This property has been accepted when image reconstruction was performed with FBP but using iterative reconstruction algorithms, image quality properties are not linear and are highly non-invariant. In such a context, the dose efficiency of imaging system can only be determined by considering the objectives of medical imaging. Thus, task-based image quality metrics should be used with medically realistic acquisition protocols.

Radiologists have to perform many tasks in their daily practice and it is impossible to define metrics that could encompass all aspects of image quality. Among the simplest tasks one can define for example the detection of a simple structure in a homogeneous background. A more realistic situation would be the use of structured background such as normal anatomy, but the generalization and standardization becomes very complicated. Other tasks could be the search and detection of simple static or moving structures, the accuracy of the shape or size of the structures presented on the images. In the framework of this project the task-based image quality metrics was the detection of simple structures in homogeneous backgrounds.

To assess these task-based metrics, one should switch from the signal theory to decision theory using mathematical model observers that can try to retrieve the maximum of information available (concept I) or that can try to mimic what a human observer could detect (concept II). Model observers applied in the image and Fourier domains have been used. The main advantage of working in the image domain is that one can check the outcomes of a model observer with his/her eyes since this requires phantoms with actual targets. Both of these concepts have been taken with no major impact on the ranking obtained on the CT units investigated. Only a shift in performance level was noted.

One of the challenges of attempting to benchmark CT units is the definition of comparable image acquisition parameters. A consensus has been reached in the framework of this project to work in a comparable framework of $CTDI_{vol}$ values. However the challenge with the most advanced technological progresses concerning the automatic tube current modulation and reconstruction processes drastically vary from one manufacturer to another. This constitutes a limit on the possibility of actual inter-manufacturer comparisons.

The proposed DEI is the integral of the outcome of a model observer together with the $CTDI_{vol}$. This definition is quite interesting when dealing with the detection of low contrast structures in the abdominal or brain region. For the abdominal region it is important to vary not only the dose levels (for example at least three $CTDI_{vol}$ values) but also the size of the phantom since performances tend to drop in the low dose range and for large anatomy. The results presented in this work with an anthropomorphic phantom are quite encouraging to pursue in this direction. For the brain region it was decided to use the MITA phantom that manufacturers have agreed on to characterize the dose reduction potential of their iterative reconstruction processes. This phantom however is larger than a standard head, and does not have a ring that would simulate beam hardening produced by the cortical bone. Moreover the use of rods (that allow obtaining precise results with a limited number of acquisitions) does not take into account the slice sensitivity profile of the images. Nevertheless, the MITA phantom is fully adequate for its intended use: relative comparisons of image quality for a given CT unit, with the aim of assessing the performance of image reconstruction strategies while reducing the dose. One interesting aspect of this phantom is the presence of very small structures that are not only sensitive to noise but also to spatial resolution. Moreover, we did not take into account a textured background since the tools used to analyze this kind of background were not yet developed and ready in order to be implemented in clinical routine.

However, in the future, a more realistic phantom, with not only circular lesions in homogeneous background should be used to benchmark clinical CT protocols. The detection of structures when dealing with small detail of high contrast remains trivial. For such structures it would be more adequate to switch to a task that check if the shape and the size presented on the image correspond to those of the structures placed in the phantom.

To conclude, this study shows that a benchmark of CT scanners or clinical protocols in terms of image quality assessment and patient exposure become feasible with model observers using a task-based approach [21]. The use of the methodology presented should be promoted among medical physicists to strengthen the communication between radiographers and radiologists. It would be quite useful to propose not only a DRL framework but also a consensual image quality level when dealing with CT protocol optimization. Model observers could be used in quality assurance programs or even during the clinical auditing of centers.

References

[1] F.R. Verdun, D. Racine, J.G. Ott, M.J. Tapiovaara, P. Toroi, F.O. Bochud, W.J.H. Veldkamp, A. Schegerer, R.W. Bouwman, I Hernandez Giron, N.W. Marshall, S. Edyvean. Image quality in CT: from physical measurements to model observers. Phys Med. 2015 Dec;31(8):823-43. doi: 10.1016/j.ejmp.2015.08.007. Epub 2015 Oct 12. Review.

[2] Richard S, Husarik D B, Yadava G, Murphy S N and Samei E 2012 Towards task-based assessment of   CT performance: System and object MTF across different reconstruction algorithms *Med. Phys.* **39** 4115-22

[3] Brunner C C and Kyprianou I S 2013 Material-specific transfer function model and SNR in CT *Phys. Med. Biol.* **58** 7447-61

[4] Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO and Verdun FR An update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. *Phys. Med. Biol.* 2014; 59:4047-4064

[5] Richard S, Li X, Yadava G and Samei E 2011 Predictive models for observer performance in CT: applications in protocol optimization *Proc. SPIE* **7961**

[6] Beutel J, Kundel HL, Van Metter RL. Handbook of Medical Imaging: Physics and psychophysics. SPIE Press; 2000

[7] Samei E, Krupinski E. The Handbook of Medical Image Perception and Techniques. Cambridge University Press; 2014.

[8] Barrett HH, Myers KJ. Foundations of image science. Wiley-Interscience; 2004.

[9] Vennart W. ICRU Report 54: Medical imaging—the assessment of image quality. Radiography 1996;3:243–4. doi:10.1016/S1078-8174(97)90038-

[10] Wunderlich A, Abbey CK. Utility as a rationale for choosing observer performance assessment paradigms for detection tasks in medical imaging. Med Phys 2013;40:111903. doi:10.1118/1.4823755

[11] Zhang Y, Leng S, Yu L, Carter RE, McCollough CH. Correlation between human and model observer performance for discrimination task in CT. Phys Med Biol 2014;59:3389–404. doi:10.1088/0031-9155/59/13/3389.

[12] Hernandez-Giron I, Geleijns J, Calzado A, Veldkamp WJH. Automated assessment of low contrast sensitivity for CT systems using a model observer. Med Phys 2011;38:S25–35. doi:10.1118/1.3577757.

[13] Yu L, Leng S, Chen L, Kofler JM, Carter RE, McCollough CH. Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: Impact of radiation dose and reconstruction algorithms. Med Phys 2013;40:041908. doi:10.1118/1.4794498.

[14] Racine D, Ba AH, Ott JG, Bochud FO, Verdun FR. Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer. Phys Med. 2016 Jan;32(1):76-83. doi: 10.1016/j.ejmp

[15] Abbey CK, Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. J Opt Soc Am 2001;18:473–88.

 [16] Evaluation of internal noise methods for Hotelling observer models Yani Zhang Binh T. Pham and Miguel P. Eckstein

[17] Damien Racine, Nick Ryckx, Alexandre Ba, Julien G. Ott, François O. Bochud, Francis R. Verdun; BENCHMARKING OF CT FOR PATIENT EXPOSURE OPTIMISATION. Radiat Prot Dosimetry 2016; 169 (1-4): 78-83. doi: 10.1093/rpd/ncw021

[18] Edyvean S. The relationship between image noise and spatial resolution of CT scanners. http://impactscan.org/slides/ukrc2002/noiseres/index.htm (accessed March 11, 2015).

[19] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. CRC Press; 1994.

[20] Abbey CK, Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. J Opt Soc Am 2001;18:473–88.

[21] Racine D and Viry A, Becce F, Schmidt S, Ba AH, Bochud FO, Edyvean S, Schegerer A, Verdun FR. Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners. Accepted to Medical Physics in January 2017.

Annexe 1

1) **Head protocol - High contrast detection (HR)**
- Typical indication: trauma
- High voltage: 120 kV;
- Bone filter
- Reconstructed slice thickness: 2.5/3 mm
- No tube current modulation
- Helical pitch close to 1.0
- Reconstructed FOV: close to 25 cm

2) **Abdomen/Pelvis protocol – High contrast detection (HR)**
- Typical indication: search for renal stones
- High voltage: 120 kV;
- Standard body filter
- Reconstructed slice thickness: 2.5/3 mm
- No Tube current modulation
- Helical pitch close to 1.4
- Reconstructed FOV: close to 25 cm

3) **Chest protocol - High contrast detection (HR)**
- Typical indication: chest diseases
- High voltage: 120 kV;
- Lung
- Reconstructed slice thickness: 2.5/3 mm
- No Tube current modulation
- Helical pitch close to 1.4
- Reconstructed FOV: close to 25 cm

4) **Head protocol – Low contrast detection (LCD)**
- Typical indication: soft tissue pathologies
- High voltage: 120 kV;
- Standard head filter
- Reconstructed slice thickness: 2.5/3 mm
- No tube current modulation
- Axial or helical pitch close to 1.0
- Reconstructed FOV: close to 25 cm

5) **Abdomen/Pelvis protocol – Low contrast detectability for fixed dose levels and automatic tube current modulation**
- Typical indication: search for diverticulis or appendicitis
- High voltage: 120 kV;
- Standard body filter
- Reconstructed slice thickness: 2.5/3 mm
- Fixed dose levels and automatic tube current modulation
- Helical pitch close to 1.3
- Reconstructed FOV: close to 32, 37 and 42 cm

6) **Pediatric head protocol (1 – 4 year old) - High level of Low contrast detection (LCD)**
- Typical indication: soft tissue pathologies
- High voltage: 100 kV;
- Standard head filter
- Reconstructed slice thickness: 2.5/3 mm
- No tube current modulation
- Helical pitch close to 1.0
- Reconstructed FOV: close to 16 cm

Depending on the CT unit, FBP or iterative reconstruction are used

Annexe 2

Concept 1 : Ideal model observers : Benchmark of CT units

1- Head protocol HR

	PW	SD	TTF	SD	NPS	SD	Q2	SD	DIN	SD
CT Aa	107.5	0.14	0.75	0.009	557.01	5.29	3.03	0.045	89.8	1.08
CT Ab	121.5	0.43	0.87	0.010	495.04	4.70	3.12	0.047	121.8	1.46
CT Ba	86.8	0.18	0.62	0.007	739.80	7.03	1.99	0.030	87.6	1.05
CT Bb	100.3	0.16	0.63	0.008	537.58	5.11	2.25	0.034	49.1	0.59
CT Ca	191.9	0.15	1.01	0.012	268.83	2.55	7.90	0.118	522.7	6.27
CT Cb	66.9	0.21	0.89	0.011	6832.85	64.91	0.33	0.005	54.1	0.65
CT Da	110.9	0.30	1.04	0.012	2133.27	20.27	1.80	0.027	160.1	1.92
CT Db	77.0	0.10	0.63	0.008	1568.72	14.90	1.62	0.024	51.6	0.62

2- Abdomen protocol HR

	PW	SD	TTF	SD	NPS	SD	Q2	SD	DIN	SD
CT Aa	132.9	103.51	0.366	0.004	40.80	0.39	6.27	0.09	198.35	2.38
CT Ab	147.3	88.98	0.429	0.005	45.32	0.43	7.69	0.12	213.53	2.56
CT Ba	100.1	73.60	0.427	0.005	167.11	1.59	3.26	0.05	96.51	1.16
CT Bb	102.4	74.98	0.270	0.003	45.70	0.43	3.98	0.06	50.91	0.61
CT Ca	221.7	185.14	0.421	0.005	19.91	0.19	9.15	0.14	293.60	3.52
CT Cb	123.7	70.35	0.378	0.005	68.98	0.66	3.96	0.06	111.34	1.34
CT Da	126.5	74.22	0.365	0.004	51.85	0.49	5.64	0.08	169.65	2.04
CT Db	88.2	67.30	0.342	0.004	98.71	0.94	3.46	0.05	124.01	1.49

3- Chest protocol HR

	PW	SD	TTF	SD	NPS	SD	Q2	SD	DIN	SD
CT Aa	83.42	0.35	0.88	0.01	1344.4	12.77	0.81	0.12	185.02	2.22
CT Ab	137.64	0.60	0.86	0.01	419.0	3.98	1.10	0.04	45.28	0.54
CT Ba	96.68	0.58	0.58	0.01	638.0	6.06	1.09	0.06	84.09	1.01
CT Bb	99.46	0.60	0.48	0.01	262.7	2.50	1.13	0.02	83.19	1.00
CT Ca	156.59	0.90	1.33	0.02	596.0	5.66	1.71	0.05	373.55	4.48
CT Cb	102.65	0.44	0.87	0.01	1765.2	16.77	0.83	0.16	102.96	1.24
CT Da	111.30	0.47	0.89	0.01	723.8	6.88	0.90	0.07	276.36	3.32
CT Db	66.56	0.29	0.37	0.004	4965.9	47.18	0.56	0.45	7.78	0.09

4 - Head protocol –LCD

45 mGy

	2 mm 9 HU		3 mm 5HU		4 mm 4 HU		5mm 3HU	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CT Aa	0.73	0.01	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	0.97	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Bb	0.97	0.00	0.99	0.00	1.00	0.00	1.00	0.00
CT Ca	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	0.97	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00

55 mGy

	2 mm 9 HU		3 mm 5HU		4 mm 4 HU		5mm 3HU	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CT Aa	0.95	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	0.98	0.00	0.99	0.00	1.00	0.00	1.00	0.00
CT Bb	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	0.97	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00

65 mGy

	2 mm 9 HU		3 mm 5HU		4 mm 4 HU		5mm 3HU	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CT Aa	0.91	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Bb	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

5- Abdomen protocol with fixed dose levels (small phantom size)

5 mGy

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU	
	Mean	SD	Mean	SD	Mean	SD
CT Aa	1.00	0.00	1.00	0.00	0.90	0.01
CT Ab	1.00	0.00	1.00	0.00	0.99	0.00
CT Ba	1.00	0.00	1.00	0.00	0.96	0.01
CT Bb	1.00	0.00	0.97	0.01	0.91	0.01
CT Ca	1.00	0.00	1.00	0.00	0.99	0.00
CT Cb	1.00	0.00	0.97	0.01	0.85	0.01
CT Da	1.00	0.00	1.00	0.00	0.99	0.00
CT Db	1.00	0.00	0.96	0.01	0.77	0.02

10 mGy

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU	
	Mean	SD	Mean	SD	Mean	SD
CT Aa	1.00	0.00	1.00	0.00	0.98	0.01
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	1.00	0.00	1.00	0.00	1.00	0.00
CT Bb	1.00	0.00	1.00	0.00	0.98	0.00
CT Ca	1.00	0.00	1.00	0.00	0.99	0.00
CT Cb	1.00	0.00			0.97	0.01
CT Da	1.00	0.00	1.00	0.00	0.99	0.00
CT Db	1.00	0.00	0.97	0.01	0.89	0.01

15 mGy

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU	
	Mean	SD	Mean	SD	Mean	SD
CT Aa	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	1.00	0.00	1.00	0.00	1.00	0.00
CT Bb	1.00	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	1.00	0.00	1.00	0.00	0.98	0.00

5- Abdomen protocol with automatic tube current modulation

Small phantom

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU		Dose (mGy)
	Mean	SD	Mean	SD	Mean	SD	
CT Aa	1.00	0.00	0.98	0.00	0.93	0.01	5.50
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00	6.30
CT Ba	1.00	0.00	0.98	0.00	0.91	0.01	4.10
CT Bb	1.00	0.00	0.97	0.01	0.86	0.01	5.70
CT Ca	1.00	0.00	1.00	0.00	0.99	0.00	3.20
CT Cb	1.00	0.00	1.00	0.00	1.00	0.00	8.70
CT Da	1.00	0.00	1.00	0.00	0.98	0.00	8.80
CT Db	1.00	0.00	1.00	0.00	0.97	0.01	11.30

Medium phantom

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU		Dose (mGy)
	Mean	SD	Mean	SD	Mean	SD	
CT Aa	1.00	0.00	0.97	0.01	0.94	0.01	8.20
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00	15.30
CT Ba	1.00	0.00	0.97	0.01	0.92	0.01	10.60
CT Bb	1.00	0.00	0.92	0.01	0.87	0.01	9.30
CT Ca	1.00	0.00	1.00	0.00	0.98	0.00	6.90
CT Cb	1.00	0.00	1.00	0.00	0.97	0.01	13.60
CT Da	1.00	0.00	0.99	0.00	0.97	0.01	12.90
CT Db	1.00	0.00	0.97	0.01	0.78	0.01	12.50

Large phantom

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU		Dose (mGy)
	Mean	SD	Mean	SD	Mean	SD	
CT Aa	1.00	0.00	0.96	0.01	0.94	0.01	13.20
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00	20.50
CT Ba	0.99	0.00	0.92	0.01	0.82	0.01	21.65
CT Bb	0.98	0.01	0.87	0.01	0.61	0.02	18.60
CT Ca	1.00	0.00	1.00	0.00	0.97	0.01	7.80
CT Cb	1.00	0.00	0.96	0.01	0.73	0.02	17.80
CT Da	1.00	0.00	0.94	0.01	0.85	0.02	15.40
CT Db	1.00	0.00	0.89	0.01	0.60	0.02	21.20

6 - Pediatric head protocol (1 – 4 year old) – High level of LCD
1 mGy

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU	
	Mean	SD	Mean	SD	Mean	SD
CT Aa	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	1.00	0.00	1.00	0.00	1.00	0.00
CT Bb	1.00	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	1.00	0.00	1.00	0.00	1.00	0.00

Concept 2 : Anthropomorphic model observers Benchmark of clinical protocols

1- **Head protocol HR**

	NPWE (mean)	SD
CT Aa	64.68	0.26
CT Ab	54.86	0.63
CT Ba	39.03	0.29
CT Bb	48.26	0.49
CT Ca	98.65	0.42
CT Cb	26.06	0.07
CT Da	36.64	0.28
CT Db	36.98	0.19

2- **Abdomen protocol HR**

	NPWE (mean)	SD
CT Aa	103.51	0.725
CT Ab	88.98	0.961
CT Ba	73.60	0.434
CT Bb	74.98	0.187
CT Ca	185.14	0.704
CT Cb	70.35	1.196
CT Da	74.22	0.564
CT Db	67.30	0.357

3- **Chest protocol HR**

	NPWE (mean)	SD
CT Aa	33.55	0.27
CT Ab	47.26	0.59
CT Ba	60.15	0.30
CT Bb	73.37	0.61
CT Ca	92.14	0.31
CT Cb	37.90	0.74
CT Da	46.17	0.41
CT Db	12.78	0.26

4- Head protocol LCD

45 mGy

	2 mm 9 HU		3 mm 5HU		4 mm 4 HU		5mm 3HU	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CT Aa	0.76	0.01	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Bb	0.95	0.01	0.99	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

55 mGy

	2 mm 9 HU		3 mm 5HU		4 mm 4 HU		5mm 3HU	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CT Aa	0.97	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	0.99	0.00	0.99	0.00	1.00	0.00	1.00	0.00
CT Bb	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

65 mGy

	2 mm 9 HU		3 mm 5HU		4 mm 4 HU		5mm 3HU	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CT Aa	0.96	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	0.99	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Bb	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

5- Abdomen protocol with fixed dose levels (small phantom size)

5 mGy

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU	
	Mean	SD	Mean	SD	Mean	SD
CT Aa	1.00	0.00	0.99	0.00	0.96	0.01
CT Ab	1.00	0.00	0.99	0.00	0.97	0.00
CT Ba	1.00	0.00	0.99	0.00	0.97	0.00
CT Bb	1.00	0.00	0.96	0.01	0.96	0.01
CT Ca	1.00	0.00	1.00	0.00	0.96	0.01
CT Cb	1.00	0.00	0.99	0.00	0.86	0.01
CT Da	1.00	0.00	0.99	0.00	0.95	0.01
CT Db	1.00	0.00	0.99	0.00	0.89	0.02

10 mGy

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU	
	Mean	SD	Mean	SD	Mean	SD
CT Aa	1.00	0.00	1.00	0.00	0.99	0.00
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	1.00	0.00	1.00	0.00	0.99	0.00
CT Bb	1.00	0.00	1.00	0.00	0.97	0.01
CT Ca	1.00	0.00	1.00	0.00	0.99	0.00
CT Cb	1.00	0.00	1.00	0.00	0.94	0.01
CT Da	1.00	0.00	1.00	0.00	0.97	0.00
CT Db	1.00	0.00	0.98	0.01	0.95	0.01

15 mGy

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU	
	Mean	SD	Mean	SD	Mean	SD
CT Aa	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	1.00	0.00	1.00	0.00	0.97	0.01
CT Bb	1.00	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	1.00	0.00	1.00	0.00	0.99	0.00
CT Da	1.00	0.00	1.00	0.00	0.97	0.00
CT Db	1.00	0.00	1.00	0.00	0.97	0.01

5- Abdomen protocol with automatic tube current modulation

Small phantom

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU		Dose (mGy)
	Mean	SD	Mean	SD	Mean	SD	
CT Aa	1.00	0.00	0.99	0.00	0.97	0.01	5.50
CT Ab	1.00	0.00	1.00	0.00	0.98	0.01	6.30
CT Ba	1.00	0.00	0.98	0.01	0.90	0.01	4.10
CT Bb	1.00	0.00	0.96	0.01	0.86	0.01	5.70
CT Ca	1.00	0.00	1.00	0.00	0.99	0.00	3.20
CT Cb	1.00	0.00	1.00	0.00	1.00	0.00	8.70
CT Da	1.00	0.00	0.99	0.00	0.97	0.01	8.80
CT Db	1.00	0.00	0.99	0.00	0.98	0.01	11.30

Medium phantom

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU		Dose (mGy)
	Mean	SD	Mean	SD	Mean	SD	
CT Aa	0.99	0.00	0.95	0.01	0.89	0.01	8.20
CT Ab	1.00	0.00	0.98	0.00	0.85	0.01	15.30
CT Ba	1.00	0.00	0.97	0.01	0.93	0.01	10.60
CT Bb	1.00	0.00	0.96	0.01	0.94	0.01	9.30
CT Ca	1.00	0.00	0.91	0.01	0.80	0.01	6.90
CT Cb	0.99	0.00	0.99	0.00	0.97	0.01	13.60
CT Da	0.98	0.00	0.90	0.02	0.77	0.02	12.90
CT Db	0.97	0.01	0.88	0.01	0.82	0.02	12.50

Large phantom

	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU		Dose (mGy)
	Mean	SD	Mean	SD	Mean	SD	
CT Aa	0.97	0.01	0.92	0.01	0.80	0.02	8.7
CT Ab	1.00	0.00	0.96	0.01	0.95	0.01	20.50
CT Ba	0.97	0.01	0.95	0.01	0.95	0.01	21.65
CT Bb	0.98	0.01	0.92	0.01	0.83	0.01	18.60
CT Ca	0.98	0.01	0.97	0.01	0.87	0.01	7.80
CT Cb	0.99	0.00	0.97	0.01	0.79	0.01	17.80
CT Da	0.99	0.00	0.92	0.01	0.67	0.02	15.40
CT Db	0.99	0.00	0.89	0.01	0.79	0.01	21.20

6- Pediatric protocol : Head LCD

1 mGy	8 mm 20 HU		6 mm 20 HU		5 mm 20 HU	
	Mean	SD	Mean	SD	Mean	SD
CT Aa	1.00	0.00	1.00	0.00	1.00	0.00
CT Ab	1.00	0.00	1.00	0.00	1.00	0.00
CT Ba	1.00	0.00	1.00	0.00	1.00	0.00
CT Bb	1.00	0.00	1.00	0.00	1.00	0.00
CT Ca	1.00	0.00	1.00	0.00	1.00	0.00
CT Cb	1.00	0.00	1.00	0.00	1.00	0.00
CT Da	1.00	0.00	1.00	0.00	1.00	0.00
CT Db	1.00	0.00	1.00	0.00	1.00	0.00